

Edition Informatik Spektrum

Edy Portmann
Sara D'Onofrio *Hrsg.*



Cognitive Computing

Theorie, Technik und Praxis

Informatik
Spektrum

EBOOK INSIDE



Springer Vieweg

Edition Informatik Spektrum

Hauptaufgabe der Fachzeitschrift „Informatik Spektrum“, einem Organ der Gesellschaft für Informatik e.V. und mit ihr assoziierter Organisationen, ist die Publikation wissenschaftlicher Ergebnisse und aktueller, praktisch verwertbarer Informationen über technische und wissenschaftliche Fortschritte aus allen Bereichen der Informatik und ihrer Anwendungen in Form von Übersichtsartikeln und einführenden Darstellungen sowie Berichten über Projekte und Fallstudien, die zukünftige Trends aufzeigen. Erhältlich ist die Zeitschrift im elektronischen Bezug über SpringerLink und Springer Professional sowie in gedruckter Form im Abonnement.

Die Buchreihe „Edition Informatik Spektrum“ greift aktuelle Schwerpunkt-Themen der Zeitschrift auf. Passende Fachbeiträge aus den Heften werden gebündelt, überarbeitet und durch neue zusätzliche Beiträge ergänzt. Somit bietet jedes Buch einen fundierten Überblick und Einblick in aktuelle technische und wissenschaftliche Entwicklungen.

Weitere Bände in der Reihe

<http://www.springer.com/series/16130>

Edy Portmann • Sara D'Onofrio
Hrsg.

Cognitive Computing

Theorie, Technik und Praxis



Springer Vieweg

Hrsg.
Edy Portmann
Human-IST Institut
Universität Fribourg
Fribourg, Schweiz

Sara D'Onofrio
Human-IST Institut
Universität Fribourg
Fribourg, Schweiz

Edition Informatik Spektrum
ISBN 978-3-658-27940-0 ISBN 978-3-658-27941-7 (eBook)
<https://doi.org/10.1007/978-3-658-27941-7>

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Springer Vieweg

© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2020

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von allgemein beschreibenden Bezeichnungen, Marken, Unternehmensnamen etc. in diesem Werk bedeutet nicht, dass diese frei durch jedermann benutzt werden dürfen. Die Berechtigung zur Benutzung unterliegt, auch ohne gesonderten Hinweis hierzu, den Regeln des Markenrechts. Die Rechte des jeweiligen Zeicheninhabers sind zu beachten.

Der Verlag, die Autoren und die Herausgeber gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag, noch die Autoren oder die Herausgeber übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen. Der Verlag bleibt im Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutionsadressen neutral.

Springer Vieweg ist ein Imprint der eingetragenen Gesellschaft Springer Fachmedien Wiesbaden GmbH und ist ein Teil von Springer Nature.

Die Anschrift der Gesellschaft ist: Abraham-Lincoln-Str. 46, 65189 Wiesbaden, Germany

Geleitwort

Die Buchreihe „Edition Informatik Spektrum“ behandelt aktuelle Themen der Informatik und ihrer Anwendungen. Ziel ist dabei die Einführung des wissenschaftlichen Hintergrunds und praktisch verwertbarer Anwendungen, die sich daraus ergeben. Die Buchreihe greift Schwerpunkt-Themen der Fachzeitschrift „Informatik Spektrum“ auf, einem Organ der Gesellschaft für Informatik e.V. und mit ihr assoziierter Organisationen. Hauptaufgabe der Zeitschrift ist die Publikation von wissenschaftlichen und technischen Fortschritten aus allen Bereichen der Informatik, ihrer Anwendungen und Wirkungen in Form von Übersichtsartikeln, einführenden Darstellungen sowie Berichten über Projekte und Fallstudien, die zukünftige Trends aufzeigen. Erhältlich ist die Zeitschrift im elektronischen Bezug über SpringerLink und Springer Professional sowie in gedruckter Form im Abonnement. Fachbeiträge aus der Zeitschrift werden in der Buchreihe gebündelt, überarbeitet und durch neue zusätzliche Beiträge ergänzt. Somit bietet jedes Buch einen fundierten Überblick und Einblick in aktuelle wissenschaftliche und technische Fragestellungen aus der sich rasant entwickelnden Informatik.

Cognitive Computing ist ein etablierter Begriff, der aber bisher ohne eine allgemein akzeptierte Definition geblieben ist. Er bezeichnet die Nachbildung kognitiver Fähigkeiten des Menschen in technischen Systemen und die Nutzung dieser Systeme in soziotechnischen Prozessen. Mithin umfasst er vieles, was auch durch Künstliche Intelligenz abgedeckt ist. Die Herausgeber und Autoren des vorliegenden Bandes zielen darauf ab, den Begriffsumfang von Cognitive Computing durch eine Reihe von Artikeln einzugrenzen und klarer zu bestimmen. Wichtiger ist es ihnen jedoch, das Potenzial des Cognitive Computings einer breiten Leserschaft nahezubringen. Ausgehend von allgemein wissenschaftlichen Betrachtungen werden spannende Anwendungsfälle des Cognitive Computings dargestellt.

Der vorliegende Band ist in vier Abschnitte gegliedert, die verschiedene Facetten des Cognitive Computings darstellen. Teil I befasst sich allgemein mit den Konzepten der Künstlichen Intelligenz und des menschlichen Gehirns mit dem Ziel, der Leserschaft einen Einordnungsrahmen für das eigene Verständnis zu geben. Drei Artikel strukturieren das Gebiet der Künstlichen Intelligenz und der Kognition auf einem hohen Abstraktionsniveau und ermöglichen somit, ein erstes Verständnis über die Thematik zu erlangen. Hier werden auch allgemeinere Aspekte wie zum Beispiel ethische Fragen und Aspekte der Ausbildung ausführlicher betrachtet. Das

Ziel ist eine Mensch-Maschine-Interaktion, bei der die Maschine die Menschen in verschiedenen Lebenszusammenhängen effektiv unterstützt.

Teil II wendet sich der maschinellen Kreativität und Problemlösungsfähigkeit zu. Technische Systeme haben Potenzial für Kreativität, jedoch sind diesem systematische Grenzen gesetzt. Zipp diskutiert die Möglichkeiten maschineller Kreativität und setzt sie in Beziehung zur menschlichen Kreativität. Gross hingegen diskutiert, wie mithilfe von kreativen Systemen die heutigen großen Probleme, beispielsweise des Klimawandels, gelöst werden können. Ein spannender Ausgangspunkt für die weiteren Betrachtungen in diesem Sammelband.

Teil III diskutiert die Methoden und Einsatzmöglichkeiten von Chatbots. Sie werden als zukunftsweisende Schnittstelle der Mensch-Maschine-Kommunikation angesehen und finden ihren Einsatz beispielsweise im Bereich von User Help Desks. Die Akzeptanz der Systeme ist eng mit einer menschlich anmutenden Interaktion verbunden. Die Autoren geben hierzu einen Einblick in ihre aktuelle Forschung und teilen ihre Erfahrungen.

Schließlich werden in Teil IV verschiedene Anwendungsszenarien des Cognitive Computings erläutert. Gemeinsam ist ihnen, dass technische Systeme Unterstützung bei Entscheidung liefern. Die einzelnen Anwendungsgebiete könnten aber unterschiedlicher nicht sein: Die Autoren betrachten die Wegeplanung bei Transportfahrzeugen, analysieren Gerichtsurteile und Rechtsquellen im juristischen Bereich, diskutieren das Skillmanagement im Personalbereich und geben Empfehlungen zum Gesundheitsmanagement im medizinischen Bereich.

Den Herausgebern gelingt es mit einer exzellenten Autorenschaft das Gebiet des Cognitive Computings in Theorie und Praxis abzudecken. Das Spektrum der Anwendungsbeispiele belegt eindrucksvoll die Durchdringung wichtiger Lebensbereiche durch Cognitive Computing.

Die Leser werden mitgenommen auf eine intellektuelle Reise durch die Gebiete der Künstlichen Intelligenz und der maschinellen Kreativität. Spannende Beispiele des Einsatzes des Cognitive Computings belegen die digitale Transformation weiter Teile unseres täglichen Lebens. Wir empfehlen das Buch all jenen zur Lektüre, die heute erkennen wollen, wie wir morgen leben werden.

Arndt Bode und Thomas Ludwig
Hauptherausgeber Informatik Spektrum

Cognitive Computing – Bedeutung, Vielfalt und Potenzial

Zusammenfassung

Dieser Artikel dient als Einleitung in die Thematik Cognitive Computing. Aufgrund der fehlenden eindeutigen Definition gibt es unterschiedliche Auffassungen, was Cognitive Computing ist und wie es in zukünftigen Technologieentwicklungen eingesetzt werden kann. Dabei ist es längst nicht mehr nur ein Forschungsthema in der Informatik, sondern hat bereits Aufsehen in unterschiedlichen Disziplinen und in der Industrie erlangt. Dieser Artikel führt die einzelnen Beiträge ein, die in diesem Sammelband zu finden sind. Der erste Teil widmet sich dem Forschungsfeld – der künstlichen Intelligenz. Der zweite Teil diskutiert, warum heutige Systeme auch kreativ sein sollten. Der dritte Teil präsentiert Chatbot-Projekte, in welchen Ansätze aus dem Cognitive Computing angewandt wurden. Zuletzt folgen praktische Anwendungsfälle in unterschiedlichen Fachbereichen.

Einleitende Worte

Wenn ich gefragt werde, was ich unter Cognitive Computing verstehe, so lautet meine Beschreibung wie folgt: Cognitive Computing ist ein Überbegriff für das komplementäre Zusammenspiel von fortgeschrittenen Informations- und Kommunikationstechnologien mit Theorien aus der Psychologie, Neurobiologie, Linguistik und anderen Disziplinen für die Schaffung eines kognitiven Systems – einem System, das ähnliche Fähigkeiten besitzen kann wie ein Mensch (schlussfolgern, lernen, kreativ sein etc.). Einige von Ihnen werden sich nun fragen, ob und wie sich dies von künstlicher Intelligenz differenzieren lässt – womöglich gar nicht. Aber dies ist für dieses Buch nebensächlich. Definitionen sind in Zeiten, in welchen sich schnell alles ändert und grundlegende Strukturen und Prozesse im Wandel sind keine Hilfe. Daher möchte dieses Buch keine Definition zum Begriff Cognitive Computing liefern, sondern eher einen Einblick in dessen Potenzial geben.

Cognitive Computing ist längst nicht mehr nur ein Forschungsthema in der Informatik, sondern hat bereits Aufsehen in unterschiedlichen Disziplinen und in der Industrie erlangt. Aufgrund der bestehenden Unschärfe, die der Begriff Cognitive Computing mit sich trägt, unter anderem aufgrund der fehlenden eindeutigen Definition, eröffnen sich zahlreiche Berührungspunkte zu unterschiedlichen Anwen-

dungsfelder. Mit anderen Worten, Cognitive Computing ist ein Thema mit grossem Potenzial. Es liegt an uns, dies auch zu nutzen.

Als Co-Herausgeberin freut es mich nun die verschiedenen Teile dieses Buches einzuführen, die einen Einblick in die mögliche Bandbreite des Einsatzes von Cognitive Computing präsentieren. Dieses Buch widmet sich zuerst der künstlichen Intelligenz, wobei hier Cognitive Computing als Synonym betrachtet wird. Der zweite Teil führt kreative Systeme ein – Systeme, die nicht nur intelligent, sondern auch kreativ sein sollten. Es folgen Beiträge zum Thema Chatbots und im Anschluss kommen unterschiedliche Bereiche zum Zug, wie die mobile Robotik, Legal Tech, Skillsmanagement und eHealth.

Bevor wir uns den einzelnen Beiträgen widmen, möchte ich Ihnen noch etwas mit auf dem Weg geben. Dieses Buch ist ein Produkt, das aus einer transdisziplinären Zusammenarbeit entstanden ist – ein kollektives Werk. Daher folgendes: nur wenn wir zusammenarbeiten, können wir Gutes erschaffen!

Teil 1: Künstliche Intelligenz

Vorbild Gehirn – Randbedingungen für eine kognitive Architektur

Der deutsche Physiker und Neurobiologe Christoph von der Malsburg zeigt in seinem Artikel auf, welche Voraussetzungen geschaffen werden müssten, um eine auf neurowissenschaftlichen Erkenntnissen basierende kognitive Architektur zu schaffen, die es ermöglicht die menschliche Kognition annähernd nachzubauen. Dabei geht er in die Diskussion ein, ob und inwiefern die menschliche Kognition mittels moderner Technologien, insbesondere durch künstliche neuronale Netze, nachgebaut werden kann. Dabei diskutiert er kritisch mögliche Vorurteile, die den Fortschritt, die Funktion des Gehirns elektronisch nachzubilden, lähmen könnten.

Um weitere Einblicke in die menschliche Kognition zu bekommen, erklärt Christoph von der Malsburg anhand alltäglicher Beispiele, wie Menschen sich einer Situation bewusst werden, wie sie Objekte wahrnehmen und diese schliesslich verstehen. Dabei geht er detailliert auf die menschliche Anwendung von Schemas ein und erklärt, wie die Fähigkeit, angeborene oder erlernte Verhaltensmuster zu aktivieren, das Entschlüsseln und Kommentieren einer aktuellen Situation ermöglicht. Diese Fähigkeit verbindet er mit dem menschlichen Lernen, welches daraus besteht, neue Eindrücke mit bestehenden, im Gedächtnis gespeicherten Informationen, abzugleichen. Dieses menschliche Lernen vergleicht er mit den möglichen Lernmechanismen der künstlichen neuronalen Netze und hebt dabei den Mangel eines inhärenten Strukturkonzepts hervor, der wiederum hemmenden Einfluss auf die Nachahmung der menschlichen Kognition hat. Schliesslich fasst er die wesentlichen Erkenntnisse zusammen, die für eine kognitive Architektur vonnöten sind.

Die detaillierte Auseinandersetzung der Entwicklung eines kognitiven Rechners aus einer kognitionswissenschaftlichen und neurowissenschaftlichen Perspektive

ist informationsreich und ein perfekter Einstieg in die Welt des Cognitive Computings.

Artificial Intelligence – The Big Picture

Was ist künstliche Intelligenz und was ist der Reifegrad von KI-Systemen in Bezug auf verschiedene Einsatzgebiete? Um diese Fragen beantworten zu können, haben sich David Abele und Sara D’Onofrio mit den überbordenden Informationen und vielen Konzepten im Zusammenhang mit künstlicher Intelligenz aus der Forschung und Praxis auseinandergesetzt. Anhand der gesammelten Informationen und gewonnenen Erkenntnisse haben sie daraus ein Framework entwickelt, das aus IT-Sicht und basierend auf praktischen Erfahrungen die wesentlichen Merkmale einer künstlichen Intelligenz darstellt. Dabei definieren sie sechs fundamentale Komponenten: *Intelligence, Utilization, Aims, Technology, Enablers* und *Robustness, Transparency and Trust*. Die Autoren gehen auf jede Komponente detailliert ein und zeigen anhand illustrativer Beispiele, welchen Einfluss diese Aspekte auf die Entwicklung einer künstlichen Intelligenz in Forschung und Praxis haben und ob und inwiefern sie sich gegenseitig beeinflussen.

Die Aufstellung der einzelnen Komponenten und die Diskussion der Bestandteile anhand passender Illustrationen helfen dem Leser künstliche Intelligenz aus Sicht der Informatik zu verstehen. Dabei halten die Autoren fest, dass dieser Artikel keine abschliessende Darstellung künstlicher Intelligenz ist, sondern ein Ansatz, um das komplexe Thema in einer übersichtlichen, technischen und praxisorientierten Weise zu erfassen. Des Weiteren ermutigen sie Forscher, Entwickler, Unternehmer und Interessierte das Framework weiterzuentwickeln, weil es aus ihrer Sicht weiterer Ausarbeitung bedarf.

Dieser Artikel bietet einen Perspektivenwechsel und ermöglicht die künstliche Intelligenz aus einem neuen, übergreifenden Blickwinkel zu verstehen. Dabei werden Sie, liebe Leserin und liebe Leser, bemerken, dass dieser Artikel, trotz anderem Fokus, Gemeinsamkeiten mit dem vorherigen Artikel aufweisen wird – oder doch nicht?

Artificial Intelligence – The Mindfire Foundation and other Initiatives

Moreno Colombo, Edy Portmann und Pascal Kaufmann setzen sich ebenfalls mit dem Begriff *Künstliche Intelligenz* auseinander. In ihrem Beitrag analysieren sie kritisch bisherige technologische Fortschritte und erklären, warum möglicherweise die Forschung und Entwicklung künstlicher Intelligenz bisher eher schleppend voranging. Dabei heben sie besonders den Aspekt der Forschungszusammenarbeit hervor, indem sie statuieren, dass für den Bau einer künstlichen Intelligenz, nicht mehr (inter)disziplinäres Arbeiten vonnöten ist, sondern nun

cher eine trans- wenn nicht gar antidisziplinäre Zusammenarbeit angestrebt werden soll.

Für das Erreichen dieses Zieles, der Bau einer künstlichen Intelligenz, führen die Autoren vier wesentliche Bausteine – *Wissenschaftliche Forschung*, *Aus- und Weiterbildung*, *Adoption* und *Ethik und Regulationen* – in Form eines Frameworks ein. Die ausführliche theoretische Auseinandersetzung wird mit einem Anwendungsfall aus der Schweiz, der *Mindfire Foundation*, untermauert. Anhand dieser Non-Profit-Organisation werden die einzelnen theoretischen Blöcke mit reellen Beispielen illustriert und es wird aufgezeigt, wie die Schweiz einen Beitrag in der Entwicklung künstlicher Intelligenz leistet und welchen Einfluss dabei die transdisziplinäre beziehungsweise antidisziplinäre Forschung hat.

Das Spannende an diesem Beitrag ist nicht nur, dass aus theoretischer und praktischer Perspektive über den aktuellen Stand der künstlichen Intelligenz gesprochen wird, sondern auch, dass die Autoren dabei den Ansatz des *Phenotomics* miteinbeziehen. Ein Ansatz, der die Konzeption von Systemen zulässt, die sich an interagierenden Systemen anpassen können, ohne dass Protokolle erforderlich sind. Womöglich ein Ansatz, der die Entwicklung der künstlichen Intelligenz in eine neue Richtung schlagen wird?

Teil 2: Kreative Systeme

Zur Grenze des Verstehens künstlicher Kreativität

Der Artikel von Jan Sebastian Zipp lässt, nach einer ausführlichen Einführung über künstliche Kreativität, einen Perspektivenwechsel zu. In seinem Beitrag zeigt er illustrativ auf, wie eng verwoben die Diskussionen um menschliche und künstliche Kreativität sind und wie schwer es ist, *Kreativität* zu definieren. Um dennoch ein Verständnis über Kreativität zu erlangen, werden die drei Ebenen nach Margaret Boden – *kombinatorisch*, *explorativ* und *transformativ* – präsentiert, wobei im Folgenden der Fokus auf die transformative Kreativität gelegt wird. Basierend auf dieser theoretischen Grundlage wird die Entwicklungsgeschichte des Begriffes *Kreativität* erläutert. Hierbei wird verdeutlicht, aus welchen Gründen bis heute noch keine eindeutige Definition dieses interdisziplinären Begriffs besteht. Um dies weiter zu untermauern, illustriert Jan Sebastian Zipp anhand alltäglicher Beispiele, wie Menschen (transformativ) kreativ sein können und erläutert, warum Menschen im Hinblick auf Kreativität gegenüber kognitiven Systemen (noch) überlegen sind.

Mit der Aufforderung die Perspektive zu wechseln, überzeugt Jan Sebastian Zipp in seiner Argumentation, dass der Versuch künstliche Kreativität zu definieren oder sie sogar skalierbar zu machen, nicht zielführend ist. Bei der Entwicklung kognitiver Systeme sollte nicht die Frage gestellt werden, wie kreativ ein System werden kann oder soll, sondern wie solche Systeme den Menschen in ihrer kreativen Entfaltung unterstützen können.

Maschinen zur Lösung der grossen Probleme – Ein formeller Rahmen

Philip Gross präsentiert in seinem Artikel einen formellen Rahmen, der helfen soll, Probleme der Weltbevölkerung (wie bspw. die des Klimawandels) zu lösen. Anhand eindrücklicher Beispiele wird den Lesern vorgeführt, wie der eigene Beitrag für die Lösungsfindung grosser Probleme durch das individuell wahrgenommene Risiko und den eigenen einzusetzenden Ressourcenbedarf beeinflusst wird. Weiter wird dargelegt, dass unsere Probleme in *well-structured* und *ill-structured* eingeteilt werden können, und abhängig von ihren Ausprägungen, verschiedene Vorgehensweisen für deren Lösung notwendig sind. Im formellen Rahmen zeigt Philip Gross auf, wie *well-structured* und *ill-structured* Probleme mithilfe von Maschinen angegangen werden können.

Die maschinelle Unterstützung ist aber gemäss Philip Gross nicht der einzige essenzielle Bestandteil für die Lösung grosser Probleme. Überzeugend erläutert er, warum ein Netzwerk verschiedener menschlicher und künstlicher Instanzen und der Zugang zum Semantic Web sowie die Elemente des funktionalen Denkens (z. B. Informationen abrufen und verbinden) vonnöten sind, um der Lösung grosser Probleme näher zu kommen. Dabei kommt er wieder auf das individuell wahrgenommene Risiko eines Menschen zurück und schlägt zusätzlich vor, die Technologie Blockchain und besonders die Anwendung Smart Contracts zu nutzen, um die partizipative Lösungsfindung zu fördern. Denn nur wenn wir bereit sind unser Wissen und unsere Kreativität für die Lösung der grossen Probleme zu teilen, können wir womöglich unsere Welt retten.

Teil 3: Chatbots

Chatbots & Cognitive Services – Ein Schritt Richtung Automatisierung im User Help Desk der Schweizerischen Post

Nicht nur die Forschung ist an der sich rasant entwickelnden Technologie interessiert, sondern auch Unternehmen wie die Schweizerische Post. Der Artikel von Julie Mackmood, Daniel Bammert Marty und Sara D'Onofrio diskutiert, wie eine Automatisierung im User Help Desk mittels eines textbasierten Chatbots angegangen werden kann. Da in technologischen Belangen in Grossunternehmen meist die *Make or Buy*-Frage gestellt wird, ist eine der Herausforderungen den geeigneten Anbieter für das jeweilige Produkt zu finden und/oder zu bestimmen, welche Funktionen selbst entwickelt werden sollen. Dies ist besonders bei der Entwicklung und Implementierung von Chatbots eine zeitaufwendige und herausfordernde Aufgabe.

Dieser Artikel adressiert diese Herausforderung, legt jedoch den Fokus auf die Evaluation von Microsoft-Produkten. In diesem Anwendungsfall führen die Autoren die Cognitive Services von Microsoft ein (die für die emergente Entwicklung

von Cognitive Computing stehen) und evaluieren diese mit einem selbst entwickelten Kriterienkatalog. Die Bewertung der Cognitive Services gibt dann Aufschluss über deren Tauglichkeit.

Dieser Artikel offeriert einen Einblick, wie Chatbot-Produkte evaluiert werden können und welche Kriterien in der Entscheidungsfindung für die Entwicklung und Implementierung eines Chatbots zu berücksichtigen sind. Ein Artikel aus der Praxis, der erlaubt, die wissenschaftliche Brille abzulegen und die Chatbot-Thematik aus der praktischen Perspektive zu erfahren.

Soziotechnische Gestaltung von Chatbots

Im Artikel von Ulrich Gnewuch, Jasper Feine, Stefan Morana und Alexander Maedche wird die Gestaltung von Chatbots aus einer soziotechnischen Perspektive betrachtet. Dabei adressieren die Autoren ein wichtiges Thema aus der Kommunikationswissenschaft: die Wirkung sozialer Signale und die Entstehung sozialer Reaktionen. Beide sind unentbehrlich für die Entwicklung heutiger Dialogsysteme geworden. Die Autoren führen verschiedene soziale Signale (wie z. B. Lächeln, Sprachstil oder Antwortgeschwindigkeit) ein, erklären die Bedeutung solcher Signale in der Interaktion mit anderen Menschen und zeigen auf, warum diese Erkenntnisse in die Entwicklung eines Chatbots einfließen sollten.

In ihrem Artikel wird deutlich, dass die Entwicklung eines Chatbots nicht nur eine technische Herausforderung ist, sondern auch ein Augenmerk auf die zwischenmenschlichen Aspekte gelegt werden sollte. Anhand des Beispiels eines chronemischen Signals – der Antwortgeschwindigkeit – zeigen die Autoren auf, welche Auswirkungen es auf die Wahrnehmung eines Chatbots und die Zufriedenheit der Nutzer hat, wenn ein Chatbot unnatürlich schnell oder etwas zeitverzögert antwortet. Des Weiteren wird kritisch über die Vermenschlichung von Chatbots diskutiert und der Frage nachgegangen, ob und inwiefern gewisse Grundzüge eines Chatbots durch soziale Signale vermenschlicht werden sollten.

Der Artikel bietet einen interessanten Einblick in ein Forschungsprojekt, das sich der soziotechnischen Gestaltung von Chatbots widmet und reflektiert kritisch die Auswirkungen einer Vermenschlichung von intelligenten Systemen – Erkenntnisse, die in der Praxis von Nutzen sein werden!

Teil 4: Praktische Anwendungsfälle

Kollektive Intelligenz von Mensch und Maschine – Fuzzy Logik zur Einbindung von Expertenwissen in die Wegeplanung von Fahrerlosen Transportfahrzeugen

Der Artikel von Sarah Uttendorf stellt einen Ausschnitt ihrer Dissertation dar, in welcher sie sich mit mobiler Robotik beziehungsweise Fahrerlosen Transportfahr-

zeugen und -systemen auseinandergesetzt hat. Sie führt in diesem Artikel nicht nur die theoretischen Grundlagen ein, sondern gewährt auch Einblicke in die Praxis – Erfahrungen, die sie während ihrer Dissertationszeit gewinnen konnte.

Zu Beginn präsentiert Sarah Uttendorf eine Übersicht verschiedener Methoden für die Erstellung von Wegenetzen, zeigt anschliessend auf, wie sich offline und online beziehungsweise manuelle und automatische Planung unterscheiden, und diskutiert schliesslich im Bereich Fahrerlose Transportfahrzeuge und -systeme den Einsatz von Cognitive Computing Ansätzen mit Schwerpunkt Fuzzy Logik. Dabei geht Sarah Uttendorf auf die allgegenwärtige Unsicherheit in der Wegeplanung ein, die mittels Fuzzy Logik (bzw. linguistischer Terme) adressiert werden kann. Mit visualisierenden Beispielen erklärt sie, wie und aus welchen Gründen Fuzzy Logik Regeln helfen, Fahrerlose Transportfahrzeuge und -systeme zu optimieren und präsentiert somit eine Möglichkeit wie die kollektive Intelligenz von Mensch und Maschine vorangetrieben werden kann.

Sarah Uttendorf führt die Leser nicht nur in die Welt der mobilen Robotik ein, sondern spricht auch über die Herausforderungen und Chancen. Dabei zeigt sie auf, wie komplex die Erstellung von Wegenetzen sein kann, und welche Entscheidungen unter welchen Umweltfaktoren, getroffen werden müssen, um geeignete Wegenetze erstellen zu können.

Legal Tech – Cognitive Computing in der juristischen Praxis

Aus Recht wird Code – die Rechtsbranche wird digitalisiert. Philip Hanke erzählt aus eigener Erfahrung, wie sich die Rechtsbranche mit den technologischen Fortschritten auseinandersetzt und visualisiert anhand von Beispielen, wie Technologie eine Unterstützung für JuristInnen, NotarInnen und RechtsanwältInnen sein kann. Dabei weist er auf die Herausforderungen, die im Rechtswesen vorhanden sind und reflektiert kritisch, ob und inwiefern moderne Technologien für bestimmte Bereiche einen Nutzen generieren können.

Das Hauptmerkmal des Rechts ist die natürliche Sprache. Klauseln in Verträgen oder Allgemeine Geschäftsbedingungen müssen wasserdicht formuliert werden. Unschärfe ist in dieser Branche nicht erwünscht. Daher führt Philip Hanke eine breite Palette von Technologien ein, welche die Sprache im Sinne der Rechtsbranche verarbeiten können: von einfachen Wenn-Dann-Aussagen bis zu künstlichen neuronalen Netzwerken. Dabei zeigt er auf, in welchen Rechtsbereichen die unterschiedlichen Technologieansätze eingesetzt werden können.

Ein weiteres Augenmerkmal legt Philip Hanke auf den zurzeit bestehenden Trend der Blockchain und dessen Anwendung von Smart Contracts. Er erläutert, aus welchen Gründen es sinnvoll erscheint, dass sich JuristInnen das Wissen im Bereich Blockchain aneignen. Des Weiteren diskutiert er den Bau sowie den Einsatz von Robo-AnwältInnen und Robo-RichterInnen, die eine Erweiterung von Chatbots darstellen. Könnte die Rechtsberatung in naher Zukunft durch eine Maschine durchgeführt werden?

Die Digitalisierung des Skillmanagements – Kognitiver Digitaler Assistent für die Wertschöpfung des Human Resource Management in der Digitalen Transformation

Bisher wurden das Verständnis von und die Erfahrungen mit Cognitive Computing mehrheitlich aus technischer oder neurowissenschaftlicher Perspektive diskutiert. Adrian Vogler offeriert in seinem Artikel die Thematik Cognitive Computing aus einer Managementsicht anzugehen. In seinem Beitrag diskutiert er den Wandel von Berufsbildern und greift dabei weit zurück, in eine Zeit, in der es noch Fischbeinreisser und Kalfater gab und wagt gleichzeitig einen Blick in die Zukunft, in eine Zeit, in der wir nicht wissen, welche neue Berufsbilder entstehen werden. Basierend auf diesem kontinuierlichen Wandel, adressiert er die Herausforderung, mit welcher die heutigen HR-Abteilungen konfrontiert sind: die Bildung und Haltung einer optimalen Workforce für das Unternehmen.

Adrian Vogler präsentiert einen kognitiven, digitalen Assistent, der über eine Cognitive-Computing-Plattform helfen soll, die optimale Workforce eines Unternehmens zu berechnen. Er zeigt anhand zweier Szenarien, wie der kognitive, digitale Assistent helfen kann, bestehende Fähigkeiten und Fertigkeiten in einem Unternehmen zu berechnen, neue benötigte Skills für die zukünftigen Herausforderungen zu erfassen, und daraus die Kosten zusammenzutragen, die für eine Transformation der (neuen) Workforce anfallen werden.

Die Managementsicht auf die Thematik Cognitive Computing führt zur Erkenntnis, dass die Digitalisierung überall im Gange ist, auch in Abteilungen, die bisher nur wenige Berührungspunkte mit Technologien hatten. Wie dieser Artikel zeigt, sind moderne Technologien nicht mehr nur ein Thema in der Informatik, sondern sollten längst in jeder Branche als wichtiger Bestandteil angesehen werden.

Health Recommendation System Framework for the Optimization of Medical Decisions

Luis Terán, Jhonny Pincay, Diana Pacheco, Martin Štěpnička und Daniel Simancas-Racines beschäftigen sich in ihrem Beitrag mit Empfehlungssystemen und zeigen auf, wie solche Systeme im Gesundheitswesen eingesetzt werden können, um die Entscheidungsfindung der Ärzte und anderer Gesundheitsfachleute zu unterstützen.

Zu Beginn erläutern die Autoren die Problemstellung im Gesundheitswesen. Dabei gehen sie auf die Informationsasymmetrie und das Agency-Problem ein. Des Weiteren präsentieren sie Erkenntnisse aus der Forschung und führen dann, für das Verständnis des Artikels, die wichtigsten Theorien ein. Darauf aufbauend, schlagen die Autoren ein eigenes Framework vor, das die medizinische Entscheidungsfindung optimieren soll. Detailliert erläutern sie die Systemarchitektur, gehen auf die unterschiedlichen Benutzerfunktionen ein, beschreiben den Impact, den sie sich durch das Framework erhoffen, adressieren soziale und ethische Aspekte und schla-

gen zuletzt noch geeignete Implementationsstrategien vor, um das Produkt in Umlauf zu bringen. Dieses Framework gehört zu einem grösseren Projekt, das den Namen *MedicGate* trägt. Es handelt sich um ein soziales Netzwerk, welches darauf abzielt, medizinische Entscheidungen basierend auf Cognitive Computing zu optimieren.

Der Artikel gibt nicht nur Einblick in ein grösseres Forschungsprojekt und präsentiert detailliert ein Framework, das für die Entscheidungsfindung für medizinische Problemstellungen eingesetzt werden kann, sondern illustriert auch wie eine Zusammenarbeit zwischen länder- und disziplinenübergreifenden Universitäten und Industrie funktionieren kann.

Danksagung

Im Namen der Herausgeber möchte ich allen Personen, die an diesem Buch beteiligt waren, herzlichst danken. Ein besonderer Dank geht an die Autorinnen und Autoren, die sich die Zeit genommen haben, ihr Expertenwissen und ihre Erfahrungen in interessanten Beiträgen zu verfassen. Die wertvollen Erkenntnisse aus Forschung und Praxis machen dieses Werk einzigartig und auch für Nicht-Expertinnen und Nicht-Experten greifbar. Des Weiteren möchte ich mich bei den Gutachterinnen und Gutachter bedanken, die sich die Zeit genommen haben, die einzelnen Beiträge kritisch und konstruktiv zu reviewen, das zur Verbesserung der Qualität, Kohärenz und inhaltlichen Darstellung des Buches geführt hat. Danke für die tolle Zusammenarbeit!

Human-IST Institut
Universität Fribourg,
Fribourg, Schweiz
sara.donofrio@unifr.ch

Sara D’Onofrio

Inhaltsverzeichnis

Teil I Künstliche Intelligenz

Vorbild Gehirn – Randbedingungen für eine kognitive Architektur	3
Christoph von der Malsburg	
1 Einleitung	3
2 Lähmende Vorurteile	5
2.1 Denken in Direkten Algorithmen	5
2.2 Agnostische Statistik	7
2.3 Die Einzellhypothese	9
3 Vom Ursprung des Geistes	10
4 Der Neuronale Code	12
5 Kognition	13
5.1 Aufmerksamkeitsfokus, Situationsbewusstsein	14
5.2 Wahrnehmung	14
5.3 Verstehen, Abstraktion	18
6 Implementierung von Schema-Verstehen	20
7 Lernen	22
8 Vergleich mit Künstlichen Neuronalen Netzen	25
9 Diskussion	27
Literatur	28
Artificial Intelligence – The Big Picture	31
David Abele und Sara D’Onofrio	
1 Introduction	32
2 Hype or Ripe?	33
3 Key Concepts of Artificial Intelligence	34
3.1 Intelligence: Maturity	35
3.2 Utilization: Fields of Application, AI Capabilities, AI Applications	38
3.3 Aims: Target Outcomes	42
3.4 Technology: Approaches to Artificial Intelligence	44
3.5 Technology: Learning Strategies	48
3.6 Enablers: Data	51
3.7 Robustness, Transparency and Trust	54

4	Discussion	57
4.1	Maturity, Technology and Approaches.	57
4.2	The Dangers of AI: Misalignment of Goals.	58
4.3	Practical Considerations.	59
5	Conclusion and Outlook.	60
	References.	61

Artificial Intelligence – The Mindfire Foundation and Other Initiatives. 67

Moreno Colombo, Edy Portmann und Pascal Kaufmann

1	Introduction	68
2	Related Work	70
2.1	Current Limitations of Artificial Intelligence.	70
2.2	Transdisciplinary Research	72
2.3	Antidisciplinary Research	73
3	Framework: The Building Blocks of an AI Initiative.	74
3.1	Scientific Research.	75
3.2	Education	76
3.3	Public and Private Sector Adoption of Artificial Intelligence	76
3.4	Ethics and Regulations Frameworks	77
3.5	Overview of AI Initiatives	77
4	Use Case: The Mindfire Foundation	78
4.1	Scientific Research.	78
4.2	Education	80
4.3	Public and Private Sector Adoption of Artificial Intelligence	80
4.4	Ethics and Regulations Frameworks	81
4.5	Overview of the Building Blocks of Mindfire	82
5	Conclusion	82
	References.	83

Teil II Kreative Systeme

Zur Grenze des Verstehens künstlicher Kreativität. 89

Jan Sebastian Zipp

1	Einleitung	90
2	Kreativität im Zeitalter des Cognitive Computings	91
3	Der Kreativitätsbegriff: Kreativität ist nicht gleich Kreativität	92
3.1	Die unterste Ebene – Die kombinatorische Kreativität	93
3.2	Die mittlere Ebene – Die explorative Kreativität.	94
3.3	Die höchste Ebene – Die transformative Kreativität	94
4	Eine kurze Entwicklungsgeschichte der Künstlichen Kreativität	95
5	Die Entwicklung des originalen Kreativitätsbegriffs.	96
6	Künstliche Kreativität: Ein Perspektivenwechsel	98
7	Charakteristika für Transformative Kreativität	101
7.1	Körperlichkeit.	101
7.2	Individualität/Subjektivität.	102

7.3	Bewusstsein	104
7.4	Das Unbewusste	104
7.5	Sinnlichkeit	105
8	Ausblick: Die Wiederverzauberung der Welt.	105
	Literatur.	107

Maschinen zur Lösung der grossen Probleme – Ein formeller Rahmen ... 111

Philip Gross

1	Grosse Probleme – Grosse Risiken	112
1.1	Maschinen für mehr Kreativität	113
1.2	Geringe Wahrscheinlichkeit eines Lösungserfolgs	114
1.3	Unvorhersehbar hoher Ressourceneinsatz	115
2	Ungesteuerte Risikominderung verschleppt die Problemlösung	117
3	Forcierte Lösungsfindung auf online Kollaborations-Plattformen.	117
3.1	Informations- und Partizipationsoffenheit	118
3.2	Polymath vs. Foldit	119
4	„Well-structured Problems“ und „Ill-structured Problems“	121
5	Zwei Systeme zur Lösung bedeutender Probleme: Ein formeller Rahmen	122
5.1	Problemstruktur-Typen: Kreativität für ISPs und Produktivität für WSPs.	123
5.2	Wirkungsziele: Wie die Maschinen zur Risikominderung beitragen.	126
5.3	Leistungsanforderungen: Was von Maschinen erwartet wird	128
6	Maschinen im Creativity System	132
6.1	Network of Neural Networks	132
6.2	Semantic Web der Kognitionen	133
6.3	Elemente funktionalen Denkens	135
7	Maschinen im Productivity System	136
7.1	Konvergenz als Evolution	136
7.2	Gaming the Limbic System	137
7.3	Informationsoffenheit mit „Smart Contracts“	138
7.4	Ein Reproduktions-Modus für die Expertin.	139
7.5	Ein Transpositions-Modus für den Citizen Scientist	140
8	Und wer baut diese Maschinen?	141
	Literatur.	141

Teil III Chatbots

Chatbot & Cognitive Services – Ein Schritt Richtung Automatisierung im User Help Desk der Schweizerischen Post

Julie Mackmood, Daniel Bammert Marty und Sara D’Onofrio

1	Der User Help Desk der Schweizerischen Post	148
2	Cognitive Computing	150
3	Chatbots	151
4	Textbasierter Chatbot für den User Help Desk	153

5	Cognitive Services von Microsoft	154
5.1	Knowledge-API	154
5.2	Language-API	155
5.3	Search-APIs	156
5.4	Speech-APIs	157
5.5	Vision-APIs	157
5.6	Erstes Fazit	157
6	Analyse der Cognitive Services	158
6.1	Kriterienkatalog	158
6.2	Bewertung der Knowledge-API und Language-APIs	161
6.3	Zweites Fazit	161
7	Diskussion	164
8	Fazit und Ausblick	166
	Literatur	167

Soziotechnische Gestaltung von Chatbots 169

Ulrich Gnewuch, Jasper Feine, Stefan Morana und Alexander Maedche

1	Einleitung	170
2	Chatbots und die Wirkung ihrer sozialen Signale	172
3	Soziotechnische Gestaltung von Chatbots am Beispiel der Antwortgeschwindigkeit	175
3.1	Antwortgeschwindigkeit von Chatbots	176
3.2	Dynamische Antwortverzögerung für Chatbots	177
3.3	Experimentaldesign und Chatbot-Szenario	179
3.4	Ergebnisse und Diskussion des Experiments	180
4	Diskussion	182
4.1	Adaptive Gestaltung der sozialen Signale von Chatbots	183
4.2	Wie menschenähnlich sollte ein Chatbot sein?	184
5	Fazit	186
	Literatur	186

Teil IV Praktische Anwendungsfälle

Kollektive Intelligenz von Mensch und Maschine – Fuzzy Logik zur Einbindung von Expertenwissen in die Wegeplanung von Fahrerlosen Transportfahrzeugen 193

Sarah Uttendorf

1	Was sind Fahrerlose Transportfahrzeuge?	193
2	Herausforderungen bei der Wegeplanung von Fahrerlosen Transportsystemen	195
3	Cognitive Computing zur Erstellung von Wegenetzen	196
4	Fuzzy Logik für die Wegeplanung von Fahrerlose Transportsysteme	200
5	Herausforderungen beim Einsatz von Cognitive Computing für Fahrerlose Transportfahrzeuge	203
6	Vorteile	204
7	Ausblick	205
	Literatur	205

Legal Tech – Cognitive Computing in der juristischen Praxis 207

Philip Hanke

1	Einleitung.	207
2	Dokumentenautomatisierung	209
3	Aus Recht wird Code.	210
4	Datenanalyse und Natural Language Processing – Vom Einzelfall zur Statistik	213
4.1	Regular Expression als Freund und Helfer.	213
4.2	Deep Learning	215
4.3	Vom Chatbot zum Robo-Lawyer?	216
4.4	Was, wenn sich der Chatbot irrt?	218
5	Smart Contracts und Decentralized Ledgers	219
6	Totale Kontrolle durch Legal Tech.	220
7	Ausblick.	221
	Literatur.	222

**Die Digitalisierung des Skillmanagements – Kognitiver Digitaler
Assistent für die Wertschöpfung des Human Ressource Management
in der Digitalen Transformation 225**

Adrian Vogler

1	Motivation	226
2	Eine (Berufs-)Welt im Wandel.	227
3	Bestehende Ansätze	228
4	Offene Fragen – Details zum Strategischen Workforce Management.	229
5	Erfolgsfaktoren & Aufwände.	230
6	Neue Digitale Assistenten für HR – Welche Rolle spielt dabei Cognitive Computing?	231
7	Wertschöpfende Verbindungen schaffen	233
8	Notwendige Kognitiv-Digitale Assistenzfunktionen	234
9	Ausbau der Plattform.	235
10	Minimum Viable Product: Workforce-Transformations-Assistent	236
11	Der Anwendungsfall	237
11.1	Beispiel: Fotografie	238
11.2	Beispiel: Automobilindustrie	243
12	Zusammenfassung	246
	Literatur.	246

**Health Recommendation System Framework for the Optimization of
Medical Decisions 249**Luis Terán, Jhonny Pincay, Diana Pacheco, Martin Štěpnička und Daniel
Simancas-Racines

1	Introduction	250
2	Problem Statement.	251
2.1	Free Markets and Market Failures	252
2.2	The Healthcare Market.	253
2.3	What Is the Empirical Evidence?	254

3	State-of-the-Art on Health Recommender Systems.....	254
3.1	Health Recommender Systems	256
4	Theoretical Background.....	257
4.1	Graph Theory	257
4.2	Fuzzy Logic	259
5	Theoretical Framework	261
5.1	System Architecture.....	263
5.2	User Registration Process	264
5.3	Users Capabilities.....	266
5.4	Impact.....	266
5.5	Social and Ethical Aspects	267
5.6	Implementation Strategies	268
6	Outlook and Conclusions.....	268
	References.....	269
	Glossar	273

Über die Herausgeber und Autoren

Über Die Herausgeber

Edy Portmann ist Professor für Informatik und Förderprofessor der Schweizerischen Post am Human-IST Institut der Universität Freiburg (Schweiz) sowie Ehrenprofessor der ESPOL Universität in Guayaquil (Ecuador). Zu seinen transdisziplinären Forschungsschwerpunkten zählt das Thema Cognitive Computing sowie dessen Anwendung auf Städte. Er studierte Wirtschaftsinformatik, Betriebs- und Volkswirtschaftslehre und promovierte in Informatik. Er war unter anderem bei Swisscom, PwC und EY tätig. Edy Portmann forschte an den Universitäten Singapur, Berkeley und Bern. Er ist wiederholt für Marquis Who's Who nominiert worden, ein ausgewähltes Mitglied des Heidelberg Laureate Forum, Mitbegründer von Archai, FMsquare und Mindfire, Mitherausgeber der Springer-Reihe Fuzzy Management Methods sowie Autor mehrerer populärer Bücher auf seinem Gebiet.

Sara D'Onofrio hat einen zweisprachigen Bachelor in Betriebswirtschaftslehre, einen Master in Betriebswirtschaftslehre mit Spezialisierung in Wirtschaftsinformatik und einen CAS in Hochschuldidaktik. Im Herbst 2019 schloss sie ihr Doktoratsstudium in der Informatik am Human-IST Institut der Universität Freiburg, (Schweiz) ab. In ihrer Dissertation beschäftigte sie sich mit dem Thema *From Simple Question-Answering Systems to „Intelligent“ Chatbots* und entwickelte in einem gestaltungsorientierten und transdisziplinären Forschungsvorgehen mehrere Prototypen. Des Weiteren besuchte sie Weiterbildungskurse an unterschiedlichen europäischen Universitäten und nahm an verschiedenen Tagungen in Europa, Südamerika und Kanada teil. Sara D'Onofrio hat bereits mehr als 20 wissenschaftliche Arbeiten veröffentlicht und zwei ihrer Arbeiten wurden als Best Papers auserkoren. Sie ist Mitglied der Stiftung FMsquare sowie Member bei EUSFLAT und IEEE. Ihre aktuellen Forschungsinteressen umfassen die Themen Cognitive Computing, Innovation Management, Mensch-Maschine-Interaktion und Smart Cities.

Über Die Autoren

David Abele ist seit mehr als 15 Jahren im Bereich Informationstechnologie-Strategie und Technologietransformationen bei multinationalen Grossunternehmen am Puls neuer Technologien tätig. Er hält Masters in Wirtschaftsinformatik und in Betriebswirtschaft und kam in seiner Tätigkeit, hauptsächlich in Grossbritannien, schon frühzeitig bei praktischen Implementierungen mit neuronalen Netzwerken, statistischen Machine Learning Systemen, Sprachverarbeitung und anderen AI Technologien in Berührung. Seit 2016 ist er beim Konzern der Schweizerischen Post (Logistik, Öffentlicher Personentransport, Banking, Immobilien, Prozess Outsourcing Dienstleistungen) als IT-Unternehmensarchitekt tätig, wo seine breit-abgestützte Expertise bei der Erstellung der konzernweiten Technologiestrategie zum Einsatz kommt. Ebenfalls zu seinem Verantwortungsbereich gehören Technologie-Innovationen, speziell im Bereich der künstlichen Intelligenz.

Daniel Bammert Marty ist Leiter des User Managements in der Informatik der Schweizerischen Post. Er ist ein Schweizer Post Urgestein und ist vor zwei Dekaden in den Bereich Informatik quer eingestiegen. Nach Führungsjobs in Operations und Direktion hat Daniel Bammert Marty als System- und Netzwerkspezialist in der Informatik gestartet. Nebenbei hat er Wirtschaftsinformatik an der IFA in Zürich studiert. Nach dem Wechsel nach Bern in den Hauptsitz der Informatik hat er mehrere Jahre als Business Process Engineer gewirkt und die Business Unit Customer Support in Boards vertreten. Seine Affinität zu neuen Technologien und Geschäftsmöglichkeiten hat er nebenbei als Innovation Coach eingebracht. Im Sommer 2019 hat er einen Executive MBA in Innovation Management erfolgreich abgeschlossen. In diesem Kontext hat er sich vertieft mit dem Einsatz von kognitiven Services für den Einsatz im Umfeld des Customer Care beschäftigt.

Moreno Colombo has a Bachelor and a Master degree in Computer Science from the University of Fribourg. During and after his studies, he had the opportunity of experimenting interdisciplinary research by working on several projects with private companies, different faculties and end-users. Those included between others the use of augmented reality glasses as a prosthesis for people with severe vision problems. Since April 2018, Moreno is doing a PhD on phenotropic human-computer interaction at the Human-IST Institute at the University of Fribourg where he collaborates for some projects with Swiss Post. On the side, he is also working on initiatives about smart cities and he is participating as a Mindfire talent to the understanding and creation of human-like artificial intelligence. His main research focus lies on human-computer interaction, augmented reality and artificial intelligence with their respective applications, especially in the context of smart cities, but this does not keep him from being fascinated by almost anything, from quantum mechanics to cognitive psychology.

Jasper Feine ist wissenschaftlicher Mitarbeiter und Doktorand am Institut für Informationswirtschaft und Marketing (IISM) des Karlsruher Institut für Technologie

(KIT). Er hat am KIT einen Bachelor- und Masterabschluss in Wirtschaftsingenieurwesen erworben. Seine Forschung konzentriert sich auf das Design anthropomorpher adaptiver Chatbots. Insbesondere untersucht er, wie Informationen über den Benutzer erfasst werden können, wie adaptiv auf den Benutzer reagiert und wie dieser Mechanismus in einem Chatbot Prototypen modelliert und instanziiert werden kann.

Ulrich Gnewuch ist wissenschaftlicher Mitarbeiter und Doktorand am Institut für Informationswirtschaft und Marketing (IISM) des Karlsruher Institut für Technologie (KIT). Er studierte Wirtschaftsinformatik an der Universität Mannheim, wo er 2014 sein Bachelor- und 2017 sein Masterstudium abschloss. Seine Forschung konzentriert sich auf die Gestaltung, Implementierung und Evaluation von Chatbots, insbesondere in Bereichen wie Kundenservice und E-Commerce. Bei der hsg Heidelberg Services AG arbeitet er als Produktmanager an der Umsetzung von Chatbots für Energieversorgungsunternehmen.

Philip Gross ist Berater und Unternehmer. Als Partner und Gründer von Axonalliance (axonalliance.com) unterstützt er seine Kunden dabei, neue Geschäftsmodelle zu entwickeln und Wachstum mit neuen Marktleistungen zu realisieren. Als Mission Engineer verantwortet er den organisatorischen Aufbau und die inhaltliche Ausgestaltung der Non-Profit-Organisation Archai (archai.xyz), deren Ziel es ist, mittels neuer Technologien der Idee der *Collective Intelligence* zum Durchbruch zu verhelfen, um so das Problem des Klimawandels zu lösen. Philip Gross hat an der ETH Zürich studiert und das Diplomstudium als Ingenieur abgeschlossen. Später hat er an den Universitäten Hannover und Zürich in Wirtschaftswissenschaften promoviert.

Philip Hanke hat an der Universität Wien Volkswirtschaftslehre und Politikwissenschaft studiert und den European Master in Law and Economics (Universitäten Bologna, Rotterdam und Haifa) sowie einen Bachelor of Laws an der University of London absolviert. Er hat zum Thema staatliche Beihilfen für Unternehmen promoviert (Universitäten Rotterdam, Bologna, Hamburg). Von 2014 bis 2018 forschte er als Postdoc am Institut für Öffentliches Recht der Universität Bern im Rahmen des Nationalen Forschungsschwerpunkts *NCCR on the move – The migration-mobility nexus*. Dort hat er Seminare zu Legal Tech sowie zu rechtlichen Fragen betreffend Künstliche Intelligenz gehalten. An der School of Management and Law der Zürcher Hochschule für Angewandte Wissenschaften in Winterthur hält er seit Herbst 2018 Kurse zu Legal Tech. Neben seinen universitären Tätigkeiten hat er Erfahrung im Bereich Politikberatung (u. a. für Lexidale International Policy Consulting, das Österreichische Institut für Wirtschaftsforschung und den Schweizerischen Wissenschaftsrat). Seit 2019 ist er Verlagsleiter der Weblaw AG, wo er unter anderem ein Blockchain-Projekt entwickelt hat und Programmierkurse für JuristInnen hält.

Pascal Kaufmann is a neuroscientist, tech-entrepreneur and the visionary behind the non-profit organization Mindfire. He is determined to unite the best talents

worldwide to tackle the *brain code*. Some of his works include connected living brains to machines developed at the Northwestern University, Illinois (USA) and he built one of the first cyborgs, feeding information directly into brain tissue. His vision *to think with the power of 1000 brains* gave rise to the concept of Corporate Brains with self-learning algorithms, opening a new field called *Brain Technology*. His vision also led to the foundation of Starmind, a solution based on the principles of brain and artificial neural network research to create an insight of the knowledge of company employees, allowing them to connect with the right people based on their needs. Pascal Kaufmann takes a sceptical view on the current state of artificial intelligence and is convinced that human-like artificial intelligence can and must be created by and for the people as an answer to the challenges of the 21st century.

Julie Mackmood hat einen Bachelor in Computerwissenschaften und einen Master in Electronic Systems Design von der University of Manchester (UK). Julie Mackmood arbeitete als Software-Ingenieurin bei der Lloyds Bank und verschiedenen Software-Beratungsfirmen, bevor sie zu Astra Zeneca (Stockholm, Schweden) wechselte. Dort war sie für die Digitalisierung wissenschaftlicher Arbeiten zuständig. Danach zog sie in die Schweiz, wo sie bei Swisscom in verschiedenen Funktionen von der Geschäftsentwicklung über die Projektleitung bis hin zur Innovation tätig war. Dort fand sie ihre wahre Passion: neue Geschäfte entwickeln. Seit etwa fünf Jahren arbeitet sie nun bei der Schweizerischen Post. Als Innovationsmanagerin bei der Informatikabteilung der Schweizerischen Post beschäftigte sie sich mit innovativen Themen, wie künstliche Intelligenz, digitale Brillen und Robotik und absolvierte nebenbei einen Executive MBA in Innovation Management, dessen Abschluss sie im Sommer 2019 erlangt hat. Seit Sommer 2019 ist sie Co-Leiterin Innovation bei PostMail.

Alexander Maedche ist Professor am Karlsruher Institut für Technologie (KIT) und Leiter der Forschungsgruppe *Information Systems & Service Design* am Institut für Informationswirtschaft und Marketing (IISM) und am Karlsruhe Service Research Institute (KSRI). Sein Forschungsschwerpunkt ist die Gestaltung interaktiver, intelligenter Systeme. Seine Forschungsergebnisse erschienen in führenden internationalen Zeitschriften, wie dem Journal of the Association of Information Systems, International Journal of Human-Computer Studies, Business Process Management Journal, Information & Software Technology, IEEE Intelligent Systems, SIGMOD Record und AI Magazine.

Christoph von der Malsburg hat in Göttingen, München und Heidelberg Physik studiert und promovierte mit einer Arbeit über Elementarteilchenphysik am CERN in Genf im Jahre 1970 in Heidelberg. Daraufhin wechselte er in die Neurowissenschaft, an das Max-Planck-Institut für biophysikalische Chemie in Göttingen. Im Jahr 1988 erhielt er den Ruf als Professor für Informatik, Neurowissenschaften, Physik und Psychologie an die University of Southern California, Los Angeles. Ab 1990 gründete er mit Werner von Seelen parallel dazu als Professor für

Systembiophysik das Institut für Neuroinformatik an der Ruhr-Universität Bochum. Beide Positionen hatte er bis 2007 inne. Seither forscht er als Senior Fellow am Frankfurt Institute for Advanced Studies, das er 2008/9 leitete. Christoph von der Malsburg ist bekannt als Pionier der Theorie der Selbstorganisation von neuronalen Netzen in der Ontogenese des visuellen Systems. In der Folge arbeitet er an der Erforschung der kognitiven Architektur des Gehirns, hat auf der Basis von konkreten Resultaten dieser Arbeit zahlreiche Artikel über Vorgänge hauptsächlich im visuellen, aber auch auditorischen und olfaktorischen Teil des Gehirns veröffentlicht und ein Gesichtserkennungssystem entwickelt, das in zwei von ihm mit-gegründeten Firmen in Bochum und Los Angeles erfolgreich eingesetzt wurde. Unter seine Ehrungen zählt der Pioneer Award des Neural Network Council der IEEE, der Karl Heinz Beckurts Foundation Award for Outstanding Achievements in the Field of Neural Computing, der Koerber-Award for European Science und der Hebb Award 2003 of the International Neural Network Society. Er ist gegenwärtig Fellow of the International Neural Network Society.

Stefan Morana ist Postdoktorand am Institut für Informationswirtschaft und Marketing (IISM) des Karlsruher Institut für Technologie (KIT). Er erwarb einen Master-Abschluss in Informatik an der Fachhochschule Darmstadt und einen Dokortitel in Wirtschaftsinformatik an der Universität Mannheim. Seine Forschungsschwerpunkte liegen auf dem Design interaktiver, intelligenter Systeme und methodischen Aspekten der Design Science Forschung. Konkret untersucht er die Gestaltung von Assistenzsystemen und natürlichsprachlichen Nutzerschnittstellen, welche die individuelle Nutzung von Informationssystemen unterstützen.

Diana Pacheco is a Research Fellow in the Department of Health Services Statistics and Economics at Swiss Paraplegic Research in Nottwil, Switzerland. She is also Lecturer at the University of Neuchâtel in the Faculty of Economics and at the University of Lucerne in the Department of Health Science and Policy. She holds a PhD in Economics from the University of Neuchâtel (2013) and a MSc in Economics with Specialization in Econometrics from the University of Lausanne (2008). As part of the Swiss Program for Young Researchers, funded by the Swiss National Science Foundation, Diana Pacheco visited the Department of Economics at the University of La Plata (2012). Her research interests are on Health Economics, Impact Evaluation, Applied Economics, Social Policy, Energy Economics and Micro-econometrics. She has worked on several research projects, and has published her work on health, energy, applied economics and health economics journals.

Jhonny Pincay is a research assistant and PhD student at the Human-IST Institute of the University of Fribourg, Switzerland. In 2016, he finished an MSc in Information Technology at the University of Stuttgart, Germany. In 2013, he received a BSc in Computer Science from Escuela Superior Politécnica del Litoral (ESPOL), Guayaquil, Ecuador. His research interests include smart cities, fuzzy logic, recommender systems, computing with words, eHealth and software engineering.

Daniel Simancas-Racines was born in Ecuador. He graduated with honours from the Medical School at UTE University. He holds a postgraduate degree in Advanced Telemedicine and Bioinformatics from the Centre for Development of Advancing Computing (CEDAC). He also holds a Master degree in Public Health from UTPL, and a Master degree in Clinical Epidemiology from Pontifical Javeriana University. He is a doctoral candidate from the Autonomous University of Barcelona in *Methodology in Biomedical Research and Public Health*. He has been commissioned several roles at UTE University, including the direction of postgraduate degrees, and the direction of the Telemedicine Centre. He is currently the Director of the research department at UTE University, and the Director of the Public Health and Clinical Epidemiology Research Centre (CISPEC). He is also the coordinator of Cochrane Ecuador, and he has been a lecturer in evidence-based practice for more than 12 years. Daniel Simancas-Racines has published more than 30 papers in high impact journals indexed in SCOPUS, LILACS and Scielo. He has been awarded the *Best Researcher UTE University award* in the category of highest impact in scientific publications for two consecutive years (2017, 2018). His work has been focused on Research Methodology, which has allowed him to explore several fields such as blood transfusion, vaccines for the common cold, patients' preferences in telehealth, causes of mortality in Ecuador and evidence-based medicine. He is currently working on Cochrane systematic reviews, and the development of scientific research at UTE University.

Martin Štěpnička obtained his MSc in Applied Mathematics in 2002 and his PhD in Applied Mathematics – Fuzzy Modeling in 2008 from the University of Ostrava. In 2012, he was habilitated to Associate Professor. He started his career as a researcher in 2004 at the Institute for Research and Applications of Fuzzy Modeling of the University of Ostrava where he keeps the vice-director position since 2005. Since 2009, he is a member of the Board of the European Society for Fuzzy Logic and Technology (EUSFLAT) for four consecutive terms and 2017, he was elected to become a President of the EUSFLAT Society. Furthermore, he is an IEEE Senior Member. He is an area editor of JCR journals Fuzzy Sets and Systems (Elsevier) and International Journal of Computational Intelligence Systems (Atlantis Press), and furthermore, a guest editor of several special issues of other JCR journals. He is a frequent member of program committees of international conferences and a reviewer of more than 10 JCR journals. His research interests include mainly fuzzy modeling, especially fuzzy inference systems and fuzzy relational equations, and computational intelligence in time series forecasting and data analysis. His publication record contains about 70 publications on Web of Science out of which 20+ are in JCR journals with ca 600 citations and h-index equal to 15. In July 2016, Martin Štěpnička has received the FUZZ-IEEE Best Paper Award (out of 350 accepted submissions) in Vancouver, Canada.

Luis Terán was working as assistant doctor and lecturer at the Information Systems Research Group, University of Fribourg, Switzerland and he is a full professor at Universidad de Las Fuerzas Armadas (ESPE), Ecuador. He earned a PhD in

Computer Science at the University of Fribourg (Information Systems Research Group). In 2009, he finished an MSc in Communication Systems from the Federal Institute of Technology (EPFL), Lausanne, Switzerland. In 2004, he received a BSc in Electronics and Telecommunications from Escuela Politécnica Nacional, Quito, Ecuador. His research interests include eGovernment, eParticipation, eCollaboration, eDemocracy, eElection, eVoting, eCommunities, ePassports, recommender systems, and fuzzy classification. He is currently a vice-chair for the IEEE e-Government Special Technical Community, program chair and main organizer for the International Conference on eDemocracy and eGovernment (ICEDEG), guest editor at IEEE Transactions on Emerging Topics in Computing, and distinguished exhibitor at IEEE Ecuador Section.

Sarah Uttendorf absolvierte ihr Bachelorstudium in der Fachrichtung Wirtschaftsingenieurwesen am Karlsruher Institut für Technologie (KIT). Anschließend studierte sie im Zuge ihres Masters ein Jahr an der University of Connecticut (USA) und schloss ihr Masterstudium zum Wirtschaftsingenieur am Karlsruher Institut für Technologie (KIT) 2014 ab. Im Zeitraum von 2014–2017 arbeitete sie am Institut für Integrierte Produktion Hannover (IPH) im Bereich der Produktionsautomatisierung. Dort leitete sie Forschungsprojekte in den Bereichen der Fahrerlosen Transportfahrzeuge und der Digitalisierung. In dieser Zeit fertigte Sarah Uttendorf zudem ihre Dissertation zu dem Thema *Automatisierte Auslegung von Wegenetzen für Fahrerlose Transportsysteme* an. Ihre Forschungsschwerpunkte liegen im Bereich mobile Robotik, Fuzzy Logik und Methoden der Künstlichen Intelligenz. Ein besonderer Fokus liegt auf der Wegeplanung. Seit 2017 arbeitet sie für die Robert Bosch GmbH im Bereich der mobilen Robotik.

Adrian Vogler ist Gründer und CEO des Startup StratWorX UG (<http://www.stratworx.de>), dass sich den Aufbau einer AI-basierten Plattform zum Ziel gesetzt hat. Als Mitglied der Peter Drucker Society Europe sowie des Expert Network Horizon 2020 der EU ist es ihm ein Anliegen, Unternehmer und verantwortliche Manager allen Organisationen, insbesondere KMU, über die Möglichkeiten der Digitalisierung verständlich und klar zu informieren und sie bei der Implementierung digitaler Instrumente in ihr Geschäftsmodell zu unterstützen. Er bringt fast 30 Jahre praktische Arbeitsmarkt- und Digitalisierungs-Erfahrung aus einer Vielzahl von Projekten mit.

Jan Sebastian Zipp ist Doktorand am Lehrstuhl für Kunsttheorie & Inszenatorische Praxis an der Zeppelin Universität in Friedrichshafen am Bodensee. Hier und in seiner mehrjährigen Tätigkeit am THINKLab des IBM-Forschungslabors in Zürich diskutiert und erforscht er den Kreativitätsbegriff in Grossunternehmen der IT-Branche. Jan Sebastian Zipp ist Absolvent des Studiengangs Communication & Cultural Management an der Zeppelin Universität. In seiner Abschlussarbeit behandelte er das Thema „Kreativität und kreativer Imperativ“. Zuvor studierte er in den Niederlanden internationale Wirtschaftswissenschaften, wovon er zwei Semester bei der Deutschen Bank mit Stationen in Frankfurt und Amsterdam verbrachte.

Seine Forschungsinteressen und -veröffentlichungen liegen im Bereich der diskursiven Kreativitätsforschung, der Technikphilosophie, der Philosophie des Geistes, der Zukunft des Arbeitens und des Lernens sowie einer kultursoziologischen Perspektive auf neue Formen von (künstlichen) Kreativitätskonzepten.

Teil I

Künstliche Intelligenz

Vorbild Gehirn – Randbedingungen für eine kognitive Architektur

Christoph von der Malsburg

Zusammenfassung

Um menschliche Kognition als emergentes Phänomen verstehen zu können, müssen zwei Fragen beantwortet werden: Was ist das gemeinsame Datenformat aller kognitiven Strukturen und durch welchen Mechanismus wird dieses Datenformat geformt? Für die Beantwortung dieser Fragen dürfte es nützlich sein, sich von Wissen über Strukturen und Vorgänge im Gehirn, dem einzigen existierenden Modell für Kognition, leiten zu lassen. Als Datenformat werden hier strukturierte Netze vorgeschlagen, die durch den gut untersuchten Vorgang der Netzwerk-Selbstorganisation gebildet werden, und wichtige kognitive Prozesse werden auf dieser Basis diskutiert. Der Artikel setzt sich mit einigen der das Gebiet beherrschenden Vorurteile auseinander.

1 Einleitung

Jeder Versuch, *Cognitive Computing* oder *künstliche Intelligenz* zu erzeugen, muss sich am Gehirn als einzigem Beispiel und Existenzbeweis orientieren. Alle bisherigen Versuche spielen sich in einem seit fünf Jahrzehnten unveränderten begrifflichen Rahmen ab und können sich in wesentlichen Punkten nicht mit den Leistungen eines auch nur dreijährigen Kindes messen. Die unbestreitbaren Erfolge dessen, was heute *künstliche Intelligenz* genannt wird, sind eng beschränkt auf Gültigkeitsbereiche rund um menschliche Vorgaben, Vorgaben entweder in der Form von Algo-

Vollständig überarbeiteter und erweiterter Beitrag basierend auf von der Malsburg, C.: Die Geschichte unserer Geburt, Informatik Spektrum (2018) 41(2): 85–87

C. von der Malsburg (✉)

Frankfurt Institute for Advanced Studies (FIAS), Frankfurt am Main, Deutschland

E-Mail: malsburg@fias.uni-frankfurt.de

rithmen oder von Massen von Beispiel-Lösungen. Sobald ein solcher enger Rahmen vorgegeben ist, kann die elektronische Datenverarbeitung mit ihrer ungeheuren Geschwindigkeit den Menschen weit überflügeln, aber eben nur in diesem Rahmen. Intelligenz, Kognition und jedes Verstehen von Inhalten sind aber weiterhin ganz beim Menschen.

Angesichts der jahrzehntelangen Stagnation in den Bereichen *künstliche Intelligenz* und *Verständnis der Hirnfunktion* drängt sich der Verdacht auf, dass der Fortschritt durch fundamentale Vorurteile blockiert ist. Was sind diese Vorurteile und welche Perspektiven öffnen sich, sobald sie einmal identifiziert und beseitigt sind?

Zunächst seien die Bedingungen betrachtet, unter denen der Geist im Kind entsteht. Unser Organismus samt Gehirn wird auf der Basis von einem Gigabyte genetischer Information erzeugt (Consortium 2001). Die Umgebung unserer Kleinkinder, die wir mit Bedacht einfach und stabil gestalten, könnte als *Virtual Reality* mit Hilfe eines Programms vom Umfang einiger Gigabyte (siehe z. B. (Crytek 2017)) simuliert werden.¹ Und nach zwei oder drei Jahren repräsentiert das Kind in seinem Gehirn die Realität der unmittelbaren Umgebung, es bewegt sich darin und handelt, spricht, unterhält komplexe soziale Wechselwirkungen und hat die Fähigkeit, nach Inspektion einmaliger Beispiele neue Objekte oder Vorgänge desselben Typs trotz grosser Variation in Detail und Perspektive zu erkennen. Ab vier oder fünf Jahren kann das Kind Situationen in ihrer Gesamtheit aufnehmen und über Tage oder gar Jahrzehnte erinnern. Es scheint, dass das Kind zu diesem Zeitpunkt dann eine innere Sprache zur Verfügung hat, mit der es die Essenz von Situationen ausdrücken und festhalten kann.

Alle Eltern wissen zudem, dass Kinder nicht als tabula rasa auf die Welt kommen, sondern Antriebe, Präferenzen und einen Willen haben, welche sie im Laufe des Aufwachsens in ihrem konkreten Umfeld mit Substanz füllen. Es gibt eine umfangreiche Literatur im Bereich der Kleinkindpsychologie oder der Verhaltensforschung (siehe z. B. (McFarland 1999)), die dieses Repertoire von Antrieben in grossem Detail beschreibt. Das Besondere daran ist, dass die Evolution in uns die einzelnen Verhaltensmuster in sehr abstrakter Form angelegt hat. Verhaltensforscher und Psychologen sprechen in diesem Zusammenhang von *Schemata*, die gerade so viel Struktur enthalten um in wirklichen Situationen ausgelöst werden zu können, darüber hinaus aber kein Detail, das ihnen relevante Situationen verschliessen würde. Diese Art des Antriebs durch ein Repertoire generischer Verhaltensmuster ist die Grundlage des autonomen Verhaltens von Tier und Mensch und unterscheidet diese von aller bisherigen künstlichen Intelligenz, die durch viel zu spezifische Zielsetzungen eingeengt ist.

Um allein die Schaltung des menschlichen Gehirns in Form einer Liste aller Verbindungen zwischen Neuronen zu beschreiben, braucht es ein Petabyte, was eine Million mal mehr ist, als der Informationsgehalt von Genom und kindlicher

¹Diese virtuelle Realität müsste als vollgültige Lernumgebung alle Sinne umfassen. Praktikabel und sinnvoll wäre ein solches Unterfangen nur für lernende Roboter. Das ändert aber nicht wesentlich die Abschätzung des benötigten Programmfangs.

Umwelt.² Wie kann ein so komplexes Gebilde durch so wenig Information erzeugt werden? Der richtige Informationsbegriff ist der von Kolmogorov (siehe (Li und Vitányi 2008)): Der Informationsgehalt einer Struktur ist gleich der bit-Länge des kürzesten Algorithmus, durch den die Struktur erzeugt werden kann. Der relevante Algorithmus für das Gehirn ist der *ontogenetische Mechanismus*. Seit einigen Jahrzehnten gibt es eine Vielzahl von experimentellen und theoretischen Untersuchungen zur genetischen Steuerung der Herstellung spezifischer Schaltungsmuster (Goodhill 2007; Grabska-Barwinska und der Malsburg 2008; von der Malsburg 1973, 1979). Der dabei entdeckte zentrale Mechanismus wird als *Netzwerk-Selbstorganisation* bezeichnet. Das Wesentliche dabei ist, dass die Menge der entstehenden Struktur-Information nicht durch die Menge der genetischen Information begrenzt ist.

Aus diesen Beobachtungen kann die Schlussfolgerung gezogen werden, dass das Gehirn zumindest des Dreijährigen zu einem überwältigenden Grade von einem regelmässigen Strukturprinzip beherrscht ist, das aus dem ontogenetischen Mechanismus, insbesondere Netzwerk-Selbstorganisation, verstanden werden kann.³

2 Lähmende Vorurteile

Der Versuch, die Funktion des Gehirns soweit zu verstehen, dass sie in wesentlichen Zügen elektronisch nachgebildet werden kann, steckt nach Meinung einer Reihe anerkannter Wissenschaftler in einem fehlangepassten Begriffsgebäude fest, einem veralteten Paradigma (Kuhn 1962) oder Denkstil (Fleck 1935). Keine Übereinstimmung gibt es dagegen, welches die Vorurteile sein könnten, die den Fortschritt lähmen. Hier sind drei mögliche Kandidaten: *Denken in direkten Algorithmen*, *agnostische Statistik* und die *Einzelzellhypothese*. Diese Vorurteile betreffen zwei fundamentale Fragen, a) wie entsprechen sich materielle Hirnzustände einerseits und mentale, introspektiv beobachtbare Zustände andererseits? und b) wie entstehen diese materiellen und mentalen Zustände? Die ersten beiden Unterabschnitte betreffen b), der dritte betrifft a).

2.1 Denken in Direkten Algorithmen

Zunächst soll auch hier das Credo akzeptiert werden, dass, was immer das Funktionsprinzip des Gehirns ist, es sich als Algorithmus auf einer universellen *Turing Maschine* (Turing 1936), auf einem Rechner genügender Kapazität und Rechenleis-

²Um eines der 1010 Neuronen zu adressieren braucht es $\text{ld}(1010) = 33$ bit. Für die 1014 Synapsen des Grosshirns braucht es daher 3.3×1015 bit.

³Es ist nicht zu erwarten, dass sich dies im Laufe der Jahre ändert, etwa das Erwachsenengehirn von von aussen absorbiert Information dominiert sein könnte. Es gibt Argumente und Abschätzungen, dass der Mensch nur etwa ein bit an Information pro Sekunde in sein permanentes Gedächtnis absorbiert (Landauer 1986).

tung, wird implementieren lassen. Alle Motivation, darüber hinaus zu gehen (etwa in Richtung auf Quantenphänomene) ist nur die Folge des allgemeinen Denkens im Rahmen eines viel zu engen Konzepts, dem *direkter* Algorithmen.

Dieses enge Konzept liegt im Wesentlichen allen digitalen Anwendungen zu Grunde. Nach ihm wird der Rechner als blinder Seiltänzer behandelt, dessen jeder Schritt durch den menschlichen Programmierer vorgegeben werden muss. Alle Vorstellungen von Sinn und Ziel des Prozesses sowie vom Umfeld, in dem er stattfindet, existieren allein im Kopf des menschlichen Programmierers, und entsprechend muss dieser auf alle zu erwartenden Wechselfälle explizit eingehen und diese berücksichtigen. Der Programmierer steuert durch den Algorithmus die auszuführenden Aktionen direkt.

Der Begriff hinter dieser Art von Algorithmen ist der der Maschine: Das System ist *passives* Werkzeug in der Hand des Nutzers und darf sich keinerlei Spontaneität erlauben. Dieses (gewollt) maschinenhafte Verhalten des im direkten Sinn programmierten Rechners und die Verinnerlichung der Tatsache, dass alles Verstehen, alle Einsichten, alle Willensentscheidungen von ausserhalb der Maschine stammen, sind der Grund für das generelle und tief in der Erfahrung verwurzelte Gefühl, dass im algorithmisch gesteuerten Computer Gefühl, Bewusstsein und autonome Intelligenz nie möglich sein werden.

Neben dieser direkten algorithmischen Prozesssteuerung lässt sich jedoch im Rechner auch ein ganz anderer Stil realisieren, ein Stil nach dem der vorgegebene Algorithmus *indirekt* wirkt und lediglich eine Architektur definiert, in der die eigentlichen konkreten Prozesse und Strukturen spontan wachsen. Diese werden dabei durch zwei Arten von sich ergänzenden Einflüssen gestaltet, durch *Vorgabe von Regeln* für die Wechselwirkung zwischen Systemelementen sowie durch *Vorgabe von generellen Zielen*. Im Falle des Gehirns wird Ersteres durch das Verhaltensrepertoire von Zellen und Synapsen definiert, letzteres durch Wertung von Gesamtsystemzuständen durch entsprechende Zentren. In einem solchen System spielen der Zufall, evolutionsartige Mechanismen und Selbstorganisation die zentrale Rolle. Wenn es mit einem reichen Sensorium und der Möglichkeit des Ansammelns von Erfahrung ausgestattet ist, wird an ihm nichts maschinenhaftes mehr sein, obwohl der zugrunde liegende Algorithmus deterministisch ist, und alle Argumente gegen den Rechner als Implementierungsmedium für Geist werden an einem solchen System abprallen.

Fixierung auf direkte Algorithmen und Ignorieren der Instrumente und Mechanismen der Erzeugung konkreter Prozesse und Lösungen durch Mechanismen der Selbstorganisation und durch Vorgabe genereller Ziele tragen zu einer allgemeinen Grundhaltung bei, nach der hinter jedem intelligenten Gebilde ein intelligenter Schöpfer stehen muss, ein Schöpfer, dessen kreative Mechanismen ausserhalb der Betrachtung bleiben, der also nicht selbst als Teil eines Gesamtsystems verstanden wird. Im Rahmen der Evolutionsdebatte wird hier von *Intelligent Design* gesprochen. Diese Haltung ist tief verwurzelt im Unterbewusstsein und ist ein starkes Hindernis auf dem Weg zum Verständnis von Gehirnfunktion und zu autonomer Intelligenz.

Ganz abgesehen von diesem grundsätzlichen Missverständnis sind direkte Algorithmen ein enorm ineffizientes Mittel, funktionelle kognitive Strukturen zu erzeugen. In diesem Stil etwas so Komplexes wie ein mit dem Menschen vergleichbares künstliches visuelles System zu schaffen, würde die ökonomische Leistungsfähigkeit der Welt überfordern, und dennoch hinter immer neuen Herausforderungen herhinken müssen. Die Evolution jedenfalls kann nach Auskunft der Knappheit der genetischen Information diesen Weg nicht beschritten haben. Es gibt zahlreiche Versuche, die Schwäche direkter Algorithmen zu überwinden.

Indirekte Algorithmen werden im Rahmen des sogenannten *Soft Computing* durch evolutionäre oder genetische Algorithmen sowie durch Schwarm-Optimierung oder -Intelligenz realisiert. Es wird versucht, die Zerbrechlichkeit digitaler Prozesse (und damit kritische Abhängigkeit von menschlicher Fehlerbehebung) abzumildern, in Ansätzen wie Fuzzy Logik oder Einbetten in Wahrscheinlichkeitsverteilungen, etwa im Rahmen von Bayes Statistik. Des Weiteren wird versucht, den Anteil direkt-algorithmischer Steuerung durch lernende Verfahren, im Rahmen neuronaler Netze oder maschinelles Lernen zu reduzieren (siehe den nächsten Abschnitt). Noch immer aber ist jede einzelne Anwendung angewiesen auf menschliche Einsicht in ihre Spezifika, etwa bei der Wahl der verwendenden Datenstruktur. Das Ziel muss daher sein, eine Architektur zu finden, die alle Arten kognitiver Prozesse zu implementieren gestattet, so wie wir es vom menschlichen Gehirn kennen.

2.2 Agnostische Statistik

Als grosser Gegenpol zum direkt-algorithmischen Ansatz künstlicher Intelligenz und als dessen Überwindung wird gegenwärtig das *maschinelle Lernen* hochgehalten. Dabei spielen indirekte Algorithmen eine wichtige Rolle, prominent darunter das sogenannte *Backpropagation of Errors* – nach dem ein neuronales System seine Verbindungen so ändert, dass die Abweichung zwischen Systemreaktionen und vorgegebenen Lösungen minimiert wird. Im Zentrum aber steht die statistische Ausnutzung grosser Datensätze von Beispiel-Eingaben, jeweils gepaart mit von Menschen (oder von Algorithmen) erzeugten korrekten Ausgaben. Im wirtschaftlichen Sinne wird diese Technologie gegenwärtig als Durchbruch angesehen und sie wird, besonders unter Einbettung in direkt-algorithmische Umgebungen, über die kommenden Jahre sicher noch grosse Erfolge feiern. Von kognitivem Rechnen oder von autonomer Intelligenz aber kann dabei nicht gesprochen werden.

Die Gültigkeitsbereiche der trainierten Systeme sind sehr eng an die Datensätze und die darin vorgegebenen Lösungen gekoppelt, und die zu verarbeitende Varianz muss voll im Trainingsdatensatz vorhanden sein. Wenn also das behandelte Phänomen in n Dimensionen variiert (bei Objekten auf Fotografien z. B. Position, Grösse, Orientierung, Farbe, Beleuchtung oder Deformation), so muss der Trainingsdatensatz diesen ganzen n -dimensionalen Raum mit Beispielen ausfüllen. Der Ansatz des maschinellen Lernens verzichtet gänzlich darauf, die Mechanismen zu modellieren, die aus wenigen Beispielen den Raum aller Varianten selbst erzeugen und

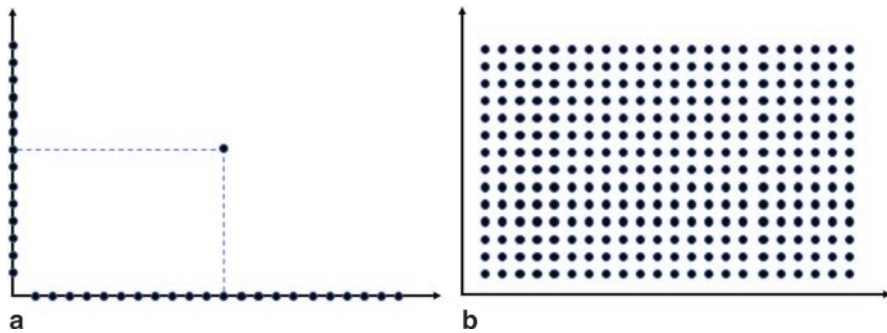


Abb. 1 Kompositionale oder ganzheitliche Darstellung

verarbeiten könnten (so, wie es die Computergrafik tut), also etwa ein einmal gesehenes Objekt in allen Positionen, Grössen, Farben und Beleuchtungen auf Grund seiner inneren Struktur wiedererkennen zu können. Bezüglich der Notwendigkeit einer solchen domänenspezifischen Modellierung, siehe Geman, Bienenstock und Doursat (1992).

Diese Fähigkeit, aus wenigen oder gar einzelnen Beispielen auf einen grossen Varianzbereich zu schliessen, ist im menschlichen Gehirn offensichtlich angelegt, etwa einen einmal gesehenen Objekttyp in allen Varianten und Posen etc. wieder erkennen zu können. Dieses *Verallgemeinerungsvermögen* ist ein wesentliches Merkmal von Intelligenz und Kognition. Es beruht auf der Fähigkeit, die verschiedenen Varianz-Dimensionen oder Aspekte eines Phänomenbereichs voneinander abzutrennen und einzeln zu modellieren (vgl. Abb. 1).

In kompositionaler Darstellung wird ein komplexes Gebilde (symbolisiert durch einen Punkt in Teil a der Abbildung) zusammengesetzt aus Teilaspekten, etwa Objektform oder -farbe (symbolisiert durch die Punkte auf den Achsen, von denen jeder selbst ein hochdimensionales Muster darstellt). Neue Kombinationen können in dem Fall aus zuvor gelernten Teilaspekten zusammengesetzt werden. Wenn dagegen, wie dies beim *Deep Learning* der Fall ist, jede Kombination von Teilaspekten ganzheitlich separat dargestellt wird (die vielen Punkte in Teil b der Abbildung), müssen alle diese Kombinationen beim Lernen mit einiger Abtastdichte einzeln zur Kenntnis genommen werden, was bei hoher Anzahl von Dimensionen völlig impraktikabel wird. Das Abtrennen von Teilaspekten bedeutet Abstraktion, indem jeder Aspekt vom Einfluss der anderen bereinigt ist. Ein spezifisches Phänomen ist dann als Verschränkung aller beteiligten Aspekte zu verstehen.

Diese Art der Modellierung ist selbstverständlich *domänenabhängig* (siehe die *no-free-lunch Theoreme* (Wolpert und Mcready 1997)). Für visuelle Szenen ist sie im Rahmen der Computergrafik konsequent durchgeführt worden. Anscheinend aber gibt es einen sehr viel grösseren Bereich von Phänomenen, die in diesem Sinne abgedeckt werden können, den Bereich der *Kognition*, das heisst, den Bereich der unserem Geist zugänglichen Phänomene. Dieser Bereich ist offensichtlich durch eine besondere Struktur charakterisiert, die von der Architektur unseres Gehirns erfasst wird. Dagegen sind aber die Werkzeuge, die im Bereich der agnostischen,

domänenunabhängigen, Statistik zur Verfügung stehen, viel zu stumpf, um dem Phänomen der Intelligenz oder der Kognition auch nur nahe zu kommen. Auch hier wieder ist es die allgemeine Vernachlässigung eines wichtigen Denkbereichs die den Fortschritt wesentlich behindert.

2.3 Die Einzellhypothese

Von fundamentaler Bedeutung für das Verständnis von Gehirn und Kognition ist der neuronale Code, die *Sprache*, in der die neuronalen Zustände und Prozesse im Gehirn die von uns erlebten mentalen Phänomene darstellen: die Frage nach dem Datenformat der Kognition. Erstaunlicherweise taucht diese Frage in Diskussionen unter Neuro- und Kognitionswissenschaftlern gegenwärtig kaum auf. Der Grund dafür ist, dass sie allgemein als längst beantwortet angesehen wird. Danach lassen sich die geistigen Phänomene in Elementarsymbole zerlegen, diese Elementarsymbole entsprechen den Nervenzellen, und der mentale Inhalt des Gehirns ist in jedem Moment vollständig durch die Menge der gerade aktiven Zellen beschrieben, deren jede ihren symbolischen Inhalt additiv beiträgt.

Für die Interpretation von Neuronen als Elementarsymbole spricht eine unübersehbare Zahl von elektrophysiologischen Versuchen, gemäss derer jeder Zelle ein auslösender Reiz oder eine durch sie ausgelöste motorische Aktion zugeordnet werden kann. Diese Version des neuronalen Codes war immer wieder der Rahmen für Versuche, Kognition oder psychophysische Effekte zu modellieren und ist die Grundlage aller künstlichen neuronalen Netze inklusive Deep Learning.

Diese Version des neuronalen Codes hat jedoch das Problem, grob unvollständig zu sein. Jedes je vom Menschen benutzte Sprach- oder Symbolsystem hat zusätzlich zu Elementarsymbolen (Phonemen, Buchstaben, Schriftzeichen, Noten, Pixel etc.) ein syntaktisches Mittel, diese Elementarsymbole zu komplexeren Symbolgebilden (Worte, Tabellen, Formeln, Partituren, Bilder etc.) zusammenzusetzen. Üblicherweise ist dieses syntaktische Mittel Anordnung im Raum oder in der Zeit. Dieses Mittel lässt sich leider aber nicht auf das Gehirn anwenden, da die Anordnung der Neuronen nicht veränderlich ist. Auf diese Schwierigkeit der künstlichen neuronalen Netze ist schon vor langer Zeit hingewiesen worden (von der Malsburg 1981) und wird heute allgemein als Bindungsproblem bezeichnet (Roskies 1999). Das einzige Mittel des Codes der künstlichen neuronalen Netze, aus elementaren Symbolen komplexere Entitäten zu bilden ist die gleichzeitige Aktivierung, das Feuern in *Assemblies* (Hebb 1949) oder *Pools* (d. h. Mengen gleichzeitig aktiver Neuronen). Das Problem der *Assemblies* oder *Pools* ist, dass das gleichzeitige Aktivieren von mehreren solcher diese unvermeidlich zu einer undifferenzierter Menge verschmelzen lässt, dass es also im Rahmen dieses neuronalen Codes kein Mittel gibt, die Zugehörigkeit zu einzelnen *Assemblies* zu markieren (die sog. *Superpositions-katastrophe* (von der Malsburg 1986)).

An der Frage nach dem neuronalen Code hat sich seinerzeit auch die heftige Diskussion *Symbol-basierte versus neuronale Verarbeitung* entzündet. Viele für die Symbolverarbeitung unverzichtbare Operationen sind im Rahmen des Codes der

künstlichen neuronalen Netze nicht möglich. Infolgedessen kann auf dieser Basis nichts realisiert werden, was in irgendeinem Sinne *Cognitive Computing* genannt werden könnte.

3 Vom Ursprung des Geistes

Wenn Intelligenz und Geist nicht das Ergebnis von Intelligent Design sind, also eingehaucht von einem separaten intelligenten Agenten (wie im Falle der direkten Algorithmen), woher kommen sie dann?

Diese Frage kann auf dieselbe Stufe gestellt werden wie die nach dem Ursprung mathematischer Konstrukte, wie etwa der natürlichen oder komplexen Zahlen, der Euklidischen Geometrie oder der Gruppentheorie. Mathematiker haben angesichts dieser Frage das Gefühl, diese Strukturen nicht zu erfinden, sondern vielmehr, sie zu entdecken, und dass diese Strukturen unabhängig von ihrem Tun vorhanden sind, so wie die Himmelskörper vorhanden sind, unabhängig davon ob sie betrachtet werden. Von dem Moment an, da die Definition der mathematischen Gruppe hingeschrieben war bis zur vollständigen Klassifikation etwa der endlichen simplen Gruppen (einem gewaltigen Stück Arbeit, dessen Veröffentlichung geschätzte 17.000 Journalseiten einnimmt), war der Weg im Detail unausweichlich vorgezeichnet. Das alle mathematischen Strukturen beherrschende Prinzip ist *Widerspruchsfreiheit, Konsistenz*. Wird eine mathematische Struktur aufgefasst als eine grosse Menge spezifischer Strukturen und Aussagen, die durch ein Netzwerk von Beweis- oder Schlussketten miteinander verbunden sind, so bedeutet Konsistenz, dass es nie vorkommen kann, dass auf zweierlei Wegen eine Aussage und ihr Gegenteil bewiesen werden.

Es gibt mathematische Strukturen, die sich *materialisieren*, etwa als elektromagnetische oder akustische Wellenmuster, als Elektronenorbitale um Atomkerne, als periodische Kristallgitter oder als kugelförmige Seifenblasen. Eine Art, diese zu verstehen ist, sie als raum-zeitliche Netze von Wechselwirkungen zu sehen, die die Tendenz haben, sich in optimalen Konfigurationen zu stabilisieren, optimal im Sinne grösstmöglicher Konsistenz und Widerspruchsfreiheit aller beteiligten Wechselwirkungsketten. Lehrbücher der theoretischen Physik sind dominiert von solchen Optimalitäts- oder Extremalprinzipien.

Das grösste Wunder dieser Welt ist, wie Einstein sagte, dass wir sie überhaupt verstehen können. Dieses Wunder kommt wohl dadurch zu Stande, dass sowohl Welt wie Geist, ihrer Natur nach, dichte Netze von Wechselwirkungen sind, die die Tendenz haben, in optimale Konfigurationen zu fallen. Das Universum dieser Konfigurationen ist hoch strukturiert, und im Ozean aller möglichen Strukturen machen sie einen verschwindend kleinen Teil aus.

Eine in diesem Zusammenhang mögliche Haltung ist, von einer *immateriellen Prä-existenz* dieser Strukturen auszugehen und ihnen die Fähigkeit zuzugestehen, unter gewissen günstigen Bedingungen materielle Systeme unter ihren Einfluss zu bringen. So gelingt es etwa der Idealgestalt Kugel, sich unter gewissen günstigen

Bedingungen auf der Basis membranbildender Moleküle zu materialisieren, oder den Idealgestalten periodischer Kristallgitter – von denen es nach einer im neunzehnten Jahrhundert durchgeführten mathematischen Klassifikation eine endliche Zahl gibt, etwa 235 – sich auf der Basis geeigneter Moleküle als Mineralien zu materialisieren.

Dieses Denkmuster ist insbesondere auch auf das Leben anzuwenden, obwohl dort sehr viel komplexere und variabelere Idealgestalten angetroffen werden, die etwa die molekularen Netze der lebenden Zelle beherrschen, oder die Baupläne der Organismen und Organismengruppen (von der Malsburg 1987). Im Bereich der letzteren hat die Biologie in den vergangenen Jahrzehnten erhebliche Fortschritte gemacht und genetische Kontrollstrukturen aufgedeckt, die die Ontogenese in Kaskaden von Differenzierungsprozessen steuern und in wesentlichen Zügen quer durch das Reich der Organismen erhalten sind (Bürglin und Affolter 2016).

Von unmittelbarer Relevanz im gegenwärtigen Zusammenhang sind die netzartigen Idealgestalten, die sich dem materiellen Substrat Nervensystem im Laufe der Ontogenese (der Entwicklung des Organismus aus dem Ei), des Lernens und des Denkens aufprägen. Dazu haben diese Strukturen im Laufe der Evolution das Verhaltensrepertoire der Neuronen dahingehend differenziert, dass es das Entstehen *kognitiver Netzstrukturen* ermöglicht, analog dem Verhalten von Seifenmolekülen, das dem Entstehen von Lamellen, Oberflächenspannung und Kugelgestalten zu Grunde liegt.

Dieser Prozess der Netzentstehung ist im Rahmen der Ontogenese an einigen prominenten Beispielen von der Biologie vor einiger Zeit mit grosser Intensität experimentell und theoretisch erforscht worden. Im Fokus stand dabei das Entstehen retinotoper Faserprojektionen (siehe z. B. (Goodhill 2007; Häussler und von der Malsburg 1983; Willshaw und von der Malsburg 1976, 1979)), neben geordneten Merkmalskarten (z. B. (Grabska-Barwinska und von der Malsburg 2008; Hubel und Wiesel 1977; von der Malsburg 1973, 1994)).

Das generelle Schema dieses Prozesses der Netzwerk-Selbstorganisation ist einfach: Eine gegebene Menge von Neuronen ist stochastisch aktiv, dieser stochastischen Aktivität werden durch die existierenden neuronalen Verbindungen Korrelationen aufgeprägt, und die Korrelationen wirken unter Nebenbedingungen plastisch auf die Stärke der Verbindungen zurück. Die kanonische Form dieser Rückwirkung ist die sogenannte *Hebbsche Plastizität*, nach der positive Signalkorrelation zweier Zellen eine Verstärkung ihrer Verbindung bewirkt. Nebenbedingungen begrenzen diesen Effekt und bewirken Spärlichkeit des Netzes, das heisst, Begrenzung auf wenige konvergierende Verbindungen pro Zelle. Als Ergebnis dieser Wechselwirkung zwischen Verbindungstärken und stochastischer Aktivität werden gewisse Netze stabilisiert. Die entstehenden Netze sind optimal im Sinne der wechselseitigen Konsistenz alternativer Signalwege, in vager Analogie zu den widerspruchsfreien Schlussketten mathematischer Aussagenetze.

4 Der Neuronale Code

Der neuronale Code stellt die materielle Basis dar, auf der kognitive Strukturen und Prozesse sich realisieren. In den vorausgegangenen Abschnitten wurden zwei Aspekte des neuronalen Codes ins Zentrum gestellt: *die von ihm verkörperte Datenstruktur* und *die Mechanismen seiner Organisation*. Auf beide habe ich bereits partielle Antworten gegeben. Danach hat die Datenstruktur die Form strukturierter Netze, für deren Organisation Netzwerk-Selbstorganisation eine wesentliche Rolle spielt. Nach wie vor spielen die einzelnen Neuronen die Rolle von Elementarsymbolen oder Merkmalserkennern, die jedoch im Rahmen von Netzwerken in definiertem Kontext auftreten.

Die permanente (durch Lernen langsam veränderliche) Verschaltung des Gehirns ist nach diesem Konzept aufzufassen als eine Überlagerung von sehr vielen durch Netzwerk-Selbstorganisation strukturierten Netzfragmenten. Diese können auf dem funktionellen Zeitmassstab, also innerhalb von Sekundenbruchteilen, aktiviert und deaktiviert werden.

Unter den sicher ungeheuer vielgestaltigen Netzstrukturen, die durch Netzwerk-Selbstorganisation erzeugt und stabilisiert werden können, stechen zwei als besonders einfach hervor: *zwei-dimensionale Felder von Neuronen mit kurz-reichweitigen erregenden Verbindungen* und *topologische (Nachbarschaft-erhaltende) Faserprojektionen* zwischen zwei solchen Feldern. Neuronale Felder werden häufig beschrieben (Amari 1977; von der Malsburg 1973; Wilson und Cowan 1972), werden allerdings üblicherweise nicht als Datenstruktur in sich angesehen, sondern nur als Substrat für die Erzeugung lokaler Aktivitätsgipfel (für ein Beispiel siehe (Erlhagen und Bicho 2006)).

Aufgefasst als Netz sind sie Attraktoren im Sinne von Netzwerk-Selbstorganisation (Jug et al. 2012; Stringer et al. 2002). Zu einer reichhaltigen Datenstruktur, insbesondere im visuellen System, werden neuronale Felder, wenn an jeder Stelle des Feldes eine Vielzahl von Neuronen zur Auswahl stehen, so dass eine kombinatorische Vielfalt von Netzen mit derselben Grundstruktur darin angesiedelt werden kann (siehe Abb. 2).

Ein einfaches Modell kann dies illustrieren. Gegeben eine eindimensionale Kette von Kolumnen, von denen jede von einer Menge von Neuronen besetzt ist. Die einzelnen Neuronen seien mit Buchstaben des Alphabets bezeichnet, und jede Kolumne enthält für jeden Buchstaben eine Anzahl alternativer Neuronen. Dann kann

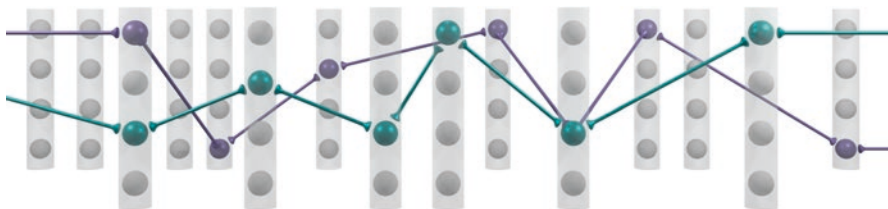


Abb. 2 Überlagerung neuronaler Felder (Abb. nach von der Malsburg 2018a)

eine grosse Zahl von Buchstabenketten in dieses Feld eingeschrieben werden, indem innerhalb jeder Kette die teilnehmenden Buchstabenneuronen über einige Distanz mit ihren Nachbarn im Sinne der Kette erregend verbunden werden. Wenn jedes einzelne Neuron an mehreren Ketten teilnehmen kann, so kann das System viele Buchstabenketten speichern. Jede einzelne von ihnen kann durch Aktivierung der teilnehmenden Neuronen aktiviert werden, ohne durch die Existenz anderer Ketten gestört zu werden (von der Malsburg 1985; von der Malsburg und Bienenstock 1987). Hilfreich ist dabei eine zwischen den alternativen Neuronen derselben Kolumne agierende *winner-take-all* Hemmung, die im Endzustand nur noch ein Neuron als aktiv zulässt.

Im Bereich des Sehens geht es um zweidimensionale neuronale Schichten (siehe Abb. 2) und um visuelle Merkmale (durch vertikale Rechtecke, *Kolumnen*, dargestellt, deren jede mehrere alternative Neuronen von identischer Merkmalspezifität enthält). Der sensorische Eingang aktiviert eine Untermenge der Merkmalskolumnen (vergrösserte Rechtecke). Ein solches Eingangsmuster kann nach (anfänglich zufälliger) Auswahl je eines Neurons pro aktiver Kolumne durch laterale Vernetzung über einige Entfernung gespeichert werden. Ein einzelnes Neuron kann mehreren dieser Netze angehören (siehe das Neuron in der fünften Kolumne von rechts in Abb. 2), und eine Kolumne kann mehreren Netzen angehören (siehe die dritte Kolumne von links in Abb. 2).

Angenommen das Gesamtsystem ist analog den kortikalen Arealen des Grosshirns in Untersysteme der gerade beschriebenen Art. Angenommen weiter, dass nach einigem Lernen und Netzwerk-Selbstorganisation die lateralen Verbindungen so strukturiert sind, dass jede Schicht als Überlagerung vieler neuronaler Felder aufgefasst werden kann, wobei jede der Zellen mehreren dieser Felder angehören kann, diese also ein verzweigtes System bilden. Vom Standpunkt des kognitiven Rechnens sind Netzfragmente strukturierte Bausteine, die zu grösseren Gebilden kombiniert werden können, wenn sie im Sinne der Konsistenz zusammenpassen.

Der Prozess der Netzaktivierung läuft im ganzen Gehirn in rascher Folge wellenartig immer und immer wieder ab. Die Konsistenz der Signalwege zwischen den feuernden Zellen in einzelnen Arealen, zwischen Arealen und insbesondere auch mit den schnell wechselnden Sinnessignalen ist vielleicht nie perfekt, wird aber ständig durch plastische Veränderung der Verbindungen verbessert. Dies ist der Prozess, der den netzartigen Idealgestalten des Geistes die Möglichkeit gibt, sich durch Netzwerk-Selbstorganisation zu etablieren.

5 Kognition

Längerfristiges Ziel des kognitiven Rechnens muss es sein, einen Überblick zu gewinnen über typische Strukturen und Prozesse auf der Basis des neuronalen Codes und diese auf elektronischer Basis nachzubilden. Hier seien kurz einige Beispiele diskutiert, die in ihrer Gesamtheit wesentliche Aspekte der Kognition wiedergeben:

5.1 Aufmerksamkeitsfokus, Situationsbewusstsein

Das Spezifische am *Aufmerksamkeitsfokus* ist, dass er eine Auswahl an mentalem Inhalt sowohl in den sensorischen Eingängen als auch im Gedächtnis aktiviert und die aktivierten Elemente funktionell und inhaltlich in kohärenter und konsistenter Weise miteinander vernetzt. Üblicherweise verfolgt das System ein momentanes Ziel, eine Agenda, und die Auswahl des Aufmerksamkeitsfokus muss insbesondere unter dem Blickwinkel dieser Agenda Sinn machen. Wenn von *Situationsbewusstsein* gesprochen wird, dann ist damit die Repräsentation und wechselseitige Vernetzung aller für eine Agenda relevanten Aspekte der gegebenen Situation gemeint.⁴

Das kognitive System muss einen langen Lernprozess durchführen, um schliesslich für alle möglichen Situationen jeweils geeignete Agenden und für diese Agenden geeignete Komponenten auswählen zu können, das heisst, umfassende Netze zur Verfügung zu haben, die situationsabhängig aktiviert und koordiniert werden können. Das geschieht nicht mit starren, globalen Netzen, da jede Situation im Allgemeinen ihr eigenes umfassendes Netz braucht, das sich selbst aus vorher gelernten Netzfragmenten verschiedener Abstraktionsstufe zusammensetzt. Das Endresultat ist die momentane Spezialisierung des Gesamtsystems auf ein konkretes Netz, das alle funktionellen Teile, Sinne, Planung und Effektoren so miteinander verbindet, dass das System gemäss der Agenda reagieren kann, das heisst, in diesem Moment als situationsspezifisches *Reiz-Reaktions-System* wirkt. So konfiguriert zum Beispiel der Fahrer eines Wagens einen grossen Teil seines Gehirns als eine Art von Autopilot, der in jeder Sekunde bereit ist, auf die gegenwärtige Situation durch Blicksteuerung und Bedienung von Steuerrad und Hebeln zu reagieren und dabei seine übergeordnete Agenda – Navigation auf ein geografisches Ziel hin – zu verfolgen.

5.2 Wahrnehmung

Zentral für das Gehirn ist seine Fähigkeit, in seinem Inneren eine Repräsentation der unmittelbaren Umwelt aufzubauen. Dies geschieht auf der Basis von Sinneseindrücken, welche jedoch nur sehr mittelbar mit den zu erschliessenden Bestimmungsstücken zusammenhängen. Das Feld reflektierten Lichtes wird zwar durch die Linsen unserer Augen zu retinalen Bildern fokussiert, deren jeder Punkt einer Richtung im Raum entspricht, aber erstens werden diese Bilder von den Bewegungen des Auges und des Kopfes mitbestimmt, und zweitens liefern die Bilder nur unvollständige Information über die Welt da draussen. Unser Tastsinn ist zwar direkter als der visuelle, aber ein ertasteter Ort der Umwelt ist in komplizierter Weise durch die Lage unseres Körpers und unserer Extremitäten kodiert. Der Zusammenhang zwi-

⁴Bewusstsein bedeutet übrigens allgemeiner die Aktivierung eines Quorums von funktionellen Komponenten (etwa Agenda und Emotionen, Repräsentation der eigenen Person, Repräsentation der physischen Umgebung, Gedächtnis, Sprache) und die Integration derselben zu einem funktionellen Ganzen (von der Malsburg 1997).

schen kinesthetischen und den visuellen Repräsentationen beim Berühren eines Punktes der Umwelt mit dem Finger oder beim Fangen eines Balles ist also sehr komplex. Und dennoch gelingt es unserem Gehirn, derartige Daten zuverlässig in einer Darstellung unserer statischen oder gar bewegten Umwelt an der richtigen Stelle einer Repräsentation derselben einzuordnen und diese Darstellung auf Grund von sensorischen Eindrücken überhaupt erst einmal aufzubauen. Wie ist das möglich?

Eine fundamentale Voraussetzung für diesen Umgang mit der Umwelt ist die Fähigkeit, Abbildungen zwischen zwei- oder drei-dimensionalen Koordinatensystemen darzustellen und diese situationsabhängig zu verändern. Dazu gehören Abbildungen zwischen retinalen, Objekt- und Szenen-zentrierten Koordinatensystemen. Diese müssen mit der Geschwindigkeit der Augenbewegungen oder von Objektbewegungen gegenüber der Szene verändert werden. Ebenso muss sich die Abbildung zwischen Tastpunkten der Hautoberfläche und Punkten der Szenen- oder Objektdarstellung entsprechend der Körperposition und -haltung ändern.

Solche flexiblen Projektionen zwischen zweidimensionalen Koordinatensystemen sind im Rahmen eines Rechnermodells der Gesichtserkennung im Gehirn realisiert worden (siehe Abb. 3 (Wolfrum et al. 2008)). Aufgrund von Experimenten ist bekannt, dass retinale Bilder im primären visuellen Cortex als zweidimensional ausgedehnte Felder von aktivierten, texturspezifischen Neuronen repräsentiert werden (schematisch dargestellt durch einige Punkte in dem linken mit Retinabild unterlegten Teil der Abb. 3). Ebenso ist bekannt, dass die Neuronen dort durch massive laterale Verbindungen über einige Distanz vernetzt sind. Um ein bestimmtes Gesicht zu erkennen setzt das Modell voraus, dass in einem anderen Teil des

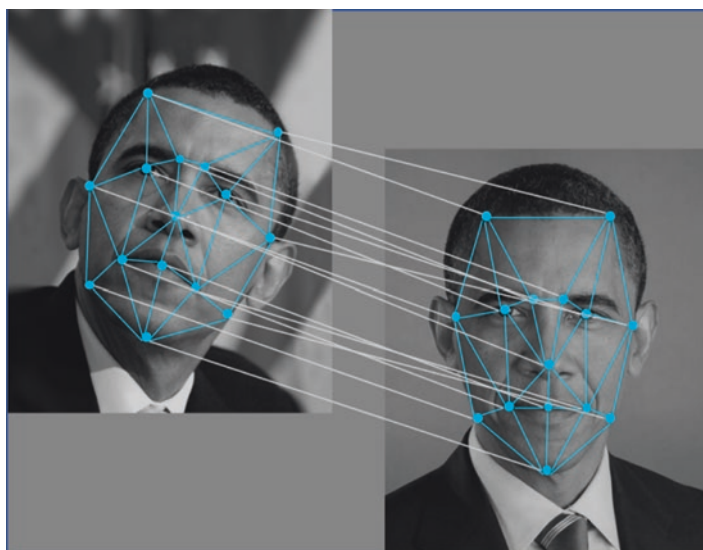


Abb. 3 Schema eines neuronalen Gesichtserkennungssystems (Abb. nach von der Malsburg 2018a)

Cortex (dem fusiformen Komplex) ein Modell dieses Gesichts in Form eines weiteren Netzes vorhanden ist, das eine dem ersteren Netz sehr ähnliche Texturverteilung darstellt (rechter Teil der Abb. 3). Die Ähnlichkeit von visuellem Reiz und gespeichertem Modell wird durch die Aktivierung eines Projektionsnetzes festgestellt (schräge Linien in Abb. 3), das korrespondierende Neuronen in den beiden Gesichtsnetzen miteinander verbindet. Insgesamt also wird der Erkennungsvorgang dadurch realisiert, dass drei Netzfragmente (das vom visuellen Reiz im primären visuellen Cortex ausgewählte, ein im Gedächtnis gebildetes im fusiformen Komplex und ein – seinerseits aus Fragmenten zusammengesetztes – Projektionsnetz) sich zu einem stabilen, in sich konsistenten Gesamtnetz zusammensetzen.

Strukturerhaltende Projektionsnetze können durch *Selbstorganisation* realisiert werden, siehe (Bergmann und von der Malsburg 2011; Fernandes und von der Malsburg 2015; Zhu et al. 2010). Ihre Aktivierung wird durch Kontrollneuronen (Anderson und Van Essen 1987; Hinton 1981; Kree und Zippelius 1988; Olshausen et al. 1993) gesteuert. Diese Kontrollneuronen stellen die Relativkoordinaten der Abbildungen dar. Zwischen den Motorsignalen der Augen- und Körperbewegungen einerseits und den Veränderungen der Relativkoordinaten der genannten Abbildungen andererseits bestehen feste Beziehungen. Auch diese können von dem System in Form von Netzen gelernt und dargestellt werden. Alle die Netze, die dieses Gewebe von Abbildungen und deren Veränderungen darstellen, bilden, wenn sie einmal korrekt eingerichtet sind, ein in sich konsistentes Gewebe von Verbindungen und Signalen. Dabei bedeutet Konsistenz, dass an allen Stellen im System Signale, die von einer Quelle ausgehen und über verschiedene Wege wieder an einer Stelle aufeinandertreffen, sich dort in Ort und Zeit genau treffen.

Dieses Gewebe hat Attraktoreigenschaft, das heisst, kleine Abweichungen von der Konsistenz werden durch den Selbstorganisationsprozess reduziert bis volle Konsistenz hergestellt ist. Wenn das System also unter genetischer Steuerung anfänglich diesem konsistenten Netzzustand schon einigermaßen nahekommt, so kann man erwarten, dass er durch Selbstorganisation schnell vollends hergestellt wird. In etwas vereinfachter Form ist dieser Prozess am Beispiel eines Roboterarms mit sieben Freiheitsgraden und einem pan-tilt Doppelkamerasystem demonstriert worden (Bolder 2005).

Zielpunkt der Netzorganisation ist, wie gesagt, Konsistenz zwischen konvergierenden Signalen. An den Schnittstellen zwischen den Modalitäten, etwa Sehen, Tasten und Kinesthetik, bedeutet Konsistenz, dass die dort von verschiedener Richtung kommenden zeitlich veränderlichen Signale miteinander übereinstimmen, sich gegenseitig vorhersagen können. An den Schnittstellen zwischen den Sinnen und dem Gehirn (also, in den primären sensorischen Kortexarealen) bedeutet Konsistenz, dass die sensorischen Signale vom Netz vorausgesagt werden. Wie ist das möglich?

Wie diskutiert, ist das Instrumentarium des Gehirns eine Überlagerung aktivierbarer Netzfragmente. Demnach müssen die genannten Vorhersagen durch die Aktivierung geeigneter Kombinationen von Netzfragmenten konstruierbar sein. Es stellen sich entsprechend folgende Fragen:

- 1) Sind solche Vorhersagen durch eine endliche Menge von Netzfragmenten darstellbar?
- 2) Wie werden im aktuellen Fall geeignete Kombinationen von Netzfragmenten aktiviert?
- 3) Wie können solche Netzfragmente entstehen?

Zu 1): Die Computergrafik zeigt, wie oben erwähnt, dass unendliche Vielfalten visueller Bildfolgen zusammengesetzt werden können aus einem sehr kompakten Vorrat von Grundstrukturen, wie Formen von Objekten, Texturen, Farben, Bewegungen, geometrischen Projektionen oder Beleuchtung. Die Computergrafik kann als Vorbild genommen werden für ein visuelles System, das die Bilder auf der Retina rekonstruiert und vorhersagt. Es ist durchaus vorstellbar, dass alle Komponenten eines Computergrafiksystems sich in die Form von Netzen übersetzen lassen, inklusive der geometrischen Projektionen. Wenn es auch für die anderen Modalitäten, Tastsinn, Motorsteuerung und Weiteres nicht ähnlich ausgearbeitete technische Modelle wie die Computergrafik gibt, so liegt es doch nahe, dass auch für diese eine Reduktion auf eine endliche Menge von Netzfragmenten möglich ist und dass auch deren kohärente Wechselwirkung sich so realisieren lässt, alles auf der Basis endlich vieler Netze und deren konsistente Zusammensetzung. Es kann daher für eine positive Antwort auf Frage 1) plädiert werden.

Die Diskussion der Frage 2) sei zunächst wieder auf das visuelle System beschränkt und auf Computergrafik bezogen. Dann stellt sich das Problem, dass die retinalen Bilder nur durch Invertierung des Grafikprozesses erklärt und vorhergesagt werden können, das heisst es muss zunächst der Weg von Bildern zu Bildinhalten gefunden werden. Dies ist in direktem Sinne nicht möglich, es handelt sich um einen sogenannten *inversen* Prozess (von der Malsburg 2018b). Die sensorischen Daten sind unvollständig und zunächst äusserst vieldeutig, und erst durch Konstruktion eines Szenenmodells, das in sich konsistent ist und auf dessen Basis die kurzfristige Veränderung sensorischen Eingänge zuverlässig vorhergesagt wird, entsteht Gewissheit bezüglich der Korrektheit der Rekonstruktion. Die Geschwindigkeit, mit der das Gehirn diese Aufgabe löst – typischerweise in einer Sekunde – zeigt, dass sie nicht durch sequenzielle Suche, sondern durch gleichzeitige Aktivierung aller mit dem sensorischen Eingang verträglichen Hypothesen und einen Auswahlprozess gelöst wird, der unter diesen eine Untermenge von miteinander konsistenten Hypothesen auswählt.

Im Rahmen des hier diskutierten neuronalen Systems nimmt dieser Prozess die folgende Form an. Der sensorische Eingang erregt eine grosse Untermenge aller Neuronen. Die Aktivität aller afferent erregter Zellen wird nach kurzer Zeit gelöscht (möglicherweise durch die experimentell beobachtete *balanced inhibition* (Isaacson und Scanziani 2011), durch die jede afferente Erregung nach kurzer Zeit annulliert wird), und nur die Zellen bleiben aktiv, die Teil aktivierter Netzfragmente sind, innerhalb derer sie sich gegenseitig anregen und unterstützen. Und auch diese Netzfragmente können sich nur halten, wenn sie Teil grösserer konsistenter Netze sind. Für eine Demonstration eines solchen Prozesses siehe Wolfrum et al. (2008).

Nun zur Frage 3), wie können die nötigen Netzfragmente entstehen? Hier sei eine vereinfachte Modellsituation betrachtet, in der es nur eine Eingangsschicht und eine kleinere Ausgangsschicht gibt, anfänglich in zufälliger Weise durch Fasern miteinander verbunden. In Modellrechnungen (Bergmann und von der Malsburg 2011; Fernandes und von der Malsburg 2015; Zhu et al. 2010) wurde gezeigt, dass ein solches System spontan oder bei Inspektion von Objekten gleichzeitig sowohl topologisch strukturierte Faserprojektionen samt Kontrolleinheiten zwischen den Schichten aufbauen kann als auch, während der Inspektion eines Objektes, ein statisches Modell der bewegten Eingangsbilder. Wesentliche Bedingung war für das System, dass die Aktivitätsverteilung in der Zielschicht während der Betrachtung eines Bildes unter Augenbewegungen fortgeschrieben wird, dass also während dieser Zeit Konsistenz zwischen den ankommenden und den in der Zielschicht fortgeschriebenen Signalen verlangt wird.

Das Prinzip also, das zuverlässige Perzeption ermöglicht, ist die Bedingung der Konsistenz von Signalwegen vom Objekt zu einer zentralen Darstellung desselben, Konsistenz sowohl im Raum (benachbarte Punkte des Objekts werden auf benachbarte Punkte der zentralen Darstellung abgebildet), als auch in der Zeit (ein Punkt des Objekts wird auch nach Augenbewegung und entsprechender Verschiebung des Bildes in Retina und Eingangsschicht auf denselben Punkt der zentralen Darstellung abgebildet).

5.3 Verstehen, Abstraktion

Ein Vorgang oder ein Objekt werden dadurch *verstanden*, dass sie auf eine auf das Wesentliche beschränkte Darstellung bezogen werden. Eine solche Darstellung wird im Allgemeinen *Schema* genannt. In der kognitiven Literatur im weiteren Sinne haben Schemata als Grundlage des Verstehens eine lange Tradition. Oft wird Kant (Kant 1781, Kap. *Vom Schematismus der reinen „Verstandesbegriffe“*) als frühe Referenz genannt, ebenso (Bartlett 1932). Die Grundideen der Gestaltbewegung der Psychologie können als Schema-basiert angesehen werden. Die Idee des Schemas war dominant in der künstlichen Intelligenz der 70er-Jahre (Arbib 1992; Riseman und Hanson 1992; Winograd 1971), manchmal auch unter anderen Bezeichnungen wie *Frames* (Minsky 1974) oder *Scripts* (Schank und Abelson 1977). Was noch immer fehlt ist eine konsequente Implementierung der Schema-Idee im Sinne des Cognitive Computing.

Um den Vorgang des Verstehens in einem kognitiven System zu realisieren braucht es eine Datenstruktur, die drei Dinge darstellen kann, *Inkarnate* (d. h. konkrete Realisationen) der zu verstehenden Dinge oder Vorgänge, *Schemata* sowie drittens die *Abbildungen* zwischen Schemata und Inkarnaten. Weiter müssen die Elemente der Datenstruktur mit *Wechselwirkungen* ausgestattet sein, unter deren Einfluss sie in Konfigurationen fallen, die dem Verstehen entsprechen. Letzteres wird vielleicht klarer anhand eines Beispiels.

Betrachtet wird ein konkretes Objekt, ein Dreirad (siehe Abb. 4), das als Bild (*Inkarnat*) in einer *Bilddomäne* erscheint. Um das Objekt zu verstehen, wird es auf

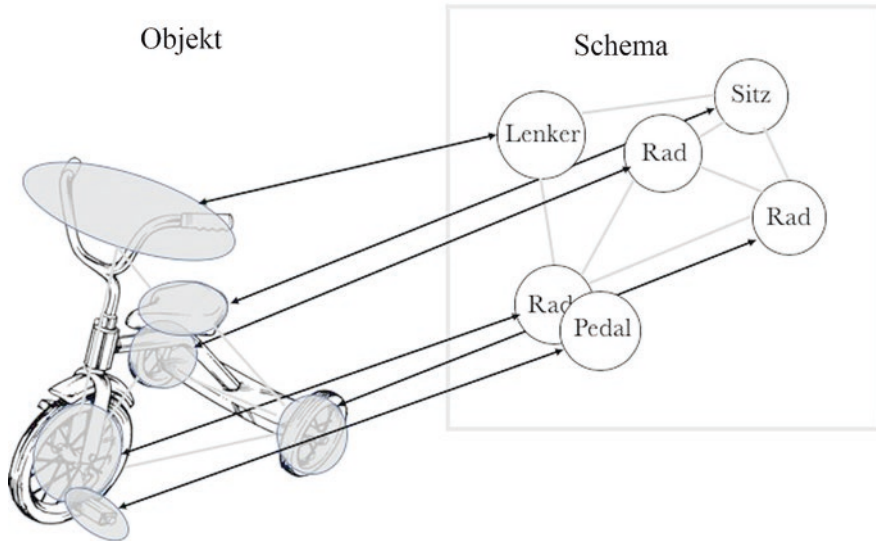


Abb. 4 Verstehen eines Objektes durch Bezug auf ein Schema

ein Schema in einer *Modelldomäne*, einem Teil des Gedächtnisses, bezogen. Schemata können als Graphen aufgefasst werden, deren Knoten sich auf Objektteile und deren Kanten sich auf Relationen zwischen ihnen beziehen. Im Prozess des Verstehens wird ein geeignetes Schema aus dem Gedächtnis aktiviert oder im Prozess aus vorhandenen Teilgraphen zusammengesetzt, und zwischen Inkarnat und Schema werden Knoten gleichen Typs so miteinander verbunden, dass in der Bilddomäne verbundene Knoten auf in der Modelldomäne verbundene Knoten zu fallen kommen. Eine solche Abbildung wird als *Homeomorphie* bezeichnet.

Das Besondere am schematischen Verstehen ist die *Abstraktion*. Nur die in gewissem Sinne wesentlichen Teile und Beziehungen des zu verstehenden Gebildes finden Eingang in das Schema. Wesentlich sind die Teile und Beziehungen, die allen Beispielen des Gebildes gemeinsam und für den Umgang mit ihm relevant sind.⁵

Schemata sind üblicherweise aus Sub-Schemata aufgebaut und in übergeordnete Schemata eingebettet. In diesem Beispiel, dem Dreirad, sind die Knoten ihrerseits Schemata für Teile (Räder, Lenkstange, Sattel etc.), und sein Schema ist selbst Inkarnat von höhergeordneten Schemata – Fahrzeug, Gebrauchsgegenstand oder, in den Augen des Kindes, Spielzeug – und es kann in Relation gesetzt werden zu dem Körperschema des Benutzers und zu Schemata für den Gebrauch.

⁵Im Fall des Verstehens von Bildern dreidimensionaler Objekte gehört dazu auch Abstraktion von der *Pose*, der dreidimensionalen Orientierung des Objekts. Nach einem Vorschlag zur Behandlung dieses Problems, siehe z. B. Biederman (1987), werden Objekte gedacht als in ein Alphabet bekannter Objektkomponenten (genannt *geons*, *primitives* oder *generalized cylinders*) zerlegt, für die das System bereits gelernt hat, wie sich ihr Aspekt und wie sich ihre Lagebeziehungen bei Tiefenrotation verändern.

Die Fähigkeit des Gehirns, konkrete Situationen auf abstrakte Begriffe abzubilden ist von der Technik weitgehend unverstanden. Dies ist ein erheblicher Stolperstein auf dem Weg zur Nachbildung autonomer Intelligenz. Als Folge dieses Unvermögens muss die Technik die zu behandelnden Situationen und auszuführenden Aktionen viel zu spezifisch beschreiben, was die Verallgemeinerung auf unerwartete neue Situationen verhindert. Tiere sind in ihrem Verhalten noch weitgehend beschränkt auf artspezifische Umgebungen und Schlüsselreize. Dies gilt anfänglich auch für menschliche Säuglinge, aber Kinder lernen dann mit ungeheurer Geschwindigkeit, die vererbten Schemata zu verallgemeinern oder zu differenzieren sowie auf die aktuelle Umgebung einzustimmen, siehe den Abschnitt *Lernen*.

Die Fähigkeit der Entschlüsselung und Kommentierung der aktuellen Situation durch Schemata erlaubt es auch, *angeborene oder erlernte Verhaltensmuster* zu aktivieren, so dass sie steuernd eingreifen können. Die angeborenen Verhaltensschemata haben zwei Aspekte, einen für das Erkennen relevanter Situationen, einen zweiten für die Aktivierung entsprechenden Verhaltens. Zu letzterem gehören direkte Aktionen, etwa die Auslösung einer Flucht- oder Angriffsreaktion, aber auch humorale und körperliche Reaktionen wie Ausschüttung von Adrenalin oder Erregung von Emotionen, und insbesondere auch Rückwirkungen auf Steuerung des Hirnprozesses und von Lernen.

Verkettete Schemata stellen ganze Programme dar, wie zum Beispiel: *Appetenzverhalten* → *Identifizieren geeigneter Beute* → *Verfolgung* → *Fang* → *Verzehr*. Schemata setzen sich zu komplexen Vorstellungen zusammen und ihre Verkettung miteinander, mit sensorischen Eingängen, mit Körperreaktionen und mit aktiven Handlungsbewegungen verleiht unserem Leben Bedeutung, Zielrichtung und emotionalen Gehalt.

6 Implementierung von Schema-Verstehen

Wie bereits erläutert, ist es im Sinne des *Cognitive Computing* erforderlich, den Mechanismus des schematischen Verstehens konkret zu implementieren. Wie leistet der neuronale Code dies?

Gegeben sei ein Inkarnat I und ein Schema S . Ausgegangen wird von dem Endprodukt, dem mit dem Inkarnat homeomorph vernetzten Schema. Dieses Gesamtgebilde wird dargestellt durch ein Netz von Neuronen, das im Sinne der Netzwerk-Selbstorganisation ein Attraktor ist. Es wird angenommen, dass sowohl S selbst als auch ein geeigneter Subgraph des I darstellenden Netzwerks Netz-Struktur haben, das heisst, für sich genommen Attraktoren von Netzwerk-Selbstorganisation sind. Es wird weiter angenommen, dass es ein Netz IS gibt, das diese beiden homeomorph miteinander verbindet. Das heisst, dass die Fasern dieses Projektionsnetzes ein-eindeutig Zellen des Subnetzes von I mit Zellen im Netz von S verbinden, die jeweils die gleiche Bedeutung (Merkmal, Etikett, ...) tragen, und zwar so, dass in I verbundene Zellen auf in S verbundene Zellen abgebildet sind.

Wenn I von aussen angeregt wird, dann erzeugt es Aktivität in all den vielen Zellen, die durch von I ausgehenden Verbindungen getroffen werden. Dazu muss

angenommen werden, dass Projektionsfasern im Ruhezustand leiten und erst im Laufe von Netzaktivierung von der Aktivierung ihrer Kontrolleinheiten abhängig werden. Wenn die eben genannten drei Netze als Teilnetze der Gesamtverschaltung existieren, dann werden sich innerhalb der anfänglichen Aktivitätswolke nach kurzer Zeit trotz der genannten *balancing inhibition* (Isaacson und Scanziani 2011) nur noch die Zellen aktiv halten, die Teil dieser drei Netze sind. Das ist der Vorgang, durch den *I* durch *S* interpretiert wird.

Eine weitere Frage ist, wie die eben vorausgesetzten Netze entstehen. Die Entstehung von Schemata ist übrigens in der kognitiven Literatur eine weitgehend offene Frage. Nach Auskunft der Verhaltensforschung (siehe z. B. (McFarland 1999)) werden Tiere bereits mit elementaren Verhaltensmustern geboren, nach unserer Interpretation also mit Schemata. Diese werden im Verlauf von Erfahrungen weiter differenziert und helfen, weitere Schemata ins Leben zu rufen. Für uns Menschen spielen für diesen sekundären Prozess Anleitung, Erziehung und Unterricht eine zentrale Rolle. Wie können neue Schemata entstehen?

Es wird angenommen, dass unter Anleitung immer wieder verschiedene Beispiele *I* desselben Phänomens in den Aufmerksamkeitsfokus gerückt werden sowie gleichzeitig eine Zellmenge *S* als Rohling für das entsprechende Schema aktiviert wird (letztere vielleicht unter dem Einfluss der Nennung eines Namens), und dass zwischen den Domänen von *I* und *S* ein System von Projektionsfasern zur Verfügung steht, von Fasern also, die unter Netzwerk-Selbstorganisation das Potenzial zur Erstellung homeomorpher Abbildungen haben. Es wird weiter angenommen, dass Netzwerk-Selbstorganisation sowohl in der Eingangsdomäne als auch in *S* zu Strukturen konvergiert, die unter *Homeomorphie* aufeinander abbildbar sind. Im Sehsystem dominieren anscheinend Netze mit zweidimensionaler Topologie, also neuronale Felder, so dass diese Bedingung dort in natürlicher Weise gegeben wäre. Es wird entsprechend zunächst sowohl für *I* als auch für *S* vorausgesetzt, dass sie von Nachbarschaftsverbindungen in zweidimensionaler Topologie dominiert sind.

Unter Einfluss der Anleitung haben die in der Eingangsdomäne aktivierten Beispiele strukturelle Gemeinsamkeiten: *gleiche Merkmale* (durch Zellen repräsentierte Untermuster) in *gleicher Anordnung*. Jedes Mal, wenn ein Beispiel erscheint und gleichzeitig ein Netz derselben homeomorphen Struktur aktiviert wird, organisieren sich die Projektionsfasern ein Stück in Richtung auf topografische (Nachbarschafts-erhaltende) Ordnung. Diese Tendenz zum Bilden topografischer Ordnung (d. h. verbundene Zellen in *I* projizieren auf verbundene Zellen in *S*) dominiert am Anfang die Entwicklung, wenn die Zellen in *S* noch keine Etiketle tragen. Dieser Prozess ist in (Bergmann und von der Malsburg 2011; Fernandes und von der Malsburg 2015; Zhu et al. 2010) modellhaft realisiert worden. Erst im Verlauf dieses Prozesses akquirieren die Neuronen des Schemas ihre Bedeutung als Etiketle oder Merkmale, nämlich durch zuverlässige Verbindung durch Projektionsfasern mit entsprechenden Zellen der Inkarnate (die ihrerseits ihre Bedeutung durch die Struktur ihrer sensorischen Eingangsverbindungen erhalten haben).

Umgekehrt entwickelt das System auch eine starke Rückwirkung des Schemas auf die Netzstruktur der Inkarnate. Da das Schema sich nur auf das beziehen kann, was allen Beispielen des gemeinsamen Phänomens gemeinsam ist, übt seine Rück-

wirkung auf die Netze I einen vereinheitlichen Einfluss aus, indem die auf das Schema bezüglichen Merkmale und Relationen dort hervorgehoben und zum Schema homeomorph gestaltet werden.

Die hier gegebene Diskussion der Entstehung der für Schema-Anwendung nötigen Netze ging noch davon aus, dass die Netze in I und S die Form zweidimensionaler Felder haben. Gibt es aber vielleicht eine allgemeinere Netzstruktur die allen Begriffen und zugehörigen Schemata gemeinsam sowie stabil unter Netzwerk-Selbstorganisation ist? Vielleicht ist hierarchische Kompositionalität hierfür die entscheidende generelle Eigenschaft aller kognitiver Strukturen: Jede auf Begriffe zu bringende kognitive Struktur – Gegenstand, Vorgang etc. – lässt sich in Teile zerlegen, die sich ihrerseits in kleinere Teile zerlegen lassen, möglicherweise durch mehrere Stufen. Nun sei zusätzlich angenommen, auf jeder dieser Stufen gäbe es mehrere verschiedene Möglichkeiten der Zerlegung, oder, von unten hergesehen, auf jeder Stufe gäbe es mehrere Möglichkeiten, die gegebenen Teile zu grösseren zusammenzufassen. Solche alternativen Zerlegungen würden dem Ganzen der kognitiven Struktur einen gewissermassen topologischen Zusammenhang geben. Die Hypothese wäre dann, dass diese Zerlegungen allen Beispielen einer kognitiven Struktur sowie ihrem Schema gemeinsam ist, so dass die Komponenten dieser Zerlegungen sich sinnvoll aufeinander abbilden und miteinander identifizieren lassen.

Es lässt sich gegenwärtig schwer entscheiden, ob sich diese Hypothese halten lässt. Jedenfalls aber müsste sich diese hierarchische Zerlegung durch eine Netzstruktur ausdrücken lassen und diese müsste unter Selbstwechselwirkung, unter Netzwerk-Selbstorganisation, stabil sein. Letzteres könnte dadurch realisiert werden, dass die Hierarchie der Teile sich widerspiegelt in einer Hierarchie der Dichte von Verbindungen innerhalb der Teile, am dichtesten innerhalb der Teile feinsten Zerlegung, und schwächer auf gröberer Stufe.

Entsprechend wäre dann bei spontan erzeugter Aktivität die Wahrscheinlichkeit der gleichzeitigen Aktivierung am grössten innerhalb der feinsten Teile, und abgestuft kleiner auf gröberer Ebene, und im Prozess der Netzwerk-Selbstorganisation müssten sich diese Wahrscheinlichkeiten dann über synaptische Plastizität umgekehrt wieder in die Dichte der Vernetzung innerhalb der Teile auswirken, so dass die Verbindungsstruktur zum Attraktor wird.

7 Lernen

Lernen, die Extraktion von Struktur aus gegebenen Situationen für spätere Verwendung, ist ein aktiver Prozess. Eine Situation muss zunächst funktionell Sinn machen, muss interpretiert werden im Rahmen eines Vorgangs, die für den Vorgang wesentlichen Teile müssen hervorgehoben werden, in zeitüberspannender Weise, bevor signifikante Strukturelemente identifiziert und zunächst im Kurzzeitgedächtnis abgelegt werden können.

Unser Kurzzeitgedächtnis, unser Arbeitsspeicher, hat sehr grosse Kapazität und erlaubt es zielgerichteten Aktivitäten über Sekunden, Minuten und Stunden ihre Form zu geben. Nach Massgabe emotionaler und langfristiger Bedeutung wird

einem sehr kleinen Teil dieser metastabilen Gedächtnisspuren erlaubt, durch stückweises Wiederholen einen Konsolidierungsprozess zu durchlaufen und ins dauerhafte Gedächtnis überführt zu werden.

Wie lassen sich diese Begriffe und diese kognitiven Prozesse im Rahmen der neuronalen Architektur interpretieren? Ein Hinweis auf die Ebene, auf der sich dies abspielt resultiert daraus, dass das menschliche bewusste Erleben und die menschlichen Gedächtnisinhalte abstrakter Natur sind, dass der Mensch sich üblicherweise an Schematisches statt Details erinnert, dass, wie ja immer wieder gesagt wird, sein Gedächtnis keine Kamera- oder Bandaufnahme ist.

Um einer Situation funktionellen Sinn zu geben, bezieht der Mensch sie auf abstrakte funktionelle Schemata. Will jemand, zum Beispiel, in der Stadt die verkehrsreiche Strasse überqueren, so engagiert er ein Navigationsschema, das in die Darstellung der Szene eine Trajektorie (eine Folge von Ort-Zeit Paaren) über die Strasse projiziert, er lenkt seine Aufmerksamkeit in die Richtungen aus denen Fahrzeuge auf Kollisionskurs kommen könnten, und plant, zeitüberspannend, seine eigene Trajektorie relativ zu denen der ankommenden Fahrzeuge so, dass er sicher drüben ankommt. All dies realisiert unser Gehirn als ein Ensemble miteinander vernetzter Schemata, die ihrerseits sich mit Hilfe von Projektionsnetzen auf die augenblickliche Szenenrepräsentation und ihre zeitliche Fortschreibung beziehen. In unserem Kurzzeitgedächtnis hat die Trajektorie eines Fahrzeugs eine bleibende Existenz, selbst wenn wir das Fahrzeug zeitweise nicht im Blick haben. Das gesamte Netz von sensorischen Signalen, Schemata, und wechselseitigen Projektionsnetzen wird dynamisch durch die Evolution der Szene aktualisiert, indem die Projektionsnetze durch Geschwindigkeitsparameter gelenkt werden.

Dieses selbe Szeneninterpretationsnetzwerk stellt den Aufmerksamkeitsfokus dar (wenigstens wenn wir konzentriert bei der Sache sind). Er wird zusammengehalten durch die Schema-darstellenden Netze und die Projektionsnetze zwischen ihnen. Es mag Aktivität ausserhalb dieses funktionellen Netzes geben, aber diese Aktivität verhallt ungehört.

Soweit ein solcher Prozess als Routine abläuft, wird das ihn darstellende Gesamtnetz lediglich kleine auftretende Inkonsistenzen reparieren. Sobald aber durch verhaltensrelevante Schemata etwas Bemerkenswertes diagnostiziert wird, Neuheit, Gefahr, persönlich oder emotional Relevantes, ein unerwarteter Erfolg oder Ähnliches, in unserer Beispielszene etwa eine drohende oder aktuelle Kollision, dann wird ein zentraler Mechanismus in Gang gesetzt (realisiert durch spezifische Kerne im Mittelhirn), durch den im ganzen System die Plastizität der aktiven Verbindungen hochreguliert wird. Das hat den Effekt, dass das die Situation darstellende Netz so weit verstärkt wird, dass es in der Folge mehrfach spontan wieder aktiviert werden kann, in Gänze oder in Teilen, als Grundlage für einen Konsolidierungsprozess und Überführung in das Langzeitgedächtnis.

Der Prozess der Gedächtnisbildung spielt sich offensichtlich auf der Ebene abstrakterer Schemata und ihrer wechselseitigen Vernetzung statt. Dies geht vielleicht darauf zurück, dass die Hochregulierung der Plastizität durch die verhaltensrelevanten Schemata auf die Teile des aktuell aktiven Netzes beschränkt ist, die in direkter Verbindung zu diesen Schemata stehen und zu ihrer Auslösung geführt haben.

Verhaltensrelevante Schemata sind im Gehirn genetisch vorgebildet, haben auf Grund ihres Einflusses auf die Regulierung der Plastizität einen Sonderstatus und definieren letzten Endes das Verhalten. Sie sind bei Geburt ursprünglich ausserordentlich abstrakter Natur, umgeben sich aber als Folge von Erfahrungen mit Schemata konkreteren Inhalts.

Dieser Prozess ist am deutlichsten im Verlauf dessen, was Verhaltensforscher als *Prägung* bezeichnen. So wird, nach Auskunft von Psychologen, die Aufmerksamkeit des neugeborenen Kindes sehr stark von der nahen Präsenz menschlicher Gesichter gefesselt. Das Gesicht, vorerst noch sehr schematisch abstrakt definiert, ist demnach ein stark verhaltensrelevantes Schema. Nach kurzer Zeit wird diese Reaktion eingeschränkt auf das Gesicht und andere Merkmale einer bestimmten Person, etwa der Mutter. Es wird also ein konkreteres, die Person spezifisches beschreibendes sekundäres Schema gebildet, das das ursprüngliche dann allein vertritt. Dies ist der Prozess der Prägung.

Mit dem Gesicht verbunden sind weitere genetisch angelegte Schemata, die emotionale Ausdrücke diagnostizieren und spezifische verhaltenssteuernde Wirkungen auslösen. Das Kind lernt schnell, den Aufmerksamkeitsfokus der Mutter zu identifizieren und bezieht dann die emotionale Wertung auf diesen. Auf diese Weise dienen also Gesicht und Verhalten der Mutter als eine Art von Antenne, als Tor zur Kultur der Erwachsenen.

Das hier beschriebene Bild könnte als *Schema-basiertes Lernen* bezeichnet werden: Gegeben eine aktuelle Szene, dargestellt in intrinsischen Koordinaten (d. h. invariant gegen Augen- und Körperbewegungen). Durch die Szene wird ein verhaltensrelevantes Schema ausgelöst, das heisst, ein Teil der Szene wird durch aktivierte Fasern homeomorph auf das Schema abgebildet. Diese homeomorphen Verbindungen zwischen Schema und Szene heben die zugehörigen Elemente der Szene als in diesem Sinne signifikant hervor. Als Folge werden nun die Verbindungen zwischen diesen Elementen verstärkt. Es ist anzunehmen, dass zwischen den Elementen bereits Verbindungen bestehen, diese annähernd Netzstruktur haben und lediglich leicht umgeordnet und verstärkt werden müssen. Auf diese Weise wird eine neue Version des ursprünglichen Schemas erzeugt, das dieses ergänzt oder ersetzt.

Im Falle der Prägung muss dieser Vorgang nur einmal oder wenige Male passieren und das ursprüngliche, angeborene verhaltensrelevante Schema wird durch das neue quasi überschrieben. Dies setzt eine entsprechend heftige Hochregulierung synaptischer Plastizität voraus. Üblicherweise wird jedoch der Vorgang mehrfach wiederholt werden müssen bevor das ursprüngliche Schema stark modifiziert oder durch eine konkretere Version ergänzt wird.

Insgesamt ist für das Lernen wichtig, dass eine gegebene Szene zunächst interpretiert wird, indem sich ein Aufmerksamkeitsfokus bildet, durch den in den verschiedenen Modalitäten Teilstrukturen ausgewählt, miteinander vernetzt und damit in funktionell sinnvoller Weise miteinander verbunden werden. Lernen findet dann innerhalb dieses im Vergleich zu der Gesamtszene sehr viel engeren Gehäuses an Neuronen und Verbindungen statt, und nur die im Sinn der so interpretierten Szene signifikanten Strukturen und Verbindungen werden verstärkt, während die grosse

Masse der übrigen Szene von Plastizität unberührt bleibt. Dies ist wohl die Lösung des *Platizitäts-Stabilitäts-Dilemmas* (Grossberg 1980).

Hinzu kommt, dass das unter einer Agenda handelnde Individuum durch seine Bewegungen die Szenen selektiert, denen es sich aussetzt, im Sinne einer verallgemeinerten Aufmerksamkeitssteuerung. Anscheinend ist schon im Säugling das Lernen in diesem Sinne einer genetisch angelegten Schema-gesteuerten Agenda unterworfen, inklusive Suche nach Neuem und Durchführung von Experimenten um etwa kausale Zusammenhänge festzustellen. Dieses Verhalten ist von Kleinkindpsychologen ernsthaft mit dem von Wissenschaftlern verglichen worden. Darüber hinaus lernen wir im Verlauf der Reifung methodisches Lernen, und erlernen die Fähigkeit, gewisse Netzwerke in unserem Inneren systematisch zu adressieren, indem wir uns zielstrebig gewissen Reizen und Übungen aussetzen.

8 Vergleich mit Künstlichen Neuronalen Netzen

Das im letzten Abschnitt beschriebene Konzept *Interpretations-gesteuerten Lernens* muss im Kontrast gesehen werden zu dem der üblichen künstlichen neuronalen Netze. Im klassischen assoziativen Gedächtnis, siehe etwa (Hopfield 1982; Willschaw und Longuet-Higgins 1970), wird beim Lernen eine Untermenge aller Neuronen aktiviert und alle Verbindungen zwischen aktiven Neurone werden plastisch verstärkt, es wird also keine Unterscheidung zwischen signifikanten und nicht-signifikanten Verbindungen getroffen. Dies führt dazu, dass der so gespeicherte Zustand monolithischen Charakter hat und nur als Ganzes wieder aufgerufen werden kann. Die ist eine ziemlich sinnlose Funktion, da sowohl die Umwelt als auch unser Inneres stets neue Konstellationen bekannter Strukturfragmente erzeugt oder erzeugen muss, da jede Szene in diesem Sinne neu ist.

In Perzeptron-artigen künstlichen neuronalen Netzmodellen, insbesondere dem *Deep Learning* (siehe z. B. (Krizhevsky et al. 2012)), wird eine Unterscheidung zwischen signifikanten und nicht-signifikanten Verbindungen lediglich auf statistischer Basis getroffen: *Signifikant* ist eine Verbindung, die immer wieder verstärkt wird oder dazu beiträgt, den Fehler des Gesamtsystems zu reduzieren. Es ist dieses rein auf Beispiel-Statistik gestützte Konzept von Signifikanz sowie der völlige Mangel eines Auswahlprinzips für sinnvolle Verbindungs-Konstellationen, die zu dem ungeheuren Bedarf des Deep Learnings an Beispielen und dessen Mangel an Verallgemeinerungsfähigkeit führt.

Der Mangel eines inhärenten Strukturkonzepts der künstlichen neuronalen Netze resultiert aus dem völligen Ignorieren der Netzwerk-Selbstorganisation, obwohl diese im Rahmen der Ontogenese vor einigen Jahrzehnten sowohl experimentell als auch theoretisch intensiv untersucht worden ist, wie oben erwähnt. Dieser Mangel eines inhärenten Strukturkonzepts hat eine fundamentale Beschränkung der Funktion der künstlichen neuronalen Netze zur Folge. Alle Verbindungen müssen unter Anleitung durch Lösungsbeispiele (*Deep Learning*), Erfolgssignale (*Reinforcement Learning*) oder unter Selektion von ganzen Systemen (*Evolutionsstrategie*) gelernt werden. Diese Lernmechanismen sind ausserordentlich ineffizient. Wenn auch

relativ kleine Systeme so noch strukturiert werden können, wie die gegenwärtige Technik zeigt, ist das für ein Gehirn mit 1014 Verbindungen nicht mehr denkbar. Dafür braucht es ein Strukturkonzept, das den Suchraum um viele Grössenordnungen reduziert, so wie Netzwerk-Selbstorganisation es tut.

Das entscheidende Dogma der künstlichen neuronalen Netze ist die *Barlowsche Neuronenhypothese* (Barlow 1972), nach der jede zu treffende Entscheidung ein darauf spezialisiertes Neuron braucht. Da ein Neuron erst einmal durch einen Lern- oder Organisationsprozess gebildet werden muss, bedeutet dieses Dogma, dass ein System künstlicher neuronaler Netze in unerwarteten Situationen keine neuen Entscheidungen treffen kann. Ohne diese Fähigkeit lässt sich aber ein System kaum intelligent nennen. Kreativität und Spontaneität beruhen stattdessen auf kollektiven Entscheidungen, in denen eine grosse Zahl von Elementen unter dem Einfluss ihrer Wechselwirkungen in einen Attraktorzustand fallen, in dem also eine Vielzahl von Teilen als zusammenpassend erkannt werden. Ein Beispiel für eine solche kollektive Entscheidung ist der in Abb. 3 dargestellte Erkennungsvorgang. Die das aktuelle Bild, links, und das im Gedächtnis befindliche Modell, rechts, darstellenden Neuronen bilden Netze mit Attraktorcharakter, und auch das System der Projektionsfasern zwischen Bild und Modell bilden ein solches Netz. Insgesamt also ist die Entscheidung, dass Modell und Bild homeomorph sind und aufeinander passen eine kollektive Schwarmentscheidung tausender von Neuronen.

Es besteht breiter Konsens dazu, dass es zur Verarbeitung von Objekten und Situationen der natürlichen Umgebung eine Hierarchie von Merkmalen bis hinauf zu solcher grossen Komplexität bedarf. Die Neuronenhypothese erzwingt nun, dass alle Merkmale dieser Hierarchie durch Neuronen repräsentiert werden, wie etwa im Deep Learning realisiert. Verglichen mit einer Darstellung von Komplex-Merkmalen durch Kombination von Merkmalen einfachster Stufe hat das Insistieren auf der Repräsentation von Merkmalen durch Einzelneuronen erhebliche *Nachteile*, von denen hier nur zwei genannt werden sollen.

Erstens, bei der Zuweisung von n Merkmalstypen zu n Neuronen im Verlauf des Lernens muss unter den $n!$ äquivalenten Möglichkeiten eine zufällig ausgewählt werden. Diese Art von spontaner Symmetriebrechung kostet sehr viel Zeit. Wenn dagegen Komplex-Merkmale durch Netze von Merkmalen niedrigerer Stufe dargestellt werden, dann gibt es keine grosse Zahl von Alternativen, unter denen ausgewählt werden muss.

Zweitens, wenn Komplexmerkmale als strukturierte Netze dargestellt werden, so kann beim Vorgang der Erkennung *Homeomorphie*, das heisst, Gleichheit der Anordnung, zu gespeicherten Modellnetzen verlangt und auf diese Weise Invarianz gegenüber Transformationen erzielt werden. Wenn Komplexmerkmale aber durch Einzelneuronen dargestellt werden, stehen diese zwar für Anordnungen von Elementarmerkmalen der Eingangsschicht, sind aber natürlich strukturlos und stellen diese Anordnungen und folglich auch deren räumlichen und Überlapp-Beziehungen in keiner Weise dar. Dies macht solche Systeme künstlicher neuronaler Netze anfällig für fälschliches Erkennen von Eingängen mit denselben Merkmalen aber anderer Anordnung (siehe z. B. (Brendel und Bethge 2019)). Ausserdem kann Invarianz,

zum Beispiel gegenüber Musterverschiebung, nur durch aufwändiges Lernen eigens für jeden zu erkennenden Mustertyp erreicht werden.

Ein ganz entscheidendes Argument für *Kompositionalität*, das heisst, die Darstellung komplexerer Gebilde durch strukturierte Zusammensetzung aus einfacheren, ist die Möglichkeit des Lernens aus einer endlichen Menge von Beispielen für den Umgang mit der potenziell unendlichen Menge zukünftiger Situationen. Wenn, wie im Deep Learning üblich, darauf verzichtet wird, Eingangsmuster in unabhängige Aspekte (z. B. Form und Farbe) oder Teile zu zerlegen, dann ist für das System jede neue Kombination dieser Aspekte oder Teile neu und muss neu gelernt werden. Wenn dagegen ein lernendes System die Fähigkeit hat, Aspekte und Teile voneinander abzutrennen und separat darzustellen und speichern, dann kann es neue, nie gesehene, Eingänge als neue Kombination von bekannten Teilbeschreibungen nachbilden und in diesem Sinne verstehen (siehe Abb. 1). Die Fähigkeit zur Konstruktion einer unendlichen Vielfalt von visuellen Szenen auf der Basis einer sehr endlichen Menge von Grundformen und Kompositionsregeln wird durch die Computergrafik demonstriert.

9 Diskussion

Seit Jahrtausenden sind Geist und Gehirn Gegenstand von Untersuchungen, seit Jahrzehnten rasant beschleunigt sowie ergänzt durch Versuche der technischen Nachbildung des Geistes. Der dabei angesammelte ungeheure Schatz an Wissen ist inzwischen zersplittert in Duzende von Fachgebieten und zugehörigen Denkstilen, die miteinander kaum noch Kontakt haben. Es wäre daher äusserst nützlich, wenn ein Territorium gefunden werden könnte, auf dem sich alle mit der Funktion von Geist und Gehirn beschäftigten Aktivitäten austauschen können. Eine als Brücke zwischen den materiellen und geistigen Aspekten des Problemkomplexes dienende kognitive Architektur könnte die Rolle eines solchen gemeinsamen Bodens spielen.

Die gegenwärtige Arbeit macht den Versuch, eine solche Architektur zu formulieren. Dies geschieht auf der Grundlage der folgenden Prinzipien:

- i. Der Kognition liegt ein einheitliches Datenformat zu Grunde.
- ii. Kognitive Strukturen sind emergente Phänomene.
- iii. Das kognitive Datenformat hat die Form aktivierbarer strukturierter Netzfragmente.
- iv. Netzfragmente setzen sich zu grösseren Netzen derselben Struktur zusammen.
- v. Netze werden durch den Vorgang der Netzwerk-Selbstorganisation gebildet.

Die Details dieser Architektur stehen überall auf dem Boden der neurowissenschaftlichen Erkenntnis. Dennoch wird es nicht einfach sein, die Gliederung der Verschaltung des Gehirns in Netzfragmente sowie deren selektive Aktivierung experimentell nachzuweisen. Ersteres könnte im Rahmen der Methodik der *Connectomics* durch Analyse der detaillierten synaptischen Verschaltung eines ausreichenden Volumens von kortikalem Gewebe gelingen, Letzteres durch gleichzeitiges Ableiten sämtlicher

neuronaler Signale eines solchen Volumens. Die sich schnell entwickelnde experimentelle Technik wird derartige Versuche innerhalb der kommenden Dekade möglich machen, allerdings nur wenn die Fragestellung genügend Aufmerksamkeit auf sich zieht.

Eine ganz andere Art experimentellen Zugangs ist die Realisation eines kognitiven Systems im Rechner. Wie in dieser Arbeit zitiert, sind die wesentlichen funktionellen Aspekte der Architektur bereits simuliert und demonstriert worden. Für einen technischen Durchbruch ist es aber erforderlich, durch Emergenz und Lernen auf der Basis dieser Architektur ein System mit überlegener Funktionalität zu erzeugen. Hierfür bietet es sich an, ein künstliches Seh-System in einer *Virtual Reality* Umgebung zu trainieren. Eine solche Umgebung würde dem System das Äquivalent der zum Lernen nötigen zusätzlichen Modalitäten, insbesondere Tastsinn und Eigenbewegung, liefern können. Die Homogenität der Architektur würde die klassische Schwierigkeit des Computersehens – dessen babylonische Sprachverwirrung verschiedener Datenstrukturen und Algorithmen unabhängig voneinander entwickelter Subsysteme – überwinden.

Die Entwicklung eines ersten Demonstrators eines solchen Seh-Systems könnte einem kleinen Team von Entwicklern in wenigen Jahren gelingen, würde die Tür zu einer neuen technologischen Ära öffnen und genügend Sog entwickeln, die Kognitionswissenschaften und Neurowissenschaften fundamental zu transformieren.

Literatur

- Amari S (1977) Dynamics of pattern formation in lateral-inhibitory type neural fields. *Biological Cybernetics*, 27(2):77–87.
- Anderson C, Van Essen D (1987) Shifter circuits: a computational strategy for dynamic aspects of visual processing. *Proceedings of the National Academy of Sciences of the United States of America* 84:6297–6301.
- Arbib M (1992) Schema theory. *The Encyclopedia of Artificial Intelligence*, 2:1427–1443.
- Barlow H (1972) Single units and sensation: A neuron doctrine for perceptual psychology. *Perception*, 1:371–394.
- Bartlett F (1932) *Remembering, A Study in Experimental and Social Psychology*. Cambridge University Press, Cambridge.
- Bergmann U, von der Malsburg C (2011) Self-organization of topographic bilinear networks for invariant recognition. *Neural Computation*, 23(11):2770–2797.
- Biederman I (1987) Recognition-by-components: a theory of human image understanding. *Psychological Review*, 94(2):115–147.
- Bolder B (2005) *Coordination of an artificial visual system with biological models*, PhD thesis, Ruhr-University Bochum.
- Brendel W, Bethge M (2019) Approximating CNNs with bag-of-local-features models works surprisingly well on imagenet. *International Conference on Learning Representations (ICLR)*.
- Bürglin T, Affolter M. (2016) Homeodomain proteins: an update. *Chromosoma* 125:497–521.
- Consortium TIHG (2001) A physical map of the human genome. *Nature*, 409:934–941.
- Crytek (2017) Robinson: The journey. <https://www.crytek.com/games/robinson>.
- Erlhagen W, Bicho E (2006) The dynamic neural field approach to cognitive robotics. *Journal of Neural Engineering*, 3, R36. <http://stacks.iop.org/1741-2552/3/i=3/a=R02>.
- Fernandes T, von der Malsburg C (2015) Self-organization of control circuits for invariant fiber projections. *Neural Computation*, 27(5):1005–1032.

- Fleck L (1935) Entstehung und Entwicklung einer wissenschaftlichen Tatsache. Schwabe, Basel.
- Geman S, Bienenstock E, Doursat R (1992) Neural networks and the bias/variance dilemma. *Neural Computation*, 4:1–58.
- Goodhill G (2007) Contributions of theoretical modeling to the understanding of neural map development. *Neuron*, 56(2):301–311.
- Grabska-Barwinska A, von der Malsburg C (2008) Perinatal ontogenesis of orientation specificity and maps in primary visual cortex of higher mammals. *Journal of Neuroscience*, 28:249–257.
- Grossberg S (1980) How does a brain build a cognitive code? *Psychological Review*, 87:1–51.
- Hebb D (1949) *The Organization of Behavior*. Wiley.
- Hinton G (1981) A Parallel Computation that Assigns Canonical Object-Based Frames of Reference. *International Joint Conference on Artificial Intelligence*, pp. 683–685.
- Hopfield J (1982) Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558.
- Hubel D, Wiesel T (1977) Ferrier-lecture: Functional architecture of macaque monkey visual cortex. *Proceedings of the Royal Society London, Series B*, 198:1–59.
- Häusser A, von der Malsburg C (1983) Development of retinotopic projections – an analytical treatment. *Journal of Theoretical Neurobiology*, 2:47–73.
- Isaacson J, Scanziani M (2011) How inhibition shapes cortical activity. *Neuron*, 72(2):231–243.
- Jug F, Cook M, Steger A (2012) Recurrent competitive networks can learn locally excitatory topologies. *Proceedings of International Joint Conference on Neural Networks*, pp. 1–8.
- Kant I (1781) Kritik der reinen Vernunft, Kap. Von dem Schematismus der reinen Verstandesbegriffe. gutenberg.spiegel.de/buch/kritik-der-reinen-vernunft-2-auflage-3502/1.
- Kree R, Zippelius A (1988) Recognition of topological features of graphs and images in neural networks. *Journal of Physics*, A21:813–818.
- Krizhevsky A, Sutskever I, Hinton G (2012) Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25:1097–1105. <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.p>.
- Kuhn T (1962) *The Structure of Scientific Revolutions*. University of Chicago Press.
- Landauer T (1986) How much do people remember? some estimates of the quantity of learned information in long-term memory. *Cognitive Science*, 10(4):477–493.
- Li M, Vitányi P (2008) Preliminaries, in: *An Introduction to Kolmogorov Complexity and Its Applications*. Springer New York, NY, pp. 1–99.
- McFarland D (1999) *Biologie des Verhaltens. Evolution, Physiologie, Psychobiologie*. Spektrum Akademischer Verlag, Heidelberg, Berlin.
- Minsky M (1974) A framework for representing knowledge. Technical Report, 306, MIT AI Laboratory.
- Olshausen B, Anderson C, Van Essen D (1993) A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *Journal of Neuroscience*, 13(11):4700–4719.
- Riseman E, Hanson A (1992) A Methodology for the Development of General Knowledge-Based Vision Systems. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 293–336.
- Roskies A (1999) The binding problem. *Neuron*, 24:7–9.
- Schank R, Abelson R (1977) *Scripts, plans, goals and understanding: An inquiry into human knowledge structures*. Erlbaum, New Jersey.
- Stringer S, Trappenberg T, Rolls E, de Araujo I (2002) Self-organizing continuous attractor networks and path integration: one-dimensional models of head direction cells. *Network: Computation in Neural Systems*, 13(2):217–242.
- Turing A (1936) On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of London Mathematical Society*, 2–42:230–265.
- von der Malsburg C (1973) Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*, 14:85–100.
- von der Malsburg C (1979) Development of ocularity domains and growth behaviour of axon terminals- *Biological Cybernetics*, 32:49–62.

- von der Malsburg C (1981) The correlation theory of brain function. Internal report, 81-2, Max-Planck-Institut für Biophysikalische Chemie, Postfach 2841, 3400 Göttingen, FRG. Reprinted in Domany E, van Hemmen JL, Schulten K (Hrsg.) *Models of Neural Networks II*, chapter 2, pp. 95–119. Springer, Berlin, 1994.
- von der Malsburg C (1985) Nervous structures with dynamical links. *Berichte der Bundengesellschaft für physikalische Chemie*, 89(6):703–710.
- von der Malsburg C (1986) Am I thinking assemblies? In *Brain Theory. Proceedings of the Trieste Meeting on Brain Theory*, 1–4 October 1984, Springer, pp. 161–176.
- von der Malsburg C (1987) Ist die Evolution blind? Küppers BO (Hrsg.) *Ordnung aus dem Chaos*, Piper, München, pp. 269–279.
- von der Malsburg C (1994) Network self-organization in the ontogenesis of the mammalian visual system. Zornetzer S, Davis J, Lau C (Hrsg.) *An Introduction to Neural and Electronic Networks*. Second Edition. Academic Press, pp. 463–447.
- von der Malsburg C (1997) The coherence definition of consciousness, in Ito M, Miyashita Y, Rolls ET (Hrsg.) *Cognition, Computation and Consciousness*. Oxford University Press, pp. 193–204. URL: <http://cogprints.org/1487/>.
- von der Malsburg C (2018a) Concerning the neural code. *Journal of Cognitive Science*, 19:511–550.
- von der Malsburg C (2018b) Vision as gateway to the brain. <https://platonite.com/vision-as-gateway-to-the-brain>.
- von der Malsburg C, Bienenstock E (1987) A neural network for the retrieval of superimposed connection patterns. *EPL (Europhysics Letters)*, 3(11):1243. <http://stacks.iop.org/0295-5075/3/i=11/a=015>.
- Willshaw D, Longuet-Higgins H (1970) Associative memory models. *Machine Intelligence*, 5:351–359.
- Willshaw D, von der Malsburg C (1976) How patterned neural connections can be set up by self-organization. 194:431–445.
- Willshaw D, von der Malsburg C (1979) A marker induction mechanism for the establishment of ordered neural mappings: its application to the retinotectal problem. *Philosophical Transactions of the Royal Society of London, Series B*:287:203–243.
- Wilson H, Cowan J (1972) Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, 12:1–24.
- Winograd T (1971) Procedures as a representation for data in a computer program for understanding natural language. Technical Report, 235, MIT AI.
- Wolfrum P, Wolff C, Lücke J, von der Malsburg C (2008) A recurrent dynamic model for correspondence-based face recognition. *Journal of Vision*, 8(7):34.
- Wolpert D, Mcready W (1997) No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1):67–82.
- Zhu J, Bergmann U, von der Malsburg C (2010) Self-organization of steerable topographic mappings as basis for translation invariance. *International Conference on Artificial Neural Networks (ICANN 2010)*, pp. 414–419.



Artificial Intelligence – The Big Picture

David Abele and Sara D’Onofrio

Abstract

This article presents an introduction to key concepts in the field of artificial intelligence and how they relate to each other. It offers a scheme to assess the maturity levels of current and future technologies or applications and lays out why today’s solutions cannot be reasonably considered intelligent. At its heart, the article includes an overview of fundamental solution approaches, the resulting capabilities, required enablers and the critical aspects of robustness, transparency and trust of produced results. Using novel ways to organize existing knowledge, it aims to enable the reader to accurately separate different notions and to gain a broad understanding of the key concepts under consideration. Based on a more comprehensive framework, it is deliberately limited in scope to provide a well-rounded overview to help the reader identify topics on which more detailed information should be sought out.

Vollständig neuer Beitrag.

D. Abele
Swiss Post, Bern, Switzerland
E-Mail: david.abele@post.ch

S. D’Onofrio (✉)
Human-IST Institute, University of Fribourg, Fribourg, Switzerland
E-Mail: sara.donofrio@unifr.ch

1 Introduction

Specialized articles and key notes about various artificial intelligence technologies,¹ their strengths and weaknesses, and discussions about their detailed implementations abound. Yet, for anyone seeking to gain a broad understanding of how different technologies and concepts in the field of artificial intelligence fit into a broader context, gathering the relevant information turns out to be a time-consuming and difficult undertaking. There simply does not seem to exist a generally accepted framework, covering how various solution approaches, AI capabilities, fields of application relate to each other, and what each of them contributes to filling the gaps in the quest for artificial intelligence. Invariably, it makes it difficult to know what to focus on, or to separate the *why* from the *what* and the *what* from the *how*. The authors aim to fill the gap by giving the reader the tools to organize different concepts encountered, and the ability to immediately understand how any new terminology fits into the big picture. In a business context, it will give the reader the faculty to structure any discussion about AI and to ask the right questions. In academic settings, the authors hope this framework will allow researchers to take a step back from the detailed problems at hand, for a broader view of how their research contributes to a vast field of desired outcomes.

While there does not seem to be any doubt that the concepts of artificial intelligence and cognitive computing overlap, there is an ongoing debate as to whether they differ at all or whether they should be used interchangeably. Historically, no such distinction can be discerned. The term artificial intelligence or AI was coined in 1955 (McCarthy et al. 1955), which John McCarthy then updated in 2007 to mean “*the science and engineering of making intelligent machines*” (McCarthy 2007, p. 2). The term *cognitive computing*, on the other hand, was introduced by IBM along with its Watson system,² which came to fame by beating the world’s best *Jeopardy!* players in 2011 (Gabbatt 2011), during a period of disillusionment when the notion of AI did not trigger either the general public’s or many an investor’s enthusiasm (Hendler 2008). The new terminology was a way to market the Watson brand without associating it with the negative perception of AI, but it basically meant the same thing. Over the years, the use of the term cognitive computing seems to have transcended its original scope, and it has gained a following that accredits it with additional connotations, such as a tighter association with cognitive sciences and a focus on human-machine interaction (D’Onofrio et al. 2018; Haun 2014). In contrast, these disciplines see AI as being focused on achieving the result of intelligence, regardless of the kinds of processes that may lead to this result. Some authors, to varying degrees, now even choose to view the term AI in a somewhat narrower light, perhaps giving in to the popularity of machine learning, which currently tends to overshadow all other aspects of AI (Ignatov et al. 2018; Silver et al. 2016).

¹Throughout this text the term *artificial intelligence* and its popular abbreviation *AI* will be used interchangeably.

²<https://www.ibm.com/watson>

The authors will not enter this debate and treat both terms as equivalent for the purposes of this article.

It is not the authors' intention to provide a rigorously peer-reviewed scientific paper but rather to offer a synthesis that condenses relevant information from diverse sources and the authors' expertise and extensive practical experience in various business contexts to provide a succinct and practical guide to AI. The goal is to help the reader understand how the key concepts relate to each other and, where necessary, to provide a few examples that illustrate their relevance in practice. Where the article introduces original thoughts and content, or novel perspectives on the topic, they do not represent a consensus view, and the reader is encouraged to challenge them against complementary information from other sources. While the article contains novel ways of organizing existing knowledge, its deliberately limited scope does not allow delving into detailed descriptions of new methods and techniques. The authors believe, however, that in many cases, the condensed introduction to key concepts in the field of AI, as described in this article, presents a more balanced view (e.g., on what constitutes AI) than can be found in many highly specialized scientific papers. The authors hope that, it will prove to be a valuable contribution to structuring the available knowledge and that it will help the reader to navigate the plethora of new concepts that come with it.

2 Hype or Ripe?

Enthusiasm for AI has been shown to be highly cyclic – it has abruptly waxed and waned several times. During an initial honeymoon period in the late 1950s and 1960s, starting with the Turing's Test (Turing 1950) that led to the first summer conference on AI (McCarthy et al. 1955) and the first prototype of a chatbot by the name of ELIZA (Weizenbaum 1966), the prospect of AI fueled the collective imagination of humanity. Various fundamentally different approaches aiming to reach the goal of systems that exhibit intelligent behavior competed and alternately became fashionable or fell out of fashion. Highly medialized successes, such as IBM Deep Blue's win over Gary Kasparov (Weber 1997), announced milestones in the maturity of AI that initiated new hypes. During the two so-called *AI winters* (Nilsson 2009) – real troughs of disillusionment in the late 1970s and early 1990s – funding for research in the field almost dried up (Hendler 2008). Interestingly, artificial neural networks had fallen particularly deeply out of favor and were unable to attract any investment or any academic interest during a couple long decades. Expert systems based on completely different fundamentals were all the craze back then and monopolized nearly all investment (McCorduck 2004). Today, AI enjoys a renewed wave of popularity, owing to recent breakthroughs mainly in the field of deep learning (Goodfellow et al. 2016). This time around, the breakthroughs have led to viable real-world applications and demonstrated the potential to generate revenue. The excitement is now accompanied by an unprecedented flow of funding into a few narrow areas, mainly deep learning and natural language processing (Mishra 2018). These systems are getting better and better and are even able to beat humans at some very specific tasks.

The newfound popularity of the term AI is such that marketing departments all over the world have embraced it as an opportunity and every piece of technology seems to have miraculously become intelligent almost overnight. However, not every system or device tagged with AI deserves the designation, and even the most hyped up technologies exhibit only certain narrow aspects of what could be perceived as intelligent behavior. Intelligent or not, certain technologies have now undoubtedly reached a threshold of usefulness that justifies the investments, and they will drive growth in many business areas, as well as render obsolete numerous current processes and even entire industries. With the already achieved traction (and bearing in mind that technological progress is exponential), one can expect within the next decade or so, if not necessarily truly intelligent machines, at least very effective, efficient and ubiquitous ones. After millennia of building increasingly more sophisticated *passive* tools, humanity is at the threshold of entering an era of *active* tools (Voicea [n.d.](#)). This is a significant step. Whether the terms of artificial intelligence and cognitive computing ultimately stick, the underlying quest for active tools capable of autonomous action in complex environments is here to stay.

3 Key Concepts of Artificial Intelligence

Despite the term artificial intelligence having originated in the 1950s, no generally accepted definition of AI exists to date. The label artificial intelligence is commonly applied to systems that create an illusion of intelligent behavior, even when such an illusion is limited to a very narrow part of the kind of intelligence humans and other animals exhibit (Li and Du [2007](#)). An information system may fall under the banner of AI at a particular moment in time and, as technologies evolve, may be considered an ordinary IT system a few years down the line. In the view of the authors, artificial narrow or general intelligence (Goertzel and Pennachin [2007](#)) requires a complex combination of capabilities, well beyond the learning and problem-solving skills of today's man-made systems. Regardless of the level of intelligence achieved by AI systems, they are all the result of some fundamental choices made by the developers of the systems, even before implementing any specific solution. These choices lead to particular strengths and limitations, but they also significantly determine a number of the other characteristics of any AI system, such as its data requirements, the type of tasks it is good at, as well as the robustness or transparency of the solution. The distinction between the technical implementation choices underlying a system and the facets pertaining to its utilization is important; it will allow the reader to ask the right questions and to focus on those aspects that matter. However, what are those things that characterize an AI system and its context? The authors have clustered the key concepts into six categories. They distinguish between aspects pertaining to the level of intelligence, the utilization of a system, the fundamental types of outcomes that an AI system seeks to achieve, the underlying technology choices, the key enablers of an AI system and, finally, concepts related to its robustness, transparency and trust (Fig. 1).

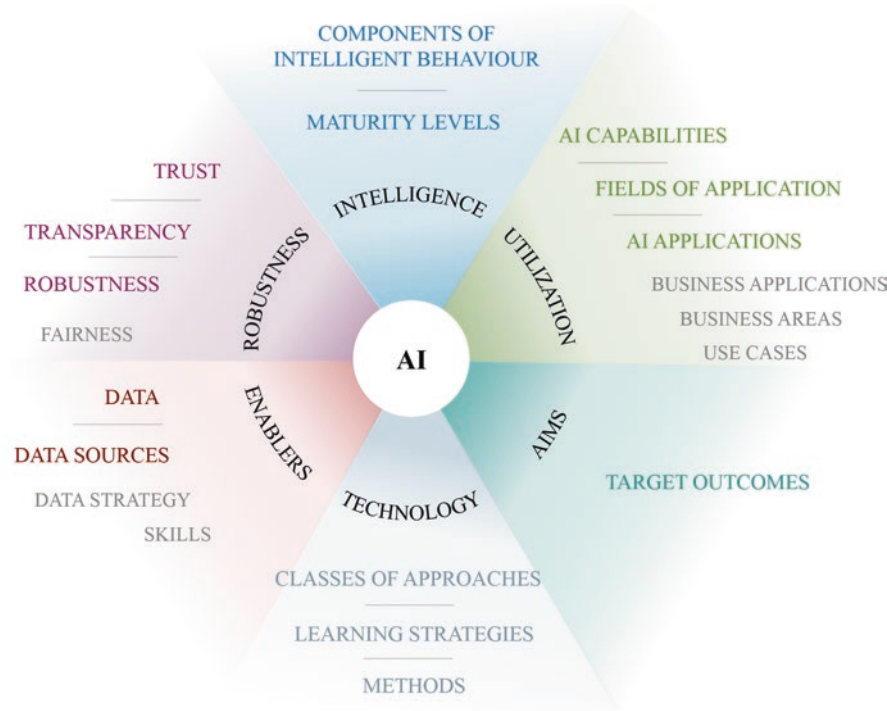


Fig. 1 Overview of key concepts that characterize an AI system and its context

3.1 Intelligence: Maturity

A maturity model for assessing AI systems is useful, since it can be used to determine the stage of development of technologies in relation to each other or to gain an understanding of how well they would do in given situations. For an initial assessment of the maturity level of autonomous systems, the authors propose a classification along two dimensions, where the x-axis represents the *ability to deal with uncertainty* and the y-axis shows the *ability to solve problems adaptively*.³

The maturity model, illustrated in Fig. 2, offers one of the many possible lenses to express the maturity of *intelligent* systems.⁴ The various gradations help classify existing systems based on their capabilities (and by extension, their level of intelligence): Each category of the y-axis represents a step-change from the previous category. As many artificial intelligence systems *only simulate* intelligent behavior, an awareness of the relative maturities of different solution approaches, as well as their evolution over time, is useful. The model also shows one view of the path to artificial general intelligence and what ground remains to be covered.

³The ability to deal with complexity is implicitly taken into account by this model. More complex environments tend to require more mature capabilities on both axis of the model.

⁴The model is applicable for both, biological and artificial systems.

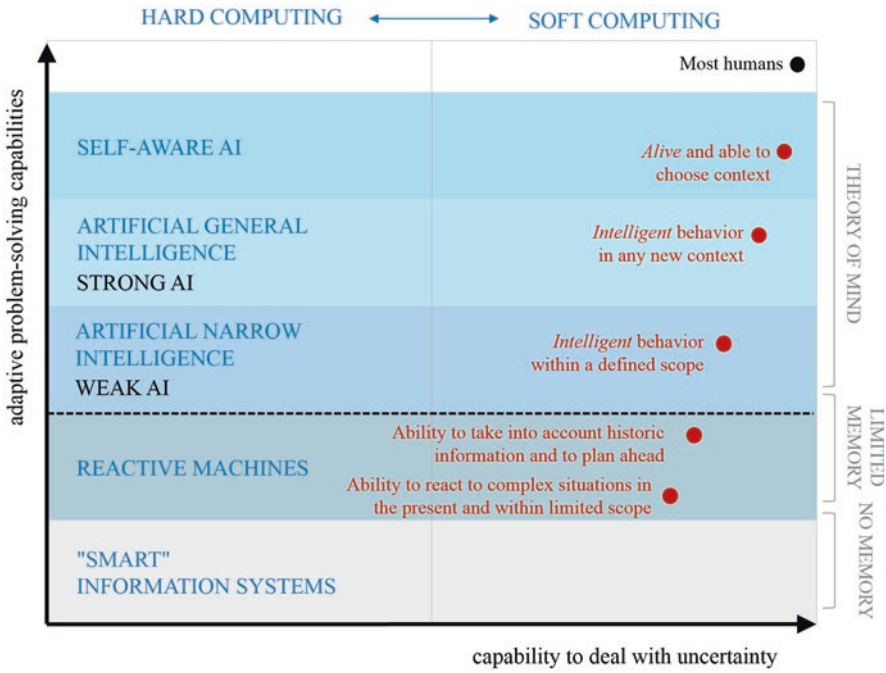


Fig. 2 Evolution of intelligence in the maturity model of artificial intelligence

The development of systems that are able to deal with uncertainty (x-axis) has led to the distinction between hard or soft computing approaches. *Hard Computing* is an umbrella term for exact models that can usually only handle structured data and whose application generates precise answers. It offers the advantage of very predictable and traceable results, which are only valid within the previously defined conditions and can only be applied to reality to a certain extent. Moreover, the required computing power tends to scale exponentially with the degree of complexity of the environment, and every possible constellation needs to be explicitly modeled (Gupta and Kulkarni 2013). *Soft Computing*, however, refers to models that permit a certain degree of uncertainty, for example, in the form of vague values and results. In comparison to the exact models of hard computing, the findings generated from soft computing are not precise but rather approximate (e.g., several possible solutions instead of just one). Such solutions are easier to apply to real-world situations (Seising and Sanz 2012). The ability to deal with imprecise information is one of the preconditions that must be met when aiming to develop AI systems.

The second dimension of the maturity model addresses the adaptive problem-solving capability (y-axis) of systems. It indicates the ability of a system to generalize and transfer its capabilities to entirely new and unknown contexts. The gradations show, by means of a hierarchy, how systems can be classified into five discrete categories. Progressing within a category may be the result of incremental

improvements, whereas moving from one category to the next typically requires a step-change in capabilities.

“Smart” Information Systems refers to systems with capabilities slightly beyond those of traditional legacy information technology (IT). They are able to accommodate variable, possibly less structured and incomplete inputs to draw inferences. However, they are usually not able to process highly complex situations and can only be used in limited and specific areas. They tend to rely on algorithms that encode each possible operation and process. Chatbots are examples of this category. While today’s chatbots may appear to react to unexpected user input in a *smart* way, they are, in fact, still just following the instructions that had been preprogrammed by the developers (Stucki et al. 2018, 2020).

Reactive Machines contain systems capable of immediately responding to current input, without necessarily being explicitly programmed to provide a deterministic response to a given, well-defined data entry. This enables reactive systems to deal with more complex or fuzzy situations. They do not, however, have to truly understand objects or situations to produce an appropriate response. Reactive machines are generally limited in their ability to consider previous experience or to strategize for future actions (Joshi 2019; Kumar 2018). The control systems of self-driving cars could be narrowly classified in this category because they have to react to situations never before modeled due to the complexity of the environment in which they evolve (Behere and Torngren 2015). Yet, autonomous vehicles are not expected to understand associated concepts, such as pedestrians.

To be considered **Artificial Narrow Intelligence** (i.e., a *weak* form of “true” artificial intelligence⁵ limited to a narrow scope) (Pennachin and Goertzel 2007), a system would have to integrate, as a minimum, sophisticated abilities of perception, learning, generalization, adaptation, reasoning, and, crucially, knowledge representation that includes an *understanding* of concepts. It should also be capable of applying these skills autonomously within a certain domain (Joshi 2019; Kumar 2018). Reaching this maturity level is still only a hypothetical possibility; assigning any future system to this category should be done with caution. *Simulations* of learning or reasoning processes do not belong in this category – the system must be able to master the required capabilities to a comparable level as humans do.

For **Artificial General Intelligence** to become a reality, a system would have to exhibit intelligent behavior across all domains (without being limited to a given context) and possess a high degree of autonomous self-control. This is also referred to as a *strong* form of “true” artificial intelligence (Pennachin and Goertzel 2007). If the development of AI were to reach sufficient maturity, one can imagine that machines could reach a threshold beyond which they would become conscious. If this happened, one would speak of **Self-Aware AI**. Self-aware systems might become more directly comparable to human intelligence (as far as human intelligence is understood) (Joshi 2019; Kumar 2018). At that stage humans would likely gradually lose control over those systems as tools for their own benefit.

⁵“True” was purposely placed in quotation marks, as it has not been clearly defined what exactly an artificial intelligence is.

Note that the step from *Reactive Machines* to *Artificial Narrow Intelligence* is a massive one. The authors do not anticipate any scenario in which this would be possible in the near future. Further evolving the technology to reach an AI capable of adapting and behaving intelligently in any new context is a tougher hurdle than might seem. Not only would this imply systems that are capable of assessing new environments and adjusting their existing capabilities to be used in different ways, but this might also mean that the system must be able to develop new learning strategies or entirely new capabilities for itself. Achieving *true* artificial intelligence is thus undoubtedly a medium- to long-term goal of researchers and developers.

Another way of considering adaptive problem-solving capabilities (the y-axis) is by the reference to *Theories of Mind*. A theory of mind describes the ability to sense the counterpart (Kidd and Castano 2013). While “*Smart*” *Information Systems* and *Reactive Machines* only require a limited amount of past information to carry out their tasks, systems with higher levels of maturity will require more advanced abilities to relate past information to future outcomes (Joshi 2019; Kumar 2018).

It is noteworthy that past developments in AI systems have mainly led to a shift towards the left-hand side of the chart. Systems have become ever better at dealing with uncertainty in a given, narrow context. Most modern AI systems would therefore be clustered in the right-hand half of the chart, mainly within the “*Smart*” *Information Systems* layer, with a few recent forays into the *Reactive Machines* layer. Progress along the y-axis has been more modest. The ability to generalize and transfer capabilities to new contexts remains very limited, and AI systems’ aptitude to *understand* real-world concepts is, at best, embryonic. Significant advances will have to be made in the quest for truly intelligent artificial systems. While recent progress in *machine learning* systems seems to be the main driver behind the current interest in AI systems, the authors believe that their overall contribution to the quest for *true* artificial intelligence remains, to date, relatively modest, as *learning* is only one of the many cognitive capabilities such as problem-solving, knowledge representation and emotional intelligence that are necessary for intelligent behavior (Li and Du 2007). Today’s cutting-edge AI systems typically implement only a couple of cognitive capabilities. Rarely do they combine more, and rarely are those cognitive capabilities fully developed. Fig. 2 shows that to create intelligent systems, capabilities along the y-axis will have to be developed, moving the systems increasingly more towards the upper right corner – starting from today’s aptitudes to solve relatively complex situations in a limited environment towards the ability to make conscious decisions.

3.2 Utilization: Fields of Application, AI Capabilities, AI Applications

AI can be understood as a set of advanced technologies or as a collection of tools that drive productivity in various business areas. Both the technology perspective and the business perspective are indispensable in driving progress. At first glance, the distinction between them seems obvious as does the consideration of which

concepts of artificial intelligence belong to which of these two complementary aspects. However, in practice, the authors have observed some level of confusion.

When the objective is to develop, optimize or deploy technologies, researchers' and IT professionals' primary focus is on *how* to achieve this goal. They will need to understand, in various degrees of detail, how the underlying mechanisms work, which solution approaches exist, how they differ from each other and how they should be implemented. Conversely, when aiming to use AI technology in a business context, the focus shifts to identifying which business problems lend themselves to the technologies' use and *what* can be achieved with them. At this point, the details of how the AI systems are built should not be a primary concern and fade into the background. Of course, the authors do not suggest that either side should be ignorant of the problems and aspirations of the other; on the contrary. What is important, however, is to ask the right types of questions in a given context. For instance, when a business unit states that it should do something with artificial neural networks, they miss the point and may well end up with a whole series of failed proofs of concepts. Thus, before examining the current state of AI systems as a set of technologies, this article views them through the lens of their potential utilization.

The authors suggest breaking down the concepts pertaining to the utilization of AI systems into three branches as follows:

1. **Objectives:** what humans want from a system.
2. **Capabilities:** what is required to fulfil the objective.
3. **Packaging:** how the capabilities are wrapped into a deployable solution.

Considering a chatbot, while IT professionals may derive satisfaction from developing a chatbot, implementing and operating a chatbot is not the aim. Rather, the *objective* is to provide the users of an information system with answers to their questions or to fulfil their requests in the most convenient way for them. Achieving this requires *capabilities*, including the processing and interpretation of natural language, placing a request into context, accessing relevant information from other systems and presenting that information in a user-friendly way (Stucki et al. 2018, 2020). Ideally, these capabilities come in a neat *package* that can be easily configured and deployed: In the case of the present example, a set of partial capabilities is wrapped into an application called chatbot.

Illustrated in Fig. 3, the corresponding levels defined by the authors are: *Fields of Application*, *AI Capabilities* and *AI Applications*.

Fields of Application represent assignments or objectives that can be ascribed to AI systems to benefit humans. For instance, in Fig. 3, *Assistance* is given as one example of a field of application as a desired result. It can be broken down into several layers of different sub forms, such as *Personal Assistance*, *Product and Service Support*, *Decision Support* and *Caring*. Fig. 4 provides a broad overview, covering some key fields of application and showing their diversity. Note that the fields of application in Fig. 4 represent a high-level view. For instance, the category *Pilot and Control* in the *Steering and Management* group includes such purposes as

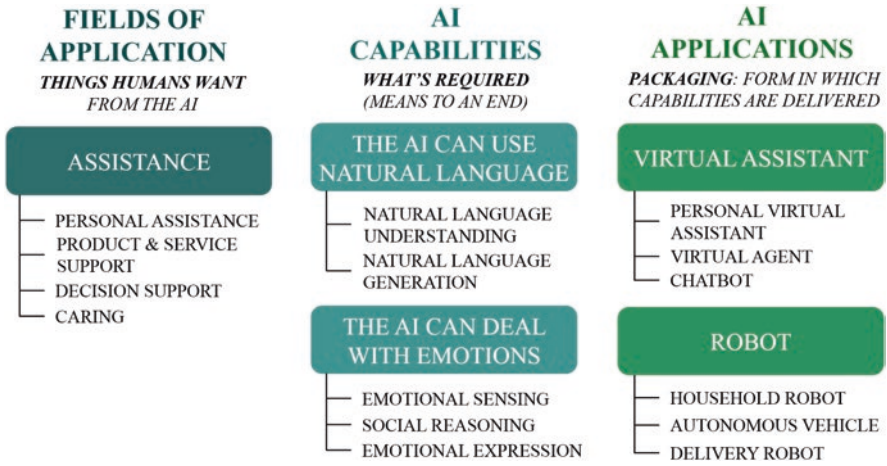


Fig. 3 Fields of application, AI capabilities and AI applications with the example of a chatbot

piloting systems, controlling processes, operating machinery, and managing and developing people.

Each field of application requires its own combination of high-level *capabilities*, resulting in a variety of AI technologies. Capabilities describe what a system is able to achieve, whereas required capabilities are the minimum set of capabilities a system must possess to fully deliver on its intended purpose. For instance, to provide assistance to a human, it is highly likely that any system would require some form of capabilities in communication and knowledge management. Depending on the level of sophistication to be reached, capabilities in dealing with emotions might also be necessary, and its communication capabilities might have to include the ability to deal with natural language in its written or spoken form and perhaps the capability to deal with accents or even the capacity to understand incorrectly formulated phrases and mispronounced words. The corresponding capabilities are commonly referred to as natural language processing (Kurdi 2016, 2017).

If the target field of application is customer support, the primary intention is to answer customer queries automatically and/or trigger actions to resolve a problem, whereas a robotic assistant for a firm’s reception desk should be able to lead guests into the correct rooms. Both cases require additional capabilities, for instance, in the latter case, the ability of spatial orientation (Suriyon et al. 2011). The exact combination and level of required capabilities must be determined on a case-by-case basis and at a low level. Once the required *AI Capabilities* have been defined, they must be implemented or purchased, preferably in an easily deployable form. The capabilities are wrapped into *AI Applications*, such as chatbots or virtual agents (Stucki et al. 2018, 2020) or a robot’s control system (Suriyon et al. 2011).

Therefore, to achieve a goal in a specific area (i.e., field of application), an *AI Application* is required, which consists of a set of *AI Capabilities*. Together, these three aspects describe the reason for deploying an AI system and what it takes to

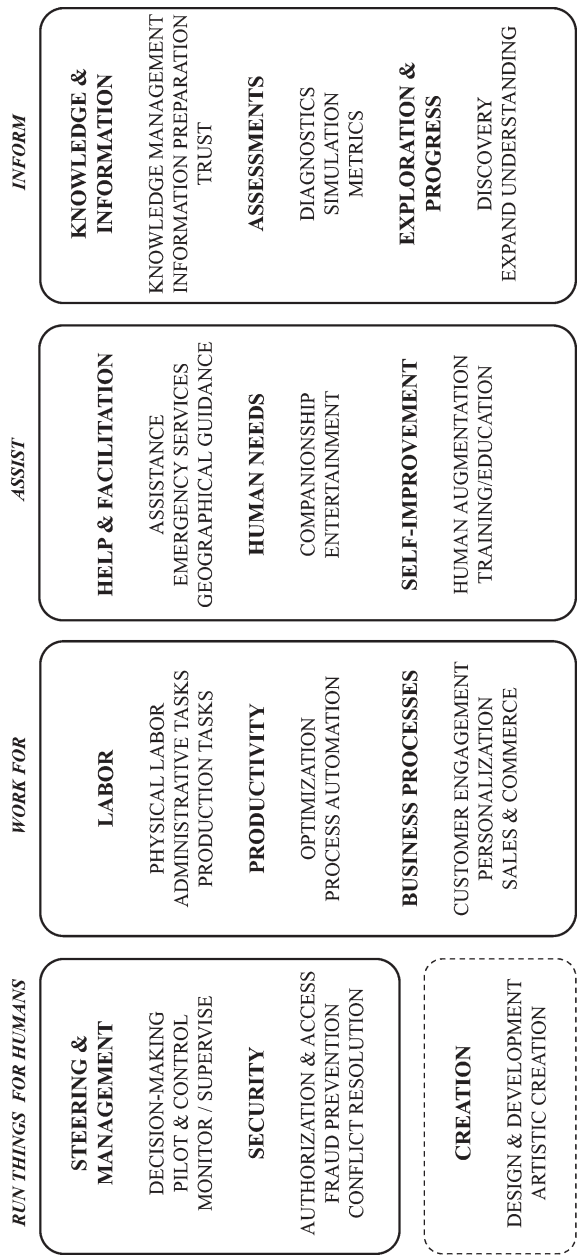


Fig. 4 Clusters of fields of application

achieve the goals. To round up the big picture view of a system’s utilization in its business context, *AI Applications* are often integrated into a *Business Application* – chatbots into customer service or a robot control system into a robot – which is deployed in a particular *Business Area*, where it serves particular use cases.

3.3 Aims: Target Outcomes

Independent of the field of application, every AI system will be deployed to perform operations that result in one or several of six types of outcomes. *Recent advances in technology have not equally benefitted each of the target outcomes.* The six fundamental types of outcomes identified by the authors are shown in Fig. 5.

Identification (1) refers to the outcome of recognizing something, verifying the identity of something, or discovering the existence of something. Examples include recognizing and naming objects or people (Liang and Hu 2015), identifying an emotion from a facial expression (Teixeira Lopes et al. 2015), retinal or vein pattern recognition (Trabelsi et al. 2016), extracting the opinion from a text (Poria et al. 2016) and anomaly detection (Malhotra et al. 2015). More generally, identification is typically characterized by the mapping of known reference information to a given situation or by discovering whether something exists in a defined space or not. Recent advances in artificial neural networks and machine learning have particularly benefitted this category of outcomes.

Understanding (2) a situation or a concept is to perceive and interpret the meaning of something. In this context, achieving the outcome of understanding requires not only the assignment of semantics to the respective objects but also the ability to correctly contextualize the information. This includes associating properties or behavior, linking concepts and analyzing complex situations (D’Onofrio et al. 2018). At different maturity levels, understanding can range from a simple context analysis to projecting complex interactions. Note that while for an AI system the underlying capabilities do not necessarily have to be equivalent to those of human understanding at all points, the leap from identification to understanding is tremendous.

Decisions (3) are the result of weighing and choosing options to determine future behavior. *Informed decisions* are reached via the process of understanding what these options entail and anticipating the consequences of different possible decisions.

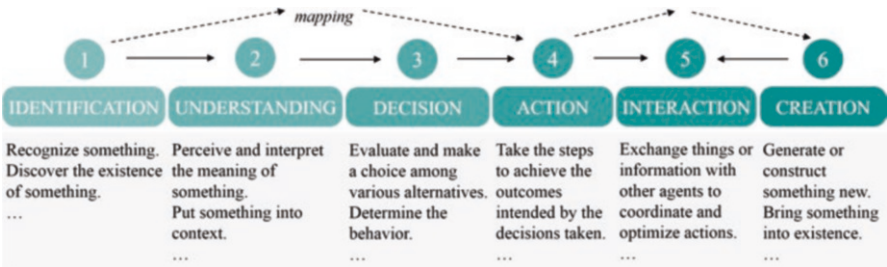


Fig. 5 Target outcomes

The authors argue that the ability to reach informed decisions and therefore to understand are prerequisites for any intelligent system. In a more primitive form of informed decisions, a number of alternatives are evaluated via a scoring function (i.e., evaluation and ranking of matching results) (Stucki et al. 2020). In this case, some of the *smarts* are not explicitly performed by the system but are implicitly included in the scoring function. Examples in this area are artificial neural networks for image recognition (Simonyan and Zisserman 2015), chess engines (Carrera et al. 2017) and medical expert systems (Korenevskiy 2015).

Actions (4) are a category of outcomes that create effects on the environment and cause a change. Today, machines autonomously do the work that people once did (Pacaux et al. 2011). Therefore, the target outcome is to do something. In the context of AI systems, action refers to carrying out explicit and informed decisions. In the absence of *intelligent* decision-making, actions can be inherent consequences of design or configurations – in which case, they would be implicit and more accurately termed reactions – or be the result of a direct mapping from *Identification* (1) to *Action* (4). In this latter case, the mapping merely creates the illusion of understanding and of decision-making. They both result from the encapsulation of the system developers' expertise into algorithms, rather than from intelligent behavior of the system itself. Carrying out actions, once a decision has been reached, can be quite complex in itself and may also require some degree of smart or even intelligent behavior, especially in complex or fast-changing environments.

In addition to being able to carry out actions independently, in certain situations, autonomous machines should also be able to cooperate with other agents (people or machines) or to respond to their actions. Thus, another target outcome is called **Interaction** (5). This is an important feature in any situation where an autonomous agent does not operate in isolation. For example, for autonomous vehicles or chat-bots, interaction is key (Behere and Torngren 2015; Stucki et al. 2018, 2020).

Finally, artificial intelligence can also be applied to generate something completely new and original. The corresponding target outcome is **Creation** (6). This outcome implies more than merely imitating something that already exists by varying some parameters.

Although the target outcomes are listed as a sequence, potentially inducing the feeling that producing one type of outcome requires the ability to produce the previous outcome on the list, this is not always the case. Often, a system only aims to achieve one type of outcome. For example, a biometric authentication system typically only produces identification. More importantly, many systems create the illusion of intelligent behavior by skipping several target outcomes, mainly by mapping an *Identification* outcome directly to an *Action* (see the dotted arrows from (1) to (4) in Fig. 5). Intelligent interaction can equally be simulated, for example, by applying a cleverly designed decision tree.

Although the second position of the sequence in Fig. 5 might suggest that the outcome *Understanding* (2) is not difficult to achieve, the authors are not aware of any AI systems that master it. Recent advances in machine learning technologies have not contributed much to achieving this outcome. It is without a doubt one of the tougher hurdles – but an essential one – in the quest for artificial intelligence.

3.4 Technology: Approaches to Artificial Intelligence

There are many ways to implement systems that may be considered artificial intelligence. Any systems developer must make a number of fundamental choices before building an AI system, in particular that of selecting the basic approach used to solve a specific problem. Some developers believe that verified facts and logical reasoning are essential to creating systems with some level of intelligent behavior and hence adopt a deterministic approach (Lazarowska 2017; Mostafavi and Barnes 2017). Other researchers and developers recognize that the world is full of uncertainty and suggest as the central tenet of their approach a way to deal with this uncertainty (Shachter et al. 2017; Yager and Zadeh 2012). Many experts believe that the most useful information can be found in historical data and that statistical analysis is the best basis for meaningful results (Ma et al. 2017; John et al. 2016). However, another starting point would be an attempt to comb through all the possible options of a problem and to assign a score to each potential outcome. Early attempts to mimic the underlying mechanics of how the human brain learns were based on different premises than any of the approaches above before evolving into modern artificial neural networks.

These fundamental choices determine from which angle to best tackle the problems at hand and define the main categories of approaches in this framework.⁶ Fig. 6 illustrates the main categories of approaches and some of their variants.

The category **Artificial Neural Networks** has evolved from attempts to mimic the basic mechanisms of the human brain – specifically, the architecture of neurons connected to other neurons via synapses and axons. The reinforcement and weakening of connections, depending on the past use of those pathways, is simulated with values allocated to nodes (i.e., neurons) and connections (i.e., synapses and axons). Mathematical functions process and adjust those values during the training phase of an artificial neural network. These networks consist of successive layers of nodes and edges (i.e., the connections between nodes). The essential task of such a system is to adapt its internal configuration to enable it to recognize patterns in data. Slight design variations, such as the way in which the nodes are regrouped and information is propagated inside the network, the mode in which errors are minimized during the training phase, the functions used to process values, or the integration of feedback loops to account for previous results, have led to a number of different forms of artificial neural networks. Variant forms of those networks include, for example, *generative adversarial networks* (Wang et al. 2017), *convolutional neural networks* (O'Shea and Nash 2015) and *recurrent neural networks* (Bodén 2002). Modern artificial neural networks' inherent strengths include a relatively high tolerance for fuzziness or uncertainty, and they can easily be trained to recognize a large variety

⁶Note that, in practice, several different approaches may be used in combination to accomplish the best results. This deliberate categorization is based on a thorough literature analysis that reflects the understanding of the authors. It is not a final state, but rather a dynamic wheel that must be continuously adjusted to incorporate future developments.

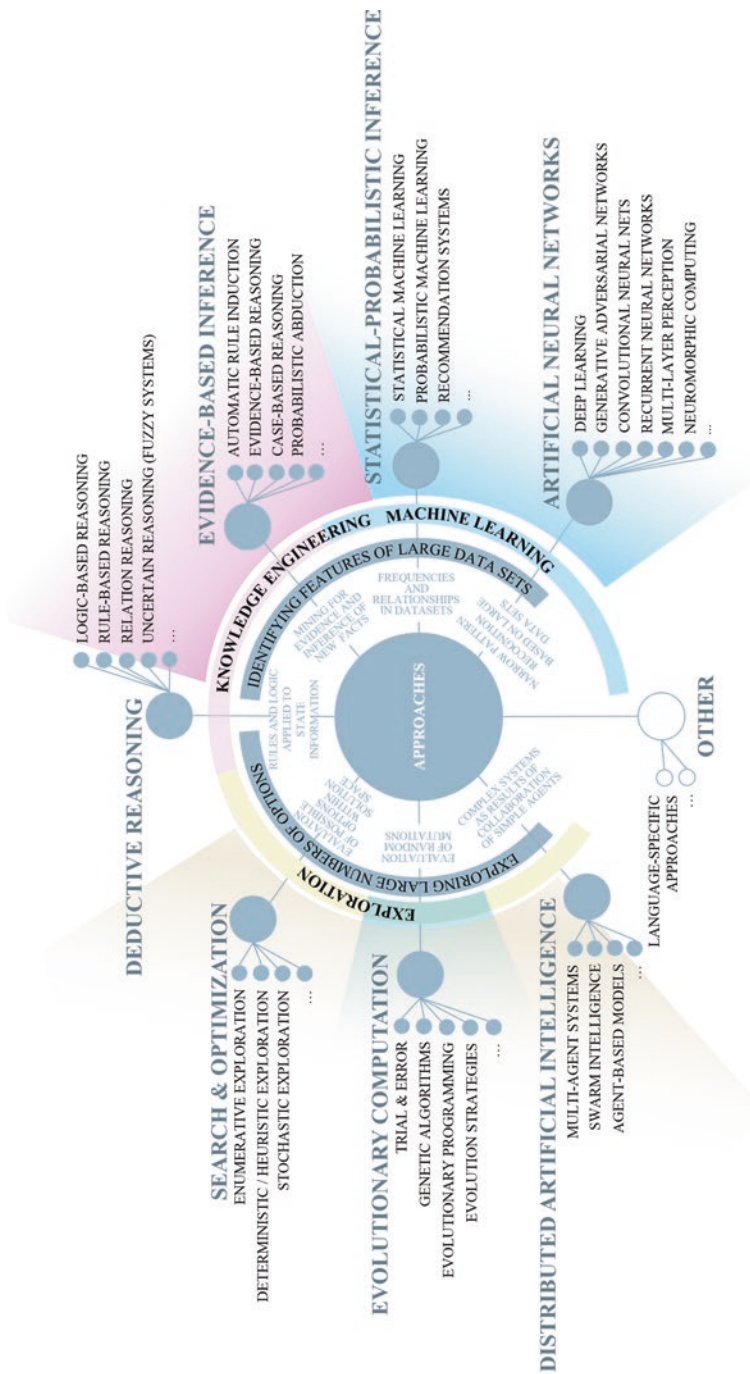


Fig. 6 Classes of approaches

of patterns. This makes them effective *machine learning* systems.⁷ In its broad sense, the term *machine learning* is applied to systems that are able to *learn* based on various learning strategies, which will be discussed later in this section.

Statistical-Probabilistic Inference is another category of approaches that can be used to implement AI systems. The process of statistical inference consists of analyzing a sample of data from a larger population and drawing conclusions from its features (Lindgren 2017). There are also models that describe samples of data using probabilistic distributions (Mitzenmacher and Upfal 2017). The main motivation for using probabilistic over nonprobabilistic approaches lies in their ability to process uncertainty or frequency. If uncertainty is a key component of a problem, probabilistic approaches allow the generation of approximative models. The authors chose to group both approaches into the same category, as they share the same fundamentals: to generate statistics from a sample of data, representing frequencies and relationships in datasets (Johnson and Bhattacharyya 2019).

Approaches using **Evidence-based Inference** form another distinct category. They have a central element of knowledge engineering that refers to the development of systems that may process, link and restructure large amounts of information from different sources, using rules and argumentation mechanisms for solving problems (Wu et al. 2015). *Evidence-based Inference* is about reasoning based on facts or on *evidence* – a weaker form of facts, and it can be vague, ambiguous and not necessarily verified (Bird 2018). Evidence gathered from various sources may be restructured and combined to create new evidence (Introne and Iandoli 2014). Although *Evidence-based Inference* systems may implement some statistical methods, the computation basis differs from the *Statistical-Probabilistic Inference* because it uses information that already offers a certain degree of certainty or interpretation (Wilkinson n.d.). For example, if the weather forecast predicts rain and planned activities are likely to involve covering some distance on foot, then one conclusion may be that taking an umbrella is an appropriate course of action. Known variants of evidence-based inference include *automatic rule induction* (Stepanova et al. 2018) and *case-based reasoning* (Kolodner 2014).

Deductive Reasoning refers to approaches that use a process of reasoning, with the aim of transforming input into output by using logical deduction techniques or predefined rules, whereby a conclusion is drawn from a set of propositions (e.g., “*if it is snowy and cold, then one dresses warmer*”) (Bradford 2017; Trillas et al. 2019). *Rule-based reasoning* (Williams and Wood 2015) and *relation reasoning* (Egenhofer 2015) are examples of approaches in this category, as is a variety of hard and soft computing approaches. While rule-based approaches tend to be based on exact models (i.e., simplified models of reality with exact data models, mostly with numerical inputs) (Williams and Wood 2015), fuzzy systems have been developed to introduce a tolerance for uncertainty and ambiguity (e.g., allowing linguistic variables) (Zadeh 1988). Although they represent two different ways of drawing conclusions – *exact* versus *approximate* – both types of systems use deductive reasoning. Note

⁷Note that, in today's machine learning systems, the learning part usually takes place in a distinct training phase and not during normal operation (depending on the chosen learning strategy). It is, thus, not equivalent to the way humans learn on the fly.

that while *Evidence-based Inference* shares some traits with reasoning approaches, its use of different learning strategies (*inductive* vs. *deductive* (Bradford 2017)), its stronger focus on knowledge engineering, its emphasis on text-based inputs rather than numerical values, as well as other characteristics, set it apart from *Deductive Reasoning* approaches.

While the approaches presented above, particularly *Artificial Neural Networks*, *Statistical-Probabilistic* and *Evidence-based* systems, emphasize analyzing *inputs* by extracting features and patterns from large datasets, the approaches of *Search and Optimization* and *Evolutionary Computation* focus more on possible *outputs* by exploring a large number of options or possible solutions. In the category ***Search and Optimization***, for instance, many problems can be solved theoretically by searching through many possible solutions in a smart way. The basic approach to optimization is to formulate a fitness function that sequentially evaluates the performance of all possible options to identify the global optimum (Tang and Wu 2011). However, with a large solution space and limited computing resources, it is often not possible to explore all available options, or at least not within a reasonable cost and time frame. In cases where the solution space is very large and the exploration of possible options cannot be exhaustive, a search strategy must determine the order in which the available options are evaluated. *Enumerative exploration* (Herceg et al. 2015) and *stochastic exploration* (Procter et al. 2019) are examples of *Search and Optimization* approaches.

Evolutionary Computation is an umbrella term for a series of problem-solving techniques based on the principle of biological evolution for optimization reasons. In general, the aim is to create a population, to evaluate it based on a function (e.g., fitness function), to select suitable candidates and, finally, to reproduce them with random mutations. This process lasts until an exit criterion is fulfilled (Xue et al. 2016). This category includes, among others, the approaches of *genetic algorithms* (Bakirtzis and Kazarlis 2016) and *evolution strategies* (Salimans et al. 2017).

Evolutionary Computation is conceptionally related to *Search and Optimization*. Both approaches explore and evaluate available options to find an optimum. However, the two approaches are listed separately for the following reasons:

- (1) The aim of *Search and Optimization* approaches is to evaluate *all* of possible options, while *Evolutionary Computation* approaches explore the solution space step by step, without concern about what or how large the space is.
- (2) While *Search and Optimization* approaches implement exploration strategies, *Evolutionary Computation* approaches do not care to define an order in which to proceed. They use randomly altered versions of models and select the best-performing variants as a basis for further mutations.

Another category of approaches regroups ***Distributed Artificial Intelligence***. It refers to multiagent systems, inspired by the way that biological insect colonies work. In principle, they are composed of simple, relatively unsophisticated agents, each implementing only a small set of rules, with a focus on interagent communication and collaboration. However, through multiagent interaction, they can form systems capable of complex behaviors or executing tasks and pursuing goals that go far be-

yond the capabilities of each individual agent. In such *Distributed Artificial Intelligence* systems, there is usually no centralized control structure dictating how the individual agents should behave. Thus, cooperation is crucial to achieving the goals and tasks set (Crowder and Carbone 2016). Local and, in many cases, quasi-random interactions between agents lead to the emergence of a smart global behavior unknown to the individual agents (Siemens 2005; Zambonelli et al. 2015).

While the authors believe that the categories described above cover most of the approaches to artificial intelligence to date, the framework also contains an open category. It includes, for example, some *Language-specific Approaches* that do not fit into any of the other groups, and it leaves room for future developments that would be based on entirely different principles. The popular notion of *Natural Language Processing* (Kurdi 2016, 2017) as such is not a technology approach but a capability and, therefore, does not figure in any of the above categories.

It is worth reiterating that the distinctions between the categories presented above are based on the *basic tenets* underlying the different approaches that they contain. The boundaries between categories are by no means sharp; common methods can be shared by approaches from several categories, which may even use implementations belonging to another category. While *Distributed Artificial Intelligence* systems' most fundamental principles are based on multiagent collaboration, it is possible that each individual agent of such a distributed system implements an artificial neural network, a rule-based reasoning system, or both. Furthermore, the distinction between these approaches is useful, as they exhibit different strengths and weaknesses and are more or less suitable to achieving certain outcomes. Every category counts different variations designed to reinforce strengths or to mitigate the inherent weaknesses of the main approach. For instance, the category of *Artificial Neural Networks* is declined into *convolutional neural networks* (i.e., strong pattern recognition, e.g., image analysis) and *recurrent neural networks* (i.e., adding a capability to process sequential events) among other variants.

In summary, the fundamental choices underlying a technology approach determine many of an AI system's characteristics and its strengths and weaknesses, even before specific solutions and methods are developed. The authors distinguish seven categories of approaches, as laid out in this section, that share similar fundamentals. Identifying the category to which a system belongs can provide a first indication of its potential and limitations. Note that composite systems that integrate several subunits with technology approaches from different categories may provide a way of combining advantages of each into stronger overall systems.

3.5 Technology: Learning Strategies

Although most modern commercial *machine learning* systems implement either *statistical-probabilistic* approaches or *artificial neural networks*, they are primarily characterized by a learning strategy that determines some of their data requirements, strengths and weaknesses. Without delving into too much detail, the authors acknowledge the importance of the choice of a learning mechanism for any learning system and present a short synoptic in Table 1. The list is not exhaustive.

Table 1 Learning strategies

	Strategy	Description
Deductive	<i>Control Learning</i>	Control learning (or iterative learning control) is a way of learning through repetition, with transfer of information from previous iterations (Bien and Xu 2012)
	<i>Curriculum Learning</i>	Curriculum learning is a process in which the contents are presented step by step, in a predetermined order that aims to facilitate the acquisition of concepts of increasing complexity (Bengio et al. 2009)
	<i>Hierarchical Learning</i>	Human problem solving (incl. learning) offers the possibility to conceptualize the world at different levels of granularity and abstraction and to treat problems hierarchically (Hobbs 1985)
	<i>Imitation Learning</i>	Imitation learning, sometimes also called <i>learning by example</i> , refers to the process by which a learner imitates the behavior of an expert ((Hussein et al. 2017)
<i>Transfer Learning</i>		Transfer learning is the process of training one instance of a system, and then copying the resulting model to a different instance or an entity for use (Weiss et al. 2016)
Inductive	<i>Supervised Learning</i>	Supervised learning uses a set of input data with known responses to the data (output) to generate a model of the perceived reality, with the aim to then generate reasonable predictions as a response to new data (Kotsiantis 2007)
	<i>Unsupervised Learning</i>	Unsupervised learning is a process of finding hidden patterns or intrinsic structures in data, and establishing a model based on inferences drawn from datasets consisting of input data without labeled responses (Chandrashekar and Sahin 2014)
	<i>Semi-supervised Learning</i>	Semi-supervised learning mainly makes use of unlabeled data for training but also adds a small amount of labeled data for verification purposes (i.e., combination of supervised and unsupervised learning) (Chandrashekar and Sahin 2014)
	<i>Reinforcement Learning</i>	Reinforcement learning is a goal-oriented learning strategy. It consists of iterative cycles of <i>observing – taking action – being rewarded or punished</i> . Agents gradually optimize actions in an environment, to maximize the rewards (Wiering and van Otterlo 2012)
	<i>Adversarial Learning</i>	Adversarial learning aims to enhance learning through competition against other agents (Huang et al. 2011)
	<i>Agency-Driven Learning</i>	Agency-driven learning (or exploration) is related to reinforcement learning, with the difference that its rewards are not determined by factors or agents outside the learning system, but by <i>intrinsic motivation</i> (i.e., system-internal evaluation of rewards) (Facer 2011)

The learning strategies belong to either the *deductive*, *inductive* or *abductive* reasoning category and apply to humans as well as artificial systems. **Transfer Learning** is a special case in which the actual learning happens in one instance and the resulting model is then copied to a different instance. While the model generated by the first system is based on a strategy belonging to one of the other two main groups of reasoning, there is no actual learning involved for the receiving entity. Transfer learning is common in *machine learning* systems, where the configuration, weights, and relationships (model) are determined in one system to then be used in one or several others (Weiss et al. 2016). An example from the natural world is the way that biological entities encode adaptations to an environment into genetic material, which is then copied and transferred to new individuals.

Deductive Reasoning starts with a set of rules (that are consistent and do not contradict each other) and, given an input, proceeds with logical reasoning to reach a conclusion that respects these rules (Bradford 2017; Trillas et al. 2019). **Inductive Reasoning**, however, starts with observing a dataset and trying to infer general rules from it. Subsequently applying those rules to a given input does not lead to a proposition that is necessarily a logical consequence. In essence, it is about extracting information from a large dataset from which knowledge can be gained, without determining in advance what is being searched for (Hayes and Heit 2017). **Abductive Reasoning** is similar to *Inductive Reasoning*, with the distinction that it uses a hypothesis of what seems to be *the most likely explanation* in a given situation (Walton 2014). In the proposed framework, *Abductive Learning* strategies have been included in the category of *Inductive Reasoning*.

Belonging to *Inductive Reasoning*, three often referenced learning strategies that are used by *machine learning* systems are *Supervised Learning* (e.g., image recognition (Carneiro et al. 2007)), *Unsupervised Learning* (e.g., predicting noise (Bojarski and Joulin 2017)) and *Semi-supervised Learning* (e.g., sequence learning with recurrent neural networks (Dai and Le 2015)). Supervised learning requires a set of input and output data, the generation of which can be an expensive process and, in many cases, the required training data simply does not exist. In general, all three learning strategies tend to require a large amount of data and a distinct *learning* phase, in which, a model is established. A quite different set of learning strategies includes *Reinforcement Learning* and *Agency-driven Learning*, which are prevalent in nature. Both learning strategies require carefully designed and tuned reward functions for effective learning but do not usually require a separate training phase, allowing an agent to adjust its behavior during normal operation (Facer 2011; Wiering and van Otterlo 2012).

Another strategy that is commonly used in *machine learning* systems is *Adversarial Learning*. For AI systems, this entails an oppositional setting in which two or more machine instances produce results that are evaluated against each other through many cycles (e.g., playing chess (Chidambaram and Qi 2017)) (Huang et al. 2011). Recently, researchers and developers in the field of AI have also gained interest in learning strategies from the category of *Deductive Reasoning*. There have been efforts to use automated *Curriculum Learning* for artificial neural networks (Graves et al. 2017) or to adapt the concepts of *Hierarchical Learning* to machine

learning (Barto and Mahadevan 2003), and several algorithms have recently been proposed for the task of *Imitation Learning* (Osa et al. 2018).

3.6 Enablers: Data

A crucial factor in deploying AI systems is the availability of *appropriate* datasets. Their importance is never to be underestimated: at least an equal emphasis should be put on ensuring the availability of suitable data to select the best AI technology or application. Commonly, the required data are not readily available – they may have to be transformed, engineered, or newly generated and this may require an update to a company’s data strategy ahead of time (DalleMule and Davenport 2017; Mazzei and Noble 2017). While all information systems rely on suitable data to function, autonomous learning systems compel additional focus on the data used to train the systems. The use of ill-formed training datasets can lead to unintended but important consequences. Data requirements should thus be considered carefully, both from a quantitative perspective and from the point of view of the quality and composition of datasets. Failure to do so may result in an inability to generate the required models, or worse, lead to bad decisions in operation. In a world, wherever more decisions are made autonomously by AI systems, it is vital that the conclusions reached by those systems are reliable and applicable (de Bruijne 2016).

All datasets represent a filtered view of reality and, as such, they all include certain *biases* or *variances*. A bias is created by any imbalance in a dataset, such as gaps in time series, missing values, absence of a category, discrepancies in the number of records with different characteristics, an under- or overrepresentation of a feature – the list goes on. Thus, the system does not work based on fully accurate assumptions. A variance represents the sensitivity to minor variations in a training dataset. A high variance causes noise to be modeled instead of relevant characteristics. Recognizing and eliminating such imbalances is extremely difficult. However, it is crucial to be aware of any such biases present in a dataset used to train a system, as they are highly likely to be reflected in the resulting models and ultimately in the decisions made by the machine while in service. Certain partialities may or may not be acceptable, whereas others can result in dangerous or illegal actions (Fitzgerald 2014).

Consider the example of loan applications evaluated by a machine. While human employees of a financial institution, as well as traditional information systems, typically use a set of rules to assess a loan application, an AI system built on an artificial neural network and using a strategy from machine learning would reach its conclusions based on patterns identified in the dataset used to train the system. Now, assume that the training data are obtained from historical data present in the business support systems of the firm – a likely scenario in a real AI project. In this case, there is a rich and complete dataset of past loan applications – information provided by the applicants, the rationale and decisions reached by the human experts – over the previous decades. Furthermore, it is thought that the data are of excellent quality, that there are no missing records, that correct decisions were reached in all

cases and that the rules used were good rules. The absence of any quality issues with the training data would suggest that it should be possible to generate an accurate AI model to automate the evaluation process for loan applications. However, the unsuspecting financial institution might be surprised to face discrimination lawsuits after a few years of operating the resulting loan assessment bot. In this instance, biases in the model could be the result of changing cultural and societal conditions: while in today's society it is expected that loan applications can be equally submitted by both genders, perhaps the historic data contained only very few and mostly unsuccessful loan applications from women and possibly even fewer applications from people with certain ethnic backgrounds. Within the black box of the AI model, gender and ethnicity may thus be important contributing factors in the decisions reached by the machine.

It is thus important to cautiously examine any training data for biases. Any corrective measures must be carefully designed and assessed for unintended side-effects, and the resulting models must be tested and remaining biases identified and quantified (Fitzgerald 2014). This is by no means trivial and requires the specialized skills of highly qualified data professionals. The required expertise to ensure that the data requirements for AI systems are met is different from that of traditional data management. However, there is a significant overlap with the skills required for advanced data analytics, which is one of the main reasons why establishing *machine learning* models and managing their life-cycle often falls within the realm of a company's data science team.

An insufficient *quantity* of adequate training data can also be a cause of failed automation attempts. For instance, while large strides have been made in processing natural language with the use of modern techniques, among others based on machine learning, the best results have without a doubt been achieved in the English language and, usually with a few years lag, some of the other widespread languages such as German, French and Spanish (Belinić 2018). However, for companies attempting to implement speech recognition systems in countries using less prevalent languages or with strong dialects, the lack of sufficiently large corpora of well-labeled training data (e.g., audio recordings with corresponding transcriptions and meta-data describing the topic, the context, the outcome and other characteristics) is often an insurmountable challenge. Sourcing or generating the necessary data for the intended context can be a lengthy and expensive process (Forsberg 2003).

Organizations that envision automating processes (e.g., call center processes, customer interactions) with the use of AI systems should have a plan for sourcing the required training data well before procuring a system. To do so may require timely adjustments to a company's data strategy. A data strategy aims to ensure that the collection, aggregation, storage, deployment, maintenance and archiving of data, as well as the preparation, manipulation, enrichment and transformation of the collected and stored data to be used or reused with value, are effective for achieving the defined targets (DalleMule and Davenport 2017). There are many papers relating to data governance and data strategy (Alhassan et al. 2016; Cheong and Chang 2007). However, traditional data strategies and data management processes are typically not optimized for the enhancement and reuse of data assets outside their

primary context, nor do they cover the additional data requirements for AI systems (Constantiou and Kallinikos 2015).

The idea that intelligent systems can simply be deployed into a new context and will then learn on their own from whatever data they encounter is surprisingly widespread. Instead, substantial efforts are required to configure such systems to adapt them to operate within a firm's processes, to select the optimal model among hundreds, to source and prepare adequate training data, and to test and measure the resulting behavior. Data preparation alone includes such tasks as data quality measures, data cleansing, data denoising, feature engineering and synthesis, and parameter- and hyperparameter-tuning, whereas verifying the resulting behavior involves checking for overfitting or underfitting and drifting generally for output biases. In many cases, such models must be periodically retrained with new data, which raises questions of tracking model versions and data lineage and, more generally, of managing machine learning life-cycles. Updated models must again be tested and redeployed into production (Han et al. 2011; Witten et al. 2016). The above lists of tasks are non-exhaustive, and while the emergence of automated machine learning can help with a number of tasks such as model selection, it is the authors' observation that even data professionals themselves will often underestimate the effort required.

Of course, learning processes are not the only aspect of AI systems where data are key. In particular, internal knowledge management (incl. evaluation of new data and extracting relevant information from it) as well as the capability of building concepts and linking them to previous knowledge are of note with regard to managing data. However, the heightened focus of this section on model training seems justified not only by its current relevance to many organizations but also by the time and responsibilities regarding data management. Internal knowledge representation capabilities are built into an AI system as part of its core design, typically under very *controlled conditions*, by the specialists building the system. Model training, on the other hand, is often the responsibility of the deploying organization and is *pragmatically* done by a project team with the data ready at hand, leading to the data challenges laid out in this article.⁸

While it is possible that future, more mature, AI systems may replace today's model training with a continuous learning approach more similar to human learning, where the internal knowledge representation and behaviors are shaped by a combination of preprogramming, perception, communication, and experience, the authors believe that it is premature to address the new data challenges that would be associated with this in this article. For live data feeds during operation, it would be expected that AI systems with a high degree of maturity should be able to handle any type of data gathered from their environment, isolate the parts relevant to their mission and appropriately deal with errors or unclear situations, thus not requiring any particular attention to data from the deploying organization. In today's systems of limited maturity, however, it may still be necessary to control the types and quality

⁸While pre-trained models are increasingly more prevalent, deploying organizations will mostly still have to train models for a particular task or context.

of data inputs, whether they originate from perception or are ingested via other interfaces. Ill-suited data may simply result in the systems' inability to produce meaningful results, while deliberately ill-formed inputs might allow misuse.

3.7 Robustness, Transparency and Trust

Advances in machine learning, in particular *deep learning*, have had the side-effect of also highlighting a number of weaknesses. While some of the observed flaws are not new, their significance certainly reaches another level with the recent commercial successes and the strongly growing use in the production of AI systems. For instance, pattern recognition with artificial neural networks has been found to be very vulnerable to the addition of minute noise patterns, opening them to a number of potential attacks. By changing a small number of pixels in an image, imperceptible to the human eye, systems can be easily manipulated to reach incorrect conclusions (Su et al. 2019). The real-world repercussions of being able to fool AI systems can be critical: autonomous cars that can be made to accelerate when perceiving a stop-sign, with just a couple of added stickers in specific places (Eykholt et al. 2018), is only one example. Although efforts are being made to mitigate such weaknesses, once they have been recognized, the authors wager that the problem will be here to stay. The sophistication of potential attacks exploiting newly discovered weaknesses is likely to evolve in parallel with any mitigating measures – much like ever new vulnerabilities are found, exploited and patched up in traditional IT applications.

The types of weaknesses presented in the example above are not the only concerns raised by AI systems. The authors suggest breaking down the issues around robustness, transparency and trust into the following topics:

- **Robustness:** Adversarial Robustness, Resilience to Errors, Reliability
- **Transparency:** Transparency, Traceability
- **Trust:** Fairness, Consistency/Predictability, Alignment of Goals

All of the above themes, which are discussed later in more detail, should be important considerations before training or deploying any AI system. Failure to adequately address them might result in exposing the owner to liabilities. The authors anticipate that the relatively young disciplines around robustness, transparency and trust in the field of AI will be gaining ever more traction.

In addition to these issues, the following topics are equally gaining in importance and should be addressed when deploying autonomous systems in the field: opportunities for potential misuse (e.g., the emergence of deep fakes (Chesney and Citron 2019)), legal challenges (particularly around liability) (Cath 2018) and ethical questions (Bostrom and Yudkowsky 2014). Some of these concerns are exacerbated by the increasing degrees of autonomy of a system. Their detailed discussion is, however, beyond the scope of this article.

3.7.1 Robustness

Adversarial Robustness describes how vulnerable or hardened a system is to external attacks. Whether a system can be hacked and illicitly controlled and whether a targeted attack requires deep expert knowledge and privileged access is one aspect of adversarial robustness. However, it also includes the notion of how easily a system can be fooled to reach the wrong decision without actually hacking it (Stutz et al. 2019). For instance, some computer vision systems can be shown pictures containing specific micropatterns that fool them into recognizing objects that are not actually present (Vincent 2017). Adversarial robustness tends to have direct implications for security and a potential for misuse (Stutz et al. 2019).

Resilience to Errors acknowledges that errors do happen and refers to the *rate*, *gravity* and *handling of such errors* (Drozdyuk 2017). For instance, in object recognition, it is not possible to know the constellation of objects in advance or to predict the noise in a live video feed (Siruvuri et al. 2009). The question then becomes how likely it is that a pedestrian will be mislabeled as an inert object in various conditions (*error rate*), what the potential damage will be if a pedestrian in motion remains unregistered (*gravity*), and how the system will deal with the margins for error (*handling*). If an autonomous vehicle's control system recognizes an object as a fire hydrant with 99.8% certainty and as a pedestrian with a 0.2% chance, and reaches the conclusion that the object is inert, estimating its *average error rate* for similar situations would be 1 in 500 instances. However, the *gravity* of such an error occurring might be high, as a potential consequence of mislabeling the object could be that the pedestrian is run over. The system may choose to take no action on account of its 99.8% certainty, or it may choose to proceed cautiously, despite the low chance of causing any damage. How a system handles potential errors and uncertainties, whether it contains fail-safes, or how it deals with actual inaccuracies contributes to defining its *resilience to errors* (Ruth and Goessling-Reisemann 2019).

Reliability generally refers to the probability that a system will not fail throughout a prescribed operating period. In this context, *failure* is an insufficiently delimited criterion, due to the complexity of situations that AI systems are designed to deal with and the variable nature of the expected outputs. The reliability of an AI system includes the ability of the system to produce consistently adequate results, regardless of the situation, or to repeatedly reach valid conclusions under varying conditions (Kolowrocki 2014). This is important for systems that are designed to autonomously make decisions in an ever-increasing array of processes, while the validity of the decisions reached in every situation is difficult to establish. They must thus not only remain functional but also produce consistent results in a variety of circumstances.

3.7.2 Transparency

Many AI systems, such as the human brain, are black boxes to any external entity and thus **Transparency** is lower or not given. With the growing complexity of systems, their inner workings become increasingly more mysterious. For example, recent progress in deep learning has given rise to a new field of research that seeks to

understand how modern artificial neural networks with thousands of hidden layers reach their conclusions. While the algorithms at work are generally known, the reality is that the detailed effects of weights and values of a trained artificial neural network cannot be understood by humans. Why is this a problem? Humans make decisions (with their own *black boxes*), and those decisions may well be (or appear to be) arbitrary in many situations. Nonetheless, the possibility usually exists to make the decision process reproduceable or render the rationale for the decisions transparent through some process: explicit rules that stipulate how certain information must be processed to lead to a conclusion or explanations after the fact. In many AI systems, the rules applied and the decision processes are either implicit, indirect, too complex to understand, or simply not known. In fact, in complex situations, it could even be difficult to assess whether a presented outcome is correct or to submit valid arguments to challenge it. Moreover, in the case of an identified error or accident, it could be excessively difficult to diagnose the reasons for the failure. Knowing that today's artificial neural networks are very performant but can, on occasion, make extremely silly mistakes that no 3-year-old child would make, attempts at poking holes in the black boxes for better transparency seem appropriate. Some researchers and developers are hard at work on the problem by attempting to integrate into AI systems algorithms that quantify the weight of each contributing factor or parameter in a decision or conclusion (Bathae 2018; Phillips-Wren 2012).

Traceability views the problem of transparency mainly through the lens of compliance. For many real-world applications, the law requires that the party accountable for a process be able to provide a step by step account of how a decision was made (Srivastava 2018). With AI technologies, this is notoriously difficult to achieve, given the abovementioned inherent lack of transparency.

3.7.3 Trust

Fairness refers to the presence or absence of biases in the models and decision processes. Such biases are usually not trivial to detect without targeted analysis, but they are a crucial element of trust and may have legal implications. If someone's application for a mortgage is rejected by a bank's machine learning algorithm, it is important to ensure that the denial is not the result of racial, gender, or any other discrimination (DefinedCrowd n.d.). If this sounds obvious to the reader, consider how easily distortions can be present in any historic dataset. In fact, it is difficult to find and prepare datasets that do not present various imbalances. During the training phase of a *machine learning* system, any such biases would typically be transferred to the model but not be apparent. A careful analysis and preparation of the training data is therefore crucial but not sufficient. Depending on the application, the models themselves, as well as the results achieved in production, must also be monitored for any biases (Fitzgerald 2014). This is a rapidly growing concern, and some organizations have already made it their field of expertise.

Consistency or Predictability also engenders *Trust*. If people can expect that similar initial conditions lead to similar decisions relevant to them (like they usually do with traditional IT systems), this may, to some extent, allow us to predict the

behavior of a machine, to understand the results, or to be able to relate to them (Kim 2018). This would help prevent sentiments of injustice and help with the acceptance of autonomous machines. For example, in the case of autonomous agents piloting physical equipment, predictability of movement is also a main security factor (Houser 2019).

Alignment of Goals is a longer-term consideration with autonomous agents. Many AI systems will have implicit objectives or explicitly programmed goals or at least a reward function. Depending on the degree of the autonomy of a system, the machine will pursue these goals or seek to maximize its rewards without further human guidance (Johnson et al. 2018). The caveat is that humans tend to have shifting values, motivations, goals, alliances and priorities, requiring changing behaviors over time. In a world where yesterday aspirations can easily become today's nightmare scenario, it is not difficult to imagine situations where autonomous machines with unchanging objectives can become a liability.

4 Discussion

Finally, the authors would like to add a few thoughts and comments, as well as their view on the current maturity level of AI systems and on possible paths of evolution.

4.1 Maturity, Technology and Approaches

Current AI technologies have undoubtedly reached a level of usefulness that justifies a renewed focus on the subject. However, there still lies a long path of evolution ahead before any AI systems can be considered *intelligent*. The authors speculate that climbing higher on the maturity ladder will lead to a shift in focus, away from monolithic learning systems and towards composite systems, integrating several subunits based on technology approaches from different categories. The question of which one of the presented fundamental approaches will eventually prevail is therefore probably moot. Composite systems, integrating several subunits with technology approaches from different categories, may provide a way to combine the strengths of each into stronger overall systems. Combining multimodule systems with a central control unit to orchestrate the capabilities of subunits using different approaches may prove to be a key step on the artificial intelligence maturity path. Considering *target outcomes*, large strides have been made to accomplish various tasks of *Identification* (1). On the other hand, much progress remains to be made to achieve systems that understand what they have identified. It seems inevitable that achieving something equivalent to an *Understanding* (2) of situations must lead via the construction and integration of complex concepts of objects and ideas. Whether this will be realized by mimicking the way that humans construct concepts or by using novel ways remains to be seen.

4.2 The Dangers of AI: Misalignment of Goals

AI has been dubbed *humankind's last invention*. Doomsday scenarios based on sentient AI systems taking over the world and fighting humankind abound. Today, many credible speakers warn about the dangers of AI, and among them are some illustrious scientists who believe that we will eventually lose control of AI entities once they reach a certain (usually undefined) threshold of *intelligence*. However, it could equally be speculated that higher levels of intelligence would possess empathy and a better understanding of other lifeforms and might therefore be less likely to develop antagonistic behaviors. Such long-term predictions are notoriously difficult, and the likelihood of any one of the possible developments occurring and leading precisely to the imagined consequences is anyone's guess.

However, the authors would like to point out some real and tangible risks associated with the deployment of AI systems, even in the short and medium term – namely, a *misalignment of goals*. More precisely, it is not AI per se but the associated larger level of autonomy combined with adaptive behavior that may present a more immediate danger. It must be assumed that any sophisticated autonomous system would pursue missions or objectives. These can be explicit or implicit within the programming or models, perhaps even via a simple optimization function. At the time of their deployment, the missions of autonomous systems are typically expected to be in alignment with the goals of their owners. The problem arises when humans change their intentions over time – as they frequently do – or when the general context changes. Five, ten or twenty years after their deployment, no one may know what the status of an AI system is and what goals it is working towards.⁹ In the case of mobile robots capable of adaptive behavior, it may not even be known where they are or if they still exist, until they emerge in unexpected situations. An obvious example would be autonomous defense systems, reappearing after a conflict is over. However, seemingly innocuous applications may also cause damage if goals become misaligned or obsolete, whether they are machines making decisions that affect people, armies of builder-bots that autonomously construct physical structures, or nanobots designed to eliminate a particular type of pollution (an element considered harmful in one environment may be a vital part of another ecosystem) (Gronlund 2019; Nield 2016).

Similar to any new technology that holds great promise, there is also the potential for misuse. In the case of AI, this should not be underestimated: The adaptive capabilities and autonomous nature of replicable agents might enable misuse on an unprecedented scale – more efficiently, more anonymously, more furtively, and more remotely, both in terms of time and space. A focus on enhancing the adversarial robustness of AI systems is crucial but insufficient to prevent misuse, and the authors believe this to be a far more tangible threat than any other.

⁹Similar difficulties can also routinely be observed in large IT organizations running traditional information technology, where it is not uncommon to have a number of legacy systems in operation that are nearly impossible to decommission, mainly for the reason that no one knows precisely what services they support and for whom.

4.3 Practical Considerations

Focus on data should precede any deployment project and ideally be the object of a data strategy that explicitly includes data for the purpose of generating AI models. The composition of (training) datasets is intrinsically linked to *Robustness*, *Transparency* and *Trust* and, in most cases, it is far from trivial to find adequate datasets on short notice. To avoid falling into the discrimination trap and manage risks, companies deploying AI systems should raise awareness among all employees involved about the topics of *Robustness*, *Transparency* and *Trust* early in the project. Datasets need to be prepared (sometimes engineered), verified for biases and measured at various points in the preparation process. The results produced by previously trained models also need to be verified for biases and measured, and there should be a reliable baseline to measure against.

Data preparation is by no means trivial and requires the specialized skills of highly qualified data professionals. The availability of the right skillset within the firm to prepare and measure data and, in some cases, to select the most adequate model among dozens available, is an important enabler. While the recent rise of *automated machine learning* (or *driverless ML*) – a small step towards automating the generation of AI systems themselves – may help with model selection, it does not alleviate the need for skillful data preparation, which still accounts for approximately 70–80% of the work. In addition, once a few operational *machine learning* solutions have been retrained and redeployed with updated models, the need for traceability and a robust process for keeping track of lineage and versioning of both data and associated models tends to become obvious (Reifschneider 2019).

Considering the relationship between data science, advanced analytics¹⁰ and AI applications, applications based on machine learning, fuzzy systems or statistical-probabilistic approaches could be considered data analytics systems insofar as they may implement a set of data analytics methods and algorithms to automate the identification of patterns in data. This does not, however, imply that all data analytics can be automated with AI systems. In fact, only very few, very specific data analytics use cases are suitable for current AI technologies. Data scientists are professionals who have the necessary skillset to carry out advanced analytics. However, the same skillset – an understanding of data and statistical methods – enables them to prepare and manage datasets for AI models. Analytics methods may be applied to datasets as part of their preparation for use with AI models. AI applications may, in turn, be used by data scientists to automate some specific analytics tasks.

¹⁰The process of combining and rendering explicit valuable information contained in various datasets, interpreting it and enhancing/enriching it with additional information.

5 Conclusion and Outlook

AI researchers, system developers and vendors' marketing departments show a tendency to present the level of intelligence embedded in their own systems in a rather favorable light. Many AI systems only create an illusion of cognition by encoding behaviors imagined by the collective human intelligence of the development team into algorithms in such a way that an intricate mapping from inputs to actions is able to create clever outcomes. While accepting the de facto reality that a variety of technologies and applications, with more or less sophisticated capabilities, are traded under the banner *artificial intelligence*, the authors emphasize that current AI systems cannot actually be considered *intelligent*. They recommend using the term *intelligence* (on its own) in a stricter sense when assessing the level of maturity of any system.

Information systems have long been able to surpass humans in some very narrowly defined tasks. This is not a new phenomenon and is in no way specific to AI systems. Technologies such as deep artificial neural networks simply add a few more domains to the list. However, neither traditional information technologies nor artificial neural networks can currently enter into competition with humans in terms of versatility, adaptive problem-solving capabilities and the ability to deal with uncertain and complex situations.

Intelligence requires a variety of cognitive capabilities, such as problem-solving and knowledge representation, only a few of which have seen significant advances in AI systems over the past decade. While machine learning is one area where substantial progress has been made, many such systems have to be configured during a distinct training phase and, after deployment, they produce good results only under very controlled and constrained conditions. Thus, even the *learning* capabilities of modern AI systems are by no means on a par with the ability of humans and animals to learn continuously and to relate those learnings to particular contexts. More importantly, in regard to understanding the world and the objects that it contains, today's AI systems are still at a very embryonic stage of development. To meaningfully progress on the maturity curve towards truly *intelligent* behavior, AI systems will likely have to acquire the ability to form concepts of things and to relate those concepts to each other (i.e., cognitive linkages (D'Onofrio et al. 2018)).

While none of today's systems can be reasonably considered intelligent, speculations on whether or when this could become a reality vary in their conclusions. There is, however, for the first time, no more doubt about the usefulness and economic value of a number of AI technologies. Sceptics who predict a third *AI Winter* may find that such a crisis would only be partial, as those technologies and fields of research that have already demonstrated their practical business value would likely benefit from continued interest and funding.

As a final note, it is understood that this framework provides an overview of aspects that are relevant to today's AI technologies. As technologies evolve and new approaches are developed, they may have to be accordingly extended or amended.

Acknowledgements The authors would like to thank their colleagues of Swiss Post and reviewers at the Human-IST Institute of the University of Fribourg for their constructive feedback and their valuable inputs.

References

- Alhassan I, Sammon D, Daly M (2016) Data governance activities: an analysis of the literature. *Journal of Decision Systems*, 25(sup1):64–75.
- Bakirtzis A, Kazarlis S (2016) Genetic Algorithms. Eremia M, Liu CC, Edris AA (eds.) *Advanced Solutions in Power Systems: HVDC, FACTS, and Artificial Intelligence*. Wiley Online Library.
- Barto AG, Mahadevan S (2003) Recent Advances in Hierarchical Reinforcement Learning. *Discrete Event Dynamic Systems: Theory and Applications*, 13:41–77.
- Bathae Y (2018) The Artificial Intelligence Black Box and the Failure of Intent and Causation. *Harvard Journal of Law & Technology*, 31(2): 879–938.
- Behere S, Tornigren M (2015) A functional architecture for autonomous driving. *Proceedings of the First International Workshop on Automotive Software Architecture (WASA2015)*, pp. 3–10. New York: ACM.
- Belinić T (2018) Text processing problems with non-English languages. Medium. <https://medium.com/krakensystems-blog/text-processing-problems-with-non-english-languages-82822d0945dd>. Created: 21.05.2018. Retrieved: 30.07.2019.
- Bengio Y, Louradour J, Collobert R, Weston J (2009) Curriculum Learning. *Proceedings of the 26th International Conference on Machine Learning*. Montreal, Canada.
- Bien Z, Xu JX (2012) *Iterative Learning Control: Analysis, Design, Integration and Applications*. Springer Science & Business Media.
- Bird A (2018) Evidence and Inference. *Philosophy and Phenomenological Research*, 96:299–317.
- Bodén M (2002) A guide to recurrent neural networks and backpropagation. The Dallas project, SICS technical report.
- Bojanowski P, Joulin A (2017) Unsupervised Learning by Predicting Noise. *International Conference on Machine Learning (ICML2017)*.
- Bostrom N, Yudkowsky E (2014) The ethics of artificial intelligence. *The Cambridge Handbook of Artificial Intelligence*.
- Bradford A (2017) Deductive Reasoning vs. Inductive Reasoning. Live Science. <https://www.live-science.com/21569-deduction-vs-induction.html>. Created: 24.07.2017. Retrieved: 30.07.2019.
- Carneiro G, Chan AB, Moreno PJ, Vasconcelos N (2007) Supervised Learning of Semantic Classes for Image Annotation and Retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(3):394–410.
- Carrera L, Morales F, Tobar J, Loza D (2017) MARTI: A Robotic Chess Module with Interactive Table, for Learning Purposes. *Proceedings of the World Congress on Engineering and Computer Science (WCECS2017)*, 25–27 October, San Francisco, USA.
- Cath C (2018) Governing artificial intelligence: ethical, legal and technical opportunities and challenges. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133):20180080. The Royal Society.
- Chandrashekar G, Sahin F (2014) A survey on feature selection methods. *Computers and Electrical Engineering*, 40:16–28.
- Cheong LK, Chang V (2007) The need for data governance: A case study. *18th Australasian Conference on Information System*, 5–7 December, Toowoomba.
- Chesney B, Citron D (2019) Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. *California Law Review*, 107.
- Chidambaram M, Qi Y (2017) Style Transfer Generative Adversarial Networks: Learning to Play Chess Differently. *CoRR*.
- Constantiou I, Kallinikos J (2015) New games, new rules: big data and the changing context of strategy. *Journal of Information Technology*, 30(1):44–57.

- Crowder J, Carbone JN (2016) An Agent-Based Design for Distributed Artificial Intelligence. 2016 International Conference on Artificial Intelligence, Las Vegas, NV, USA.
- Dai AM, Le QV (2015) Semi-supervised sequence learning. *Advances in Neural Information Processing Systems* (NIPS 2015).
- DalleMule L, Davenport TH (2017) What's Your Data Strategy? *Harvard Business Review*.
- de Bruijne M (2016) Machine learning approaches in medical image analysis: from detection to diagnosis. *Medical Image Analysis*, 33:94–97.
- DefinedCrowd (n.d.) AI bias and data scientists' responsibility to ensure fairness. DefinedCrowd. <https://www.definedcrowd.com/ai-bias-and-data-scientists-responsibility-to-ensure-fairness>. Created: n.d. Retrieved: 30.07.2019.
- D'Onofrio S, Portmann E, Franzelli M, Bürki C (2018) Cognitive Computing – Theoretische Grundlagen und Praxisbeispiele der Schweizerischen Post. *Informatik-Spektrum*, 41(2):113–122. Springer Berlin Heidelberg.
- Drozdyuk A (2017) What the heck is resilience? Medium. <https://medium.com/@drozzy/what-the-heck-is-resilience-de7f758586f4>. Created: 30.06.2017. Retrieved: 30.07.2019.
- Egenhofer MJ (2015) Qualitative Spatial-Relation Reasoning for Design. Gero JS (ed.) *Studying Visual and Spatial Reasoning for Design Creativity*, pp. 153–175. Springer.
- Eykholt K, Evtimov I, Fernandes E, Li B, Rahmati A, Xiao C, Prakash A, Kohno T, Song D (2018) Robust Physical-World Attacks on Deep Learning Visual Classification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1625–1634.
- Facer K (2011) *Learning futures*. Education, technology and social change. London: Routledge.
- Fitzgerald J (2014) Bias and Variance Reduction Strategies for Improving Generalisation Performance of Genetic Programming on Binary Classification Tasks. Doctoral Thesis. University of Limerick.
- Forsberg M (2003) Why is speech recognition difficult.
- Gabbatt A (2011) IBM computer Watson wins Jeopardy clash. *The Guardian*. <https://www.theguardian.com/technology/2011/feb/17/ibm-computer-watson-wins-jeopardy>. Created: 17.02.2011. Retrieved: 30.07.2019.
- Goertzel B, Pennachin C (2007) *Artificial General Intelligence*. Springer-Verlag.
- Goodfellow I, Bengio Y, Courville A (2016) *Deep Learning*. Cambridge: The MIT Press.
- Gronlund K (2019) State of AI: Artificial Intelligence, the Military and Increasingly Autonomous Weapons. *Future of Life*. <https://futureoflife.org/2019/05/09/state-of-ai/?cn-reloaded=1>. Created: 09.05.2019. Retrieved: 30.07.2019.
- Graves A, Bellemare MG, Menick J, Munos R, Kavukcuoglu K (2017) Automated Curriculum Learning for Neural Networks. *Proceedings of the 34th International Conference on Machine Learning (PMLR)*, pp. 1311–1320.
- Gupta P, Kulkarni N (2013) An Introduction of Soft Computing Approach over Hard Computing. *International Journal of Latest Trends in Engineering and Technology (IJLTET)*, 3(1):254–258.
- Han J, Pei J, Kamber M (2011) *Data Mining: Concepts and Techniques*. Amsterdam: Elsevier.
- Haun M (2014) *Cognitive Computing*. Steigerung des systemischen Intelligenzprofils. Springer Vieweg.
- Hayes BK, Heit E (2017) *Inductive Reasoning 2.0*. Wiley Online Library.
- Hendler J (2008) Avoiding Another AI Winter. *Intelligent Systems*, IEEE, 23(2):2–4.
- Herceg M, Jones C, Kvasnica M, Morari M (2015) Enumeration-based approach to solving parametric linear complementarity problems. *Automatica*, 62:243–248.
- Hobbs JR (1985) Granularity. *Proceedings of the 9th International Joint Conference on Artificial Intelligence*, pp. 432–435. Los Angeles, California.
- Houser K (2019) These 12 principles are shaping the future of autonomous cars. *World Economic Forum*. <https://www.weforum.org/agenda/2019/07/white-paper-these-12-principles-are-shaping-the-future-of-autonomous-cars>. Created: 08.07.2019. Retrieved: 30.07.2019.
- Huang L, Joseph AD, Nelson B, Rubinstein BIP, Tygar JD (2011) Adversarial Machine Learning. *Proceedings of the 4th ACM Workshop on Security and Artificial Intelligence (AISec'11)*, pp. 43–58, 21 October, Chicago, Illinois, USA.
- Hussein A, Gaber MM, Elyan E, Jayne C (2017) Imitation learning: a survey of learning methods. *ACM computing surveys*, 50(2), article 21.

- Ignatov A, Timofte R, Chou W, Wang K, Wu M, Hartley T, Van Gool L (2018) AI Benchmark: Running Deep Neural Networks on Android Smartphones. *Proceedings of the European Conference on Computer Vision (ECCV)*.
- Introne J, Iandoli L (2014) Improving decision-making performance through argumentation: An argument-based decision support system to compute with evidence. *Decision Support Systems*, 64:79–89.
- John SN, Okonigene RE, Samuel PC, Okokpujie K (2016) Intelligent Plant Watering System for Rural Farmers. *Proceedings of the International Conference on Software Engineering Research and Practice (SERP)*, pp. 194–199.
- Johnson RA, Bhattacharyya GK (2019) *Statistics: Principles and Methods*. John Wiley & Sons.
- Johnson B, Floyd MW, Coman A, Wilson MA, Aha DW (2018) Goal Reasoning and Trusted Autonomy. Abbass HA, Scholz J, Reid DJ (eds.) *Foundations of Trusted Autonomy. Studies in Systems, Decision and Control*, 117. Springer International Publishing.
- Joshi N (2019) 7 Types of Artificial Intelligence. *Forbes*. <https://www.forbes.com/sites/cognitiveworld/2019/06/19/7-types-of-artificial-intelligence/#2d8cd595233e>. Created: 19.06.2019. Retrieved: 30.07.2019.
- Kidd DC, Castano E (2013) Reading Literary Fiction Improves Theory of Mind. *Science*, 342:377–380.
- Kolodner JL (2014) *Instructional Design: Case-Based Reasoning*. Morgan Kaufmann.
- Kolowrocki K (2014) *Reliability of Large and Complex Systems*. Amsterdam: Elsevier.
- Korenevskiy NA (2015) Application of Fuzzy Logic for Decision-Making in Medical Expert Systems. *Biomedical Engineering*, 49(1):46–49.
- Kotsiantis SB (2007) Supervised Machine Learning: A Review of Classification Techniques. *Informatica*, 31:249–268.
- Kumar C (2018) Artificial Intelligence: Definition, Types, Examples, Technologies. Medium. <https://medium.com/@chethankumargn/artificial-intelligence-definition-types-examples-technologies-962ea75c7b9b>. Created: 31.08.2018. Retrieved: 30.07.2019.
- Kurdi M (2016) *Natural Language Processing and Computational Linguistics 1: Speech, Morphology, Syntax*. Wiley.
- Kurdi M (2017) *Natural Language Processing and Computational Linguistics 2: Semantics, Discourse and Applications*. Wiley.
- Larsen K (Dr. Kim) (2018) Trust Thou AI? AI Strategy & Policy. <https://aistrategyblog.com/2018/12/03/trust-thou-ai>. Created: 03.12.2018. Retrieved: 30.07.2019.
- Lazarowska A (2017) A new deterministic approach in a decision support system for ship's trajectory planning, *Expert Systems with Applications: An International Journal*, 71(C):469–478.
- Li D, Du Y (2007) *Artificial Intelligence with Uncertainty*. CRC Press, Taylor and Francis Group.
- Liang M, Hu X (2015) Recurrent Convolutional Neural Network for Object Recognition. *Computer Vision and Pattern Recognition*, pp. 3367–3375.
- Lindgren B (2017) *Statistical Theory*. Routledge.
- Ma Z, Xie J, Li H, Sun Q, Si Z, Zhang J, Guo J (2017) The Role of Data Analysis in the Development of Intelligent Energy Networks. *IEEE Network*, 31(5):88–95.
- Malhotra P, Vig L, Shroff G, Agarwal P (2015) Long Short Term Memory Networks for Anomaly Detection in Time Series. *Proceedings in European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 22–24 April, Bruges, Belgium.
- Mazzei MJ, Noble D (2017) Big data dreams: A framework for corporate strategy. *Business Horizons*, 60(3):405–414.
- McCarthy J (2007) From here to human-level intelligence. *Artificial Intelligence*, 171(18):1174–1182.
- McCarthy J, Minsky M, Rochester N, Shannon C (1955) A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence.
- McCorduck P (2004) *Machines who think: a personal inquiry into the history and prospects of artificial intelligence*. CRC Press.
- Mishra S (2018) A Guide to Paperwork in AI/Deep Learning. Medium. https://medium.com/@srishti_56/invaluable-paperwork-in-ai-deep-learning-d91e680c9df1. Created: 12.12.2018. Retrieved: 30.07.2018.

- Mitzenmacher M, Upfal E (2017) *Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis*. Cambridge University Press.
- Mostafavi B, Barnes T (2017) Evolution of an intelligent deductive logic tutor using data-driven elements. *International Journal of Artificial Intelligence in Education*, 27(1):5–36.
- Nield D (2016) These Graphene-Based Nanobots Can Suck Pollutants From Our Oceans in Under an Hour. *Science Alert*. <https://www.sciencealert.com/graphene-based-nanobots-could-clean-up-the-metal-from-our-oceans>. Created: 18.04.2016. Retrieved: 30.07.2019.
- Nilsson NJ (2009) *The Quest for Artificial Intelligence*. Cambridge University Press.
- Osa T, Pajarinen J, Neumann G, Bagnell JA, Abbeel P (2018) An Algorithmic Perspective on Imitation Learning. *Foundations and Trends in Robotics*, 7(1–2):1–179.
- O'Shea KT, Nash R (2015) *An Introduction to Convolutional Neural Networks*. <https://arxiv.org/abs/1511.08458>.
- Pacaux MP, Debernard Godin SA, Rajaonah B, Anceaux F, Vanderhaegen F (2011) Levels of automation and human-machine cooperation: Application to human-robot interaction. *IFAC Proceedings Volumes*, 44(1):6484–6492.
- Pennachin C, Goertzel C (2007) *Contemporary Approaches to Artificial General Intelligence*. Goertzel C, Pennachin C (eds.) *Artificial General Intelligence*. Berlin Heidelberg: Springer-Verlag.
- Phillips-Wren G (2012) AI Tools in Decision Making Support Systems: a Review. *International Journal of Artificial Intelligence Tools*, 21(2).
- Poria S, Cambria E, Gelbukh A (2016) Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-based Systems*, 108:42–49.
- Procter E, Wilkinson D, Maynard J, Becker J (2019) 1D Stochastic Basin Modelling as a Frontier Exploration Tool in the Peniche Basin, Portugal. 81st EAGE Conference and Exhibition 2019.
- Reifschneider J (2019) Google's AutoML vs. a Hand-Made Model in Identifying Pneumonia. *Towards Data Science*. <https://towardsdatascience.com/googles-automl-vs-a-hand-made-model-in-identifying-pneumonia-f818902d24fc>. Created: 19.07.2019. Retrieved: 08.08.2019.
- Ruth M, Goessling-Reisemann S (2019) *Handbook on Resilience of Socio-Technical Systems*. Edward Elgar Publishing.
- Salimans T, Ho J, Chen X, Sidor S, Sutskever I (2017) Evolution Strategies as a Scalable Alternative to Reinforcement Learning. *arXiv:1703.03864v2*, pp.1–13.
- Seising R, Sanz V (2012) From Hard Science and Computing to Soft Science and Computing – An Introductory Survey. Seising R, Sanz V (eds.) *Soft Computing in Humanities and Social Sciences*, pp. 3–36. Springer, Heidelberg.
- Shachter RD, Kanal LN, Henrion M, Lemmer JF (2017) *Uncertainty in Artificial Intelligence*. Elsevier.
- Siemens G (2005) Connectivism: A learning theory for the digital age. *International Journal of Instructional Technology and Distance Learning*, 2(1):3–10.
- Silver D, Huang A, Maddison CJ, Guez A, Sifre L, van den Driessche G, Schrittwieser J, Antonoglou I, Panneershelvam V, Lanctot M, Dieleman S, Grewe D, Nham J, Kalchbrenner N, Sutskever I, Lillicrap T, Leach M, Kavukcuoglu K, Graepel T, Hassabis D (2016) Mastering the game of Go with deep neural networks and tree search. *Nature*, 529:484–489.
- Simonyan K, Zisserman A (2015) Very Deep Convolutional Networks for Large-Scale Image Recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Siruvuri LR, Salama P, Kim DS (2009) Adaptive Error Resilience for Video Streaming. *International Journal of Digital Multimedia Broadcasting*.
- Srivastava S (2018) The path to explainable AI. *CIO*. <https://www.cio.com/article/3274566/the-path-to-explainable-ai.html>. Created: 21.05.2018. Retrieved: 30.07.2019.
- Stepanova D, Gad-Elrab MH, Tinh Ho V (2018) Rule Induction and Reasoning over Knowledge Graphs. *Reasoning Web International Summer School*, pp. 142–172.
- Stucki T, D'Onofrio S, Portmann E (2018) Chatbot – Der digitale Helfer im Unternehmen: Praxisbeispiele der Schweizerischen Post. *HMD Praxis der Wirtschaftsinformatik*, 55(4):725–747.
- Stucki T, D'Onofrio S, Portmann E (2020) Chatbots gestalten mit Praxisbeispielen der Schweizerischen Post. *HMD Best Paper Award 2018. Essentials*. Springer Vieweg, Wiesbaden.
- Stutz D, Hein M, Schiele B (2019) Disentangling Adversarial Robustness and Generalization. *Computer Vision and Pattern Recognition (CVPR)*.

- Su J, Vasconcellos Vargas D, Sakurai K (2019) One Pixel Attack for Fooling Deep Neural Networks. *IEEE Transactions on Evolutionary Computation*.
- Suriyon T, Keisuke H, Choempol B (2011) Development of guide robot by using QR code recognition. The Second International Conference on Mechanical Engineering (TSME 2011), pp. 1–6.
- Tang WH, Wu QH (2011) Condition Monitoring and Assessment of Power Transformers Using Computational Intelligence. Springer-Verlag London.
- Teixeira Lopes A, de Aguiar E, Oliveira-Santos T (2015) A Facial Expression Recognition System Using Convolutional Networks. 28th Conference on Graphics, Patterns and Images (SIB-GRAPI2015).
- Trabelsi R, Masmoudi A, Masmoudi D (2016) Hand vein recognition system with circular difference and statistical directional patterns based on an artificial neural network. *Multimedia Tools & Applications*, 75(2):687–707.
- Trillas E, D'Onofrio S, Portmann E (2019) An Exploration of Creative Reasoning. Portmann E, Tabacchi M, Seising R, Habenstein A (eds.) *Designing Cognitive Cities*, pp. 85–106. Cham: Springer.
- Turing AM (1950) Computing Machinery and Intelligence. *Mind*, 49:433–460.
- Vincent J (2017) Magic AI: These are the Optical Illusions that Trick, Fool, and Flummox Computers. *The Verge*. <https://www.theverge.com/2017/4/12/15271874/ai-adversarial-images-fooling-attacks-artificial-intelligence>. Created: 12.04.2017. Retrieved: 30.07.2019.
- Voicea (n.d.) Passive A.I. vs. Active A.I. Voicea. <https://www.voicea.com/passive-a-i-vs-active-a-i>. Created: n.d. Retrieved: 30.07.2019.
- Walton D (2014) *Abductive Reasoning*. University of Alabama Press.
- Wang K, Gou C, Duan Y, Lin Y, Zheng X, Wang FY (2017) Generative Adversarial Networks: Introduction and Outlook. *IEEE/CAA Journal of Automatica Sinica*, 4(4):588–598.
- Weber B (1997) Swift and Slashing, Computer Topples Kasparov. *The New York Times*. <https://www.nytimes.com/1997/05/12/nyregion/swift-and-slashing-computer-topples-kasparov.html>. Created: 12.05.1997. Retrieved: 30.07.2019.
- Weiss K, Khoshgoftaar TM, Wang DD (2016) A survey of transfer learning. *Journal of Big Data*, 3:article 9.
- Weizenbaum J (1966) Eliza – a computer program for the study of natural language communication between man and machine. *Communication of the Association for Computing Machinery (ACM)*, 9(1):36–45.
- Wiering M, van Otterlo M (2012) *Reinforcement Learning: State-of-the-Art*. Springer Science & Business Media.
- Wilkinson D (n.d.) What's the difference between data and evidence? Evidence-based practice. *The Oxford Review*. <https://www.oxford-review.com/data-v-evidence>. Created: n.d. Retrieved: 30.07.2019.
- Williams DW, Wood MS (2015) Rule-based Reasoning for Understanding Opportunity Evaluation. *Academy of Management Perspectives*, 29(2):218–236.
- Witten IH, Frank E, Hall MA, Pal CJ (2016) *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann.
- Wu X, Chen H, Wu G, Liu J, Zheng Q, He X, Zhou A, Zhao ZQ, Wei B, Gao M, Li Y, Zhang Q, Zhang S, Lu R, Zheng N (2015) Knowledge Engineering with Big Data. *IEEE Intelligent Systems*, 30(5):46–55.
- Xue V, Zhang M, Browne WN, Yao X (2016) A Survey on Evolutionary Computation Approaches to Feature Selection. *IEEE Transactions on Evolutionary Computation*, 20(4):606–626.
- Yager RR, Zadeh LA (2012) *An Introduction to Fuzzy Logic Applications in Intelligent Systems*. Springer Science & Business Media.
- Zadeh LA (1988) Fuzzy Logic. *Computer*, 21(4):83–93.
- Zambonelli F, Omicini A, Anzengruber B, Castelli G, DeAngelis FL, Di Marzo Serugendo G, Dobson S, Fernandez-Marquez JL, Ferscha A, Mamei M, Mariani S, Molesini A, Montagne S, Nieminen J, Pianini D, Risoldi M, Rosi A, Stevenson G, Viroli M, Ye J (2015) Developing Pervasive Multi-Agent Systems with Nature-Inspired Coordination. *Pervasive Mobile Computing Journal*, 17:236–252.



Artificial Intelligence – The Mindfire Foundation and Other Initiatives

Moreno Colombo, Edy Portmann and Pascal Kaufmann

Abstract

Artificial intelligence being one of the buzzwords of the decade, with an increasingly growing number of new initiatives and investments, the scope of this article is that of investigating the current situation of research in this field. In particular the reasons why experts believe that research in artificial intelligence is currently stuck and other of its problems, as well as some possible solutions are analyzed. Moreover, a framework describing the fundamental building blocks of AI initiatives, based on an analysis of already existing solutions, is defined, and the innovative structure and ideas of one of those initiatives, the Mindfire Foundation, are presented in detail. The Mindfire Foundation is a non-profit organization with the goal of understanding and replicating the human mind, with a focus on application solving problems affecting humanity. To allow this, its fulcrum is an innovative blockchain-based system providing incentivization of transdisciplinary and antidisciplinary collaborations, combined with a solid framework for the handling of ethical and regulatory problems.

Vollständig neuer Beitrag.

M. Colombo (✉) · E. Portmann
Human-IST Institute, University of Fribourg, Fribourg, Switzerland
E-Mail: moreno.colombo@unifr.ch; edy.portmann@unifr.ch

P. Kaufmann
Mindfire Foundation, Zug, Switzerland

1 Introduction

Cognitive computing can be defined as the task of simulating human thought processes in a computerized model to better solve complex problems. This definition describes a set of methods that can be used for solving problems in an intelligent way, but, even if most likely very valid, these methods are not the only ones that can be considered for this kind of task. In this article, the more general term *artificial intelligence* (also *AI*) – the ability of solving complex methods in an intelligent way – is thus used as a superset including between others the cognitive computing approach, to create a more general discourse on past, present and future of intelligent machines.

The term artificial intelligence was first used by John McCarthy in 1955, when he wrote with Marvin Minsky, Nathaniel Rochester and Claude Shannon “*A proposal for the Dartmouth Summer Research Project on artificial intelligence*” (McCarthy et al. 1955), the first text where the term *intelligence* was related to machines, and not only to the humans programming or using them. In the same text, the first definition of the artificial intelligence problem was also given:

The artificial intelligence problem is taken to be that of making a machine behave in ways that would be called intelligent if a human were so behaving. (McCarthy et al. 1955, p. 11)

So, artificial intelligence was strongly related to human intelligence and its imitation inside a machine since its very first appearance. At first, by reading McCarthy’s definition, one could think that a machine that is able to play chess and win a game would be sufficient for solving the stated problem of artificial intelligence. Indeed, that sort of machine would be behaving in a way that many humans would judge as intelligent. But in the same text, McCarthy went much further with its ambition: He wrote about the possibility for computers to simulate *many of the higher functions of the brain*. In other words, an artificial intelligence should be able to simulate (almost) completely humans’ cognitive capabilities!

From that first appearance of artificial intelligence more than sixty years have passed, and yet the common vision of what it should be has not changed much. Indeed, most science fiction movies’ artificially intelligent characters have human-like skills and behaviours, and can sometimes even be mistaken for real humans, similarly to what has been envisioned by McCarthy. But if this is still in science fiction movies, it means that it is not yet part of reality – which is interesting, since Alan Turing predicted in 1950 the creation of machines who could think like humans by the year 2000 (Turing 1950)! However, mass medias seem to tell a different story. Indeed, the fact that artificial intelligence is one of this decade’s most popular buzzwords, and the continuous discussions about its dangers and ethical problems, seem to trick people into believing that artificial intelligence is almost ready to substitute humans in any field – or even to take over the world, in what is commonly called the *Terminator scenario*. But why is there this misconception?

It is estimated that in 2016 companies have invested between \$26 billion and \$39 billion in artificial intelligence (Bughin et al. 2017), and one can imagine that with

such enormous investments, a solution to a problem – no matter how complex – can be found. So, one can clearly see from where the positivity on the successful solution of the artificial intelligence problem comes from. However, what is often not clear, is that most of these investments, especially the ones coming from private companies, focus on ways to solve specific problems, and the less risky and more direct way to do this is not investing in a general-purpose, self-learning and human-like artificial intelligence (also *artificial general intelligence*, or AGI), but investing in *weak* artificial intelligence, machine learning solutions expressly tailored to solve only those specific problems.

This way of thinking and investing is a big problem for the research in human-like artificial intelligence, developing a custom solution for each single problem is indeed probably not the most efficient approach. From a research perspective, it would be much more effective to focus more on creating a single artificial intelligence able to adapt to different situations and solve various kind of tasks, instead of focusing only on very specific tasks. One can take as an example DeepMind's AlphaGo,¹ the artificial intelligence which was able to beat the world champion Lee Sedol in a match of the extremely complex game of Go. AlphaGo is being described by medias as one of the most significant advances in artificial intelligence of the last years (Weidman 2018), but not only it does not satisfy the definition of artificial intelligence of McCarthy in most cases – it is only suited for the task of playing Go – it also cannot adapt to minor changes in the rules of the same game (e.g., changing the grid size from 19×19 to 18×18 (Urech 2017)), whereas humans have no problem in adapting to this kind of situations. AlphaGo is an evident example of the limits of this strongly problem-related approach for research in artificial intelligence: It shows the complete lack of plasticity and adaptability of the most advanced artificially intelligent machines of nowadays.

Moreover, Hans Moravec discovered that, contrarily to what people traditionally believe, reproducing high-level reasoning requires relatively little computation, whereas low-level sensorimotor functions require enormous amounts of computational resources. Or, as he expressed the paradox, “*it is comparatively easy to make computers exhibit adult level performance on intelligence tests or playing checkers, and difficult or impossible to give them the skills of a one-year-old when it comes to perception and mobility*” (Moravec 1988, p. 15). This enforces the theory that the currently most popular, computation-based, approaches to solve the problem of human-like artificial intelligence are not the best possible ones, since they tend to give complex solutions to problems that for humans are simple, and simpler solutions to problems that humans consider complex.

From these examples, one can see that research is still far from creating human-level artificial intelligence and can thus understand why many AI pioneers agree that research in this field is currently stuck and in a dead end. For example, Michael I. Jordan, the most influential computer scientist according to Science (Bohannon 2016), stated in an interview that “*we should stop using brain metaphors when we talk about computing*” (Gomes 2014). After him also Andrew Ng, co-founder of

¹<https://deepmind.com/research/alphago>

Google Brain, founder of Coursera and former Chief Scientist of Baidu AI group, wrote: *“Despite the media hype, we’re nowhere near being able to build human-level intelligence. Because we fundamentally don’t know how the brain works, attempts to blindly replicate what little we know in a computer also has not resulted in particularly useful AI systems.”* (Ng 2016), clearly expressing his critical view of the current AI research situation. Another example of this sentiment of AI research being stuck in a dead end came from Geoffrey Hinton, one of the first researchers who demonstrated the use of generalized backpropagation algorithm for training multi-layer neural networks, now professor emeritus at the University of Toronto and researcher at Google Brain, who said: *“My view is throw it all away and start again”*, speaking of the current state of artificial intelligence (LeVine 2017).

In the next section, the main current limitations of artificial intelligence with respect to human intelligence and the main discoveries that have been done in the past years are analyzed. After that, related works on transdisciplinary and antidisciplinary research, the approaches to research the authors believe are the most important to be considered to solve extremely complex problems, including that of human-like artificial intelligence, are presented. These aspects are then joined with an analysis of several recently launched initiatives to create a framework describing the fundamental building blocks of complete AI initiatives. This framework is then used to describe the innovative structure and ideas of the Mindfire Foundation. An outlook with some concluding remarks and future work proposals concludes this article.

2 Related Work

AI research being stuck, to unlock the situation one would better identify the main limitations of what today is considered artificial intelligence with respect to human intelligence. Two of the problems that have been identified as the principal ones are presented in the following paragraphs. Furthermore, transdisciplinary and antidisciplinary research is introduced.

2.1 Current Limitations of Artificial Intelligence

As already discussed, most of funding for research in artificial intelligence nowadays is not aiming at solving the problem of human-level artificial intelligence, but it is focusing on the solution of very specific problems with machine learning techniques. These techniques include for example support vector machines (Cortes and Vapnik 1995), artificial neural networks (Haykin 1994) and k-nearest neighbor classification (Peterson 2009). The only common characteristic of all these methods is the fact that they require large amounts of data and computing power to be able to accurately solve the problems they are designed to solve (Mitchell 1997; Taniguchi et al. 2018). So, if for solving each of a virtually infinite set of specific problems one not only needs to develop an ad-hoc solution, but also needs huge datasets, one can imagine that creating an artificial general intelligence with currently popular

methods would require an infinite amount of data and an enormous quantity of computing power.

On the other side, the human brain is able to learn with minimal amounts of data (Lake et al. 2015, 2017). For example, infants who see for their first time a cat are able to obtain enough information from the seen object to know that it is an animal and can tell it from other animals. All of this without the need of negative samples (Tenenbaum 1999) and with a brain consuming only about 20 watts of power (Ling 2001)!

Some research to understand and replicate the small data learning processes of humans have been carried out with good results (Lake et al. 2015, 2017; Taniguchi et al. 2018; Tenenbaum 1999), but those methods are unfortunately still limited in accuracy compared to the ones using large datasets. This makes them not yet mainstream in the AI research landscape, a field of study where task accuracy is the main indicator of the quality of a system.

The second big problem of today's situation in research comes from the fact that artificial intelligence as a research field was born in a computer science setting, and therefore was, and still is, strongly related to computation (Sieber 2018). Modern computers have incredibly high computing power, which can make them superior to humans in solving various kind of tasks. Thinking and computation, however, are not the same thing. The human brain is indeed able to do much more than just performing arithmetic operations. Not only it can understand the meaning of complex language artefacts, predict the impact of actions on the future, recognize familiar objects in visual scenes, and show complex behaviors as attention and emotion, important for the notion of cognition itself (Picard 1997), but it does these complex things very easily, compared to its ability to perform arithmetic computations. It is thus believed that to emulate all human functions, simple computation is not enough, there is the need for a new language, or a data structure to express and replicate these concepts: *the brain code* (Malsburg 2018). On the way of human-level artificial intelligence, there is indeed first the need to understand what non-artificial intelligence is and how it works, which corresponds exactly to the task of decoding the brain code.

An important factor for being able to recreate human intelligence in machines could be the use of a new paradigm in software architecture, where instead of using precise yet brittle temporal protocols for all operations, one could use more approximative and fail-safe surface-based communication between different modules, which is a paradigm much more similar to what happens in nature, where, for example, eyes detect many different light sources at the exact same instant and are able to cope with incomplete data (e.g., the blind spot created by the optic nerve). This concept takes the name of *phenotropic computing*, meaning *the interaction of surfaces* (Lanier 2003).

Other than the illustrated limitations of today's conception of artificial intelligence, one must note that "*virtually all human achievements have been made by groups of people, not lone individuals*" (Malone 2018, p. 34). AI research is no exception, like all other researches it has to be conducted in groups to get good results. The problem being very complex and stuck in a dead end, one can see that the

approaches that have been taken until now are not as effective as they are in other research fields. One solution could be to use technology (e.g., blockchain, see Sect. 4) to coordinate bigger and bigger groups of researchers in a way that could never be done before, allowing the creation of more powerful *superminds* – superorganisms composed of people and machines collaborating to reach results impossible for single researchers (Malone 2018). But increasing the size and improving coordination of the group working on the artificial intelligence problem alone is probably not a sufficient solution, a new way of handling it could also be necessary. In fact, what can be seen today all around the world are groups of computer scientists, neuroscientists, psychologists, engineers and many others trying to solve the artificial intelligence problem only from their discipline’s perspective, without considering findings and techniques from other disciplines. One important aspect to consider when researching artificial intelligence can therefore be taking a transdisciplinary and/or antidisciplinary research approach.

2.2 Transdisciplinary Research

Transdisciplinary research is when “members of the different cultures interact to co-produce knowledge” (Pohl 2008, p. 47), where different cultures can be understood as people working in different settings (e.g., governments, private companies, different sectors of academia, different countries).

The key components of a transdisciplinary research process are now presented, according to the model of Lang et al. (2012), which has many similarities with other conceptual models (Carew and Wickson 2010; Jahn 2008; Stauffacher et al. 2012; Stokols et al. 2010; Talwar et al. 2011). This model, illustrated in Fig. 1, starts with societally relevant problems, which generate research questions, and then relies on mutual learning processes between different branches of science and society. In this way, the real-world problem can be solved and at the same time new interdisciplinary approaches and methods can be developed.

The transdisciplinary research process in the model of Lang et al. (2012) is divided in three phases: Collaborative framing of the problem and creation of a collaborative and transdisciplinary research team, co-production of solution-oriented and transferable knowledge (i.e., goal-oriented collaboration) and finally application of knowledge to both solve the societal problem and improve scientific practice.

Transdisciplinary research has a big potential, but often the connections between the *real world*, academia, and even different academic disciplines do not exist. An example of a successful creation of the kind of network necessary for transdisciplinary research resides in the Human-IST Institute at the University of Fribourg (Switzerland).² This interfaculty institute provides a collaborative setup and technological infrastructures, and serves as a point of contact between companies, government and researchers from several academic disciplines, all working together in a transdisciplinary way on solutions to real-world problems.

²<http://human-ist.unifr.ch>

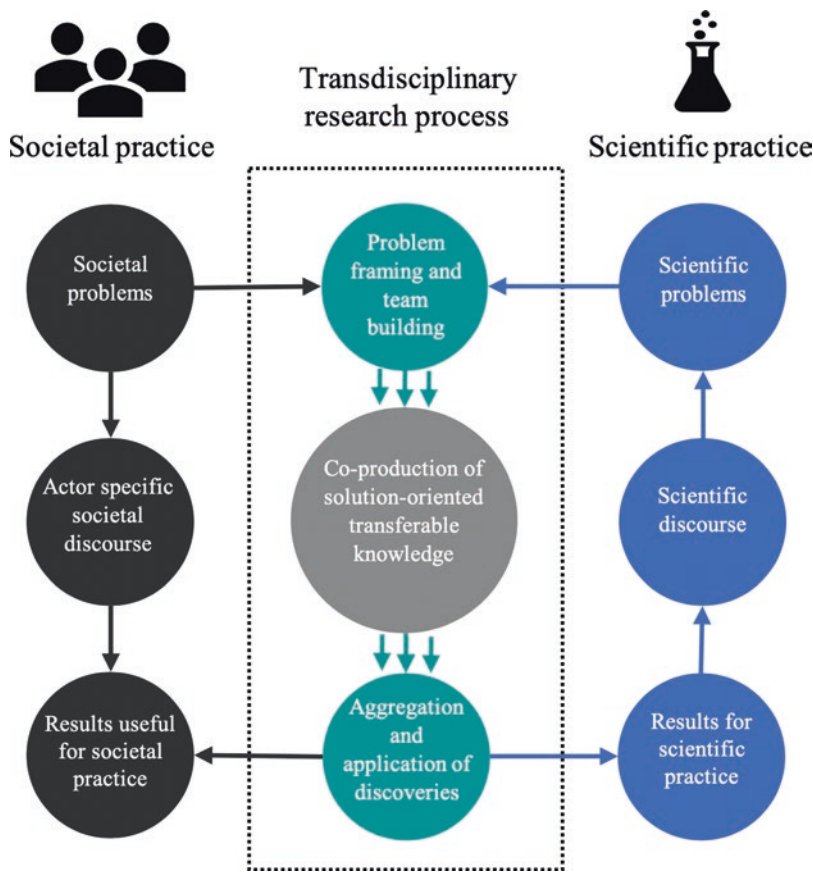


Fig. 1 Conceptual model of a typical transdisciplinary research process (adapted from Lang et al. 2012)

2.3 Antidisciplinary Research

A slightly different approach to solve important problems for the society, instead of conducting research strictly inside a traditional academic discipline, with its technical words, frameworks and methods, consists in what is called *antidisciplinary research* (Ito 2012). This approach is not yet particularly popular, despite its power to address real world problems – usually not completely related to academic disciplines – because of the difficulties of finding research funds outside traditional disciplines. Traditional academia, in fact, cares unfortunately more about citations and publications than about improving human life, and most journals are strictly related to academic disciplines, factor making the publication of antidisciplinary works – not fitting in traditional academic disciplines by definition – extremely hard. Luckily, things are slowly changing also in this direction. The MIT Media Lab³ is

³<https://www.media.mit.edu>

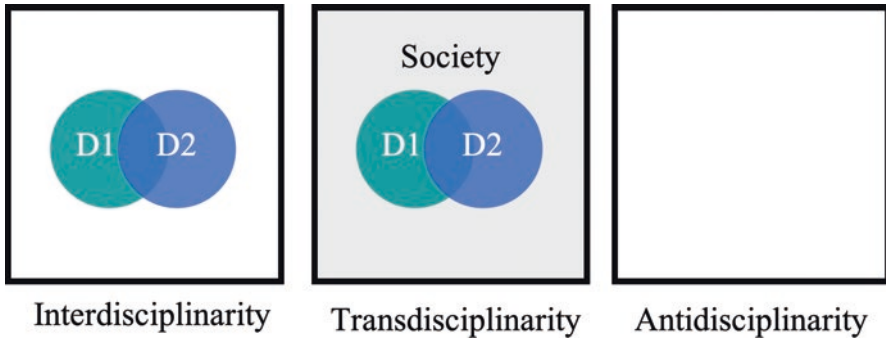


Fig. 2 Differences between interdisciplinary, transdisciplinary and antidisciplinary research

one of the first labs in academic research who gives to all researchers who do not fit into traditional academic disciplines an occasion to work on their own antidisciplinary projects. The MIT Media Lab thus gives the possibility of working on a project not because it can end up in a publication, but because it is challenging and useful (Ito 2016). This is interesting, since antidisciplinary research makes it potentially easier – especially for real world problems – to achieve the research goal, by removing restrictions on the frameworks and methodologies of traditional academic disciplines. In other words, antidisciplinary research consists in seeing things as they are, without labels (Ramchandani 2017). Therefore, antidisciplinary research in artificial intelligence could be a good and innovative approach to find new solutions to the problem.

The following fundamental differences between interdisciplinarity, transdisciplinarity and antidisciplinarity are illustrated in Fig. 2: Interdisciplinarity implies collaborations between different disciplines, transdisciplinarity is a combination of interdisciplinarity with participatory approach, and antidisciplinary means working in the spaces that do not fit inside academic disciplines.

In this section, research allowing the contextualization of recent AI initiatives with respect to the current problems of this research field has been presented. Additionally, transdisciplinary and antidisciplinary research practices have been defined, and will be included in the analysis of the following AI initiatives, particularly in that of the Mindfire Foundation.

3 Framework: The Building Blocks of an AI Initiative

The potential of artificial intelligence for improving human lives and solve some of the biggest problems of humanity being very high and globally recognized, many initiatives and strategies centered on the development of artificial intelligence have been developed over the last years. The most important of said strategies have been laid out by nations all around the world, with the intent of becoming world leaders

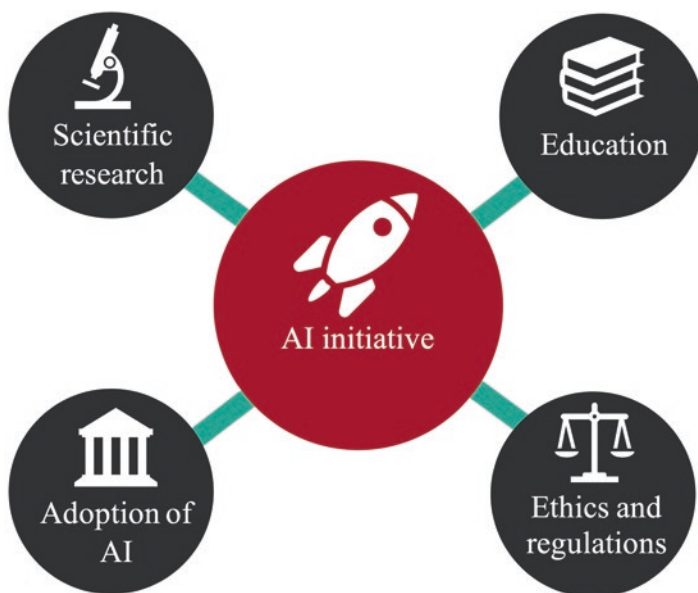


Fig. 3 The four main building blocks of an AI initiative

in artificial intelligence and improve the quality of life of their population. After an accurate analysis of several of those initiatives (Dutton 2018), one can see that they all have some points in common, which could be used to create a framework for AI initiatives and their building blocks. What have been identified as the four main building blocks of such an initiative are now explained (see Fig. 3).

3.1 Scientific Research

A first important factor for developing new technologies is scientific research, which allows the creation of both the theoretical and practical basis necessary for solving concrete problems. What is generally essential for excellent quality scientific research are competent researchers, infrastructures and material. All of these points being potentially expensive, one can clearly see that scientific research requires important funding. Because of this, most of the existing initiatives for artificial intelligence are mainly centered around the allocation of a budget to ensure an increasing economic support of research in artificial intelligence (e.g., the Government of Canada is investing C\$ 125 million in research over 5 years (Stark 2018)).

Investments for scientific research often also consist in the creation of technological infrastructures acting as a foundation for collaborating centers of competence in different locations, which ensure coverage of different regional contexts, universal accessibility to good research facilities, and provide a large number of infrastructures facilitating collaborations and experiments. An interesting example of

this aspect of research is well illustrated by Ford (2018), where the institution of a network of high competence research labs around all Europe is proposed, with many connected, collaborating laboratories researching artificial intelligence and sharing data connected to one another as a big, distributed research lab.

Since every institution has different priorities and problems to be solved, it is sometimes important to channel research in one main direction, by providing researchers with either goals to be achieved, either restrictions on their researches. For example, the common goal of all laboratories of CLAIRE (Ford 2018) is the research of human-centered artificial intelligence (i.e., the augmentation of human intelligence, instead of its replacement).

3.2 Education

Another important aspect to be considered when building an initiative for artificial intelligence is the support of education. This aspect is important in mainly three fields of application: First, artificial intelligence related lectures introduced in all school levels and considering different perspectives and approaches to the problem can increase the competency and interest of young people towards this discipline, which is fundamental for the constant provision and generational change of researchers in artificial intelligence. Second, learning some applications of artificial intelligence basics is potentially useful also for people working in either public or private sectors, in such a way that they can estimate what are the possible benefits of artificial intelligence applied to their business. Third, an education of the population on benefits and dangers of artificial intelligence is necessary to make it better accepted by society in people's daily life and making it less prone to be misused, reducing the dangers caused by its advent. This type of education is fundamental for helping people better adapt to the societal changes driven by the development of artificial intelligence, which is without any doubt one of the most important points to consider when developing artificial intelligence with the goal of improving people's life.

An example of education plan for the population is proposed in "*A Next Generation Artificial Intelligence Development Plan*" (Karch et al. 2018), where the Chinese State Council proposes the introduction of artificial intelligence related courses in primary and secondary schools.

3.3 Public and Private Sector Adoption of Artificial Intelligence

The third factor to be considered when building an initiative for artificial intelligence, is the incentivization of AI techniques at the service of the community. This means, the introduction of artificial intelligence in both public and private sector to simplify and increase the quality of people's work and life. Each strategy should ideally either work directly to particularly improve one or more aspects of life, either incentivize the use of newly developed AI techniques in public and private sec-

tors. For example, France’s AI strategy (Dillet 2018) aims at developing an open data policy to facilitate the introduction of solutions making use of artificial intelligence in sectors where those are mature enough to be used on concrete cases without risks.

3.4 Ethics and Regulations Frameworks

The last important thing to consider in the foundation of initiatives for artificial intelligence, is a way to ensure that the developed technology and its applications are not going to have a negative impact on society. To do this, one has to ensure the creation of internal, national or international regulations, laws and ethical frameworks. Things like the legal and societal status of intelligent machines, the regulation of interactions with organic forms of life, and the ownership of intelligence principles can, for example, be considered in this context. Most of the initiatives analyzed in the work of Dutton (2018) take in account, in slightly different ways, the problem of ethics and regulations.

3.5 Overview of AI Initiatives

An overview of the strategies some initiatives adopt to solve the artificial intelligence problem according to the defined framework can be found in Table 1.

Table 1 Overview of AI initiatives according to the defined framework

Initiative	Scientific research	Education	Adoption of AI	Ethics and regulations
<i>Pan-Canadian Artificial Intelligence Strategy</i> (Stark 2018)	Three inter-connected AI labs Investments in AI research	Funding for lecturers and students Public information	Study impact of AI on economy, government and society	Creation of an ethical and legal AI framework
<i>Next Generation Artificial Intelligence Development Plan</i> (Karch et al. 2018)	Goals for R&D set by the State Establishment of a central AI lab	Introduction of AI courses in primary and secondary school	Tax incentives to high-tech enterprises	Creation of an ethical and legal AI framework
<i>France’s Artificial Intelligence Strategy</i> (Dillet 2018)	Establishment of a network of 4–5 main labs across the country	Creation of a national AI education program	Open data policy to facilitate adoption of AI in some sectors	Creation of an ethical and legal AI framework
<i>CLAIRE</i> (Ford 2018)	Network of labs Focus on human-centered AI	Funding of student projects	Provision of a contact point between industry and research	Creation of an ethical and legal AI framework

4 Use Case: The Mindfire Foundation

A promising initiative for artificial intelligence, which is following the same framework as many others, but in a different, innovative and more complete way, specifically focusing on the problem of artificial general intelligence, is the *Mindfire Foundation*, a Swiss non-profit organization (Sieber 2018). The mainly recurrent analogy used by the Mindfire team, is between *artificial intelligence* and *artificial flight*. Indeed, in the early days of artificial flight, humans who wanted to fly took inspiration from nature, and managed to solve the problem only once they understood the underlying principles of flight, while all attempts to blindly copy natural flight failed miserably. Similarly, the main idea of Mindfire for solving the problem of human-like artificial intelligence is based on the decoding of the brain code – the understanding of the principles reigning intelligence per se (Malsburg 2018) – instead of blindly copying of the human brain.

The structure of Mindfire with respect to the previously defined AI initiatives' building blocks (i.e., scientific research, education, facilitation of AI adoption, ethics and regulations) is now presented.

4.1 Scientific Research

The main building block of Mindfire consists in scientific innovation and several innovative methods are used to get the most out of research. Mindfire being composed by as many competent contributors as possible, it is important to periodically gather these for updates, brainstorming, organizational tasks and ideas exchanges. This is done once to twice a year with the *Mindfire Missions*, a couple of weeks where experts from all over the world gather together in a fixed location to discuss what has been done, what has still to be done, and work together to advance of one or more of the little steps finally leading to decoding the brain code.

At the end of the mission, milestones and goals for the following mission are created and updated, in such a way that the work to be done between missions is clear. The base idea is that, in the time between the missions, proposition and outcomes of the previous are either elaborated, either implemented in prototypes by groups of experts and the results are presented in the following mission. Moreover anybody – participating to the missions or not – can discuss at any time their opinions and ideas on human-like artificial intelligence on the *Mindfire Network*,⁴ an online platform powered by *Starmind*,⁵ a system powered by an artificial intelligence allowing to route questions and observations to the best suited experts within the network. Mindfire Network is also an interesting point of contact between participants to the missions and world-leading experts from several disciplines.

⁴<https://network.mindfire.global>

⁵<http://www.starmind.com>

Fundamental aspects of work in all of contexts within Mindfire (i.e., missions and between-mission work) are transdisciplinarity and antidisciplinarity. Indeed, the innovative way of handling the artificial intelligence problem of Mindfire, consists in having the widest and most diverse group of people, the Mindfire talents, which are world-class experts in different fields related to artificial intelligence, all working together partly within their discipline's context and partly out of any discipline, in what is considered to be an antidisciplinary way. The advantage of this is being able to tackle any kind of problem in the best possible way, and not only with the best way that can be achieved with methods from a specific discipline. The selection of the best talents considers thus not only their level of expertise in their field of research, but especially their multidisciplinary and their ability to adapt to seeing and explaining ideas from different perspectives, preventing them from being too much close-minded and stuck on traditional approaches for problem solving.

An *AI Lab* is also proposed for both providing an infrastructure supporting collaborations between missions, as well as a contact point between Mindfire, other research institutions, society and business, facilitating transdisciplinary research.

To support and incentivize collaborations as well as trans- and antidisciplinarity, Mindfire uses a blockchain-based token, the *Mindfire Token* (see Fig. 4), in an attempt to ensure the correct recognition and reward to people contributing to new ideas and discoveries, since the lack of individual recognition is one of the main problems of collaborations nowadays (Zucker 2012).

The Mindfire tokens are allocated to investors, administrative staff and talents, based on their contributions to Mindfire's missions, and allows the creation of a decentralized autonomous organization, where token holders have control on the actions, revenue and distribution of intellectual property of the foundation. Revenue of the Mindfire Foundation is distributed between token holders, as a reward for their contributions. For what concerns intellectual property management, the intel-

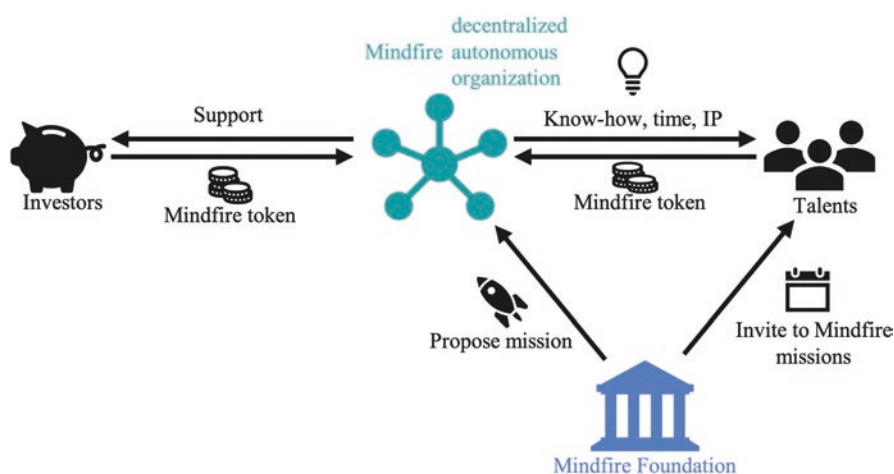


Fig. 4 The Mindfire token distribution (Sieber 2018)

lectual property on all ideas and discoveries produced during Mindfire Missions is held by the Mindfire Foundation. However, “*each participant’s contributions and discoveries are stored on a blockchain allowing fair and direct recognition for new ideas and concepts developed during Mindfire Missions*” (Sieber 2018, p. 9).

4.2 Education

For what concerns education, the Mindfire Foundation proposes three main settings, with different goals: education of people working in close contact with artificial intelligence, provision of expertise for talents, and public education. Some more details about these three initiatives are presented in the following lines.

Recently, the Mindfire Foundation and *SingularityNET* – a full-stack AI solution powered by a decentralized protocol that lets anyone create, share, and monetize AI services at scale – announced a strategic partnership focusing, among others, on education in artificial intelligence, by implementing practical courses and lectures in artificial intelligence and its business applications (Ramesh 2018). Through the dedicated *AI University*, offering courses and lectures to discuss impact of artificial intelligence on society and the latest developments in the field, Mindfire plans to ensure an active dialogue with society about artificial intelligence. This also has the goals of providing access to the latest developments of the *Mindfire Talent Network*, as well as recruiting and training new talents (Sieber 2018).

Moreover, the structure of the Mindfire Foundation is organized in such a way that allows the continuous education of talents through the advisory board and external experts transmitting their knowledge in different fields via presentations and discussions with talents. The advisory board itself is composed by experts in various disciplines, but one of the most important things to consider for Mindfire, is having at least one expert in trans- and antidisiplinary research, who can lead and organize at best the work of talents within Mindfire.

Last but not least, Mindfire provides to the general public information and education programs, mostly through popular media outlets. This factor aims at reducing the common misconceptions and fears about artificial intelligence, helping the world learn how to properly understand the impact it will have on everyday life, as well as how to deal with its growing presence in society. Moreover, a good presence in the medias increases the chances of reaching as many potential talents as possible, increasing the number and quality of talent candidatures.

4.3 Public and Private Sector Adoption of Artificial Intelligence

To jump-start the early adoption in industry, government or society of new developed techniques, discoveries and ideas, the Mindfire Foundation engages in developing prototypes and concrete applications using the outcomes from missions and other works. This usually happens between missions, and has many advantages compared to producing strictly theoretical work.

First, concrete applications of new concepts are interesting from an economic and intellectual property protection point of view, these have indeed the possibility of being patented, providing thus an income to Mindfire, useful to support its various activities and ensure its survival and growth throughout the years. Second, prototypes are an easy way to showcase the produced work to the world, making Mindfire’s discoveries both more attractive and more understandable to the general public and potential investors. Third, concrete patented applications have the potential of showing to industry and government leaders some of the possible applications of a certain newly developed method or discovery. Patents can also serve as a guideline on the procedures to follow when putting a certain idea in practice, making it easier for companies and institutions to adopt newly developed methods of artificial intelligence on real world use cases.

4.4 Ethics and Regulations Frameworks

Mindfire takes a different approach to ethics and regulations, compared to the presented national initiatives for artificial intelligence: the idea of creating an ethical framework and produce regulations and laws related to artificial intelligence is indeed not present. Mindfire proposes a distributed approach to solving the problem of making sure artificial intelligence is not going to become evil.

The power of supporting and incentivizing collaborations of the Mindfire token illustrated before, is not the only use of this token. Mindfire token holders have indeed control on Mindfire actions through a voting system (see Fig. 5). Every person or organization holding Mindfire tokens can vote to choose if to accept or reject proposals – usually intellectual property actions, for example, selling of intellectual property or cession of exclusive usage rights – and given the distribution of tokens

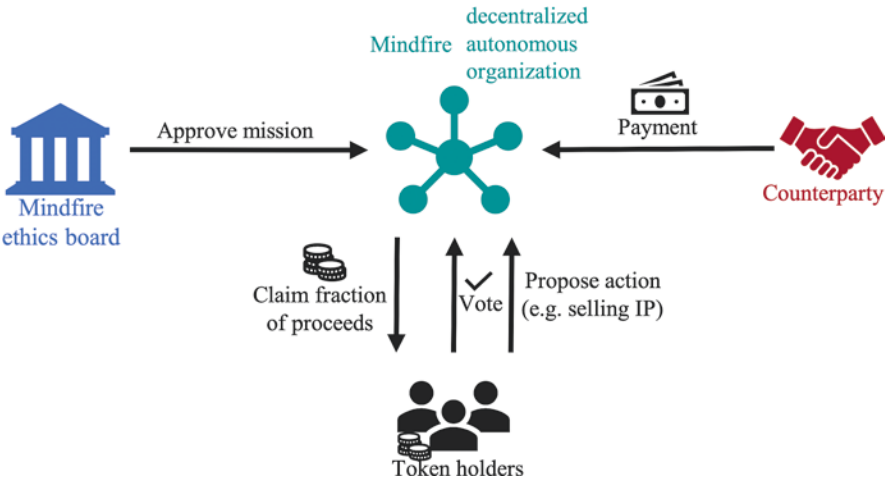


Fig. 5 Basic process of proposal voting process of Mindfire (Sieber 2018)

between people and organizations from different countries and cultures, this ensures that all actions done by Mindfire are ethically correct. To further increase the effectiveness of this strategy, before voting all propositions are analyzed by Mindfire’s ethics board, who can straight-up reject proposal which are not complying with the Mindfire Foundations’ by-laws (Sieber 2018).

4.5 Overview of the Building Blocks of Mindfire

A summary of the building blocks according to the defined framework can be found in Table 2.

5 Conclusion

In conclusion, in this text, the importance of artificial intelligence for society and its actual limits have been illustrated. This was followed by the explanation of the need of shifting the focus of research from narrowly focused AI solutions to a more general artificial intelligence, with the ability to learn and adapt to solving several different problems.

This idea being strictly related by the concept of adaptability to phenotropics (Lanier 2003) – the conception of systems able to adapt to interacting or connected systems without the need of protocols allowing mutual understanding – the two research fields should inspire one another, as well as being inspired by others, to accelerate their progresses.

Additionally, the building blocks of initiatives for artificial intelligence have been analyzed, as well as the details of a promising initiative, the Mindfire Foundation. What is innovative about this solution are not only the methods used to handle the problem – trans- and antidisciplinary research, the Mindfire token, decentralized ethics control, etc. – but especially the final goal. Understanding the principles of intelligence and how simple molecules, organized in the right way can create something as complex as intelligence, can indeed not only be used to create a human-like artificial intelligence, but also to better understand humans and solve organizational problems. This could, for example, have an impact on our understanding of

Table 2 Summary of the building blocks of Mindfire

Scientific research	Education	Adoption of AI	Ethics and regulations
Mindfire missions	AI University	Creation of prototypes	Ethics board
Mindfire network	Education of talents during missions	Selling of intellectual property	Mindfire token
AI lab	Public information	AI lab: Contact point between research and industry or government	holders voting system
Mindfire token			
Anti- and transdisciplinary research			

mental illnesses, or even be transposed to society, using the brain's organization as a model to improve the structure of companies, cities, road networks, and even communication networks.

Additionally, one important thing to consider in future works to improve human-like artificial intelligence is natural human-computer interaction. Indeed, making machines able to interact with humans without the need for protocols by interpreting their words, gestures, body language and emotions to understand the interacting human's needs will be fundamental for the development of intelligent machines. A novel proposed way for reaching this goal is through the adaptation of the concept of phenotropics (Lanier 2003), the interaction through surfaces and pattern recognition, that was presented in Sect. 2.1, to human-machine interaction. If successful, this could lead to natural communication between machines and humans in a human-human fashion.

References

- Bohannon J (2016) Who's the Michael Jordan of computer science? New tool ranks researchers' influence. *Science | AAAS*. <http://www.sciencemag.org/news/2016/04/who-s-michael-jordan-computer-science-new-tool-ranks-researchers-influence>. Created: 19.04.2016. Retrieved: 13.08.2018.
- Bughin J, Hazan E, Ramaswamy S, Chui M, Allas T, Dahlström P, Henke N, Trench M (2017) How artificial intelligence can deliver real value to companies | McKinsey & Company. <https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/how-artificial-intelligence-can-deliver-real-value-to-companies>. Created: 06.2017. Retrieved: 19.06.2018.
- Carew AL, Wickson F (2010) The TD Wheel: A heuristic to shape, support and evaluate transdisciplinary research. *Futures*, 42(10):1146–1155.
- Cortes C, Vapnik V (1995) Support-vector networks. *Machine learning*, 20(3):273–297.
- Dillet R (2018) France wants to become an artificial intelligence hub. *TechCrunch*. <http://social.techcrunch.com/2018/03/29/france-wants-to-become-an-artificial-intelligence-hub>. Created: 29.03.2018. Retrieved: 13.08.2018.
- Dutton T (2018) An Overview of National AI Strategies. *Politics + AI*. <https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd>. Created: 28.06.2018. Retrieved: 02.08.2018.
- Ford A (2018) CLAIRE – a new European confederation for AI research | Science, Technology & the Future. *Science Technology Future*. <http://www.scifuture.org/claire-a-new-european-confederation-for-ai-research>. Created: 26.06.2018. Retrieved: 13.08.2018.
- Gomes L (2014) Machine-Learning Maestro Michael Jordan on the Delusions of Big Data and Other Huge Engineering Efforts. *IEEE Spectrum: Technology, Engineering, and Science News*. <https://spectrum.ieee.org/robotics/artificial-intelligence/machinelearning-maestro-michael-jordan-on-the-delusions-of-big-data-and-other-huge-engineering-efforts>. Created: 20.10.2014. Retrieved: 25.06.2018.
- Haykin S (1994) *Neural Networks: A Comprehensive Foundation*. Prentice Hall PTR.
- Ito J (2016) *Design and Science*. Journal of Design and Science.
- Ito J (2012) Antidisciplinary. Joi Ito's Web. <https://joi.ito.com/weblog/2014/10/02/antidisciplinar.html>. Created: 02.10.2012. Retrieved: 26.06.2018.
- Jahn T (2008) Transdisciplinarity in the practice of research. *Transdisziplinäre Forschung: Integrative Forschungsprozesse verstehen und bewerten*. Campus Verlag, Frankfurt/Main, Germany, 21–37.

- Karch T, Kaja A, Luo Y (2018) Covington Artificial Intelligence Update: China's Vision for The Next Generation of AI. Inside Privacy. <https://www.insideprivacy.com/artificial-intelligence/chinas-vision-for-the-next-generation-of-ai>. Created: 24.03.2018. Retrieved: 13.08.2018.
- Lake BM, Salakhutdinov R, Tenenbaum JB (2015) Human-level concept learning through probabilistic program induction. *Science*, 350(6266):1332–1338.
- Lake BM, Ullman TD, Tenenbaum JB, Gershman SJ (2017) Building machines that learn and think like people. *Behavioral and Brain Sciences*, 40.
- Lang DJ, Wiek A, Bergmann M, Stauffacher M, Martens P, Moll P, Swilling M, Thomas CJ (2012) Transdisciplinary research in sustainability science: practice, principles, and challenges. *Sustainability Science*, 7(S1):25–43.
- Lanier J (2003) Why Gordian software has convinced me to believe to the reality of cats and apples. *Edge.org*. <https://www.edge.org/conversation/why-gordian-software-has-convinced-me-to-believe-in-the-reality-of-cats-and-apples>. Created: 18.11.2003. Retrieved: 26.07.2018.
- LeVine S (2017) Artificial intelligence pioneer says we need to start over. *Axios*. <https://www.axios.com/artificial-intelligence-pioneer-says-we-need-to-start-over-1513305524-f619efbd-9db0-4947-a9b2-7a4c310a28fe.html>. Created: 15.09.2017. Retrieved: 25.06.2018.
- Ling J (2001) Power of a Human Brain. *The Physics Factbook*. <https://hypertextbook.com/facts/2001/JacquelineLing.shtml>. Created: 2001. Retrieved: 19.06.2018.
- Malone TW (2018) How Human-Computer 'Superminds' Are Redefining the Future of Work. *MIT Sloan Management Review*, 59(4):34–41.
- Malsburg C (2018) The Neural Code: Roadblock on the way to AI. *Platonite*. <https://platonite.com/the-neural-code-roadblock-on-the-way-to-ai/> Created: 20.02.2018. Retrieved: 26.07.2018.
- McCarthy J, Minsky M, Rochester N, Shannon CE (1955) A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence.
- Mitchell TM (1997) *Machine Learning*, McGraw-Hill series in computer science. McGraw-Hill, New York.
- Moravec H (1988) *Mind children: The future of robot and human intelligence*. Harvard University Press.
- Ng A (2016) What does Andrew Ng think about Deep Learning? – Quora. <https://www.quora.com/What-does-Andrew-Ng-think-about-Deep-Learning>. Created: 03.02.2016. Retrieved: 25.06.2018.
- Peterson LE (2009) K-nearest neighbor. *Scholarpedia*, 4(2):1883.
- Picard RW (1997) *Affective computing*. MIT Press.
- Pohl C (2008) From science to policy through transdisciplinary research. *Environmental Science & Policy*, 11(1):46–53.
- Ramchandani J (2017) What is 'transdisciplinary'? We Learn, We Grow. <https://medium.com/we-learn-we-grow/what-is-transdisciplinary-13c16eacf57d>. Created: 24.01.2017. Retrieved: 16.08.2018.
- Ramesh P (2018) SingularityNET and Mindfire unite talents to explore artificial intelligence. *Packt Hub*. <https://hub.packtpub.com/singularitynet-and-mindfire-unite-talents-to-explore-artificial-intelligence>. Created: 09.08.2018. Retrieved: 13.08.2018.
- Sieber L (2018) White Paper. Mindfire Global. <https://mindfire.global/wp-content/uploads/2018/08/Mindfire-Whitepaper-V5.pdf>. Created: 08.2018. Retrieved: 27.08.2018.
- Stark L (2018) Canada's risky bet on AI. *The Globe and Mail*.
- Stauffacher M, Spreng D, Flüeler T, Scholz RW (2012) Learning from the Transdisciplinary Case Study Approach: A Functional-Dynamic Approach to Collaboration Among Diverse Actors in Applied Energy Settings. In: P. Krütti, T. Flüeler, D.L. Goldblatt, J. Minsch (Eds.), *Tackling Long-Term Global Energy Problems*. Springer Netherlands, Dordrecht, 227–245.
- Stokols D, P. Moser R, Hall K, Feng A (2010) Evaluating Cross-Disciplinary Team Science Initiatives: Conceptual, Methodological, and Translational Perspectives, *Oxford handbook on interdisciplinarity*. Oxford University Press, New York.
- Talwar S, Wiek A, Robinson J (2011) User engagement in sustainability research. *Science and Public Policy*, 38(5):379–390.

- Taniguchi H, Sato H, Shirakawa T (2018) A machine learning model with human cognitive biases capable of learning from small and biased datasets. *Scientific Reports*, 8(1).
- Tenenbaum JB (1999) Bayesian Modeling of Human Concept Learning, *Advances in neural information processing system*, 59–65.
- Turing AM (1950) Computing Machinery and Intelligence. *Mind*, 49:433–460.
- Urech M (2017) Cognitive Computing und künstliche Intelligenz am 17. CNO-Panel in Bern. Netzwoche. <https://www.netzwoche.ch/news/2017-11-01/cognitive-computing-und-kunstliche-intelligenz-am-17-cno-panel-in-bern>. Created: 01.11.2017. Retrieved: 13.08.2018.
- Weidman S (2018) The 3 Tricks That Made AlphaGo Zero Work. Hacker Noon. <https://hacker-noon.com/the-3-tricks-that-made-alphago-zero-work-f3d47b6686ef>. Created: 07.01.2018. Retrieved: 19.06.2018.
- Zucker D (2012) Developing Your Career in an Age of Team Science. *Journal of Investigative Medicine*, 60:779–784.

Teil II

Kreative Systeme

Zur Grenze des Verstehens künstlicher Kreativität

Jan Sebastian Zipp

Zusammenfassung

Dieser Artikel diskutiert das theoretische kreative Potential kognitiver Systeme. Er zeigt auf, wie eng verwoben die Diskussionen um menschliche und künstliche Kreativität sind und wie schwer der Kreativitätsbegriff einzuordnen ist. Aus diesem Grund wird die bekannte Unterscheidung dreier Ebenen genutzt, um das kreative Vermögen auf Basis seiner Komplexität in kombinatorische, explorative und transformative Kreativität zu unterteilen. Es werden fünf menschliche Charakteristika identifiziert, die für die höchste, transformative Ebene notwendig scheinen und über die kognitive Systeme jedoch nicht verfügen: Körperlichkeit, Individualität & Subjektivität, Bewusstsein, das Unbewusste sowie Sinnlichkeit. Anhand eines Beispiels wird verdeutlicht, warum es keinen Unterschied für jemandes Empfinden macht, ob kognitive Systeme tatsächlich kreativ sind oder es nur so wirkt, als ob sie es wären. Daher wird ein Perspektivenwechsel vorgeschlagen: Wenn akzeptiert wird, Kreativität nicht gänzlich verstehen zu können und das Erstaunen hierüber zuzulassen, kann die künstliche Kreativität des Cognitive Computings der eigenen Fähigkeit zur transformativen Kreativität augmentativ zur Seite stehen.

Überarbeiteter Beitrag basierend auf Zipp, J.S. & Vey, K.: Das kreative System – Überlegungen zur künstlichen Kreativität, *Informatik Spektrum* (2018) 41(1): 27–37

J. S. Zipp (✉)
Zeppelin Universität, Düsseldorf, Deutschland
E-Mail: j.zipp@zeppelin-university.net

1 Einleitung

Immer schon hat das Neue den Menschen in Erstaunen versetzt. Allein die Verwunderung darüber, dass überhaupt etwas ist, und nicht etwa nichts ist, wie es der Philosoph Martin Heidegger (1983) ausdrückte, beschäftigt die Menschen mindestens seit der Antike (Frankfurt 2016). Die Verwunderung über etwas und der Wunsch, es zu verstehen oder nachempfinden zu können, haben beständig zum Fortschritt der Menschheit beigetragen, indem durch den Blick aufs Neue wiederum Neues in die Welt gesetzt wurde; sei es in Form von Schriften, Kunstwerken, Dingen oder Objekten. Im derzeitigen Zeitalter der fortgeschrittenen Digitalisierung hinkt das Erstaunen über das Neue hinter der schier unglaublichen Geschwindigkeit hinterher, mit der, nahezu konstant, bahnbrechende Innovationen geschaffen werden. Innovationen, die zugleich als Einzelnes und diesseits der schieren Masse an Neuheiten auch von Experten nur schwerlich zur Gänze durchdrungen und begriffen werden können. Die Möglichkeit, dass Computer nun auch zu kreativen Prozessen fähig sind, bildet hier keine Ausnahme. Im Gegenteil: Bis vor kurzem schien Kreativität dem Menschen vorbehalten. Nun aber treten mehr und mehr kognitive Systeme an, den Gegenbeweis zu erbringen und durch Kreativität etwas Neues zu erschaffen. Kognitive Computer unterscheiden sich von traditionellen Computersystemen dadurch, dass sie von ihren Interaktionen mit Daten und Menschen lernen, um sich mit dem Ziel, neue Aufgaben zu meistern, im begrenzten Rahmen neu programmieren können und den Menschen augmentativ zur Seite stehen (Kelly III und Hamm 2013; Sebastian 2014). IBMs Chef Watson etwa, der Zutaten auf ungewohnte Weise kombinierte und mit einem Baltic Apple Pie oder dem Ecuadorian Strawberry Dessert nie dagewesene Rezepte kreierte (Colvin 2015); Googles Magenta-Projekt lässt eine künstliche Intelligenz Popsongs komponieren (Bauer 2016) und andere Systeme kreieren zeitgenössische Kunst, die beispielsweise von Besuchern der Art Basel besser angenommen werden als jene, die von menschlichen Künstlerinnen und Künstlern erschaffen wurden (Voon 2017). Hierdurch ist Kreativität nicht mehr nur ein Imperativ, der als Softskill auf keinem Arbeitszeugnis fehlen darf, sondern darüber hinaus auch Gradmesser für die Abgrenzung des Menschen von kognitiven Systemen geworden: *Künstliche Kreativität* (engl. *Artificial* od. *Computational Creativity*) hält den Menschen in vielfacher Hinsicht den Spiegel vor. Wenngleich originär künstlich und virtuell, ist die Diskussion um künstliche Kreativität im höchsten Maße vermenschlicht. Nicht nur, weil bisher Menschen die Systeme entwickeln und bauen, sondern weil sie auch immer sich selbst mit ihnen vergleichen, abgleichen und sich in ihnen sehen. Mehr noch: man sieht das, was man ist und was man eben nicht ist oder sein kann. Je beeindruckender die Resultate des Cognitive Computings ausfallen, desto mehr wächst auch der Zweifel an unserem scheinbar exklusiven Vermögen. Es scheint paradox: der Diskurs um künstliche Kreativität ist im höchsten Maße ein Diskurs um menschliche Kreativität, um das Menschsein an sich. Eben jenes Menschsein wird in der Diskussion um das kreative Vermögen kognitiver Systeme aber allzu oft in den Hintergrund gedrängt. Im Zentrum steht meist die Herausforderung, Kreativität technisch-wirtschaftlich kompatibel zu machen, sie zu messen, zu vergleichen und final verstehen zu können. Der nachfolgende

Beitrag möchte Impulse liefern, dies zu überdenken. Er bemüht sich aufzuzeigen, warum ein Perspektivenwechsel sinnvoll erscheint, der nicht nur die originär menschliche, sondern auch die künstliche Kreativität vom Menschen herdenkt.

Dieser Text stellt eine Erweiterung des Artikels *Das kreative System – Überlegungen zur künstlichen Kreativität*, veröffentlicht im Informatik-Spektrum (Februar 2018, Volume 41, Issue 1, 27–37) dar. Von diesem Informatik-Spektrum Artikel sind zu einem Teil Abschnitte übernommen worden. Diese werden im Verlauf des Textes jeweils im Vorfeld als solche genannt. Nach einer kurzen Darstellung der Schwierigkeit, künstliche Kreativität einzuordnen, wird eine Präzisierung des Kreativitätsbegriffs vorgenommen, indem drei Ebenen von Kreativität vorgestellt werden, indes sich der Text im Anschluss lediglich auf die höchste, nämlich transformative Ebene von Kreativität bezieht. Im Anschluss wird ein kurzer Abriss über die Entwicklung der künstlichen Kreativität sowie des originären, menschlichen Kreativitätsbegriffs gegeben. Diese beiden Darstellungen helfen dann, jenen Perspektivenwechsel anzugehen, der das Menschliche in der künstlichen Kreativität unterstreicht, welches notwendig für ein Verständnis transformativer Kreativität bleibt. Es wird folgend dargestellt, was dieses Menschliche ausmacht und warum kognitive Systeme hierzu aller Voraussicht nach auch in Zukunft nicht fähig sein werden. Abschliessend wird anhand der Grenze des Verstehens künstlicher Kreativität eine Form des Umgangs mit Cognitive Computing aufgezeigt, die den Menschen in den Mittelpunkt rückt und die augmentativen Stärken kognitiver Systeme unterstreicht.

2 Kreativität im Zeitalter des Cognitive Computings

Die Verwunderung über das Neue und Kreative ist von doppelter Natur. Es ist zunächst das – neugierige wie ängstliche – Erstaunen über das noch Unbekannte selbst, über das Neue *an sich*, dessen Existenz zuvor nicht vorstell- oder denkbar gewesen ist. Mit der Verwunderung einher geht aber auch die Faszination, wie dieses Neue überhaupt in die Welt kommen konnte. Der Prozess des kreativen Schaffens erstaunt im gleichen Masse wie das Neue selbst. Und wie das Neue wird auch die Kreativität, die zu diesem Neuen führt, mit Neugierde und Erstaunen, teils aber auch mit Argwohn und Befremdnis betrachtet und aufgenommen. So sind es nicht nur die Resultate, sondern häufig auch die Biografien und Lebensentwürfe beteiligter Wissenschaftler, Künstler, Entdecker und Innovatoren, die Verwunderung hervorrufen. Ein gewisses, positiv wie negativ konnotiertes Unverständnis schwingt in der subjektiven Rezeption dieser Menschen häufig mit. So beispielsweise beim Maler Vincent van Gogh, dessen abgeschnittenes Ohr zu zahlreichen posthumen Diagnosen aber auch filmischen Dokumentationen führte, welche die Umstände seiner Psyche zur Ursache haben, oder Albert Einstein, der häufig ikonisch mit zerzausten Haaren und herausgestreckter Zunge dargestellt wird. Der Ausspruch des *verrückten Genies* spiegelt diese ambivalente Haltung gut wider.

Diese Dualität von Argwohn und Neugierde gegenüber dem Neuen und Unbekannten wird in der medialen Auseinandersetzung des Cognitive Computings auf besonders eindrückliche Weise offenbar: Apokalyptische Endzeitszenarien skizzie-

ren Science-Fiction-Welten, in denen Computersysteme sich dank ihrer Superintelligenz über den Menschen erheben und konsequenterweise meist auch die Weltherrschaft anstreben (etwa in Filmen wie Stanley Kubricks 2001 über Terminator bis hin zu Matrix). Auf der anderen Seite der Skala zeichnen Vertreter der technischen Singularität ein idealisiertes Bild des (derzeit noch hypothetischen) Moments, ab dem kognitive Systeme in nahezu allen Belangen dem menschlichen Geist und Körper überlegen sind und sich sowie die Menschheit selbstständig weiterentwickeln. Das Erstaunen über die Entwicklung reicht also von kategorischer Verteufelung bis hin zur apologetischen Vergötterung (Goertzel 2018). Hierbei scheinen sich die Ansichten nur selten in der ausgewogenen Mitte aufzuhalten, sondern an den äusseren Rändern dieser zwei Pole. *Freund oder Feind* heisst häufig die Devise (FAZ 2018). Gerade weil der Kreativitätsbegriff sich nicht klar definieren lässt, stellt er ein geeignetes Vehikel dar, um je nach Standpunkt vom einen oder anderen Extrem angeeignet zu werden. Dies zeigt, wie eng verwoben die Forschung zu menschlicher sowie künstlicher Kreativität eigentlich ist und wie schwer greifbar ein umfassendes und vor allem einheitliches Verständnis von Kreativität dennoch bleibt. Denn längst nicht immer ist klar, worüber eigentlich gesprochen wird, wenn im Namen der Kreativität Argumente vorgebracht werden. Letztlich besitzt jeder Theoretiker, der sich mit diesem Feld beschäftigt, seine oder ihre favorisierte Kreativitätstheorie (Veale et al. 2010). So schwimmt der Begriff in Interpretationen und Zuschreibungen und ermöglicht hierdurch eine nahezu willkürliche Aufladung des Wortes, die in nahezu alle Bereiche des Lebens einsickert. Der Kreativitätsbegriff ist heutzutage ein omnipräsenter Gebrauchsgegenstand, der sowohl die Schaffung eines nie zuvor gedachten Meisterwerks (wie etwa Johann Sebastian Bachs Klavierfugen oder Romane wie James Joyces Ulysses) als auch trivialste Alltagshandlungen (wie das Umdekoration eines Schaufensters oder die Variation eines Spaziergangs durch Nutzung einer Seitengasse) umfassen kann (Reckwitz 2012). Ist also von künstlicher Kreativität die Rede, ist keineswegs klar, was überhaupt im jeweils speziellen Fall mit Kreativität gemeint ist. Genau dies soll für diesen Text im Folgenden durch eine Annäherung an eine bestimmte Deutung des Kreativitätsbegriffs vermieden werden.

3 Der Kreativitätsbegriff: Kreativität ist nicht gleich Kreativität

Zunächst ist es wichtig herauszustellen, welche *Form* oder *Variante* der enormen Bandbreite des Kreativitätsbegriffs gemeint ist, wenn hier von Kreativität die Rede ist. Es geht dabei nicht um das Anliegen, den Begriff um einen weiteren Definitionsversuch zu bereichern, sondern lediglich um eine Fokussierung, die es erlaubt, sich mit jenen Aspekten des Kreativen auseinanderzusetzen, welche die Verwunderung über das Neue am ehesten provoziert. Dies setzt voraus, dass es weniger darum geht, Kreativität zu *verstehen*, sondern vielmehr zu *erfahren*; eine Unterscheidung, die später im Kapitel erneut aufgegriffen und in den Mittelpunkt gestellt werden wird. Hierzu ist es notwendig, über das Gros an kreativen Zu-

schreibungen hinauszublicken, wie beispielsweise der Journalist Geoff Colvin (2015), der Computern zwar grundsätzlich die Fähigkeit zur Kreativität bescheinigt, dem Menschen aber die *High-value Creativity* (deut. in etwa *hochkarätige Kreativität*) vorbehält.

Was hiermit genau gemeint sein kann, zeigt die Kognitionswissenschaftlerin Margaret Boden, die eine Aufteilung und Hierarchisierung des Kreativitätsbegriffs in drei Ebenen vorschlug (Boden 2004), was ein breites Echo im Feld der Kreativitätsforschung fand (Ram et al. 1995) und die hier auf Basis des eingangs erwähnten Informatik-Spektrum Artikels aus dem Februar 2018 verkürzt dargestellt werden soll. Boden teilt Kreativität in drei Hauptkategorien ein, die sich massgeblich durch ihren Grad an Unkonventionalität und Komplexität unterscheiden (Boden 2004).

3.1 Die unterste Ebene – Die kombinatorische Kreativität

Sowohl Unkonventionalität wie auch Komplexität sind auf dem *untersten* Level vergleichsweise gering ausgeprägt. Diese basale Ebene bezeichnet Boden als *kombinatorische Kreativität*. Innerhalb bekannter und vertrauter Ideen werden hier ungewohnte Kombinationen hergestellt, die auf ihre Art und unter Berücksichtigung bestehender Regeln und Normen neue Herangehensweisen oder ungewöhnliche Zugänge zu bestehenden Inhalten aufzeigen. Dies kann zu verblüffenden und schönen Ergebnissen führen, etwa durch neu aufgezeigte Analogien oder ein gemaltes Bild, das einen bekannten Stil auf eindrückliche Weise imitiert, wie es etwa Googles Magentaprojekt anstrebt (Brandom 2016). Überhaupt ist diese Form von Kreativität nicht zu unterschätzen. Durch die kreative Fähigkeit, Analogien zu bilden, lassen sich zum Beispiel zukünftige Szenarien und passende Verhaltensmuster entwickeln. Bereits Gelerntes kann auf unbekannte Situationen angewandt werden, da durch Analogien Gemeinsamkeiten zu vertrauten Situationen entstehen. Ein berühmtes Beispiel hierfür ist Duncikers *Radiation Problem*, bei dem die Frage im Raum steht, wie ein Tumor bestrahlt und zerstört werden kann, ohne zu viel gesundes Gewebe zu vernichten. Ein möglicher Lösungsansatz ergibt sich, wenn eine Verbindung zu einer BURGbelagerung gezogen wird, bei der die angreifenden Truppen die Mauer lediglich an einer Stelle überwinden, die BURG einnehmen und so die eigenen Verluste minimieren (Gick und Holyoak 1980). Der Kognitionswissenschaftler Jeff Hawkins stellt in seinem Bestseller *On Intelligence* fest, dass die Herstellung von Querverbindungen (in diesem Fall eben zwischen Tumorbestrahlung und BURGbelagerung) äusserst komplex und mit grossen Schwierigkeiten verbunden sei. Da sie aber alltäglich auftrete, falle uns ihre Besonderheit gar nicht mehr wirklich auf (Hawkins und Blakeslee 2007). Ein Grossteil dessen, was als künstliche Kreativität bezeichnet wird, scheint auf dieser untersten Ebene der Kreativität eingeordnet werden zu können. Die zwei weiteren Arten von Kreativität in Bodens Unterteilung unterscheiden sich grundlegend von der kombinatorischen Variante, da sie das tatsächlich Neue und das Moment des Überraschenden in den Fokus der Kreativität rücken (Boden 2004).

3.2 Die mittlere Ebene – Die explorative Kreativität

Zunächst wäre da als nächste Ebene die *explorative Kreativität*. Gemeint ist hier die Fähigkeit, innerhalb eines bestimmten begrifflichen Konzeptes (engl. *Conceptual Space*), welcher zum Beispiel für eine musikalische Stilrichtung, wissenschaftliche Theorie oder die Nouvelle Cuisine stehen kann, auf Entdeckungsreise zu gehen und neue, überraschende Ideen innerhalb des existenten Feldes zu kreieren. Hierdurch wird ein konzeptuelles Feld, ein Stil, eine Denkschule weiterentwickelt, verbessert, ihre Grenzen erkundet und bisweilen sogar ausgereizt. Die meisten Künstler und Wissenschaftlerinnen tun genau das und drücken sich mit ihrer Kreativität im explorativen Sinne aus. Der Wert der explorativen Kreativität liegt im Aufzeigen neuer Möglichkeiten innerhalb eines bestehenden Rahmens, die vorher niemand vermutete oder kannte (Pearce 2010). Boden nennt als Beispiel Strassenkünstler, die täglich neue Karikaturen und Portraits zeichnen. Durch kontinuierliches Austesten und Abändern ihres Stils verbessern sie nach und nach ihre Fähigkeiten und schaffen immer wieder überraschende, neuartige Kunstwerke. Allerdings verbleiben sie trotz aller Neugierde am Entdecken letzten Endes eben doch in ihrem wohlvertrauten Stil; sie reizen *lediglich* das Potential aus, welches schon immer vorhanden war (Boden 2011).

3.3 Die höchste Ebene – Die transformative Kreativität

Nun kommt es gelegentlich, wenn auch äusserst selten, vor, dass jemand etwas denkt, was auf Grundlage der diesem Menschen bekannten konzeptuellen Muster eigentlich nicht gedacht werden könnte. Eine *verrückte* Idee also, welche die vorherrschenden Denkmuster radikal herausfordert und zu einem neuen Stil, einer neuen Theorie, also neuen konzeptuellen Räumen führen kann. Diese dritte Form kreativen Schaffens bezeichnet Boden als *transformative Kreativität* (Boden 2004). Eine Transformation rüttelt bestehende Strukturen auf und stellt das Gewohnte infrage, mehr noch, sie strebt die Auswechslung des Alten durch etwas grundlegend Neues an. Damit ist sie zugleich Gefahr für alles Etablierte. So erschwert der sehr hohe Grad des Neuen als Ausdruck transformativer Kreativität ein allgemeines Verständnis denn auch im höchsten Masse. Zu ungewohnt und anders scheinen bisweilen die neuen Ideen, als dass man sie direkt begeistert aufnehmen und adaptieren könnte. Aus diesem Grund erscheint diese Form der Kreativität bisweilen irrational und oftmals nicht nachvollziehbar: Weshalb malte Picasso nicht sein Lebtage weiter vertraute und gefällige Motive, anstatt abstrakte Werke zu schaffen? Weshalb gehen Köche das enorme Risiko ein, sich und ihren Ruf zu ruinieren, nur um etwas anders zu machen, einen neuen Stil zu etablieren, der vertraute Geschmäcker irritiert und nicht zuvorderst Gefälligkeit anstrebt? Wie und warum entstand eine solch komplexe und für ungeübte Ohren kaum gefällige Musik wie die Zwölftonmusik? Die Chance, insbesondere in der Kunst und ihren verwandten Disziplinen, mit einer revolutionären Idee offene Türen einzurennen und tosenden Applaus zu ernten, ist bekanntermassen sehr gering. Ein interessantes Beispiel hierfür gibt das skandi-

navische Restaurant *Noma* mit seinem Chefkoch René Redzepi ab, welches 2010, 2011, 2012 und 2014 als bestes Restaurant der Welt ausgezeichnet wurde.¹ Diese Auszeichnung erhalten nicht etwa Restaurants, die vorhandene Gerichte handwerklich perfektioniert haben. Wichtiger ist vielmehr die Weiterentwicklung der Kochkunst, das überraschende, neue Element in der Kulinarik. Dennoch stand das Noma in seiner Geschichte mehrmals kurz davor, bankrott zu gehen (Eisenhart 2016). Kritiker schrecken jedoch nicht davor zurück, Redzepi mit Mozart zu vergleichen und ihm einen *automatischen Zugang zum Genialen* zu attestieren (Deschamps 2015). Oder, wie jemand unter dem Eindruck seiner Gerichte, die auch schon mal Ameisen beinhalten und auf Laub serviert werden können, schrieb: „*Sie servieren Dinge, von denen ich keine Ahnung hatte, dass sie existieren*“ („*They serve things I didn't know existed*“) (Dravecky 2016).

Ist im weiteren Verlauf des Textes von *Kreativität* die Rede, so ist deshalb lediglich jene transformative Kreativität gemeint, die nach Boden die höchste Stufe des kreativen Denkens darstellt. Von Relevanz ist hier also die transformative Kreativität und nicht die kombinatorische sowie explorative Form, da diese Variante es am ehesten vermag, jenes anfänglich beschriebene Erstaunen über das wirklich Neue hervorzurufen, welches für den hier angestrebten Perspektivenwechsel unerlässlich scheint.

4 Eine kurze Entwicklungsgeschichte der Künstlichen Kreativität

Doch wieso ist es selbst dann so schwer, Kreativität genauer zu greifen und verständlich zu machen, wenn lediglich radikal-transformative in Bodens Sinne gemeint ist? Hierzu lohnt sich ein Blick in die Entwicklungsgeschichte der künstlichen Kreativität und des parallel verlaufenen Diskurses um ihre Deutungshoheit, zu dessen Zweck erneut auf den eingangs erwähnten Informatik-Spektrum Artikel zurückgegriffen wird.

Allgemein gesprochen beschäftigt sich die künstliche Kreativität mit der Entwicklung und grundsätzlichen Möglichkeit, ein kognitives System Kreativität simulieren oder replizieren zu lassen. Wenngleich die multidisziplinäre Forschung in letzter Zeit durch bahnbrechende Entwicklungsschritte auf dem Gebiet der künstlichen Intelligenz vermehrte Aufmerksamkeit erfahren hat, begann die Auseinandersetzung mit dem Thema der künstlichen Kreativität schon vor über 60 Jahren. Bereits 1950 dachte der amerikanische Psychologe Joy Paul Guilford darüber nach, dass der einzig verbliebene ökonomische Wert des menschlichen Gehirns das kreative Denken sei, da Maschinen nicht nur die Muskelkraft, sondern auch das logische Denken in weiten Teilen obsolet werden liessen, es aber immer noch des Menschen bedürfe, um diese Maschinen zu leiten und weiterzuentwickeln (Guilford 1950). Diese Aussage behielt ihre Gültigkeit für mehr als ein halbes Jahrhundert, bevor sie insbesondere in letzter Zeit durch die Entwicklung potentiell kreativer Systeme

¹ www.theworlds50best.com.

vermehrt unter Druck geriet. Gleichzeitig verweigert sich der Kreativitätsbegriff weiterhin hartnäckig einer klaren Definition, wenngleich es an Angeboten nur so wimmelt.

Trotz des schier endlosen Dickichts an Definitionsangeboten zum Kreativitätsbegriff, bemühen sich zuvorderst Institutionen und Verbände um ein klareres Begriffsverständnis, die sich durch Forschung und Entwicklung kognitiver Systeme unmittelbar mit dem Phänomen der künstlichen Kreativität auseinandersetzen. Aufmerken lässt insbesondere die *Association for Computational Creativity*,² ein Verband zum Thema künstliche Kreativität, der sich sowohl um die Weiterentwicklung als auch um die Deutungshoheit dieses Feldes bemüht (Jordanous 2014). Bis 2013 schlug dieser Verband vor, künstliche Kreativität sei etwas, dass sich damit beschäftige, (natürliche und künstliche) Leistungen und Verhaltensweisen, die bei Menschen als kreativ bezeichnet würden, mit computerunterstützten Methoden zu modellieren und zu verstehen. Die genaue Formulierung lautete: „*Computational Creativity is the study and simulation, by computational means, of behaviour, natural and artificial, which would, if observed in humans, be deemed creative*“ (Jordanous 2012, S. 248). Kreativität ist also, was kreativ anmutet. Der aktualisierte Definitionsansatz geht noch einen Schritt weiter und fokussiert die Umschreibung des interdisziplinären, wissenschaftlichen Feldes. So sei die künstliche Kreativität in erster Linie das wissenschaftliche Bestreben, ein Programm zu entwickeln, welches über ein dem Menschen ebenbürtiges kreatives Vermögen verfügen solle (computationalcreativity.net). Im Zuge dessen sei es ebenfalls Aufgabe der *künstlichen Kreativitätsforschung*, mehr über das Phänomen der menschlichen Kreativität zu lernen und Systeme zu entwickeln, die jene menschliche Kreativität verstärken (ohne notwendigerweise selbst kreativ sein zu müssen). Ähnlich hält es das *PROSECCO* (*Promoting the Scientific Exploration of Computational Creativity*) Forschungsnetzwerk:³ Es beschreibt *Computational Creativity* als wissenschaftliches Feld, welches sich unter anderem zum Ziel gesetzt hat, die Frage zu beantworten, was es bedeute, kreativ zu sein und merkt an, dass diese Frage (und viele weitere verwandte Fragen) im gleichen Masse für die künstliche, wie auch menschliche Kreativitätsforschung Gültigkeit besitzt. Um sich der Semantik einer künstlichen Kreativität anzunähern, erscheint es daher sinnvoll, in aller Kürze die Herkunft und Entwicklung des originalen Kreativitätsbegriffs nachzuzeichnen, der schliesslich in der heutigen omnipräsenten Faszination mündete.

5 Die Entwicklung des originalen Kreativitätsbegriffs

Laut der IBM Global CEO Study sahen bereits im Jahr 2010 60 % von 1541 Führungskräften aus aller Welt die Fähigkeit zum kreativen Handeln als die wichtigste Eigenschaft einer erfolgreichen Führungspersönlichkeit an (Lombardo und Roddy

²<http://computationalcreativity.net/home/about/computational-creativity/>.

³<http://prosecco-network.eu/introduction-computational-creativity/>.

2010) – was auch immer jede(r) Einzelne von ihnen genau darunter verstand. Wie anfangs angedeutet, sind jedoch weder die Faszination noch das Streben nach Kreativität ausschliesslich Phänomene unserer Tage. Relativ neu ist nur der Begriff: 1973 taucht die Kreativität im deutschsprachigen Raum erstmalig im Rechtschreibduden auf, und zwar von Beginn an ohne klare Definition⁴ (duden.de). Vielmehr wird sie dort als *schöpferische Kraft* und *kreatives Vermögen* bezeichnet, verweist also im Ersteren mit der Verwendung des Schöpfungsbegriffs auf ihren Ursprung in der Theologie, in dem der *Creator*, der Schöpfergott, aus Nichts etwas Neues schafft (Schmidt 1985) – und im Letzteren auf sich selbst. Geht man in der Geschichte weiter zurück, lässt sich erkennen, dass bis ins 17. Jahrhundert hinein die Fähigkeit, Neues zu schaffen, einer göttlichen Instanz vorbehalten blieb. Erst als der Geniebegriff unter dem Rückgriff auf antike Ideale im Auftrieb begriffen war, wurde auch der Mensch quasi offiziell kreativ und befähigt, etwas vollkommen Neues zu erschaffen, also *Schöpfer* zu sein. Das Wort *Mensch* sollte hier allerdings äusserst vorsichtig und keineswegs allgemeingültig betrachtet werden. Schliesslich war das geniale Schöpferische nur wenigen individuellen (und männlichen) Menschen vorbehalten und keineswegs eine allgemein erreichbare Befähigung.

Die Darstellung des einsam-kreativen Genies steht im starken Kontrast zur modernen Kreativitätsforschung, die in den 1950er-Jahren in der Psychologie eine erste wissenschaftliche Heimat fand und sich schnell vom theologisch-philosophischen Erbe emanzipierte. Bereits 1950 stellte der Psychologe Guilford in einem Fachbeitrag die berühmt gewordene Behauptung auf, dass jeder Mensch kreativ sei. Joseph Beuys' ebenfalls so bekanntes Diktum *jeder Mensch ein Künstler* aus den 70er-Jahren (Stachelhaus und Foglia 1991) und das vom US-Ökonomen Richard Florida (2002) entwickelte Konzept *der kreativen Klasse* öffneten der Kreativität den Weg aus dem psychologischen Wissenschaftsdiskurs und demokratisierten den Begriff weiter. Zugleich werden alle drei Postulate in diesem Zusammenhang aber oft verfälscht, zumindest aber verkürzt dargestellt.

Guilford geht davon aus, dass es bestimmter Fähigkeiten (engl. *abilities*) bedarf, um kreativ zu sein. Da alle Menschen zu einem bestimmten Masse alle Fähigkeiten (exkl. der pathologischen) besässen, könne auch von nahezu allen Menschen eine grundlegende Befähigung zur Kreativität erwartet werden (Guilford 1950). Beuys wiederum bezog sich mit seiner Aussage auf einen erweiterten Kunstbegriff, welcher der Kunst eine gesellschaftsverändernde Kraft zuspricht und nicht mehr nur das plastische, materielle Werk umfasst, sondern etwa auch Performances und künstlerische Interventionen. Durch sein soziales Handeln kann dementsprechend jeder Mensch künstlerisch wirken, also auch ein Künstler sein, ohne jedoch gleich die Befähigung zur grossen Malerin oder Komponistin in sich zu tragen (Harlan 2001). Auch Floridas kreative Klasse bezieht sich nicht auf die Vorstellung einer nur zum Selbstzweck agierenden Künstlergemeinschaft, sondern unterstreicht wirtschaftstheoretisch die (vermutete) Bedeutung von Innovationen für ein ökonomisches Wirtschaftswachstum und untersucht etwa die hierfür benötigten infrastrukturellen Voraussetzungen (Florida 2002).

⁴<http://www.duden.de/suchen/dudenonline>.

Nichtsdestotrotz kann die heutige inflationäre Konjunktur der Kreativität als eine direkte Folge dieser Entwicklungen angesehen werden (Reckwitz 2012) und zeigt ihre unklare Verwobenheit zwischen zahlreichen Wissenschaften, der Kunst und der Ökonomie auf. Genau dieser Klimax der diffusen Indifferenz erlaubt einen nahezu beliebigen und hochgradig flexiblen Gebrauch für unterschiedlichste Anwendungsgebiete. Diese Flexibilität mag ein massgeblicher Grund für die ungeheure und anhaltende Attraktivität der Kreativität sein. Zugleich ist sie damit aber auch Garant für zahlreiche Missverständnisse im Umgang mit dem Begriff, was aus Sicht der Kunst, welche die Kreativität als einen erweiterten Kunstbegriff betrachtet, zu einer *Absage* an den heutigen Gebrauch von Kreativität führte (van den Berg 2009). Ungeachtet weitergehender Definitionsangebote scheint es in der Auseinandersetzung mit künstlicher Kreativität sinnvoll, sich dieses Spannungsfeldes bewusst zu bleiben – insbesondere in Hinblick auf die vorher erwähnten interdisziplinären und immer auch die menschliche Kreativität berücksichtigenden Definitionsansätze.

6 Künstliche Kreativität: Ein Perspektivenwechsel

So sind es denn auch jene Projekte zur künstlichen Kreativität, die in den doch eigentlich exklusiv dem Menschen vorbehaltenen Bereichen der Kunst angesiedelt sind, welche massgeblich die Diskussion um die Frage befeuern, was künstliche Kreativität vermag und was eben nicht. Ein bekanntes Beispiel ist das Computerprogramm *AARON*, das seit den 1970er-Jahren vom britischen Künstler Harold Cohen entwickelt und immer weiter verbessert wurde und als Inspiration für Projekte wie Googles Magenta gilt. Das System ist in der Lage, abstrakte Bilder herzustellen, die es bis in die Tate Gallery geschafft haben. Interessanterweise sieht der *Schöpfer* Cohen selbst das Programm gar nicht als kreativ an; unter anderem deshalb, weil es neue Stile und Darstellungsweisen nicht selbstständig erlernen oder entdecken kann (Cohen 1995). Vielmehr müssen diese vom Menschen in passender Codeform geschrieben werden. Dennoch fordert Cohen den kreativen Exklusivitätsanspruch des Menschen heraus, indem er fragt, was es denn dann sei, was *AARON* produziert und wie sich dies denn überhaupt vom *Real Thing*, dem vom Menschen produzierten Kunstwerk, unterscheiden lässt (Cohen 1995).

Eine gute Frage, die jedoch recht missverständlich aufgefasst wurde. Denn aus entwicklungstechnischer Sicht mangle es zur Beantwortung dieser Frage bisher an der Möglichkeit einer systematischen Evaluation künstlicher Kreativität. Es werde die Befähigung zur Messbarkeit benötigt, *wie* kreativ etwas oder jemand sei, um dies unter- und miteinander zu vergleichen. Denn nur durch eine klare Vergleichbarkeit sei nachvollziehbar, wie gerechtfertigt die Behauptung wäre, eine Maschine sei zu kreativer Leistung fähig. Was fehle, sei Präzision, Standardisierung und ein systematischer Zugang (Jordanous 2012). Und tatsächlich gibt es einen Versuch des Department of Computer Science der Rutgers University, Kreativität auch in der Kunst zu quantifizieren (Elgammal und Saleh 2015), indem Aspekte wie Originalität und der historische Wert ihres Einflusses berücksichtigt und in Skalen überführt werden. So interessant diese Perspektive anmutet, kann jener Versuch nicht die letztendliche *Unfassbarkeit* des vorher eingeführten radikal-transformativen Kreati-

vitätsbegriffs berücksichtigen. Um dies nachzuvollziehen, hilft womöglich ein Perspektivenwechsel weiter: Nicht die Frage, *was* oder *wieviel* Kreativität sei, erscheint sinnvoll, sondern eben *wie* Kreativität sich ausdrückt, manifestiert und formt. Kreativität ist auch Nicht-Wissen, ein Zauber, das in Worten nicht Greifbare. Es ist vor allem und zuvorderst nicht vergleichbar, nicht verhandelbar und nichts final Verstehbares. Kreativität ist eigentlich nicht erklärbar. Und das lässt sich erklären.

Mit welchem Recht beansprucht Kreativität (und deshalb auch die Kunst) überhaupt irgendeine Form von Wissen zu produzieren (Mersch 2015)? Das Problem liegt nicht im Begriff, sondern in unserem Verständnis von Wissen, indem man (in diesem Falle fälschlicherweise) von faktischen und damit vergleichbaren Begebenheiten ausgehen. Traditionell spielt sich das Denken in der Sprache ab und äussert sich in Begriffen und Argumenten. Jeder Gedanke ist der sinnvolle Satz (Wittgenstein 1984). Über die Sprache gelingt es, Wahrgenommenes zu verstehen, indem man es benennt und damit klassifiziert. Ein Tisch beispielsweise ist ein Tisch und kein Stuhl (wenngleich soziokulturelle Unterschiede hier durchaus zu Missverständnissen führen können). Diese scheinbar simple Feststellung hilft, die wahrgenommene Welt zu verstehen und ermöglicht es sich in dieser auszudrücken. Das eigentliche Medium scheint also die Sprache zu sein. Dass ein Verständnis nicht nur für den Menschen, sondern insbesondere für kognitive Systeme deutlich schwieriger wird, wenn wir diese Welt mit Adjektiven beschreiben wollen und etwa von einem schönen oder hässlichen Tisch reden, zeigen die Arbeiten zur Fuzzy Logik, die sich um die präzise Erfassung des Unpräzisen bemüht (siehe z. B. (D’Onofrio und Portmann 2015)). Was mit Zuschreibungen exakt gemeint ist, bleibt also ungewiss und wird von jedem Menschen unterschiedlich interpretiert. Dies zeigt, wie ungenau und auch unsicher Sprache letzten Endes ist (vgl. z. B. (Trillas et al. 2019)). Noch weiter geht der französische Maler René Magritte, indem er offensichtlich eine Pfeife malt und darunterschreibt, dies sei eben keine Pfeife, wie es Abb. 1 zeigt. Hier ist das abgebildete Objekt nicht mit dem sprachlich ausgedrückten Objekt selbst identisch. Andere Werke entziehen sich vollends jeder sprachlichen Interpretation. Die Klarheit der Sprache gerät beim Versuch des Verstehens kreativer Prozesse ins Wanken. Denn kreatives Denken vollzieht sich anders.

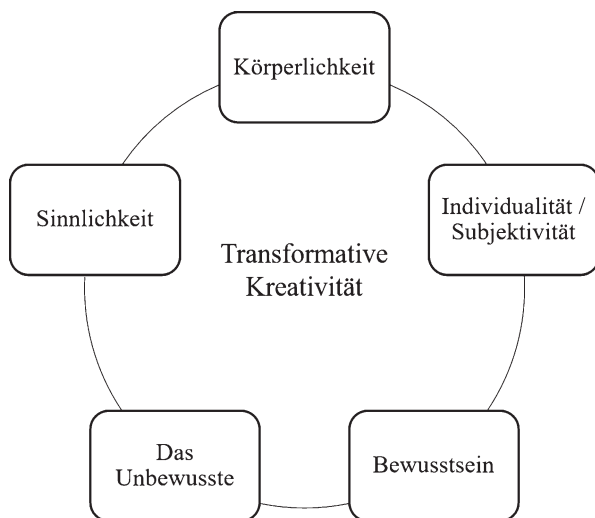
Abb. 1 René Magritte: La trahison des images (1929)



Dieses operationale Missverständnis ist der Grund, warum man endlos debattierend vor einem Bild im Museum stehen kann, versuchend diesem scheinbar gedankenlosen Werk aus dem durch Spracheexistenten Wissensfundus durch sprachliches Nachdenken über dieses Werk einen Sinn zu vermitteln. Das reguläre Denken sucht immer, Begriffe zu formulieren. Kunst und Kreativität aber verkörpern Ausdrücke, die sich nicht versprachlichen und auch nicht in *unscharfe Mengen* einbringen lassen, wie sie in der Fuzzy Logik bemüht werden. Eine Gestalt verstörender Rätsel, die sich anders nicht aufzeigen lässt; die Verkörperung eines anderen Wissens also, welches nur auf ihre jeweilige Weise ausgedrückt werden kann. In der besonderen Konstruktion einer bestimmten Reflexivität wird etwas arrangiert, das etwas aufdeckt und eine Evidenz entstehen lässt, die nur im Sinnlichen erfahrbar ist. Sie drückt sich über das wahrgenommene Empfinden aus, nicht durch Sprache und lässt sich auch nicht in eine solche übersetzen. Das Produkt der (transformativen) Kreativität ist kein Werk, sondern ein Gewirke. Die Duplizität von Spaltung und Verbindung, entstehend aus der zufälligen (aber nicht beliebigen) Begegnung zweier disparater Dinge (Mersch 2015).

Der Diskurs um künstliche Kreativität berücksichtigt diese Perspektive jedoch nicht im ausreichenden Masse, da eine irgendwie doch quantifizierbare Kreativität vorausgesetzt wird. Für ein kombinatorisches und vielleicht sogar exploratives Verständnis von Kreativität mag dies von Vorteil sein, die höchste Kreativität, das *Real Thing* also, wie es Cohen beschrieb (Cohen 1995, S. 10) lässt sich auf diese Weise nicht verstehen. Der Mensch besitzt einige Charakteristika, die für ein Verständnis notwendig scheinen, also gleichermassen für die Beobachtung als auch Produktion transformativer Kreativität unabdingbar sind und die für kognitive Systeme aber trotz ihrer exponentiellen Entwicklungssprünge wohl kaum implementierbar und noch weniger kompatibel wirken. Diese hierfür notwendigen menschlichen Eigenschaften hängen eng zusammen und formen in gegenseitiger Abhängigkeit das, was Menschen sind und was ihre Persönlichkeit ausmacht, wie Abb. 2 illustriert: *Körperlichkeit, Individualität, Bewusstsein, das Unbewusste und Sinnlichkeit*.

Abb. 2 Die fünf notwendigen Charakteristika transformativer Kreativität (eigene Darstellung)



7 Charakteristika für Transformative Kreativität

Für eine Kurzdarstellung dieser fünf Charakteristika wird erneut aus dem vorhergehenden Text in leicht gekürzter und abgewandelter Form zitiert.

7.1 Körperlichkeit

Seit den 1950er-Jahren bemüht sich die Wissenschaft, künstlichen Systemen die Welt durch virtuelle Modelle verständlich zu machen, die auf Symbolen beruhen (Medlock 2017). Verständnis ist jedoch mehr, als die Fähigkeit Symbole zu deuten, um die Welt objektiv zu erkennen. Schon seit der griechischen Antike beschäftigt sich der Mensch mit dem Verhältnis von geistigen und körperlichen Zuständen, dem sogenannten *Leib-Seele-Problem*. Für die überwältigende Mehrheit aller bisherigen Theorien lautet der allgemeine Tenor: *ohne Körper kein Geist* (Beckermann 1999). Wie sich diese zwei Bereiche nun aber genau zueinander verhalten, dass sie gemeinsam ein bewusstes Sein hervorbringen, ist noch immer nicht gänzlich geklärt. Einen wichtigen Beitrag zur Beantwortung dieser Frage liefert die sogenannte Embodimentforschung, die sich seit über 30 Jahren mit der Idee auseinandersetzt, der Körper sei eine notwendige Voraussetzung für Intelligenz, diese also ohne einen Körper überhaupt nicht möglich sei (Pfeifer und Bongard 2007). Neurowissenschaftliche Untersuchungen haben etwa ergeben, dass Bewegungen im Körper initiiert werden, noch bevor wir uns dem Wunsch oder Willen einer Bewegung bewusstwerden. Der bewusste Wille, eine Bewegung ausführen zu wollen, entsteht erst etwa 300 Millisekunden nachdem im Gehirn die jeweiligen Areale begannen, diese Bewegung vorzubereiten (Libet et al. 1983).

Auch die Art und Weise, wie sich die Welt offenbart und welchen Teil von ihr man auf welche Weise wahrnimmt, wird vom Körper in Form der zwar vielfältigen, aber doch limitierenden sensorischen Organe beeinflusst. Denkt man an eine Katze, kann man auf einen wahren Schatz an sensorischen Eindrücken zurückgreifen, um sich dieses Tier vorzustellen: Ihr Aussehen, das Fell, der Geruch, die Laute, aber auch die Erfahrungen, welche man mit einer Katze gemacht hat. Das Konzept *Katze* besteht nicht nur aus abstrakten Informationen, sondern ist massgeblich durch unsere Körperlichkeit geprägt. Demnach denkt man mit dem ganzen Körper, nicht nur mit dem Gehirn (Damasio 1994). Beispielsweise nimmt man mit der Nase den Geruch der Katze wahr, durch die Hände bemerkt man ihr weiches Fell, bevor die Erinnerung zurückkehrt, wie das Tier einen anfauchte, als man es das letzte Mal streicheln wollte, sodass man die Hände schnell zurückzieht.

Erdenkt ein Mensch ein neues Konzept, so ist ein grosser Teil der Arbeit bereits getan, da man sich auf ein bestehendes Weltmodell berufen kann, welches zahlreiche evolutionäre, sowie individuelle Erfahrungen und Situationen kennt und beinhaltet. Ein kognitives System wiederum muss jedes Mal von Neuem beginnen. Zwar gibt es Forschungsbestreben im Bereich des maschinellen Lernens (engl. *machine learning*), bereits erlerntes maschinelles Wissen neuen Situationen zugänglich zu machen (Medlock 2017), auch bekannt als *Transfer Learning* oder *Inductive Transfer*.

Ziel ist es, das bei der Lösung eines Problems gewonnene Wissen zu speichern, um es auf ein anderes, aber verwandtes Problem anzuwenden (Pratt und Thrun 1997). Doch scheint dieses Modell auf absehbare Zeit nicht an die Vielfalt und Möglichkeiten heranzureichen, die ein menschlicher Körper durch evolutionäre und persönliche Erfahrung ermöglicht (Medlock 2017).

7.2 Individualität/Subjektivität

Die Umwelt wird von jedem Menschen subjektiv beurteilt. Die Welt ist demnach in der höchsten Masse einer individuellen Wahrnehmung unterworfen. So zeigt etwa das vorherige Beispiel zu Vincent van Gogh oder Albert Einstein die Diskrepanz zwischen ihrer möglichen eigenen Wahrnehmung der Welt und der vieler anderer Menschen. Ein Zusammenhang zwischen Individualität und Kreativität ist für die Forschung seit jeher gleichermassen anerkannt wie undurchsichtig, was zahlreiche Veröffentlichungen zum Thema belegen (Leach 2007; Montuori und Purser 2006; Sawyer 2011). Viele kreative Werke lassen sich ohne das individuelle Empfinden eines Künstlers nicht fassen: Immer wieder ist es nicht die abstrakte Beobachtung, sondern die unmittelbare, subjektive Erfahrung, die Kreativität erst ermöglicht (Hoffmann 2004). Subjektivität ist gleichermassen Voraussetzung für Kreativität, wie diese es für die Entfaltung der Subjektivität ist (Ruppert 1998). Die berühmte mexikanische Malerin Frida Kahlo etwa ist insbesondere durch ihre Selbstportraits bekannt, auf denen sie auch immer wieder Schicksalsschläge wie einen Busunfall oder wie in Abb. 3 eine Affäre ihres Mannes mit ihrer Schwester thematisiert.

Auch Künstlerkollektive und ähnliche Verbünde bestehen aus einer Vielzahl an Individuen mit ihrer jeweils eigenständigen Subjektivität, sodass letztlich auch hier ein Zusammenhang zwischen Individualität und Kreativität nicht verleugnet werden kann.

Dank der Robotik und virtuellen Assistenten wie Apple's *Siri*⁵ oder Amazons *Alexa*⁶ ist es nun möglich, kognitiven Systemen eine physische Präsenz zu verleihen. Diese können Daten wie Namen oder Hobbies aus ihrem Umfeld abspeichern, was eine gewisse Kontextualisierung ermöglicht und das Gefühl vermittelt, es mit einer individuellen, dadurch einzigartigen Einheit zu tun zu haben. Aber diese Daten sind nicht auf einen Roboter beschränkt, sondern können mehr oder minder beliebig auf andere übertragen werden. Es ist ein ungleiches Verhältnis: der Mensch als individuelles Subjekt ist auf seinen Körper beschränkt und empfindet *lediglich* seine wahrnehmbare Umwelt, während sich jederzeit ein neuer Roboter organisieren liesse, der die simulierte *Persönlichkeit* des alten Systems annähme. Eine exklusive Individualität, wie sie jeder Mensch besitzt, kann für ein kognitives System allein

⁵<https://www.apple.com/siri>.

⁶<https://developer.amazon.com/alexa>.

Abb. 3 Frida Kahlo:
Memory (The Heart)
(1937)



deshalb schon nicht entstehen, weil diese cloud-basiert sind. *Cloud Computing* umschreibt ein Modell für den allgegenwärtigen, komfortablen On-Demand-Netzwerkzugriff auf einen gemeinsamen Pool von konfigurierbaren Computerressourcen (z. B. Netzwerke, Server, Speicher, Anwendungen und Dienste), die mit minimalem Verwaltungsaufwand schnell bereitgestellt und freigegeben werden können. Der Speicher kognitiver Systeme ist also nicht lokal vorhanden, sondern kann von zahllosen Systemen abgerufen werden. Das *Wissen* eines kognitiven Systems ist eben nicht auf eine Einheit beschränkt. Die gesammelten Informationen können nicht nur von einer einzelnen Entität beherbergt und repräsentiert werden, sondern in ein Netzwerk zahlreicher Schnittstellen eingespeist und durch sie ausgedrückt werden. Sie münden nicht in einer individuellen *Erfahrung* (Banavar 2016). Ein kognitives System führt zig Gespräche parallel, wertet Unmengen von Daten zahlloser Situationen aus und lässt diese in laufende Konversationen einfließen, während Menschen sich in jeweils nur einer Situation befinden. Selbst wenn diese

Gespräche parallel etwa über Smartphone oder *Skype*⁷ ablaufen, ist es nicht möglich, diese gleichzeitig und vollständig isoliert, also ohne gegenseitige Interferenzen, wahrzunehmen. Man bleibt in unserer subjektiven Empfindung verhaftet.

7.3 Bewusstsein

Das wesentliche Merkmal eines Bewusstseins ist zunächst die bloße Tatsache seiner Existenz. Zu wissen, dass der Mensch *ist*, ein individuelles Subjekt in einer Umwelt, welches diese beeinflusst und von ihr beeinflusst wird, ist gleichermassen eine grosse Sensation, wie auch Selbstverständlichkeit. Ganz unterschiedliche Aspekte wie Verantwortung oder ein Zugehörigkeitsgefühl wären ohne Kenntnis des eigenen Daseins wohl kaum vorstellbar. Warum scheinen dann aber nur wenige Menschen die Fähigkeit zu besitzen, grosse kreative Ideen zu entwickeln? Warum ist nicht jeder ein Genie und hat bahnbrechende Ideen? Der Psychologie Mihály Csíkszentmihályi (2007), der als erster das *Flow*-Erleben entdeckte und beschrieb, kommt zu dem Schluss, dass kreative Individuen auf psychologischer Ebene komplexer aufgestellt seien und ihr Bewusstsein differenzierter und integrierter agiere. Dies bedeute einerseits, dass ein solcher Mensch besser zwischen sensorischen Signalen und Gedanken differenzieren könne und also im höheren Masse die eigene Gefühlswelt von der gefühlten Umwelt unterscheide. Dies steigere die Neugierde und das Interesse, was wiederum zu einer höheren Auffassungsgabe und deutlich spürbareren Emotionen führe. Andererseits schaffe es ein kreatives Individuum laut Csíkszentmihályi (2007), die vielfältig wahrgenommenen Eindrücke untereinander zu vernetzen, also zu integrieren, um so neue Informationen und Gedanken zu generieren. Der im direkten Verhältnis zur gesamten Menschzahl eher bescheidene Anteil genial-kreativer Personen lässt erahnen, wie schwer zu erreichen diese Form der kreativen Komplexität selbst mit *echtem* Bewusstsein offenbar zu sein scheint (Giddings 1907), sodass selbst die Möglichkeit eines simulierten Bewusstseins bei kognitiven Systemen nicht ausreiche, um zu transformativen Leistungen befähigt zu sein.

7.4 Das Unbewusste

Ein Grossteil der Informationsverarbeitung im Gehirn geschieht automatisch, unbewusst, ohne, dass wir dazu aktiv etwas tun (Jaencke 2016). Viele Probleme löst man, ohne über sie nachzudenken. Nicht selten trifft einen ein unerhörter Einfall, nimmt eine neue Idee Gestalt an, wenn man etwas anderes tun. Der Effekt lässt sich experimentalpsychologisch klar belegen (Jaencke 2016). Englische Physiker empfahlen im 2. Weltkrieg die drei *B's*: *Bus*, *Bad*, *Bett*. Geniale Einsichten und Einfälle lassen sich nicht planen, sie passieren oft genau dann, wenn man sie nicht herbeizwingen sucht. Hirnforscher verorten die Quelle kreativer Gedanken etwa im

⁷<https://www.skype.com>.

Erfahrungsschatz des impliziten Wissens, welches all jene zufällig aufgenommenen und wieder vergessenen Informationen beinhaltet, die wir unbewusst mit uns herumtragen. Wer den Schatz des Unbewussten aktivieren und zu neuen Einsichten verknüpfen möchte, muss aufhören, zielgerichtet zu denken. Denn die Lösung liegt oft nicht dort, wo wir den Scheinwerfer des bewussten Denkens hinlenken. Zudem ist das bewusste Denken in seiner Kapazität beschränkt, kann nur wenige Dinge gleichzeitig verarbeiten (Jaencke 2016). Das Unbewusste dagegen ist Meister der Parallelverarbeitung, kann hunderttausende von kleinen Dingen gleichzeitig verarbeiten und miteinander in Beziehung setzen. Besonders kreativ ist man aber oft dann, wenn man das Gehirn einfach machen lässt (Jaencke 2016). Wie die Existenz eines Bewusstseins scheint auch das Unbewusste für kognitive Systeme in weiter Ferne zu liegen.

7.5 Sinnlichkeit

Wenn man einen Abend in einem mit höchsten Auszeichnungen versehenen Restaurant verbringt, geht es nicht primär darum, satt zu werden. Im Vordergrund steht eine sinnliche Erfahrung des Genusses. Man will überrascht werden, neue Geschmackserlebnisse austesten und Raffiniertes mit allen Sinnen erfahren. Obgleich ein Gericht objektiv betrachtet für jeden identisch scheint, fallen die Urteile hierüber unterschiedlich aus und werden jeweils individuell wahrgenommen und dementsprechend verschieden empfunden. Geschmäcker sind schliesslich verschieden. Auch Gedichte oder Romane können ganz unterschiedliche subjektive Zuschreibungen erfahren. Die eine denkt bei der Beschreibung einer Blumenwiese eventuell an ein zurückliegendes Picknick, der nächste nur an seinen Heuschnupfen. Ebenso verhält es sich mit der Schöpfung kreativer Ideen: Ohne sinnlichen Zugang, ohne das Gefühl, scheint eine transformative Kreativität undenkbar. Und ohne die übrigen Charakteristika wiederum erscheint die Sinnlichkeit nicht erleb- und nachvollziehbar, weswegen sie für kognitive Systeme ebenfalls nicht erreichbar scheint.

8 Ausblick: Die Wiederverzauberung der Welt

Wenn ausschliesslich über Sprache ausgedrückt wird, was eben nur über Sinnlichkeit zur Erkenntnis führen kann, werden jene Aspekte, die den Menschen von kognitiven Systemen noch immer unterscheiden, *vom Segen zum Fluch*. Körperlichkeit, Individualität, Bewusstsein, das Unbewusste und schliesslich Sinnlichkeit werden in Zukunft nur weiterhin als Distinktionsmerkmal nutzen, wenn man sie kultiviert und lernt, sich wieder neu zu wundern, zu staunen und irritieren zu lassen. Tut man dies nicht, verbleibt das Sinnliche als dauerhaft Unbestimmtes auf ewig im Unbekannten, und das Unbekannte verbleibt eine Gefahr, die durch ihren Mangel an Verständnis für den Einzelnen zur anfangs erwähnten Angst und Dystopie führen kann. Die Frage, *was* (künstliche) Kreativität sei, ist letzten Endes nicht zielführend; ebenso wenig wie der daraus resultierende Versuch, Kreativität skalierbar zu

machen. Es geht weniger um das *als*, denn um eine bestimmte Dimension des Empfindens. Dieses Empfinden akzeptiert den Umstand, dass wir unmöglich sagen können, A sei kreativer als B. Vielmehr verhalten sich A und B inkommensurabel zueinander, sind also nicht vergleichbar. Zentral ist die Wirkung, die A oder B haben können, unabhängig davon, was sie letzten Endes sind. Die Relevanz künstlicher Kreativität liegt nicht zuletzt in ihrem Vermögen, die eigene Kreativität anzu-*stossen*. Konkret bedeutet dies, dass es bei der Frage nach künstlicher Kreativität vermehrt um ein *als ob* gehen sollte, um eine Fiktionalisierung also (Mersch 2015). Lässt man es zu, sich von Ergebnissen kognitiver Systeme beeindrucken und verzaubern zu lassen, ebenso *als ob* diese *wirklich* (also im Sinne des vorher dargestellten Verständnisses) kreativ seien, erübrigt sich die Frage, ob dies stimmt oder nicht. Relevant ist der Eindruck auf das subjektive Empfinden und das damit ausgelöste Verhalten, nicht die objektive Einordnung. Diese zunächst kontrovers anmutende Denkweise hat ihren Ursprung in der Diskussion um die Frage, ob ein kognitives System jemals ein Bewusstsein erlangen wird. Für den Philosophen John Searle führt eben diese Frage am Kern der Lösung vorbei, da sie nicht den potentiellen Nutzen für den Menschen in den Mittelpunkt stellt, sondern das kognitive System. Nach Searle ist es vielmehr zielführend zu hinterfragen, wie der Effekt eines künstlichen Bewusstseins hervorgerufen werden kann. Ob dieses dann real ist, oder nur so tut, *als ob*, ist für die menschliche Reaktion weniger entscheidend – ganz so, wie es vorher für die Frage nach künstlicher Kreativität formuliert wurde. Um dies zu veranschaulichen stellte Searle im Jahr 1980 das Gedankenexperiment des *Chinesischen Zimmers* vor. Das Experiment ist etwas verkürzt folgendermassen vorstellbar: In einem Raum sitzt ein Mensch, vor der Tür steht ein weiterer. Der Mensch im Raum versteht keinerlei chinesische Schriftzeichen, der Mensch ausserhalb des Raumes aber schon. Beide kommunizieren über Zettel, die sie unter der Tür hindurchschieben. Zwar versteht der Mensch im Raum nichts des Geschriebenen, er bekommt aber Anleitungen in seiner Muttersprache, wie er die erhaltenen Texte sinnvoll beantworten kann, die er dann unter der Tür hindurchschiebt. Für den anderen Menschen ausserhalb des Raumes macht es dann keinen Unterschied, ob derjenige im Raum tatsächlich chinesisch versteht, oder dies nur *simuliert*. Ebenso kann es sich mit künstlicher Kreativität verhalten. Wenn etwas wirkt, *als ob* es kreativ sei, dann sind die Effekte dieselben, gleichgültig, ob es sich um tatsächliche oder eben fiktionalisierte Kreativität handelt. So gesehen war der vorher vorgestellte Definitionsansatz der *Association for Computational Creativity*, nachdem künstliche Kreativität sei, was auch beim Menschen als kreativ angesehen würde, bereits sehr weitsichtig.

Die Forschung zeigt dann auch, dass kognitive Systeme Werke reproduzieren und simulieren, die wirken, als ob ein Mensch in kombinatorischer Weise kreativ tätig wäre (Müller et al. 2018; D’Onofrio et al. 2018, 2019), künstliche Kreativität also zumindest auf dieser ersten Ebene existiert. Mehr (noch) nicht, aber eben auch nicht weniger. Durch unsere Fähigkeit, uns vom Neuen und Unerwarteten begeistern zu lassen, können wir einerseits diesen erstaunlichen Leistungen kognitiver Systeme Rechnung tragen, andererseits aber die künstliche Kreativität als Potential für die eigene Kreativität nutzen. Hier liegt die Grenze des Verstehens künstlicher Kre-

ativität: in der Erkenntnis, dass auch das Greifbare und numerisch oder sprachlich Vergleichbare sich nicht in sich selbst und seiner Vergleichbarkeit erschöpft, sondern über die eigene Faszination das eigene Potential transformativer Kreativität beinhaltet. Dieses Potential kann nur dann aktiviert werden, wenn man akzeptiert, es nicht verstehen zu können, sondern vor allem zu empfinden, dem damit einhergehenden Erstaunen seinen Platz einzugestehen und der Welt und der menschlichen Rolle hierin einen final nicht verstehbaren Zauber zuzusprechen (Weichert 2016).

Gewissermassen muss man sich selber austricksen, um das Erstaunen überhaupt zuzulassen. Gewohnte Denkmuster zu verlassen (Tschudi und D’Onofrio 2018) und zu akzeptieren, (künstliche wie menschliche) Kreativität nicht im versprachlichten, sondern empfindsamen Sinne verstehen zu können, ist zweifelsohne eine grosse und anstrengende Herausforderung, die es erst einmal auszuhalten gilt. Gerade weil kognitive Systeme den Menschen in vielen Aspekten jenseits der vorher genannten bereits jetzt überlegen sind und etwa immens grosse Datenmengen in wesentlich kürzerer Zeit prozessieren können als Menschen, eröffnet sich die nur scheinbar paradoxe Chance und auch Notwendigkeit, inmitten einer zunehmend technisierten aber auch entzauberten Welt mithilfe des augmentativen Einsatzes kognitiver Systeme eine der ureigensten Fähigkeiten des Menschen freizulegen und wieder mehr zu kultivieren: *den empfindsamen Umgang mit der eigenen Kreativität*. Nur Menschen sind und bleiben aller Voraussicht nach in der Lage, wirklich Neues auf einer transformativen Ebene zu erschaffen. Das Cognitive Computing bietet uns die Chance, den Voraussetzungen hierfür ein ganzes Stück näherzukommen. Diese Chance sollte genutzt werden.

Literatur

- Banavar G (2016) Learning to Trust Artificial Intelligence Systems in the Age of Smart Machines. IBM Research.
- Bauer V (2016) Google Magenta-Team veröffentlicht erstes KI-komponiertes Musikstück. mobilegeeks. <https://www.mobilegeeks.de/news/google-magenta-team-veroeffentlicht-erstes-ki-komponiertes-musikstueck/>. Erstellt: 04.06.2016. Letzter Zugriff: 17.06.2018.
- Beckermann A (1999) Leib-Seele-Problem. In: Enzyklopädie der Philosophie 1:766–774.
- Boden MA (2004) The Creative Mind: Myths and Mechanisms. Routledge, London.
- Boden MA (2011) Creativity and Art: Three Roads to Surprise. Oxford University Press, Oxford.
- Brandom R (2016) Google’s art machine just wrote its first song. The Verge. <https://www.the-verge.com/2016/6/1/11829678/google-magenta-melody-art-generative-artificial-intelligence>. Erstellt: 01.06.2016. Letzter Zugriff: 22.11.2018.
- Cohen H (1995) The Further Exploits of AARON, Painter. Stanford Humanities Review, 4(2):141–158.
- Colvin G (2015) Humans are Underrated. What High Achievers Know That Brilliant Machines Never Will. Penguin, New York.
- Csikszentmihályi M (2007) Kreativität. Klett-Cotta, Stuttgart.
- Damasio AR (1994) Descartes’ Error: Emotion, Rationality and the Human Brain. Putnam, New York.
- Deschamps P (2015) Noma: My Perfect Storm Clip – Mozart. YouTube. <https://www.youtube.com/watch?v=9sQbmY7qQhA/>. Erstellt: 08.12.2015. Letzter Zugriff: 20.07.2017.

- D'Onofrio S, Müller SM, Papageorgiou E, Portmann E (2018) Fuzzy Reasoning in Cognitive Cities – An Exploratory Work on Fuzzy Analogical Reasoning Using Fuzzy Cognitive Maps. 2018IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'18).
- D'Onofrio S, Müller SM, Portmann E (2019) A Fuzzy Reasoning Process for Conversational Agents in Cognitive Cities. Hammoudi S, Smialek M, Camp O, Filipe J (Hrsg.) Enterprise Information Systems. ICEIS 2018. Lecture Notes in Business Information Processing, vol. 363. Springer, Cham.
- D'Onofrio S, Portmann E (2015) Von Fuzzy-Sets zu Computing-with-Words. Informatik-Spektrum, 38(6):543–549.
- Dravecky M (2016) Dinner at Noma. YouTube. <https://www.youtube.com/watch?v=oYcT2oiQ-MeE/>. Erstellt: 25.12.2016. Letzter Zugriff: 20.07.2017.
- Eisenhart A (2016) Chef Rene Redzepi Finds Incredible Flavors in the Most Unexpected Places. YouTube. https://www.youtube.com/watch?v=mNz-hWPF_dY/. Erstellt: 18.11.2016. Letzter Zugriff: 20.07.2017.
- Elgammal A, Saleh B (2015) Quantifying Creativity in Art Networks. The 6th International Conference on Computational Creativity (ICCC'15).
- FAZ (2018) Große Skepsis gegenüber Künstlicher Intelligenz. Frankfurter Allgemeine Zeitung. <http://www.faz.net/aktuell/wirtschaft/diginomics/grosse-skepsis-gegenueber-kuenstlicher-intelligenz-15781806.html> Erstellt: 11.09.2018. Letzter Zugriff: 11.09.2018.
- Florida R (2002) The Rise of the Creative Class. Basic Books, New York.
- Frankfurt HG (2016) Gründe der Liebe. Suhrkamp, Frankfurt am Main.
- Gick ML, Holyoak KJ (1980) Analogical Problem Solving. Cognitive Psychology, 12:306–355.
- Giddings FH (1907) Elements of Sociology. Macmillan, New York.
- Goertzel B (2018) AI can help create a better world – if we build it right. SingularityHub. <https://singularityhub.com/2018/08/13/ai-can-help-create-a-better-world-if-we-build-it-right/> Erstellt: 13.08.2018. Letzter Zugriff: 15.08.2018.
- Guilford JP (1950) Creativity. American Psychologist, 5:444–454.
- Harlan V (2001) Was ist Kunst? Werkstattgespräch mit Joseph Beuys. Urachhaus, Stuttgart.
- Hawkins J, Blakeslee S (2007) On Intelligence: How a New Understanding of the Brain Will Lead to the Creation of Truly Intelligent Machines. Macmillan, New York.
- Heidegger M (1983) Einführung in die Metaphysik. Klostermann V (Hrsg.), Gesamtausgabe, Band 40.
- Hoffmann B (2004) Strahl und Strom – Wanders Sturmlied als dramatisierte Reflexion von Subjektivität und künstlerischer Kreativität. Deutsche Vierteljahresschrift für Literaturwissenschaft und Geistesgeschichte, 78(2):229–260.
- Jaencke L (2016) Ist das Hirn vernünftig?. Hogrefe, Bern.
- Jordanous A (2012) A Standardised Procedure for Evaluating Creative Systems: Computational Creativity Evaluation Based on What it is to be Creative. Cognitive Computation, 4(3):246–279.
- Jordanous A (2014) What is Computational Creativity?. The Creativity Post. http://www.creativitypost.com/science/what_is_computational_creativity/. Erstellt: 10.04.2014. Letzter Zugriff: 04.03.2017.
- Kelly III JE, Hamm S (2013) Smart Machines: IBM's Watson and the Era of Cognitive Computing. Columbia University Press, New York.
- Leach J (2007) Creativity, Subjectivity and the Dynamic of Possessive Individualism. Hallam E, (Hrsg.), Creativity and Cultural Improvisation.
- Libet B, Gleason CA, Wright EW, Pearl DK (1983) Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-potential) the Unconscious Initiation of a Freely Voluntary Act. Brain, 106(3):623–642.
- Lombardo BJ, Roddy DJ (2010) Cultivating Organizational Creativity in an Age of Complexity. IBM Institute of Business Value.
- Medlock B (2017) The Body is the Missing Link for Truly Intelligent Machines. aeon. <https://aeon.co/ideas/the-body-is-the-missing-link-for-truly-intelligent-machines>. Erstellt: 14.03.2017. Letzter Zugriff: 01.04.2017.
- Mersch D (2015) Epistemologien des Ästhetischen. diaphanes, Zürich und Berlin.

- Montuori A, Purser R (2006) Search of Creativity: Beyond Individualism and Collectivism. Western Academy of Management Conference, Hawaii.
- Müller SM, D'Onofrio S, Portmann, E (2018) Fuzzy Analogical Reasoning in Cognitive Cities: A Conceptual Framework for Urban Dialogue Systems. Proceedings of the 20th International Conference on Enterprise Information Systems (ICEIS'18).
- Pearce M (2010) Boden and Beyond: The Creative Mind and its Reception in the Academic Community. London: City University.
- Pfeifer R, Bongard J (2007) How the Body Shapes the Way We Think. A New View of Intelligence. The MIT Press, Cambridge.
- Pratt L, Thrun S (1997) Guest Editors' Introduction. Machine Learning, 28(1):5–5.
- Ram A, Wills L, Domeshek E, Nersessian N, Kolodner J (1995) Understanding the Creative Mind: a Review of Margaret Boden's Creative Mind. Artificial Intelligence, 79:111–128.
- Reckwitz A (2012) Die Erfindung der Kreativität. Suhrkamp, Frankfurt am Main.
- Ruppert W (1998) Der moderne Künstler.: Zur Sozial-und Kulturgeschichte der kreativen Individualität in der kulturellen Moderne im 19. und frühen 20. Jahrhundert. Suhrkamp, Frankfurt am Main.
- Sawyer R (2011) Explaining Creativity: The Science of Human Innovation. Oxford University Press, Oxford.
- Sebastian A (2014) The paradigm shift in computing. [Online]. https://www.zurich.ibm.com/pdf/sto/Y2014_TechnologyFuture.pdf. Erstellt: n.d. Letzter Zugriff: 26.11.2018.
- Schmidt J (1985) Die Geschichte des Genie-Gedankens in der deutschen Literatur, Philosophie und Politik 1750–1945. Band 1. Wissenschaftliche Buchgesellschaft, Darmstadt.
- Searle JR (1980) Minds, brains, and programs. Behavioral and brain sciences, 3(3):417–424.
- Stachelhaus H, Foglia B (1991) Joseph Beuys. Abbeville Press, New York.
- Trillas E, D'Onofrio S, Portmann, E (2019). An Exploration of Creative Reasoning. Portmann E, Seising R, Tabacchi M, Habenstein A (Hrsg.), Designing Cognitive Cities. Studies in Systems, Decision and Control, Volume 176. Springer, Cham.
- Tschudi F, D'Onofrio S (2018) Wie wir durch unsere Denk- und Handlungsmuster beeinflusst werden. Meier A, Seising R (Hrsg), Vague Information Processing, 567–471.
- van den Berg K (2009) Kreativität. Drei Absagen der Kunst an ihren erweiterten Begriff. Jansen SA, Schröter E, Stehr N (Hrsg.), Rationalität der Kreativität? 207–224.
- Veale T, Gervás P, Pérez y Pérez R (2010) Computational Creativity: A Continuing Journey. Minds & Machines, 20:483–487.
- Voon B (2017) Humans Prefer Computer-Generated Paintings to Those at Art Basel. Hyperallergic. <https://hyperallergic.com/391059/humans-prefer-computer-generated-paintings-to-those-at-art-basel/>. Erstellt: 31.07.2017. Letzter Zugriff: 05.07.2018.
- Weichert N (2016) Die Wiederverzauberung der Welt. Transform Magazin. <https://www.transform-magazin.de/soziologie-wiederverzauberung-aufklaerung-utopie/>. Erstellt: 13.10.2016. Letzter Zugriff: 24.11.2018.
- Wittgenstein L (1984) Werkausgabe, Band 1: Tractatus logico-philosophicus/Tagebücher 1914–1916/Philosophische Untersuchungen. Suhrkamp, Frankfurt am Main.

Maschinen zur Lösung der grossen Probleme – Ein formeller Rahmen

Philip Gross

Zusammenfassung

Einige der bedeutendsten Probleme dieser Welt müssen jetzt gelöst werden. In diesem Artikel wird dargelegt, was von den Maschinen – also von modernen Technologien – erwartet wird, wenn sie einen Beitrag zur Bewältigung grosser und drängender Probleme leisten sollen. Der vorgeschlagene formelle Rahmen für die Herleitung dieser Erwartungen baut darauf auf, dass das individuelle wahrgenommene Risiko einer Wissenschaftlerin, eines Ingenieurs oder jeder anderen potenziellen Mitwirkenden eine kollektive Problemlösung verhindern und die Lösungsfindung verschleppen kann. Es wird aufgezeigt, wie dieses Risiko mit der strukturellen Güte von Problemen zusammenhängt und es wird vorgeschlagen, je ein System zur Bewältigung der beiden grundlegenden Problemstruktur-Typen aufzubauen. Die Bewältigung von „ill-structured problems“ soll dabei über ein *Creativity System* erfolgen, während „well-structured problems“ über ein *Productivity System* abgewickelt werden sollen.

Dieser Artikel stellt eine Erweiterung eines bereits in der Zeitschrift Informatik Spektrum erschienenen Artikels dar (Gross 2018). Während im Artikel das Augenmerk auf das Creativity System gelegt wurde, findet im vorliegenden Artikel auch eine Auseinandersetzung mit der Frage statt, wie ein Productivity System im Rahmen der kollektiven Problemlösung zur Reduktion des von potenziellen Mitwirkenden wahrgenommenen individuellen Risikos beiträgt und somit deren Partizipation fördert. Es werden Leistungsanforderungen an ein solches System formuliert und in Ansätzen wird diskutiert, was das System zur Erfüllung dieser Anforderungen leisten muss. Die entsprechenden Ausführungen zum

Überarbeiteter Beitrag basierend auf Gross, P.: Zur Lösung der wirklich bedeutenden Probleme: Was von Maschinen erwartet wird, Informatik Spektrum (2018) 41(1): 38–51

P. Gross (✉)

Archai – Open Science for an Enjoyable Planet Earth, Holzhäusern, Schweiz

E-Mail: philip.gross@archai.xyz

Creativity System wurden praktisch unverändert aus dem Informatik-Spektrum-Artikel übernommen.

1 Grosse Probleme – Grosse Risiken

Wer sich im Forschungs- und Entwicklungsbetrieb an die grossen Probleme heranwagt, geht grosse Risiken ein. Damit sind ganz elementare und oft existenzielle Risiken gemeint. Risiken nämlich, die ein Molekularbiologe wie Robert Bywater beispielsweise dann eingeht, wenn er zusammen mit dem Musikwissenschaftler Jonathan Middleton Melodien zur Erforschung von Proteinfaltungsmechanismen heranzieht (Bywater und Middleton 2016).

In diesem Beispiel versuchen die beiden Forscher, ein komplexes und lange ungelöstes Problem auf einem sehr unkonventionellen Weg zu lösen. Damit setzen sie sich zwei Risikodeterminanten aus, die jedes Forschungsvorhaben zu einem Wagnis machen. Um ein molekularbiologisches Problem zu lösen ist Bywater bereit, den Suchraum für einen Lösungsansatz weit über die Grenzen seines angestammten Fachgebiets, der Molekularbiologie nämlich, hinaus zu dehnen. Die *Wahrscheinlichkeit eines Lösungserfolgs* wird sich damit für ihn *deutlich verringern*. Obwohl es für ihn viel naheliegender wäre, einen Ansatz in seiner naturwissenschaftlichen Stammdomäne zu suchen und damit zumindest ein Teilproblem zu lösen, lässt er sich auf das musikalische Abenteuer mit ungewissem Ausgang ein.

Zur geringen Erfolgswahrscheinlichkeit kommt ein *kaum vorhersehbares Mass an erforderlichen Ressourcen* hinzu. Der Aufwand für die Erarbeitung publikationswürdiger Erkenntnisse kann in einer so unkonventionellen Kooperation nicht antizipiert werden. Es ist aber zu vermuten, dass er höher ausfällt als der Aufwand, den Bywater auf dem konventionellen Weg betreiben müsste; insbesondere, wenn er sich mit einem inkrementellen Erkenntnisgewinn begnügen würde. Dieser Zusatzaufwand stellt in der individuellen Risikoevaluation einen negativen Nutzen dar. Im Zusammenwirken mit der verringerten Erfolgswahrscheinlichkeit determiniert dieser negative Nutzen das grosse Risiko, welches Bywater eingeht (und mit umgekehrten Vorzeichen natürlich auch Middleton).

Das hier zur Formulierung eines formellen Rahmens zum geforderten Beitrag von Maschinen zur Lösung grosser Probleme herangezogene Verständnis von Risiko – insbesondere von *individuellem wahrgenommenem Risiko* – soll nicht auf eine ökonomische Kosten-/Nutzen-Abwägung beschränkt sein. Insbesondere die Risiko-Komponente *Ressourceneinsatz* (siehe Abschn. 1.3) meint neben der Unvorhersehbarkeit eines zeitlichen oder monetären Aufwandes auch das kaum absehbare Ausmass der Einbringung der eigenen Glaubwürdigkeit und der hart erarbeiteten Anerkennung in relevanten Peer-Gruppen. Ressourceneinsatz kann hier auch heissen, wie viel persönliches *Machtkapital* jemand einzusetzen bereit ist. Es ist zu vermuten, dass insbesondere im akademischen Umfeld solche nicht-ökonomischen Parameter der individuellen Risikowahrnehmung oft schwerer wiegen als die rein monetären. Menschliche Grundmotive wie der Wunsch nach Zugehörigkeit (Leary

und Baumeister 2017) und der Drang nach Erfolg und Leistung (McClelland et al. 1953) sind die Treiber einer solchen Ressourcen-Evaluation.

1.1 Maschinen für mehr Kreativität

Im Unterschied zu einem Problem, für welches der Lösungsweg den Regeln deduktiv-inferentiellen Denkens folgt, sind komplexe Probleme der realen Welt oft Probleme, die sich nicht durch regelbasierte, deterministische Lösungsmechanismen bewältigen lassen. Sie verlangen nach kreativen Ansätzen. Anders als die Deduktion beruht kreatives Problemlösen nicht nur auf logischem Schlussfolgern, sondern ebenso sehr auf Analogieschlüssen (Boden 2004) und auf Abduktion (Thagard und Shelley 1997). Es sind nicht die mechanistischen *Rules of Inference*, die zu neuen Ideen führen, sondern das Zusammenbringen vormals beziehungsloser Konzepte im Sinne von Arthur Koestler's *Bisociations* (Koestler 1964). Die kombinatorische Natur des kreativen Denkens meint, ein Problem auf sein strukturelles Fundament zu reduzieren, um dann ausschliesslich von dieser Abstraktion ausgehend nach neuen Lösungsmustern zu suchen. Diese Muster können demzufolge aus gänzlich unverwandten Wissensgebieten stammen.

Der vorliegende Artikel soll Anregungen geben, wie Denkmuster dieser nicht logisch-schlussfolgernden Art durch Systeme und Technologien unterstützt werden können: Wie hätten Melodien ihren Weg in die Molekularbiologie finden können, wenn sich Bywater und Middleton nicht zufällig getroffen hätten? Oder, etwas allgemeiner gefasst: Wie können die Problemlösungsanstrengungen einer einzelnen Wissenschaftlerin, eines Ingenieurs, eines Entwicklers oder vielleicht einer Designerin technologisch befruchtet werden mit potenten Lösungsmustern die zwar passend, deren Wissens-Domänen diesen Akteuren aber unbekannt sind? Wie würden also Maschinen den Zugang zu diesen Lösungsmustern unterstützen, wenn eine Problemlöserin deren Träger nicht persönlich kennt?

Der Grundgedanke ist dieser: Maschinen sollen die kognitiven Prozesse zur Lösung eines Problems aus der Beengtheit eines einzelnen Gehirns befreien und ein *Denken über mehrere Gehirne hinweg* ermöglichen. Die Synchronizität mehrerer neuronaler Aktivitätsmuster, die jedes für sich genommen vormals unabhängige mentale Konzepte reflektieren, kann deren Verknüpfung bewirken, wenn zeitlich derart ineinanderfliessende neuronale Aktivitäten wiederholt gemeinsam ausgelöst werden. Solche neuen Verknüpfungen können einer einzelnen Problemlöserin kreative Einsichten bringen (Jung-Beeman et al. 2004; Thagard und Stewart 2011). Mit geeigneten Technologien könnte die Kreativität noch gesteigert werden, wenn Maschinen es ermöglichen, Lösungskonzepte *aus den Gehirnen vieler* gleichzeitig zu aktivieren und ineinander zu verweben. Die Maschinen sollen also ein *Network of Neural Networks* bilden, um damit die Bewältigung grosser Probleme zu unterstützen, wenn diese nach Inspiration aus unvermuteten Quellen verlangt.

Um die leistungsfähigsten Gehirne in einem solchen Netzwerk zu vereinen und um deren kollektive Intelligenz zur Bewältigung der grossen Probleme abzuschöpfen ist es erfolgsentscheidend, dass das individuelle Risiko für jede potenzielle

Akteurin so weit wie möglich gesenkt wird. Nur dann wird sie sich an einem massiv kollaborativen Effort mit so unkonventionellem Charakter beteiligen.

Das *individuelle wahrgenommene Risiko* ist ein Mass, das auf einer *subjektiven Einschätzung zu Auswirkung und Eintretenswahrscheinlichkeit eines Ereignisses* beruht (Sjöberg 2000). Dieses Risiko bildet das Fundament für den formellen Rahmen, der zur Formulierung von Anforderung an problemlösungs-unterstützende Systeme und Technologien dienen wird. Ohne Risikokontrolle lassen sich nicht genügend Knoten zum Aufbau eines Network of Neural Networks (siehe Abschn. 6.1) mobilisieren und die Problemlösungsmaschine bleibt auf singuläre biologisch-neuronale Netzwerke reduziert.

1.2 Geringe Wahrscheinlichkeit eines Lösungserfolgs

Wenn das Aufeinandertreffen eines Molekularbiologen mit einem Musikwissenschaftler auf dem Zufall beruht und wenn das Auffinden ganz neuer und bedeutender Ideen nicht-lineare, vielleicht sogar unwillkürliche kognitive Prozesse fordert, dann ist die Beute derart kreativen Denkens zwangsläufig eine magere. Der Erfolg tritt schlicht nicht oft genug ein. Anhand des modellhaften *Radiation Problem* (Dunker 1945) soll aufgezeigt werden, warum das so ist.

Das Radiation Problem stellt sich beispielsweise einer Ärztin, die einen im Körper befindlichen Tumor mit einer bestimmten minimalen Strahlungsintensität unschädlich machen soll. Sie weiss aber, dass die geforderte Intensität mit einer Zerstörung des umliegenden Gewebes einhergeht. Einfach drauflos bestrahlen geht also nicht! Zur Lösung des Problems könnte die Ärztin aber versuchen, die Strahlung so durch den Körper zum Tumor zu schicken, dass möglichst wenig umliegendes Gewebe Schaden nimmt; indem Sie vielleicht den kürzesten Weg zum Tumorzentrum findet und in dieser Richtung bestrahlt. Gleichwohl würde gesundem Gewebe Schaden zugefügt. Sie könnte die Strahlungsintensität auch auf ein für gesundes Gewebe harmloses Mass reduzieren und dafür länger bestrahlen. Aber auch das funktioniert nicht, weil die minimal geforderte Intensität zu keinem Zeitpunkt erreicht wird. Sie merkt also: die rein domänenspezifische Lösungssuche, die medizinische also, bringt sie nicht weiter.

Wenn die Ärztin das Problem nun aber abstrahiert und es auf die wesentlichen konzeptionellen Aspekte reduziert, dann wird sie nicht mehr nur nach einer medizinischen Lösung suchen. Die Abstraktion erweitert ihren potenziellen Lösungsraum und sie ermöglicht, Lösungsansätze zu erkennen, die von strukturell isomorphen Problemen und nicht von medizinischem Wissen abgeleitet sind. *Strukturell isomorph* bedeutet hier, dass sich die Abstraktion des Problems auf die Abstraktion des Lösungsansatzes projizieren lässt (Sowa 1976). Ohne auf die technischen Erfordernisse für eine isomorphe Projektion (bei Konzept-Graphen) im Detail einzugehen ist leicht einzusehen, dass ein valider Lösungsansatz auch von einem nicht-medizinischen Konzept kommen könnte, wenn dieses jede Einzel-Abstraktion des ursprünglichen Problems (wie z. B. Tumor, unschädlich machen, Strahlungs-dosis)

und deren strukturelle Relationen erfasst und wenn die abstrahierten Konzept-Bausteine von medizinischem und nicht-medizinischem Problem eindeutig korrespondieren. So kann auch eine mittelalterliche Kriegstaktik zum Problemlösungsansatz werden: Eine Festung (der Tumor) kann eingenommen (unschädlich gemacht) und die notwendige militärische Kraft (geforderte minimale Strahlungs-dosis) auf den Zielort konzentriert werden, indem mehrere kleine Heereseinheiten (reduzierte Intensitäten) gleichzeitig von verschiedenen Seiten angreifen (bestrahlt wird) (Duncker 1945).

Indem die Ärztin eine zur Zielstruktur passende isomorphe Abbildung in einer unvermuteten Quelldomäne findet, erschliesst sie sich eine neue und medizinisch wirksame Lösung aus einem fachfremden Kontext. Ausserhalb des Geltungs-bereichs von Modellproblemen ist es nun aber eher unwahrscheinlich, auf eine isomorphe Abbildung einer gegebenen Problemstruktur in einer nicht verwandten Wissensdomäne zu treffen, um daraus dann eine kreative Lösung zu konstruieren. Das hat mehrere Gründe. Der wichtigste davon ist ein probabilistischer. Je weiter entfernt die Quellstruktur von der Zielstruktur ist, desto kleiner ist die Wahrscheinlichkeit, dass diese ganz unerwartete Verbindung von einer Problemlöserin hergestellt und im Sinne einer Inspiration genutzt wird. In einer grossen Menge von teilweise völlig unverwandten Wissensgebieten gibt es zu viele potenzielle Quellstrukturen, als dass man eine realistische Chance hätte, eine isomorph passende zu finden. Ohne Such-Einschränkungen, die über die strukturell limitierende Anforderung eines Isomorphismus hinausgehen, ist der Suchraum schlicht zu gross (Holyoak und Thagard 1989). Unglücklicherweise sind es aber oft die weit entfernten Ansätze, diejenigen, die ausser struktureller Ähnlichkeit keinen offensichtlichen Bezug haben zur Problemstellung, von denen die wirklich guten Ideen herrühren (Mednick 1962). Genau hier könnten Maschinen helfen. Es gibt heute zwar noch keinen praktisch funktionsfähigen Algorithmus, der verborgene Strukturbeziehungen zwischen zwei Konzepten erkennt. Aber die Experten sind zuversichtlicher als auch schon, dass es einen theoretisch effizienten (zumindest quasi-polynomialen) Algorithmus gibt, der einen Isomorphismus zwischen zwei vordergründig unterschiedlichen Konzept-Graphen feststellen kann, sollte denn einer vorliegen (Babai 2017; Klarreich 2017a, b).

Das ist ein Anfang – allerdings ein bescheidener. Solange es keinen wirklich tauglichen Algorithmus gibt, der einem im Stile von Amazon die „*You may also be interested in this*“-Idee zur Inspiration vorschlägt, bleibt die einzelne Akteurin mit beschränkten Ressourcen auf den Zufall angewiesen, will sie im weiten Feld der potenziellen Quellstrukturen auf eine gute Idee für die Lösung ihres Problems stossen. Diese Unabdingbarkeit des Zufalls birgt ein enorm hohes individuelles Risiko.

1.3 Unvorhersehbar hoher Ressourceneinsatz

Unter gewissen Umständen liesse sich dieses weite Feld auch einfach absuchen. Insbesondere dann, wenn Such-Restriktionen den Lösungsraum genügend stark

eingrenzen. Das Softwareunternehmen Autodesk¹ bietet beispielsweise Tools an, welche Fahrgestelle für Automobile selbständig designen, wenn ihnen einige wenige Parameter vorgegeben werden. Etwas vereinfacht ausgedrückt erfolgt die Anweisung für das Tool etwa wie folgt: „*Design mir ein Fahrgestell für ein 4-plätziges Fahrzeug, welches weite Fahrten bei hohem Tempo zulässt und bei dem Aerodynamik wichtiger ist als Ästhetik*“ und die Software berechnet einen Design-Vorschlag (Conti 2017). An und für sich ist das eine Problemstellung, die sich wegen der enorm grossen Zahl von Lösungsvarianten auch mit massiver Rechenkapazität kaum lösen lässt. Werden die Systemgrenzen aber weit genug eingeschränkt, weiter als es die formulierten Bedingungen vermuten lassen, zum Beispiel indem der Software Grundkenntnisse zu Möglichkeiten und Grenzen materialwissenschaftlicher oder aerodynamischer Natur als Begleitrestriktionen beigebracht werden, reduziert sich die Lösungssuche auf eine Optimierung. Diese hat zwar weiterhin einen hohen *Branching Factor* – das heisst, jeder festzulegende Faktor lässt noch immer eine Vielzahl möglicher Ausprägungen zu – aber der Prozess der Optimierung ist endlich. In einem komplett offenen Problemlösungskontext hingegen fehlen solche kollateralen Such-Restriktionen. Oft fehlen nicht nur grundlegende Systemgrenzen, die sich zum Beispiel aus der Physik ergeben. In vielen Fällen ist sogar der Zielzustand unklar.

Ohne derartige Restriktionen könnte nun rohe Rechengewalt bei der Suche in einem weiten Feld helfen. Dagegen sprechen allerdings zwei Dinge. Erstens wird der Rechenaufwand rasch so hoch, dass ihn kein Computer mehr bewältigen kann. Zweitens, und wichtiger noch, ist diese Form der Lösungssuche nicht praktikabel – selbst wenn es diesen Hochleistungsrechner gäbe. Denn eine notwendige Voraussetzung für das Gelingen ist die *komplette Abbildung der Welt* in abstrahierten Konzept-Graphen und das *Ausprogrammieren aller natürlicher Ursache/Wirkungs-Beziehungen*. Das geht für das Designen von Fahrgestellen innerhalb eingrenzbarer fachlicher Domänen gerade so. Das scheitert aber im Fall der Suche nach kreativen und nicht bloss deterministisch optimierten Lösungen auf jeden Fall. Schliesslich soll ja für die Lösung des Radiation Problem eben gerade nicht an den Grenzen der Medizin Halt gemacht werden. Vielmehr soll die ganze Welt als Suchraum erschlossen sein und das umfasst auch den Teil der Welt, zu dem eben nicht vorher schon Informationen gespeichert und Funktionsregeln programmiert wurden.

Je weiter der Suchradius also geöffnet wird, desto grösser ist die Ungewissheit über den notwendigen Aufwand für die Lösungsfindung. Im Unterschied zu einer Datenbankabfrage mit vorhersehbarer Verarbeitungszeit würde man ja selbst in einem auf die Medizin eingeschränkten Szenario im Radiation Problem die nicht naheliegende Theorie oder die musterhaft ähnliche und doch fremde Struktur von ausserhalb der Radiologie suchen wollen. Der Ressourcenbedarf kann so leicht sehr hoch werden und es ist dieses potenzielle Ausmass, welches (neben der geringen Wahrscheinlichkeit eines Lösungserfolgs) die zweite relevante Determinante des individuellen wahrgenommenen Risikos bildet.

¹ [autodesk.com](https://www.autodesk.com).

2 Ungesteuerte Risikominderung verschleppt die Problemlösung

In vielen Fällen ist das wahrgenommene Risiko dann so gross, dass es den Risikoappetit des Einzelnen übersteigt. Wenn der Molekularbiologe und der Musikwissenschaftler nicht die Möglichkeit und die Bereitschaft gehabt hätten, ungewöhnliche Wege zu gehen und grosse Unsicherheiten auf sich zu nehmen, dann wäre heute vielleicht nicht bekannt, dass das menschliche Musikgehör Mutationen und andere Unregelmässigkeiten in den Faltpatterns von Proteinen aufdecken kann (Bywater und Middleton 2016).

Der menschliche Drang zur Risikominderung hätte dazu geführt, dass das Problem nie angegangen oder die Problemlösung zumindest stark verschleppt worden wäre. Ist die Erfolgswahrscheinlichkeit nämlich gering, wird sich jede Wissenschaftlerin, jeder Ingenieur oder jeder Entwickler und jede Designerin vielversprechenderen Problemen zuwenden. Soll der Ressourceneinsatz unter Kontrolle gehalten werden, wird das Problem vielleicht angegangen, aber die Lösungsfindung bleibt auf Fragmente des eigentlichen Problems beschränkt. Bei knappen Mitteln werden nur Teillösungen mit inkrementellem Wert für den Erkenntnisgewinn erarbeitet. Ohne Koordination zwischen den fragmentarischen Anstrengungen einzelner Individuen und Arbeitsgruppen ist das Ziel kaum je zu erreichen. Der Fortschritt bleibt geringfügig und die ungesteuerte Risikominderung verschleppt die Lösung des Gesamtproblems.

Es gibt Probleme, bei denen es nebensächlich ist, wann genau sie gelöst werden. Natürlich wäre es grossartig, wenn es bald gelänge, die Relativitätstheorie und die Quantenmechanik in einer *Great Unification* zusammenzuführen. Aber es ist für den Fortbestand des Lebens auf der Erde nicht existenziell wichtig, ob das tatsächlich schon bald oder erst in fünf oder zehn oder 100 Jahren gelingt. Aber dann gibt es Probleme, die wirklich bedeutend sind – und diese verlangen jetzt nach Lösungen.

3 Forcierte Lösungsfindung auf online Kollaborations-Plattformen

Grosse und facettenreiche Probleme können besser bewältigt werden, wenn sich viele unterschiedliche Menschen an der Lösungsfindung beteiligen. Aber selbst bei Problemen von so grosser Tragweite und von so hoher Komplexität wie dem Klimawandel oder der Armut grosser Teile der Weltbevölkerung wird niemand einen Beitrag leisten, der sein individuelles wahrgenommenes Risiko für untragbar hält.²

²Bei vielen Umweltschutzproblemen – auch beim Klimawandel – verschärft eine intergenerationale Diskontierung des Nutzens von Gegenmassnahmen die individuelle Risikowahrnehmung. Die Risikoevaluation ist eine Kosten/Nutzen-Evaluation bei gegebener Wahrscheinlichkeitsverteilung in Bezug auf das Eintreten des risikobehafteten Ereignisses. Umweltschutzprobleme zeichnen sich oft dadurch aus, dass die zur Lösung notwendigen Kooperations-Kosten heute getragen

Wenn dieses Risiko nicht wie bei einigen privatwirtschaftlichen oder staatlichen Grossorganisationen durch Reputations- oder Salär-Anreize reduziert wird, dann braucht es andere Mittel zur Risikominderung.

Ein naheliegender Ausgangspunkt für Überlegungen zu den Möglichkeiten einer wirksamen Risikominderung bei Forschungs- und Entwicklungsvorhaben sind *on-line Kollaborations-Plattformen*. Es ist zwar nicht anzunehmen, dass auf einer dieser Plattformen eine technologisch wirkungsvolle und wirtschaftlich sowie gesellschaftlich relevante Lösung für die Bewältigung des Klimawandels erarbeitet wird. Aber diese Plattformen zeigen beispielhaft, wie entweder die Wahrscheinlichkeit eines Lösungserfolges gesteigert wird, oder wie jede Mitwirkende ihren persönlichen Ressourceneinsatz kontrolliert reduzieren und trotzdem einen wertvollen Beitrag leisten kann. Diese virtuellen Arbeitsgemeinschaften – oder online Kollaborations-Plattformen – zeichnen sich aus durch eine grosse Wissensvielfalt und sie erlauben eine Arbeitsteilung im Kollektiv.

3.1 Informations- und Partizipationsoffenheit

Zwei Dinge sind all diesen Plattformen gemein. Erstens zeichnen sie sich aus durch ihre *Offenheit in Bezug auf Daten sowie hinsichtlich methodischer Aspekte im Problemlösungsvorgehen*. Es ist ein eigentliches Prinzip, dass Informationen frei zugänglich sind.³ Diese Informations-Offenheit führt dazu, dass vielversprechende Lösungsansätze vom Kollektiv rasch identifiziert werden und so die bereitstehenden Ressourcen natürlicherweise da zum Einsatz kommen, wo sie am meisten bewirken. Der allzu akademischen Maxime der kumulativen Wissensbildung wird bei so konsequenter Offenheit in höchstem Masse Genüge getan. Und nicht nur das. Oft vollzieht sich die Konvergenz zur besten Lösung auch schneller, als dies in geschlossenen Systemen der Erkenntnisgewinnung der Fall ist (Goldbloom 2016). Offenheit bewirkt also auch hier nicht nur den ressourcenschonenderen Lösungserfolg, sondern auch einen wahrscheinlicheren.

Die zweite Gemeinsamkeit von online Kollaborations-Plattformen ist die, dass sie meist komplett offen für jede und jeden sind und für eine grosse Zahl interessierter Amateure einen direkten Zugang zum sonst unerreichbaren Wissenschaftsbetrieb schaffen. Damit ermöglicht diese *Offenheit in Bezug auf die Partizipation* die Teilnahme vieler unterschiedlicher Menschen. Diversität und Arbeitsteilung sind denn auch zwei wesentliche Erfolgsfaktoren der Formate, die auf die *Crowd* als

werden müssen, während der Nutzen zukünftigen Generationen zuteilwird (Jacquet et al. 2013). Die Kosten/Nutzen-Evaluation fällt in einem solchen Szenario also – unabhängig von der Wahrscheinlichkeitsverteilung – umso schlechter aus und damit wird das Risiko als noch höher eingestuft.

³Diese Art der Offenheit gilt natürlich nicht bei Plattformen, auf welchen Forschungs- und Entwicklungsprobleme für kommerziell agierende Unternehmen gegen Bezahlung, zum Beispiel in Form von Preisgeldern, gelöst werden.

Arbeitskraft bauen.⁴ Diversität und Arbeitsteilung halten nämlich das wahrgenommene Risiko für jede einzelne Akteurin niedrig und erlauben eine wirksame Risikokontrolle. Auf der einen Seite bewirkt die hohe Diversität unter den Mitstreitenden eine Erhöhung der Wahrscheinlichkeit, in einem weit gefassten Suchraum auf eine passende Quellstruktur zu stossen und so eine kreative Lösung zu finden. Auf der anderen Seite erlaubt es die grosse Zahl von Mitwirkenden, den individuellen Ressourceneinsatz auf ein akzeptables Mass zu begrenzen, ohne gleich das ganze Vorhaben zu unterminieren.

In der Praxis sind zwei Grundtypen von Kollaborations-Plattformen weit verbreitet – aber beide setzen auf Informations- und Partizipationsoffenheit. Die wesentlichen Unterschiede zwischen den beiden Typen kommen in der Gegenüberstellung der Formate *Polymath* und *Foldit* zum Ausdruck: Ersteres baut auf ein breites Spektrum sehr spezieller Kompetenzen der Mitstreiter, zweiteres lebt von der Masse von Akteurinnen respektive von deren kumulierter Arbeitskraft.

3.2 Polymath vs. Foldit

*Polymath*⁵ ist eine online Kollaborations-Plattform mit hohem wissenschaftlichem Anspruch. Das Ziel jedes einzelnen Polymath-Projekts ist die Erbringung eines Beweises für ein mathematisches Theorem. Um es genau zu nehmen besteht die Plattform nur aus einem Projekt-Blog⁶ und einer Wiki-Page.⁷ Tim Gowers hat Polymath aus Neugierde gestartet. Im Jahr 2009 hat er in einem Post auf seinem persönlichen Blog die Frage aufgeworfen „*Is massively collaborative mathematics possible?*“ (Gowers 2009c) und um eine Antwort zu bekommen hat er auch gleich ein bis dahin ungelöstes mathematisches Problem als Testfall vorgeschlagen: den Beweis für einen Spezialfall des Density Hales – Jewett (DHJ) Theorems. Am 1. Februar 2009 hat Gowers die Beschreibung des Problems gepostet. Innerhalb von 37 Tagen haben 27 unterschiedliche Personen insgesamt 800 relevante Kommentare geschrieben (Gowers und Nielsen 2009). Schon am 30. März 2009 liess Gowers verlauten, dass ein Beweis vorliege. Im Jahr 2012 wurde dieser in den *Annals of Mathematics* unter dem Pseudonym D.H.J. Polymath publiziert (Polymath 2012).

Bei den Polymath-Projekten ist die Vielfalt unter den Teilnehmenden erfolgentscheidend. Um einen konstruktiven Beitrag leisten zu können ist in jedem Projekt sehr spezialisiertes Wissen nötig. Besonders wichtig für die erfolgreiche Problemlösung ist aber, dass Teile des Kollektivs Spezialwissen aus mathematischen Disziplinen einbringen, die peripher sind zur Disziplin des zu beweisenden Theorems. Auch wenn am Ende eine mathematische Beweisführung steht, also eine Erklärung, die den Regeln der deduktiven Logik gehorcht, so ist die Beweisfindung lange nicht so deterministisch, wie man das als Nicht-Mathematiker vermuten könnte. Die

⁴Einen Eindruck von Anzahl und Varietät dieser Plattformen vermittelt citizenscience.org.

⁵polymathprojects.org.

⁶The polymath blog: <https://polymathprojects.org/>.

⁷Hauptseite polymath wiki: http://michaelnielsen.org/polymath1/index.php?title=Main_Page.

Entwicklung einer neuen Idee zur Lösung eines Problems erfordert auch in der Mathematik viel Kreativität. In vielen Fällen kommen entscheidende Hinweise von ausserhalb der eigentlichen Problemdomäne.

*Foldit*⁸ ist ein online Computergame, bei welchem die Spieler denaturierte Proteinstrukturen so zu falten versuchen, dass daraus virtuelle Bilder potenziell natürlicher Proteine entstehen. Die Natur faltet denaturierte Proteinstrukturen immer so, dass das fertige native Protein in einen energetisch optimalen Zustand kommt (Nicholls et al. 1991). Diesen Prozess zu simulieren und die finale Proteinstruktur aus einer offenen Molekülkette abzuleiten ist aber selbst für Computer mit hoher Rechenleistung nicht zu bewältigen. Wissenschaftler am Center for Game Science und am Department of Biochemistry der University of Washington haben Foldit entwickelt, um sich die menschliche *Puzzle-solving Intuition* zunutze zu machen und diese mit Fachwissen und algorithmischem Problemlösen mit Hilfe von Computern zu verbinden (Foldit 2017).

Im Unterschied zu Polymath hängt der Erfolg von Foldit von der grossen Zahl von Spielern ab und nicht von deren spezieller Expertise. Jede und jeder mit einem guten Gespür für komplementäre Muster, Symmetrien, Harmonien, natürliche Transformationen und ähnlichen Gaben kann mittun. Das Spiel ist leicht zu verstehen und die Aufgabe ist immer die gleiche: Proteine falten. Die Lösung des einen Problems wird auch nicht von der Lösung eines anderen Problems bestimmt. Das macht es möglich, dass die Spieler unabhängig voneinander tätig sind und viele Puzzles parallel abgewickelt werden können.

Die Gegenüberstellung von Polymath und Foldit reflektiert die zwei grundsätzlichen Mechanismen, mit denen das individuelle wahrgenommene Risiko von Problemlöserinnen gesenkt werden kann. Polymath baut auf ein breites Spektrum an Kompetenzen, das die Mitstreiter einbringen. Dementsprechend zielen die Kollaborationsmethoden bei diesem Typ auf die Erhöhung der Wahrscheinlichkeit eines Lösungserfolges durch Diversität. Foldit lebt von der Arbeitskraft, die eine sehr grosse Zahl von Akteurinnen bereitstellt. Die Werkzeuge dieses Typs bewirken also eine Limitierung des Ressourceneinsatzes für jede einzelne von ihnen. Auch wenn das weder bei Polymath noch bei Foldit ein explizites Ziel ist, so ist die positive Risikoevaluation (aufgrund des wahrscheinlicheren Lösungserfolges resp. des limitierten Ressourceneinsatzes) in beiden Fällen eine Voraussetzung dafür, dass sich genügend unterschiedliche respektive sehr viele Mitstreitende aktiv beteiligen.

Die beschleunigte Lösungsfindung, wie sie auf online Kollaborations-Plattformen so gut funktioniert, hängt also wesentlich davon ab, dass individuelle Risiken mit geeigneten Mitteln entschärft werden und damit der Problem-Verschleppung Einhalt geboten ist. Genau das ist auch für die Lösung der drängenden Probleme gefordert.

⁸fold.it.

4 „Well-structured Problems“ und „Ill-structured Problems“

Das spielerische Falten denaturierter Proteinstrukturen in potenziell natürlich vorkommende Proteine ist ein „*well-structured problem*“ (WSP). WSPs zeichnen sich dadurch aus, dass die Ausgangslage klar und beschreibbar ist (eine lange Kette unterschiedlicher Aminosäuren, angeordnet in einer bestimmten Sequenz), dass der formelle Rahmen des Problems abgegrenzt werden kann (Proteinfaltung) und dass sich die Validität einer Lösung anhand bekannter Kriterien testen lässt (ein thermodynamisch stabiler Zustand bei gegebenen physiologischen Bedingungen). Zudem lassen sich bei WSPs die notwendigen (biochemischen und molekularbiologischen) Wissensgrundlagen annähernd abschliessend erfassen und die zulässigen Transformationsmechanismen von einem intermediären Zustand der Problemlösung in einen anderen, also die formellen Lösungsregeln, sind bekannt (und entsprechen den Gesetzen und Abläufen der Natur) (Simon 1973).

Zwei ergänzende Kriterien zur Prüfung des Vorliegens eines WSP's, vor allem im Kontext der technologischen Unterstützung, sind die Beschränkung der Laufzeit des Problemlösungs-Algorithmus' auf ein akzeptables Mass sowie die Abbildung allen zur Lösungsfindung benötigten deklarativen und prozeduralen Wissens in einer Datenbank. Ohne diese beiden Bedingungen könnte jedes Problem als „well-structured“ angesehen werden. Eine Problemlösungs-Maschine mit unbeschränkter Rechen- und Speicherkapazität erlaubte es nämlich, dass alle vorgängig genannten Kriterien für WSPs immer erfüllt würden. Wäre beispielsweise zur vollständigen Beschreibung der Ausgangslage im Proteinfaltungs-Problem das gesamte Zellmilieu in der Maschine abzubilden und nicht bloss die Aminosäuresequenz, so würde diese Zusatzinformation im (beliebig erweiterbaren) Speicher festgehalten und damit ein „*ill-structured problem*“ (ISP) (Ausgangslage nicht vollständig erfasst) zum WSP gemacht (Ausgangslage vollständig erfasst).

Diese beliebig erweiterbare Maschine existiert aber nicht und damit sind nicht alle Probleme „well-structured“. Im Gegenteil: kaum je werden alle oben angeführten Kriterien für WSPs vollständig erfüllt sein. Fehlte dem Fahrgestell-Design Tool von Autodesk beispielsweise die Kenntnis auch nur eines der Gesetze der Aerodynamik, dann würde aus dem „well-structured“ Optimierungsproblem rasch ein „ill-structured“ Kreativitätsproblem. Eher ist die Welt also voll von ISPs. Herbert Simon ging so weit zu sagen, dass die Struktur eines Problems kaum fassbar sei und die Grenzen zwischen „well-structured“ und „ill-structured“ immer vage: „*In general, the problems presented to problem solvers by the world are best regarded as ISPs. They become WSPs only in the process of being prepared for the problem solver*“ (Simon 1973, S. 186). Wichtig im Kontext von WSPs ist das Verständnis, dass es klar fassbare und mit deterministisch laufenden Algorithmen lösbare Probleme sind. Man könnte auch sagen: Lösungen für WSPs lassen sich berechnen.

Lösungen für ISPs lassen sich nicht berechnen. Mit dem Lösen von ISPs verhält es sich in etwa so, wie wenn ein Komponist zu erklären versuchte, wie er eine Fuge kreiert. Das Kreieren einer Fuge lässt sich nicht erklären und genau so wenig lässt sich eine Fuge berechnen. Den ISPs fehlt die Struktur, die eine künstliche Problemlö-

sungs-Maschine braucht, um sie aufbauend auf gespeicherten Inhalten und nach programmierbaren (oder auch selbst erlernten) Regeln zu lösen. Die Ausgangslage ist bei ISPs oft alles andere als klar und es fehlen Kriterien für die Identifikation und die Validierung einer Lösung. Das macht den Suchraum unscharf und potenzielle Lösungsansätze schwer fassbar. ISPs verschliessen sich einem antizipierbaren Lösungsweg. Das liegt aber nicht daran, dass durch eine komplexe (wohl aber deterministische) Folge von Regeln manövriert werden müsste, um zu einem Ergebnis zu kommen. Nein, wie Keith Holyoak es treffend ausdrückte ist der Lösungsweg für ISPs deshalb schwer zu errahnen „*because we have trouble thinking of even one idea worth pursuing*“ (Holyoak 1990, S. 285). Oft ist es bei ISPs auch kaum möglich zu sagen, welches Wissen für die Lösung des Problems notwendig sein wird. Führt beim Radiation Problem rein medizinisches Wissen zum Ziel oder ist Wissen aus der Optik oder vielleicht sogar aus der mittelalterlichen Kriegsführung gefragt? Zudem sind die formellen Problemlösungsregeln nicht einmal annähernd mit den tatsächlichen Gesetzen der realen Welt in Übereinstimmung zu bringen (was ja beim Fahrgestelldesign noch möglich schien). Die den ISPs zugrunde liegenden physikalischen und natürlichen Basismechanismen spielen in unermesslicher Komplexität zusammen und sind teilweise nicht einmal bekannt.

5 Zwei Systeme zur Lösung bedeutender Probleme: Ein formeller Rahmen

Im vorliegenden Text wird Wert gelegt auf die Herleitung eines ganzheitlichen formellen Rahmens. Diese Ganzheitlichkeit soll helfen, über eigentliche Systeme – also über aus mehreren Einzelteilen zusammengesetzte Komplexe – nachzudenken und sich so nicht im spezifischen Nutzen einzelner Technologien zu verlieren. Publikationen mit lose aneinandergereihten Wunsch-Funktionalitäten problemlösungsunterstützender Technologien führen nämlich nicht weiter. Der eine wünscht sich dies (vielleicht einen Automaten zur Generierung neuer Hypothesen) und die andere das (einen künstlichen Assistenten, der Forschungsergebnisse sammelt und zusammenfasst). Beides kann Sinn machen – nur der Zusammenhang fehlt. Ohne formellen Rahmen, der die Formulierung solcher Anforderungen leitet, bleibt alles Stückwerk. Die Anforderungen müssen *richtungsweisenden und handlungsleitenden Charakter* haben und nicht enumerativ-deskriptiven. All denjenigen, die bereit sind, beizutragen zur Entwicklung von Systemen und Technologien der kollektiven Lösung grosser Probleme, dient ein Kompass mehr als eine Landkarte.

Systeme und Technologien, die der zeitnahen Lösung der bedeutenden Probleme dienen sollen, müssen primär das individuelle wahrgenommene Risiko jeder potenziellen Problemlöserin mindern. Ist das nicht erreicht, wird eine Partizipation verweigert. Es ist dieses Risiko, welches die Wissenschaftlerin, den Ingenieur, den Entwickler, die Designerin oder auch den Künstler davon abhält, den ungewöhnlichen Weg auszuprobieren und sich Lösungsansätze aus weit entfernten oder gänzlich fremden Wissensgebieten zu erschliessen. Um den inkrementellen und diskontinuierlichen Charakter herkömmlicher Problemlösungsmechanismen in einen

progressiven und zielgerichteten umzuwandeln, ist diese *massive Exploration* aber erfolgsentscheidend. Darum steht die Risikominderung im Zentrum.

Der vorgeschlagene formelle Rahmen für die Entwicklung unterstützender Systeme und Technologien baut sich über drei Komponenten auf (siehe Abb. 1). Die erste Komponente erfasst den *Problemstruktur-Typen* und es trennt ISPs von WSPs. Diese grundsätzliche Zweiteilung ist wichtig, weil, wie später ausgeführt werden wird, die Wirkungsziele und die Leistungsanforderungen an Maschinen ganz unterschiedlicher Art sind. In der zweiten Komponente werden die *Wirkungsziele* dargestellt; und zwar in Bezug auf die Reduktion des individuellen wahrgenommenen Risikos jeder Problemlöserin. Dies soll bei WSPs durch eine systemgeführte Abwicklung regelbasierter Lösungsprozedere reduziert werden. Die Kontrolle (oder vielmehr die Limitierung) des Ressourceneinsatzes für die individuelle Akteurin ist hier die relevante Wirkgrösse. Bei ISPs soll das Risiko durch die Unterstützung des kreativen Problemlösungsprozesses reduziert werden. Als Wirkgrösse steht die Steigerung der Wahrscheinlichkeit im Zentrum, ein aufschlussreiches Lösungskonzept in einem Suchraum zu finden, der keine Domänengrenzen kennt. Die dritte Komponente beinhaltet dann *Leistungsanforderungen*, welche zur Erreichung der Wirkungsziele an die Maschinen im System gestellt werden. Mit Hilfe dieses modulhaften formellen Rahmens lässt sich folglich ganzheitlich darüber nachdenken, welcher Problemstruktur-Typ mit welchem Wirkungsziel bewältigt werden kann und welche Leistungsanforderungen dabei an die Technologien gestellt werden.

5.1 Problemstruktur-Typen: Kreativität für ISPs und Produktivität für WSPs

Entsprechend der Dichotomie der *Problemstruktur-Typen* (ISPs und WSPs) werden zwei Systeme zur Problembewältigung vorgeschlagen. In einem System sind Technologien und technologieunabhängige Bausteine eingefasst, die helfen, neue und wegweisende Ideen für die Lösung von ISPs zu evozieren. Es wird hier *Creativity System* genannt und damit die Forderung in den Mittelpunkt gestellt, dass etwas ganz Neues und Nützliches resultieren soll. Ins andere System gehen Technologien und weitere Bausteine ein, die den Ressourcenaufwand in der Lösungsfindung für WSPs kontrollieren helfen. Es wird hier *Productivity System* genannt, um zu betonen, dass die Optimierung des Ressourceneinsatzes im Zentrum steht.

Da die Lösungssuche bei ISPs in einem unscharfen Suchraum mit schwer fassbaren Lösungsansätzen und einem nicht zu antizipierenden Lösungsweg mit wenig Aussicht auf ein positives Ergebnis erfolgt, sind im Creativity System in erster Linie Technologien gefragt, die den Lösungserfolg begünstigen. Jede Methode, die zur Erhöhung der Wahrscheinlichkeit beiträgt, irgendwo im weiten Feld der möglichen Quellkonzepte auf eines zu stossen, welches die entscheidende Idee im Kontext des Zielkonzepts birgt, hilft, Probleme vom „ill-structured“-Typ zu bewältigen. Damit sinkt für den Nutzer eines solchen Systems das Risiko, seine Exploration erfolglos beenden zu müssen. Technologien zur Bewältigung von ISPs wirken also über die Erhöhung der Lösungswahrscheinlichkeit risikomindernd. Demgegenüber wirken

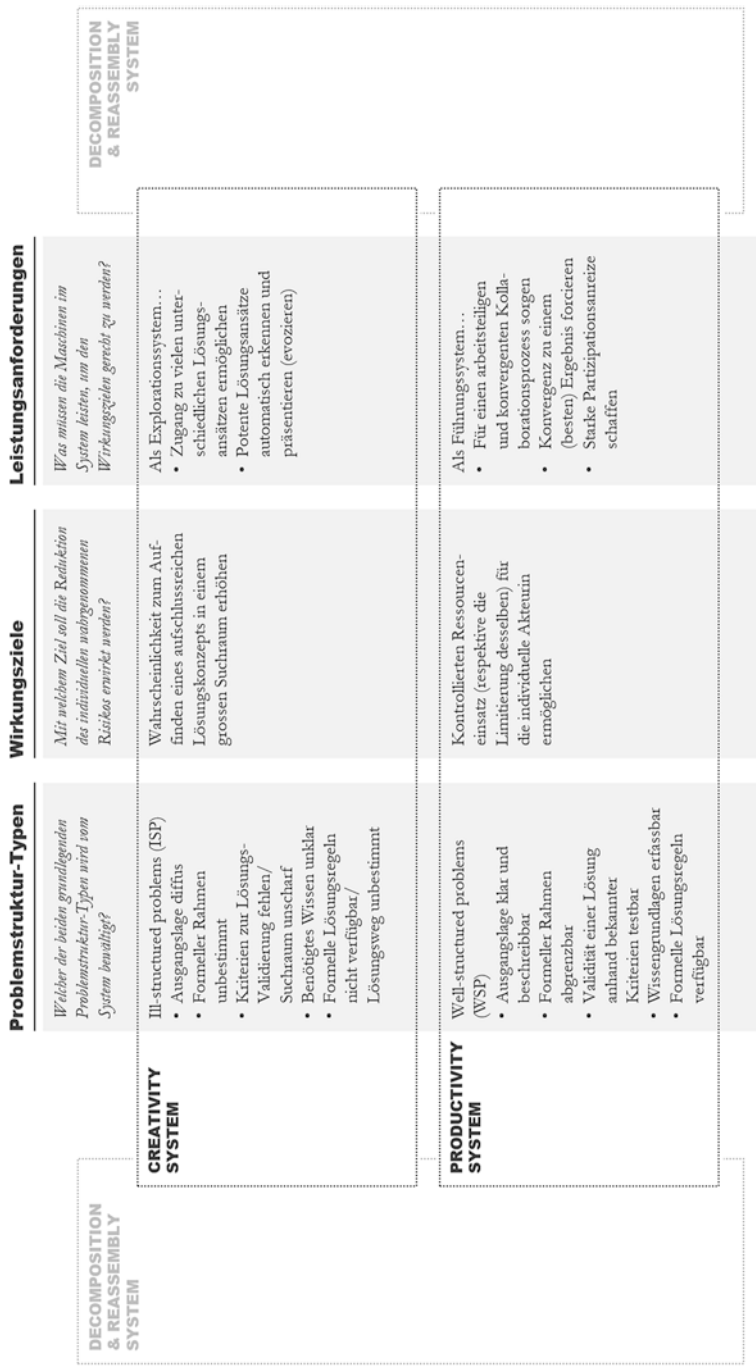


Abb. 1 Formeller Rahmen

Technologien zur Bewältigung von WSPs risikomindernd, weil sie den zu erwartenden individuellen Ressourceneinsatz vorhersehbar machen respektive der einzelnen Akteurin die Kontrolle über ihren Ressourceneinsatz ermöglichen, ohne die Problemlösung insgesamt zu verlangsamen. Die Wissenschaftlerin kann sich genauso wie der *Citizen Scientist* darauf verlassen, dass das Productivity System sie zum richtigen Zeitpunkt auf die richtige Fragestellung ansetzt und so ihre zeitlichen und intellektuellen Ressourcen produktiv genutzt sind.

Insgesamt wäre diesen zwei Systemen ein drittes hinzuzufügen. Dieses dritte System ist mit Technologien zu versehen, die einer Problem-Dekomposition sowie der Wiederherstellung einer ganzheitlichen Lösung dienen. Vielleicht könnte man es *Decomposition & Reassembly System* nennen. Seine Hauptaufgabe bestünde darin, ein komplexes Problem in der Dekompositions-Phase in Sub-Probleme zu unterteilen. Je nach dem, von welchem Strukturtyp diese Sub-Probleme sind (ISP oder WSP), werden sie entweder im Creativity System oder im Productivity System bearbeitet. Lägen dann für alle Sub-Probleme Sub-Lösungen vor, würden diese in der Wiederherstellungs-Phase zur Gesamtlösung zusammengefügt.

Zwei Gründe sprechen gegen eine weitere Auseinandersetzung mit dem Decomposition & Reassembly System im vorliegenden Text. Zwar stellt es konzeptionell gesehen ohne Frage eine sinnvolle Ergänzung zu den beiden oben genannten Systemen dar (und deswegen findet es auch Aufnahme in den formellen Rahmen, siehe Abb. 1). Schliesslich kann kein wirklich bedeutendes Problem in seiner vollen Komplexität einem der beiden Hauptsysteme zugeführt werden. In jedem grossen Problem stecken Sub-Probleme, zu deren Lösung es viel Kreativität und die sequentielle Integration unerwarteter, teilweise marginaler, Lösungsansätze bedarf. Und es lassen sich auch Sub-Probleme abgrenzen, zu deren Lösung einfach ein Mechanismus für die parallele, arbeitsteilige *Produktion* von Einzelerkenntnissen nötig ist. Eine Dekomposition macht bei komplexen Problemen also immer Sinn, um deren Einzelteile entweder mittels der explorativen Kraft eines Creativity Systems zu lösen, oder um die koordinativen Routinen eines Productivity Systems darauf anzusetzen. Aber, und das ist der erste Grund für den Verzicht auf eine weitere Auseinandersetzung mit dem dritten System, es lässt sich darüber streiten, ob insbesondere die Dekomposition nicht eigentlich der erste Schritt bei der Lösung von ISPs darstellt und damit das Decomposition & Reassembly System einfach ein Teil des Creativity Systems ist. Komplexität wird teilweise sogar mit „*ill-structuredness*“ gleichgesetzt (siehe bspw. (Franzoni und Sauermaun 2014)). Die Dekomposition ist bei einem Problem mit diffuser Ausgangslage und einem nicht zu antizipierenden Lösungsweg vom Typ ISP schliesslich um ein Vielfaches wichtiger, als es bei WSPs ist, die sich durch Klarheit und formelle Regeln zur Lösung auszeichnen. Zweitens darf angenommen werden, dass zum Beispiel ein genügend qualifizierter Wissenschaftler (oder Ingenieur, Entwickler, Designer, Künstler, etc.) über Dekompositions- und Wiederherstellungs-Heuristiken verfügt, die eine zweckmässige Komplexitätsreduktion seines Problems auch auf technologie-unabhängigem Weg erlaubt. Allein die zwingend ihm zufallende Aufgabe das Problem klar zum Ausdruck zu bringen bringt es mit sich, dass er sich über Dinge wie Abgrenzung, Kontext, Erkenntnisdefizit oder Lösungsansätze Gedanken gemacht hat. Diese Auseinandersetzung

führt unweigerlich zu einer Komplexitätsreduktion. Das anfänglich unfassbar grosse Problem wird fassbarer und typischerweise kleiner. So wird also das Decomposition & Reassembly System in den weiter unten vorgeschlagenen formellen Rahmen zur Formulierung von Anregungen aufgenommen, weil es sinnvoll erscheint, die Notwendigkeit zur Komplexitätsreduktion (Decomposition) respektive zur Erkenntnisintegration (Reassembly) darzustellen. Auf eine weitere Auseinandersetzung wird aber verzichtet, da die praktische Notwendigkeit dieses dritten Systems als *eigenständige Maschine* fraglich ist.

5.2 Wirkungsziele: Wie die Maschinen zur Risikominderung beitragen

Um nicht auf die speziellen Eigenschaften und Fähigkeiten bestimmter Technologien oder Methoden eingehen zu müssen und um nicht permanent spezifizieren zu müssen, ob nun eine Web-basierte Technologie oder ein selbstlernender Algorithmus oder eine Visualisierungs-Software oder gar tatsächlich ein Apparat zum Anfassen und Knopfdrücken gemeint ist, soll in der Folge etwas plakativ einfach von *Maschinen* die Rede sein. Das Ziel des vorliegenden Beitrages ist es nicht, konkrete Technologien vorzuschlagen, sondern es soll eine Grundlage geschaffen werden – eine Art *Denkmodell* – auf deren Basis dann Technologien und Methoden so entwickelt werden können, dass letzten Endes ein ganzheitlich funktionierendes System zur Lösung der wirklich bedeutenden Probleme entsteht. Unter einer solchen Maschine kann man sich denn auch eine spezifische Technologie mit einer einzigartigen Fähigkeit vorstellen (z. B. ein künstliches neuronales Netz zur Erkennung von Katzenbildern) oder eine relative breit gefasste *Über-Technologie* (z. B. die Web-Technologie). Mit *System* ist dann ein gemeinsam wirkendes und aufeinander abgestimmtes Gefüge von Maschinen, Menschen und Methoden gemeint. Der System-Gedanke geht also über die Bündelung von Technologien hinaus und er erfasst die Maschinen als lediglich einen Bestandteil eines weiter gefassten Gefüges. In der Folge geht es aber vornehmlich um diesen einen Bestandteil.

Das Creativity System soll da nach Lösungen suchen helfen, wo individuelles Wissen nicht hinreicht oder wo ein Gedankengang blockiert ist, weil er der eigenen Intuition widerspricht: „*Creativity is the epitome of cognitive flexibility. The ability to break conventional or obvious patterns of thinking, adopt new and/or higher order rules, and think conceptually and abstractly is at the heart of any theory of creativity*“ (Dietrich 2004, S. 1014). Das System soll die zur Lösung von ISPs benötigten kreativen Denkmuster aktiv evozieren und damit die Wahrscheinlichkeit erhöhen, dass potente Lösungsansätze auch dann zu ihrer Bestimmung finden, wenn sie aus fachlich weit entfernten Domänen stammen. Es soll betont werden, dass damit das konsequente Überschreiten des eigenen kognitiven Horizonts gemeint ist und nicht die Form von akademischer Extravaganz, bei der die Grenzen von einem wissenschaftlichen Fachgebiet zum benachbarten lediglich ausgelotet werden. Was

gemeinhin als Transdisziplinarität bezeichnet wird geht nicht annähernd weit genug! Wegweisende Erkenntnisse werden nicht dann gewonnen, wenn der Molekularbiologe mit einem Biochemiker zusammenarbeitet. Erst wenn er den Musikwissenschaftler trifft erkennt er, dass Melodien der Erforschung von Proteinfaltungsmechanismen dienlich sind und dass sich biologische und musikalische Konzepte zu neuer Erkenntnis zusammenfügen lassen. Weil dieses Zusammenfügen wirkungsvoll ist, dessen Zufälligkeit aber hohe Risiken für die beteiligten Akteure birgt, soll das Creativity System ein Zustandekommen dieser (Zusammen-)Fügung forcieren.

Das Productivity System hilft, die zur Lösung von WSPs benötigte arbeitsteilige Kollaboration zielgerichtet zu gestalten, um der einzelnen Akteurin Kontrolle über ihren Ressourceneinsatz zu geben ohne damit das Unterfangen insgesamt zu kompromittieren. Der Ressourceneinsatz wird abhängig vom Risikoappetit erhöht oder gesenkt. Arbeitsteilung ergibt sich zwar auch, wenn jedes Individuum eines potenziellen Kollektivs nach eigenem Gutdünken nur einen kleinen Teilaspekt des Gesamtproblems bearbeitet. Weil derart unkoordiniertes Verhalten im höchsten Grad unproduktiv ist und die Lösungsfindung verschleppt (wenn nicht gar verunmöglichlicht), soll ein Productivity System das intentionale Vorgehen im Kollektiv bringen. Die produktive Zusammenarbeit vieler Individuen bedingt denn auch mehr als nur prozessuale Steuerungsmechanismen. Wo immer in dezentralen Organisationen an einem gemeinsamen Ziel gearbeitet wird, ist auch gegenseitiges Vertrauen und eine gemeinsame Absicht wichtig. Das Productivity System muss demnach auch zu einer *shared consciousness* (McChrystal et al. 2015) beitragen können. Fehlt dieses geteilte Bewusstsein oder wird es nicht gut gepflegt, dann leidet die Bereitschaft zur Teilnahme am gemeinsamen Effort. Ebenso leidet die Partizipation, wenn die persönliche Anstrengung nicht gewürdigt wird oder wenn das Risiko besteht, dass jemand anderes daraus Nutzen schlägt. Auch wenn dem juristisch nicht so ist, so geht jede Akteurin mit dem Kollektiv einen Vertrag ein, der besagt: „*Ich investiere einen Teil meiner Ressourcen (z. B. Daten, Glaubwürdigkeit) und im Gegenzug wird mir eine Form von Kompensation zugesichert (z. B. die Aufnahme in die Autorenliste, wenn eine Publikation resultiert)*“. Starke Partizipationsanreize sind für ein potentes Productivity System von grösster Bedeutung. Allein die Möglichkeit zur Kontrolle des individuellen Ressourceneinsatzes und damit zur persönlichen Risikominderung allein nützt noch nichts. Im Normalfall bewirkt dies nämlich eine Limitierung des Ressourceneinsatzes. Diese Limitierung muss dann kompensiert werden durch die Rekrutierung einer genügend hohen Zahl von Akteuren; und genau dazu dienen Partizipationsanreize.

Das Productivity System hat im Kontext der Risikominderung also koordinierende und animierende Aufgaben. Es ist ein *Führungssystem*, welches den *konvergenten Prozess* hin zu der einen (besten) Lösung für ein WSP steuert. Die Aufgaben des Creativity Systems hingegen sind eher darin zu sehen, neue Ideen zu evozieren und *Structural Holes* (Burt 1992, 2004) – also Lücken zwischen einzelnen Wissensgebieten – zu überbrücken. Es ist ein *Explorationssystem*, welches den *divergenten Weg* hin zu einer Vielfalt möglicher Lösungskonzepte für ein ISP eröffnet.

5.3 Leistungsanforderungen: Was von Maschinen erwartet wird

Das Creativity System muss als Explorationssystem also den Zugang zu vielen unterschiedlichen Lösungsansätzen ermöglichen und es muss diejenigen Ansätze automatisch erkennen und vorschlagen können, die für die Lösung eines bestimmten Problems in Frage kommen.

Durchschlagskräftige Lösungen entstehen dann, wenn es starke Kontraste zwischen kollaborierenden Individuen gibt und wenn Wissensträger einbezogen werden, die üblicherweise marginalisiert und vom Problemlösungsprozess ausgeschlossen sind. Individuen an der fachlichen oder sozialen Peripherie befruchten den Problemlösungsprozess mit ihrem fremdartigen Wissen. Damit machen sie den Problemlösungserfolg wahrscheinlicher (Jeppesen und Lakhani 2010). Auf der Innovationsplattform *InnoCentive*⁹ sind es oft Personen, die mit den fachlichen Details der Problemdomäne kaum vertraut sind, welche die wirkungsvollsten Vorschläge zur Lösung hartnäckiger Forschungs- und Entwicklungsprobleme einbringen: zum Beispiel der Rundfunktechniker Bruce Cragin aus New Hampshire. Er hat für die NASA ein Modell zur Vorhersage von Sonneneruptionen entwickelt, an dem die NASA selbst während 35 Jahren ohne nennenswerten Erfolg gearbeitet hat. Cragin selbst betont, dass ihn sein Wissen in einem speziellen Gebiet der Plasmaphysik sowie seine visuell analytischen Fähigkeiten zu der Lösung geführt haben, zu der die Solarphysiker der NASA keinen Zugang fanden (Bingham und Spradlin 2011).

Für den wissenschaftlichen Fortschritt ist die soziale Interaktion und die gelebte Uneinigkeit innerhalb einer Forschungs-Gemeinschaft aber genauso wichtig wie die Diversität selbst. Individuelles Abweichen von sozialen Konventionen oder vom empirischen Konsens fördert die Erkenntnisgewinnung (Baliotti et al. 2015), während Herdenverhalten innerhalb von Wissensgebieten das Risiko von Konvergenz hin zu einem potenziell falschen Resultat birgt (Park et al. 2014). Eine zentrale Aufgabe, die von Maschinen bewältigt werden kann, ist es also, Blockaden zwischen bestehenden Wissenskomplexen aufzulösen, um so potente Ideen in Fluss zu bringen. Die potentesten Ideen müssen aber nicht nur zugänglich gemacht werden. Wenn sie sich zeigen, müssen sie auch gesehen werden. Das ist gerade bei ISPs keine leichte Aufgabe, denn es fehlt ja typischerweise an den Bemessungsgrundlagen, anhand derer Lösungen erkannt und validiert werden. Das Creativity System muss also eine Art Erkennungs-Automatismus für vielversprechende Quellstrukturen haben. Insbesondere soll der Automatismus diejenigen Quellstrukturen erkennen, die im Kern *gerade analog genug sind zur Zielstruktur*,¹⁰ sich ausserhalb des Kerns aber möglichst stark von dieser unterscheiden. Ansätzen mit zu umfassender Ähnlichkeit fehlt das Potenzial für etwas ganz Neues (Wolverton und Hayes-Roth 1994).

⁹innocentive.com.

¹⁰Eine leicht nachvollziehbare technische Interpretation dieser Strukturanalogie gibt John Sowa mit der Definition der „*maximal common projection* [of two structural graphs] with respect to [a] kernel“ (1976, S. 342).

Kein *General Problem Solver* (Newell et al. 1959) dieser Welt ist aber imstande, eine gute Intuition für potenziell passende Lösungsansätze von ISPs zu entwickeln. Es gibt wohl Algorithmen, die Beweise für ausgewählte Theoreme finden können (Ganesalingam und Gowers 2016). Aber es gibt keine Maschine, die eine Lösung für komplexe ISPs findet, indem sie intuitiv beurteilt, wie informativ und plausibel eine kaum verwandte Quellstruktur im Hinblick auf eine gegebene Zielstruktur ist. Maschinen sind unfähig, ausserhalb der deduktiven Logik zu denken und Lösungsansätze *einfach zu sehen*. Intuition ist nicht ihr Ding. Das können nur Menschen. Menschen können in Analogien denken, abduktiv schlussfolgern und sich mental gleichzeitig auf den Ebenen von abstraktem und konkretem Wissen bewegen. Damit entkommen sie dem Korsett deduktiv-inferenzlogischer und deterministischer Lösungsansätze, in welchem Maschinen gefangen sind. Das menschliche Gehirn kann aus einem diffusen Gesamtkontext heraus auf taugliche Lösungsmuster schliessen. Es gelingt uns, in der *Knowledge Soup* die besten Ideen zu sehen (Sowa 2006) und gute Quellstrukturen auch dann zu erkennen, wenn deren Verbindung zur Zielstruktur erstmals nur vage ist.

Weil also der Mensch die Intuition hat und die Maschinen über die nüchterne Suchkraft verfügen, scheint das gemeinsame Annähern an die besten Ideen ein aussichtsreicher Weg zu sein. Eine Problemlöserin könnte auf der Grundlage einer Suchanfrage von einer Maschine verschiedene Leit motive für Lösungen präsentiert bekommen. Dabei könnte eine solche Suchanfrage wegen der schwierigen Beschreibbarkeit der Ausgangslage bei ISPs durchaus ein mehrere Seiten langes geschriebenes Dokument sein oder gar das gesamte von der Problemlöserin initial recherchierte Text-, Bild- und weitere Datenmaterial umfassen. Unter *Leitmotiv* wird eine grundsätzliche Lösungsrichtung verstanden. Beispielsweise würde die Ärztin im Radiation Problem als mögliche Leit motive die Festungseinnahme und die Optik präsentiert bekommen. Ausgehend von diesen richtungsweisenden Vorschlägen schränkt die Problemlöserin die Suche weiter ein und die Maschine präsentiert im folgenden Suchzyklus bessere Ideen – und so weiter. Diese kooperative Form des *Progressive Deepening* (De Groot 1978) im Zusammenspiel von Mensch und Maschine ermöglichte zumindest eine semi-automatische Erkennung der besten Lösungsansätze.

Das Creativity System baut also in wesentlichen Teilen auf menschliche Kognitionen wie dem Problemlösen und dem Entscheiden. Damit ist es alles andere als eine autonom agierende, künstliche Intelligenz, die selbst lernt und Probleme löst. Der maschinelle Teil des Systems schafft in erster Linie einen Zugang zum massiv vielfältigen Wissen, das in Millionen von menschlichen Gehirnen und künstlichen Informationsspeichern verwahrt ist. Menschen sind dann gefordert, den Suchradius so einzugrenzen, dass am Ende die kreativste Idee als Lösungsansatz übrig bleibt.

Das Productivity System muss als Führungssystem Mechanismen und Instrumente bereitstellen, die für einen arbeitsteiligen und konvergenten Kollaborationsprozess sorgen und es muss Partizipationsanreize derart schaffen, dass sich möglichst viele Menschen an der Problemlösung beteiligen. Einerseits muss es also dafür sorgen, dass die verfügbaren Ressourcen sinnvoll eingesetzt werden und es dabei den einzelnen Akteuren überlassen, wie hoch ihr jeweiliger Ressourceneinsatz

sein soll. Andererseits muss das System insgesamt genügend Ressourcen rekrutieren, um die erfolgreiche Problemlösung sicherzustellen.

Um wirklich produktiv zu sein, muss das Productivity System Ergebnisse hervorbringen. Mehr noch: der gemeinschaftliche Problemlösungsprozess muss gegen *eine* beste Lösung konvergieren. Theoretisch haben WSPs nämlich beste Lösungen, da sie sich schlussfolgernd lösen lassen. Nur existieren reine WSPs eigentlich nicht. Wenn, dann sind es eher wohlpräparierte ISPs und damit idealisierte Ausnahmen (Simon 1973). Und zudem können Schlussfolgerungen auch falsch sein und die Problemlösung, auch die Lösung eines WSPs, wird fehlgeleitet. Weil sich das rein inferenzlogische Vorgehen also in den meisten Fällen nicht anbietet und damit die Problemlösung niemals komplett deterministisch abläuft, muss das System durch Ausprobieren alternativer inkrementeller Schritte in jeder Phase des Problemlösungsprozesses dafür sorgen, dass nur die besten Ansätze überleben und dass die diesen Ansätzen inhärenten Grundsätze (Prinzipien, Methoden, Annahmen, etc.) im gesamten nachfolgenden Prozess ihre Gültigkeit bewahren. Das wäre vergleichbar mit den drei Prinzipien, die Darwins Evolutionstheorie begründen: *Variation, Vererbung, Selektion*. So wie die Natur gegen eine (momentan) beste Lösung *konvergiert* und dabei auch immer wieder sub-optimale Zwischenlösungen hervorbringt, so konvergiert der Lösungsprozess für ISPs kaum geradlinig zum besten Ergebnis.

Die Arbeitsteilung innerhalb von Netzwerken, welche sich durch massive Kollaboration auszeichnen, wie zum Beispiel Linux¹¹ oder GitHub,¹² lässt sich kaum systemgestützt forcieren. Sie ist eher eine *emergent property* von Systemen, die in erster Linie Kollaborationsmechanismen zur Verfügung stellen. Alle, die sich an einer solchen kollektiven Entwicklung beteiligen, haben vollständige Kontrolle über ihren persönlichen Ressourceneinsatz. Wie bei der Arbeitsteilung selbst, gilt auch für die Kontrolle des Ressourceneinsatzes, dass diese nicht technologisch gewährleistet wird, sondern dass sie dem simplen *opt-in*-Prinzip solcher Plattformen geschuldet ist: „*Ich beteilige mich entweder mehr oder weniger; wie ich gerade will*“. Der Technologie kommt dann aber die wichtige Aufgabe zu, die fragmentarischen Leistungsbeiträge zu integrieren und aufeinander abzustimmen.

Genügend viele Ressourcen verfügbar zu haben bedingt die Partizipation vieler Akteure. Und diese Partizipation wiederum setzt Offenheit voraus. Jede massive Kollaboration wie sie vor allem in der virtuellen Welt zu finden ist (Linux oder GitHub wurden bereits beispielhaft genannt), wie sie teilweise aber auch in der realen Welt von Grossorganisationen anzutreffen ist (als gutes Beispiel sei hier das CERN¹³ genannt), lebt vom Prinzip der Offenheit. Ohne dass jede und jeder zu jeder Zeit Zugang zu allen Informationen hat, kann keine echte Kollaboration entstehen. Offenheit ist die entscheidende Grundlage für psychologische Sicherheit, also die Sicherheit in Bezug auf zwischenmenschliche Risiken. Die leistungsfähigsten Teams zeichnen sich zuallererst durch ein hohes Mass an psychologischer Sicherheit aus

¹¹ linux.org.

¹² github.com.

¹³ home.cern.

(Duhigg 2016). Offenheit bringt aber auch ein gewisses Mass an rechtlicher Sicherheit, zum Beispiel in Bezug auf den Schutz geistigen Eigentums. Insgesamt lässt sich also festhalten, dass Offenheit Vertrauen schafft. Hinzu kommt, dass nur mit Offenheit gemeinsame Ziele für alle erkennbar werden; selbst dann, wenn diese nicht explizit ausgesprochen sind. Über Vertrauen und gemeinsame Ziele befeuert Offenheit die *shared consciousness*, die für einen kollaborativen Effort unabdingbar ist.

Bei Offenheit geht es natürlich immer um den Zugang zu Rohdaten und Methoden, so dass Resultate plausibilisiert und deren Interpretation kritisiert werden kann. Im besten Fall unterliegen nicht nur Rohdaten und Methoden der Idee des *Open Access*, sondern sämtliche Daten: graphische Darstellungen, Messdaten, Konversationen, vertragliche Vereinbarungen und vieles mehr. Jede Akteurin muss die Möglichkeit haben, den vollen Informationsgehalt zu einer Problemlösungsanstrengung in ihre eigenen Überlegungen einfliessen zu lassen. Dieser Art der Offenheit liesse sich als *Informationsoffenheit* bezeichnen. In Ergänzung zur Informationsoffenheit ist hohe Produktivität in der Problemlösung – insbesondere in der Lösung grosser Probleme, die vieler unterschiedlicher Ressourcen bedarf – nur dann zu haben, wenn es auch maximale Partizipationsoffenheit gibt. *Partizipationsoffenheit* meint die Möglichkeit für jede und jeden, einen Beitrag leisten zu können. Dabei sind zwei Dinge wichtig.

Erstens bringt Offenheit nur dann etwas, wenn sich eine potenzielle Mitstreiterin all die Daten und Methoden, alle Prinzipien, Beweise, Vermutungen, offenen Fragen, ja den gesamten aktuellen Erkenntnisstand zu einem Problem und den Weg dahin mit nützlichem Aufwand erschliessen kann. Open Access allein bringt ihr noch nichts. Wenn sie diesen äussersten Rand der Erkenntnisgewinnung zuerst stückchenweise (und im schlechtesten Fall bruchstückhaft) nachbilden muss, bis sie sich involvieren kann, wird sie nicht zur Problemlösung beitragen, auch wenn sie vielleicht genau die Expertin wäre, die an dieser *Frontier* das entscheidende Wissen einbringen könnte. Der Aufwand ist schlicht zu gross. In einem System mit sehr hoher Produktivität wird es zur Herausforderung, den Gang der Erkenntnisgewinnung einfach nur mitzugehen – selbst für diejenigen, die an der Frontier aktiv mit-tun. Das System muss also über eine Art *Reproduktions-Modus* verfügen, der den Werdegang der bisherigen Erkenntnisgewinnung zusammenfasst und nachvollziehbar macht.

Zweitens sind es neben den Experten eines Fachs oft auch die Nicht-Experten, die an dieser Frontier wichtige Arbeit leisten. Speziell deren Integration setzt Partizipationsoffenheit voraus, denn die Hürden, die einem Beitrag im Weg stehen sind bei Nicht-Experten noch viel grösser als bei Experten. Auch wenn Laien vielleicht nicht das entscheidende Fachwissen einbringen, so erledigen sie doch wertvolle repetitive Arbeiten oder geben aufschlussreiche Hinweise für die weitere Entwicklung. Sie müssen den Weg hin zum äussersten Rand der Erkenntnisgewinnung nicht unbedingt nachvollziehen können. Aber sie müssen verstehen können, welchen Beitrag sie an der Frontier leisten können und warum dieser Beitrag wichtig ist. Viele Nicht-Akademiker freuen sich, wenn sie Teil des wissenschaftlichen Forschungsbetriebs sein

können. Und wie das Beispiel von drei EteRNA¹⁴-Spielern zeigt, werden sie manchmal selbst zu publizierten Wissenschaftlern: zu *Citizen Scientists* (Ohlheiser 2016). Offenheit bedeutet hier die Vermittlung eines komprimierten Erkenntnisstandes durch Übersetzung in eine Informationsform, die der Citizen Scientist verstehen kann. Ihm nützt der offene Zugang zu den Rohdaten nichts. Er muss ohne fachliches Grundlagenwissen den aktuellen Erkenntnisstand und die offenen Fragen intuitiv verstehen können. Einem Foldit-Spieler bringt der Zugang zu GenBank¹⁵ nichts. Er kann mit dem Wissen über die Aminosäuresequenz eines bestimmten Proteins nichts anfangen. Aber mit einem System, das für ihn viel fachliches Detailwissen komprimiert und es für seine Art der Informationsverarbeitung aufbereitet, indem es beispielsweise visualisiert wird, kann er sehr wohl einen wertvollen Beitrag zur Proteinforschung leisten. Das System muss also auch eine Art *Transpositions-Modus* haben, mit dem der aktuelle Erkenntnisstand in eine Form überführt werden kann, die dem Nicht-Experten eine Informationsverarbeitung ermöglicht. Es wird nicht ganz einfach sein, bestehendes Wissen adäquat aufzubereiten, denn abhängig vom Vorwissen jeder individuellen Nicht-Expertin und abhängig von der konkreten Aufgabenstellung, muss der Zugang anders erfolgen. Es stellt sich die Frage, ob sich Fachwissen zur Proteinfaltung mit ähnlichen Methoden zugänglich machen lässt, wie Wissen zur Entstehung von Galaxien. Könnten also Foldit und GalaxyZoo¹⁶ auf dem gleichen Transpositions-Modus aufbauen?

6 Maschinen im Creativity System

Die zentrale Aufgabe von Maschinen im Creativity System ist es, zwischen ganz fremden fachlichen Domänen zu vermitteln und Verbindungen herzustellen, wo vorher keine waren. Diese Aufgabe liesse sich auch so verstehen: solange der Musikwissenschaftler nicht weiss, was der Molekularbiologe wissen muss, weiss er auch nicht, wie diesem mit musikalischem Wissen geholfen wäre. Maschinen haben dafür zu sorgen, dass sich der Molekularbiologe und der Musikwissenschaftler finden und sie sollen zwischen den beiden moderieren und so deren Wissen zusammenfügen helfen.

6.1 Network of Neural Networks

Zum Fördern potenter Lösungsansätze müssen Maschinen ein Netzwerk bilden, welches möglichst viele menschliche Gehirne und deren gespeichertes Wissen vernetzt. In diesem *Network of Neural Networks* sollen nicht nur die Gehirne von Wissenschaftlern miteinander verbunden werden, um damit dann wissenschaftliche Probleme zu lösen. Das Netzwerk soll, etwas idealisiert ausgedrückt, alle Menschen

¹⁴ eternagame.org.

¹⁵ ncbi.nlm.nih.gov/genbank/.

¹⁶ galaxyzoo.org.

verbinden und es spricht nichts dagegen, dass in ein voll integriertes und funktionales Nervensystem dieser Art auch nicht-biologische, also artifizielle, Wissensträger eingebunden sind. Die Vielfalt der kognitiven Landschaft, die ein solches *Supra-Nervensystem* darstellen kann und die Weite dieser Landschaft sind denn auch wichtige Erfolgsfaktoren. Es ist dabei ganz nebensächlich, ob die Kognitionsinhalte biologischen oder artifiziellen Ursprungs sind.

Die Maschinen haben dann die Aufgabe, zielgerichtetes Navigieren „*through the hills and valleys of that landscape in search of a high-value solution*“ (Jeppesen und Lakhani 2010, S. 1018) zu ermöglichen. Bei der Lösung von Problemen des Typs „ill-structured“ ist es ja eben gerade nicht vorhersehbar, woher die erhellende Idee kommen wird. Dank maschineller Unterstützung sollen aber Suchen vermieden werden, die unergiebig sind. Maschinen müssen in dieser Landschaft der kognitiven Inhalte rasch zu *Points of Interest* (POI) navigieren können, um da Ortsmarkierungen zu setzen. Das Resultat wäre beispielsweise die Origami-Technik, identifiziert als POI mit Bezug zu einem möglichen Lösungsansatz zum bereits mehrfach beispielhaft geschilderten Problem der Proteinfaltung.¹⁷

Ähnlich einem einzelnen menschlichen Gehirn muss dieses Network of Neural Networks immer wieder neue Verbindungen herstellen und bestehende Verbindungen lösen können. Es muss plastisch sein und anfänglich unzusammenhängende Konzepte zu neuen Lösungsmustern verknüpfen, wenn diese mehrfach in zeitlicher Abhängigkeit und im gleichen Kontext stimuliert wurden. Damit korrespondiert diese Forderung nach kognitiver Flexibilität im künstlichen System mit neurowissenschaftlichen Erkenntnissen darüber, wie ein menschliches Gehirn sich strukturell verändert, wenn es lernt. Synaptische Verbindungen zwischen Nervenzellen werden da gestärkt, wo prä- und postsynaptische neuronale Aktivität wiederholt in kurzer zeitlicher Folge auftritt (Song et al. 2000). Wenn man sich anstatt eines neuronalen Netzwerks ein soziales Netzwerk von Menschen vorstellt – und schliesslich basiert das geforderte Nervensystem in Teilen auf menschlichen Wissensträgern – dann kommt der Überbrückung von Lücken in diesem Geflecht eine Bedeutung zu, die vergleichbar ist mit der Bildung neuer synaptischer Verbindungen. Maschinen können helfen, diejenigen Knoten im sozialen Gefüge zu identifizieren, an denen der Informationsfluss unterbrochen ist und die kreative Exploration stockt (Chen et al. 2009; Lou und Tang 2013). Gute Ideen können sich dann entfalten, wenn bestimmte Individuen an neuralgischen Punkten im Netzwerk *Structural Holes* schliessen und zwischen isolierten Clustern homogenen Wissens vermitteln.

6.2 Semantic Web der Kognitionen

Die Maschinen im Creativity System sind zudem darauf auszulegen, für hohe *Kompatibilität unterschiedlicher Formen der Wissensrepräsentation* zu sorgen und damit *Interoperabilität zwischen diesen* sicherzustellen. Ohne die Möglichkeit zum

¹⁷ Origami-Techniken werden schon erfolgreich für die platzsparende Faltung von Airbags eingesetzt (Bruton et al. 2016).

Informationsaustausch ist jedes Netzwerk wertlos. Das riesige neuronale Netzwerk im menschlichen Gehirn wäre ohne Informationsaustausch nicht mehr als ein Fett- und Proteinklumpen. Dank der schnellen, sehr präzisen und über grosse Distanzen funktionierenden Kommunikation zwischen Nervenzellen, wird dieser Klumpen zu einem ausserordentlich leistungsfähigen Computer. Auch wenn viele neuronale Kommunikationswege hoch spezialisiert sind, so funktioniert die Signalübertragung zwischen den Nervenzellen an den Synapsen durch jeweils einen von nur zwei Mechanismen: eine *chemische Reaktion* oder ein *elektrischer Impuls* (Kandel und Siegelbaum 2000). Damit sprechen alle Nervenzellen im Gehirn entweder die Sprache von chemischen Neurotransmittern oder die Sprache ionischer Ladungen. Wenn sich also beispielsweise bei Lernprozessen neue synaptische Verbindungen bilden und sich das Zusammenwirken bestimmter Gehirnregionen anpasst, dann stellt sich die Frage der Interoperabilität eigentlich nie, denn sämtliche Gehirnregionen sind beinahe uneingeschränkt kompatibel.

Sollen nun sehr unterschiedliche Wissenskomplexe in einem Network of Neural Networks zusammenwirken, dann müssen Maschinen eine Beziehung zwischen der kognitiven Struktur eines Problems und den kognitiven Strukturen vieler möglicher Inspirationsquellen herstellen können. Zu diesem Zweck hat es immer wieder Anstrengungen gegeben, grosse Wissensgebiete standardisiert zu erfassen und abzubilden – mit allen Daten und mit sämtlichen Ursache/Wirkungs-Prinzipien, die darin enthalten sind. Abgesehen von einigen sehr domänenspezifischen *Repositories* oder *Ontologien* (z. B. GenBank; siehe dazu (Benson et al. 2010)) sind sämtliche umfassend angelegten Standardisierungs-Versuche wie Cyc (Lenat 1995), WordNet (Miller 1995) oder das Electronic Dictionary Research Project (Yokoi 1995) gescheitert. Standardisierung wird erst recht der falsche Weg sein, wenn nicht nur explizites Wissen, sondern auch Überzeugungen, Begabungen, Handlungsweisen, Entscheidungsmuster, Argumente oder Motivstrukturen ins Creativity System eingespielt werden sollen.

Es ist unwahrscheinlich, dass es eine einheitliche Struktur- und Symbolsprache für all diese Kognitionen geben wird. Um Kompatibilität und Interoperabilität unterschiedlicher mentaler Prozesse und Zustände zu erreichen werden Ontologien und semantische Netzwerke benötigt, wie sie im World Wide Web mit dem Semantic Web schon erfolgreich eingesetzt sind (Shadbolt et al. 2006). Wie im Web muss auch in der Problemlösung im Network of Neural Networks dafür gesorgt werden, dass die Translation zwischen unterschiedlichen Formen der Wissensrepräsentation dem Auffinden isomorpher Abbildungen einer Problemstruktur in fremden Wissensdomänen nicht im Wege steht. Dazu ist es gar nicht nötig, dass in der Wissensdomäne der Zielstruktur die gleiche Sprache gesprochen wird wie in der Wissensdomäne der Quellstruktur. Aber es ist wichtig, dass eine Analogie im Sinngehalt erkannt wird. Tim Berners-Lee, der Erfinder des World Wide Web's, hat die Bedeutung des Semantic Web's in der *Unterstützung der Evolution des Wissens* gesehen und über die Notwendigkeit von Konzept-Kompatibilität und -Interoperabilität für die produktive Kollaboration verschiedener Subkulturen gesagt: „*Often two groups independently develop very similar concepts, and describing the relation between them brings great benefits ... [T]he relations allow communication and collaboration*

even when the commonality of concept has not (yet) led to a commonality of term“ (Berners-Lee et al. 2001, S. 37). Die Maschinen werden für den Betrieb eines *Semantic Web der Kognitionen* gebraucht.

6.3 Elemente funktionalen Denkens

Wenn das Network of Neural Networks steht und wenn Problem- und Lösungskonzepte mit einem Semantic Web der Kognitionen in Beziehung gesetzt werden können, trotz grösster Artenvielfalt, dann fehlen einem System, das die kreative Problemlösung unterstützen soll, noch *Elemente funktionalen Denkens*.

Nur wenn Netzwerk- und Inhaltsinformationen in die kognitiven Prozesse des problemlösenden Menschen eingewoben werden, können sie auch wirksam werden. Sowohl die spontane als auch die bewusst reflektierende Form der Problemlösung im menschlichen Gehirn setzt voraus, dass ausgehend von einem für die Informationsverarbeitung vorbereiteten Gesamtkonzept des Problems Assoziationen zu nahe und entfernt verbundenen Konzepten hergestellt werden (Kandel und Siegelbaum 2000). Die aktivierten Konzepte werden aus dem Langzeitgedächtnis ins Kurzzeitgedächtnis befördert, wo sie in die Informationsverarbeitung eingehen. Dabei erfolgt die Aktivierungsausbreitung entlang der assoziativen Verbindungen bis hin zu dem Punkt, an dem die Aktionspotenziale der neuronalen Signalübertragung versiegen. Bei allen aktivierten und task-relevanten Informationen wird diese *Spreading Activation* (Anderson 1983; Collins und Loftus 1975) neu initiiert.

Demnach wird von den Maschinen erwartet, dass sie *zwei wesentliche limitierende Faktoren der kognitiven Informationsprozessierungs-Kaskade biologischer Gehirne positiv beeinflussen*. Erstens ist die Assoziationsausbreitung bei einer einzelnen Problemlöserin auf ein einzelnes neuronales Netzwerk beschränkt. Konzepte können aus dem Langzeitgedächtnis nur dann aktiviert werden, wenn sie da auch gespeichert sind. Was die Problemlöserin nicht weiss, kann sie nicht abrufen. Maschinen aber könnten die Spreading Activation ausweiten auf Konzepte, die im erweiterten Network of Neural Networks gespeichert sind, das heisst in den Langzeitgedächtnissen anderer Personen oder auch in künstlichen Informationsspeichern. Damit kann der Suchraum massiv vergrössert werden. „*The more knowledge is readily available, the more relevant items can be ‘juggled’ in working memory“* (Dietrich 2004, S. 1019).

Um die Suche in diesem Raum produktiv zu gestalten, sollte die Aktivierungsausbreitung durch Feedbacks von Seiten der Problemlöserin gesteuert werden (Crestani 1997). Ihr Wissen zum Problem und ihre Fähigkeit „*to see through the gist“* können helfen, die initial von einem System präsentierten Leit motive auf deren Analogiepotenzial zu beurteilen und die Suche mittels Rückmeldung richtungsweisend zu unterstützen (Wolverton und Hayes-Roth 1994). Kreativität braucht Einschränkungen, wenn sie wirkungsvoll sein soll.

Der zweite limitierende Faktor ist die Anzahl der parallel im Kurzzeitgedächtnis verarbeitbaren Konzepte. Je weniger Konzepte verglichen, miteinander verbunden und aneinander gerieben werden können, desto weniger neue Ideen entstehen. Diese

Kapazitätslimite schränkt das kreative Potenzial enorm ein. Von den Maschinen wird die Unterstützung des *Working Memory* erwartet, entweder indem sie helfen, mehr Konzepte präsent zu halten oder indem sie gewisse Informationsverarbeitungsschritte gleich selbst übernehmen.

Auf der Basis von Netzwerkbildung, dem Semantic Web der Kognitionen und Elementen funktionalen Denkens sollen Maschinen also dieses Network of Neural Networks ermöglichen, das die Möglichkeiten des individuellen menschlichen Verstandes deutlich übersteigt. Nur mit einer derart übermenschlichen kreativen Kraft können die wirklich bedeutenden Probleme in der geforderten Frist gelöst werden. Ein Zusammenwirken von Wissenschaftlerinnen, Ingenieuren, Entwicklern, Designerinnen und vielen anderen mehr in der Art des vorgeschlagenen Netzwerks kommt aber nur zustande, wenn sich jede Akteurin in der Lage sieht, das individuelle wahrgenommene Risiko ihrem Risikoappetit anzupassen. Die adäquate Risikokontrolle ist die zentrale Voraussetzung dafür, dass jede von ihnen ihren Beitrag an eine *kollektive Intelligenz* leisten kann.

7 Maschinen im Productivity System

Dass ein Productivity System zur kollaborativen Bewältigung von WSPs eine grosse Zahl von Akteuren über ein Netzwerk verbinden muss, ist naheliegend. Und ein Netzwerk zu etablieren um die Kommunikation zwischen zwei beliebigen Knoten, also zwischen zwei Akteuren, in einem solchen Netzwerk zu gewährleisten, ist technisch gesehen keine grosse Herausforderung. Die *Internet-Maschine* macht das. Nicht so einfach ist es allerdings, dafür zu sorgen, dass diese Knoten arbeitsteilig kooperieren und insgesamt auf ein Ergebnis hinarbeiten und es ist ebenfalls alles andere als einfach, genügend Knoten in einem solchen Netzwerk zusammenzufügen. Für den erfolgreichen Aufbau eines Productivity Systems stellen sich also zwei Fragen: Wie Sorge ich für einen arbeitsteiligen und konvergenten Arbeitsprozess und wie rekrutiere und motiviere ich genügend Akteure?

7.1 Konvergenz als Evolution

Die Evolution führt deshalb zu Ergebnissen, die unter den gegebenen Umständen die besten sind, weil jede Entwicklung in Bezug auf ihre Leistungsfähigkeit an genau diesen Umständen gemessen wird. Leistungsfähige Entwicklungen überleben – nicht-leistungsfähige Entwicklungen überleben nicht. Weil Produktivität in der Problemlösung auch heisst, zu guten Ergebnissen zu gelangen, müssen die Maschinen das Prinzip des *Survival of the Fittest* auf die Domäne der Problemlösung übertragen. *Kaggle*,¹⁸ eine Kollaborationsplattform für Data Scientists, nutzt im Grunde genau dieses Evolutions-Prinzip, um rasch eine Konvergenz zur best-

¹⁸ [kaggle.com](https://www.kaggle.com).

möglichen Lösung zu erreichen (Goldbloom 2016). Jede Kaggle-Coderin misst die Leistungsfähigkeit ihres Algorithmus' immer wieder auf der Basis eines realen Datensatzes. Sie sieht sofort, ob ihr Algorithmus gute Überlebenschancen hat oder nicht und sie kann auch sehen, welche Algorithmen von anderen Coderinnen vielleicht noch fitter sind als ihr eigener. Weil die Algorithmen aller Wettbewerberinnen öffentlich sind, nimmt sie sich nun, was sie glaubt für die Verbesserung ihres eigenen Codes zu brauchen. Wenn die Leistung des Algorithmus steigt, wird sie nach einem weiteren Snippet zur Optimierung suchen, wenn die Leistung sinkt, wird sie die Anpassung wieder verwerfen. Vielleicht, und auch das ist erlaubt, nimmt sie sich einfach den bislang leistungsfähigsten Algorithmus und versucht, diesen noch ein kleines Stück zu verbessern. Auf diese Weise wird für ein hohes Mass an Variation gesorgt, es ist sichergestellt, dass Code-Sequenzen, die gut funktionieren, weitergetragen werden, und es erfolgt eine natürliche Selektion der leistungsfähigsten Algorithmen – und das alles sehr schnell.

Auch wenn nicht jedes WSP ein Coding-Problem und sich intermediäre Ergebnisse nicht immer so direkt auf ihre Leistungsfähigkeit hin testen und dann verbessern lassen, so sind die *Prinzipien von Variation, Vererbung und Selektion* doch sehr natürliche Bausteine in vielen formalisierten Problemlösungsmethoden. In der Software-Entwicklung heisst das *Agile Development* und im Produktdesign oder in der Produktentwicklung wird von *Design Thinking* und *Rapid Prototyping* gesprochen. Gemeint ist im Grunde immer dasselbe: iterative Entwicklung durch Ideation, Kreation, Testing und Anpassung in hoher Frequenz. Wenn Maschinen diesen Prozess unterstützen und diesem in der wissenschaftlichen Forschung oder im Engineering zu Akzeptanz und breiter Anwendung verhelfen können, dann kann das Productivity System dem Anspruch, zeitnah Lösungen hervorzubringen, gerecht werden.

Wie schon an anderer Stelle ausgeführt, ist die Arbeitsteilung keine Eigenschaft, die vom System forciert werden kann. Sie entsteht als *emergent property* sozusagen von innen heraus, wenn die Kollaboration vom System gut orchestriert wird. Die Technologie schafft lediglich zielführende Entwicklungs- oder Problemlösungsmechanismen.

7.2 Gaming the Limbic System

Bevor eine potenzielle Problemlöserin darüber nachdenkt, einen Beitrag zu leisten und auch bevor sie beginnt, irgendwelche individuellen Risiken bewusst wahrzunehmen, wird sie sich von einem Gefühl geleitet einem Productivity System entweder zuwenden oder sich ihm gegenüber verschliessen. Diese Gefühlsreaktion erfolgt im Limbischen System, einem entwicklungsbiologisch gesehen sehr alten Teil des Gehirns. Das Limbische System kontrolliert Funktionen, die der Selbsterhaltung und der Erhaltung der eigenen Spezies dienen und es spricht speziell gut auf emotionale Stimuli an. Im Gegensatz zum *thinking brain* wird das *Limbische System als feeling and reacting brain verstanden*. Aufgrund seiner zentralen Erhaltungsfunktionen – darunter fallen Funktionen wie die Fight-or-Flight Response,

die Nahrungssuche, die Fortpflanzung oder der Schutz der Nachkommen – sind die Signale des Limbischen Systems oft viel stärker, als diejenigen des Neocortex (in welchem, etwas vereinfacht ausgedrückt, *das Denken* stattfindet). Damit wird ein wirkungsvolles Triggering des Limbischen Systems zu einem zentralen Erfolgsfaktor für jedes Productivity System, das möglichst viele Akteure möglichst stark einbinden soll. Maschinen müssen also Menschen auf der Ebene der Gefühle ansprechen können, noch bevor sie deren Forderung nach Informations- und Partizipationsoffenheit Genüge tun.

Alle erfolgreichen Plattformen für massive Kollaboration wie Foldit, Kaggle, GalaxyZoo, Linux, GitHub oder auch das CERN bauen darauf auf, dass sie das Limbische System gekonnt ansprechen und Gefühle wecken: Gefühle der Zugehörigkeit, Selbstwertgefühle, Erfolgs- und Machtgefühle oder Gefühle ästhetischen Gefallens. Sie tun das mittels bewährter Gestaltungselemente aus dem *Game-Design* wie Player-Communities, Leaderboards, Play-Levels, Reward-Systemen oder Batches-of-Honor (siehe bspw. (Interaction Design Foundation [n.d.](#); Kumar und Herger [2013](#)), zu Mechanismen des Game-Designs). Zudem wirkt sich gutes Interaktions-Design positiv auf die Motivation potenzieller Akteure aus. Wenn der Umgang mit den Maschinen intuitiv ist, wenn Navigationsmuster menschlichen Denkmustern entsprechen, wenn Medienbrüche und nur mit hohem Aufwand überwindbare Schnittstellen zu Maschinen vermieden werden und wenn eine Maschine hohen ästhetischen Ansprüchen genügt, dann ist die Bereitschaft höher, eine Leistung zu erbringen. Games sind also wichtige und in vielfacher Hinsicht auch bereits verfügbare Maschinen für den Aufbau eines Productivity Systems.

7.3 Informationsoffenheit mit „Smart Contracts“

Die komplette Offenheit in Bezug auf alle möglichen Arten von Rohdaten und prozessierten Daten in allen möglichen Formaten kann mit modernen Speichern und Datenbanktechnologien prinzipiell gewährleistet werden. Selbst wenn die Nutzbarmachung dieser Daten überaus komplex ist und Wissenschaftszweige wie die Astronomie oder die Genomik die Grenzen der Computertechnologie austesten werden (Stephens et al. [2015](#)), so stimmen Open-Source Entwicklungen wie Hadoop¹⁹ oder Spark,²⁰ die die Abfrage, Übertragung oder Verarbeitung grosser Datenmengen stark vereinfachen, zuversichtlich. Solche Maschinen gewährleisten heute schon, dass auch Nicht-Experten eine grosse Vielfalt von Daten nutzen können.

Eine entscheidende Voraussetzung zur Realisierung der universellen Informationsoffenheit ist aber der Abbau rechtlicher Hemmnisse. Ein Informationstransfer ist letzten Endes eine Transaktion – auch unter der Prämisse der Offenheit: *„Ich gebe Informationen und ich nehme etwas dafür“*. Genau wie bei einer Transaktion von Gütern oder finanziellen Mitteln muss bei Transaktionen von Informationen der ursprüngliche Informationsinhaber geschützt werden. Bei voller Informationsoffen-

¹⁹ hadoop.apache.org.

²⁰ spark.apache.org.

heit ist dieser Schutz besonders wichtig, denn im Normalfall erhält ein Informationsgeber keine direkte Gegenleistung. Diese ist für ihn natürlich nicht minder wichtig, denn ohne Gegenleistung würde er einer Transaktion nicht zustimmen. Nur erfolgt die Gegenleistung nicht sofort und in welcher Form sich diese zeigen wird, ist auch nicht immer abzusehen. Umso mehr will der Informationsgeber die Sicherheit haben, beispielsweise in die Autorenliste einer allfällig resultierenden Publikation aufgenommen zu werden oder an einem Patent teilzuhaben. *Smart Contracts* (siehe bspw. (Gopie 2018)) können diese Sicherheit geben, indem sie die Modifikations-Resistenz Blockchain-gesicherter Aufzeichnungen – und eben auch quasi-juristischer Vereinbarungen – nutzen. Dank der Blockchain-Technologie (oder einer anderen Distributed-Ledger-Technologie) kann der ursprüngliche Informationsinhaber oder auch ein Ideengeber, ein Programmcode-Schreiber und jeder andere Beitragsleistende zweifelsfrei und für alle – also auch für juristische Instanzen – identifiziert werden, sollte die Notwendigkeit dafür entstehen (siehe bspw. (3Blue1Brown 2017; Hülsbömer und Genovese 2018)). *Smart Contracts* könnten so ein Schlüssel zur Realisierung wirklicher Informationsoffenheit werden. Das macht sie dann zu einer zwingenden und zentralen Komponente jedes Productivity Systems, denn nur so kann der Anspruch an komplette Offenheit in Bezug auf Daten mit den Restriktionen des Immaterialgüterrechts und des Persönlichkeitsrechts (unterschiedlicher Länder) vereinbart werden.

7.4 Ein Reproduktions-Modus für die Expertin

Wenn nun das Gefühl für eine Partizipation an einer massiv kollaborativen Problemlösungsanstrengung spricht (*Gaming the Limbic System*) und die Informationsoffenheit mittels *Smart Contracts* Tatsache ist, dann muss sich eine potenzielle Akteurin immer noch involvieren *können*. Bei einer Diskussion muss man deren Verlauf sowie den aktuellen Stand kennen, um sich sinnvoll beteiligen zu können. Für die kollaborative Problemlösung gilt das sinngemäss: nur wer den Problemlösungsverlauf und den aktuellen Stand kennt, kann mittun. Wie schon erwähnt, wird von einem Productivity System eine Art Reproduktions-Modus für Experten und ein Transpositions-Modus für nicht Experten gefordert (siehe zu Letzterem den nächsten Abschnitt).

Um zu konkretisieren, welche Erwartungen diejenigen Maschinen zu erfüllen haben, die den Reproduktions-Modus ermöglichen sollen, muss geklärt sein, was genau reproduziert wird. Die Forderung, dass ein „*intelligent science assistant*“ relevante Forschungsergebnisse „*across the worldwide multilingual spectrum of blogs, preprint archives, and discussion forums*“ (Gil et al. 2014, S. 171) zusammenfassen soll, setzt voraus, dass der Absender einer solchen Anfrage auch klar macht, über welche Facette der *worldwide Forschungsergebnisse* er gerne eine Zusammenfassung hätte. Es ist eine fundamental andere Aufgabe, den fachlich eingrenzenden Diskussionsverlauf zur Lösung eines Polymath-Problems zu reproduzieren, als den gesamten Wissensstand zur Kombinatorik oder nur schon zu den verschiedenen Lösungsansätzen für das Density Hales-Jewett Theorem übersichtlich

nachzubilden. Zudem werden zwei potenzielle Polymath-Mitstreiter kaum eine identische Reproduktion sehen wollen. Wenn ihr Vorwissen, ihre methodischen Vorlieben und ihre Überzeugungen im Hinblick auf den kombinatorischen Ansatz der Beweisführung zum Density Hales-Jewett Theorem nicht deckungsgleich sind, wird sich die eine mehr mit den Diskussionsdetails mit Bezug auf die Nutzung von Szemerédi's Regularity Lemma befassen wollen (Gowers 2009a), während der andere sich intensiver mit denjenigen Threads auseinandersetzt, die die Imitation eines Beweises für ein ähnliches Problem mit Hilfe des Triangle-Removal Lemmas behandeln (Gowers 2009b). Aufgrund der schweren Eingrenzbarkeit des Reproduktions-Umfangs und wegen der hohen Unterschiedlichkeit der potenziellen Nutzer wird hier vorgeschlagen, die Maschinen nicht primär auf die Reproduktion von Inhalten anzusetzen, sondern auf die Reproduktion des Entwicklungspfad zur Lösung eines ganz bestimmten Problems. Dieser Entwicklungspfad soll (beispielsweise) visuell darstellen, auf welchen Annahmen der Lösungsansatz aufbaut, welche inferenzlogischen Schritte wann von wem ausprobiert worden sind, welche davon verworfen wurden und warum, ob sich unterschiedliche Stossrichtungen herausgebildet haben und wer welche verfolgt, wann und auf welcher Basis die zentralen Erkenntnisse gewonnen wurden, welche Ansätze aktuell die vielversprechendsten sind und welche es einmal waren usw. Die Inhalte selbst, die an jedem Punkt dieses Pfades natürlich sehr relevant sind, kann sich dann jede und jeder nach eigenem Bedarf interaktiv und mittels erweiterter Such-Services erschliessen (Shneiderman 2007). Solche Technologien zur Reproduktion eines Entwicklungspfad würden nicht nur für diejenigen von Bedeutung sein, die sozusagen *on the run* in ein Projekt einsteigen wollen oder die den Anschluss zur Frontier verloren haben. Die Reproduktionstechnologien würden bei Bedarf auch eine geordnete Rückabwicklung hin zu einem Punkt ermöglichen, an welchem eine falsche Entscheidung getroffen wurde.

7.5 Ein Transpositions-Modus für den Citizen Scientist

Visualisierungen eignen sich hervorragend, um einen grossen Informationsgehalt *auf einen Blick* erfassbar zu machen. Die Datenübertragungsrate von der Umwelt ins kognitive System eines Menschen ist enorm hoch, die herausragendsten Elemente der Gesamtinformation sind sofort erkennbar, so auch die Fehler, und die Information kann je nach Aufgabe auf sehr unterschiedlichen Aggregationsniveaus verarbeitet werden (Ware 2013). Wenn die Visualisierung des Problemlösungspfad eine der wesentlichen Forderung ist, die an die Reproduktionstechnologien für die Expertin gestellt wird, dann muss Visualisierung ein zentraler Anspruch an die Entwicklung von Technologien sein, die den Transpositions-Modus für eine wirkungsvolle Integration aller Citizen Scientists bilden, die nur allzu gerne einen Beitrag leisten, wenn ihnen ein Zugang zu einem Problem eröffnet wird. Gerade diejenigen, die die fachlichen Details nicht kennen und die Sprache, wie sie in einem bestimmten Fachgebiet gesprochen wird, nicht verstehen, sind auf die universelle Gültigkeit von visuellen Symbolen und deren räumlichem Arrangement angewiesen.

Nur leider ist Visualisierung eine sehr kontextbezogene Kunst. Die Struktur eines Proteins zu visualisieren bedarf ganz anderer Techniken (z. B. der Farbgebung, der Farbsättigung oder der Hell-Dunkel-Kontrastierung) und anderer stilistischer Mittel (z. B. Symmetrien, Perspektiven oder Figur-Grund-Beziehung), als es die Visualisierung eines Computer-Netzwerkes bedarf.

Vielleicht, aber das ist nicht mehr als intuitions-getriebene Spekulation, wird im Productivity System der Zukunft die Visualisierungs-Kompetenz, wie sie hier sowohl für den Reproduktions-Modus als auch für den Transpositions-Modus vorgeschlagen wurde, mittels *Virtual Reality* eingelöst (siehe bspw. (Rubin 2018)). Die Möglichkeiten dieser Technologie scheinen fast ohne Grenzen zu sein, wenn es darum geht, neue Formen der Kommunikation zu finden. Diese Kommunikation muss dabei nicht zwingend eine Kommunikation zwischen Menschen sein. Genauso gut geeignet scheint Virtual Reality als Kommunikationsinstrument zwischen Menschen und Maschinen zu sein (Rautaray und Agrawal 2015).

8 Und wer baut diese Maschinen?

Ob tatsächlich ein Semantic Web der Kognitionen in der skizzierten Art Teil eines zukünftigen Creativity Systems ist und ob Smart Contracts auf der Basis einer Distributed-Ledger-Technologie zu einem Productivity System beitragen werden bleibt ungewiss. Die Inhalte zur Ausgestaltung der Maschinen im Creativity respektive im Productivity System sollen Inspiration sein, über diejenigen Technologien nachzudenken, die uns helfen können, die grossen und die wirklich bedeutenden Probleme der heutigen Zeit zu lösen. Dabei sind die präsentierten Ansätze als Aufforderung zu verstehen, den formellen Rahmen mit fundiertem Wissen zu den Möglichkeiten moderner Technologien, aber unbedingt auch mit visionären Ideen, zu bestücken.

Da wir auf deren Entwicklung nicht beliebig lange warten können – die grosse Bedeutung der Probleme meint ja eben auch die Dringlichkeit der Lösungsfindung – sei hier abschliessend noch eine Prioritätenordnung vorgeschlagen: Wir sollten uns in erster Linie derjenigen Maschinen annehmen, die uns im Creativity System helfen, „ill-structured problems“ zu lösen. Denn wie schon erläutert sind eigentlich alle Probleme, mit denen wir es zu tun haben, „ill-structured“. Wie Werner Simon sich ausdrückte, werden diese misslich strukturierten Probleme höchstens zu etwas weniger Misslichem „*in the process of being prepared for the problem solver*“ (Simon 1973, S. 186). Und dann ist noch die Frage zu klären, wer sich um die Technologieentwicklung kümmern soll: Wer baut diese Maschinen? Die Antwort ist klar: Wir alle – in einer massiv kollaborativen Problemlösungsanstrengung!

Literatur

- 3Blue1Brown (2017) Ever wonder how Bitcoin (and other cryptocurrencies) actually work?
- Anderson JR (1983) A Spreading Activation Theory of Memory. *Journal of Verbal Learning and Verbal Behavior*, 22(3):261–295.

- Babai L (2017) Quasipolynomial claim restored. <http://people.cs.uchicago.edu/~laci/update.html>.
- Baliotti S, Mäs M, Helbing D (2015) On disciplinary fragmentation and scientific progress. *PloS one*, 10(3):e0118747.
- Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2010) GenBank. *Nucleic Acids Research*, 38(Database issue):D46–D51.
- Berners-Lee T, Hendler J, Lassila O (2001) The semantic web. *Scientific American*, 284(5):28–37.
- Bingham A, Spradlin D (2011) NASA's New Innovation Framework. Upper Saddle River, NJ: FT Press.
- Boden MA (2004) *The creative mind: Myths and mechanisms* (2nd ed.). New York, NY: Routledge.
- Bruton JT, Nelson TG, Zimmerman TK, Fernelius JD, Magleby SP, Howell LL (2016) Packing and deploying Soft Origami to and from cylindrical volumes with application to automotive airbags. *Royal Society Open Science*, 3(9):1–15.
- Burt RS (1992) Structural holes: The social structure of competition. Cambridge, MA and London, UK: Harvard University Press.
- Burt RS (2004) Structural holes and good ideas. *American Journal of Sociology*, 110(2):349–399.
- Bywater RP, Middleton JN (2016) Melody discrimination and protein fold classification. *Heliyon*, 2(10):e00175.
- Chen C, Chen Y, Horowitz M, Hou H, Liu Z, Pellegrino D (2009) Towards an explanatory and computational theory of scientific discovery. *Journal of Informetrics*, 3(3):191–209.
- Collins AM, Loftus EF (1975) A Spreading-Activation Theory of Semantic Processing. *Psychological Review*, 82(6):407–428.
- Conti M (2017). The incredible inventions of intuitive AI. TED Talk. https://www.ted.com/talks/maurice_conti_the_incredible_inventions_of_intuitive_ai.
- Crestani F (1997) Application of spreading activation techniques in information retrieval. *Artificial Intelligence Review*, 11(6):453–482.
- De Groot, A. D. (1978). *Thought and choice in chess*, 2nd ed., vol. 4. The Hague: Mouton Publishers.
- Dietrich A (2004) The cognitive neuroscience of creativity. *Psychonomic Bulletin & Review*, 11(6):1011–1026.
- Duhigg C (2016) What Google learned from its quest to build the perfect team. https://www.nytimes.com/2016/02/28/magazine/what-google-learned-from-its-quest-to-build-the-perfect-team.html?_r=0. Erstellt: 25.02.2016. Letzter Zugriff: 11.08.2019.
- Duncker K (1945) On problem-solving. *Psychological Monographs*, 58(5):i–113.
- Foldit (2017) The science behind Foldit. <http://fold.it/portal/info/about>.
- Franzoni C, Sauermann H (2014) Crowd science: The organization of scientific research in open collaborative projects. *Research Policy*, 43(1):1–20.
- Ganesalingam M, Gowers T (2016) A Fully Automatic Theorem Prover with Human-Style Output. *Journal of Automated Reasoning*, 58:253–291.
- Gil Y, Greaves M, Hendler J, Hirsh H (2014) Amplify scientific discovery with artificial intelligence. *Science*, 346(6206):171–172.
- Goldbloom A (2016). What has Kaggle learned from 2 million machine learning models? <https://youtu.be/8KzjARKIgTo>. Erstellt: 05.02.2016. Letzter Zugriff: 11.08.2019.
- Gopie N (2018) What are smart contracts on blockchain? <https://www.ibm.com/blogs/blockchain/2018/07/what-are-smart-contracts-on-blockchain/>. Erstellt: 02.07.2018. Letzter Zugriff: 11.08.2019.
- Gowers T (2009a) DHJ – Possible Proof Strategies. <https://gowers.wordpress.com/2009/02/13/dhj-possible-proof-strategies/>. Erstellt: 13.02.2009. Letzter Zugriff: 11.08.2019.
- Gowers T (2009b) DHJ – The Triangle-Removal Approach. <https://gowers.wordpress.com/2009/02/06/dhj-the-triangle-removal-approach/>. Erstellt: 06.02.2009. Letzter Zugriff: 11.08.2019.
- Gowers T (2009c) Is massively collaborative mathematics possible? <https://gowers.wordpress.com/2009/01/27/is-massively-collaborative-mathematics-possible/>. Erstellt: 27.01.2009. Letzter Zugriff: 11.08.2019.

- Gowers T, Nielsen M (2009) Massively collaborative mathematics. *Nature*, 461(7266):879–881.
- Gross P (2018) Zur Lösung der wirklich bedeutenden Probleme: Was von Maschinen erwartet wird. *Informatik-Spektrum*, 41(1):38–51.
- Holyoak KJ (1990) Problem solving. Smith EE, Osherson DN (Hrsg.) *Thinking: An invitation to cognitive science*, 2 ed., 3:267–296. MIT Press.
- Holyoak KJ, Thagard P (1989) Analogical Mapping by Constraint Satisfaction. *Cognitive Science*, 13(3):295–355.
- Hülsbömer S, Genovese B (2018) Definition, Vorteile, Nachteile: Was ist Blockchain? computer-woche / IDG Business Media GmbH.
- Interaction Design Foundation (n.d.). Gamification. <https://www.interaction-design.org/literature/topics/gamification>. Erstellt. n.d. Letzter Zugriff: 11.08.2019.
- Jacquet J, Hagel K, Hauert C, Marotzke J, Röhl T, Milinski M (2013) Intra- and intergenerational discounting in the climate game. *Nature climate change*, 3(12):1025–1028.
- Jeppesen LB, Lakhani KR (2010) Marginality and Problem-Solving Effectiveness in Broadcast Search. *Organization Science*, 21(5):1016–1033.
- Jung-Beeman M, Bowden EM, Haberman J, Frymiare JL, Arambel-Liu S, Greenblatt R, Reber PJ, Kounios J (2004) Neural activity when people solve verbal problems with insight. *PLoS Biology*, 2(4):0500–0510.
- Kandel ER, Siegelbaum SA (2000) Overview of Synaptic Transmission. Kandel ER, Schwartz JH, Jessell TM (Hrsg) *Principles of Neural Science*, 4th ed.. New York, NY: McGraw-Hill.
- Klarreich E (2017a) Complexity Theory Problem Strikes Back. <https://www.quantamagazine.org/20170105-graph-isomorphism-retraction/>. Erstellt: 05.01.2017. Letzter Zugriff: 11.08.2019.
- Klarreich E (2017b) Graph Isomorphism Vanquished – Again. <https://www.quantamagazine.org/graph-isomorphism-vanquished-again-20170114/>. Erstellt: 14.01.2017. Letzter Zugriff: 11.08.2019.
- Koestler A (1964) *The act of creation*. London, UK.
- Kumar JM, Herger M (2013) *Gamification at Work: Designing Engaging Business Software*. Lightning Source Inc.
- Leary MR, Baumeister RF (2017.) The need to belong: Desire for interpersonal attachments as a fundamental human motivation. Zuckauskiene R (Hrsg.), *Interpersonal Development*, pp. 57–89. Routledge.
- Lenat DB (1995) CYC: A large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38(11):33–38.
- Lou T, Tang J (2013) Mining structural hole spanners through information diffusion in social networks. *Proceedings of the 22nd International Conference on World Wide Web*, Rio de Janeiro, Brazil.
- McChrystal S, Collins T, Silverman D, Fussell C (2015) *Team of teams: New rules of engagement for a complex world*. New York, NY: Penguin.
- McClelland DC, Atkinson JW, Clark RA, Lowell EL (1953) *The achievement motive*. East Norwalk, CT: Appleton-Century-Crofts.
- Mednick SA (1962) The associative basis of the creative process. *Psychological review*, 69(3):220.
- Miller GA (1995) WordNet: A lexical database for English. *Communications of the ACM*, 38(11):39–41.
- Newell A, Shaw JC, Simon HA (1959) Report on a general problem solving program. IFIP congress.
- Nicholls A, Sharp KA, Honig B (1991) Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Structure, Function, and Genetics*, 11(4):281–296.
- Ohlheiser A (2016) An addictive online puzzle game is turning its players into published scientists. https://www.washingtonpost.com/news/the-intersect/wp/2016/02/17/an-addictive-online-puzzle-game-is-turning-its-players-into-published-citizen-scientists/?noredirect=on&utm_term=.8325326528fe. Erstellt: 17.02.2016. Letzter Zugriff: 11.08.2019.

- Park IU, Peacey MW, Munafo MR (2014) Modelling the effects of subjective and objective decision making in scientific peer review. *Nature*, 506(7486):93–96.
- Polymath DHJ (2012) A new proof of the density Hales-Jewett theorem. *Annals of Mathematics*, 175(3):1283–1327.
- Rautaray SS, Agrawal A (2015) Vision based hand gesture recognition for human computer interaction: a survey. *Artificial Intelligence Review*, 43(1):1–54.
- Rubin P (2018) The Wired Guide to Virtual Reality. <https://www.wired.com/story/wired-guide-to-virtual-reality/>. Erstellt: 09.11.2018. Letzter Zugriff: 11.08.2019.
- Shadbolt N, Berners-Lee T, Hall W (2006) The semantic web revisited. *IEEE Intelligent Systems*, 21(3):96–101.
- Shneiderman B (2007) Creativity support tools: Accelerating discovery and innovation. *Communications of the ACM*, 50(12):20–32.
- Simon HA (1973) The structure of ill structured problems. *Artificial intelligence*, 4(3–4):181–201.
- Sjöberg L (2000) Factors in risk perception. *Risk Analysis*, 20(1):1–11.
- Song S, Miller KD, Abbott LF (2000) Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nature Neuroscience*, 3(9):919–926.
- Sowa JF (1976) Conceptual graphs for a data base interface. *IBM Journal of Research and Development*, 20(4):336–357.
- Sowa JF (2006) The challenge of knowledge soup. *Research trends in Science, Technology and Mathematics Education*, Goa, India.
- Stephens ZD, Lee SY, Faghri F, Campbell RH, Zhai C, Efron MJ, Iyer R, Schatz MC, Sinha S, Robinson GE (2015) Big Date: Astronomical or Genomical? *PLOS Biology*, 13(7).
- Thagard P, Shelley C (1997) Abductive Reasoning: Logic, Visual Thinking, and Coherence. Dalla Chiara ML, Doets K, Mundici D, van Benthem J (Hrsg.) *Logic and Scientific Methods: Volume One of the 10th International Congress of Logic, Methodology and Philosophy of Science*, Florence, pp. 413–427. Dordrecht: Springer.
- Thagard P, Stewart TC (2011) The AHA! Experience: Creativity Through Emergent Binding in Neural Networks. *Cognitive Science*, 35(1):1–33.
- Ware C (2013) *Information visualization: perception for design*, 3rd ed.. Waltham, MA: Elsevier.
- Wolverton M, Hayes-Roth B (1994) Retrieving semantically distant analogies with knowledge-directed spreading activation. *Proceedings of the 12th National Conference on Artificial Intelligence*, Menlo Park, CA.
- Yokoi T (1995) The EDR Electronic Dictionary. *Communications of the ACM*, 38(11):42–44.

Teil III

Chatbots

Chatbot & Cognitive Services – Ein Schritt Richtung Automatisierung im User Help Desk der Schweizerischen Post

Julie Mackmood, Daniel Bammert Marty und
Sara D’Onofrio

Zusammenfassung

Der User Help Desk der Informatikabteilung der Schweizerischen Post ist über verschiedene Kommunikationskanäle erreichbar. Das Telefon ist der meist verwendete Kanal, gefolgt von der E-Mail und dem Webformular für die Erfassung von Tickets. Dabei ist der Automatisierungsgrad der Verarbeitung von technischen Problemfällen und Anfragen gering. Mitarbeiter des User Help Desks müssen sich oft mit sich wiederholenden (und allgemeinen) Fragen auseinandersetzen und Aufgaben ausführen, die der Benutzer selbst erledigen könnte. Aus diesem Grund verfolgt die Informatikabteilung der Schweizerischen Post das Ziel, den Automatisierungsgrad mithilfe eines textbasierten Chatbots basierend auf Microsofts Cognitive Services als zusätzlichen Kanal zu erhöhen. Der Chatbot soll den Benutzern jederzeit die Informationen liefern können, die für die Problemlösung relevant sind. Deshalb wird in diesem Artikel diskutiert, ob und welche Cognitive Services von Microsoft für den Einsatz eines textbasierten Chatbots geeignet sind.

Vollständig neuer Beitrag

J. Mackmood · D. Bammert Marty
Die Schweizerische Post, Bern, Schweiz
E-Mail: julie.mackmood@post.ch; daniel.bammert@post.ch

S. D’Onofrio (✉)
Human-IST Institut, Universität Freiburg, Fribourg, Schweiz
E-Mail: sara.donofrio@unifr.ch

1 Der User Help Desk der Schweizerischen Post

Die Schweizerische Post besteht aus drei Konzerngesellschaften: Post CH AG,¹ PostAuto AG² und PostFinance AG.³ Die erste Gesellschaft ist aktiv in den Bereichen Kommunikation, Logistik und Immobilien; die zweite ist im Personenverkehr tätig; die letzte ist eine systemrelevante Bank im Schweizer Finanzsektor. Für alle Bereiche stellt die Informatikabteilung der Post (IT Post) die grundlegenden Informations- und Kommunikationstechnologien bereit.

Bei Anfragen oder Problemen im Bereich der Hard- und Software und IT Services kann sich jede Mitarbeiterin und jeder Mitarbeiter an den User Help Desk wenden. Mögliche Anfragen stellen Zugriffsberechtigungen, neue Bestellungen von Geräten und Hilfestellung in der Installation von Software dar. Probleme hingegen sind Fälle, in denen gewisse Anwendungen nicht wie gewünscht funktionieren, defekte Geräte zu ersetzen sind oder keine Verbindungen zum Internet oder anderen IT Services gemacht werden können. Somit ist der User Help Desk die zentrale Anlaufstelle für Anfragen und Störungen.

Der User Help Desk kann über verschiedene Kommunikationskanäle und in drei Sprachen (Deutsch, Französisch und Italienisch) kontaktiert werden. Die Wahl des Kanals hat Einfluss auf die Synchronität der Kommunikation (synchroner vs. asynchroner Informationsaustausch). Der häufigste genutzte Kanal ist das Telefon (synchron), gefolgt von E-Mail und dem Webformular (d. h. einer strukturierten Anfrageaufzeichnung mittels Formulars auf einer internen Webseite) (asynchron). Aufgrund der Vielfalt der Kanäle und der Mehrsprachigkeit scheint es sinnvoll, sich Gedanken über eine mögliche Automatisierung zu machen, um die Prozesse im User Help Desk effizienter zu gestalten. In diesem Kontext bedeutet Effizienz Zeit- und Kostenaufwände in der Bearbeitung von Anfragen und Problemfällen bestmöglich zu reduzieren und die Agentinnen und Agenten des User Help Desk in ihrer Funktion zu entlasten.

Wie in der Fallstudie *Chatbots for Internal Customer Care* von Mackmood, Bammert Marty und Klopfenstein (2017) beschrieben, ist der Grad der automatisierten Interaktionen zwischen dem User Help Desk und seinen Kunden (d. h. den Mitarbeitenden der Schweizerischen Post) eher gering. Agentinnen und Agenten des User Help Desk müssen sich oft mit repetitiven Fragen auseinandersetzen und Aufgaben ausführen, die der Benutzer selbst erledigen könnte. Häufig werden Fragen gestellt, die ohne Hilfe des User Help Desk (aber mit den richtigen Informationen) alleine gelöst werden könnten, wie zum Beispiel die Aktivierung des internen Wireless Local Area Network (WLAN), das Zurücksetzen eines Passwortes oder die Bestellung von Produkten über den IT-Shop.

Aufgrund der sich wiederholenden Anfragen sowie der Vielzahl angebotener Kommunikationskanäle sind die aktuellen Kosten des User Help Desk und der Zeit-

¹ www.post.ch.

² www.postauto.ch.

³ www.postfinance.ch.

aufwand für die Verarbeitung eines Problems sehr hoch. Pro Tag erhalten die Agentinnen und Agenten des User Help Desk mehr als 600 Serviceanfragen über das Telefon; gekoppelt mit den anderen Kommunikationskanälen sind es fast 1000 Serviceanfragen. Verteilt über den Arbeitstag bedeutet dies, dass in jeder Minute eine bis mehrere Anfragen eingehen, die berücksichtigt werden sollten. Zurzeit beträgt die Lösungsquote des User Help Desk etwa 65 %.⁴ Die Rate ist eher tief, so dass neue Ansätze benötigt werden, um die grosse Anzahl von Anfragen pro Tag bewältigen zu können. Schliesslich geht es um die Sicherstellung der grundlegenden Informations- und Kommunikationstechnologien für alle Bereiche der Schweizerischen Post, damit sie ihre Aufgaben wahrnehmen können.

Ein möglicher Lösungsansatz, um den Bearbeitungsprozess einer Anfrage zu automatisieren, ist die Entwicklung und Implementierung eines textbasierten Chatbots. Für dieses Projekt wurde der Schwerpunkt auf die Cognitive Services von Microsoft⁵ gelegt. Andere Unternehmen, die ebenfalls Chatbot-Produkte anbieten, wurden vorerst nicht näher betrachtet. Die Autoren haben aber Kenntnis über andere Chatbot-Projekte und deren grundlegenden Produkten (vgl. (Stucki et al. 2018, 2020)).

Die Diskussion ob und inwiefern der Einsatz von textbasierten Chatbots (mit Cognitive Services) für den User Help Desk von Nutzen ist, hilft der strategischen Ausrichtung der Schweizerischen Post, besonders dem Aspekt der systematischen Einfachheit, näher zu kommen. Die Strategie baut auf der Vision auf, die wie folgt lautet: *„Einfach mit System – Die Post. Wir verbinden die physische und die digitale Welt und setzen mit unseren Produkten und Systemlösungen neue Massstäbe. Damit vereinfachen wir unseren Kunden das Handeln in einem komplexen Umfeld und helfen ihnen zu mehr Freiräumen.“* (Die Schweizerische Post AG 2016). Nebst dem funktionalen Wert sollte der Chatbot also auch der strategischen Ausrichtung der Schweizerischen Post gerecht werden.

Im folgenden Abschnitt werden die Begriffe *Cognitive Computing* und *Chatbot* erklärt. Dann folgt die Beschreibung des Anwendungsfalls *User Help Desk*. Ausgehend vom Lösungsansatz *textbasierter Chatbot*, wird eine Übersicht angebotener Cognitive Services von Microsoft aufgezeigt. Basierend auf dem Anwendungsfall werden (möglich) passende Kategorien für die Evaluation selektiert und dann mithilfe eines Kriterienkatalogs bewertet. Die Selektion und Evaluation der Cognitive Services erfolgt in Zusammenarbeit mit dem Human-IST Institut der Universität Freiburg (Schweiz). Schliesslich wird kritisch über die Tauglichkeit eines textbasierten Chatbots für die Schweizerische Post diskutiert.

⁴Es handelt sich hierbei um einen Mittelwert aller eingegangenen Anfragen sowie derer Verarbeitungsquote, verteilt über ein Jahr. Diese Zahlen erhielten die Autoren vom Projektcontroller des User Help Desk.

⁵www.microsoft.com.

2 Cognitive Computing

Kognitive Systeme stellen eine neue Ära dar, um Erkenntnisse aus einer Vielzahl von Datentypen und -quellen zu gewinnen. Mit Methoden aus der Text-, Bild- oder Videoverarbeitung (z. B. Natural Language Processing (Kurdi 2016, 2017), Deep Learning (Schmidhuber 2015)) können kognitive Systeme Daten in den jeweiligen Kontext stellen und ein bestimmtes Wissensgebiet verstehen. Sie generieren Hypothesen und liefern mögliche Antworten (mit den entsprechenden Vertrauenswerten (engl. *confidence score*)). Dabei stellt die Entwicklung eines kognitiven Systems einen iterativen Prozess dar, welcher stetig mit jeder verarbeiteten Information lernt (Hurwitz et al. 2015).

Cognitive Computing kann also als eine Erweiterung des Semantischen Computing angesehen werden. Systeme basierend auf Cognitive Computing erkennen nicht nur die Bedeutung der zu verarbeitenden Information, sondern schaffen auch kognitive Verbindungen zwischen den einzelnen Informationsstücken. Solche Systeme werden in der Lage sein, kognitive Fähigkeiten eines Menschen (z. B. lernen, schlussfolgern, sich erinnern) und somit auch die Sprachbeherrschung in einer bestimmten, abgegrenzten Domäne nachahmen zu können. Hier wird besonders das Zusammenspiel von Soft Computing (Methoden, die tolerant gegenüber unpräzisen, unsicheren und vagen Daten sind) und Kognitions- und Lerntheorien (aus der Kognitionspsychologie) hervorgehoben, welche die Entwicklung kognitiver Systeme fördern (D’Onofrio und Portmann 2017; D’Onofrio et al. 2018). Aber nicht nur in der Forschung ist Cognitive Computing ein prominenter Begriff, sondern auch in der Industrie wird es als wichtiges Asset behandelt.

IBM⁶ gilt als eines der ersten Unternehmen, welches den Begriff Cognitive Computing für die Industrie etabliert hat. Virginia Marie Rometty, Chief Executive Officer von IBM, definiert Cognitive Computing als Technologie, die den Systemen die Fähigkeit gibt, mit Nutzern zu interagieren und aus den Interaktionen zu lernen (Ray 2016). Seit der Veröffentlichung von IBM Watson⁷ – das kognitive System von IBM – ist das Portfolio von IBM im Bereich Cognitive Computing gewachsen. Es bietet ein grosses Spektrum an Möglichkeiten an: von der deskriptiven, prädiktiven bis zur präskriptiven Analytik für kognitive Systeme. Diese werden vor allem verwendet, um grosse Datenmengen zu verarbeiten und aus der Interaktion mit Anwendern und Systemen zu lernen (IBM Geschäftsbericht 2017).

Auch andere Unternehmen beschäftigen sich mit dieser Technologie. Microsoft bietet beispielsweise eine breite Palette kognitiver Dienste (sog. *Cognitive Services*) an, die sich in die Kategorien *Knowledge* (Wissensmanagement), *Language* (Textverarbeitung), *Search* (Suchmaschinen), *Speech* (Sprachverarbeitung) und *Vision* (Bildverarbeitung) unterteilen lassen. Eine Reihe von Programmierschnittstellen (sog. *APIs*) stehen den Entwicklern zur Verfügung, um eigene Anwendungen, wie beispielsweise Chatbots, mit kognitiven Fähigkeiten (Texte übersetzen, Bilder erkennen etc.) auszustatten (Microsoft Azure 2019a). Jede API kann einzeln imple-

⁶<https://www.ibm.com>.

⁷<https://www.ibm.com/watson>.

mentiert werden und ist auch kompatibel mit anderen APIs, sei es von Microsoft selbst oder von anderen Unternehmen, wie Google⁸ oder Amazon.⁹

Auch Hewlett Packard (HP) bietet kognitive Programmierschnittstellen zur Optimierung von Diensten an. HP betrachtet Cognitive Computing, inspiriert durch die Funktionsweise des menschlichen Gehirns, als eine massiv parallele, kognitive Architektur, die einfach zu programmieren ist, adaptives Lernen unterstützt und auf Millionen von Kernen skaliert werden kann. Gemäss diesem Unternehmen wird der Bau von intelligenten Systemen aufgrund der kognitiven Komponenten möglich (HP Lab 2012).

Interessanterweise werben Google und Amazon nicht mit dem Begriff Cognitive Computing. Stattdessen verwenden sie den Begriff des maschinellen Lernens (engl. *machine learning*) und bieten ähnliche Produkte und Dienste wie die oben genannten Unternehmen an. Unter *maschinellern Lernen* sind Algorithmen zu verstehen, die aus Daten Muster erkennen, daraus lernen und dadurch die eigene Leistung verbessern können (Hurwitz et al. 2015).

Die kurze (nicht vollständige) Übersicht über Unternehmen, die sich bereits mit Cognitive Computing oder Ähnlichem auseinandersetzen, zeigt auf, dass aus ihrer Sicht Cognitive Computing eine verbesserte Informationsverarbeitung (vom Input bis zum Output) durch (Lern-)Algorithmen in verschiedenen Bereichen (Text, Bild, Video etc.) ermöglicht. Aus Forschungssicht ist der wesentliche Bestandteil des Cognitive Computing jedoch nicht (nur) maschinelles Lernen, sondern eine Kombination mehrerer Methoden und Disziplinen, die zweckgebunden auf und abhängig von der Art und Qualität der Daten eingesetzt werden. Cognitive Computing ist sowohl in der Forschung als auch in der Praxis ein emergentes Phänomen, das viele Definitionsversuche besitzt und dessen Bedeutung erst noch am Entstehen ist.

3 Chatbots

Ein Chatbot ist ein Dialogsystem, das über eine Benutzeroberfläche (z. B. Webseite) aufgerufen wird, und mit dem ein Benutzer in einen Dialog eintreten kann, um sein Anliegen zu lösen. Über natürliche Sprachbefehle kann der Chatbot vom Benutzer aktiviert werden. Gewünschte Informationen können auch vom Chatbot in natürlicher Sprache dem Benutzer zur Verfügung gegeben werden (Stucki et al. 2018, 2020). Daher ist der wesentliche Aspekt eines Chatbots die natürliche Sprache.

Für die Bearbeitung der natürlichen Sprache können Methoden des *Natural Language Processing* eingesetzt werden. Gemäss Stucki et al. (2018) können im textbasierten Umfeld sowohl syntaktische als auch semantische Methoden angewendet werden, die zum einen die Struktur des Inputs bearbeiten und zum andern die Bedeutung des Inputs extrahieren. Für die Syntax kann beispielsweise *Tokenization* angewendet werden, um die Eingabe in einzelne Wortteile zu zerlegen. Eine weitere

⁸<https://cloud.google.com/products/ai>.

⁹<https://aws.amazon.com/machine-learning>.

Methode stellt *Lemmatization* dar, die dafür sorgt, dass jedes Wort in seine lexikalische Grundform zurückgeführt wird. Für die Semantik hingegen kann zum Beispiel *Named Entity Recognition* eingesetzt werden, um einzelne Wortteile generell (z. B. Person, Ort, Beruf) und *Word Sense Disambiguation*, um dieselben Wortteile kontextspezifisch (konkretes Unternehmen, Familie XY etc.) zu klassifizieren (Barrière 2016).

Die Anwendung von Natural Language Processing stellt jedoch nur ein Teilaspekt der Funktionsweise des Chatbots dar. Sobald die Inputs maschinenleserlich vor- und aufbereitet wurden, findet das *Intent-Matching* (entweder regelbasiert oder mithilfe von Ansätzen des maschinellen Lernens) statt. Dort wird der eingegangene Input mit bestehenden Input-Output-Paaren abgeglichen, und die Kombination mit dem höchsten Vertrauenswert wird als bestmöglicher Kandidat zur Verfügung gestellt (d. h. der Benutzer erhält die Antwort, die am besten zum Input passt) (Stucki et al. 2018, 2020).

Wie in Abb. 1 illustriert, beginnt der Dialog mit einer Frage des Benutzers (das zugleich den Input darstellt). Der Input wird daraufhin mithilfe von Natural Language Processing Methoden vor- und aufbereitet. Abhängig vom Kontext, von den verfügbaren Daten und der Sprache wird die bestmögliche Kombination an syntaktischen und semantischen Methoden ausgewählt und verwendet (Stucki et al. 2018). Dieser Schritt ermöglicht den Input maschinenleserlich zu machen, wodurch das Intent-Matching stattfinden kann. Abhängig von der Absicht (sog. *Intent*), das aus dem Abgleich zwischen eingegangenen Input und bestehenden Input-Output-Paaren hervorkommt, holt sich der Chatbot die notwendigen Informationen oder Aktionen

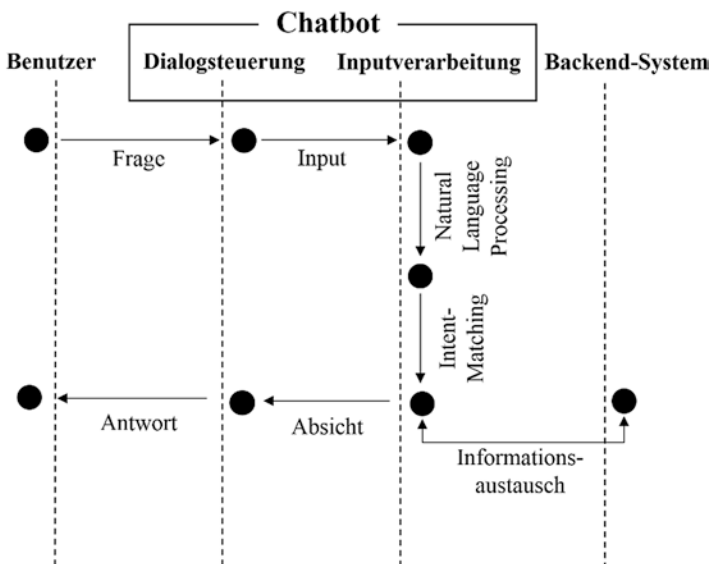


Abb. 1 Interaktionsfluss eines Chatbots in Anlehnung an Stucki et al. (2018, 2020)

aus den Backend-Systemen. Informationen sind zum Beispiel *Frequently Asked Questions* (FAQ) aus einer Wissensdatenbank und Aktionen können Prozessflüsse darstellen, wie beispielsweise die Erstellung eines Tickets. Der Output mit dem höchsten Vertrauenswert wird dem Benutzer zur Verfügung gestellt. Benötigt der Benutzer weitere Informationen, beginnt der Prozess von Neuem (Stucki et al. 2020).

Besonders interessant wird der Einsatz von (textbasierten) Chatbots basierend auf Cognitive Computing in dialogintensiven Bereichen, wie beispielsweise im Kundencenter.

4 Textbasierter Chatbot für den User Help Desk

Für die Automation des Bearbeitungsprozesses von Anfragen im User Help Desk ist die Entwicklung und Implementierung eines textbasierten Chatbots eine mögliche Lösung. Für den ersten Versuch wurde der Schwerpunkt auf die Produkte und Dienste von Microsoft gesetzt. Diese Entscheidung hat unter anderem folgenden Grund: Die einzelnen Cognitive Services stellen Module dar, die nach Bedürfnis ausgewählt und zum Chatbot-Framework hinzugetan werden können. Aufgrund dieser simplen Bedienung lassen sich die verschiedenen Funktionalitäten der Cognitive Services sowie der Chatbot als Ganzes schnell und einfach testen und evaluieren. Das ermöglicht in kurzer Zeit erste Erfahrungswerte im Bereich Chatbot zu sammeln. Daher wurden vorerst andere Chatbot-Anbieter ausgeschlossen. Die Autoren haben aber Kenntnis über andere Chatbot-Produkte.

Das Microsoft Bot-Framework kann als Grundlage verwendet werden (siehe Abb. 2 (Zientara 2016)). Die Grundstruktur des Chatbot-Frameworks ermöglicht es, in kurzer Zeit funktionale und benutzerdefinierte Modelle zu erstellen. Dabei ist *LUIS* (*Language Understanding Intelligence Service*)¹⁰ die erste grundlegende Komponente der kognitiven APIs von Microsoft im Bereich der Textverarbeitung. Dieser Dienst ermöglicht eingehende Informationen seitens der Nutzer zu verarbeiten und passend darauf zu reagieren (z. B. Fragen beantworten, eine Transaktion auslösen). Zusätzlich kann die Entwicklung des Chatbots und insbesondere seine Dialogstruktur (z. B. Message Parsing, Konversationsstruktur und Formularfluss mit Eingabevalidierung) mit der *Standard Web API* von Microsoft realisiert werden. Der *Bot Connector* ermöglicht die Kommunikation zwischen dem entwickelten Bot und dem Benutzer. Dazu gehören Bot-Registrierung (d. h. Veröffentlichung des Bots), Konnektivität (d. h. Verbindung zwischen Bot und Backend-Systemen sowie zwischen Bot und Benutzern) und Message-Routing (d. h. Umleitung von Nachrichten an den entsprechenden Ziel-Endpunkten) mit einer Reihe von Kommunikationskanälen. Sein Vorteil ist, dass er den Bot mit *LUIS* und anderen Cognitive Services, die im nächsten Abschnitt eingeführt werden, verbinden kann. Der Zugriff kann über diverse Kanäle erfolgen, beispielsweise via Webchat, Skype, Facebook Messenger, E-Mail oder SMS.

¹⁰<https://www.luis.ai/home>.

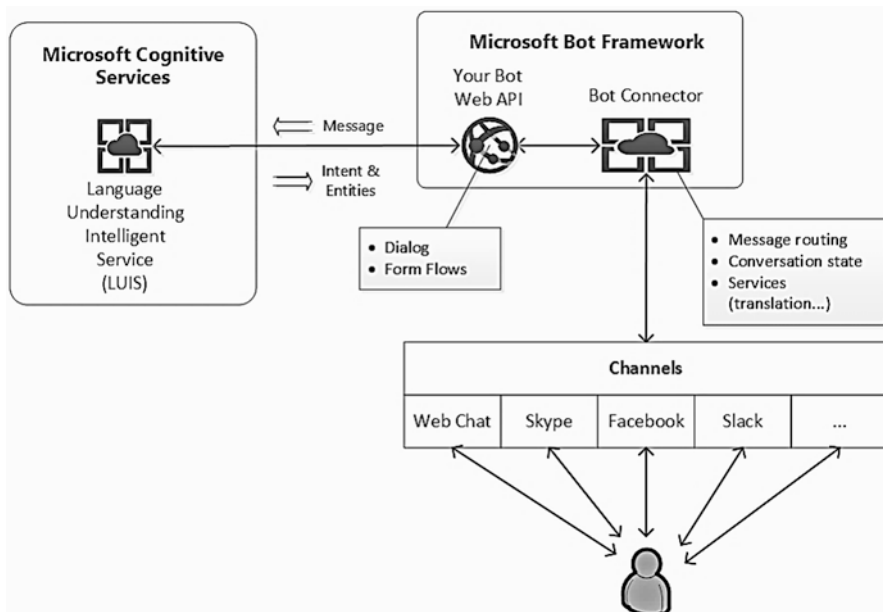


Abb. 2 Azure Bot-Framework (Zientara 2016)

5 Cognitive Services von Microsoft

Microsoft bietet eine Reihe verschiedener Cognitive Services an, um Kommunikationsdienste mittels maschinellen Lernens zu erweitern. Diese Cognitive Services können in *Knowledge*, *Language*, *Search*, *Speech* und *Vision* kategorisiert werden, wie in Abb. 3 illustriert (Microsoft Azure 2019a).

Die nachfolgenden Abschnitte führen die einzelnen APIs grob ein, um ein Verständnis der einzelnen Funktionen zu bekommen. Detailliertere Informationen sind auf den jeweiligen Webseiten von Microsoft verfügbar. Die Informationen über die verschiedenen Cognitive Services wurden im Juli 2018 erhoben. Die folgende Aufzählung ist daher nicht abschliessend.

5.1 Knowledge-API

In der Kategorie *Knowledge* wird nur eine API angeboten: den *QnA Maker* (Microsoft Azure 2019b). Der webbasierte Dienst repräsentiert eine Wissensdatenbank mit Fragen und Antworten. Diese kann selbst mit Informationen (z. B. FAQ, Dokumentationen, URL) befüllt werden. Der *QnA Maker* dient zum einen als Informationsspeicher und zum andern als Antwortgenerator. Beim Eingang einer Frage wird diese mit gespeicherten Frage-Antwort-Paaren abgeglichen und die bestmögliche Antwort wird dann über ein Interface wiedergegeben.

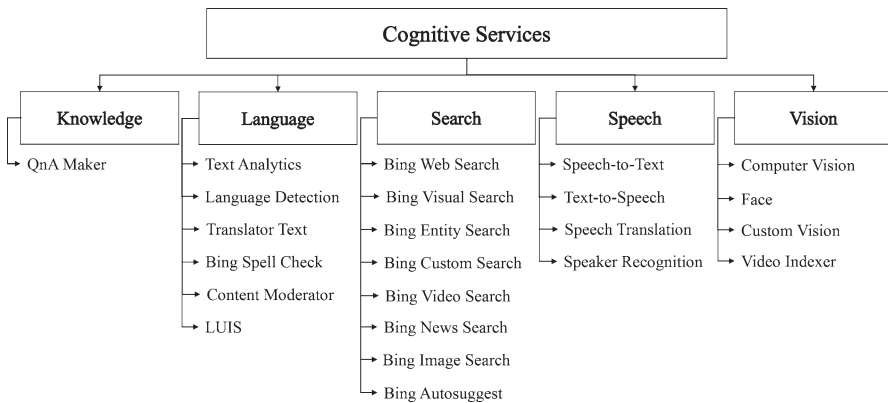


Abb. 3 Cognitive Services – Ein Überblick (eigene Abbildung)

5.2 Language-API

In der Kategorie *Language* werden mehrere APIs angeboten (Microsoft Azure 2019b). Die API *Text Analytics* stellt ein Bündel dar, welches aus den folgenden Funktionen besteht: *Sentiment Analysis*, *Key Phrase Extraction*, *Language Detection* und *Entity Linking*. Bei der Funktion *Sentiment Analysis* wird mithilfe Klassifizierungstechniken analysiert, welchen Stimmungsscore (1 = positiv, 0 = negativ) die Texteingabe besitzt. Bei der *Key Phrase Extraction* werden die wichtigsten Zeichenketten (meist in Form von Schlüsselwörtern) aus der Eingabe zurückgegeben, welche helfen, die Absicht der Eingabe (besser) zu identifizieren. *Language Detection* wird verwendet, um die Sprache des eingegebenen Textes zu erkennen. Dies ist besonders in mehrsprachigen Anwendungen sowie für Übersetzungszwecke von Vorteil. Die letzte Funktion, *Entity Linking*, welche zurzeit nur in einer Vorschauversion verfügbar ist, ermöglicht die Identität einer Entität zu bestimmen und sie korrekt in Kontext zu setzen (bspw. die Entscheidung, ob *Bank* für eine Sitzmöglichkeit oder für ein Finanzinstitut im eingegebenen Text steht).

Eine weitere Language-API ist *Translator Text*. Dabei handelt es sich um einen cloud-basierten, maschinellen Übersetzungsdienst, welcher auf neuronalen Netzen basiert. Die neuronalen Netze ermöglichen den vollen Kontext einer Texteingabe zu nutzen, um Wörter zu übersetzen (anstelle von einzelnen Wörtern, wie es früher das statistische Übersetzungsmodell gemacht hatte).

Bing Spell Check korrigiert Rechtschreibfehler oder verbessert grammatikalische Punkte (z. B. Gross- und Kleinschreibung, einfache Interpunktion), um die Dokumentenerstellung oder den Abgleich zwischen Input und Output zu verbessern. Bei dieser API können zwei Modi ausgewählt werden: *Proof* und *Spell*. Im Modus *Proof* wird die Gross- und Kleinschreibung, die grundlegende Interpunktion sowie andere grammatikalische Funktionen überprüft. Im Modus *Spell* konzentriert sich die Überprüfung mehrheitlich auf die Gross- und Kleinschreibung.

Grammatikalische Fehler werden nur selten erkannt. Im deutschsprachigen Raum ist momentan nur der Modus *Spell* verfügbar.

Die *API Content Moderator* wird angewendet, um anstössige, unerwünschte und riskante Inhalte (z. B. pornographische oder rassistische Inhalte) in Bildern, Text und Videos zu vermeiden. Dabei handelt es sich um eine präventive Vorkehrung sowie eine Möglichkeit, die generierten Inhalte fortlaufend kontrollieren zu können. Diese API ist ebenfalls in der Kategorie *Vision* vorzufinden.

Schliesslich und wie im vorherigen Kapitel erläutert, ist *LUIS* eine *Language-API*, welche die Eingabe untersucht, daraus die Absicht extrahiert und darauf die entsprechende Antwort generiert (falls ein erfolgreicher Abgleich zwischen Input und Output möglich ist). Mit den Daten, die in *LUIS* gespeichert und aktualisiert werden, kann der Bot die Anfragen des Benutzers analysieren, dessen Absichten (d. h. Fragen, Probleme) verstehen und die bestmögliche Antwort zu den jeweiligen Eingaben generieren. Gemäss Microsoft lernt *LUIS* im Laufe der Zeit, welche Antworten passend und richtig sind. Das heisst, die Vertrauenswerte zwischen Absicht (extrahiert aus der Frage) und der Antwort werden angepasst. So kann *LUIS* ein eigenes Muster erstellen und die Verarbeitung von Benutzereingaben ständig verbessern.

5.3 Search-APIs

In der Kategorie *Search* bietet Microsoft sieben APIs im Bereich *Suchfunktion* und eine weitere API namens *Autosuggest* an, alle basierend auf der Suchmaschine Bing¹¹ (Microsoft Azure 2019b). Die Search-APIs *Bing Web Search*, *Bing Visual Search*, *Bing Custom Search*, *Bing Entity Search*, *Bing Video Search*, *Bing News Search* und *Bing Image Search* üben dieselbe Funktion aus (d. h. etwas suchen) und unterscheiden sich in der Granularität *wonach* sie suchen. Dabei ist *Bing Web Search* die API, die basierend auf den eingegebenen Schlüsselwörtern das Internet nach allem Möglichen wie Webseiten, Bilder, Videos und weiteres durchsucht. Die anderen Search-APIs fokussieren sich hingegen auf den Typ des gesuchten Inhalts (z. B. *Bing Video Search* sucht nur nach Videos, *Bing Image Search* ist zuständig für Bilder). *Bing Custom Search* ist die API, bei welcher die Suchfunktion personalisiert werden kann. Das heisst, nebst dem Typ der Datei lässt sich hierbei auch bestimmen, was gesucht werden kann/soll und was nicht. Mit anderen Worten handelt es sich hier um eine personalisierte, kontextspezifische Search-API.

Bing Autosuggest ist eine API in der Kategorie *Search*, die automatisiert die Eingabe mit vollständigen Suchvorschlägen ergänzt, und somit die Suche für den Benutzer vereinfacht.

¹¹<https://www.bing.com>.

5.4 Speech-APIs

In der Kategorie *Speech* stellt Microsoft verschiedene APIs zur Verfügung, die als Ergänzung zu den Language-APIs angewendet werden können, sobald die gesprochene Sprache zum Zug kommt (Microsoft Azure 2019b). Dabei sind besonders zwei APIs von Nutzen: *Speech to Text* und *Text to Speech*. Die erste wandelt gesprochene Sprache in Text um, während die zweite das Gegenteil tut (d. h. geschriebener Text wird in gesprochenen Text konvertiert). Beide gehören zum Bündel der *Speech Services*, das wie *Text Analytics* aus den Language-APIs als Paket vermarktet wird. Hinzu kommt die *API Speech Translation*, welche wie die *API Translation Text* sprachliche Texteingaben automatisch basierend auf neuronalen Netzen über einen Cloud-Server übersetzt.

Eine weitere API, die in der Kategorie *Speech* angeboten wird, ist *Speaker Recognition*. Diese kann für die Identifizierung und Verifizierung von Personen eingesetzt werden. Dabei verwendet die API die menschliche Stimme als eindeutiges Charakteristikum und versucht mithilfe in der Wissensdatenbank gespeicherten Stimmabdrücken zu identifizieren, wer nun spricht.

5.5 Vision-APIs

In der Kategorie *Vision* werden APIs angeboten, die den Schwerpunkt auf die Bearbeitung und Analyse von Bildern und Videos setzt (Microsoft Azure 2019b). Die *API Computer Vision* ist hierbei ein Set von Algorithmen, welche zum einen die eingehenden Bilder verarbeitet und zum andern daraufhin dem Nutzer Informationen bezüglich des Bildes zurückgibt (bspw. Meta-Daten). Dabei kann es unter anderem erkennen, ob die Inhalte jugendfrei sind. *Face* hingegen konzentriert sich ausschliesslich auf Gesichter. Dabei kann sie gesichtsbezogene Attribute, wie Geschlecht, Gesichtsbehaarung und Brille, extrahieren oder das Gesicht für die Authentifizierung einer Person nehmen, wobei mindestens ein Bild der zu verifizierenden Person in der Wissensdatenbank vorhanden sein muss. Passend dazu bietet die *API Custom Vision* die Möglichkeit, eigene Klassifizierer (engl. *classifier*) für Bilder zu erstellen und somit die Zuordnung der Bilder zu personalisieren.

Video Indexer besteht aus einem Bündel von Cognitive Services (mehrheitlich aus den Bereichen *Vision* und *Speech*), um unter anderem in Videos Gesichter, Stimmen und die gesprochene Sprache zu erkennen und diese zu identifizieren, Beschriftungen für visuelle Objekte bereitzustellen, die Stimmung innerhalb eines Videos einzuordnen und vieles mehr. Die breite Vielfalt der Funktionalitäten erlaubt die Bearbeitung und Analyse von Videos von verschiedenen Seiten anzugehen und somit unterschiedliche Aspekte zu adressieren.

5.6 Erstes Fazit

Zusammenfassend lässt sich sagen, dass die Funktionen der Cognitive Services breit gefächert sind. Die *Knowledge-API* ermöglicht sowohl die Abbildung als auch die Speicherung von Informationen (zurzeit in Form von Frage-Antwort-Paaren).

Language-APIs ermöglichen die Verarbeitung natürlicher Sprache. Mit vordefinierten Skripten werden die Eingaben des Benutzers analysiert, klassifiziert und mit geeigneten Antworten wird darauf reagiert. Nebst der eigentlichen Extraktion der Absicht einer Benutzereingabe bieten *Language-APIs* von automatischen Rechtschreibkorrekturen, über automatische Übersetzungen bis hin zur Erkennung der Stimmung eine Vielfalt an Funktionalitäten an, um die textbasierte Eingabe zu verarbeiten. *Search-APIs* von Bing ermöglichen den Abruf von Milliarden von Webseiten, Bildern, Videos und Nachrichten und zielen darauf ab, nur relevante Inhalte an den Benutzer zurückzugeben. *Speech-APIs* wandeln gesprochene Wörter in Text um, um sie dann weiter mit den *Language-APIs* verarbeiten zu können. Nebst der Verarbeitung bieten diese APIs Verifizierungs- und Identifikationsdienste an, wobei die Sprache als eindeutiges Merkmal für die Erkennung der Person angewendet wird. Schliesslich bieten *Vision-APIs* Funktionalitäten für die Bearbeitung und Analyse von Bildern und Videos an, sei es das Bearbeiten von Inhalten, Erkennen von Gesichtern oder Objekten oder Erstellen von Bildbeschriftungen. Abhängig von der Anwendung, den Interessengruppen und der Qualität der verfügbaren Daten sollten entsprechende APIs eingesetzt werden. Aus diesem Grund folgt im nächsten Abschnitt die Selektion sowie die Bewertung der Cognitive Services.

6 Analyse der Cognitive Services

Microsoft's Cognitive Services werden in diesem Kapitel anhand eines Kriterienkatalogs bewertet und auf ihre Tauglichkeit im Anwendungsfall User Help Desk evaluiert. Da es sich in diesem Anwendungsfall um einen textbasierten Chatbot handelt, werden die Dienste für die gesprochene Sprache und die Dienste für Bilder und Videos (d. h. die APIs der Kategorien *Speech* und *Vision*) nicht näher angeschaut. Die APIs des Bereichs *Search* werden ebenfalls nicht näher angeschaut, weil ihre Funktion auf der Suchmaschine Bing basieren und sich dies mit den momentan geltenden Datenschutzbestimmungen nicht verträgt.

6.1 Kriterienkatalog

Um die Cognitive Services möglichst genau und objektiv zu vergleichen, wurde ein Katalog von 28 Kriterien, eingeteilt in acht Kategorien, entwickelt. Der Katalog wurde zuerst mit einer Online-Literaturanalyse erstellt (u. a. basierend auf den Tipps von Softguide (2018) in Anlehnung an Hänggeli und Zurwerra (2017)). Dann wurde der Katalog iterativ mit den Mitarbeitern des Chatbot-Projektes ausgearbeitet, wobei jeweils im Team entschieden wurde, welche Kriterien passend für den Anwendungsfall und für das Unternehmen sind und für die Evaluation der Cognitive Services eingesetzt werden sollen. Anhand der folgenden Kriterien wurde beurteilt, ob die Cognitive Services im User Help Desk eingesetzt werden können. Die letzte Spalte zeigt die Gewichtung der Wichtigkeit pro Kriterium; pro Block wird eine Gewichtung von 100 % vergeben. Die Gewichtungen sind von den Autoren definiert worden.

Tab. 1 Geschäftsanforderungen

Personalisierung		100 %
Anpassbarkeit im Allgemeinen	Die API kann angepasst werden (z. B. Begriffe der Schweizerischen Post).	30 %
Corporate Identity	Die API kann der Corporate Identity der Schweizerischen Post angepasst werden.	50 %
Mehrsprachigkeit	Die API unterstützt mehrere europäische Sprachen.	20 %
Benutzerfreundlichkeit		100 %
Benutzerfreundlichkeit	Die API ist einfach und intuitiv zu bedienen.	50 %
Benutzerführung	Dokumentationen helfen dem Anwender, die API korrekt zu nutzen.	50 %
Preismodell		100 %
Kostenlose Version	Die API hat eine kostenlose Version.	30 %
Kostenpflichtige Version	Die API hat eine Standardfunktion mit Kosten.	30 %
Testversion	Die API kann kostenlos oder zu einem reduzierten Preis getestet werden.	40 %

Zunächst werden Kriterien für die Aspekte der Personalisierung, Benutzerfreundlichkeit und Kosten eingeführt, die als wichtige, grundlegende Geschäftsanforderungen gelten. Sowohl die User Help Desk Agentinnen und Agenten als auch die Geschäftspartner müssen in der Lage sein das System (inkl. den Cognitive Services) effektiv nutzen zu können (Tab. 1).

Die technischen Anforderungen wurden in vier Kategorien zusammengefasst, um festzustellen, welche Programmierkenntnisse erforderlich sind, welche Zugangspunkte und Kooperationsmöglichkeiten zur Verfügung stehen, wie die Daten verwaltet werden und in welchem Umfang technische Unterstützung angeboten wird. Diese Kriterien sind für die Programmierer, Betreiber und Anwender dieser kognitiven APIs relevant (Tab. 2).

Schliesslich wird geprüft, ob und inwieweit die kognitiven APIs für die Schweizerische Post geeignet sind. Ziel ist es herauszufinden, ob die APIs für den Anwendungsfall anwendbar sind, ob sie drei von vier Landessprachen der Schweiz (d. h. Deutsch, Französisch und Italienisch) unterstützen, ob sie überhaupt in die Systemlandschaft integriert werden können und nicht zuletzt, ob sie die strategische Ausrichtung der Post „*Einfach mit System – Die Post*“ unterstützen (Tab. 3).

Die in der Abb. 4 illustrierten fünfstufige Skala dient der Bewertung der APIs. Mit einer Punktzahl von 5 erfüllt die API das Kriterium vollständig. Mit einer Punktzahl von 1 erfüllt die API das Kriterium überhaupt nicht. Alle Werte dazwischen stellen einen gewissen Grad der Erfüllung oder Nichterfüllung des Kriteriums dar. Die Option *nicht anwendbar* (N/A) wird für den Fall hinzugefügt, dass ein Kriterium auf eine API nicht anwendbar ist. Es wurde bewusst darauf verzichtet, eine Bewertung mit nur zwei Optionen durchzuführen (z. B. *erfüllt* und *nicht erfüllt*). Die breitere Abstufung ermöglicht eine bessere Beurteilung der Eignung der APIs. (Hängeli und Zurwerra 2017).

Die Cognitive Services wurden je von zwei Autoren gleichermassen (nach den vorgegebenen Kriterien) und unabhängig voneinander bewertet. Mit der fünfstufigen Ratingskala konnten die Testpersonen für jede API entscheiden, inwieweit sie

Tab. 2 Technische Anforderungen

Programmierung		100 %
Implementierung	Die Implementierung der API ist schnell und einfach.	40 %
Programmiersprache	Die API unterstützt verschiedene Programmiersprachen.	30 %
Autonomie	Die API arbeitet eigenständig, unabhängig von anderen APIs.	20 %
Upgrade	Bei Versionsänderungen kann die API aktualisiert werden.	10 %
Zugriff		100 %
Asynchronität	Die API kann unabhängig von einer anderen Microsoft-API ausgeführt werden.	30 %
Synchronität	Die API kann gleichzeitig mit anderen APIs verwendet werden.	30 %
Kollaboration	Die API kann mit anderen APIs als die von Microsoft verwendet werden.	20 %
Internet-Basis	Zugriff auf die API über Intranet und Internet möglich.	15 %
Mobiler Zugriff	Zugriff auf die API von einem mobilen Endgerät aus ist möglich.	5 %
Daten		100 %
Import	Die API unterstützt den Import von Daten.	33,3 %
Export	Die API unterstützt den Export von Daten.	33,3 %
Protokollierung	Die API protokolliert Benutzereingaben.	33,3 %
Hilfestellung		100 %
Online Support	Unterstützung in Form von Dokumentation, Tutorials und weiteres wird angeboten.	30 %
Lokaler Support	Bei Problemen steht ein lokaler Support zur Verfügung (z. B. in der Schweiz).	30 %
Supportsprache	Mehrsprachiger Support (u. a. für Europa).	10 %
Best Practices	Best Practices sind verfügbar (d. h. bestehende Implementierungen, die wiederverwendet werden können).	30 %

Tab. 3 Unternehmensspezifische Anforderungen

Die Schweizerische Post		100 %
Tauglichkeit für den User Help Desk	Die API unterstützt die Entwicklung eines textbasierten Chatbots.	25 %
Schweizer Landessprachen	Die API unterstützt die Schweizer Landessprachen (Deutsch, Französisch, Italienisch).	10 %
Technische Integration	Die API kann in die technische Infrastruktur und Architektur der Post integriert werden.	40 %
Strategie der Schweizerischen Post	Die API unterstützt die strategische Ausrichtung der Schweizerischen Post.	25 %

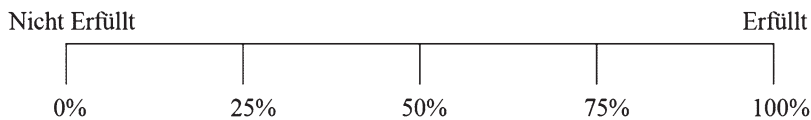


Abb. 4 Bewertungsskala

das Kriterium erfüllt (von 1 bis 5). Die Bewertungen wurden miteinander verglichen und pro Kriterium wurde der Mittelwert berechnet. Anschliessend wurde der Mittelwert mit der jeweiligen Gewichtung multipliziert. Schliesslich zeigt die Summe aller gewichteten Mittelwerte, wie gut die API nach diesem Kriterienkatalog abgeschnitten hat. Die maximale Punktzahl aller Kriterien, wenn eine API alle Anforderungen vollständig erfüllen würde, beträgt 40. Die Bewertung beziehungsweise Evaluation der kognitiven APIs wurde im Juli 2018 vorgenommen.

Bei den folgenden Bewertungen wird der Mittelwert für jedes Kriterium (zwischen 1 und 5) pro API ohne Berücksichtigung der Gewichtung angezeigt. Die Kursivschrift zeigt dann die Summe aller Bewertungen, die mit dem jeweiligen Gewicht multipliziert wurden. Schliesslich fasst die Gesamtzahl alle gewichteten Zwischensummen zusammen.

6.2 Bewertung der Knowledge-API und Language-APIs

Die APIs der Kategorien *Knowledge* und *Language* wurden anhand des vordefinierten Kriterienkatalogs bewertet. Die Einzelbewertungen pro Kriterium werden in den Abb. 5 und 6 angezeigt. Die Gesamtpunktzahl der zurzeit einzigen *Knowledge*-API beträgt 32,17. Die Gesamtpunktzahl der *Language*-APIs variiert zwischen 25,05 und 34,09.

6.3 Zweites Fazit

Abb. 7 zeigt das Ranking der Cognitive Services (aus den Bereichen *Knowledge* und *Language*) basierend auf den gegebenen Bewertungen (d. h. Gesamtpunktzahl aller Punkte multipliziert mit den jeweiligen Gewichten).

Wie in Abb. 7 ersichtlich, sind nicht alle Cognitive Services, die für die Evaluation selektiert wurden, für den Anwendungsfall User Help Desk geeignet. Hierbei ist wichtig zu erwähnen, dass die Ergebnisse auf dem aktuellen Kenntnisstand (Juli 2018) und dem gegebenen Anwendungsfall (User Help Desk) basieren. Andere Faktoren, wie zum Beispiel technologische Fortschritte bei Microsoft oder andere Zielsetzungen im Anwendungsfall, können natürlich die zukünftige API-Integration beeinflussen.

Knowledge API	QnA Maker
Kriterien	
Personalisierung	4,25
Anpassbarkeit i.A.	4,50
Corporate Identity	4,00
Mehrsprachigkeit	4,50
Benutzerfreundlichkeit	4,25
Benutzerfreundlichkeit	4,00
Benutzerführung	4,50
Preismodell	3,75
Kostenlose Version	3,50
Kostenpflichtige Version	5,00
Testversion	3,00
Programmierung	3,75
Implementierung	3,50
Programmiersprache	3,50
Autonomie	4,50
Upgrade	4,00
Zugriff	4,05
Asynchronität	3,00
Synchronität	4,00
Kollaboration	5,00
Internet-Basis	5,00
Mobiler Zugriff	4,00
Daten	4,00
Import	5,00
Export	3,50
Protokollierung	3,50
Hilfestellung	3,65
Online Support	4,50
Lokaler Support	2,00
Supportsprache	5,00
Best Practices	4,00
Die Schweizerische Post	4,48
Tauglichkeit für UHD	5,00
Landessprachen	5,00
Technische Integration	4,00
Strategie der Post CH AG	4,50
Total	32,17

Abb. 5 Bewertung der Knowledge-API

Language APIs Kriterien	Text Sentiment Analysis	Key Phrase Extraction	Entity Linking	Language Detection	Translator Text	Bing Spell Check	Content Moderator	LUIS
Personalisierung	2,70	2,05	2,80	1,70	5,00	1,90	4,20	4,55
Anpassbarkeit i.A.	3,00	1,50	2,50	1,00	5,00	1,50	4,50	5,00
Corporate Identity	2,00	2,00	2,50	1,00	5,00	1,50	4,50	4,50
Mehrsprachigkeit	4,00	3,00	4,00	4,50	5,00	3,50	3,00	4,00
Benutzerfreundlichkeit	3,25	3,25	3,25	4,00	3,75	3,00	4,00	4,25
Benutzerfreundlichkeit	2,50	2,50	2,50	3,00	3,50	2,50	3,50	4,00
Benutzerführung	4,00	4,00	4,00	5,00	4,00	3,50	4,50	4,50
Preismodell	2,60	2,60	1,50	2,30	3,95	2,95	2,55	4,40
Kostenlose Version	2,00	2,00	1,50	2,00	3,50	1,50	1,50	5,00
Kostenpflichtige Version	4,00	4,00	1,50	3,00	5,00	5,00	5,00	5,00
Testversion	2,00	2,00	1,50	2,00	3,50	2,50	1,50	3,50
Programmierung	2,95	4,00	3,70	4,20	4,10	3,80	3,75	4,00
Implementierung	3,00	4,00	3,50	4,00	4,00	3,50	3,50	4,00
Programmiersprache	3,50	4,00	4,00	4,00	4,00	4,00	3,50	4,00
Autonomie	1,50	4,00	3,50	5,00	4,50	4,00	4,50	4,00
Upgrade	4,00	4,00	4,00	4,00	4,00	4,00	4,00	4,00
Zugriff	3,30	4,45	3,90	4,80	4,15	3,90	4,05	4,15
Asynchronität	2,00	4,50	3,50	5,00	4,00	3,50	5,00	4,00
Synchronität	3,50	4,50	4,00	4,50	4,00	4,00	3,00	4,00
Kollaboration	3,50	4,00	3,50	5,00	4,00	3,50	3,50	4,00
Internet-Basis	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00
Mobiler Zugriff	4,00	4,00	4,00	4,00	4,00	4,00	4,00	4,00
Daten	4,00	4,00	4,00	4,00	2,83	3,00	3,83	4,67
Import	5,00	5,00	5,00	4,50	5,00	5,00	5,00	5,00
Export	4,50	4,50	4,50	4,50	2,00	2,50	4,00	4,50
Protokollierung	2,50	2,50	2,50	3,00	1,50	1,50	2,50	4,50
Hilfestellung	3,20	3,20	3,20	2,90	3,35	3,20	3,35	3,65
Online Support	4,00	4,00	4,00	3,00	4,00	4,00	3,50	4,50
Lokaler Support	2,00	2,00	2,00	2,00	2,00	2,00	2,00	2,00
Supportsprache	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00
Best Practices	3,00	3,00	3,00	3,00	3,50	3,00	4,00	4,00
Die Schweizerische Post	3,05	4,50	3,60	2,60	3,23	4,18	3,38	4,43
Tauglichkeit für UHD	1,50	5,00	2,50	1,00	1,00	5,00	3,00	5,00
Landessprachen	4,50	4,00	4,50	5,00	5,00	4,00	4,00	4,50
Technische Integration	4,00	4,00	3,50	4,00	4,00	3,50	4,00	4,00
Strategie der Post CH AG	2,50	5,00	4,50	1,00	3,50	4,50	2,50	4,50
Total	25,05	28,05	25,95	26,50	30,36	25,92	29,11	34,09

Abb. 6 Bewertung der *Language-APIs***Abb. 7** Ranking der Cognitive Services

Kategorie	API	Total
Language	LUIS	34,09
Knowledge	QnA Maker	32,17
Language	Translator Text	30,36
Language	Content Moderator	29,11
Language	Key Phrase Extraction	28,05
Language	Language Detection	26,50
Language	Entity Linking	25,95
Language	Bing Spell Check	25,92
Language	Text Sentiment Analysis	25,05

7 Diskussion

Ein möglicher Ansatz einen User Help Desk zu automatisieren, ist ein textbasierter Chatbot als zusätzlichen Kommunikationskanal anzubieten. Dabei wird das Ziel verfolgt, dem Benutzer die Möglichkeit zu geben, schnell und einfach allgemeine Anfragen zu beantworten und Problemstellungen selbst zu lösen. Es ist daher essenziell zu bestimmen, welche Funktionen der Chatbot benötigt, um das gesetzte Ziel zu erfüllen. Beginnend mit der eintreffenden Eingabe, muss der Chatbot in der Lage sein, die Informationen zu verarbeiten, sie in den richtigen Kontext einzuordnen, die wesentliche Absicht zu extrahieren und darauf die richtige Antwort zu generieren (vgl. Interaktionsfluss eines Chatbots (Stucki et al. 2018, 2020)). Eine Möglichkeit, einem Chatbot solche (kognitiven) Funktionen bereitzustellen, ist es, die von Microsoft angebotenen Cognitive Services, welche im Abschn. 5 präsentiert wurden, zu nutzen. Basierend auf der durchgeführten Analyse mit einer Vielzahl von Kriterien wurden die kognitiven APIs im Hinblick auf Eignung und Machbarkeit in Bezug auf den User Help Desk bewertet.

Die API mit der höchsten Punktzahl ist *LUIS* aus der Kategorie *Language*. Hinsichtlich des Anwendungsfalls ist es naheliegend, dass eine API, welche die natürliche Sprache bearbeitet, am besten abschneidet. Sie ist essenziell für die Entwicklung eines Interaktionsflusses zwischen Mensch und Maschine. Dabei können erstellte Modelle (d. h. mit einer Auswahl an *Intents* und *Entities*) angewendet werden, welche sich auf den eigenen Anwendungsfall anpassen lassen, oder von Anfang an selbst Modelle entwickelt werden, wobei die *Intents* (Kategorien von Absichten) und *Entities* (Merkmale von Intents) selbst eingepflegt werden müssen. Der auf *Machine Learning* basierte Dienst extrahiert aus einer Benutzereingabe die mögliche Absicht (*Intent*) und generiert daraus die bestmögliche Antwort. Die Antwort ist im System hinterlegt, wodurch *LUIS* nur wissen muss, zu welchem *Intent* die Eingabe gehört, um die dazu gespeicherte Antwort wiederzugeben. Eine weitere nützliche Funktion ist das Logging: Eingaben seitens der Benutzer werden gespeichert und dadurch kann das Modell kontinuierlich trainiert werden. Im Anwendungsfall User Help Desk musste ein eigenes Modell entwickelt werden, da die verfügbaren Modelle die Sprache Deutsch nicht enthalten. Das heisst, es gibt zurzeit keine vortrainierten *Intents* und *Entities* in deutscher Sprache, die übernommen werden können. Aber eigene deutsche *Intents* und *Entities* werden von *LUIS* unterstützt.

LUIS kann autonom agieren, ohne eine Abhängigkeit zu anderen APIs zu besitzen. Diese API kann aber auch mit anderen Cognitive Services oder APIs von anderen Unternehmen kombiniert werden, um die gesamte Interaktion zwischen Mensch und Maschine zu verbessern. Eine Erweiterung und Ergänzung zu *LUIS* stellt *QnA Maker* dar, welcher in dieser Analyse den zweiten Platz hält.

QnA Maker ist für das Management von Wissen zuständig und fungiert als Wissensdatenbank für den Chatbot. *LUIS* selbst kann ebenfalls Frage-Antwortpaare hinterlegen. Jedoch ist dies nur in den sogenannten *Intents* möglich, welche zum einen den Administrator zwingt, Kategorien von Absichten zu bilden und darin Fragen zu formulieren. Zum andern sind sie in ihrer Anzahl limitiert und können bei

steigender Anzahl unübersichtlich werden, wodurch die Benutzerfreundlichkeit der API sinkt. Beim *QnA Maker* gibt es keine vorgefertigten Modelle. Dort werden alle Daten in Form von Fragen und Antworten selbst eingepflegt, wobei keine Programmierkenntnisse benötigt werden.

QnA Maker agiert ebenfalls autonom, kann jedoch keinen Dialog mit einem Benutzer führen. Die Funktion besteht rein darin, die eintreffende Eingabe mit den abgespeicherten Frage-Antwort-Paaren abzugleichen, und die Antwort wiederzugeben, welche am meisten mit der Eingabe übereinstimmt. Deshalb macht es Sinn, den *QnA Maker* mit *LUIS* zu kombinieren, da *LUIS* in der Lage ist, einen Dialog mit dem Benutzer zu führen. Zugleich können vordefinierte Absichten (wie z. B. Begrüßungsfloskeln oder automatische Ticketgenerierung) über *LUIS* abgerufen werden, und alles andere kann mittels Kategorie *Non-Intents* zum *QnA-Maker* weitergeleitet werden. Der *QnA-Maker* antwortet dann auf die Fragen und liefert die Antwort, welche den höchsten Vertrauenswert (basierend auf den eingegebenen Schlüsselwörtern) generiert.

Als nächstes folgen die APIs *Translator Text* und *Content Moderator*. Beide haben in der Bewertung gut abgeschnitten, sind jedoch für den momentanen auf Deutsch fokussierten Anwendungsfall nicht relevant. Falls in naher Zukunft weitere Sprachen im Chatbot implementiert werden sollen, so ist der *Translator Text* sicherlich in Betracht zu ziehen. Bei steigender Komplexität des Inhaltes könnte in Zukunft auch ein *Content Moderator* Sinn machen.

Die *Key Phrase Extraction*-API, mit der Funktion relevante Schlüsselwörter aus der Texteingabe zu extrahieren, kann je nach Zweck sinnvoll sein, um *LUIS* und *QnA Maker* zu unterstützen. Sie kann zum Beispiel zu Beginn eingesetzt werden, um die Absichtsextraktion in *LUIS* beziehungsweise *QnA Maker* schneller und womöglich besser zu machen, oder am Schluss, falls beide Cognitive Services zu keiner möglichen Antwort kommen und sie den Kunden zur Agentin oder zum Agent weiterleiten müssen. So kann der Agentin beziehungsweise dem Agenten bereits im Voraus mitgeteilt werden, welche Schlüsselwörter relevant sind.

Der Einsatz von *Language Detection* wird interessant, sobald der Chatbot mehrere Sprachen verarbeiten muss/soll. Dabei wird es hilfreich sein, im Voraus zu bestimmen, welche Sprache der User verwendet, um die richtigen Datensätze abzurufen. Da der textbasierte Chatbot sich zurzeit nur mit der Sprache Deutsch auseinandersetzt, bietet diese API für den aktuellen Anwendungsfall keinen Mehrwert, ist aber für zukünftige Erweiterung im Hinterkopf zu behalten.

Die darauffolgende API *Entity Linking* befindet sich noch in der Vorschauphase, wodurch deren Potenzial sich nur erahnen lässt. Mit wachsendem Volumen und steigender Komplexität der Daten (und somit grösserer Wahrscheinlichkeit auf Mehrdeutigkeit innerhalb einer Eingabe) werden Funktionalitäten, die in der Lage sind, mehrdeutige Wörter und Aussagen in den für den Benutzer richtigen Kontext zu setzen, relevant sein. Dies verspricht die API *Entity Linking* gemäss der Beschreibung von Microsoft. Besonders in komplexeren Domänen wird diese API von grossem Nutzen sein und sollte unbedingt in die Entwicklung eines Chatbots einfließen.

Der *Bing Spell Check* wäre in der eigentlichen Funktion von Vorteil, da er die Eingabe des Benutzers (bei Rechtschreibfehlern oder unvollständigen Wörtern)

automatisch korrigiert und somit die Suche nach der richtigen Antwort vereinfacht. Zwei Argumente sprechen jedoch dagegen: Erstens verwendet *LUIS* bereits etwas Ähnliches, bevor der Frage-Antwort-Abgleich stattfindet; zweitens basiert diese API auf der Suchmaschine von Microsoft (Bing). Das heisst, die Rechtschreibprüfung greift auf eine Vielzahl von Web-Recherchen und Dokumenten zurück. Es ist daher anzunehmen, dass die eingegebenen Informationen von Microsoft abgespeichert werden. Im Anwendungsfall des User Help Desk kann es vorkommen, dass Benutzerspezifische Informationen vonnöten sind (z. B. Loginname, Arbeitsplatz, Arbeitsgerät) und basierend auf SR 235.1 Bundesgesetz vom 19. Juni 1992,¹² gilt es die Daten der Benutzer zu schützen.

Text Sentiment Analysis wird mittelfristig ebenfalls eine wichtige Funktion in der Interaktion zwischen Benutzer und Maschine werden. Da die Funktion von Microsoft zurzeit nur darin besteht, zwischen positiv und negativ zu unterscheiden und nicht auf die einzelnen Gefühlsgruppen eingeht, wie beispielsweise vorgeschlagen von Plutchik (2002), wird auf diese API nicht weiter eingegangen.

8 Fazit und Ausblick

Ziel ist es, über einen Chatbot zu verfügen, der als zusätzlicher Kommunikationskanal erreichbar ist und (repetitive) einfache Anfragen selbst bearbeiten kann, während die komplexeren und schwierigen Probleme von den Agentinnen und Agenten des User Help Desk bearbeitet und gelöst werden. Dabei wird in erster Linie das Bedürfnis der Kunden (d. h. den Mitarbeiterinnen und Mitarbeitern der Schweizerischen Post) adressiert, die richtigen Informationen für die Lösung des Problems so schnell und einfach wie möglich zu erhalten. Die wesentlichen Vorteile bestehen darin, dass zum einen die Kunden immer gleich angesprochen werden und auch immer dieselbe Antwort erhalten (d. h. keine Bevorzugung von Kunden bzw. keine unterschiedlichen Informationsausgaben); zum andern führt die Automation potenziell zu einer Verkürzung der Kundenwartezeit (besonders in einfachen Anliegen) und erhöht damit die Kundenzufriedenheit.

Der User Help Desk ist für Mitarbeitende der Schweizerischen Post die zentrale Anlaufstelle für Anfragen und Störungen. Aufgrund der Vielfalt der Kommunikationskanäle und der Mehrsprachigkeit scheint es sinnvoll, sich Gedanken über eine mögliche Automatisierung zu machen. Ein möglicher Lösungsansatz die Verarbeitung der Service Calls des User Help Desk zu automatisieren, ist die Entwicklung und Implementierung eines textbasierten Chatbots als zusätzlichen Kommunikationskanals. In diesem Kapitel wurde daher die Tauglichkeit der Cognitive Services aus den Kategorien *Knowledge* und *Language*, angeboten von Microsoft, mithilfe eines Kriterienkataloges evaluiert. Die Bewertung zeigte, dass besonders die zwei APIs *LUIS* und *QnA Maker* für die Entwicklung eines solchen Bots essentiell sind und daher diese Kombination zu empfehlen ist. Weitere APIs, wie zum Beispiel *Language Detection* oder *Translator Text* sind für den jetzigen Anwendungsfall

¹²<https://www.admin.ch/opc/de/classified-compilation/19920153/index.html>.

nicht relevant, werden aber von grossem Nutzen sein, sobald der Chatbot mehrsprachig agieren muss.

Da im Anwendungsfall User Help Desk der Fokus auf die Verarbeitung geschriebener Sprache basiert, sind APIs, welche die gesprochene Sprache verarbeiten zurzeit nicht in Frage gekommen und wurden deshalb auch nicht evaluiert. In Anbetracht der rasanten Entwicklung der virtuellen Assistenten, wie Siri¹³ von Apple oder Alexa¹⁴ von Amazon, ist es dennoch wichtig, die Entwicklung von sprachbasierten Chatbots für zukünftige Anwendungsfälle der Schweizerischen Post in Betracht zu ziehen.

Danksagung Die Autoren dieses Kapitels danken insbesondere Markus Bacher, Joseph Beer, Nikolay Borissov, Hans-Rudolf Burri, Mik Häfliger, Matthias Hartl, Daniel Huber, Robert Käpeli, Friedrich Klopfenstein und Nicole Schuler für ihre wertvollen Inputs. Ohne sie hätte dieser Artikel in dieser Form nicht geschrieben werden können. Vielen Dank!

Literatur

- Barrière C (2016) Natural Language Understanding in a Semantic Web Context. Springer International Publishing.
- Die Schweizerische Post (2016) Vision – Einfach mit System – Die Post. <https://www.post.ch/de/ueber-uns/unternehmen/unsere-grundsaeetze/vision>. Letztes Update: 28.01.2016. Letzter Zugriff: 03.01.2019.
- D’Onofrio S, Portmann E (2017) Cognitive Computing in Smart Cities. Informatik Spektrum. 40(1):46–57.
- D’Onofrio S, Portmann E, Franzelli M, Bürki C (2018) Cognitive Computing: Theoretische Grundlagen und Praxisbeispiele der Schweizerischen Post. Informatik Spektrum. 41(2): 113–122.
- Hänggeli L, Zurwerra M (2017) Softwares zur Innovations- und Businessmodellgenerierung – Eine Handlungsempfehlung für Grossunternehmen. Masterthesis. Universität Bern.
- HP Lab (2012) Intelligent Infrastructure. http://www.hpl.hp.com/research/intelligent_infrastructure_m.html. Letztes Update: 2012. Letzter Zugriff : 03.01.2019.
- Hurwitz JS, Kaufman M, Bowles A (2015) Cognitive Computing and Big Data Analytics. John Wiley and Sons, Inc., Hoboken, New Jersey.
- IBM (2017) 2017 Annual Report. https://www.ibm.com/annualreport/2017/assets/downloads/IBM_Annual_Report_2017.pdf. Letzter Zugriff: 16.09.2018.
- Kurdi M (2016) Natural Language Processing and Computational Linguistics 1: Speech, Morphology, Syntax. Wiley.
- Kurdi M (2017) Natural Language Processing and Computational Linguistics 2: Semantics, Discourse and Applications. Wiley.
- Mackmood J, Bammert Marty D, Klopfenstein F (2017) Chatbots for internal customer care. Berner Fachhochschule.
- Microsoft Azure (2019a) Cognitive Services. <https://azure.microsoft.com/en-us/services/cognitive-services>. Letztes Update: 02.01.2019. Letzter Zugriff : 03.01.2019.
- Microsoft Azure (2019b) Cognitive Services Directory. <https://azure.microsoft.com/en-us/services/cognitive-services/directory>. Letztes Update: 02.01.2019. Letzter Zugriff: 03.02.2019.
- Plutchik R (2002) Emotions and Life: Perspectives from Psychology, Biology, and Evolution. American Psychological Association. First edition.

¹³<https://www.apple.com/siri/>.

¹⁴<https://developer.amazon.com/alexa>.

- Ray T (2016) IBM's Rometty Wants You to Know They're a 'Cognitive Solutions Cloud Platform Company'. Barron's. <https://www.barrons.com/articles/ibms-rometty-wants-you-to-know-theyre-a-cognitive-solutions-cloud-platform-company-1452139085>. Erstellt: 06. Januar 2016. Letzter Zugriff: 16.09.2018.
- Schmidhuber J (2015) Deep learning in neural networks: An overview. *Neural Networks*. 61:85–117.
- Softguide (2018) 1. Einführung und Grundlagen – Einführung und Grundlagen zum Leitfaden „Tipps zur Auswahl betrieblicher Software“. <https://www.softguide.de/software-tipps/grundlagen>. Letztes Update: 15.05.2018. Letzter Zugriff: 03.01.2019.
- Stucki T, D'Onofrio S, Portmann E (2018) Chatbot – Der digitale Helfer im Unternehmen: Praxisbeispiele der Schweizerischen Post. *HMD Praxis der Wirtschaftsinformatik*. 55(4):725–747.
- Stucki T, D'Onofrio S, Portmann E (2020) Chatbots gestalten mit Praxisbeispielen der Schweizerischen Post. *HMD Best Paper Award 2018. Essentials*. Springer Vieweg, Wiesbaden.
- Zientara D (2016) Blog: The Bot Framework. *Applied Information Sciences (ais)*. <https://www.appliedis.com/the-bot-framework/>. Erstellt: 16.08.2016. Letzter Zugriff: 18.09.2018.



Soziotechnische Gestaltung von Chatbots

Ulrich Gnewuch, Jasper Feine, Stefan Morana und
Alexander Maedche

Zusammenfassung

Chatbots sind softwarebasierte Systeme, die mittels natürlicher Sprache mit Menschen interagieren. Viele Unternehmen setzen zunehmend auf Chatbots, um Kunden bei der Suche nach Informationen über Produkte oder Dienstleistungen und bei der Durchführung einfacher Prozesse zu unterstützen. Nichtsdestotrotz ist die Akzeptanz von Chatbots bei vielen Nutzern derzeit noch gering. Ein Grund dafür ist, dass sich die Interaktion mit Chatbots nur selten natürlich und *menschlich* anfühlt. Es wächst deshalb die Erkenntnis, dass neben einer guten technischen Plattform auch weitere Faktoren bei der Gestaltung von Chatbots beachtet werden sollten. Die soziotechnische Gestaltung von Chatbots fokussiert sich daher auf die sozialen Signale eines Chatbots. Diese sozialen Signale (z. B. Lächeln, Sprachstil oder Antwortgeschwindigkeit) spielen nicht nur in der zwischenmenschlichen Kommunikation, sondern auch in der Interaktion mit Chatbots eine grosse Rolle. Dieser Artikel erläutert die Grundlagen der soziotechnischen Gestaltung von Chatbots, verdeutlicht die Wirkung sozialer Signale anhand eines Forschungsbeispiels und diskutiert kritisch die Vermenschlichung von Chatbots.

Vollständig neuer Beitrag

U. Gnewuch (✉) · J. Feine · S. Morana · A. Maedche
Karlsruher Institut für Technologie (KIT), Institute of Information Systems and Marketing
(IISM), Karlsruhe, Deutschland
E-Mail: ulrich.gnewuch@kit.edu; jasper.feine@kit.edu; stefan.morana@kit.edu; alexander.maedche@kit.edu

1 Einleitung

Chatbots und Sprachassistenten liegen derzeit im Trend (vgl. (Gartner 2018)). Da sie eine Interaktion über natürliche Sprache ermöglichen, wird ihnen grosses Potenzial als einfache zu bedienende Benutzerschnittstelle zugeschrieben. Neben den populären Sprachassistenten, wie zum Beispiel Amazon's Alexa,¹ werden zunehmend auch sogenannte Chatbots, mit denen ein Nutzer über Textnachrichten interagieren kann, auf Webseiten oder Messaging-Plattformen eingesetzt. Aufgrund der jüngsten technologischen Fortschritte und der Popularität von Textnachrichten als Kommunikationsform ist bei Unternehmen insbesondere das Interesse an Chatbots gestiegen. In vielen Branchen werden Chatbots bereits produktiv eingesetzt, um Nutzer bei der Suche nach relevanten Informationen über Produkte und Dienstleistungen zu unterstützen und einfache Prozesse, wie etwa bei der Flug- oder Hotelbuchung, zu automatisieren (Morana et al. 2017). Die Schweizerische Post setzt beispielsweise einen Chatbot auf ihrer Webseite ein, um häufige Kundenanfragen zu Produkten und Dienstleistungen von PostFinance automatisiert zu beantworten (D'Onofrio et al. 2018; Stucki et al. 2018, 2020). Von dem Einsatz eines Chatbots versprechen sich Unternehmen sowohl Kosteneinsparungen als auch eine Rund-um-die-Uhr-Verfügbarkeit für ihre Kunden (Gnewuch et al. 2017). Allein auf Facebook Messenger² wurden daher in einem Jahr bereits über 100.000 Chatbots erstellt (Johnson 2017).

Trotz des grossen Interesses von Unternehmen und Technologiekonzernen Chatbots zu entwickeln und einzusetzen, ist die Akzeptanz von Chatbots bei vielen Nutzern derzeit noch gering. Zwar kann sich mittlerweile jeder Vierte vorstellen, einen Chatbot zu nutzen, doch gibt es noch viele Vorbehalte gegenüber Chatbots (Bitkom 2018). Ein Grund für die mangelnde Akzeptanz bei Nutzern ist, dass sich die Interaktion mit einem Chatbot nur selten natürlich und *menschlich* anfühlt (Moore et al. 2017). Aufgrund der Komplexität der menschlichen Sprache gibt es noch immer Schwierigkeiten beim Verstehen von mehrdeutigen Nachrichten eines Nutzers. Zum Beispiel kann die Nachricht „*Das ist ja toll*“ je nach Situation als Lob oder als ironische Bemerkung verstanden werden. Aktuell tun sich Chatbots jedoch schwer damit, solche Nuancen in der Sprache zu erkennen. Die Forschung hat jedoch gezeigt, dass es nicht ausschliesslich eine technische Herausforderung ist, dem Nutzer das Gefühl einer natürlichen Konversation zu vermitteln. In der Interaktion mit einem Chatbot gibt es diverse Gestaltungselemente, die das Nutzererlebnis (engl. *user experience*) massgeblich beeinflussen. Zum Beispiel haben Chatbots oft ein visuelles Erscheinungsbild (bspw. einen menschenähnlichen Avatar), benutzen Emojis oder Smalltalk. Studien haben gezeigt, dass Nutzer – auch wenn sie wissen, dass sie es mit einem Chatbot zu tun haben – diese Elemente in einer Interaktion erwarten, da sie diese aus der Kommunikation mit anderen Menschen gewohnt sind (Nass und Moon 2000). Viele dieser Elemente entsprechen daher den sozialen Si-

¹<https://developer.amazon.com/de/alexa>.

²<https://www.messenger.com>.

gnalen (engl. *social cues*), die aus der Kommunikationsforschung bekannt sind. So haben nonverbale Signale, wie etwa ein Lächeln, Blickkontakt oder die Stimmlage, einen grossen Einfluss auf die zwischenmenschliche Kommunikation (Burgoon et al. 2010). Ähnlich wie in der zwischenmenschlichen Kommunikation nehmen Nutzer auch soziale Signale von Chatbots wahr und reagieren unterbewusst darauf (Nass et al. 1994). Diese unterbewussten Reaktionen können die Empfindungen der Nutzer und mitunter auch ihr Verhalten beeinflussen. Beispielsweise bauen viele Nutzer des chinesischen Chatbots Xiaoice³ eine emotionale Bindung zu dem Chatbot auf und senden täglich die Nachricht „*ich liebe dich*“ (Markoff und Mozur 2015). Die Forschung hat gezeigt, dass dies unter anderem durch ein attraktives Erscheinungsbild des Chatbot Avatars ausgelöst werden kann (Chae et al. 2016). Nutzer verwenden in der Interaktion mit einem Chatbot auch oft Höflichkeitsfloskeln oder führen Smalltalk, ungeachtet der Tatsache, dass Chatbots Computerprogramme sind (Stucki et al. 2018, 2020). Dieses Verhalten verdeutlicht, dass Chatbots oft nicht nur als reine Werkzeuge für die Erledigung bestimmte Aufgaben wahrgenommen werden, sondern von den Nutzern unterbewusst *personalisiert* werden (Nass und Moon 2000).

Viele Studien deuten daher darauf hin, dass bei der Entwicklung von Chatbots nicht nur der Einsatz modernster Technologien und Methoden, wie die des *Natural Language Processing*, im Mittelpunkt steht, sondern dass auch soziale Faktoren einen massgeblichen Einfluss auf den Erfolg eines Chatbots haben können (vgl. z. B. (Araujo 2018; Gnewuch et al. 2018a; Liao et al. 2018)). Folglich zielt die soziotechnische Gestaltung von Chatbots darauf ab, sowohl technische als auch soziale Faktoren und Erkenntnisse möglichst gleichberechtigt in den Gestaltungsprozess einzubringen (Mumford 2006). Allerdings gibt es bisher kaum konkrete Richtlinien für die soziotechnische Gestaltung von Chatbots. Im Internet finden sich viele Webseiten (bspw. *chatbots.org* oder *chatbotsmagazine.com*) und Blogartikel (z. B. *botsociety.io/blog*), die eher allgemeine Tipps und Anregungen geben, auf was bei der Gestaltung von Chatbots zu achten ist (McTear 2017). Beispielsweise weisen viele dieser Artikel darauf hin, dem Chatbot eine Persönlichkeit zu geben und soziale Signale, wie zum Beispiel Emojis oder Smalltalk, einzusetzen, um die Interaktion mit dem Chatbot natürlicher zu gestalten. Diese Ratschläge sind jedoch meist vage oder zu abstrakt, um sie leicht umsetzen zu können.

In diesem Artikel wird die Gestaltung von Chatbots aus einer soziotechnischen Perspektive betrachtet. Komplementär zur Entwicklung von Chatbots mit dem Fokus auf technischen Problemstellungen, konzentriert sich dieser Artikel auf die Gestaltung der sozialen Signale eines Chatbots. Zunächst werden Chatbots kurz eingeführt und die theoretischen Grundlagen, warum Menschen auf soziale Signale von Chatbots reagieren, erläutert. Danach wird anhand eines aktuellen Forschungsprojekts gezeigt, wie sich die Gestaltung eines bestimmten sozialen Signals (d. h. der Antwortgeschwindigkeit) auf die Wahrnehmung eines Chatbots und die Kundenzufriedenheit auswirkt. Abschliessend wird erörtert, wie die sozialen Signale eines Chatbots an bestimmte Umgebungsvariablen angepasst werden können. Ausserdem

³<https://www.msxiaobing.com>.

werden die Implikationen einer Vermenschlichung von Chatbots durch die Gestaltung sozialer Signale kritisch diskutiert.

2 Chatbots und die Wirkung ihrer sozialen Signale

Chatbots sind textbasierte Dialogsysteme, die sich mit einem Nutzer in natürlicher Sprache unterhalten (*Chat*) und eigenständig einfache Aufgaben ausführen können (*Bot*) (Stucki et al. 2018, 2020). Während in der Praxis viele verschiedene Bezeichnungen, wie etwa digitaler oder virtueller Assistent genutzt werden, konzentriert sich die Wirtschaftsinformatikforschung auf den Begriff *Conversational Agent*. Dieser Begriff bezeichnet üblicherweise alle Softwaresysteme, die eine Interaktion über natürliche Sprache ermöglichen (Dale 2016), und umfasst neben den textbasierten Chatbots auf Webseiten und Messaging-Plattformen auch Sprachassistenten, wie zum Beispiel Amazon's Alexa (Gnewuch et al. 2018a).

Obwohl 2016 zum *Jahr des Chatbots* ernannt wurde, ist diese Technologie nicht gänzlich neu (Dale 2016). Die Forschung beschäftigt sich schon seit einigen Jahrzehnten mit den Grundlagen dieser Technologie (Følstad und Brandtzæg 2017). Der erste Chatbot ELIZA wurde bereits 1966 entwickelt und ahmte den Konversationsstil eines Psychotherapeuten nach (Weizenbaum 1966). Mit einem einfachen Musterabgleich (engl. *pattern matching*) erkannte ELIZA bestimmte Begriffe in den Nachrichten des Nutzers. Anschliessend antwortete ELIZA mit vordefinierten Antworten, die Platzhalter enthielten und mit entsprechenden Worten ersetzt werden konnten. Auf den ersten Chatbot folgten viele weitere, wie zum Beispiel PARRY, der einen paranoiden schizophrenen Menschen nachahmte (Colby et al. 1971), oder A.L.I.C.E., ein Web-Chatbot der auf der Open-Source Programmiersprache AIML von Richard Wallace beruht (Wallace 2009). Seitdem hat sich die Technologie jedoch deutlich weiterentwickelt. Insbesondere durch technische Fortschritte auf dem Gebiet der künstlichen Intelligenz können moderne Chatbots den Menschen immer besser verstehen und automatisch dazulernen. Beispiele dafür sind die intelligenten Sprachassistenten wie Amazon's Alexa und Google's Assistant.⁴

Während sich ein Grossteil der Forschung auf die Lösung technischer Problemstellungen (Verbesserung der Natural Language Processing Algorithmen, Entwicklung neuer Architekturen, usw.) fokussiert, wächst die Erkenntnis, dass neben einer guten technischen Plattform auch weitere Faktoren bei der Gestaltung von Chatbots beachtet werden müssen. Da Nutzer mit Chatbots über natürliche Sprache kommunizieren, müssen bei der Gestaltung viele Faktoren und Elemente aus der zwischenmenschlichen Kommunikation berücksichtigt werden.

In der zwischenmenschlichen Kommunikation spielen soziale Signale eine grosse Rolle und sind für den Aufbau und Erhalt erfolgreicher sozialer Beziehungen unerlässlich (Burgoon et al. 2010). Bei Menschen sind soziale Signale implizite Hinweisreize, wie zum Beispiel Lächeln, Blickkontakt oder Stimmlage, aus denen sich soziale Informationen ableiten lassen. Diese Signale werden oft unterbewusst

⁴<https://assistant.google.com>.

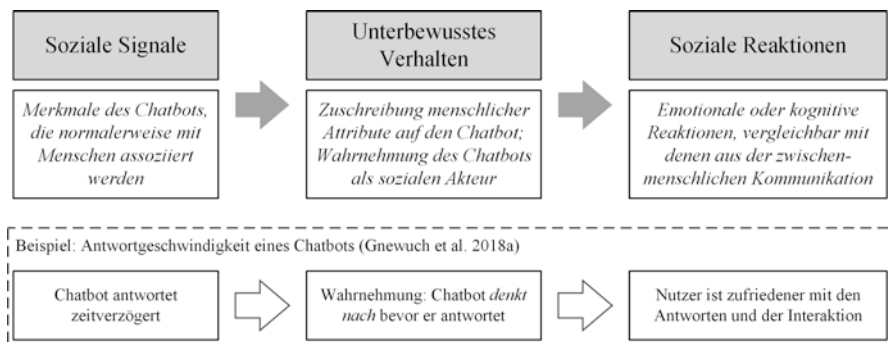


Abb. 1 Entstehung von sozialen Reaktionen in der Interaktion mit Chatbots

verwendet, um die Gedanken und Gefühle anderer Personen richtig zu interpretieren. Die Forschung hat gezeigt, dass Menschen auch in der Interaktion mit Computern und anderen Technologien diese sozialen Signale unterbewusst wahrnehmen und automatisch darauf reagieren (Nass et al. 1994). Dieses Verhalten lässt sich darauf zurückführen, dass Menschen als soziale Wesen evolutionär darauf programmiert sind, auf alle Merkmale zu reagieren, die als soziale Signale verstanden werden können (Nass und Moon 2000). Selbst relativ subtile und unauffällige soziale Signale reichen bereits aus, um bei Menschen eine soziale Reaktion hervorzurufen. Diese ausgelösten sozialen Reaktionen werden vom Menschen meist nicht bewusst wahrgenommen. Aus diesem Grund weisen viele Studien darauf hin, dass auch Computer und andere Technologien als *Soziale Akteure* angesehen werden können (Nass et al. 1994).

Abb. 1 zeigt den Prozess der Entstehung von sozialen Reaktionen in der Interaktion mit Chatbots (basierend auf Krämer (2005) sowie Nass und Moon (2000)). Zunächst nimmt ein Nutzer unterbewusst Merkmale eines Chatbots wahr, die normalerweise mit Menschen (z. B. Name, Geschlecht) assoziiert werden und daher als soziale Signale interpretiert werden können. Neben der Verwendung von natürlicher Sprache gibt es bei Chatbots eine Vielzahl dieser Signale. Diese sozialen Signale führen dazu, dass der Nutzer dem Chatbot unterbewusst menschliche Attribute zuschreibt und gelerntes Verhalten aus der zwischenmenschlichen Kommunikation auf den Chatbot anwendet. Somit wird der Chatbot als sozialer Akteur angesehen und ruft soziale Reaktionen beim Nutzer hervor. Dabei handelt es sich um emotionale und kognitive Reaktionen, die vergleichbar sind mit den Reaktionen während der Interaktion mit anderen Menschen.

Viele Studien haben untersucht, wie sich bestimmte soziale Signale auf die Wahrnehmung eines Computers oder anderer Technologien und auf das Nutzerverhalten auswirken. In ihren ersten Experimenten in der 1990er-Jahren zeigten Nass, Steuer und Tauber, dass die Teilnehmer zwei verschiedene Stimmen desselben Computers als zwei verschiedene soziale Akteure wahrnahmen und je nach Geschlecht der Stimme andere Stereotypen auf den Computer anwendeten (Nass et al. 1994). So bevorzugten die Teilnehmer eine Leistungsbewertung von einem Compu-

ter mit männlicher Stimme und lehnten einen weiblichen Computer mit einer dominanten Persönlichkeitsausprägung ab. Ausserdem konnten Nass, Fogg und Moon zeigen, dass ein Computer als Teammitglied wahrgenommen werden kann, wenn er durch farbliche Kennzeichnung dem Team eines Nutzers zugeordnet wird (Nass et al. 1996).

Viele Studien über Chatbots haben sich mit den sozialen Signalen beschäftigt, die vom Erscheinungsbild des Avatars eines Chatbots ausgehen. Einige Forscher haben den Einfluss visueller sozialer Signale, wie zum Beispiel eines lächelnden Avatars, untersucht. So erwidern Menschen ein Lächeln deutlich öfter, wenn ein Avatar sie anlächelt (Krämer et al. 2013) und interagieren deutlich länger mit einem lächelnden Avatar (Cafaro et al. 2016). Daneben gibt es aber auch viele andere Arten sozialer Signale, wie etwa verbale Signale. So hat die Verwendung von Smalltalk Einfluss auf die wahrgenommene Persönlichkeit eines Chatbots (Li et al. 2017) und Chatbots, die Witze machen, werden als humorvoller angesehen (Dybala et al. 2009).

Im Rahmen eines Forschungsprojekts wurde eine systematische Identifizierung und Klassifizierung der sozialen Signale von Chatbots durchgeführt. Diese Klassifizierung basiert auf der Einteilung der menschlichen Kommunikation in vier grundlegende Kommunikationssysteme: *Verbal*, *visuell*, *auditiv* und *unsichtbar* (Leathers 1976). Wie in Abb. 2 dargestellt, lassen sich die sozialen Signale von Chatbots anhand dieser Kategorien aus der zwischenmenschlichen Kommunikation klassifizieren.

Verbale Signale umfassen alle sozialen Signale, die durch die Verwendung von Worten erzeugt werden. Diese verbalen Signale können dahingegen unterschieden werden, dass der gleiche verbale Inhalt einer Nachricht (d. h. *was* gesagt wird) in unterschiedlichen Stilen (d. h. *wie* etwas gesagt wird) ausgedrückt werden kann (Collier 2014). Visuelle Signale sind alle nonverbalen sozialen Signale, die visuell wahrnehmbar sind und auf drei verschiedene Arten erzeugt werden können: Durch das Aussehen (z. B. Farben), durch Bewegungen (bspw.

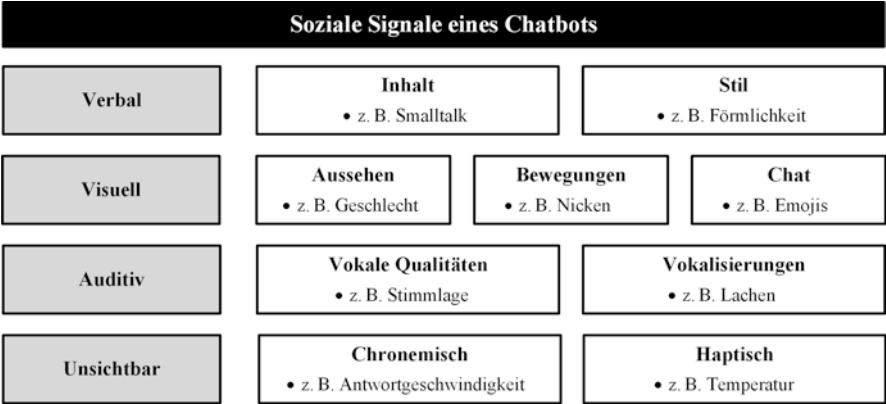


Abb. 2 Klassifikation sozialer Signale von Chatbots

Mimik) und durch die Chat-Umgebung (d. h. besondere Elemente der textbasierten Kommunikation, wie etwa Emojis) (Leathers 1976; Walther 1992). Auditive Signale umfassen alle hörbaren sozialen Signale, die durch nonverbale Laute erzeugt werden (d. h. nicht durch Worte). Diese bestehen aus den vokalen Qualitäten (d. h. permanente und variable Eigenschaften der gesprochenen Sprache, wie etwa Stimmlage und Sprechtempo) und den Vokalisierungen (nicht-sprachliche Stimmlaute oder Geräusche, wie z. B. Husten) (Trager 1958). Unsichtbare Signale sind alle sozialen Signale, die nicht gesehen oder gehört werden können. Diese bestehen aus den chronemischen Signalen (d. h. zeitliche Aspekte der Kommunikation, wie bspw. Antwortzeiten) und haptischen Signalen (z. B. Wärmeempfindungen).

Durch die Vielzahl an sozialen Signalen und deren mögliche Ausprägungen gibt es viele Gestaltungsmöglichkeiten, die die Wahrnehmungen und das Verhalten der Nutzer beeinflussen können. Daher können bestimmte Gestaltungsentscheidungen eine grosse Auswirkung auf den Erfolg eines Chatbots haben. Eine smarte Gestaltung der sozialen Signale kann dazu beitragen, das Vertrauen in einen Chatbot zu steigern (de Visser et al. 2016), die Nutzerzufriedenheit mit einem Chatbot zu erhöhen (Verhagen et al. 2014) und den Erfolg einer langfristigen Beziehung zwischen einem Chatbot und einem Menschen sicherzustellen (Bickmore und Picard 2005). Im Gegensatz dazu kann sich eine ungünstige Gestaltung der sozialen Signale eines Chatbots negativ auf den Nutzer und die Interaktion auswirken. So wird ein Chatbot mit einer sehr direkten Ansprache als unhöflich empfunden (Mayer et al. 2006) und ein zu menschenähnliches Erscheinungsbild kann den Nutzer davon abhalten, Informationen über sich selbst preiszugeben (Sah und Peng 2015).

3 Soziotechnische Gestaltung von Chatbots am Beispiel der Antwortgeschwindigkeit⁵

Im vorangegangenen Kapitel wurden die theoretischen Grundlagen, warum Menschen auf soziale Signale von Chatbots reagieren, erklärt und die grosse Vielfalt der möglichen sozialen Signale von Chatbots aufgezeigt. Im folgenden Kapitel wird nun auf ein konkretes soziales Signal eines Chatbots (d. h. die Antwortgeschwindigkeit) eingegangen und anhand dieses Beispiels die soziotechnische Gestaltung von Chatbots illustriert. Zunächst wird der aktuelle Forschungsstand zur Antwortgeschwindigkeit von Chatbots dargestellt, bevor ein Gestaltungsprinzip und eine Formel zur Berechnung einer dynamischen Antwortverzögerung hergeleitet werden. Anschliessend werden die Ergebnisse eines Experiments zur Evaluierung dieser dynamischen Antwortverzögerung vorgestellt und diskutiert.

⁵Zentrale Teile dieses Kapitels basieren auf Gnewuch et al. (2018a).

3.1 Antwortgeschwindigkeit von Chatbots

In der zwischenmenschlichen und computergestützten Kommunikation ist die Antwortgeschwindigkeit (d. h. wie schnell eine Person auf die Aussage einer anderen Person antwortet) ein wichtiges soziales Signal, das die Wahrnehmung einer Person und die Beurteilung von dessen Antwort massgeblich beeinflusst (Walther und Tidwell 1995; Moon 1999). So kann eine lange Pause vor einer Antwort (d. h. eine langsame Antwortgeschwindigkeit) ein Zeichen für ein sorgfältiges Nachdenken sein, aber in bestimmten Situationen auch den Eindruck erwecken, dass eine Person lügt, da sie sich die passende Antwort zunächst zurechtlegen muss. Es wird daher vermutet, dass auch in der Interaktion zwischen Mensch und Chatbot die Antwortgeschwindigkeit des Chatbots eine wichtige Rolle spielt (Walther und Tidwell 1995).

Forscher sind sich jedoch uneinig darüber, welchen Einfluss die Antwortgeschwindigkeit eines Chatbots auf seine Wahrnehmung beim Nutzer hat. Generell wird angenommen, dass blitzschnelle Antworten einen Chatbot *unmenschlich* erscheinen lassen und dem Nutzer nicht das Gefühl einer natürlichen Konversation vermitteln (Appel et al. 2012; Holtgraves und Han 2007; Stucki et al. 2018, 2020). Da moderne Chatbots in der Lage sind, fast unmittelbar auf Nachrichten des Nutzers zu antworten, werden Antworten daher oft künstlich verzögert. Die Verzögerung simuliert dabei die Zeit, die ein Mensch benötigen würde, um eine Nachricht zu lesen und darauf zu antworten (Appel et al. 2012). Beispielsweise hat Lufthansa als Reaktion auf Beschwerden der Kunden über die zu schnellen Antworten ihres Chatbots *Mildred*⁶ eine künstliche Verzögerung eingebaut (Crozier 2017). Andere Studien haben jedoch gezeigt, dass Verzögerungen sich auch negativ auf die Nutzerzufriedenheit auswirken können (Hoxmeier und DiCesare 2000), insbesondere dann, wenn der Nutzer eine schnelle Antwort erwartet (wie bspw. im Kundenservice).

Bisher gibt es nur wenige empirische Studien, die die Auswirkungen unterschiedlicher Antwortgeschwindigkeiten auf die Wahrnehmung von Chatbots beim Nutzer untersucht haben. Ausserdem kommen diese Studien zu widersprüchlichen Ergebnissen über den Einfluss verzögerter Antworten eines Chatbots. Während Moon (1999) feststellt, dass Nutzer schnelle Antworten negativ bewerten, zeigt Holtgraves et al. (2007), dass ein schneller Chatbot positiver wahrgenommen wird als ein Chatbot mit verzögerten Antworten. Ferner werden in diesen Studien nur statische Verzögerungen untersucht, die unabhängig von der Länge oder Komplexität der gesendeten Nachricht sind (d. h. jede Nachricht des Chatbots wird nach einer bestimmten Anzahl von Sekunden gesendet). Aufgrund der widersprüchlichen Forschungsergebnisse und des zunehmenden Einsatzes künstlicher Verzögerungen bei realen Chatbots wurde daher der Frage nachgegangen, wie die Antwortgeschwindigkeit eines Chatbots natürlicher gestaltet werden kann. Aufbauend auf den bisherigen Forschungsergebnissen und unter Berücksichtigung von Erkenntnissen aus der zwischenmenschlichen Kommunikation haben die Autoren eine Formel zur

⁶https://www.lufthansa.com/us/en/lufthansa_bot.

Berechnung einer dynamischen Verzögerung hergeleitet und anschliessend in einem Chatbot implementiert.

3.2 Dynamische Antwortverzögerung für Chatbots

Für die Herleitung der dynamischen Antwortverzögerung wurde ein gestaltungsorientierter Forschungsansatz (engl. *design science research*) angewendet. Der gestaltungsorientierte Forschungsansatz zielt auf die iterative Gestaltung und Evaluation neuer und innovativer Artefakte (z. B. Softwareprototypen oder Modelle) zur Lösung praktischer Probleme ab (Hevner et al. 2004). Ein spezieller Fokus liegt dabei auf der Einbeziehung von existierenden wissenschaftlichen Erkenntnissen (bspw. Theorien aus anderen Disziplinen) als Basis zur fundierten Herleitung von Gestaltungsprinzipien. Dieser Forschungsansatz bietet sich daher besonders für die soziotechnische Gestaltung von Chatbots an, da so bestehendes Wissen aus der zwischenmenschlichen Kommunikation in die Gestaltung und Evaluation einer dynamischen Antwortverzögerung integriert werden kann. Im Folgenden werden drei Meta-Anforderungen (MA) für die Gestaltung von Chatbots mit einer dynamischen Antwortverzögerung hergeleitet und anschliessend in einem Gestaltungsprinzip zusammengefasst.

Forscher und Anwender argumentieren, dass schnelle Antworten einen Chatbot unmenschlich wirken lassen und Nutzern nicht das Gefühl einer natürlichen Konversation geben (Appel et al. 2012; Crozier 2017; Holtgraves et al. 2007). Daher haben viele Chatbots eine Verzögerung eingebaut, um die Zeit zu simulieren, die ein Mensch für seine Antwort bräuchte. Die Länge dieser Verzögerung ist jedoch meist fest definiert oder zufällig generiert und orientiert sich nicht an den individuellen Eigenschaften einer Nachricht. In der textbasierten Kommunikation setzt sich die Antwortzeit eines Menschen in der Regel aus zwei Komponenten zusammen: Zum einen aus der Zeit, die der Mensch benötigt, um die Nachricht einer anderen Person zu lesen beziehungsweise zu verstehen, und zum anderen aus der Zeit, die der Mensch benötigt, um eine Antwort zu formulieren und einzugeben (Derrick et al. 2013). Konsequenterweise sollte eine dynamische Antwortverzögerung eines Chatbots beide Komponenten, das heisst sowohl die Nachricht des Nutzers (MA1) als auch die Antwort des Chatbots (MA2) berücksichtigen. Darüber hinaus hängt die Antwortzeit eines Menschen von den individuellen Eigenschaften einer Nachricht ab. Insbesondere die Komplexität spielt dabei eine grosse Rolle, da komplexere Nachrichten länger zum Lesen und komplexere Antworten länger zum Formulieren brauchen (Khawaja et al. 2010). Demzufolge sollte eine dynamische Antwortverzögerung eines Chatbots die Komplexität der Nachricht beziehungsweise der Antwort einkalkulieren (MA3). Basierend auf diesen drei MAs wird folgendes Gestaltungsprinzip für eine dynamische Antwortverzögerung formuliert:

Gestaltungsprinzip: *Integriere eine dynamische Antwortverzögerung in den Chatbot, die basierend auf der (1) Komplexität der eingehenden Nachricht (des Nutzers) und der (2) Komplexität der Antwort des Chatbots eine Antwortzeit berechnet, die sich für den Nutzer natürlicher anfühlt.*

Im Folgenden wird nun die Umsetzung dieses Gestaltungsprinzips in einer Formel zur Berechnung einer dynamischen Antwortverzögerung beschrieben. Zur Berechnung der Komplexität einer Nachricht wurde ein Lesbarkeitsindex – *Flesch-Kincaid-Grade-Level* – verwendet, der bereits zur Analyse von textbasierter Kommunikation angewendet wurde. Prinzipiell berechnet dieser Lesbarkeitsindex die Anzahl der Schuljahre, die ein Leser absolviert haben muss, um einen Text zu verstehen (Kincaid et al. 1975).

Anhand folgender Formel des *Flesch-Kincaid-Grade-Level* wurde die Sprachkomplexität (K) einer Nachricht (N) berechnet:⁷

$$K(N) = 0,39 * \left(\frac{\text{Anzahl Wörter}}{\text{Anzahl Sätze}} \right) + 11,8 * \left(\frac{\text{Anzahl Silben}}{\text{Anzahl Wörter}} \right) - 15,59$$

Die Komplexitätswerte reichen von $-3,40$ bis positiv unendlich. Darauf aufbauend wird eine Antwortverzögerung (A) in Millisekunden für eine beliebige Nachricht (N) unter Verwendung des Komplexitätswertes ($K(N)$) als Input berechnet:

$$A(N) = 0,5 * \ln(K(N) + 0,5) + 1,5, K(N) > 0$$

$$A(N) = 0$$

$$K(N) \leq 0$$

Mit dieser Formel wird eine Antwortverzögerung für (1) die vorherige Nachricht (N_{n-1}), die entweder vom Nutzer oder vom Chatbot gesendet wurde, und (2) die vom Chatbot gesendete Antwort (N_n) berechnet. Dies liegt daran, dass die Antworten des Chatbots auch in mehrere, aufeinander folgende Nachrichten aufgeteilt werden können und nicht zwingend als eine lange, aggregierte Nachricht gesendet werden müssen. Letztlich werden beide Verzögerungen addiert, um die gesamte Antwortverzögerung (A_{gesamt}) in Millisekunden zu berechnen:

$$A_{\text{gesamt}}(N_n) = A(N_{n-1}) + A(N_n)$$

Basierend auf dieser Formel werden komplexe Antworten des Chatbots nach einer komplexen Nachricht des Nutzers signifikant verzögert (d. h. ca. bis zu 4 Sekunden). Weniger komplexe Nachrichten hingegen werden – wenn überhaupt – nur minimal verzögert. Zwar kann das Ergebnis aus der Berechnung mit dieser Formel nicht als *optimale* Verzögerung verstanden werden, aber dennoch stellt die Formel eine ausreichende Annäherung für die Gestaltung einer dynamischen Antwortverzögerung dar.

⁷Für weitere Informationen zur Herleitung dieser Formel verweisen die Autoren auf das Originalwerk von Kincaid et al. (1975).

3.3 Experimentaldesign und Chatbot-Szenario

Um die dynamische Antwortverzögerung zu evaluieren und die Auswirkungen unterschiedlicher Antwortgeschwindigkeiten auf die Wahrnehmung von Chatbots beim Nutzer zu untersuchen, wurde ein Online-Experiment in einem Kundenservice-Szenario durchgeführt. Dieses Experiment verwendete ein Zwischensubjekt-Design (engl. *between-subject-design*) mit zwei experimentellen Bedingungen: Schnelle Antwortgeschwindigkeit (Kontrollgruppe) und dynamisch verzögerte Antwortgeschwindigkeit (Treatmentgruppe).

In der Kontrollgruppe interagierten die Teilnehmer mit einem Chatbot, dessen Antworten nicht künstlich verzögert wurden und daher in weniger als einer Sekunde erschienen. In der Treatmentgruppe interagierten die Teilnehmer mit einem Chatbot, der – wie im Abschn. 3.2 beschrieben – die Antworten erst nach einer zusätzlichen, dynamisch berechneten Verzögerung sendete. In beiden Fällen wurden die Teilnehmer vor dem Experiment darüber informiert, dass sie mit einem Chatbot und nicht mit einem Menschen interagieren. Die Teilnehmer wurden zufällig einer der beiden Gruppen zugeteilt. Das Experiment wurde auf Englisch durchgeführt. Für die Durchführung des Experiments wurden zwei Chatbots mit dem Microsoft Bot Framework⁸ und dem *Language Understanding Intelligence Service* (LUIS)⁹ entwickelt. Beide Chatbots waren identisch bis auf ihre jeweilige Antwortgeschwindigkeit.

Zu Beginn des Online-Experiments wurde den Teilnehmern eine fiktive Kopie ihrer Mobilfunkrechnung aus dem letzten Monat gezeigt. Aus dieser Rechnung ging hervor, dass ihr aktueller Mobilfunktarif zu teuer ist und sie bei einem Wechsel Geld sparen könnten. Die Teilnehmer wurden aufgefordert, sich von dem Chatbot zu einem neuen Tarif beraten zu lassen. In der Konversation fragte der Chatbot in einem strukturierten Dialog nach den Anforderungen und der Zahlungsbereitschaft des Teilnehmers (siehe Abb. 3). Anschliessend empfahl er einen fiktiven Tarif, der besser zu den Anforderungen des Teilnehmers passte.

Nach der Interaktion mit dem Chatbot wurden die Teilnehmer gebeten, den Chatbot zu beurteilen und ihren Eindruck von der Konversation zu beschreiben. In dem dafür entwickelten Fragebogen wurden folgende etablierte Messinstrumente verwendet: Wahrgenommene Menschlichkeit (engl. *perceived humanness*) (Holtgraves und Han 2007; Holtgraves et al. 2007), soziale Präsenz (Gefen und Straub 1997) und Nutzerzufriedenheit (Verhagen et al. 2014). Um zu messen, wie menschlich der Chatbot wahrgenommen wird, wurden Teilnehmer danach gefragt, wie gut bestimmte Adjektive (bspw. höflich, menschlich oder ansprechend) auf den Chatbot zutreffen (Holtgraves und Han 2007; Holtgraves et al. 2007).

84 Teilnehmer wurden über Mailinglisten und persönlichen Kontakt rekrutiert. Fünf Antworten wurden aus der Stichprobe entfernt, da sich die jeweiligen Teilnehmer nicht an die Aufgabenstellung des Experiments gehalten oder den Fragenbogen nur durchgeklickt hatten. Insgesamt bestand die Stichprobe daher aus 79 Beobachtungen (25 Frauen

⁸<https://dev.botframework.com>.

⁹<https://www.luis.ai>.

Chat

How much did you pay in total last month?

Chatbot

I paid about 32 €

You

Got it. How many minutes have you spent on the phone last month (national calls)?

Chatbot

652 Minutes

You

How many text messages did you send last month?

Chatbot at 14:23:32

Type a message...

Your Mobile Phone Bill

Your Mobile Phone Plan

Monthly charges:	
• Yellow Basic 1000	19.99 €
Other charges:	
• 2x Data packages 300 MB	6.00 €
• International calls	6.30 €
Total charges:	32.29 €

Your Usage Details

Telephony	
• National calls	652 minutes
• International calls (non EU)	21 minutes
Text	
• Text messages sent	23
Data	
• Included data volume	1000 MB
• Data packages	450 MB
	1450 MB

Abb. 3 Experimentaldesign und Chatbot Szenario (auf Englisch)

und 54 Männer; Durchschnittsalter = 28,83 Jahre), von denen 35 in der Kontrollgruppe und 44 in der Treatmentgruppe waren. Die ungleiche Gruppengröße ist darauf zurückzuführen, dass die zufällige Zuordnung der Teilnehmer über die Experimentalplattform nicht berücksichtigte, dass einige Teilnehmer das Experiment nicht bis zum Ende durchgeführt hatten.

3.4 Ergebnisse und Diskussion des Experiments

Tab. 1 zeigt die deskriptiven Ergebnisse (d. h. den Mittelwert und die Standardabweichung in Klammern) der Auswertung der erhobenen Messinstrumente für die Kontroll- und Treatmentgruppe. In der Analyse wurde auf einen signifikanten Unterschied zwischen den Mittelwerten der beiden Gruppen getestet. Alle Tests wurden einseitig durchgeführt, um zu untersuchen, ob sich eine dynamisch verzögerte Antwortgeschwindigkeit (Treatmentgruppe) positiv auf die Wahrnehmung eines Chatbots auswirkt (d. h. im Vergleich zu einer schnellen Antwortgeschwindigkeit in der Kontrollgruppe).

Die Teststatistik zeigt, dass eine dynamisch verzögerte Antwortgeschwindigkeit eines Chatbots zu einer signifikant höheren wahrgenommenen Menschlichkeit und sozialen Präsenz des Chatbots führt. Darüber hinaus sind Nutzer insgesamt zufriedener mit einem Chatbot mit dynamisch verzögerter Antwortgeschwindigkeit.

Tab. 1 Ergebnisse des Online-Experiments (für weitere Details siehe Gnewuch et al. 2018a)

Gruppe/Chatbot	Wahrgenommene Menschlichkeit ¹	Soziale Präsenz ²	Nutzer-zufriedenheit ²
Treatment: Dynamisch verzögerte Antwortgeschwindigkeit (n = 44)	5,53 (1,67)	3,70 (1,51)	5,17 (0,97)
Kontroll: Schnelle Antwortgeschwindigkeit (n = 35)	4,46 (1,88)	2,95 (1,43)	4,61 (1,34)
Teststatistik	$t(77) = -2,69$, $p = 0,004$, $r = 0,293$	$t(77) = -2,22$, $p = 0,02$, $r = 0,246$	$t(60,04) = -2,088$, $p = 0,02$, $r = 0,260$

¹gemessen mit einer semantischen 9-Punkte-Differenzskala | ²gemessen mit einer 7-Punkt-Likert-Skala

Die Ergebnisse des Online-Experiments deuten darauf hin, dass sich eine dynamische Antwortverzögerung positiv auf die Wahrnehmung des Chatbots beim Nutzer auswirkt. Diese Erkenntnis stützt die Annahme, dass dynamische Antwortverzögerungen eine effektive Möglichkeit sind, um einen Chatbot zu *vermenschlichen* und Konversationen für den Nutzer natürlicher zu gestalten. Darüber hinaus zeigt dieses Experiment, dass relativ subtile soziale Signale, wie zum Beispiel die Antwortgeschwindigkeit eines Chatbots, soziale Erwartungen und Prozesse beim Nutzer auslösen und die Wahrnehmung des Chatbots wesentlich beeinflussen können. Interessanterweise zeigt das Experiment auch, dass Verzögerungen die Zufriedenheit der Nutzer mit der Chatbot-Interaktion erhöhen. Frühere Studien weisen eher darauf hin, dass Verzögerungen in der Reaktionszeit eines Systems zu einer geringeren Nutzerzufriedenheit führen (bspw. (Hoxmeier und DiCesare 2000)). Diese Ergebnisse erscheinen daher auf den ersten Blick widersprüchlich und kontraintuitiv: Warum sollte jemand länger als nötig auf die Antwort eines Chatbots warten wollen?

Im Gegensatz zum hier beschriebenen Experiment haben andere Studien jedoch die Antwortgeschwindigkeiten von Systemen mit grafischen Benutzeroberflächen (bspw. von webbasierten Anwendungen) untersucht, bei denen die Interaktion durch Klicks, Scrollen oder andere Mausbewegungen erfolgt. Durch die Verwendung natürlicher Sprache können Chatbots jedoch mehr oder andere soziale Erwartungen, Prozesse und Reaktionen auslösen als die meisten Anwendungen mit klassischen grafischen Benutzeroberflächen. Die Autoren nehmen daher an, dass das *menschliche Element* von Chatbots ihre Wahrnehmung beim Nutzer stark prägt, wodurch sie bei der Evaluation der Antwortgeschwindigkeit eines Chatbots andere Maßstäbe anlegen als bei der Evaluation von Systemen mit grafischer Benutzeroberfläche. Da Menschen mit anderen Menschen chatten, die in der Regel nicht augenblicklich antworten können, scheinen sie automatisch die gleichen Regeln und Erwartungen auf ihre Interaktion mit einem Chatbot zu übertragen. Dieser Effekt wird noch dadurch verstärkt, dass sich viele Chatbots auf Messaging-Plattformen wie Facebook

Messenger, Slack¹⁰ oder Telegram¹¹ befinden, die sonst nur zur Kommunikation mit anderen Menschen verwendet werden.

Obwohl noch mehr Forschung nötig ist, zeigen die Ergebnisse, dass Nutzer in diesem speziellen Kontext einen Chatbot mit einer verzögerten Antwortgeschwindigkeit, die dynamisch basierend auf der Komplexität berechnet wurde, bevorzugen. Das vorgestellte Gestaltungsprinzip und die Formel zur Berechnung einer dynamischen Antwortverzögerung können daher helfen, Chatbots so zu gestalten, dass sie von ihren Nutzern besser angenommen werden und auch eher genutzt werden.

4 Diskussion

Die vorangegangenen Kapitel verdeutlichen, dass soziale Signale nicht nur in der zwischenmenschlichen Kommunikation, sondern auch in der Interaktion mit Chatbots eine grosse Rolle spielen können. Nutzer reagieren unterbewusst auf bestimmte Signale eines Chatbots, wie zum Beispiel die Attraktivität seines Avatars oder seine unnatürlich schnellen Antworten. Im Rahmen verschiedener Forschungsprojekte konnten die Autoren feststellen, dass es nahezu unmöglich ist, den Einfluss sozialer Signale auf die Interaktion mit Chatbots zu unterbinden. Auch wenn den Nutzern bewusst war, dass sie es mit einem Chatbot zu tun hatten, fanden sich in den Chatverläufen oft Hinweise darauf, dass sie sich dessen noch einmal vergewissern wollten. Während einige Nutzer den Chatbot aktiv nach seiner Identität fragten (z. B. „*Bist du ein Mensch?*“), gab es andere, die den Chatbot mit besonders komplizierten Fragen herausforderten, um sicherzugehen, dass die Antworten nicht doch von einem Menschen kommen.

Aufgrund der Vielzahl an möglichen sozialen Signalen ist es für Forscher und Anwender wichtig, die soziotechnische Gestaltung zu verstehen. Die soziotechnische Gestaltung von Chatbots kann daher auf Erkenntnissen aus der zwischenmenschlichen Kommunikation aufbauen, um soziale Signale von Chatbots sowohl isoliert als auch im Zusammenspiel bestmöglich zu gestalten. Beispielsweise kann die verzögerte Antwortgeschwindigkeit eines Chatbots mit verschiedenen Varianten einer Eingabeanzeige (engl. *typing indicator*) kombiniert werden. Diese Eingabeanzeigen, die dem Nutzer aus Messaging-Plattformen bekannt sind (bspw. drei animierte Punkte in Facebook Messenger), werden ebenfalls als ein soziales Signal wahrgenommen (Gnewuch et al. 2018b). Die soziotechnische Gestaltung von Chatbots sollte daher auch die Abhängigkeiten zwischen verschiedenen sozialen Signalen berücksichtigen.

In der Regel wird im Laufe des Entwicklungsprozesses die konkrete Gestaltung einzelner sozialer Signale, wie etwa der Sprachstil oder das Erscheinungsbild des Chatbots, festgelegt und nach Einführung des Chatbots nicht mehr geändert. Nichtsdestotrotz sind die technischen Voraussetzungen für eine Adaption der sozialen Signale eines Chatbot auf verschiedene Nutzer, Kontexte oder Aufgaben bereits

¹⁰<https://slack.com>.

¹¹<https://telegram.org>.

heute gegeben. Ein Chatbot kann sich nicht nur mit jeder Konversation neues Wissen aneignen, sondern sich auch an einen bestimmten Nutzer oder Kontext anpassen. Im folgenden Kapitel werden daher die Möglichkeiten der adaptiven Gestaltung sozialer Signale eines Chatbots diskutiert.

4.1 Adaptive Gestaltung der sozialen Signale von Chatbots

Aus der zwischenmenschlichen Kommunikation ist bekannt, dass Menschen ihre sozialen Signale beeinflussen, um sich an andere Menschen und ihre Umgebung anzupassen. Beispielsweise verwenden Menschen im beruflichen Umfeld oft das distanzierte *Sie*, während sie im privaten Umfeld in aller Regel das intimere *Du* einsetzen (Wrede-Grischkat 2007). Analog wird eine derartige Anpassung an äussere Umgebungsvariablen oft auch von Chatbots gefordert, um als intelligente Akteure wahrgenommen zu werden (Knote et al. 2018). Ein Grund dafür ist, dass die Wirkung einer Mensch-Chatbot Interaktion unter anderem durch den (1) einzelnen Nutzer, (2) den Kontext und (3) die Aufgabe beeinflusst werden (Benyon et al. 2005). Daher können Erkenntnisse zur Gestaltung eines sozialen Signals nicht ohne Weiteres auf einen anderen Nutzer, einen anderen Kontext oder eine andere Aufgabe übertragen werden. Dieser Aspekt soll anhand der Ergebnisse einiger Studien in den folgenden Abschnitten verdeutlicht werden.

Potentielle Nutzer von Chatbots können sich hinsichtlich ihrer demografischen (Alter, Geschlecht, usw.) oder geografischen Daten (z. B. Herkunftslands), aber auch hinsichtlich ihres Nutzungsverhaltens (sporadische Nutzer, erfahrene Nutzer, usw.) stark unterscheiden (Catrambone et al. 2004). Aus diesem Grund können sich soziale Signale von Chatbots auch unterschiedlich auf verschiedene Nutzergruppen auswirken, weshalb ihre individuellen Eigenschaften bei der soziotechnischen Gestaltung berücksichtigt werden sollten. Braslavski et al. (2018) haben beispielsweise gezeigt, dass Witze eines Chatbots abhängig von Alter, vom Geschlecht und den Sprachkenntnissen des jeweiligen Nutzers unterschiedlich wahrgenommen werden. Endrass, Rehm und André (2011) untersuchten den Einfluss von Smalltalk abhängig vom kulturellem Hintergrund des Nutzers und zeigten, dass Nutzer in westlichen Kulturen Smalltalk von Chatbots anders wahrnehmen als Nutzer aus asiatischen Kulturen. Weiter fanden Isbister und Nass (2000) heraus, dass Chatbots als angenehmer bewertet werden, wenn ihre Persönlichkeitseigenschaften mit denen des Nutzers übereinstimmen.

Weiter beeinflussen die Aufgabe und der Kontext, in dem die Aufgabe stattfindet, die Wirkung sozialer Signale. So bevorzugen Nutzer in einer Banking-Umgebung ein Erscheinungsbild mit einer klassischen und formaleren Kleidung, wohingegen die Kleidung des Assistenten in einem virtuellen Reisebüro oder Kinokasse keine Rolle zu spielen scheint (McBreen 2002). Auch ist ein Nutzer bereit, mehr intime Informationen in einem Gesundheitskontext offenzulegen, wenn er direkt und persönlich angesprochen wird (Sah und Peng 2015). Eine direkte und persönliche Ansprache wirkt sich zudem auch positiv auf die Entwicklung einer vertrauensvollen Beziehung zwischen dem Nutzer und einem Fitness-Chatbot aus (Bickmore und

Picard 2005). Wenn aber finanzielle Informationen vom Nutzer abgefragt werden, wirkt sich eine direkte Ansprache negativ auf die Offenlegung dieser Informationen aus (Puzakova et al. 2013). Fogg (2002) argumentiert daher, dass das Ausmass der Ausgestaltung von sozialen Signalen reduziert werden sollte, wenn es die Hauptaufgabe des Chatbots ist schnell und effektiv Informationen zur Verfügung zu stellen. In solchen Fällen lenken soziale Signale seiner Argumentation nach von der sinnvollen Nutzung des Chatbots ab. Wenn es jedoch um Freizeit-, Unterhaltungs- und Bildungsaufgaben geht, ist eine ausgiebige Gestaltung von sozialen Signalen sinnvoll und sogar vom Nutzer gewünscht (Fogg 2002).

Abschliessend lässt sich feststellen, dass es für eine erfolgreiche soziotechnische Gestaltung von sozialen Signalen notwendig ist, diese an die Umgebung anzupassen. Da sich die Nutzer, der Kontext, aber auch die Aufgabe in einer Interaktion mit einem Chatbot kontinuierlich verändern können, ist es sinnvoll den Chatbot so zu gestalten, dass er sich adaptiv an Veränderungen der Umgebung anpassen kann. Die notwendigen Umgebungsinformationen kann der Chatbot hierfür aus verschiedenen Quellen auslesen. So kann er zum Beispiel auf die Informationen aus Nutzerprofilen von sozialen Plattformen zugreifen (z. B. Facebook) oder den Nutzer direkt nach wichtigen Informationen fragen (bspw. nach Geschlecht, Alter oder Erfahrung). Weiter kann der Chatbot die vom Nutzer geschriebenen Nachrichten in Echtzeit analysieren, um Erkenntnisse über die Stimmung (engl. *sentiment*), die Emotionen oder sogar über die Persönlichkeit des Nutzers zu gewinnen und zu nutzen (Clavel und Callejas 2016). Die Nutzung dieser Erkenntnisse, ihre Modellierung und die anschliessende Durchführung von adaptiven Anpassungsstrategien ermöglichen es, die sozialen Signale des Chatbots an die Bedürfnisse, die Verhaltensweisen und die Situation des individuellen Nutzers anzupassen und somit die sozialen Reaktionen positiv zu beeinflussen.

4.2 Wie menschenähnlich sollte ein Chatbot sein?

Sowohl in der Forschung als auch in der Praxis wird oft darüber diskutiert, wie sehr Technologie dem Menschen ähnlich sein sollte beziehungsweise sein darf (z. B. (Wirtschaftswoche 2016)). Auch bei der Gestaltung von Chatbots muss zwangsläufig über den Grad der Menschlichkeit entschieden werden. Insbesondere die Gestaltung der sozialen Signale eines Chatbots hat einen grossen Einfluss darauf, wie menschenähnlich ein Chatbot wahrgenommen wird. Wie in Abschn. 3 dargestellt, reicht es bereits aus, die Antwortgeschwindigkeit eines Chatbots zu verzögern, um die Wahrnehmung des Nutzers dahingehend zu beeinflussen. Analog dazu können beliebige soziale Signale eingesetzt werden, um den Chatbot weiter zu vermenschlichen. Dies wirft die Frage auf, ob eine Vermenschlichung des Chatbots sinnvoll ist. So führte die starke Vermenschlichung des bekannten Chatbots Anna von IKEA¹² dazu, dass viele Nutzer keine inhaltlichen Fragen rund um IKEAs Produkte stellten, sondern sich hauptsächlich zum äusseren Erscheinungsbild von Anna

¹²Den Chatbot Anna von IKEA (www.ikea.com) gibt es seit 2016 nicht mehr.

äusserten. Andererseits gibt es auch Nutzer, die sich über die Roboterhaftigkeit von Chatbots beschwerten (Brandtzaeg und Følstad 2018). Auch die Präsentation von Google's Sprachassistenten *Duplex*¹³ hat zu grossen Diskussionen darüber geführt, ob es ethisch vertretbar ist, einen Menschen darüber im Unklaren zu lassen, ob er mit einer Maschine oder einem Menschen interagiert (Kremp 2018). Auch Joseph Weizenbaum (1976), der Entwickler von ELIZA, hat schon frühzeitig kritisiert, dass Maschinen viele menschlichen Qualitäten wie Mitgefühl, Kreativität und Intuition niemals besitzen werden und somit auch nicht in Kontexten, in denen diese erforderlich sind, eingesetzt werden sollten.

In der Forschung gehen die Meinungen darüber auseinander, ob die Vermenschlichung eines Chatbots sinnvoll und ethisch vertretbar ist. Die eine Seite argumentiert, dass es bei Chatbots primär um die Funktion gehen sollte und eine Vermenschlichung durch soziale Signale nur von der eigentlichen Aufgabe ablenkt. Daher sollten Entwickler das Ausmass der sozialen Signale eines Chatbots soweit wie möglich reduzieren. Häufig berufen sich Forscher auf die Theorie des *Uncanny Valley* (deut. Unheimliches Tal), die besagt, dass künstliche Figuren, wie etwa Roboter oder Chatbots, mit zunehmender Vermenschlichung positiver evaluiert werden und eine zunehmend positive Reaktion beobachtbar ist (Mori 1970). Allerdings wird mit steigendem Grad der Menschenähnlichkeit ein Punkt erreicht, an dem sich diese positive Reaktion in eine negative Reaktion umkehrt. Ab diesem Punkt löst die unvollkommene Menschlichkeit Unbehagen beim Nutzer aus und die Akzeptanz steigt erst wieder an, wenn die künstliche Figur nahezu identisch mit dem Menschen ist. Übertragen auf die Gestaltung von Chatbots bedeutet dies, dass bei dem Versuch, einen Chatbot möglichst menschenähnlich wirken zu lassen, die Gefahr besteht, dass der Nutzer durch seine unvollkommene Menschlichkeit abgeschreckt wird. Diese Gefahr besteht insbesondere dann, wenn der Nutzer im Unklaren darüber gelassen wird, ob er mit einem Menschen oder einem Chatbot interagiert (vgl. (Jentsch 1997)). Diese Unklarheit führt nicht selten dazu, dass sich Nutzer im Kundenservice-Chat erst einmal vergewissern, ob sie es mit einem Menschen oder Chatbot zu tun haben (Mooallem 2017). Folglich empfehlen diese Forscher, dem Nutzer bereits zu Beginn klarzumachen, dass sie es mit einem Chatbot zu tun haben und in der Gestaltung des Chatbots auf soziale Signale zu verzichten, um keine falschen Erwartungen beim Nutzer zu wecken oder sogar Unbehagen bei ihnen auszulösen.

Andererseits erwarten viele Nutzer, obwohl sie wissen, dass ein Chatbot eine Maschine ist, dass er soziale Signale sendet, da sie diese aus der Interaktion mit Menschen gewohnt sind (Nass und Moon 2000). Aus diesem Grund wird argumentiert, dass eine Vermenschlichung des Chatbots vom Nutzer gewünscht und folglich auch ethisch vertretbar ist. Hinzukommt, dass sich allein durch die Verwendung von natürlicher Sprache als Interaktionsform eine gewisse Assoziation mit menschlichem Verhalten nicht komplett vermeiden lässt. Zusammenfassend lässt sich daher sagen, dass es vermutlich keine pauschale Antwort auf die Frage nach dem optimalen Grad der Menschenähnlichkeit eines Chatbots gibt. Der Ansatz einer soziotechnischen Gestal-

¹³<https://www.youtube.com/watch?v=D5VN56jQMWM>.

tung lässt sich aber auch hier anwenden, um die Wirkung der sozialen Signale eines Chatbots zu untersuchen und gegebenenfalls Anpassungen vorzunehmen, um den Grad der Menschenähnlichkeit zu reduzieren oder zu erhöhen.

5 Fazit

Soziale Signale sind nicht nur in der zwischenmenschlichen Kommunikation wichtig. Auch bei der Interaktion mit Chatbots nehmen Nutzer dessen soziale Signale (Antwortgeschwindigkeit, Emojis, usw.) unterbewusst wahr, wodurch soziale Reaktionen bei ihnen ausgelöst werden. In der Gestaltung von Chatbots gibt es eine Vielzahl an sozialen Signalen, die einen grossen Einfluss auf die Wahrnehmung des Chatbots beim Nutzer haben können. Für Forscher und Anwender ist es daher wichtig, die soziotechnische Gestaltung von Chatbots zu verstehen. Die soziotechnische Gestaltung versucht, sowohl technische als auch soziale Faktoren und Erkenntnisse im Gestaltungsprozess eines Chatbots zu berücksichtigen. So können soziale Signale von Chatbots optimal gestaltet und adaptiv im Laufe der Konversation angepasst werden, um für eine positive Wahrnehmung des Chatbots beim Nutzer zu sorgen, ohne den Chatbot zu stark zu vermenschlichen.

Danksagung Die Autoren dieses Kapitels danken allen Teilnehmern der hier vorgestellten Studien und allen Kolleginnen und Kollegen am Institute of Information Systems and Marketing für ihre Unterstützung. Ein besonderer Dank gilt Marc T. P. Adam für seine Mitarbeit in diesem Forschungsprojekt.

Literatur

- Appel J, von der Pütten A, Krämer NC, Gratch J (2012) Does Humanity Matter? Analyzing the Importance of Social Cues and Perceived Agency of a Computer System for the Emergence of Social Reactions during Human-Computer Interaction. *Advances in Human-Computer Interaction*, 2012:1–10.
- Araujo T (2018) Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior*, 85: 183–189.
- Benyon D, Turner P, Turner S (2005) *Designing interactive systems: People, activities, contexts, technologies*. Pearson Education.
- Bickmore TW, Picard RW (2005) Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction*, 12(2): 293–327.
- Bitkom (2018) Jeder Vierte will Chatbots nutzen. Bitkom Research. <https://www.bitkom.org/Presse/Presseinformation/Jeder-Vierte-will-Chatbots-nutzen.html>. Erstellt: 18.01.2017. Letzter Zugriff: 08. August 2018.
- Brandtzaeg PB, Følstad A (2018) Chatbots: Changing User Needs and Motivations. *Interactions*, 25(5): 38–43.
- Braslavski P, Blinov V, Bolotova V, Pertsova K (2018) How to Evaluate Humorous Response Generation, Seriously? In: *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval – CHIIR '18*. New York, New York, USA.

- Burgoon J, Guerrero L, Floyd K (2010) Nonverbal communication. Routledge, New York, NY, USA.
- Cafaro A, Vilhjálmsson HH, Bickmore T (2016) First Impressions in Human--Agent Virtual Encounters. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 23(4): 24.
- Catrambone R, Stasko J, Xiao J (2004) ECA as user interface paradigm. In: *From brows to trust*. Springer, 239–267.
- Chae SW, Lee KC, Seo YW (2016) Exploring the Effect of Avatar Trust on Learners' Perceived Participation Intentions in an e-Learning Environment. *International Journal of Human-Computer Interaction*, 32(5): 373–393.
- Clavel C, Callejas Z (2016) Sentiment analysis: from opinion mining to human-agent interaction. *IEEE Transactions on Affective Computing*, 7(1): 74–93.
- Colby KM, Weber S, Hilf FD (1971) Artificial paranoia. *Artificial Intelligence*, 2(1): 1–25.
- Collier G (2014) Emotional expression. Psychology Press, New York, NY, USA.
- Crozier R (2017) Lufthansa delays chatbot's responses to make it more 'human'. *iNews Australia*. <https://www.itnews.com.au/news/lufthansa-delays-chatbots-responses-to-make-it-more-human-462643>. Erstellt: 24. Mai 2017. Letzter Zugriff: 30. August 2018.
- D'Onofrio S, Portmann E, Franzelli M, Bürki C (2018) Cognitive Computing: Theoretische Grundlagen und Praxisbeispiele der Schweizerischen Post. *Informatik-Spektrum*, 41(2): 113–122.
- Dale R (2016) The return of the chatbots. *Natural Language Engineering*, 22(05): 811–817.
- de Visser EJ, Monfort SS, McKendrick R, Smith MABB, McKnight PE, Krueger F, Parasuraman R (2016) Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology: Applied*, 22(3): 331–349.
- Derrick DC, Meservy TO, Jenkins JL, Burgoon JK, Nunamaker JF (2013) Detecting Deceptive Chat-Based Communication Using Typing Behavior and Message Cues. *ACM Transactions on Management Information Systems*, 4(2): 1–21.
- Dybala P, Ptaszynski M, Rzepka R, Araki K (2009) Activating Humans with Humor--A Dialogue System That Users Want to Interact with. *IEICE TRANSACTIONS on Information and Systems*, 92(12): 2394–2401.
- Endrass B, Rehm M, André E (2011) Planning Small Talk behavior with cultural influences for multiagent systems. *Computer Speech, Language*, 25(2): 158–174.
- Fogg BJ (2002) Computers as Persuasive Social Actors. In: *Persuasive Technology: Using Computers to Change What We Think and Do*. Morgan Kaufmann Publishers, San Francisco, CA, USA, 89–120.
- Følstad A, Brandtzæg PB (2017) Chatbots and the new world of HCI. *Interactions*, 24(4): 38–42.
- Gartner (2018) Gartner Says 25 Percent of Customer Service Operations Will Use Virtual Customer Assistants by 2020. <https://www.gartner.com/newsroom/id/3858564>. Erstellt: 19. Februar 2018. Letzter Zugriff: 11. September 2018.
- Gefen D, Straub DW (1997) Gender Differences in the Perception and Use of E-Mail: An Extension to the Technology Acceptance Model. *MIS Quarterly*, 21(4): 389–400.
- Gnewuch U, Morana S, Adam MTP, Maedche A (2018a) Faster is Not Always Better: Understanding the Effect of Dynamic Response Delays in Human-Chatbot Interaction. In: *Proceedings of the European Conference on Information Systems (ECIS)*. Portsmouth, UK.
- Gnewuch U, Morana S, Adam MTP, Maedche A (2018b) , The Chatbot is typing ... ' – The Role of Typing Indicators in Human-Chatbot Interaction. In: *Proceedings of the 17th Annual Pre-ICIS Workshop on HCI Research in MIS*. San Francisco, CA, USA.
- Gnewuch U, Morana S, Maedche A (2017) Towards Designing Cooperative and Social Conversational Agents for Customer Service. In: *Proceedings of the 38th International Conference on Information Systems (ICIS)*. Seoul, South Korea.
- Hevner AR, March ST, Park J, Ram S (2004) Design Science in Information Systems Research. *MIS Quarterly*, 28(1): 75–105.
- Holtgraves T, Han T-L (2007) A procedure for studying online conversational processing using a chat bot. *Behavior Research Methods*, 39(1): 156–163.

- Holtgraves T, Ross SJ, Weywadt CR, Han T-L (2007) Perceiving artificial social agents. *Computers in Human Behavior*, 23(5): 2163–2174.
- Hoxmeier JA, DiCesare C (2000) System Response Time and User Satisfaction : An Experimental Study of Browser-based Applications. In: *Proceedings of the 6th Americas Conference on Information Systems (AMCIS)*. Long Beach, CA, USA.
- Isbister K, Nass C (2000) Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics. *International Journal of Human-Computer Studies*, 53: 251–267.
- Jentsch E (1997) On the psychology of the uncanny (1906). *Angelaki: Journal of the Theoretical Humanities*, 2(1): 7–16.
- Johnson K (2017) Facebook Messenger hits 100,000 bots. <https://venturebeat.com/2017/04/18/facebook-messenger-hits-100000-bots/>. Erstellt: 18. April 2017. Letzter Zugriff: 08. August 2018.
- Khawaja MA, Chen F, Marcus N (2010) Using language complexity to measure cognitive load for adaptive interaction design. In: *Proceedings of the 15th International Conference on Intelligent User Interfaces (IUI '10)*.
- Kincaid JP, Fishburne RP, Rogers RL, Chissom BS (1975) Derivation Of New Readability Formulas (Automated Readability Index, Fog Count And Flesch Reading Ease Formula) For Navy Enlisted Personnel. Branch Report, 8–75.
- Knote R, Janson A, Eigenbrod L, Söllner M (2018) The What and How of Smart Personal Assistants: Principles and Application Domains for IS Research. In: *Multikonferenz Wirtschaftsinformatik*. Lüneburg, Germany.
- Krämer NC (2005) Social Communicative Effects of a Virtual Program Guide. In: R. T. Panayiotopoulos T., Gratch J., Aylett R., Ballin D., Olivier P. (ed.), *Intelligent Virtual Agents. IVA 2005*. Springer, Berlin, Heidelberg.
- Krämer NC, Kopp S, Becker-Asano C, Sommer N (2013) Smile and the world will smile with you – The effects of a virtual agent's smile on users' evaluation and behavior. *International Journal of Human-Computer Studies*, 71(3): 335–349.
- Kremp M (2018) Künstliche Intelligenz: Google Duplex ist gruselig gut. Spiegel. <http://www.spiegel.de/netzwelt/web/google-duplex-auf-der-i-o-gruselig-gute-kuenstliche-intelligenz-a-1206938.html>. Erstellt: 09. Mai 2018. Letzter Zugriff: 30. August 2018.
- Leathers DG (1976) Nonverbal communication systems. Allyn and. Bacon, Inc., Boston, MA, USA.
- Li J, Zhou MX, Yang H, Mark G (2017) Confiding in and Listening to Virtual Agents: The Effect of Personality. In: *Proceedings of the 22nd International Conference on Intelligent User Interfaces – IUI '17*.
- Liao QV, Hussain MM, Chandar P, Davis M, Crasso M, Wang D, Muller M, Shami NS, Geyer W (2018) All Work and no Play ? Conversations with a Question-and-Answer Chatbot in the Wild. In: *CHI 18*.
- Markoff J, Mozur P (2015) For Sympathetic Ear, More Chinese Turn to Smartphone Program. *New York Times*. <https://www.nytimes.com/2015/08/04/science/for-sympathetic-ear-more-chinese-turn-to-smartphone-program.html>. Erstellt: 31. Juli 2015. Letzter Zugriff: 05. Dezember 2018.
- Mayer RE, Johnson WL, Shaw E, Sandhu S (2006) Constructing computer-based tutors that are socially sensitive: Politeness in educational software. *International Journal of Human-Computer Studies*, 64(1): 36–42.
- McBreen H (2002) Embodied conversational agents in e-commerce applications. In: *Socially Intelligent Agents*. Springer, 267–274.
- McTear MF (2017) The Rise of the Conversational Interface: A New Kid on the Block? In: *FETLT 2016*. Springer, Cham, 38–49.

- Mooallem J (2017) Mr. Know-It-All: Is it OK For Me to Ask Customer Service Reps if They're Robots? <https://www.wired.com/story/mr-know-it-all-are-customer-service-reps-robots/>. Erstellt: 10. März 2017. Letzter Zugriff: 30. August 2018.
- Moon Y (1999) The effects of physical distance and response latency on persuasion in computer-mediated communication and human-computer communication. *Journal of Experimental Psychology: Applied*, 5(4): 379–392.
- Moore RJ, Arar R, Ren G-J, Szymanski MH (2017) Conversational UX Design. In: Proceedings of the SIGCHI Conference: Extended Abstracts on Human Factors in Computing Systems (CHI '17). Denver, CO, USA.
- Morana S, Friemel C, Gnewuch U, Maedche A, Pfeiffer J (2017) Interaktion mit smarten Systemen – Aktueller Stand und zukünftige Entwicklungen im Bereich der Nutzerassistenz. *Wirtschaftsinformatik, Management*, 9(5): 42–51.
- Mori M (1970) The uncanny valley. *Energy*, 7(4): 33–35.
- Mumford E (2006) The story of socio-technical design: reflections on its successes, failures and potential. *Information Systems Journal*, 16(4): 317–342.
- Nass C, Fogg BJ, Moon Y (1996) Can computers be teammates? *International Journal of Human Computer Studies*, 45(6): 669–678.
- Nass C, Moon Y (2000) Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, 56(1): 81–103.
- Nass C, Steuer J, Tauber ER (1994) Computers are social actors. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Boston, MA, USA.
- Puzakova M, Rocereto JF, Kwak H (2013) Ads are watching me: A view from the interplay between anthropomorphism and customisation. *International Journal of Advertising*, 32(4): 513–538.
- Sah YJ, Peng W (2015) Effects of visual and linguistic anthropomorphic cues on social perception, self-awareness, and information disclosure in a health website. *Computers in Human Behavior*, 45: 392–401.
- Stucki T, D'Onofrio S, Portmann E (2018) Chatbot – Der digitale Helfer im Unternehmen: Praxisbeispiele der Schweizerischen Post. *HMD Praxis Der Wirtschaftsinformatik*, 55(4): 725–747.
- Stucki T, D'Onofrio S, Portmann E (2020) Chatbots gestalten mit Praxisbeispielen der Schweizerischen Post. HMD Best Paper Award 2018. Essentials. Springer Vieweg, Wiesbaden.
- Trager GL (1958) Paralanguage: A first approximation. *Studies in Linguistics*, 13: 1–11.
- Verhagen T, van Nes J, Feldberg F, van Dolen W (2014) Virtual Customer Service Agents: Using Social Presence and Personalization to Shape Online Service Encounters. *Journal of Computer-Mediated Communication*, 19(3): 529–545.
- Wallace RS (2009) The Anatomy of A.L.I.C.E. In: R. Epstein, G. Roberts, G. Beber (eds.), *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*. Springer Netherlands, Dordrecht, 181–210.
- Walther JB (1992) Interpersonal Effects in Computer-Mediated Interaction. *Communication Research*, 21(4): 460–487.
- Walther JB, Tidwell LC (1995) Nonverbal cues in computer-mediated communication, and the effect of chronemics on relational communication. *Journal of Organizational Computing*, 5(4): 355–378.
- Weizenbaum J (1966) ELIZA – A Computer Program for the Study of Natural Language Communication between Man and Machine. *Communications of the ACM*, 9(1): 36–45.
- Weizenbaum J (1976) *Computer power and human reason: From judgment to calculation*. W. H. Freeman, Co, New York, NY, USA.
- Wirtschaftswoche (2016) Menschenähnliche Computerwesen: Wie die Grenzen zwischen Menschen und Robotern verschwimmen. <https://www.wiwo.de/technologie/forschung/menschen-aehnliche-computerwesen-wie-die-grenzen-zwischen-menschen-und-robotern-verschwimmen/14684090.html>. (Erstellt: 16. Oktober 2016. Letzter Zugriff: 08. August 2018).
- Wrede-Grischkat R (2007) *Manieren und Karriere: internationale Verhaltensregeln für Führungskräfte*. Springer.

Teil IV

Praktische Anwendungsfälle

Kollektive Intelligenz von Mensch und Maschine – Fuzzy Logik zur Einbindung von Expertenwissen in die Wegeplanung von Fahrerlosen Transportfahrzeugen

Sarah Uttendorf

Zusammenfassung

Dieser Artikel beschäftigt sich mit der Wegeplanung von Fahrerlosen Transportfahrzeugen. Es wird eine Übersicht über die verschiedenen Gruppen von Wegeplanungsalgorithmen, wie zum Beispiel Geometrische Ansätze oder Algorithmen der Computational Intelligence gegeben. Ein besonderes Augenmerk liegt auf der Kombination von menschlichem Wissen mit Wegeplanungsalgorithmen. Ein konkretes Beispiel für ein Expertensystem wird vorgestellt, das menschliches Wissen in Form eines Fuzzy Inferenz Systems speichert. Der Output des Fuzzy Inferenz Systems interagiert mit einem A*-Algorithmus. Als Ergebnis entstehen Wegenetze, die sowohl anwendbar, als auch effizient sind. Der Artikel endet mit einer Evaluierung, wann Ansätze des Cognitive Computings sinnvoll für die Wegeplanung von Fahrerlosen Transportfahrzeugen eingesetzt werden können.

1 Was sind Fahrerlose Transportfahrzeuge?

Im Zuge der fortschreitenden Automatisierung innerhalb der Logistik und der Produktion, hat sich in den letzten Jahrzehnten auch der Materialtransport weitergehend automatisiert. Immer häufiger sind auf modernen Produktions- und Kommissionierungsflächen anstatt Gabelstapler und manuellen Flurförderzeugen sogenannte

Dieser Beitrag basiert auf „Uttendorf, S.: Automated Generation of Roadmaps for Automated Guided Vehicle Systems. In: Behrens, B.-A.; Nyhuis, P.; Overmeyer, L. (Hrsg.): Berichte aus dem IPH, Band 3/2018, TEWISS – Technik und Wissen GmbH, Garbsen 2018. ISBN: 978-3-95900-218-9“.

S. Uttendorf (✉)
Hannover, Deutschland
E-Mail: sarah.uttendorf@web.de

Fahrerlose Transportfahrzeuge zu sehen. Abb. 1 zeigt ein Beispiel für ein kleinskaliges Fahrerloses Transportfahrzeug, das für den automatisierten Transport von Kleinladungsträgern innerhalb einer Produktionsumgebung zuständig ist.

Laut dem Verein Deutscher Ingenieure (VDI) sind Fahrerlose Transportfahrzeuge „*flurgebundene Fördermittel mit eigenem Fohrantrieb, die automatisch gesteuert und berührungslos geführt werden*“ (VDI 2510 (2005)). Sie sind für den Materialtransport, nicht aber für den Personentransport zuständig. Für den störungsfreien, automatisierten Materialtransport sind die Fahrerlose Transportfahrzeuge mit verschiedenen Arten von Sensoren ausgestattet. Diese Sensorik lässt sich unterteilen in *Sensorik zum Personenschutz* und *Sensorik zur Lokalisierung und Navigation*. Vor allem Laserscanner werden häufig in Fahrerlose Transportfahrzeuge verbaut, da sie simultan zum einen für den Personenschutz und zum anderen zur Navigation verwendet werden können. Häufig wird die Lokalisierung und Navigation mittels des SLAM-Verfahrens gelöst. SLAM steht für *Simultaneous Localization and Mapping* (Nüchter 2009). Das Verfahren versucht das Henne-Ei-Problem der Navigation von Fahrerlosen Transportfahrzeuge zu lösen: Zur Navigation-Erstellung einer Karte ist das Wissen über den Standort des Fahrerlosen Transportfahrzeugs nötig, gleichzeitig ist aber für die Bestimmung des Standorts eine initiale Karte notwendig (Ullrich 2014). Das SLAM-Verfahren löst dieses Problem, in dem es einen initialen Startpunkt des Fahrerlosen Transportfahrzeugs festsetzt und von diesem dann anfängt, mittels Sensoren (z. B. Ultraschall, Laserscanner) eine Karte aufzunehmen bis die Umgebung vollständig kartiert wurde.

Meist werden mehrere Fahrerlose Transportfahrzeuge verwendet, um den gesamten intralogistischen Materialtransport innerhalb eines Abschnitts zu realisieren. Sie sind dann Teil eines Fahrerlosen Transportsystems. Fahrerlose Transportsysteme bestehen meist aus den folgenden fünf Bestandteilen, die in Abb. 2 dargestellt sind.

Abb. 1 Beispiel eines kleinskaligen Fahrerlosen Transportfahrzeugs (Rexroth AG)



Bestandteile eines Fahrerlosen Transportsystems				
Ein oder mehrere Fahrerlose Transportfahrzeuge	Leitsteuerung	Einrichtungen zur Standort- und Lageerfassung	Einrichtungen zur Datenübertragungen	Infrastruktur und periphere Einrichtungen

Abb. 2 Bestandteile eines Fahrerlosen Transportsystems (VDI 2510)

Ein Fahrerloses Transportsystems besteht aus einem oder mehreren Fahrerlosen Transportfahrzeuge. Die Koordination der Fahrzeuge und die Allokation der Aufträge an die Fahrzeuge erfolgt über eine Leitsteuerung. Sie bildet die zentrale Intelligenz des Systems, die massgeblich für die Effizienz des Fahrerlosen Transportsystems verantwortlich ist. In ihr sind unter anderem die Algorithmen für ein effizientes Routing hinterlegt, als auch Algorithmen zur geeigneten Disposition der Transportaufträge.

Weiterhin gibt es noch Einrichtungen zur Standort- und Lageerfassung, die für die Navigation der Fahrzeuge von zentraler Bedeutung sind. Einrichtungen zu Datenübertragungen sind zudem notwendig, um mit den einzelnen Fahrzeugen zu kommunizieren, Ablade- und Abholaufträge zu generieren oder Wegabschnitte zu sperren. Zum Einsatz von Fahrerlosen Transportsysteme werden ausserdem spezielle Infrastrukturen und periphere Einrichtungen, wie beispielsweise Übergabestationen, für Transportkisten benötigt.

2 Herausforderungen bei der Wegeplanung von Fahrerlosen Transportsystemen

Zentralen Einfluss auf die Effizienz des Fahrerlosen Transportsystems hat das Wegenetz, auf dem die Fahrzeuge sich bewegen. Die Herausforderung darin liegt, dass es zwar Richtlinien und Normen, zum Beispiel zur Einhaltung von Mindestabständen gibt, aber keine Grundsätze für eine Gestaltung von effizienten Wegenetzen existieren. Für ein effizientes Wegenetz muss zwingend eine Vielzahl von verschiedenen Inputfaktoren betrachtet werden. Dazu gehören unter anderem:

- **Transportmatrix:** Die Transportmatrix gibt Aufschluss darüber, zu welchem Zeitpunkt welche Menge von Material von verschiedenen Startstationen zu Zielstationen transportiert werden muss. Aus diesen Informationen können wichtige Eigenschaften des zukünftigen Wegenetzes, wie etwa die Auslastung einzelner Wege oder die Notwendigkeit von Alternativwegen, berechnet werden.
- **Layout der Umgebung:** Das Layout der Hallenumgebung gibt an, wo Freiflächen zur Wegenerierung vorhanden sind. Zudem gibt es wertvolle Informationen über Risikobereiche (z. B. Kantinenumgebungen, Ein- und Ausgänge), die bei der Wegenerierung bei Möglichkeit vermieden werden sollten.

- **Fahrzeugeigenschaften:** Die Eigenschaften der verwendeten Fahrzeuge haben einen grossen Einfluss auf Kenngrössen, wie etwa der Auslastung der Wege. Werden beispielsweise langsame Fahrzeuge verwendet, blockieren diese länger Wegabschnitte als schnellere Fahrzeuge. Ausserdem haben die Kinematikeigenschaften, wie etwa die Breite und Länge des Fahrzeugs oder der Wenderadius, einen Einfluss auf die Hüllkurve des Fahrzeugs. Die Hüllkurve beeinflusst die notwendigen Mindestabstände zu anderen Wegen oder Wänden.

Zurzeit werden die Wegenetze vor allem durch menschliche Systemplaner erstellt, die ein breites Erfahrungswissen darüber haben, wie ein effizientes Wegenetz aussehen muss. Die manuelle Planung durch einen Systemplaner ist allerdings nur in kleinen Umgebungen möglich, in denen die oben genannten Inputfaktoren noch alle von einem Menschen berücksichtigt werden können. Ausserdem hat die manuelle Planung den Nachteil, dass sie wenig standardisiert ist und damit nur schwer reproduzierbar ist. Ein weiterer Nachteil ist die zeitliche Dauer der manuellen Planung. Für mittelgrosse Layouts kann die Planung mehrere Wochen dauern, was die Erprobung von verschiedenen Varianten eines Wegenetzes sehr umständlich macht. Die menschlichen Planer haben zudem nur eine begrenzte zeitliche Verfügbarkeit und bei einem Wechsel des Planers in ein anderes Unternehmen geht dieses Know-How verloren.

3 Cognitive Computing zur Erstellung von Wegenetzen

Die Aufgabe ein Wegenetz für Fahrerlose Transportfahrzeuge zu erzeugen, lässt sich definieren als die Aufgabe einen kollisionsfreien Weg von einem Startpunkt zu einem Endpunkt zu finden, während ein bestimmter Transportauftrag in der Zwischenzeit erfüllt werden muss (Turban und Aronson 1998). Die ersten Verfahren beschränkten sich zunächst darauf, den kürzesten Weg zwischen zwei Punkten im Raum zu finden (z. B. mithilfe des A*-Algorithmus (Hart et al. 1968) oder des Dijkstra-Algorithmus (Dijkstra 1959)).

Es existiert keine offizielle Unterteilung der unterschiedlichen Wegenetzalgorithmen. Grob lassen sich die Methoden aber in *offline* und *online* Planung (Raja und Pugazhethi 2012) unterteilen. Bei der offline Planung wird global ein Wegenetz mit allem bis dahin verfügbaren Wissen gestartet. Dieses Wissen kann unter anderem stationäre Objekte beinhalten, aber auch Sperrzonen. In diesem Fall liegen keine aktuellen Sensorwerte vor, die beispielsweise dynamische Objekte in den Wegen mitberücksichtigen und ein kurzfristiges Ausweichen realisieren. Bei der online Planung ist es zudem möglich Echtzeit-Inputs zu erhalten. Darunter würden etwa Signale der Laserscanner oder Ultraschallsensoren fallen. Mit deren Hilfe könnte zum Beispiel eine lokale Generierung eines Alternativweges zur Hindernisumfahrung ermöglicht werden. Dieser Artikel beschäftigt sich mit der offline Planung. Abb. 3 zeigt eine Kategorisierung der verschiedenen Wegenetzalgorithmen, die keinen Anspruch auf Vollständigkeit hat.

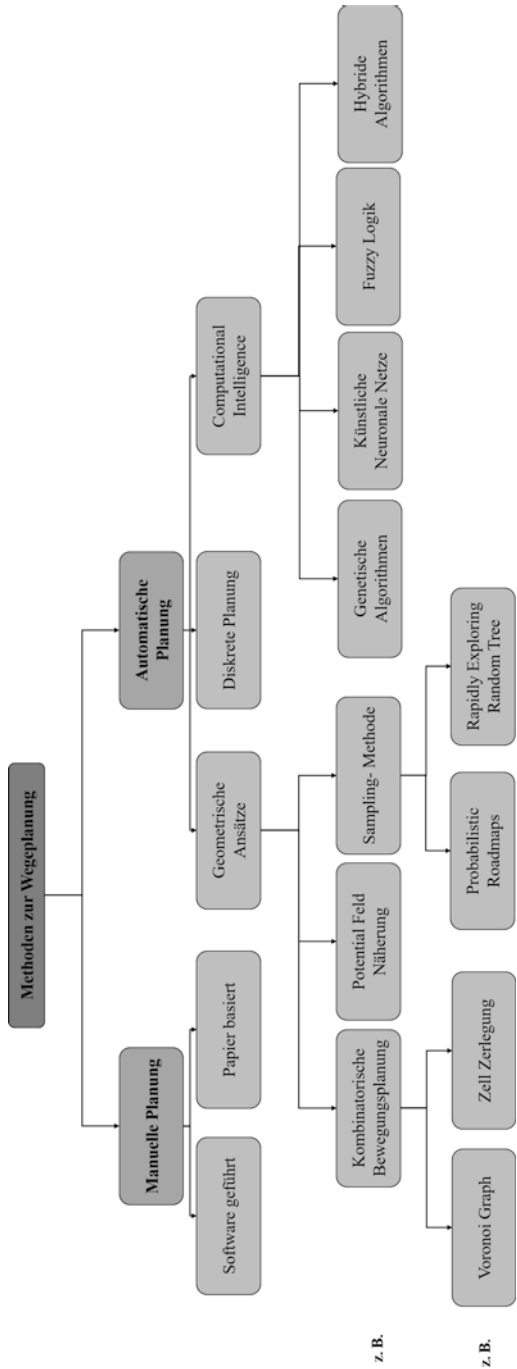


Abb. 3 Übersicht über verschiedene Methoden zur Wegenetzgenerierung für Fahrerlose Transportfahrzeuge (Uttendorf und Overmeyer 2018)

Die Verfahren zur offline Planung lassen sich grob in die zwei Kategorien unterteilen: *manuelle Planung* und *automatisierte Planung*. Die manuelle Planung ist im Bereich der Wegenetze für Fahrerlose Transportfahrzeuge noch weit verbreitet. Gerade in kleinen Produktionsumgebungen kann ein erfahrener Systemplaner von Fahrerlosen Transportsysteme durch seinen Erfahrungsschatz auch ohne aufwändige Berechnungen ein effizientes und anwendbares Wegenetz einzeichnen. Meist handelt es sich hierbei um zyklische Wegenetze, in denen nur einfache Kreisverkehre ohne viele Kreuzungen realisiert werden. Einige Hersteller¹ bieten zudem kleine Softwaretools an, die den Benutzer bei der Planung unterstützen. Diese gehen allerdings oft nicht über das manuelle Einzeichnen des Weges auf einer grafischen Benutzeroberfläche und dem anschliessenden Herunterladen des Wegenetzes in Form von Koordinaten hinaus (Demuth 2016).

Die automatisierte Planung lässt sich in die Bereiche *geometrische Ansätze*, *diskrete Planung* und *Ansätze des Cognitive Computing* unterteilen. In den letzten Jahren gab es einen Wandel von den geometrischen Methoden hin zu heuristischen Methoden (Mac et al. 2016). Dies liegt vor allem daran, dass geometrische Methoden häufig sehr zeitintensiv sind.

Die diskreten Wegenetzalgorithmen bilden die Grundlage für viele weitere Ansätze im Bereich der Wegeplanung. Sie finden meist Anwendung in Bereichen, in denen die Umwelt komplett bekannt ist und somit gut für die offline Planung geeignet sind. Normalerweise unterteilt ein diskreter Wegenetzalgorithmus die gegebene Umwelt zunächst in verschiedene Zellen. Der Algorithmus behält immer den Zustand der jeweiligen Zellen im Blick, diese können entweder *tot* (d. h. Zelle wurde besucht und ist nicht mehr relevant), *unbesucht* (d. h. Zelle muss noch besucht werden) oder *lebendig* (d. h. Zelle wurde besucht und ist auch weiterhin relevant) sein. Die zwei geläufigsten Algorithmen in dieser Klasse sind der A*-Algorithmus (Hart et al. 1968) und der Dijkstra-Algorithmus (Dijkstra 1959). Der A*-Algorithmus ordnet jeder Zelle einen Kostenwert zu. Anschliessend sucht der Algorithmus den Weg mit den geringsten Kosten von einer Startzelle zu einer Zielzelle. Anders als der A*-Algorithmus, arbeitet der Dijkstra Algorithmus graphenbasiert. Auf Basis einer Kostenfunktion berechnet er den kürzesten Pfad zwischen einem Startknoten und allen übrigen Knoten. Die jeweiligen Kosten entsprechen den Gewichten an den Kanten des Graphen.

Die geometrischen Ansätze lassen sich in die *kombinatorische Bewegungsplanung* (z. B. Voronoi Graph oder Zell Zerlegung (Klein 2005)), *Potential Feld-Näherungen* (Ge und Cui 2002) und *Sampling-Methode* unterteilen. Die geometrischen Ansätze versuchen das Problem der langen Berechnungszeiten zu reduzieren und das Zick-Zack-Verhalten von diskreten Algorithmen zu überwinden. Anstatt gleich grosse Zellen zu verwenden, wird bei der kombinatorischen Bewegungsplanung und den Sampling- Methoden versucht, eine intelligente Unterteilung für die

¹z. B. STILL iGO (www.still.de/fileadmin/stillwww/content/de/04_Intralogistik-Systeme/03_Automatisierung/Prospekte/iGo-Prospekt_2015_HHd_2015_web.pdf) und Götting Track Editor (<http://www.goetting.de/news/2015/track-editor>).

Umwelt zu finden. Anschliessend wird ein diskreter Algorithmus (bspw. A*) verwendet, um den tatsächlichen Weg zu planen.

Der Forschungsschwerpunkt der letzten Jahre lag im Bereich der Ansätze des Cognitive Computings. Diese können entweder alleine für sich verwendet werden oder zusammen mit einem Algorithmus aus den vorher genannten Methoden. Methoden des Cognitive Computings orientieren sich an menschlichen Lern- und Denkprozessen. Methoden aus dem Bereich des Cognitive Computings werden vor allem in Umwelten verwendet, in denen es wenig a-priori Wissen über ein Problem gibt. Sie eignen sich auch besonders gut zur Einbindung von aktuellen Sensorwerten von Fahrerlosen Transportfahrzeuge und damit auch zur online Wegeplanung. Wichtige Vertreter in dieser Gruppe sind *genetische Algorithmen*, *künstliche neuronale Netze*, *Fuzzy Logik* und *hybride Methoden*.

Genetische Algorithmen arbeiten ähnlich wie die Natur, die Gene im menschlichen Körper weiterentwickelt, mit den drei Operationen *Selektion*, *Kreuzung* und *Mutation* (Kruse et al. 2011). Zu Beginn wird eine Startpopulation gewählt. Die Startpopulation ist die erste Generation einer Population mit zufälligen Merkmalen. Diese erste zufällige Population wird dann sukzessive nach einer Fitnessfunktion bewertet wird. Die Fitnessfunktion stimmt mit der Zielfunktion eines Optimierungsproblems überein. Je nach Ergebnis der Fitnessfunktion wird die Population mutiert oder gekreuzt. Der Algorithmus stoppt, wenn ein gewisses Level von Fitness, also einer ausreichend guten Bewertung, erreicht wurde.

Künstliche neuronale Netze orientieren sich auch an einer Funktion im Körper, dem menschlichen Nervensystem. Die Idee ist, Informationen ähnlich wie die Infrastruktur des Gehirns zu verarbeiten (Heaton 2015). In einem ersten Schritt wird ein Netzwerk mit Neuronen aufgesetzt. Zwischen den Neuronen gibt es Verbindungen. Diese Verbindungen zwischen den Neuronen sind der wichtigste Schritt für die Effizienz des künstlichen neuronalen Netzwerks. Im nächsten Schritt findet eine Trainingsphase statt, in der neue Verbindungen hinzugefügt und unter anderem die Gewichte der Verbindungen geändert werden (Goodfellow et al. 2017). Die Komplexität der künstlichen neuronalen Netze kann durch das Hinzufügen von weiteren Neuronen-Schichten erhöht werden. Ein künstliches neuronales Netz besteht mindestens aus einer Input-Schicht und einer Output-Schicht (Buduma 2017). Künstliche neuronale Netze können Systeme mit hoher Komplexität beschreiben. Allerdings führt das Training dieses Netzes oft zu hoch dimensionalen und nonlinearen Optimierungsproblemen, dessen Lösung schwierig ist und viel Zeit in Anspruch nehmen kann. Auch in diesem Fall ist a-priori Wissen notwendig, um ein initiales Netz aufzusetzen. Wichtig ist auch die Auswahl einer geeigneten Datenbasis, da ansonsten Probleme wie einer Überanpassung (engl. *overfitting*) auftreten können.

Eine weitere Möglichkeit der Wegeplanung in der Kategorie der Cognitive Computing Methoden ist die Fuzzy Logik. Fuzzy Logik bietet eine Möglichkeit, menschliches Wissen in einem maschinenlesbaren Format zu beschreiben (Seising 2005). Lotfi Zadeh (1965) hat diese Methodik für Umwelten eingeführt, die sich nicht in der traditionellen binären Logik beschreiben lassen. Anstatt für eine Variable x nur die binären Werte $x \in \{0,1\}$ zuzulassen, kann die Variable auch graduell einem Attribut angehören $x \in [0,1]$. Neben diesem Vorteil bietet Fuzzy Logik zudem auch die

Möglichkeit mit linguistischen Variablen zu arbeiten. Dadurch kann menschliches Expertenwissen in linguistischer Form aufgenommen werden und in maschinenlesbarer Form verarbeitet werden.

Hybride Ansätze versuchen die Vorteile der einzelnen Ansätze zu kombinieren (Heaton 2014). Ein Beispiel für die online Planung wäre die Einbindung von verschiedenen Sensor-Inputs (Abstand von einem Hindernis, Distanz zum Ziel, Lenkwinkel etc.) und deren Verbindung in Form von Regeln in einem Fuzzy Logik Controller. Die Zugehörigkeitsfunktionen zu *kleinem Lenkwinkel* oder *geringer Distanz* können dann über künstliche neuronale Netze trainiert werden.

4 Fuzzy Logik für die Wegeplanung von Fahrerlose Transportsysteme

Gerade in Bereichen, in denen viel Expertenwissen vorliegt, ist es schwierig komplexe Probleme in mathematischen Funktionen auszudrücken. Allerdings kann menschliches Wissen sehr gut dazu genutzt werden, um Schlussfolgerungen zu ziehen. Dagegen ist es schwierig, dieses Wissen in einer präzisen mathematischen Form zu speichern. Mit der Definition von Zugehörigkeitsfunktionen ist es möglich festzuhalten zu welchem Grad eine Variable zu einer Klasse gehört. Systemexperten können leicht Aussagen treffen wie „*Wenn zu viel Auslastung auf einem Weg ist, dann sollte ein zweiter Weg geplant werden*“ oder „*Wenn Transporte sehr dringlich sind, dann sollten die Wege möglichst kurz sein*“. In der klassischen Informatik wäre diese Aussage schwer darstellbar. Fuzzy Logik hingegen kann mit vagen Begriffen wie *zu viel* und *zu wenig* arbeiten. Der Input *Auslastung* könnte also mit verschiedenen linguistischen Variablen belegt werden und je nachdem, wie die tatsächliche Auslastung aussieht, können mehrere Ausprägungen dieses Wertes, beispielsweise *wenig* und *mittel* gleichzeitig zu einem gewissen Grad zu treffen.

Das Herzstück jedes Fuzzy Logik Expertensystems ist die Regelbasis. In dieser werden die verschiedenen Regeln in der Form *WENN ... DANN* gespeichert (Mendel 2017). Es gibt verschiedene Möglichkeiten eine Regelbasis aufzustellen. Alonso und Madgalena (2011) unterscheidet die zwei Möglichkeiten, die Regelbasis entweder *durch menschliche Experten* oder *basierend auf einer Datenbasis* aufzustellen. Bei der Erstellung von Wegenetzen ist es schwierig, eine einheitliche und vergleichbare Datenbasis zu erhalten, da die verschiedenen Einsatz-Szenarien sehr voneinander abweichen. So unterscheiden sich, abhängig von der Größe der Layouts oder der Art des eingesetzten Fahrzeugtyps, Einschätzungen über Kriterien, wie die Auslastung des Wegenetzes oder die Länge von Wegen. Aus diesem Grund wurden in der Arbeit von Uttendorf und Overmeyer (2018) Experten zur Erstellung von Fuzzy Regeln befragt. Dazu wurden sieben Systemexperten verschiedene Wegenetze und Szenarien vorgestellt. Klir und Yuan (1995) haben eine Systematik erstellt, wie ein geeignetes Fuzzy-Inferenz-System aufgesetzt werden kann. Zuerst werden relevante Input und Output Variablen identifiziert und für diese geeigneten linguistischen Terme und Fuzzy Sets definiert. Danach werden passende Fuzzifizierungsmethoden für die Inputvariablen ausgewählt, die den Grad der Unsicher-

heit widerspiegeln. Dann werden geeignete Inferenzregeln durch die Experten aufgesetzt. Die zuvor definierten Input-Variablen mit den zugehörigen Fuzzy Regeln werden verbunden und geeignete Outputs werden definiert. Zuletzt wird eine geeignete Defuzzifizierungsmethode gewählt. Hierfür bietet sich unter anderem die Schwerpunkt-Methode an, bei der die Hüllkurve der Zugehörigkeitsfunktionen gebildet und anschliessend der Flächenschwerpunkt bestimmt wird. Der Flächenschwerpunkt stellt dann den defuzzifizierten Zahlenwert dar (Klir und Yuan 1995).

Im Folgenden werden einige Beispiele für Fuzzy Regeln zur Wegeplanung von Fahrerlose Transportfahrzeuge vorgestellt, die zusammen mit Systemexperten erstellt wurden.

Abb. 4 zeigt stark generalisiert ein Beispiel für eine Fuzzy Regel, die die *Generierung eines zusätzlichen Weges* darstellt. Eine Regel zu diesem Szenario könnte beispielsweise lauten: „WENN hohe Auslastung UND NICHT wenig Freifläche DANN hoher Bedarf an Alternativweg“.

Im linken Bild ist zu sehen, dass verschiedene Fahrwege zusammengefasst wurden um eine Station zu beliefern. Dies hat allerdings zur Folge, dass der gemeinsam genutzte Weg sehr stark ausgelastet ist. Um die Station herum ist allerdings noch viel Freifläche vorhanden. Die Experten würden in einem solchen Fall entscheiden, dass bei hoher Auslastung des gemeinsamen Weges ein Alternativweg generiert werden soll. Die Fuzzy Logik Regel würde in diesem Fall das Erstellen eines Alternativweges empfehlen.

Ein weiteres Beispiel für eine Regel ist in Abb. 5 zu sehen. Hier geht es um die Entscheidung über das Kreuzen von Fahrwegen. Diese Abbildung zeigt generalisiert

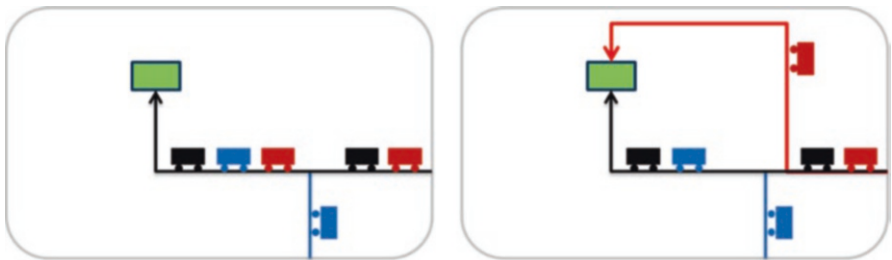


Abb. 4 Visualisierung der Regel *Generierung eines zusätzlichen Weges* (Uttendorf und Overmeyer 2018)

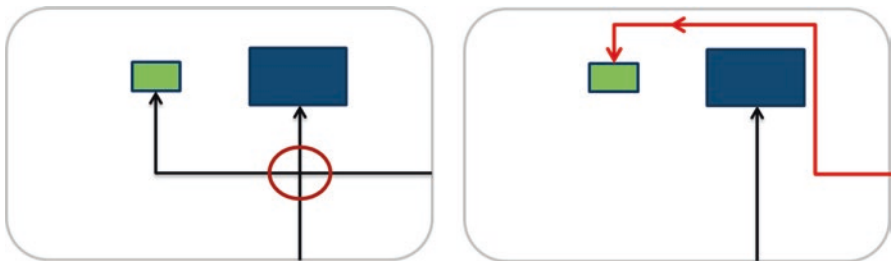


Abb. 5 Visualisierung der Regel *Reduzierung von Kreuzungen* (Uttendorf und Overmeyer 2018)

ein Szenario, in dem zwei Stationen angefahren werden. Im linken Bild schneiden sich die beiden Wege noch vor der Anfahrt in die Stationen. Im rechten Bild wurde ein Ausweichweg um eine der Station herumgeplant. Eine dazugehörige Regel lautet zum Beispiel: „*WENN die Fahrzeuggeschwindigkeit niedrig ist UND KEINE geringe Auslastung besteht DANN wenige Kreuzungen*“.

Im Falle von langsamen Fahrzeugen ist die Kreuzung länger belegt als bei schnellen Fahrzeugen. Wenn nun zusätzlich noch eine hohe Auslastung auf den Wegen herrscht, ist die Gefahr von gegenseitigen Behinderungen (d. h. Deadlocks) rund um den Bereich der Kreuzungen noch höher.

Die Experten können zu jedem der für sie relevanten Entscheidungskriterien für ein effizientes Wegenetz Regeln aufstellen. Die verschiedenen Regeln zu einer Entscheidung werden zu Regelsets zusammengefasst. Tab. 1 zeigt ein Fuzzy Regel Set für die vorherig beschriebene Regel zur Minimierung von Kreuzungen.

In der zweiten und dritten Spalte ist jeweils die Ausprägung der beiden Inputs *Geschwindigkeit von Fahrerlosen Transportfahrzeuge* und *Auslastung* zu sehen. In der letzten Spalte wird der Fuzzy Output β präsentiert. Dieser Output kann dazu verwendet werden, ihn mit anderen Algorithmen zu verknüpfen und auf diese Weise menschliches Systemplanerwissen mit mathematischer Optimierung zu kombinieren. In Uttendorf und Overmeyer (2018) wurde der Output der Fuzzy Regeln dazu genutzt, um die Kostenmatrix des A*-Algorithmus zu beeinflussen.

Abb. 6 zeigt, wie die Outputs der Fuzzy Regeln die Kostenmatrix des A*-Algorithmus beeinflussen können. Wie im Abschn. 3 beschrieben, ist der A*-Algorithmus ein diskreter Wegeplanungsalgorithmus, der zu Beginn das zugehörige Layout in ein Gitter mit gleich grossen Zellen unterteilt. Anschliessend ordnet er jeder Zelle einen Kostenwert aufgrund einer Heuristik zu. In der Abb. 6 wird nun gezeigt, dass nach einer Fuzzy Logik Evaluation, die Outputs der Regeln in Form von α , β , γ ,... auf die ursprüngliche Kostenmatrix addiert werden können. Durch dieses Verfahren kann das Expertenwissen automatisiert die Wegeplanung des A*-Algorithmus beeinflussen.

Tab. 1 Fuzzy Regel Set – Minimierung der Anzahl von Kreuzungen

Regel Set 2	Geschwindigkeit von Fahrerlosen Transportfahrzeuge	Auslastung	β
a	sehr niedrig	niedrig	niedrig
b	niedrig	niedrig	niedrig
c	hoch	nicht hoch	niedrig
d	hoch	hoch	mittel
e	mittel	mittel	nicht hoch
f	niedrig	hoch	hoch
g	niedrig	mittel	mittel
h	mittel	hoch	mittel

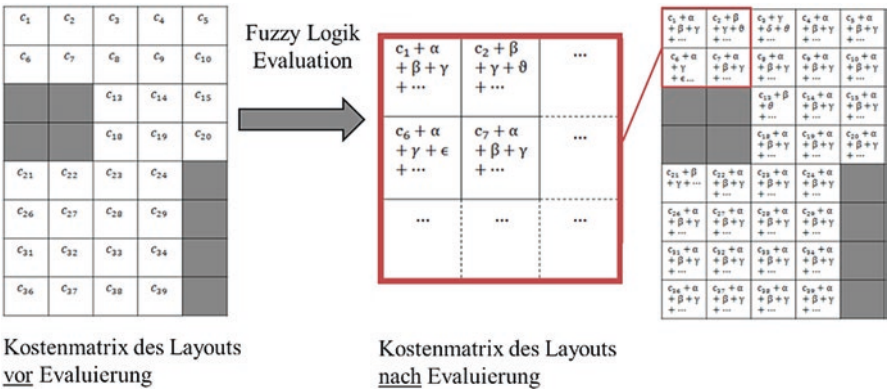


Abb. 6 Kombination der Fuzzy Evaluation mit dem A*-Algorithmus (Uttendorf und Overmeyer 2018)

5 Herausforderungen beim Einsatz von Cognitive Computing für Fahrerlose Transportfahrzeuge

Die Methoden des Cognitive Computings eignen sich gut dazu, erste Algorithmen aufzusetzen, auch wenn kein grosses a-priori Wissen über ein Szenario zu Beginn zur Verfügung steht. Die Methoden sind oft in der Lage, bei einer aktuelleren Datenlage selbst Anpassungen durchzuführen und sich stetig zu verbessern. Allerdings birgt dieses Vorgehen gerade in der Extrahierung von Regeln aus Expertenwissen die Gefahr, dass die generierten Regeln stark subjektiv sind und sich nicht auf jedes neue Szenario anwenden lassen können. Deshalb muss sichergestellt werden, dass eine geeignete Trainingsphase stattfindet, um die Regeln so zu beeinflussen, dass sie auf neue Szenarien angewendet werden können. Bei der Erstellung der Methode muss beachtet werden, dass gerade bei einer anfänglich kleinen Datenbasis die Gefahr einer statistisch nicht signifikanten oder stark subjektiven Wissensbasis vorliegt. Im Bereich der Fuzzy Logik können bei durch Expertenwissen aufgestellten Regeln und Fuzzifizierungs- und Defuzzifizierungsmethoden die Gewichte und Regeln durch den Experten stark beeinflusst sein. Im realen Betrieb kann sich dann herausstellen, dass die Regeln auf einen neuen Anwendungsfall nicht in gleicher Weise anwendbar sind. Erst nach Anwendung auf mehrere Situationen und adaptiver Anpassung der Gewichte kann sichergestellt werden, dass der entwickelte Algorithmus effizient genutzt werden kann.

Zudem sollte vor dem Einsatz und Lernen der Algorithmen getestet werden, ob sich ein Einsatz von Methoden des Cognitive Computings lohnt. Oft besteht ein Nachteil dieser Methoden darin, dass sie längere Rechenzeiten unter anderem gegenüber diskreten Planungsalgorithmen haben. Sollte bei einer zu grossen Änderung der Trainingsdaten zum Beispiel erneut ein künstliches neuronales Netz trainiert werden, beansprucht dies einiges an Rechenzeit und eignet sich nur bedingt für eine online Planung während des Betriebs des Fahrzeugs.

Auch erscheinen einige der Verfahren des Cognitive Computings dem Benutzer als *Blackbox*, das heisst für den Benutzer ist unter Umständen nicht ohne vertieftes mathematisches Verständnis klar, wie beispielsweise die Verbindungen und Gewichte in künstlichen neuronalen Netzen aufgestellt wurden. Das verringert unter Umständen das Vertrauen in die Methode. Dieses Problem kann durch regelmässige Abfragen des Benutzers gelöst werden oder durch manuelle Bestätigung von Gewichtsänderungen oder Schwellwerten durch den Benutzer.

6 Vorteile

Ansätze des Cognitive Computing im Bereich der Fahrerlose Transportfahrzeuge zeichnen sich dadurch aus, dass sie in einem hohen Mass *adaptiv* sind. Das heisst, sie sind in der Lage ein gewisses Mass an Ungewissheit zu tolerieren. Es müssen nicht alle Informationen zu Beginn vorliegen, um eine erste lauffähige Version zu erhalten. Im Beispiel des im Abschn. 4 aufgezeigten Fuzzy-Inferenz-Systems reicht es aus, wenn Experten erste Regeln für den Algorithmus in linguistischer Form definieren. Die Regeln können dann nach mehreren Testdurchläufen automatisiert angepasst oder auch die Gewichtung der verschiedenen Optimierungsziele aktualisiert werden. Cognitive Computing Ansätze sind daher in der Lage, sich auch an neue Informationen automatisiert zu adaptieren.

Weiterhin bieten Ansätze des Cognitive Computing die Möglichkeit, Algorithmen *interaktiv* zu gestalten. So gibt es die Möglichkeit, zum Beispiel durch Abfragen, direkt mit dem Benutzer zu interagieren. Im Beispiel der Fuzzy Logik ist diese Abfrage sogar in linguistischer Form möglich, was die Verständlichkeit des Systems noch weiter erhöht.

Zudem bieten Ansätze des Cognitive Computings die Möglichkeit *iterativ* abzu-
laufen. Das bedeutet, dass der verwendete Ansatz durch stetiges Anwenden immer neu trainiert werden kann und von einer grossen Anzahl von historischen Werten lernen kann. Hierdurch kann das menschliche Urteilsvermögen immer schrittweise weiter erhöht und, im Falle der Fuzzy Regeln für Wegenetze, neue Regeln stetig hinzugefügt werden. Ein automatisiertes Anlernen von Fuzzy Regeln wäre auch möglich.

Die *kontextuelle* Funktionsweise der Methoden ermöglicht es, ein vertieftes Verständnis von Verhaltensmustern zu erlangen. So kann beispielsweise durch das häufige Generieren eines Alternativwegs an einer bestimmten Stelle, der Algorithmus automatisch angepasst werden, sodass der Weg an dieser Stelle permanent geändert wird, um die Zeit für eine lokale Umplanung zu reduzieren beziehungsweise die lokale Umplanung nicht für jeden einzelnen Roboter realisieren zu müssen.

Zudem bieten die Methoden die Möglichkeit menschliches Wissen standardisiert und wiederverwertbar zu speichern. Das automatisiert generierte oder abgefragte Wissen liegt in maschinenlesbarer Form dar und ist nicht mehr abhängig von der Verfügbarkeit eines Systemexperten.

7 Ausblick

Zusammenfassend kann gesagt werden, dass Ansätze des Cognitive Computing in der Lage sind, in kürzester Zeit Sachverhalte zu erkennen und darauf zu reagieren. Methoden des Cognitive Computing bieten im Bereich der Fahrerlosen Transportsysteme eine gute Möglichkeit, ein selbstlernendes Verhalten zu implementieren und auf diese Weise auch in Zukunft anwendbar zu sein. Gerade im Bereich der Fahrerlosen Transportsysteme befindet sich die Wirtschaft in einer schnell wandelnden Umgebung, in der eine manuelle Anpassung der Algorithmen sehr zeitintensiv wäre.

Da der grossflächige Einsatz der Fahrzeuge in einer hochautomatisierten Umgebung noch eine relativ neue Disziplin ist, besteht Expertenwissen für kleinere Spezialfälle. Gerade der Aspekt des Sammelns von neuen Informationen und das kontinuierliche, iterative Lernen der Algorithmen über die Zeit ermöglicht eine stetige Verbesserung der Methoden. Besonders der Ansatz der *kollektiven Intelligenz*, also der Einbindung von menschlichem Wissen in die Wegeplanung, ist hier vielversprechend.

Um Cognitive Computing Ansätze gezielt einzusetzen, ist es wichtig Experten mit in den Prozess der Gestaltung einer Methode einzubeziehen, um zum einen von bestehendem Wissen zu profitieren, aber auch zum anderen zu gewährleisten, dass die ausgewählten Methoden für die zuvor definierten Ziele richtig sind.

Die weiteren Entwicklungen von Fahrerlosen Transportfahrzeuge bewegen sich immer weiter weg von spurgeführten Systemen hin zu einer mobilen Robotik, in der die Wege nicht starr vorgegeben sind, sondern die Fahrzeuge komplett frei navigieren. Gerade in diesen sehr wandelnden Umgebungen ist eine situative Entscheidung bei zu ändernde Wege sehr wichtig. Ansätze des Cognitive Computings können gut mit dieser Ungewissheit umgehen und haben daher eine grosse Relevanz in der Zukunft in diesem Bereich.

Literatur

- Alonso JM, Magdalena L (2011) Special issue in interpretable fuzzy systems. *Information Sciences*, Elsevier, 181(20):4331–4339.
- Buduma N (2017) *Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms*. O'Reilly.
- Demuth R (2016) Götting Track Editor jetzt online verfügbar. <http://www.goetting.de/news/2015/track-editor>. Letzter Zugriff 13.07.2016.
- Dijkstra EW (1959) A note on two problems in connexion with graphs. In: *Numerische Mathematik*, Springer-Verlag, 1(1):269–271.
- Ge S, Cui Y (2002) *Autonomous Robots*, 13:207.
- Goodfellow I, Bengio Y, Courville A (2017) *Deep Learning (Adaptive Computation and Machine Learning)*. The MIT Press.
- Hart P, Nilsson N, Raphael B (1968) A formal basis for the heuristic determination of minimum cost paths. *IEEE Transaction on System Science and Cybernetics*, IEEE Press, 4(2):100–107.
- Heaton J (2014) *Artificial Intelligence for Humans, Volume 2: Nature-Inspired Algorithms*. CreateSpace Independent Publishing Platform.

- Heaton J (2015) Artificial Intelligence for Humans, Volume 3: Deep Learning and Neural Networks. CreateSpace Independent Publishing Platform.
- Klein R (2005) Algorithmische Geometrie- Grundlagen, Methoden, Anwendungen, Springer Verlag Berlin.
- Klir GJ, Yuan B (1995) Fuzzy Sets and Fuzzy Logic. Theory and Applications. Prentice Hall Verlag, Upper Saddle River.
- Kruse R, Borgelt C, Braune C, Klawonn F, Moewes C, Steinbrecher M (2011) Computational Intelligence. Eine methodische Einführung in Künstliche Neuronale Netze, Evolutionäre Algorithmen, Fuzzy-Systeme und Bayes-Netze. Vieweg+Teubner Verlag, Wiesbaden.
- Mac T, Copot C, Tran D, Keyser R (2016) Heuristic approaches in robot path planning: A survey. Robotics and Autonomous Systems, 86:13–28.
- Mendel J (2017) Uncertain Rule-Based Fuzzy Systems: Introduction and New Directions, 2nd Edition, Springer.
- Nüchter A (2009) 3D Robotic Mapping- The Simultaneous Localization and Mapping Problem with Six Degrees of Freedom. Springer.
- Raja P, Pugazhenth S (2012) Optimal path planning of mobile robots. A review. In: International Journal of Physical Sciences, Academic Journals, 7(9):314–1320.
- Seising R. (2005) Die Fuzzifizierung der Systeme. Die Entstehung der Fuzzy Set Theorie und ihrer ersten Anwendungen – Ihre Entwicklung bis in die 70er Jahre des 20. Jahrhunderts. Stuttgart: Franz Steiner Verlag (Boethius: Texte und Abhandlungen zur Geschichte der Mathematik und der Naturwissenschaften Band 54).
- Turban E, Aronson JE (1998) Decision Support Systems and Intelligent Systems. 5. Aufl., Pearson Education (US), New York.
- Ullrich G (2014) Fahrerlose Transportsysteme. Eine Fibel – mit Praxisanwendungen – zur Technik – für die Planung, 2. Aufl., Springer Vieweg Verlag, Wiesbaden.
- Uttendorf S, Overmeyer L (2018) Automated Generation of Roadmaps for Automated Guided Vehicle Systems, Berichte aus dem IPH, 3/2018, Garbsen: TEWISS Verlag.
- VDI 2510 (2005) Automated Guided Vehicle Systems (AGVS). Beuth Verlag, Berlin.
- Zadeh LA (1965) Fuzzy sets. Information and Control, Elsevier, 8(3):338–353.



Legal Tech – Cognitive Computing in der juristischen Praxis

Philip Hanke

Zusammenfassung

Die Rechtswissenschaft und juristische Praxis zeichnen sich durch das Beherrschen der natürlichen Sprache aus. Sachverhalte und Argumente müssen möglichst präzise beschrieben werden. Das Recht, bestehend aus Gesetzestexten und Gerichtsurteilen, stellt einen grossen, weitgehend unstrukturierten Datensatz dar. Der Artikel zeichnet nach, wie es dennoch möglich ist, Recht maschinenlesbar zu machen beziehungsweise wie Computersysteme das Recht interpretieren und JuristInnen unterstützen können. Methoden, die im Sinne des Cognitive Computings natürliche Sprache ohne viel menschliches Zutun auswerten, stecken noch in den Kinderschuhen, weisen aber Potenzial auf. Mittels Analyse von Metadaten und gezielten Suchen ist es möglich, Zusammenhänge zwischen Rechtsquellen herzustellen oder Gerichtsurteile zu einem gewissen Grad vorherzusagen. Das Kapitel zeigt dabei die eingesetzten datengetriebenen oder regelorientierten Ansätze auf und setzt sich mit den Möglichkeiten, Grenzen und Risiken dieser Methoden auseinander.

1 Einleitung

Das Handwerk der JuristInnen ist die natürliche Sprache. Klauseln in Verträgen oder Allgemeinen Geschäftsbedingungen müssen wasserdicht formuliert werden. Vor Gericht muss im Interesse der Mandantin oder des Mandanten argumentiert

Vollständig neuer Beitrag

P. Hanke (✉)
Weblaw AG, Zürich, Schweiz
E-Mail: philip@hanke.com

werden. Der vom Gesetzgeber intendierte Wille muss in Gesetzen und Verordnungen geschrieben werden. All dies sind tendenziell Tätigkeiten, die – zumindest bis jetzt – Menschen deutlich besser können als Maschinen.

Während sich Rechtsinformatik und Computerlinguistik in der wissenschaftlichen Forschung entwickelt haben, versteht man unter *Legal Tech* die Verwendung von Software in der juristischen Praxis. Der Einsatz von Cognitive Computing kann hierbei zielführend sein. Denn dieses bedient sich des Semantic Computings, um die Bedeutung natürlicher Sprache zu interpretieren. Dafür wird unter anderem Natural Language Processing angewendet, um den Input aufzubereiten und somit maschinenlesbar zu machen. Aufbauend auf der semantischen Verarbeitung stellt Cognitive Computing kontextbasierte Verknüpfungen her und versucht, anhand der bestehenden Verlinkungen Schlussfolgerungen zu ziehen (D’Onofrio und Portmann 2017; D’Onofrio et al. 2018) – dies ist in einem juristischen Kontext ganz besonders wichtig.

Aufgrund der Produktivitätssteigerungen und Konkurrenz durch gänzlich neue Geschäftsmodelle ist die Rechtsbranche gezwungen, sich mit dem rasanten Technologiefortschritt auseinanderzusetzen. Sie hat dabei in den letzten Jahren ähnliche Entwicklungen wie andere Industrien mitgemacht: die eingesetzten Computersysteme sind gewissermaßen *smarter* geworden. Während IT-Tools bislang vor allem als Dokumentenablage eingesetzt wurden (also etwa zum Abspeichern von gescannten Dokumenten) oder als Datenbanken, so haben moderne Machine-Learning-Verfahren langsam aber doch in der juristischen Praxis Einzug erhalten. Dennoch sind die primären Tools der JuristInnen von heute immer noch in erster Linie Microsoft Word, eine E-Mail-Applikation, ein Case-Management-Tool,¹ eventuell Onlinedatenbanken (von Gerichtsurteilen und Gesetzestexten) und der elektronische Zugang zu Fachzeitschriften. Bücher und gedruckte Dokumente, wie zum Beispiel (unterschiedene) Formulare, die danach wieder gescannt und als E-Mail verschickt werden (quasi als Ersatz zum Fax), haben immer noch einen grossen Stellenwert. Anwaltskanzleien, Behörden, Gerichte, juristische Verlage und andere in der Rechtspraxis tätige AkteurInnen sind daher in einem Prozess der Digitalisierung, welcher zum Einsatz von neuen Verfahren führen kann.

Die Rechtsinformatik unterscheidet im Wesentlichen zwischen Ansätzen, die an Regeln orientiert sind (engl. *rules-driven*) und jenen, die an Daten orientiert sind (engl. *data-driven*). Ersterer übersetzt rechtliche Regeln in Code, während der zweite Ansatz Daten mithilfe statistischer Methoden analysiert. In der Praxis verschmelzen diese beiden Ansätze: Systeme verwenden etwa maschinelles Lernen (engl. *machine learning*) (das Erkennen von Mustern in Daten), um Nutzereingaben zu interpretieren, verwenden aber Regeln, um Auskünfte zu geben (bspw. im Szenario eines Chatbots).

Der folgende Artikel gibt einen Überblick, in welche Richtungen sich Legal Tech entwickelt und an welchen Stellen Cognitive Computing von Bedeutung ist. Zuerst

¹ Im Folgenden wird nicht weiter auf diese Tools eingegangen, da sie tendenziell eher der Verwaltung, Arbeitszeiterfassung oder dem Projektmanagement zuzurechnen sind und nicht der computergestützten Rechtsberatung.

werden die Entwicklungen im Bereich der Dokumentenautomatisierung präsentiert. Anschliessend wird diskutiert, ob und wie rechtliche Regelungen in Code übersetzt werden können. Nach diesen Regel-basierten Ansätzen, wird gezeigt, wie, mit geeigneten Daten, *Natural Language Processing* und *Deep Learning* eingesetzt werden (können). In Kenntnis der Regel- und Daten-orientierten Ansätzen kann dann gefragt werden, ob eine automatisierte Rechtsberatung (also im Sinne eines *Robo-Lawyers*) machbar ist. Ein nächster Abschnitt zeigt die Entwicklungen im Bereich der *Smart Contracts* auf der Blockchain auf. Schliesslich werden noch ein paar Schattenseiten von Legal Tech aufgezeigt und ein Ausblick gegeben.

2 Dokumentenautomatisierung

Viele Tätigkeiten in der juristischen Praxis sind von einer gewissen Eintönigkeit geprägt. Dazu gehört zum Beispiel das Erstellen von Dokumenten. Mehrere Tools können hier Abhilfe verschaffen. In der Schweiz bietet etwa [digitalCounsels.com](https://digitalcounsels.com) ein Onlinetool für EndkundInnen, das nach Eingabe der verlangten Daten beziehungsweise Auswahl der gewünschten Optionen automatisch eine Datenschutzerklärung für die eigene Website, eine Patientenverfügung, ein Schreiben an die Hausverwaltung zur Mietzinsreduktion, einen Arbeitsvertrag, eine Vertraulichkeitsvereinbarung oder die notwendigen Dokumente zur Gründung einer Aktiengesellschaft oder Gesellschaft mit beschränkter Haftung erstellt. Dies ist ein praktisches Tool, um rechtlich korrekte Dokumente ohne Hinzuziehen einer Anwältin oder eines Anwalts zu erstellen. Allerdings geht die Kundin oder der Kunde ein gewisses Risiko ein: Ganz ohne anwaltliche Beratung können Fehler passieren, etwa, dass für einen bestimmten Fall notwendige Elemente fehlen oder auf gewisse Dokumente gänzlich vergessen wird. Ähnliche Systeme existieren auch in anderen Ländern.² Jedoch wird meistens das Angebot noch verbunden mit der Möglichkeit, direkt über die Website Kontakt mit einer Anwältin oder einem Anwalt aufzunehmen.

Andere Tools setzen daher eher auf Ebene der JuristInnen oder Unternehmensabteilungen an. *DocEngine*,³ entwickelt von der Weblaw AG,⁴ ist ein ähnliches Dokumentenautomatisierungstool, das sich jedoch primär an Kanzleien und Unternehmen (z. B. sogenannte In-House-Counsels, also JuristInnen in Rechtsabteilungen) richtet und somit darauf abzielt, juristische Tätigkeiten beziehungsweise deren Prozesse produktiver zu machen.

Die erwähnten Dokumentenautomatisierungstools führen jedoch noch keine oder nur kaum Interpretation aus. Das Computersystem stellt die eingegeben Daten nicht oder nur kaum in einen Kontext. Die Erstellung von Dokumenten wird immer noch vom Menschen, der dem Computer sagt, wie er die Lücken in diesen füllen soll, vollzogen. Diese Vorgehensweise hat jedoch den Vorteil, dass solche Lösungen

²<https://www.rocketlawyer.com> oder <https://www.legalzoom.com/country/ch> in den USA.

³<https://www.docengine.ch>.

⁴<https://www.weblaw.ch>.

rasch und ohne viel Forschung und Entwicklung entwickelt werden können. Dies hat sich beispielsweise während der Einführung der Datenschutzgrundverordnung⁵ (DSGVO bzw. GDPR auf Englisch) als vorteilhaft erwiesen. Lösungen, um die Dokumentationspflichten im neuen europäischen Datenschutzrecht zu erfüllen, konnten mit relativ kleinem Aufwand schnell eingesetzt werden.⁶ Anders als die Tools, welche mit statistischen Modellen arbeiten, unterliegt in der Dokumentenautomatisierung der Output keinen Wahrscheinlichkeitsrechnungen. Sie ist somit weniger fehleranfällig und erfüllt die hohen Anforderungen an Genauigkeit, welche die juristische Arbeit prägen.

3 Aus Recht wird Code

Für Regel-basierte Applikationen muss zuerst gefragt werden, inwieweit rechtliche Regeln als Code abgebildet werden können. Wenn dies gelingt, so können nützliche Tools entwickelt werden, die etwa Rechtsauskünfte geben. Allerdings gibt es grosse Schwierigkeiten, das gesamte Recht, welches aus Sicht der Informatik abgesehen von wenigen Metadaten ein im Wesentlichen unstrukturierter Datensatz ist, computer-gerecht umzuformulieren oder automatisiert zu interpretieren. Hinzu kommt noch, dass das Recht nicht nur ein Datensatz ist, sondern eigentlich auf mehrere Quellen verstreut ist.

Meistens ohne es zu wissen, arbeiten JuristInnen und InformatikerInnen mit ähnlicher Logik. Vieles in der juristischen Praxis kann als *Wenn-Dann-Statement* ausgedrückt werden. Wenn beispielsweise jemand eine andere Person absichtlich tötet, dann ist die Straftat des Mordes erfüllt (vereinfacht gesagt). Absicht ist ein auf die inneren Gegebenheiten abzielender, subjektiver Tatbestand (Was hat der/die Angeklagte während der Tat gedacht?) und wird auf Lateinisch *mens rea* (*schuldiger Geist*) genannt. Töten hingegen ist ein objektives Kriterium (Was ist passiert?) und wird auf Lateinisch *actus reus* (*schuldige Handlung*) genannt.⁷ Mord verlangt beide Elemente. Zusätzlich müssen die rechtfertigenden oder entschuldigenden Strafausschlussgründe (bzw. *defences* im anglo-amerikanischen Rechtssystem), wie zum Beispiel Notwehr oder Schuldunfähigkeit, fehlen.

Diese Regel könnte folgendermassen in Code geschrieben werden:

```
if (actus_reus == 1) & (mens_rea == 1) & (defences == 0):
    murder = 1
# Tötung ohne Absicht könnte Totschlag sein:
```

⁵Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG.

⁶Siehe bspw. <https://zoa-gdpr.com>, eine Softwarelösung, die den gesamten Datenmanagementprozess in einer Applikation bündelt.

⁷Je nach Land ist das Vokabular unterschiedlich. Die Grundidee ist aber in allen modernen Rechtssystemen vorhanden.

```
elif (actus_reus == 1) & (mens_rea == 0) & (defences == 0):  
    check_manslaughter()  
# usw.
```

Eine grundsätzliche Frage, die sich dann stellt, ist: Könnte man das gesamte Recht in Codeform darstellen? Und falls ja: warum wird es nicht gemacht? Der entscheidende Faktor ist, ob die Regel eine sogenannte strenge Haftungsregel ist oder ob sie doch etwas mehr an Interpretation verlangt.

Der Strassenverkehr zeichnet sich durch strenge Haftungsregel aus. Wer bei Rot über die Kreuzung fährt, zahlt eine Strafe. Hierbei ist es unerheblich, ob eine Absicht vorlag oder es irrtümlich passiert ist. Abweichungen von der Regel würden nur in Ausnahmefällen akzeptiert (bspw. wenn nachweislich ein technischer Defekt vorlag). Solche Regeln können leicht in Code umgesetzt werden.

Es wäre daher vorstellbar, das mit einer weiteren Entwicklung des *Internet-of-Things* (deut. Internet der Dinge) (vgl. z. B. (D’Onofrio et al. 2018a)) das Strassenverkehrsrecht komplett automatisiert durchgesetzt werden könnte. In dieser Vision (und es bleibt dem Leser überlassen, ob er das als Utopie oder Dystopie erachtet) könnten vernetzte Fahrzeuge (sofern sie noch nicht komplett autonom fahren und daher von Haus aus sämtliche Regeln beachten) automatisch Verkehrsübertretungen an die Behörden melden und gegebenenfalls von selbst die anfallende Busse mit der Kreditkarte des Fahrers begleichen. Selbstfahrende Autos müssen die Verkehrsregeln sowieso kennen und befolgen können. *Computational Law* funktioniert also in beide Richtungen: Es kann einerseits *ex post* feststellen, ob Regeln eingehalten wurden, oder aber auch Angelegenheiten so einrichten, dass die Regeln eingehalten werden (Genesereth 2015).

In anderen Rechtsgebieten sieht die Sache ganz anders aus. Artikel 1 des deutschen Grundgesetzes sagt etwa: „*Die Würde des Menschen ist unantastbar. Sie zu achten und zu schützen ist Verpflichtung aller staatlichen Gewalt*“.⁸ Die grundsätzliche Logik wäre zwar als Code darstellbar, der Nutzen aber sehr eingeschränkt. Was ist die Würde des Menschen? Und wann ist sie angetastet? Wann hat der Staat sie nicht ausreichend geachtet und geschützt? Auch das oben genannte Programm für Tötungsdelikte funktioniert nur, wenn die Parameter (*actus_reus*, *mens_rea*, *defences*) eingegeben werden. In anderen Worten: es bildet die rechtliche Logik ab, ist aber auf die Eingabe des Zustandes der Welt angewiesen.

Wie eingangs erwähnt, unterscheidet sich der Schwierigkeitsgrad der Übersetzung von Recht in Code erheblich je nach Rechtsgebiet. Dort, wo es einfacher ist, gibt es nun Unternehmen und Behörden, die daran arbeiten. Erwähnenswert ist das Startup *APIAX*,⁹ welches es sich zum Ziel gesetzt hat, finanzmarktrechtliche Regeln im Bankwesen als Programmierschnittstelle (engl. *application programming interface* bzw. *API*) zur Verfügung zu stellen. Aber auch Finanzaufsichtsbehörden gehen den Weg, die Fülle an Gesetzen, Verordnungen und Rundschreiben als gebündeltes Handbuch zusammenzustellen. Dies ist zwar noch kein Code, der von Computern interpretiert werden kann, aber es ist ein wesentlicher Schritt in die Richtung Digitalisierung.

⁸ Grundgesetz für die Bundesrepublik Deutschland, 23. Mai 1949.

⁹ <https://www.apiax.com>.

Zumindest wurde in den letzten Jahren daran gearbeitet, Rechtsquellen besser zu verknüpfen, also dass zum Beispiel eine offizielle Seite, die eine Verordnung anzeigt, auch einen Link zu dem Gesetz, auf Basis dessen diese Verordnung erlassen wurde, enthält. Nützlich wäre, wenn leicht nachvollziehbar wäre, welche Änderungen an einem Gesetz vorgenommen wurden (also eine Art Versionskontrolle). Moderne Gesetzes- und Rechtssprechungssammlung ähneln, von ihrer Struktur her, Seiten wie Wikipedia,¹⁰ die sich ebenfalls durch einen hohen Grad an Verknüpfungen auszeichnet. Diese Verknüpfungen sind mit relativ geringem Aufwand erstellbar und erleichtern wesentlich die Rechtsrecherche. Ein modernerer Ansatz könnte das Recht und die Verbindungen zwischen Rechtsquellen dann noch anschaulich graphisch darstellen.¹¹

Zum jetzigen Zeitpunkt ist die primäre Aufgabe also, die Substanz des Rechts strukturiert, beispielsweise in Form eines Entscheidungsbaumes abzubilden (siehe Abb. 1). Diese Struktur könnte dann, verknüpft mit einer Schnittstelle (engl. *Interface*), wie einem Chatbot, als Basis für Rechtsauskünfte dienen (siehe weiter unten zu existierenden Chatbot-Projekten).

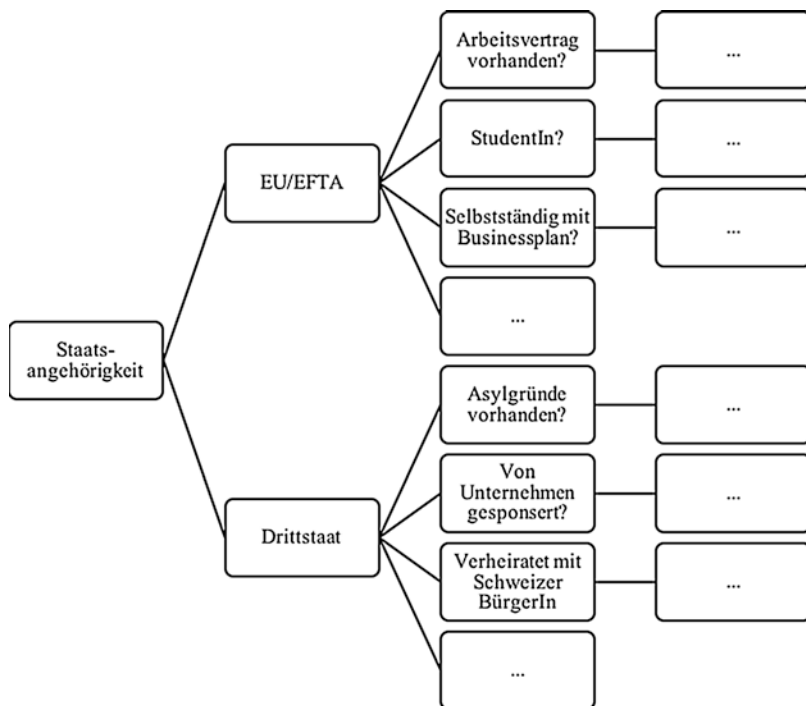


Abb. 1 Beispiel einer Rechtsauskunft im Bereich des Schweizer Migrations- und Asylrechts

¹⁰ <https://www.wikipedia.org>.

¹¹ Die deutsche Webseite <https://www.juralib.de> bietet die graphische Darstellung des Rechts als Lernhilfe für Studierende der Rechtswissenschaft an.

4 Datenanalyse und Natural Language Processing – Vom Einzelfall zur Statistik

Die folgende Situation zeigt, wie Datenanalysen die juristische Praxis verändern kann: Eine ägyptische Asylwerberin, deren Gesuch auf Asyl vom Staatssekretariat für Migration soeben abgelehnt wurde, fragt ihre Anwältin, wie sie ihre Chancen einschätzt, dass das Bundesverwaltungsgericht einer Beschwerde gegen die Entscheidung des Staatssekretariats für Migration stattgibt. Gegenwärtig kann die Anwältin nur eine ungefähre Einschätzung auf Basis der Umstände und ihres Bauchgefühls geben. Mit einer systematischen Analyse vergangener Bundesverwaltungsgerichtsentscheide könnte jedoch eine viel genauere Vorhersage getroffen werden. Doch der Weg dorthin ist steinig. Die vergangenen Entscheide sind derzeit nicht auf Knopfdruck herunterladbar. Es braucht einen *Webcrawler* (d. h. web-basiertes automatisches Suchprogramm): Mittels eines kurzen Scripts zum Beispiel unter Verwendung von Frameworks wie *Selenium*¹² (ein Paket zur Steuerung des Browsers) kann der Vorgang automatisiert werden. Das Script klickt sich automatisch durch die 30.000 Entscheide der letzten zehn Jahre durch und speichert diese lokal ab. Eine wünschenswerte Entwicklung wäre daher ein zentralisiertes Repository für alle Gerichtsentscheide und Gesetze, mit einer einfachen Möglichkeit, die Materialien zu exportieren.

Der Schweizer Föderalismus macht die ganze Angelegenheit noch einmal komplizierter, da die Rechtsquellen auf 26 Kantone und die Bundesebene verteilt sind. Für manche Fragen wäre Gemeinderecht interessant, aber dieses ist nicht zentral verfügbar. Die Open-Data-Bewegung¹³ hat also im juristischen Bereich noch viel Arbeit vor sich und kann einen konkreten Nutzen bringen.

4.1 Regular Expression als Freund und Helfer

Viele wesentliche Merkmale können relativ leicht aus einem Gerichtsentscheid ausgelesen werden. Im Fall von Asylentscheiden des Bundesverwaltungsgerichts etwa können *Regular Expressions*¹⁴ eingesetzt werden, um trotz Anonymisierung Daten, wie das Datum des Entscheids, die Sprachregion, die Namen der RichterInnen und AnwältInnen oder die zitierten Rechtsquellen (Gesetze, Verordnungen, internationale Abkommen und Gerichtsentscheide) auszulesen. Trotz Anonymisierung der Namen der Beschwerdeführer ist es möglich, festzustellen, ob es sich um einen Mann (der *Beschwerdeführer*), eine Frau (die *Beschwerdeführerin*) oder eine Familie (die *Beschwerdeführer*) handelt.

¹² <https://www.seleniumhq.org>.

¹³ Siehe z. B. www.opendata.ch oder www.opendata.swiss.

¹⁴ *Regular Expressions* sind eine Folge von Zeichen, die ein Suchmuster definieren. Zum Beispiel findet das Muster `[0-9]{2}` alle Instanzen in einem Text, in dem zwei Ziffern zwischen 0 und 9 aufeinander folgen.

Juristische Zitate folgen strengen Formatierungsregeln. Was Studierenden der Rechtswissenschaften oft Kopfschmerzen bereitet, kommt für systematische Auswertungen sehr gelegen. Der folgende Code beispielsweise liest sämtliche Zitate von Urteilen des Europäischen Gerichtshofs aus:

```
EU_decisions = re.compile(r'C-[0-9]{1,3}/[0-9]{4}')
```

Von Menschen geschriebene Urteile enthalten Fehler. Erfahrungsgemäss kommen zum Beispiel immer wieder Tippfehler im Datum vor. Wird also in einem Dokument mit Hilfe von *Regular Expressions* gesucht, so wird eine gute Kenntnis des Datensatzes sowie ein Reagieren auf fehlende Werte und dementsprechendes Nachjustieren der Suchmuster verlangt.

Der folgende Code findet die Zeile, in welcher das Bundesverwaltungsgericht erklärt, ob eine Beschwerde angenommen oder abgewiesen wird:¹⁵

```
verdict_pattern_de = re.compile(r'erkent das Bundesverwaltungsgericht\s*:\s*1.([^.]*)')
verdict_pattern_it = re.compile(r'Tribunale amministrativo federale pronuncia\s*:\s*1.([^.]*)')
verdict_pattern_fr = re.compile(r'le Tribunal administratif fédéral prononce\s*:\s*1.([^.]*)')
```

Läuft dieser Code über beispielsweise die Entscheide der letzten zehn Jahre, so entsteht ein strukturierter Datensatz, der es erlaubt, Erfolgswahrscheinlichkeiten vor Bundesverwaltungsgericht auszurechnen. In einem von Datenjournalisten des schweizerischen Tagesanzeigers durchgeführtem Projekt wurde beispielsweise die Auswirkung der Parteizugehörigkeit von RichterInnen auf das Ergebnis untersucht. Das Ergebnis war, dass politisch rechtsstehende RichterInnen Beschwerden gegen den Asylentscheid des Staatssekretariats für Migration eher ablehnen (Hosp 2016; Tagesanzeiger 2016) als RichterInnen, die sich zu linken Parteien bekennen.¹⁶

Wenn die Daten strukturiert sind, können mittels Regressionsmodellen Analysen vorgenommen werden (für die Prädiktion binärer Variablen, z. B. bei Beschwerden vor dem Bundesverwaltungsgericht „angenommen/abgewiesen“, eignen sich logistische Regressionen). Je nach Rechtsstreitigkeit sind die Urteile mehr oder weniger kompliziert – viele juristische Details müssen aber berücksichtigt werden, zum Beispiel reicht es nicht, sich nur anzusehen, ob eine Beschwerde angenommen oder abgewiesen wurde, sondern es ist auch wichtig zu verstehen, worauf sich die Beschwerde bezogen hat. Aus diesem Grund verlangt solch eine Übung eine enge Zusammenarbeit von juristischen FachexpertInnen und den *Data Scientists*.

¹⁵ Bundesverwaltungsgerichtsentscheide sind entweder auf Deutsch, Französisch oder Italienisch verfasst.

¹⁶ Dies ist eine Schweizer Besonderheit, dass die Parteizugehörigkeit von Richter öffentlich bekannt ist.

In den USA gibt es nun Modelle, die mittels eines *Random-Forest-Classifiers* Entscheide des amerikanischen Höchstgerichts (*Supreme Court*) vorhersagen können und dabei eine Trefferquote von circa 70 % aufweisen (vgl. (Katz et al. 2017)). Aletras et al. (2016) verwenden einen *Bag-Of-Words*-Ansatz und einen *Support-Vector-Machine-Classifer* um eine Trefferquote von 79 % für Voraussagen von Entscheiden des Europäischen Gerichtshofs für Menschenrechte (EGMR) in Strassburg zu erzielen.

4.2 Deep Learning

Die oben genannten Ansätze sind sehr limitiert – und wären technisch schon vor Jahrzehnten machbar gewesen. Jedoch hat der Boom an Machine-Learning-Methoden auch die Rechtsbranche erreicht und neue Möglichkeiten eröffnet, die deutlich weiter gehen. Dabei sind ganz besonders die Themen *Deep Learning* und, als Methode dessen, *Künstliche Neuronale Netzwerke* zu erwähnen. Diese stellen einen Zweig der *Künstlichen Intelligenz* dar und haben ihren Ursprung in der Neuroinformatik, welche es zum Ziel hat, die Funktionsweise des menschlichen Gehirns zu emulieren.

Herkömmliche Regressionsmodelle weisen ganz klare Koeffizienten für jeden Parameter aus. Eine lineare Regression beispielsweise kann folgendermassen beschrieben werden:

$$\text{Output} = \alpha + \beta_1 * \text{Parameter 1} + \dots + \beta_n * \text{Parameter n}$$

Das Modell schätzt also die Parameter α (die Konstante) und einen Koeffizientenvektor β . Damit kann klar gesagt werden, wie eine Veränderung in Parameter X (z. B. die Parteizugehörigkeit der RichterIn oder des Richters) sich auf das Gesamtergebnis auswirkt. Bei künstlichen neuronalen Netzwerken ist die Interpretation nicht mehr so einfach. Dies wird allerdings dadurch kompensiert, dass solche Ansätze ganz neue Arten der Mustererkennung ermöglichen. So ist es zum Beispiel mit heutiger Technologie Versicherungen möglich, auf einem Foto automatisch zu erkennen, ob das Auto auf dem Foto beschädigt ist oder nicht. Gleichzeitig ermöglichen sie neue Anwendungen auf dem Gebiet des *Natural Language Processing*. Das englische Unternehmen Case Crunch¹⁷ ist hier vermutlich führend.

In einem 2018 durchgeführten Wettbewerb stellte sich Case Crunch 100 WirtschaftsanzwältInnen in London. Im Lauf einer Woche haben die AnwältInnen 750 Vorhersagen getroffen in Fällen, die mit Beschwerden betreffend Zahlungen von Restschuldversicherungen zu tun hatten. Case Crunch hatte eine Genauigkeit von 86,6 %, während die AnwältInnen nur auf 62,3 % kamen. Solche Zahlen sind jedoch mit Vorsicht zu geniessen: ExpertInnen für die verwendeten Rechtsfälle hätten vermutlich deutlich besser abgeschnitten als die in unterschiedlichen Bereichen praktizierenden WirtschaftsanzwältInnen. Laut Case Crunch kam der Vorteil der

¹⁷<https://www.case-crunch.com>.

Software dadurch zu tragen, dass diese besser als JuristInnen nicht-juristische Parameter in ihre Schätzungen einbeziehen kann.

Natural Language Processing wird neuerdings auch von Verwaltungsbehörden eingesetzt, wenngleich nicht zur direkten Analyse des Rechts sondern zum Beispiel zur Unterstützung der Beamten in der Überprüfung der Angaben von Asylsuchenden. Das deutsche Bundesamt für Migration und Flüchtlinge hat gemeinsam mit IBM eine Anwendung entwickelt, die es erlaubt, den Dialekt von AntragstellerInnen zu erkennen. Gemeinsam mit Bilderkennungsprogrammen wird diese eingesetzt, um ihre Herkunftsangaben zu überprüfen (Strerath und Gremer 2017). So kann es etwa mit einer 80 %igen Genauigkeit erkennen, ob eine Person ägyptisches oder syrisches Arabisch spricht – ein entscheidender Faktor um festzustellen, ob eine Person Anrecht auf Asyl hat. Das System erlaubt es angeblich, die notwendige Zeit für die Erstanmeldung von Asylsuchenden von drei Stunden auf 20 Minuten zu reduzieren und hätte dazu geführt, dass 99 Prozent der Asylbewerber korrekte Angaben zu ihrer Herkunft machen (Richter 2018).

Diese Verfahren, nun auch gepaart mit dem Auslesen von Informationen aus den Smartphones (Geodaten und Fotos) stellen jedoch unter Umständen einen schwerwiegenden Eingriff in die Privatsphäre der Antragstellern dar und muss daher weiter grundrechtlich diskutiert werden (zu den rechtlichen Aspekten von High-Tech im Bereich Migrationskontrolle, vgl. (Hanke und Vitiello 2019)).

4.3 Vom Chatbot zum Robo-Lawyer?

Mit den oben genannten Schritten in Bezug auf das Kodieren und Interpretieren von Regeln stellt sich nun die Frage der Schnittstellen (bzw. Interfaces). Eine Möglichkeit wäre, einen Chatbot mit den NutzerInnen interagieren zu lassen, um nach einem Frage-und-Antwort-Spiel eine Rechtsauskunft geben zu können.

Tatsächlich arbeiten mehrere Teams an solchen Implementierungen. In der Schweiz hat die coop Rechtsschutzversicherung mit ihrem in IBM Watson Assistant entwickelten *BlitzBot* eine gewisse Medienaufmerksamkeit erlangt.¹⁸ Der BlitzBot kennt die schweizerischen Strassenverkehrsregeln und unterstützt derzeit die JuristInnen der Versicherung. Die Inspiration zum BlitzBot kam von *ROSS*,¹⁹ einem ebenfalls mit IBM arbeitenden *Natural-Language-Interface*, das Auskünfte zu verschiedenen Teilen des amerikanischen Rechts gibt.

Die Bemühungen der coop Rechtsschutzversicherung haben aber auch KritikerInnen auf den Plan gerufen. Wenngleich die Stiftung für Konsumentenschutz die Entwicklung zwar als durchaus positiv erachtet, so wird vor allem darauf hingewiesen, dass solche Tools von Versicherungen eingesetzt werden könnten, um sehr persönliche Daten zu sammeln und danach individuelle Risikoprofile zu erstellen (Schweizer Radio und Fernsehen 2018).

¹⁸ https://legaltech.weblaw.ch/podcasts/ai_coop_ibm.

¹⁹ www.rossintelligence.com.

Die derzeit bestehenden Lösungen sind noch weit davon entfernt, die Juristin oder den Juristen zu ersetzen. Selbst die weit entwickelten Chatbots scheitern, sobald das Gespräch nicht dem von den EntwicklerInnen vorgegebenem Muster folgt. Sagt eine Person beispielsweise einem (derzeit in Entwicklung befindlichen) Chatbot zum Schweizer Migrationsrecht,²⁰ dass sie in die Schweiz ziehen möchte (Eingabe: „*I want to move to Switzerland*“), so erkennt der Chatbot die Absicht (engl. *intent*) und kann gezielt die migrationsrechtlich relevanten Fragen stellen („*What is your country of citizenship?*“;²¹ „*Do you have a signed contract with a Swiss employer?*“). Idealerweise kann der Chatbot das Gespräch so lange aufrechterhalten, bis er alle notwendigen Informationen gesammelt hat, eine Rechtsauskunft geben und dann auch noch selber alle nötigen Dokumente ausfüllen kann.

Das System scheitert, sobald die Nutzerin oder der Nutzer eine längere Frage mit mehreren Aspekten eingibt:

„Ich bin vor zehn Jahren von Deutschland in die Schweiz gekommen. Zwischen 2011 und 2013 habe ich in Frankreich gelebt. Unlängst bin ich von Bern nach Zürich gezogen. Ich wurde jetzt wegen Steuerbetrugs verurteilt. Muss ich mit einem Landesverweis rechnen? Was kann ich dagegen tun?“

Mit dieser Frage hätte wohl jeder Chatbot seine Mühe, obwohl sie gut formuliert ist und nach ein paar Nachfragen juristisch relativ leicht zu beantworten wäre. Ein Dialog ist mit so einer Frage kaum möglich. Eine Lösung wäre in so einem Fall nur noch das Anzeigen von Resultaten einer Suche in einer Wissensdatenbank. In dem konkreten Beispiel wäre es sinnvoll, Informationen zum Themenkomplex Ausschaffungsinitiative anzuzeigen. Hier kommt also zum Tragen, dass rechtliche Probleme im Vergleich zu Alexa²²- oder Siri²³-Anfragen („*Wie wird das Wetter heute?*“, „*Ruf meine Frau an*“, etc.) eher zu langen Nutzereingaben mit verschachtelten Sätzen einladen. Zur Komplexität kommt auch noch hinzu, dass es relativ wenige Daten (historische Anfragen) gibt, mit denen Machine-Learning-Algorithmus trainiert werden können – das Problem ist umso gravierender, wenn wichtige Informationen aus den Trainingsdaten entfernt werden müssen, um den Datenschutzregeln zu genügen.

Neben Chatbots, die dem Zweck der anwaltlichen Beratung dienen, sollte noch das Thema Schiedsgerichtsbarkeit erwähnt werden. Angenommen *Robo-Lawyers* wären technisch machbar, so wäre ein naheliegender nächster Schritt der *Robo-Judge* (deut. *Robo-RichterIn*). Auch wenn in den letzten Jahren der Bereich Online-Streitbeilegung als alternative zum Gerichtsverfahren – angeheizt durch die EU-ADR-Richtlinie²⁴ – einen Boom erlebt hat, so scheint Automatisierung in diesem Bereich (aus guten Gründen) bis auf Weiteres kein Thema zu sein. Da Streitschlichtungsverfahren privat-

²⁰ <http://swissmigrationlaw.mybluemix.net>.

²¹ Eine wichtige Frage, da das Schweizer Recht zwischen EU/EFTA-Bürgern, die unter das Personenfreizügigkeitsabkommen fallen, und sogenannten Drittstaatsangehörigen unterscheidet.

²² <https://developer.amazon.com/de/alexa>.

²³ <https://www.apple.com/siri>.

²⁴ Richtlinie 2013/11/EU des Europäischen Parlaments und des Rates vom 21. Mai 2013 über die alternative Beilegung verbraucherrechtlicher Streitigkeiten und zur Änderung der Verordnung (EG) Nr. 2006/2004 und der Richtlinie 2009/22/EG (Richtlinie über alternative Streitbeilegung in Verbraucherangelegenheiten).

rechtlich und den Wünschen der Vertragsparteien entsprechend organisiert werden und somit dynamischer sind als der Staat, wäre der Einsatz von Robo-Judges (in welcher Form auch immer) zuerst in diesem Bereich zu erwarten, bevor die Technologie die ordentlichen, staatlichen Gerichte erreicht. Somit fallen Online-Schiedsgerichte zwar aufgrund ihres innovativen Charakters in den Bereich Legal Tech, die rechtliche Analyse wird aber von menschlichen JuristInnen ausgeführt.

4.4 Was, wenn sich der Chatbot irrt?

Eine Frage, die sich mit der Entwicklung von Chatbots stellt, ist jene der Haftung im Fall von Fehlauskünften. Mangels Präzedenz gilt diese Frage derzeit noch als ungeklärt. Wenn nun also eine Anwaltskanzlei ihren KlientInnen eine Rechtsauskunft mittels Chatbot anbietet, wird es Aufgabe der Gerichte sein, zu klären, ob der Fehler aufseiten der Anwälte lag, die den Chatbot implementiert und angeboten haben, oder ob ein Teil der Schuld beispielsweise auf das System im Hintergrund, konkret IBM Watson Assistant²⁵ (vormals Conversation) abgewälzt werden kann.

Abgesehen von der Haftungsfrage ist Legal Tech noch mit Problemen der Berufsregulierung konfrontiert. Der Anwaltsberuf ist in allen Ländern ein regulierter Beruf. Üblicherweise folgen auf die akademische Ausbildung noch ein oder zwei Praxisjahre und meist noch eine weitere praktische Ausbildung (z. B. der *Legal Practice Course* in England). Danach können RechtsanwaltsanwärterInnen die Anwaltsprüfung machen (das *Bar Exam* in englischsprachigen Ländern bzw. das Zweite Staatsexamen in Deutschland).

Üblicherweise dürfen nur Juristen mit einem sogenannten Anwaltspatent Personen vor Gericht vertreten. Anders als etwa in Deutschland²⁶ darf in der Schweiz hingegen die Rechtsberatung (also die juristische Praxis mit Ausnahme der Vertretung vor Gerichten sowie die notarielle Tätigkeit) von jedem und jeder, also auch von Personen ohne juristische Ausbildung und ohne Anwaltspatent vorgenommen werden.²⁷ Vor Gericht herrscht keine Anwaltspflicht, das heisst Personen dürfen sich selber vertreten. Gerade dann, wenn sich JuristInnen verstärkt mit InformatikerInnen selbstständig machen (in Form eines Startups), dann sind diese rechtlichen Regelungen entscheidend für die Möglichkeiten solcher Projekte.

²⁵ <https://www.ibm.com/cloud/watson-assistant>.

²⁶ Siehe die gerichtlichen Verfahrensordnungen für die Vertretung vor Gericht sowie das Gesetz über außergerichtliche Rechtsdienstleistungen (Rechtsdienstleistungsgesetz) vom 12.12.2008 für den rechtlichen Rahmen des selbstständigen Erbringens aussergerichtlicher Rechtsdienstleistungen.

²⁷ Siehe das Bundesgesetz über die Freizügigkeit der Anwältinnen und Anwälte (Anwaltsgesetz, BGFA) vom 23. Juni 2000 sowie die kantonalen Anwaltsgesetze.

5 Smart Contracts und Decentralized Ledgers

Die *Blockchain*-Technologie, mit der dezentrale Datenbanken betrieben werden können, ist derzeit ein grosses Thema unter JuristInnen. Dafür gibt es primär zwei Gründe. Der erste Grund ist, dass die Finanzierungsform des *Initial Coin Offerings*, bei welchem eine neu emittierte Kryptowährung im Tausch gegen staatlich emittierte Währungen oder andere Kryptowährungen verkauft werden um ein Projekt zu finanzieren, einher geht mit einer Menge an aufsichtsrechtlichen Anforderungen. Um diese erfüllen zu können, müssen JuristInnen schon in der Designphase des Projekts integriert sein und bei der Planung des *Tokens* (der neu emittierten Kryptowährung) die rechtliche Einordnung beispielsweise in *Payment Tokens*, *Utility Tokens* und *Security Tokens* vorzunehmen. Die Verzahnung von Rechtsberatung und technologischem Wissen ist hier ein besonderes, neues Merkmal.

Zweitens üben *Smart Contracts*, die auf der Blockchain laufen, eine besondere Faszination aus. Smart Contracts sind kleine Programme, die Verträge abbilden oder die Abwicklung eines Vertrags technisch unterstützen (siehe z. B. (Gantner 2018)). Tatsächlich steht hier unter Umständen eine grosse Umwälzung an, wenn Verträge automatisch exekutiert werden. So wäre es denkbar, dass eines Tages Testamente ohne Zutun von Notaren ausgeführt werden. Gleichzeitig können sogenannte parametrische Versicherungen in Smart Contracts implementiert. So könnte der Contract von selbst Wetterdaten herunterladen und zum Beispiel im Rahmen eines Hochwasserversicherungsvertrags bei starkem Regen die Versicherungszahlung auslösen.

Derzeit ist es noch lange nicht möglich, rechtliche Verträge als Smart Contracts abzubilden. Eine Zwischenlösung sind sogenannte *Ricardian Contracts*,²⁸ welche als Konzept auf Grigg (2004) zurückgehen und somit noch aus der Ära vor Blockchain-basierten Smart Contracts stammen. In so einem Vertrag wird ein traditioneller, natürlichsprachiger Vertrag kryptographisch mit einem maschinellen System verknüpft. Der Text des Vertrags, welcher weiterhin für JuristInnen lesbar bleibt, wird mit *Markup* versehen und somit parsierbar. Somit entsteht ein Tupel aus Prosa, Parametern und Code. Ein Versuch, eine Markup-Sprache für juristische Texte zu entwerfen ist das Projekt Legal XML,²⁹ ein Teil der *Organization for the Advancement of Structured Information Standards (OASIS)*.³⁰ Obwohl es schon Mitte der 1990er-Jahre ins Leben gerufen wurde, hat es sich jedoch seitdem nicht wirklich durchsetzen können und führt nach wie vor eher ein Schattendasein.

Während Smart Contracts darauf ausgelegt sind, den Willen der Vertragsparteien auszuführen, stellen Ricardian Contracts nur ein Design-Muster zur Abbildung der Intention dar. Blockchain-Technologie allerdings erlaubt es, solche

²⁸ Der Name bezieht sich auf das Zahlungssystem Ricardo von Systemics, welches wiederum nach dem englischen Wirtschaftswissenschaftler David Ricardo benannt ist, der massgeblich zur Handelstheorie beigetragen hat.

²⁹ <http://www.legalxml.org>.

³⁰ <https://www.oasis-open.org>.

Verträge quasi-notariell zu beglaubigen. Diese Technologie steht somit eher in Konkurrenz zu NotarInnen und nicht direkt zu den JuristInnen.

Eine Schwierigkeit, bis jetzt, ist, dass die Welt der Smart Contracts für JuristInnen noch sehr schwer zugänglich ist. Brücken zwischen den beiden Welten wären wünschenswert, sei es, dass JuristInnen *Solidity*, die Programmiersprache für Smart Contracts auf der *Ethereum*-Blockchain, lernen (eher unwahrscheinlich), oder dass Anwendungen entstehen, die etwa Templates für Smart Contracts zur Verfügung stellen.³¹ Angesichts des technischen Schwierigkeitsgrads braucht es dann *Code Audits*, welche die gewünschte Funktionsfähigkeit des Programms überprüfen. Als Reaktion auf den Initial Coin Offerings-Boom haben sich solche Audit-Unternehmen mittlerweile herausgebildet, welche sowohl die Smart Contracts des Initial Coin Offerings als auch des *Minimum Viable Products* prüfen.³²

6 Totale Kontrolle durch Legal Tech

Während Legal Tech ein Mittel zur Demokratisierung der Gesellschaft sein kann, so kann von staatlicher Seite eingesetzte Technologie, die Verknüpfungen zwischen unterschiedlichen Datensätzen herstellt, ins absolute Gegenteil umschlagen. Aktuelles Anschauungsbeispiel ist das chinesische *Sozialkredit-System* (engl. *social credits*).

Das Sozialkredit-System ist ein nationales Reputationsbewertungssystem, das derzeit von der chinesischen Regierung entwickelt wird (für eine englische Übersetzung des chinesischen Regierungsdokuments, siehe (Creemers 2014)).³³ Ziel ist, bis 2020 den *social credits* von Bürgern und Unternehmen zu bewerten. Das System ist derzeit noch in Entwicklung und nur wenige Details stehen derzeit fest. Das System sammelt Informationen über Bürger, unter anderem aus Verbindungsdaten und sozialen Medien. Dazu kommen Daten über die Nutzung von Spielkonsolen, bevorzugten Zeitungen und weiteren, sehr persönlichen Daten. Algorithmen strukturieren diese Daten und berechnen die Übereinstimmung mit beziehungsweise Abweichung von politisch gewolltem Verhalten. Negative Konsequenzen sind derzeit, dass Personen mit einem schlechten Score der Zugang zu Flügen, Hochgeschwindigkeitszügen und Privatschulen eingeschränkt wird, dass die Internetgeschwindigkeit für ihre Nutzung gedrosselt wird und das negative Stigma, beispielsweise auf Online-Kontaktbörsen (eine zeigt mittlerweile den Score auf den Nutzerprofilen an). Belohnungen für gutes Verhalten sind ein einfacherer Zugang zu Krediten, Jobs und schnellere Behandlung von Verwaltungsabläufen (Botsman 2017).

³¹ Zum Zeitpunkt des Drucks dieses Kapitels war die Contract Vault GmbH ein Early-Stage-Startup auf Finanzierungssuche, das versucht, eine Plattform für Smart-Contract-Templates zu entwickeln und sich gezielt an Juristen richtet.

³² In der Schweiz ist etwa die ChainSecurity AG (www.chainsecurity.com) ein Startup in diesem Bereich.

³³ Das chinesische Original ist auf www.gov.cn/zhengce/content/2014-06/27/content_8913.htm abrufbar.

Das Sozialkredit-System ist somit ein auf grossen Daten gestütztes System zur Massenüberwachung, welches allem Anschein nach tief in das chinesische Rechtssystem eingebettet sein wird. Auch dies ist eine Form von Legal Tech, die jedoch als Warnung dienen sollte. Es zeigt eindrücklich, wie wichtig die Hinweise auf etwaige Datenschutz- und grundrechtliche Probleme im Zusammenhang mit grossen Datensätzen sind und dass Konzepte wie die *Smart City* (d. h. ein System der permanenten Interaktion zwischen StadtbewohnerInnen und der sie umgebenden Technologie) unter den falschen gesellschaftlichen Vorzeichen schnell zur Dystopie führen kann.

7 Ausblick

Die juristische Praxis durchläuft derzeit einen Digitalisierungsschub. Dabei kommen unterschiedliche Technologien zum Einsatz. Einerseits ist da etwa der Bereich Dokumentautomatisierung, der zwar technologisch nicht sonderlich innovativ ist (in dem Sinn, dass er ohne komplizierte statistische Modelle und Natural Language Processing auskommt), nun aber einen Boom erlebt. Andere Bereiche sind hingegen viel näher am *State of the Art*, kommen aber mit wesentlichen Einschränkungen. Dazu zählen sämtliche Entwicklungen, die sich maschinelles Lernen zu eigen machen. Durchaus Fortschritte gibt es etwa im Bereich *Legal Analytics*, also etwa dem Vorhersagen von Gerichtsentscheiden. Es hat sich gezeigt, dass schon mit dem Verwenden von Metadaten und den wenigen leicht strukturierbaren Daten zu einem gewissen Grad durchaus nützliche Ergebnisse erzielbar sind. Erste Versuche, intelligente Chatbots zu entwickeln, die Rechtsauskünfte geben können, waren zwar noch nicht sonderlich erfolgreich, zeigen aber Potential. Es wird sich also zeigen, inwieweit beziehungsweise wie lange die natürliche Sprache, die im Zentrum der juristischen Praxis steht, noch eine technologische Hürde darstellt. Angesichts der Möglichkeit von Bussen und Gefängnisstrafen können Fehler schwerwiegende Folgen für Personen und Unternehmen haben, weshalb die Anforderungen an Legal-Tech-Systeme höher sind als etwa jene an Suchmaschinen oder Produktempfehlungsalgorithmen.

Über die nächsten Jahre werden auf jeden Fall viele der einfachen, repetitiven Tätigkeiten, welche derzeit vor allem von jüngeren AnwältInnen ausgeübt werden, automatisiert werden. Komplizierte Tätigkeiten wie zum Beispiel das Argumentieren vor Gericht – oder gegebenenfalls die Entwicklung neuer kognitiver Systeme – wird auf absehbare Zeit aber Menschen vorbehalten bleiben. Um Ian Dodd von *Premonition*,³⁴ einer grossen Datenbank von Gerichtsurteilen, zu zitieren (Lee 2017):

„The knowledge jobs will go, the wisdom jobs will stay.“

³⁴ <https://www.premonition.ai>.

Andere sehen die JuristInnen der Zukunft als *Transaction Engineers* und betonen daher die Wichtigkeit der Coding-Ausbildung schon während des rechtswissenschaftlichen Studiums (Fenwick et al. 2018).

Viele der genannten Beispiele, seien es Chatbots oder Online-Schiedsgerichte, erfüllen einen Zweck der weit über die einfache Produktivitätssteigerung von AnwältInnen hinausgeht. Letztlich ermöglichen viele dieser Systeme den leichteren Zugang zu Recht und Gerechtigkeit. Nachdem jede Bürgerin und jeder Bürger angehalten ist, sich an das Gesetz zu halten („*Unwissenheit schützt vor Strafe nicht*“ ist Bestandteil aller modernen Rechtssysteme und war sogar schon Teil des Römischen Rechts), muss dieses auch für alle verständlich und zugänglich sein – in vielen Fällen ist dies aufgrund der Komplexität moderner Rechtssystem nur unter Einbeziehung eines Rechtsbeistands möglich. Somit ist Legal Tech ein wesentlicher und notwendiger Bestandteil der *Smart City* oder *Digital Society* von morgen.

Literatur

- Aletras N, Tsarapatsanis D, Preoțiu-Pietro D, Lamos V (2016) Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective. *PeerJ Computer Science*, 2(e933).
- Botsman R (2017) *Who Can You Trust?: How Technology Brought Us Together – and Why It Could Drive Us Apart*. Penguin UK.
- Creemers R (2014) Planning Outline for the Construction of a Social Credit System (2014–2020). China Copyright and Media (blog). <https://chinacopyrightandmedia.wordpress.com/2014/06/14/planning-outline-for-the-construction-of-a-social-credit-system-2014-2020>. Erstellt: 14.06.2014. Letzter Zugriff: 14.08.2019.
- D’Onofrio S, Franzelli SM, Portmann E (2018) Advancing Cognitive Cities with Web of Things. Yager RR, Pascual Espada J (Hrsg.) *New Advances in the Internet of Things. Studies in Computational Intelligence: Vol. 715*. Springer International Publishing.
- D’Onofrio S, Portmann E (2017) Cognitive in Smart Cities. *Informatik-Spektrum*, 40(1):46–57.
- D’Onofrio S, Portmann E, Franzelli M, Bürki C (2018b) Cognitive Computing – Theoretische Grundlagen und Praxisbeispiele der Schweizerischen Post. *Informatik-Spektrum*, 41(2):113–122.
- Fenwick M, Kaal WA, Vermeulen EPM (2018) Legal Education in a Digital Age: Why ‘Coding for Lawyers’ Matters. SSRN Scholarly Paper ID 3227967. Lex Research Topics in Corporate Law & Economics Working Paper No. 2018-4. Rochester, NY: Social Science Research Network.
- Ganter F (2018) „Code Is Law“ aber „Is Law Code“? Jusletter IT, IRIS. http://jusletter-it.weblaw.ch/issues/2018/IRIS/-code-is-law%2D%2Daber%2D%2D_c692026243.html__ONCE. Erstellt: 22.02.2018. Letzter Zugriff: 14.08.2019.
- Genesereth M (2015) Computational Law: The Cop in the Backseat. CodeX: The Center for Legal Informatics Stanford University. The Stanford Center for Legal Informatics.
- Grigg I (2004) The Ricardian Contract. *Proceedings of the First IEEE International Workshop on Electronic Contracting*, pp. 25–31.
- Hanke P, Vitiello D (2019) High-Tech Migration Control in the EU and beyond: The Legal Challenges of ‘Enhanced Interoperability’. Carpanelli E, Lazzerini N (Hrsg.) *Use and Misuse of Technology: Contemporary Challenges for International and European Law*. Springer.
- Hosp J (2016) Das Parteibuch der Richter entscheidet mit. *Tages-Anzeiger*. <https://www.tages-anzeiger.ch/schweiz/standard/das-partreibuch-der-richter-entscheidet-mit/story/28944144>. Erstellt: 10.10.2016. Letzter Zugriff: 14.08.2019.

- Katz DM, Bommarito MJII, Blackman J (2017) A General Approach for Predicting the Behavior of the Supreme Court of the United States. PLOS ONE, 12(4):e0174698.
- Lee R, CJ D (2017) The Robot Lawyers Are Here and Winning. BBC News. <https://www.bbc.com/news/technology-41829534>. Erstellt: 01.11.2017. Letzter Zugriff: 14.08.2019.
- Richter M (2018) Transforming Immigration Operations by Applying AI and Big Data. Client Success Field Notes. IBM. <https://www.ibm.com/blogs/client-voices/immigration-operations-ai-big-data>. Erstellt: 01.03.2018. Letzter Zugriff: 11.07.2019.
- Schweizer Radio und Fernsehen (2018) «Busse? Fragen Sie den Blitzbot – Roboter als digitale Juristen». www.srf.ch/news/panorama/busse-fragen-sie-den-blitzbot-roboter-als-digitale-juristen. Erstellt: 25.07.2018. Letzter Zugriff: 14.08.2019.
- Strerath F, Gremer H (2017) Gegen falsche Flüchtlinge: BAMF prüft Asylanträge mit neuer Technik. Bayrischer Rundfunk. <https://www.br.de/nachrichten/mittelfranken/inhalt/bamf-asyl-digitale-assistenzsysteme-100.html>. Erstellt: 25.07.2017. Letzter Zugriff: 14.08.2019.
- Tagesanzeiger (2016) Das sind die härtesten Asylrichter in der Schweiz. Datenblog. <https://blog.tagesanzeiger.ch/datenblog/index.php/12556/je-nach-richter-dreimal-hoehere-erfolgchancen>. Erstellt: 08.10.2016. Letzter Zugriff: 14.08.2019.

Die Digitalisierung des Skillmanagements – Kognitiver Digitaler Assistent für die Wertschöpfung des Human Ressource Management in der Digitalen Transformation

Adrian Vogler

Zusammenfassung

Die Digitalisierung ist überall und im Zentrum dieser neuen Welle der Digitalisierung geht es (wieder einmal) um den Einsatz von künstlicher Intelligenz um eine Wertschöpfung zu erzielen. Im Rahmen der durch die Digitalisierung ausgelösten Transformationen werden nicht nur neue digitale Wertschöpfungsketten gebildet. Innerhalb dieser neuen, digitalen Wertschöpfungsketten werden die Aufgabenteilungen zwischen Mensch und Maschine neu definiert. Dieser Artikel behandelt die gleichermassen aktuelle wie relevante Frage nach der digitalen Transformation im Personalbereich. Er beschäftigt sich mit dem dortigen Nutzen einer Cognitive-Computing-Plattform, als eine Ausprägung für Systeme künstlicher Intelligenz, wie man sie heute bauen kann, sowie ihre Auswirkungen auf Unternehmen und Mitarbeiter. Von dieser Transformation sind alle Bereiche eines Unternehmens betroffen. Jeder Bereich muss für eine erfolgreiche Umsetzung seinen Beitrag leisten. Der Einsatz von künstlicher Intelligenz im Human Ressource Management (HR, deut. Personalmanagement) unterstützt darüber das Aufsetzen auf bereits vorhandene Ansätze für ein strategisches Workforce Management.

Überarbeiteter Beitrag basierend auf Vogler, A.: Die Digitalisierung des Skillmanagements, Informatik Spektrum (2018) 41(2): 97–104.

A. Vogler (✉)
StratWorX UG, Winkelhaid, Deutschland
E-Mail: adrian.vogler@fachliteratur.onmicrosoft.com

Dieser Artikel stellt eine Erweiterung eines bereits in der Zeitschrift Informatik Spektrum erschienenen Artikels *Die Digitalisierung des Skillmanagements – Digitale kognitive Assistenten für die Wertschöpfung des Human Resource Managements in der Digitalen Transformation* (Vogler 2018) dar.

„Nur wenige Veränderungen beeinflussen die Zivilisation derart nachhaltig wie eine Änderung des Prinzips, auf dem die Organisation der Arbeit beruht.“¹

„Die vierte Industrielle Revolution verändert nicht nur die Arbeitswelt, sie erzeugt auch enormen Bedarf an neuen Fähigkeiten. Damit steht uns eine globale Talentkrise ins Haus.“²

1 Motivation

Eine wesentliche Folge der digitalen Transformation des Wirtschaftslebens besteht darin, dass durch ein noch höheres Mass an Automatisierung auch nicht-manueller Tätigkeiten etablierte Berufsbilder verschwinden und durch neue ersetzt werden. Von diesem Wandel wird erwartet, dass der zukünftige Einsatz intelligenter Technologien die Arbeitsergebnisse weitreichend verbessert. Das ist insbesondere dann wichtig, wenn ein Unternehmen im Rahmen der digitalen Transformation sein Geschäftsmodell zur Adressierung neuer Kundenanforderungen ändert. Alle Beteiligten sind aufgrund der neuen Anforderungen vor die Frage gestellt, inwiefern beziehungsweise wie lange die durch die aktuelle Arbeitsteilung festgelegten Rollen und die dahinterstehenden Kompetenzprofile noch von Bedeutung sind. Oder, ob während und nach der Transformation neue beziehungsweise geänderte Fähigkeiten, Fertigkeiten und Erfahrungen (im Folgenden kurz *Skills* genannt) erforderlich sind. Wenn das qualitativ der Fall ist, dann stellt sich natürlich auch die Frage, wie sich das quantitativ auswirkt. Die Unternehmensleitung ist vor einer Umsetzung von konkreten Transformationsschritten daran interessiert, wie sich die Anforderungen des digital transformierten Geschäftsmodells auf die *Workforce* abbilden. Die Mitarbeiter müssen sich dann aufgrund der neuen Anforderungen mit der Frage befassen, inwiefern beziehungsweise wie lange die von ihnen aktuell bekleideten Rollen und die dahinterstehenden Kompetenzprofile noch von Bedeutung sind. Und wie gut passen die Skills, über die sie verfügen, zu dem neuen, digital transformierten Wertschöpfungsprozess und den neuen, dort benötigten Rollen?

¹Edersheim EH (2007) Peter F. Drucker – Alles über Management. Redline Verlag.

²http://www.spiegel.de/karriere/human-capital-index-2017-wef-studie-zu-humankapital-a1167348.html?xing_share=news.

2 Eine (Berufs-)Welt im Wandel

Was ist ein Beruf eigentlich?

- „Ein Beruf ist eine Tätigkeit, die man dauerhaft ausübt, für die man eine bestimmte Ausbildung besitzt und mit der man seinen Lebensunterhalt verdient.“³
- „Ein Beruf ist die im Rahmen einer arbeitsteiligen Wirtschaftsordnung aufgrund besonderer Eignung und Neigung systematisch erlernte, spezialisierte, meistens mit einem Qualifikationsnachweis versehene, dauerhaft und gegen Entgelt ausgeübte Betätigung eines Menschen.“⁴
- „Unter Beruf versteht man verfassungsrechtlich jede auf Dauer angelegte, der Einkommenserzielung dienende menschliche Betätigung.“⁵

Ein Blick zurück... Mit diesen Definitionen ausgestattet, werfen wir erst einmal einen Blick zurück in die Vergangenheit der Berufe. Wer kennt noch die folgenden Berufe oder kann sich vorstellen, was ein Abdecker, ein Fischbeinreisser (Spiegel Online 2017), ein Kalfater, ein Lustfeuerwerker oder ein Planetenverkäufer machte? Und wie sieht es aus, wenn wir einen Blick nach vorn tun? Wer hat bereits heute ein klares Bild davon, was ein Mangel-Analysator oder ein Bioabfall-Optimierer zukünftig tun müssen? (Frey o. J.)

Wir tun uns um ein Vielfaches leichter, wenn stattdessen *nur* die Frage zu beantworten ist, was ein Taxifahrer oder ein Sportreporter machen. Hier fällt einem die Antwort darauf sicher leichter. Warum ist das so? Das sind Berufe, die man kennt. Manche aus eigenem Erlebten. Oder wer hat sich bereits nicht mit einem Taxifahrer unterhalten oder Sportnachrichten gesehen oder gehört? Deshalb kann man sich gut vorstellen, was jemand, der diesen Beruf ausübt, tagtäglich leisten muss.

Was will uns der Blick zurück und der Blick nach vorn zeigen? Die Tatsache, dass der Wandel immer mit – manchmal rapiden – Veränderungen in der Arbeitswelt einhergeht. Diese rapiden Veränderungen der Arbeitswelt haben in der Vergangenheit bereits Hunderte von ausgestorbenen Berufen hinterlassen (Palla 2014). Viele unserer Vorfahren haben ihr Leben lang Berufe ausgeübt, von denen wir heute rein gar nichts mehr wissen. Geschweige denn, von dem Beruf schon mal etwas gehört haben. Unsere Enkel und Urenkel werden mit Sicherheit eine Vielzahl von Berufen ausüben können, die erst noch neu entstehen werden.

Insbesondere die Digitalisierung, die in allen Bereichen der Wirtschaft unterwegs ist, beschleunigt diesen Prozess der Transformation der Berufswelt eher, als dass sie ihn abschwächt. So werden diese Veränderungen, die sich in der Vergangenheit manchmal über Generationen erstreckt haben, für uns plötzlich innerhalb einer

³ https://www.bing.com/search?q=was+ist+ein+beruf&form=EDGTCT&q=DA&cvid=354d-d7ce1bae41309868904cf750078d&cc=DE&setlang=de-DE&elv=AXXfrEiqqD9r3GuelwApul-pUHcuBbuwNzeNYZP4%21qLTthV84cAXtdyNQBpufvsWAbMT3Jmuab8kl*8ap09o7Ds6a-4tYX7ff0jT4Jd%2142aSzB.

⁴ <https://de.wikipedia.org/wiki/Beruf>.

⁵ BVerfGE 7, 377, 397.

Generation wahrnehmbar. Selbst wenn sich diese Transformation vielleicht noch nicht in einem grösseren Umfang in den Berufsbezeichnungen widerspiegelt, so läuft diese Veränderung bereits *auf der Ebene darunter*. Man muss nur einen Blick auf die Inhalte der Stellenanzeigen werfen. Bei den Skills und Fertigkeiten, über die man verfügen sollte, ist die Digitalisierung allgegenwärtig.

Thomas Frey, seines Zeichens „Futurist Speaker“, hat sich im Rahmen seiner Arbeit insbesondere mit der Auswirkung auf die Berufswelt beschäftigt. Dabei hat er eine Liste von 101 Berufe erstellt, die es aus seiner Sicht in der Zukunft so wohl nicht mehr geben wird (Frey 2014). Seine zweite Liste enthält die Berufe der Zukunft. Diese Liste umfasst 162 Einträge (Frey o. J.). Folglich sind 101 Berufe vom Aussterben bedroht. Dafür werden aber 162 neue Berufe geschaffen. Auf den ersten Blick ist diese Bilanz also positiv.

3 Bestehende Ansätze

Aktuell werden im Bereich des Skillmanagements Werkzeuge eingesetzt, die die persönlichen Fähigkeiten der Mitarbeiter in einer Matrix mit unterschiedlichem Detaillierungsgrad erfassen. Die Fähigkeiten werden einem Reifegrad zugeordnet; insbesondere bei Unternehmen mit >200 Mitarbeitern. Diese Informationen werden genutzt, um geeignete Mitarbeiter für diverse Projekte zu identifizieren. Das Problem ist, dass zukünftige Skills hier nicht abgebildet werden. Die Weiterentwicklung der Mitarbeiter wird meist auf einer separaten Ebene organisiert. Mittlerweile existieren erste Werkzeuge, die Kursinhalte und Karrierepfade vorschlagen können (z. B. *Cornerstone OnDemand*,⁶ *PeopleFluent*).⁷ Diese greifen aber häufig zu kurz. Eine Vielzahl sehr bedeutsamer Faktoren wird bei der Vorschlagsgenerierung leider vernachlässigt.

Die Unternehmensstrategie, Markt- und technologische Entwicklungen und persönliche Präferenzen der Mitarbeiter sollten nicht ausser Acht gelassen werden. Das liegt unter anderem auch daran, dass diese Werkzeuge noch auf Technologien beruhen, die einen ganzheitlichen Ansatz nicht unterstützen. Darüber hinaus hält die Steuerung externer und interner Mitarbeiter besonders in Zeiten der Globalisierung vielfältige Herausforderungen bereit. Diese können mit den genannten Werkzeugen überhaupt nicht adressiert werden. Ferner kommt es in Zeiten permanenter Transformation und Restrukturierung darauf an, die benötigten Kompetenzen schnell und flexibel mobilisieren zu können. Ein Planungsmodell für das strategische Workforce Management ist in Abb. 1 zu sehen.

⁶<https://www.cornerstoneondemand.de>.

⁷<https://www.peoplefluent.com>.

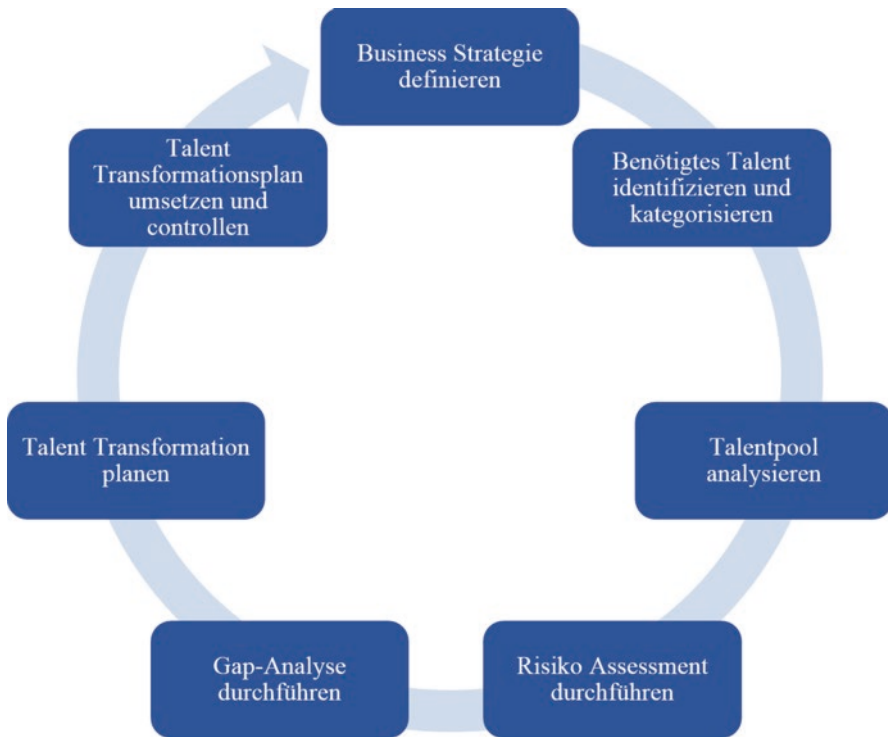


Abb. 1 Das Planungsmodell zum strategischen Workforce Management

4 Offene Fragen – Details zum Strategischen Workforce Management

Die Fragen, die es beantworten will, lauten deshalb unter anderem:

- Wo muss die Wettbewerbsfähigkeit – möglicherweise global operierenden – in Unternehmensteilen durch einen geänderten *Workforce-Mix* verbessert werden?
- Werden durch die Technologiesprünge mehr oder weniger Mitarbeiter benötigt?
- Wo muss ein tiefgreifender Skillshift erfolgen, um neue Geschäfte und Innovationen zu ermöglichen?
- Und wie kann ein solch notwendiger bereichsübergreifenden Skillshift organisiert werden?

Nutzen keine Frage der Unternehmensgrösse

Dabei spielt die Unternehmensgrösse, wenn es um den prinzipiellen Einsatz des strategischen Workforce Management geht, keine entscheidende Rolle. Grosse und globale Unternehmen weisen meist eine Mitarbeiteranzahl im sechsstelligen Bereich auf. Insbesondere die Entwicklung und Neuverteilung der bestehenden Workforce wird hier im Fokus stehen. Doch auch bei mittleren und kleineren Unternehmen schafft das Instrument Transparenz über die vorhandenen personellen Ressourcen. Der Fokus kann hier zum Beispiel in der Ableitung von massgeschneiderten Karrierepfaden liegen. Ausserdem wird durch den Einsatz des Instrumentes schnell deutlich, wo vielleicht in bestimmten Kernkompetenzen keine ausreichenden Ressourcen vorhanden sind. Mit geeigneten HR-Massnahmen kann so frühzeitig gegengesteuert werden.

Ein Instrument für die Digitale Transformation?

Es empfiehlt sich mit dem Einsatz des strategischen Workforce Managements in Bereichen zu beginnen, die vor tiefgreifenden Veränderungen stehen – beispielsweise durch neue Technologien – und die einen signifikanten Anteil an den Personalressourcen des Unternehmens ausmachen. Dort scheint das Instrument ideal auf das Adressieren der Herausforderungen zugeschnitten zu sein, vor die die digitale Transformation ein Unternehmen stellt.

5 Erfolgsfaktoren & Aufwände

Da die Einführung des strategischen Workforce Managements alle Unternehmensteile betrifft, ist die Unterstützung durch das Top-Management eine zwingende Voraussetzung. Ausserdem leiten sich die Fragen direkt aus der vom Top-Management vorgegebenen Unternehmensstrategie ab. Auch gilt es, die unterschiedlichen Interessengruppen (engl. *stakeholder*) vom Mehrwert dieses aktiven, ganzheitlichen Ansatzes zu überzeugen. Neben einem professionellen Projektmanagement spielen natürlich weitere Aspekte, wie zum Beispiel eine Abstimmung mit Betriebsrat und Datenschutz eine Rolle.

Das Instrument verspricht eine transparente und nachvollziehbare Personalplanung zur Verfügung zu stellen – abgeleitet aus der Unternehmensstrategie. Die Akzeptanz bei den beteiligten Sozialpartnern stellt somit kein unlösbares Problem dar. Der Einbezug unterschiedlichster gesetzlicher Regelungen (z. B. im Arbeits- und Datenschutzrecht) in unterschiedlichen Ländern, hat sich vom Aufwand allerdings als nicht zu unterschätzen herausgestellt. Das A und O, damit das strategische Workforce Management sein Nutzenversprechen einlösen kann, ist der kontinuierliche Einbezug von verlässlichen Daten aus dem Unternehmen und seinem Umfeld. Dieses ist – wie bei jedem Instrument aus dem Data Analytics Bereich – essentiell.

Da das Instrument auf eine grosse Menge an Daten angewiesen ist, ist dafür eine konzernweite, einheitliche und strukturierte Erfassung notwendig. Dazu muss man sich mit Themen beschäftigen, wie zum Beispiel:

- Welche Daten werden benötigt?
- Welche Daten sind bereits vorhanden?
- Welche Unterschiede gibt es bezüglich der Bedarfseinschätzung, des Verhaltens und der Bewertung in (bei global agierenden Unternehmen) in unterschiedlichsten Kulturkreisen?

Da ist der Anfangsaufwand, weil grundlegende Strukturen und Prozesse geschaffen werden müssen. Ebenso werden Methodik und Tools eingeführt. Diese müssen anschliessend in Schulungen vermittelt werden. In der Folge muss jeder *Business Case* auch einen *People Case* beinhalten. Dieser muss den notwendigen Skillaufbau, -umbau und -abbau, sowie die damit verbundenen Kosten und Risiken klar aufzeigen.

Beurteilung Ist-Situation

Gerade der letztgenannte Punkt konnte von den Tools und Werkzeugen nicht zufriedenstellend beantwortet werden. Zumindest nicht, wie sie für einen Einsatz im Rahmen eines strategischen Workforce Management bislang zur Verfügung standen. Die zu leistenden manuellen Aufwände sind auf Dauer zu hoch. Selbst, wenn man bereit ist, entsprechende personelle Ressourcen für den Einsatz zur Verfügung zu stellen, dann ist das Gesamtsystem in vielen Fällen immer noch nicht performant genug. Jedenfalls was Hypothesenbildung und das Ableiten von Trade-off-Szenarien anbelangt.

6 Neue Digitale Assistenten für HR – Welche Rolle spielt dabei Cognitive Computing?

In diesem Artikel geht es um den Nutzen von *Cognitive Computing*, als eine Ausprägung von Systemen künstlicher Intelligenz. Wir betrachten Cognitive Computing jedoch nicht aus der Perspektive des Technikers (siehe z. B. (D’Onofrio und Portmann 2017; D’Onofrio et al. 2018)), sondern aus der Perspektive des Managers und seiner Aufgaben. Die Aufgabe, die für uns im Vordergrund steht, ist der Einsatz des Instruments, um die Effektivität der Wissensarbeiter in der Organisation zu erschliessen. Beim Einsatz von Cognitive Computing geht es im Kern um die Automatisierung dessen, was der Wissensarbeiter tut, um zu Ergebnissen zu kommen. Die Ergebnisse des Wissensarbeiters sind im Verknüpfen von Daten zu Informationen zu sehen. Dabei nutzt der Wissensarbeiter unterschiedlichste Datenquellen. Er nutzt Datenquellen, bei denen die Daten aus Datenbanken kommen. Doch er nutzt auch andere Quellen, bei denen diese Daten in unstrukturierter Art vorliegen. Er nutzt Datenquellen, die textbasiert sind, genauso wie Datenquellen, die Bildmaterial, Audio beziehungsweise Video zur Verfügung stellen. Diese Daten verknüpft der Wissensarbeiter im Rahmen der Wissensarbeit zu Informationen, die dabei helfen, die ihm gestellte Problemlösung zu finden.

Digitale Systeme sind heute bereits in der Lage beliebig grosse Datenmengen unter Anwendung von programmierten Algorithmen auszuwerten. Mit Cognitive Computing bin ich in der Lage, Wissensarbeit zu automatisieren; und zwar den Teil der Wissensarbeit, der wiederkehrende Routinen auf einer Masse von Daten abbildet. Weder bereitet diesen Systemen das Einlesen grosser Datenmengen Schwierigkeiten, noch

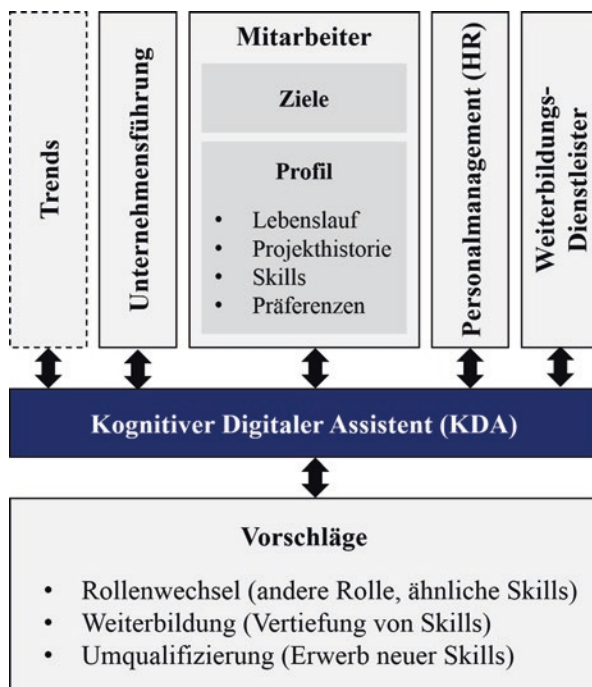


Abb. 2 Ein kognitiver, digitaler Assistent, der Vorschläge zu Rollenwechsel, Weiterbildung und Umqualifizierung bietet, unter Berücksichtigung von Trends am Markt, der Strategie der Unternehmensführung, den Erfahrungen und Talenten der Mitarbeiter

haben sie Probleme mit komplexen Berechnungen und liefern dazu Ergebnisse, die unsere menschlichen Fähigkeiten um ein Vielfaches übertreffen.

Bestehende Tools setzen dabei, wie zuvor auch für das strategische Workforce Management beschrieben, allerdings darauf, dass Daten in einer strukturierten Art und Weise vorliegen. Diese Daten werden für eine Auswertung in einer relationalen Datenbank hinterlegt. Der Schlüssel für die Erschliessung, der in der Datenbank abgelegten Informationen, ist das Datenbankschema. Die herkömmlichen Programme funktionieren nur, wenn sie auf Basis einer solchen Datenbank mit Daten versorgt werden. Ein kognitives System unterliegt dieser Restriktion nicht. Es ist in der Lage, sowohl mit strukturierten, das heisst in einer Datenbank hinterlegten Daten, als auch mit unstrukturierten Daten, wie zum Beispiel einem Textdokument, umzugehen. Dazu lernt das kognitive System in der Trainingsphase die Semantik der Domäne, in der es eingesetzt werden soll. Dadurch wird es in die Lage versetzt, die Daten, die relevant sind, entsprechend der Fragestellung des Wissensarbeiters zu verarbeiten und aufbereiten zu können. Egal ob diese Daten strukturiert, semi-strukturiert oder unstrukturiert vorliegen. Schliesslich ist ein kognitives System ein lernendes System. Je mehr immens grossen Datensätze es für seine Aufgabe zur Verfügung gestellt bekommt, desto besser.

Ein kognitives System (siehe z. B. Abb. 2) ist in der Lage in dem Datenrauschen die Muster zu identifizieren, auf die es dem Wissensarbeiter ankommt. Noch dazu

kann es neue Muster in diesen immens grossen Datensätzen identifizieren und finden, die mit den beschränkten Möglichkeiten des mentalen Systems des Wissensarbeiters nicht so offensichtlich erkennbar waren.

7 Wertschöpfende Verbindungen schaffen

Wie können nun die bestehenden Werkzeuge wertschöpfend über eine Cognitive-Computing-Plattform (siehe Abb. 3) verbunden werden? Im Fall der Digitalisierung des Skillmanagements sind das die Datenquellen, über die das Unternehmen bezüglich seiner Workforce bereits verfügt. Angereichert um die Daten aus weiteren Datenquellen. All das, um die Wissensbasis aufzubauen, die zur Beantwortung der Fragestellungen im Rahmen der digitalen Transformation der Workforce notwendig sind.

Die kognitive Plattform (siehe Abb. 3) kann, Zug um Zug, um diese zusätzliche Datenquellen ergänzt werden. Über jede, von den Domänenexperten identifizierte und integrierte Datenquelle, wird die kognitive Plattform mit ihren Empfehlungen und Prognosen besser, da es sich um ein lernendes System handelt.

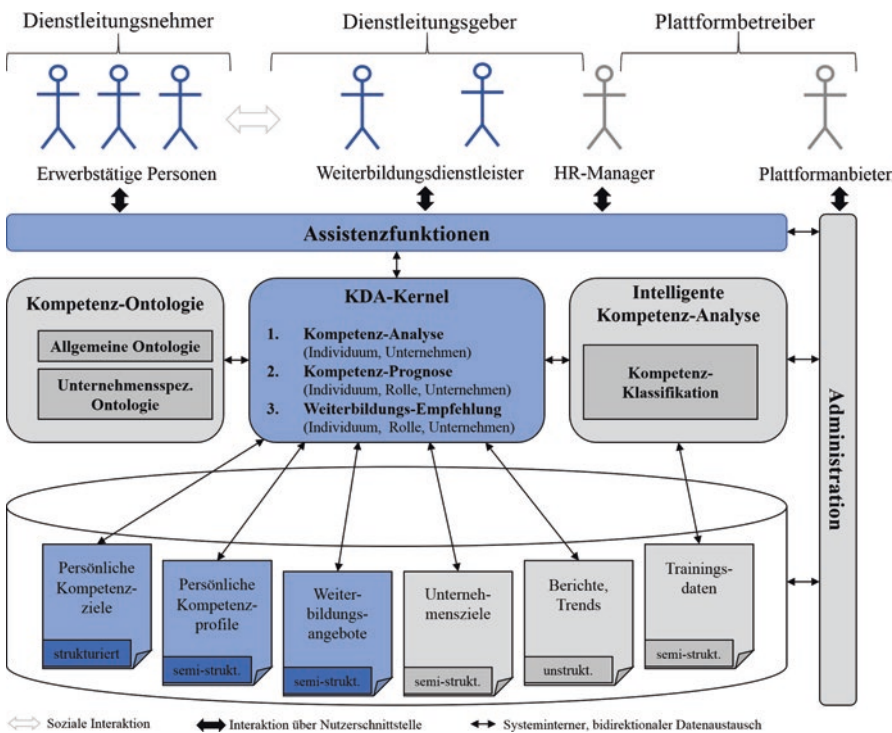


Abb. 3 Die Plattform, auf der die kognitiven, digitalen Assistenten für das Skillmanagement zur Verfügung gestellt werden können

Transformationsszenarien vorab durchspielen

Die Cognitive-Computing-basierte Plattform ist dann in der Lage in unterschiedlichster Transformationsszenarien wertschöpfend eingesetzt zu werden, um zum Beispiel die, mit dem jeweiligen Szenario verbundene Trade-offs zu visualisieren. Hier einige mögliche Fragestellungen, auf die die kognitiven, digitalen Assistenten einer solchen Plattform trainiert werden können:

- Welche Skills werden für die digitale Transformation des Unternehmens in welcher Quantität und in welcher Qualität gebraucht?
- Wie kann die digitale Transformation des Unternehmens in der kürzesten Zeit erfolgen?
- Wie kann die digitale Transformation des Unternehmens mit den niedrigsten damit verbundenen Kosten erfolgen?
- Wie kann die digitale Transformation des Unternehmens unter Priorisierung der Um- und Weiterbildung der bestehenden Workforce erfolgen?

8 Notwendige Kognitiv-Digitale Assistenzfunktionen

Prinzipiell müssen die digitalen Assistenten dabei drei Grundfunktionalitäten zur Verfügung stellen, die nachfolgend skizziert werden:

Leitfrage: Über welche Kompetenzen verfügt die Workforce aktuell?

Basierend, auf im Unternehmen verfügbaren Angaben klassifiziert das System, mittels der intelligenten Kompetenz-Analyse die Kompetenzen (das sog. *Unternehmens-Skill-Set*). Verfügbare Angaben könnten beispielsweise folgende sein:

- zu den aktuellen Rollen
- zu Lebensläufen
- Projekthistorien und Skills
- gegebenenfalls auch speziell erarbeiteten psychologischen Fragebögen zu Soft Skills

Dies erfolgt durch die Zuweisung von Skills aus der Kompetenz-Ontologie. Dieses, mittels des Verfahrens des überwachten maschinellen Lernens (d. h. vordefinierte Input-Outputpaare), die auf Trainingsdaten aufbauen. Diese Analyse kann bei Bedarf auch auf Rollen oder Personen heruntergebrochen werden, um die entsprechenden Rollen- beziehungsweise Personen-Skillsets zu bearbeiten.

Leitfrage: Passen die aktuellen Skills in meinem Unternehmen zur geplanten digitalen Transformation und dem angestrebten digitalen Ziel-Geschäftsmodell?

Auf Basis der Unternehmensziele und der in der Kompetenz-Ontologie hinterlegten Information zur Veränderung von Kompetenzen wird prognostiziert, wie sich ein Skill-Set im Laufe der Zeit verändern wird. Veränderungen können dabei die Zu- oder Abnahme der Bedeutung von Kompetenzen sein. Aber auch das Hinzutreten neuer, erforderlicher Kompetenzen.

Eine Marktsicht erlaubt, den Ressourcenbedarf eines Unternehmens zu prognostizieren. So kann die jeweilige HR-Strategie weiterentwickelt werden. Die Kompetenz-Prognose kann damit auf der jeweils interessierenden Ebene (für einen einzelnen Mitarbeiter, in Bezug auf Rollen oder auf Unternehmen) angewendet werden.

Die Definition einer Rolle reduziert sich nicht nur auf einen Namen, sondern beinhaltet eine Summe der bekannten notwendigen Skills und Fähigkeiten, die erforderlich sind, um eine Rolle erfolgreich verkörpern zu können. Sobald die Suche nach neuen Rollen auf Basis von Skills, Fähigkeiten und Fertigkeiten gemacht wird, wird das Suchergebnis nicht mehr binär – es passt oder es passt nicht. Dadurch wird man in die Lage versetzt, Abdeckungsgrade zu einer gewünschten Position darstellen und mögliche Gaps (Lücken) hervorheben zu können.

Leitfrage: Welche Um- oder Weiterbildung der bestehenden Workforce muss auf Basis der Kompetenz-Prognose und im Abgleich mit den persönlichen Zielen und Präferenzen der Beteiligten für die Umsetzung der Transformation geplant werden?

Basierend auf der Kompetenz-Analyse und -Prognose, Markttrends, den persönlichen Kompetenzzielen und Präferenzen der Beteiligten und im Abgleich mit Weiterbildungsangeboten werden Szenarien entwickelt und entsprechende Transitions-pfade berechnet.

9 Ausbau der Plattform

Die fortlaufende Aktualisierung der Kompetenzen, aber auch der Vorlieben und Präferenzen der Mitarbeiter ist für eine zielgerichtete Weiterentwicklung oder Einarbeitung neuer Mitarbeiter erforderlich. In der Praxis gestaltet sich das jedoch schwierig. Ein Grund hierbei ist, neben der Komplexität des Skillmanagements durch Werkzeuge, die fehlende Zeit im Tagesgeschäft und die Auslastung der HR-Abteilungen. Eine individualisierte Beratung kann nicht vollumfänglich geleistet werden. Es ist aus diesem Grunde erstens eine stärkere Automatisierung erforderlich; zweitens, die Verbesserung der Motivation der Mitarbeiter und Aktualität ihrer Profile zu halten; und drittens müssen die Werkzeuge hierzu leicht benutzbar sein. Nur mit aktuellen Profilen können zielgerichtete Mitarbeitergespräche initiiert und durchgeführt werden.

Vorschläge für die Um- und Weiterbildung (oder Vertiefung) mittels intern oder extern erbrachter Weiterbildungsdienstleistungen können unter Beachtung aller relevanten Faktoren generiert werden. Dazu zählen, neben der Unternehmensstrategie, die aktuelle und zukünftige Marktsituation sowie die Bedürfnisse, Präferenzen, Ziele (z. B. Qualifikationsziele, Lebenssituation) und Persönlichkeit des Mitarbeiters. Bei einer derart hohen Anzahl von Einflussfaktoren ist eine IT-Unterstützung zur Generierung dieser Vorschläge dringend erforderlich, um jedem Mitarbeiter die bestmögliche individuelle Weiterentwicklung zu bieten. Dies ist insbesondere auch im Hinblick auf die Gewinnung und Bindung hochqualifizierter Mitarbeiter von Bedeutung.

In der aktuellen Praxis existieren viele Insellösungen, mit denen teils redundante Daten verwaltet werden. Zum Beispiel Kompetenzprofile, Informationen zu Weiterbildungsangeboten, persönliche CV-Daten und Protokolle zu Status- und Bewerbungsgesprächen. Durch den Aufbau einer solchen Plattform ergeben sich natürlich nicht nur für den Bereich HR, sondern auch für die anderen Beteiligten an und auf dieser Plattform Nutzenpotentiale, was durch diese Aufzählung sicher deutlich geworden ist.

10 Minimum Viable Product: Workforce-Transformations-Assistent

Im Rahmen der Digitalisierung ändern sich die Geschäftsmodelle und Produkte in nahezu allen Domänen. Diese Änderung hat eine fundamentale Auswirkung bezüglich der Anforderungen an die Workforce der Unternehmen. Bestimmte Kernkompetenzen fallen weg, andere Kompetenzen entwickeln sich zu Kernkompetenzen oder es werden plötzlich ganz neue Kernkompetenzen (d. h. digitale Fähigkeiten und Fertigkeiten) benötigt. Die Daten, die für eine strategische und ganzheitliche Sicht auf die Workforce des Unternehmens notwendig sind, liegen verteilt im Unternehmen in einer Vielzahl von Quellen. Dabei haben wir es mit strukturierten und unstrukturierten Daten zu tun, die bislang nur manuell aufwendig einander verknüpft und für das jeweils zu diskutierende Szenario vom Personalmanagement (zusammen mit dem jeweiligen Business Partner und der Finanzplanung) erstellt wurde. Selbst wenn Sie in Ihrem Unternehmen über einen entsprechenden manuellen Prozess der Zusammenführung dieser verteilten Daten zu verwertbaren Informationen verfügen, so ist ein solcher Prozess unter dem Gesichtspunkt der Wettbewerbsfähigkeit stark optimierungsbedürftig.

Ein massgeblicher Treiber für alle Unternehmen zum aktuellen Zeitpunkt ist die Digitalisierung. Jedes Unternehmen, das sich durch die Herausforderungen der Digitalisierung zu einer digitalen Transformation entschliesst, sei es auf Ebene der Produkte, der Dienstleistungen oder Geschäftsprozesse, sieht sich vor die Herausforderung gestellt, dass es zu einem möglicherweise entscheidenden Skillshift innerhalb der Workforce des Unternehmens kommt. In den Unternehmen ist dann insbesondere die HR-Abteilungen betroffen, die in bezüglich des Themas Personal die Federführung besitzt.

Wir müssen in unserem Datenkorpus Informationen zusammenstellen, die die Anforderungen beschreiben, welche im Rahmen der Digitalisierung in einem jeden Unternehmen entstehen. Diese Erhebung müssen wir im Rahmen der Erstellung des Datenkorpus an den Anfang stellen. Was sind die notwendigen Skills einer Workforce, die erfolgreich im Kontext der Digitalisierung unterwegs ist? Diese Informationen können wir als Grundlage für die weiteren Arbeiten nehmen, egal mit welcher fachlichen Domäne wir uns in der Folge beschäftigen.

Natürlich werden sich die Anforderungen bezüglich des Themas Digitalisierung dann Domäne für Domäne unterscheiden. Dabei geht es jedoch dann eher darum,

herauszufinden, an welcher Stelle und in welchem Umfang im Rahmen einer fachlichen Domäne durch die Digitalisierung entsprechende Änderungen hervorgerufen werden. Wir müssen uns für unser Projekt und den Gesamtdurchstich überlegen, welche fachliche Domäne wir in unserem Datenkorpus dann ergänzen wollen. Genau an dieser Schnittstelle ist der nachfolgend beschriebene *Workforce-Transformations-Assistent* platziert. Im Rahmen der Digitalisierung werden die für die Entscheidungsfindung massgeblichen Daten im Rahmen des *Workforce-Transformation-Datenkorpus* zusammengeführt und der Aufbau der entsprechenden Informationsbasis wird bereits in dieser Phase durch Mechanismen der kognitiven Plattform als lernendes System von manuell auf halb automatisiert und schliesslich auf automatisiert umgestellt.

Der manuelle Prozess der in aufwendiger Abstimmung zwischen den einzelnen Unternehmensbeteiligten für ein konkretes Szenario erfolgte, wird (wenn er noch nicht vorliegt) dokumentiert und digitalisiert.

11 Der Anwendungsfall

Wenden wir uns nun einem Beispiel zu, das viele der Dinge, die wir in den Abschnitten zuvor hergeleitet haben, aufgreift und umsetzt, und das als *Minimum Viable Product* unter folgender Internet-Adresse verfügbar ist: <http://cm.stratworx.de/live-ausprobieren.html>.

Der zugrundeliegende Anwendungsfall, der in diesem Prototyp rudimentär umgesetzt ist, stammt aus dem Kontext des strategischen Workforce Management und befasst sich mit der Transformation der Workforce aus einem bestehenden Prozess in einen neuen Prozess. Für das erste Szenario haben wir eine Domäne ausgewählt, die bezüglich der Digitalisierung und der entsprechenden Auswirkungen bereits gut erschlossen ist (z. B. digitale Fotografie, Verlagswesen).

Um den entsprechenden Datenkorpus aufzubauen, muss die Fragestellung, die von dem Workforce-Transformations-Assistenten beantwortet werden soll, klar sein. Denn über die Fragestellung erschliesst sich, welche Informationen ich im Datenkorpus verknüpft vorhalten muss, um entsprechende Antworten geben zu können. Für den Kontext unseres Beispiels lautet diese Fragestellungen insofern:

- Unterstützen die Fähigkeiten, Fertigkeiten und Skills der gesamten Workforce den Wechsel auf ein neues digitalisiertes Geschäftsmodell / Produkt?
- Welche Skills sind bereits im Unternehmen verfügbar?
- Welche Skills sind im Unternehmen bereits an der richtigen Stelle?
- Welche Verlagerung von Skills muss durchgeführt werden?
- Welche Aufwände sind mit der Umstellung intern und extern verbunden?
- An welcher Stelle muss die Workforce durch externe Ressourcen ergänzt werden?
- Welche Optionen stehen zur Verfügung, wenn man den Cluster von Skills betrachtet, die für das neue Produkt möglicherweise nicht mehr benötigt werden?

Ein Unternehmen wird eher nicht von heute auf morgen komplett auf ein neues, digitalisiertes Geschäftsmodell oder Produkt umsteigen. Das heisst, verbunden mit der einen Frage, ob der Einstieg in ein neues digitalisiertes Geschäftsmodell mit der bestehenden Belegschaft möglich ist, ist dann auch die Fragestellung, was das für Konsequenzen bezüglich des bestehenden Produkts hat.

Das heisst, bezüglich jeder Alternative, was das neue Produkt anbelangt, ist auch ein entsprechendes Szenario für die bestehende Produktpalette auszuarbeiten.

- Wie lange kann ich in der bestehenden Produktpalette, die mir zum momentanen Zeitpunkt meine Einkünfte sicher noch verbleiben?
- Welche Bewegungen innerhalb des Skillsets der bestehenden Produktpalette muss ich in Abhängigkeit von der Bewegung in Richtung des neuen Skillsets auf den Weg bringen, um beide Welten für einen bestimmten Zeitraum parallel zu bedienen?
- Zu welchem Zeitpunkt muss ich mich komplett von der alten Produktpalette lösen?

11.1 Beispiel: Fotografie

Für unser erstes Beispiel greifen wir zurück auf das Szenario aus dem Bereich der Fotografie, wie es wohl den meisten allein durch das KODAK-Beispiel bereits bekannt ist. Die Datengrundlagen zu Berufen und Fertigkeiten bildet in unserem Beispiel das ESCO-Projekt der EU in der Version 1.0.⁸

Szenario 1: Von der analogen Fotografie (Kernkompetenz Chemie) zur digitalen Fotografie (Substitution des Kernkompetenz):

Schritt 1: Die analoge Fotografie basierte zu einem wesentlichen Anteil auf den Kernkompetenzen im Bereich der Chemie. Die bestehende Workforce mit den entsprechenden Schwerpunkten im Bereich der Chemie kann in Schritt 1 des digitalen Assistenten erfasst werden. Dazu fragt der Prototyp auf der ersten Seite die Rollen und Fertigkeiten, sowie die Mengengerüste zum bestehenden Prozess ab (Abb. 4).

Schritt 2: Nun werden die Fähigkeiten und Fertigkeiten erfasst, die für den neuen, angestrebten Prozess benötigt werden. In unserem Beispiel bleiben wir bewusst für den neuen Prozess auf der Ebene der Fertigkeiten. Das Aggregieren der Fähigkeiten und Fertigkeiten zu neuen Rollen ist ein separater, optionaler Schritt, der an dieser Stelle im Prototyp noch nicht enthalten ist. Wir fügen also für das Szenario 1 Fertigkeiten im Kontext der digitalen Fotografie hinzu. Dabei geben wir auch an, wie viel wir in einem optimalen Fall von dieser Fertigkeit benötigen werden und wie viel wir, zum Beispiel, um damit zu beginnen, mindestens von einer bestimmten Fertigkeit brauchen (Abb. 5).⁹

⁸Mehr zu ESCO finden Sie hier: <https://ec.europa.eu/esco/portal>.

⁹Hinweis: Diese Informationen können zum Beispiel im Rahmen der Prozessmodellierung des neuen Prozesses gewonnen werden.

Schritt 1

Lade

Rolle

Skills

+

Abteilung

Rolle	Skills	Nummer
Laborleiter/in (chemisches Labor)	Arbeitsplanung (Daten)	50
Chemiker/in	Analytische Chemie	450

Abb. 4 Szenario 1 – Schritt 1

Schritt 2

Speichern

Der NEUE Prozess

Skills

+

Abteilung

Name	Ziel	Minimum	
Fotografie (Digitalkamera)	50	10	✓✕
Datenanalyse	20	2	✓✕
Arbeitsplanung (Daten)	10	1	✓✕

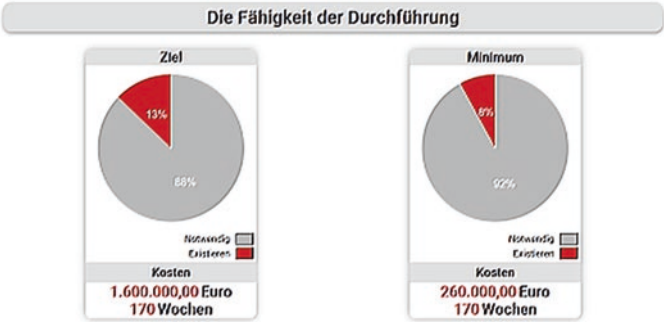
Zurück

Weiter

Abb. 5 Szenario 1 – Schritt 2

Schritt 3: Der digitale Assistent zeigt dann die Kosten an, die anfallen, um die bestehende Workforce, basierend auf den Mengengerüsten, die in Schritt 2 erfasst wurden, zu qualifizieren. Die Informationen, die unter Auswertung entsprechender Schulungsmassnahmen nach Zeitdauer und Kosten ermittelt werden, beziehen noch nicht die Kenntnisse ein, über die die Workforce bereits verfügt. Diese Transformation leistet dann Schritt 4 (Abb. 6).

Schritt 3

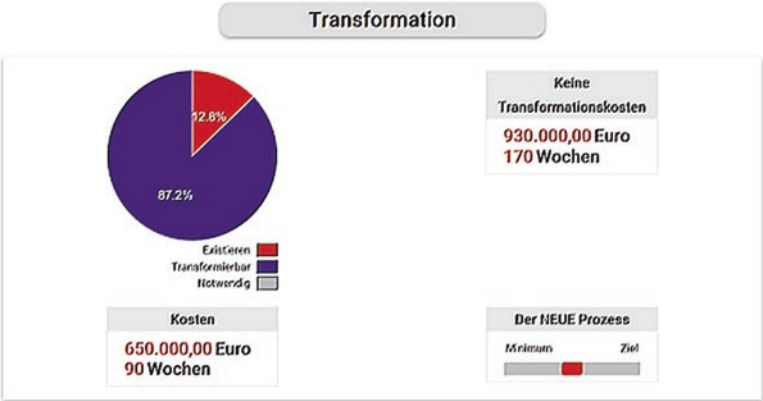


Zurück

Transformation

Abb. 6 Szenario 1 – Schritt 3

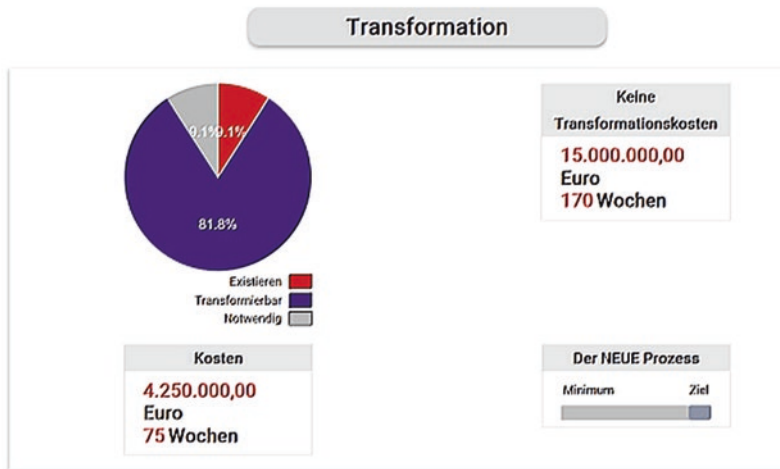
Schritt 4



Zurück

Abb. 7 Szenario 1 – Schritt 4

Schritt 4: Die Fähigkeiten und Fertigkeiten der bestehenden Workforce werden überprüft und mit den benötigten Fähigkeiten und Fertigkeiten im Rahmen des neuen Prozesses abgeglichen. Wenn Fähigkeiten und Fertigkeiten der Workforce verfügbar sind, auf denen im Rahmen der Transformation aufgebaut werden kann, so wird das in diesem Schritt geleistet. Insofern sind die im Rahmen der Transformation der bestehenden Workforce ermittelten Daten für Kosten und Qualifizierungsdauer kleiner gleich den in Schritt 3 ermittelten Kosten. Das so dokumentierte Szenario lässt sich bereits mit dem aktuellen Stand des Prototyps speichern (Abb. 7).

Schritt 4**Abb. 11** Szenario 2 – Schritt 4

Transformationsoptionen ist genau das jedoch hilfreich. Denn so werden die einzelnen Auswertungsstufen des Prototyps zu diesem ersten Anwendungsfall deutlicher herausgearbeitet.

11.2 Beispiel: Automobilindustrie

In einem weiteren Beispiel wenden wir uns einem ähnlichen Substitutionsprozess zu, den es bezüglich Kernkompetenzen im Bereich der Automobilindustrie geben wird beziehungsweise schon gibt. Mit der Substitution eines herkömmlichen Benzin- beziehungsweise Dieselmotor durch einen Elektromotor fallen die Kernkompetenzen weg, die von einem grossen Anteil der Workforce der Unternehmen dieser Industrie abgedeckt werden.

- Wie sieht die direkte Transformation der Workforce hin zur neuen Kernkompetenz, der Erstellung von Elektromotoren, aus?
- In welchen Schritten kann eine solche Transformation der Workforce erfolgen?
- Mit welchen Kosten ist eine solche Transformation der Workforce verbunden?
- In welcher Reihenfolge lassen sich welche Anteile der Workforce hin zu den neuen Fähigkeiten und Fertigkeiten vollziehen?

- Welche Szenarien lassen sich noch bilden, wenn ich zum Beispiel an Produkte denke, die den bestehenden Motor nicht durch einen neuen Motor ersetzen, sondern ihn ergänzen (Hybridmotor)?

Schritt 1 (Abb. 12):

Schritt 1

Laden

Rolle

Skills

+

Abteilung	Rolle	Skills	Nummer	
	Chefingenieur/in	Arbeitsplanung (Daten), Benzinmotor (Überholung)	10	 
	Ingenieur/in für Mechanik/Elektronik	Benzinmotor (Überholung)	100	 

Weiter

Abb. 12 Szenario 3 – Schritt 1

Schritt 2 (Abb. 13):

Schritt 2

Speichern

Der NEUE Prozess

Skills

+

Abteilung	Name	Ziel	Minimum	
	Digitale Elektronik	100	20	 
	Arbeitsplanung (Daten)	10	2	 
	Elektromotoren (Überholung)	20	2	 

Abb. 13 Szenario 3 – Schritt 2

Schritt 3 (Abb. 14):

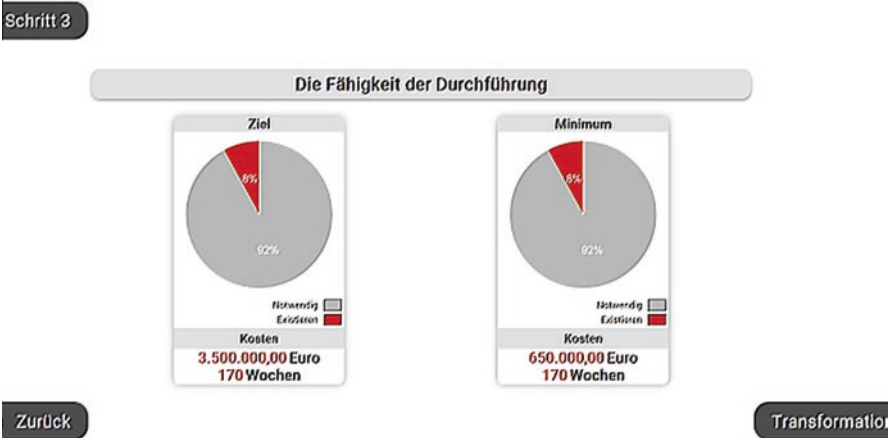


Abb. 14 Szenario 3 – Schritt 3

Schritt 4 (Abb. 15):

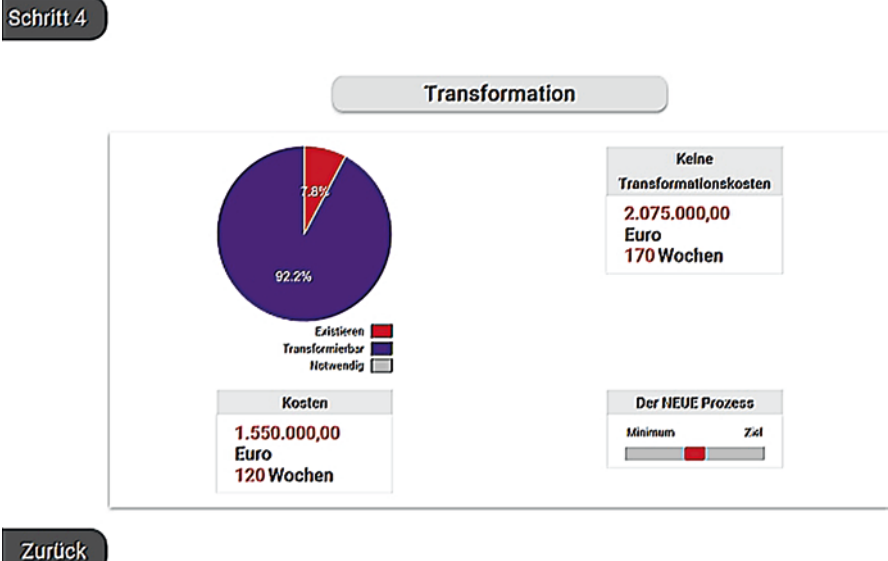


Abb. 15 Szenario 3 – Schritt 4

Auf der Grundlage dieser beiden Beispiele sollte deutlich geworden sein, dass sich der erstellte minimale Prototyp mit der bereits implementierten Anwendungslogik in einer jeden Industrie, die einen solchen Wandel, eine Transformation durchmacht, zur Anwendung bringen lässt. Industrie- und unternehmensspezifische Daten speisen dabei die grundsätzlichen Algorithmen. Die Ermittlung von Fähigkeiten und Fertigkeiten, sowie entsprechender Abhängigkeiten zwischen Fähigkeiten und Fertigkeiten, beruhen auf dem standardisierten Framework ESCO Version 1.0 der EU.

Der zweite Schwerpunkt, der zum Aufbau des entsprechenden HR-spezifischen Datenkorpus berücksichtigt wird, bezieht sich auf die digitalen Fähigkeiten und Fertigkeiten, die ein Unternehmen, gleich welcher Industrie es angehört, benötigt, um die digitale Transformation von Prozessen und Produkten durchzuführen. Für die Strukturierung der Fertigkeiten und Fähigkeiten bezieht das Beispiel dabei Ergebnisse aus anderen nationalen und auch internationalen Projekten mit ein.

12 Zusammenfassung

Diese Ausarbeitung hat sich konzentriert auf den Nutzen einer Cognitive-Computing-Plattform als Instrument im Kontext des strategischen Workforce Management. Der Artikel hat sich im Schwerpunkt mit der Erschliessung des Nutzens eines bestehenden strategischen Workforce Management-Ansatz durch den nun möglich gewordenen Einsatz Cognitive Computing als einer Ausprägung für ein System mit künstlicher Intelligenz beschäftigt. Dabei ist das grundsätzliche Nutzenpotential, das eine Cognitive-Computing-Plattform für den Bereich Personalmanagement bietet, von einem vorhandenen Konzept zum strategischen Workforce-Management durchaus unabhängig. Um dieses Nutzenpotential zu erschliessen, müsse man nicht erst ein solches Konzept erstellen. Doch wenn man im Unternehmen bereits über ein solches Konzept verfügt, dann ist das ein Eingangsartefakt, das ideal von der aufzubauenden Cognitive-Computing-Plattform verwertet werden kann. Da es sich bei dieser Plattform um ein lernendes System handelt, können Schritt für Schritt angrenzende Einsatzfelder erschlossen werden.

Literatur

- D'Onofrio S, Portmann E (2017) Cognitive Computing in Smart Cities. Informatik Spektrum, 40(1):46–57. Berlin Heidelberg: Springer-Verlag.
- D'Onofrio S, Portmann E, Franzelli M, Bürki C (2018) Cognitive Computing – Theoretische Grundlagen und Praxisbeispiele der Schweizerischen Post. Informatik Spektrum, 41(2):113–122, Berlin Heidelberg: Springer-Verlag.
- Frey T (o. J.) Future Jobs, Future Skills. Futurist Speaker Thomas Frey. <http://www.futuristspeaker.com/business-trends/162-future-jobs-the-video>. Erstellt: n.d. Letzter Zugriff: 17.06.2017.
- Frey T (2014) 101 Endangered Jobs by 2030. Futurist Speaker Thomas Frey. <https://futuristspeaker.com/business-trends/101-endangered-jobs-by-2030>. Erstellt: 07.11.2014. Letzter Zugriff: 17.06.2017.
- Palla R (2014) Verschwundene Arbeit: das Buch der untergegangenen Berufe. Wien: Brandstätter.

Spiegel Online (n.d.) Wirtschaftsgeschichte: Als Fischbeinreisser ein Top-Job war. Spiegel Online. <http://www.spiegel.de/wirtschaft/alte-berufe-der-fischbeinreisser-zerkleinerte-die-barten-von-walen-a-835788.html>. Erstellt: n.d. Letzter Zugriff: 17.06.2017.

Vogler A (2018) Die Digitalisierung des Skillmanagements – Digitale kognitive Assistenten für die Wertschöpfung des Human Ressource Managements in der Digitalen Transformation. Informatik-Spektrum, 41(2):97–104.



Health Recommendation System Framework for the Optimization of Medical Decisions

Luis Terán, Jhonny Pincay, Diana Pacheco, Martin Štěpnička and Daniel Simancas-Racines

Abstract

The introduction of recommendation systems for the health sector has opened additional channels to improve the decision-making of physicians and administrators of the healthcare system. Web-based and mobile applications, social networks and virtual communities could improve these processes. The Internet provides a huge amount of information; therefore, the creation of intelligent mechanisms and tools, such as the ones proposed, allows physicians and the health sector to be better informed in the decision-making process while improving communication channels. In this work, a theoretical framework for the design and implementation of the so-called project *MedicGate*, a smart social network to optimize medical decisions using cognitive computing is presented. The

Vollständig neuer Beitrag

L. Terán (✉)

Universidad de las Fuerzas Armadas ESPE, Sangolquí, Ecuador

Human-IST Institute, University of Fribourg, Fribourg, Switzerland

E-Mail: lfieran1@espe.edu.ec; luis.teran@unifr.ch

J. Pincay

Human-IST Institute, University of Fribourg, Fribourg, Switzerland

E-Mail: jhonny.pincaynieves@unifr.ch

D. Pacheco

Swiss Paraplegic Research, Nottwil, Switzerland

E-Mail: diana.pachecobarzalto@paraplegie.ch

M. Štěpnička

CE IT4Innovations, IRAFM, University of Ostrava, Ostrava, Czech Republic

E-Mail: martin.stepnicka@osu.cz

D. Simancas-Racines

Universidad Tecnológica Equinoccial, Quito, Ecuador

E-Mail: dsimancas@ute.edu.ec

expectation is that this initiative will have an impact on those responsible for the generation of health policies, by providing information and resources for decision-making through informed knowledge and feedback from the medical and academic sector.

1 Introduction

Health information systems refer to systems that capture, store, manage and/or transmit information related to the health of patients or organizations within the health sector. Huge amounts of data are collected in clinical databases, such as test results, treatments, and reports. Patient-oriented decision-making processes have increased considerably with the rise of electronic services (*eServices*) for both public and private sectors. These services adapted to the field of health are often distributed through different websites, thus increasing their complexity in providing information that could be of great importance to physicians, patients and medical researchers. In this context, health recommender systems are presented as intelligent and complementary tools for the decision-making processes.

In this project, new mechanisms for the development of smart health systems based on cognitive computing, graph theory, and fuzzy logic will be addressed. The term cognitive computing generally refers to hardware and/or software that mimics the functioning of the human brain to improve human decision-making and are part of the scientific disciplines of artificial intelligence including machine learning, recommender systems, reasoning, natural language processing, human–computer interaction, dialog and narrative generation, among other technologies.

In this project, cognitive computing is used to support the integration given by physicians to optimize medical decisions. The expected output of this project is the introduction of a smart social network to enhance decision-making and discussion of medical doctors and researchers that could be used as a vehicle for collecting, gathering, and analyzing data while taking the local contexts into consideration. This platform will be put at the service of the society and research communities in the fields of recommender systems' design, cognitive computing, science of organization, social networks, social behavior, and social memory. This project is expected to face the following challenges

- the analysis of large heterogeneous information networks,
- the measurement of influence,
- calculation of similarities and classification,
- detection of dependence on latent concepts and relationships,
- use and extraction of the schemes of networks, and
- approaches for guided consultation to the user given the complexity of health information systems.

The use of patient-dynamic profiles will be developed in this project. Dynamic profiles are created based on the activity with the physicians, as well as the context of the patients. The proposed dynamic profiles will include a series of modules including *context-awareness* (i.e., location data), *user interaction* (i.e., interaction of physicians with the system), and *privacy settings* (i.e., level of disclosure of each user profile). The data will be collected from different sources, such as patient data, interactions with the system (comments, ratings, posts, etc.), and the knowledge base provided by the *Cochrane Network*.¹

The system will have three types of users: medical students, physicians, and medical researchers. Physicians will feed the system with relevant information from patients manually and using an interface, maintaining the privacy of personal data (privacy settings will be handled by the system to avoid exposure of personal data), so that they can make inquiries directed to the medical and academic sector, to improve decision-making and to tackle problems of overtreatment and undertreatment. The types of inquiries that physicians can perform will depend on the needs of the user, such as similar cases, experts in the field, and up to date evidence-based medical recommendations, among others.

Researchers will enrich the knowledge base with state-of-the-art information, recent discoveries and treatments. The proposed system will learn from the profiles of users, their interactions, but also their activities. This tool will also serve for teaching in the health area, as a channel that allows the generation of consultancy for medical students, physicians, and researchers.

This article is structured as follows: The problem statement is introduced in Sect. 2. Sect. 3 presents the state-of-the-art on health recommender systems. Then, the theoretical background is described in Sect. 4. The main concepts for the implementation are presented in the theoretical framework, in Sect. 5. Finally, concluding remarks and the outlook is presented in Sect. 6.

2 Problem Statement

Too much testing of well people and not enough care for the sick worsens health inequalities and drains professionalism, harming both those who need treatment and those who don't. (McCartney 2013)

The following section explains why some markets fail to achieve optimal outcomes and require external interventions to improve their results. The healthcare market is no exception; indeed, the asymmetry of information is more harmful than in other markets and can result in suboptimal care.

¹<https://www.cochrane.org/>

2.1 Free Markets and Market Failures

Free markets are assumed to work under perfect competition, meaning that every consumer gets what he/she needs, and every buyer sells what he/she has. An efficient market exists when there is no waste of resources. Nevertheless, perfect competition is the exception rather than the rule. In fact, for perfect competition to exist the next conditions need to hold (Pindyck and Rubinfeld 2013):

1. Perfect information, where both parts supply and demand know the characteristics of the goods/services they are selling/buying. In this way, consumers can make informed decisions.
2. The market has no barriers concerning entry/exit, meaning that anyone willing to provide a good/service can enter the market.
3. There are many buyers and sellers, and the offered product is similar among suppliers.

In the real world, however, many markets fail in one, two, or all of these conditions, asymmetric information being the most common market characteristic. This means, that even if there are free markets, transactions not always result in efficient outcomes, as some of the parts have more information than other parts. In fact, in some markets asymmetric information can be so problematic that the transaction does not take place. This is known as a market failure, which requires external intervention for the transaction to take place (Pindyck and Rubinfeld 2013).

Akerlof (1978) was the first author to formalize the analysis of markets beyond the assumption of perfect competition. He showed that even in simple markets, asymmetry of information could lead to adverse selection. He used the market of used cars in the US to illustrate market failures. Akerlof explained that because the car seller has more information about the quality of the car, the market will be left only with bad-quality cars (Akerlof 1978).

In a market where you can find good-quality cars and bad-quality cars, it is very unlikely that buyers and sellers agree on a price. As the cars in the market are heterogeneous, and only the owner of the car knows about the car's quality, it is very likely that no transaction may occur. The buyer will always offer a lower price since he/she is taking into account the probability of getting a bad-quality car. The owner of a good-quality car will never sell at a low price. As a result, sellers of good-quality cars will leave the market, as no one is willing to pay a fair price.² Thus, the market will be left with the bad-quality cars only.

²In economic theory, a "fair price" is the price at which both the buyer and the seller are happy with the transaction. Even though it is a subjective notion, in a transaction both parts are clear about the price they are willing to accept. With no agreement in the price (e.g., too high/too low) the transaction will not take place.

2.2 The Healthcare Market

Akerlof's theory can be applied to the healthcare market, where information gaps and asymmetries exist and are even more serious than in other markets. When seeking for treatment, for example, the patient knows less than the physician about the treatment, the expected outcomes and relative prices. At the same time, the physician faces uncertainties, because the patient may not tell all-important information, or some patients may ask for unnecessary treatments. The asymmetry of information can result in too much treatment (*overtreatment*) or in too little treatment (*undertreatment*).

In an ideal world, where information is not a problem, a physician prescribes only what a patient needs. Nevertheless, in the real world, where information is far from perfect, the objective is to minimize errors in a given treatment, which can be too costly in monetary and non-monetary terms.

2.2.1 The Agency Problem

Agency theory is related to the relationship that arises, when a principal delegates authority to an agent due to asymmetry of information. The principal delegates because he/she does not have enough information to make a decision.

The healthcare market is the best example of the agency problem. As the patient is unable to make a decision about a treatment, so he/she asks the physician to decide on his/her health status. The agency problem arises, because the physician is the one who recommends the treatment and who is the provider of the service.

This patient-physician relationship may raise some conflict of interests, as the physician may not only prescribe a treatment, because it is in the best interest of the patient, but also because it is in his/her own interest to influence the patient (i.e., profit expectations). Even if the physician is willing to act in the best interest of the patient, it is rather difficult, as the physician has to be fully aware of the patient's preferences and needs.

To solve the agency problem, the *economic theory* suggests creating a contract, which maximizes the utility of the principal and the agent. The contract should setup a payment system that motivates the physician to maximize the patients' utility, where the guiding principles should be the best possible use and allocation of resources to guarantee the best treatment (Cerović et al. 2012).

Aware of the agency problem, healthcare systems around the world apply different payment schemes designed to optimize the system's performance. There exist different types of contracts such as the fee-for services, capitation, salary, global budget and diagnosis-related groups (OECD 2016). Depending on the payment scheme, the incentives for the physician/patient can influence the *amount* of a treatment and can lead to undertreatment or overtreatment situations.³

³ Undertreatment is defined as to inadequately treated or not frequently enough (Dayal and Hort 2015). Overtreatment refers to unnecessary medical treatment (Brownlee 2010).

2.3 What Is the Empirical Evidence?

Overtreatment and undertreatment are suboptimal care result of asymmetric information. Nevertheless, documenting them is challenging, as it is difficult to define *optimal* care.

In the case of overtreatment, empirical evidence has used two methods to document it: one, by analyzing individual records, and the second by comparing different areas to detect geographical variation in the use of services or treatments. Depending on the type of treatment and the analysis, estimates of overtreatment vary considerably in the literature going from 1 % to 80 % (Brownlee et al. 2017). The most common studies have focused on the overuse of antibiotics (Van Boeckel et al. 2014), screening tests (Ahn et al. 2014), diagnostic tests (Keyhani et al. 2013), therapeutic procedures (Schwartz et al. 2014), intensity of care (McDonagh et al. 2000), and end-of-life care (Bekelman et al. 2016).

In low-income countries, even if the rates of overtreatment are significantly lower than in rich countries, considering the economic development, the trend of use and overuse of certain medicines is concerning. There exist some examples where, even when the medicine has been proven ineffective, some health systems still provide them as part of a treatment. For example, Bevacizumab, which is proven ineffective for breast cancer in the US and in the UK, is still prescribed in some developing countries (Brownlee et al. 2017).

As for undertreatment, which sometimes coexists with overtreatment, especially in low-income countries, measuring is more complex, as it is necessary to estimate the number of people who would need treatment, but never got it. The reasons for undertreatment could be related to financial issues of the patient, access to health-care services and strict medical guidelines. Extensive evidence exists on undertreatment of pain (Deandrea et al. 2008; Shen et al. 2017), asthma (Braman 1993; Gana and Fitzgerald 2018; Vasilopoulou et al. 2015), hypertension (van Breukelen-van der Stoep et al. 2016), and depression (Davidson and Meltzer-Brody 1999; Morichi et al. 2015).

Related literature has provided some guidance to reduce suboptimal care. Nevertheless, the general agreement is, that the provision of more and better information to both physicians and patients, is the best way for improvement (Glasziou et al. 2013).

3 State-of-the-Art on Health Recommender Systems

Recommender systems have been around for quite some time, they became a relevant research field in the mid-90s when the first article about collaborative filtering was published (Park et al. 2012). At their very beginnings, they were mostly used in e-commerce and streaming portals to suggest new items such as movies, music tracks or products to users. Two platforms using recommender systems are for

example *Amazon*⁴ and *Spotify*⁵. In the first case, it is customary, that after the user browses through the site, sees details of items or performs a purchase, Amazon immediately suggests other related items that might interest the customer. In a similar manner, Spotify gives recommendations for tracks and playlists based on the listening history of the user, as well as suggestions of trending artists and songs.

In practical terms, the objective of recommender systems is to discover items that may be of interest for users or customers and to suggest further consuming options. This recommendation is usually built upon the users' previous decisions or behaviors and the decisions made by other similar users (Calero Valdez et al. 2016; Park et al. 2012). Several approaches help identify items, which could be relevant for a particular user. The main approaches found in the literature are now introduced:

- *Collaborative filtering*: Normally, users' profiles are built taking into consideration their behaviors (e.g., purchasing history) and usually, rating of products. The system compares the different users' profiles and recommends items that have been positively rated or purchased by similar customers. Among its drawbacks, there is the performance degradation, which can occur, if the number of items and users increases (Park et al. 2012; Son and Kim 2017).
- *Content-based filtering*: In this model, items are represented as a set of attributes, which are used to compute similarities between items and perform the recommendation in conjunction with the user profile. Nevertheless, if the item's information cannot be properly categorized or if it is incorrect, then the recommendations will be unsuitable (Park et al. 2012), being this the main drawback of this method.
- *Hybrid filtering*: This approach combines the benefits of the collaborative and content-based filtering with the objective of finding more suitable recommendations (Park et al. 2012).

One aspect that the aforementioned methods have in common is that they need to relate items or users with each other. There are several techniques that allow finding such connections: they go from relatively simple statistical methods, such as Euclidean distance and Cosine similarity, to more complex techniques and theories, like sentiment analysis, regression, graph theory, and fuzzy logic (Danilova and Ponomarev 2017).

Currently, the field spectrum where recommender systems are being applied is becoming broader and broader. Some domains where recommender systems are being developed nowadays are among others politics, healthcare delivery, education and learning support and tourism (Lu et al. 2015). One field that in the last few years is taking special advantage from them is healthcare services and medicine. When recommender systems are used in the health domain, they are called *health recommender systems*. The applications that recommender systems can have in the health field are multiple, they can go from recommending food/nutritional information and

⁴<https://www.amazon.com>

⁵<https://www.spotify.com>

physical exercise to patients, to even assisting doctors and physicians in the diagnosis of diseases and in the prescription of medication (Calero Valdez et al. 2016; Helman et al. 2015). Nevertheless, decisions that users have to make in the health and medicine domain are extremely risky and potentially could have irreparable effects, which therefore adds more complexity to the design and implementation of health recommender systems. Furthermore, the data describing the health condition of a person is heterogeneous and unstructured, since it could be described in terms of simple attribute-value pairs, natural language, laboratory results, image-based exams, and many more.

3.1 Health Recommender Systems

One of the first publications that suggested the employment of digital data for clinical, educational and administrative purposes came by the hand of Mitchell (1999). Later on, Eysenbach and Jadad (2001) also suggested the idea of using information systems to assist in the task of making medical decisions. The authors put in manifest the possibility of a *“linkage of the personal online-accessible health record with general health information from evidence-based resources”*, to provide a seamless access to this knowledge and that it can be used by consumers according to their needs, at the right moment and time.

The aforementioned works inspired several authors to develop ways to improve medical decision making and health services. Probably one of the most exhaustive works in the decade of the 00s was the proposal of Pattaraintakorn, Zaverucha and Cercone (2007), who aimed to develop a web-based system capable to provide accurate and low-cost medical recommendations. The researchers tested their system using a geriatric dataset and, although the results were satisfactory, the system as a whole needed further evaluation. On the side of systems, aiming to provide suggestions to patients when it comes to selection of healthcare services, an interesting work is the one of Lee et al. (2008). The researchers came up with a recommendation framework capable of gathering patients' health status data, calculating medical similarities between healthcare services and patients' needs and recommending suitable healthcare services.

Another initiative comes from Wiesner and Pfeifer (2014) who proposed to use health recommender systems as an enhancement to *personal health record systems*. Personal health record systems are repositories of patients' health data that include information related to laboratory exams, purchases from pharmacies, insurance data, etc. The authors carried out a comprehensive work in which they discussed basic concepts, scenarios, details that need to be taken into account when implementing a recommender system into personal health records, considerations for evaluating the system as well as challenges and issues.

More recent works target the concern of improving the recommendation algorithms, to obtain more accurate suggestions. Mezrigui, Theljani and Laabidi (2017) proposed the implementation of a classification model based on the employment of a modified version of a support vector domain description algorithm endowed with

a tool based on minimal kernelized distance to tackle the multiclassification problem of diagnosing a disease. The researchers aimed to provide an efficient method to diagnose four conditions: heart disease, diabetes, hepatitis and fetal risks; they characterized every disease in a set of attributes and made use of a dataset of 1006 records to train their system and 2023 cases to test it. As a result, the authors found that their methodology offered more accurate results than the ones obtained by means of using the classification algorithms of radial basis function networks, support vector machine and standard support vector domain description (Mezrigui et al. 2017). Nevertheless, their approach was yet to be tested on a real-life scenario.

On the other hand, besides assisting health professionals in the decision-making process and providing information to patients, health recommender systems can also be employed for educational training and research purposes. The authors Fernandez-Luque, Karlsen and Vognild (2009), are one of the firsts who proposed the usage of recommender systems for health education improvement and as a possible solution to deal with the huge amount of information available on the Internet and the task of identifying relevant resources. Furthermore, the researchers suggested building a *computer-tailoring health education system* and enhancing it with a recommender system, the recommender system being in charge of resource selection and the computer-tailoring health education system being in charge of adapting the resources to the users' needs. Duan, Street and Xu (2011) developed a clinical recommender system targeted to assist nurses in patient care duties. They found that the developed system was also helpful when it came to educate and improve the nurses' patient care knowledge.

Despite the several approaches that can be found in the literature, as per the authors' knowledge, there is still a lack of comprehensive health recommender systems capable of assisting physicians in decision making processes, which provide training for medical students, that are build-up upon real data, and which serve as a knowledge base for health researchers.

4 Theoretical Background

4.1 Graph Theory

As previously stated, in a recommendation system it is crucial to keep track, hold and register all the relations between and among items and users. The usage of graph theory in the development of recommender systems is attracting a lot of attention and currently many works are being developed in regards of representing data in terms of graph models (Hti and Desarkar 2018; Robinson et al. 2013).

A graph can be defined as a set of nodes and edges, where each node represents an entity and each edge represents a relationship between entities. An example of a graph representation is presented in Fig. 1. This structure allows to represent and model different kinds of real-life scenarios, for example, a road system, the distribution network of certain product and members of a social network. For instance, a social network can perfectly be modelled by means of a graph representation, where

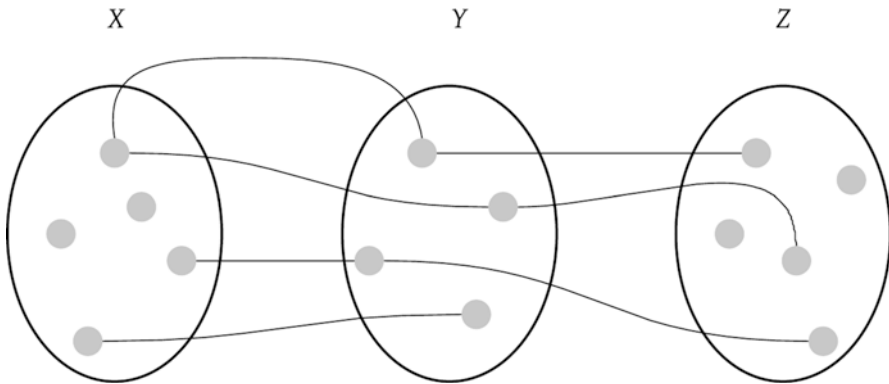


Fig. 1 Graphical illustration of sets X , Y , Z and the relations between elements

nodes are the users and the edge between two users means that they are friends or are following each other (Robinson et al. 2013).

Graphs embrace relationships, since they allow to effectively represent the connection among data. These relationships form paths. By exploring these paths, it is possible to establish different levels of connectivity between elements which, at first instance, do not seem to be related (Robinson et al. 2013).

When it comes to recommender systems, understanding the connections among entities and their strength is crucial. Graphs are a suitable option to represent the highly dense data connectivity, since they can be expressed in terms of graphs' properties (i.e., nodes and edges).

Several works regarding the usage of graph theory in the design and implementation of recommender systems can be found in the literature. One of them is the work performed by Huang et al. (2002), where the authors describe a graph-based recommendation approach to integrate content-based filtering and collaborative-filtering in the development of a recommender system for a digital library. As a first stage, similarities between customers and between books were computed. As a second stage a two-layered graph was constructed, which consisted of a book layer, where each node represented a book and the links between them corresponded to their similarity; in the second layer every node represented a customer and their links, depicted their similarities. The layers were connected. An edge from a customer to a book meant that the customer purchased the book. Having such a structure, the recommendation activity becomes solely a graph search task, which according to the researchers, works better than an exclusive content-based or a pure collaboration-based recommendation.

Hti and Desarkar (2018) proposed a graph-based approach to recommend travel packages. In this work, the location of a given user is captured as well as the time, budget and the number of places he wants to visit; his preferences regarding venues are obtained from the check-in history in social networks. After a filtering process, the candidate places are modelled as nodes and edges between them represent the cost of moving from point A to point B. The system provides the most suitable

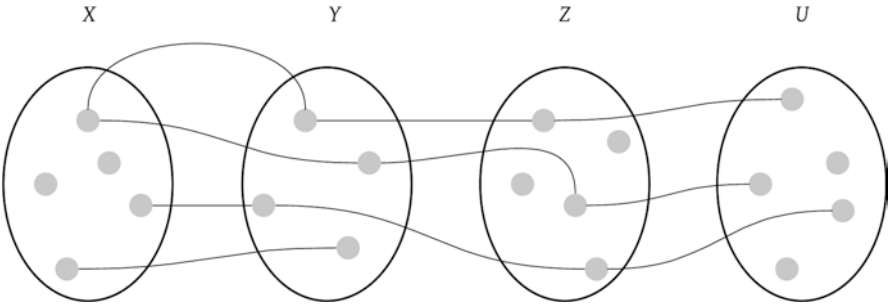


Fig. 2 Graphical illustration of sets X, Y, Z, U and the relations between elements

options of travel sequences for users. The researchers found, that using graph theory and algorithms for the building of the system was convenient in terms of accuracy of result and performance.

Graphs could also be used to model data, problems and relations in the health domain. For example, Bandler and Kohout (1986) studied in the late 1970s and 1980s relational compositions (also called *products*) and they showed how this concept may be helpful in medical diagnostics (Bandler and Kohout 1986).

Indeed, consider for instance, three finite and non-empty sets X, Y and Z , where X corresponds to a set of patients, Y to a set of symptoms and Z is a set of diseases. It is possible to model each patient, symptom and disease as graph nodes and establish relationships among them. An edge from X to Y represents that the patient x has a symptom y , whereas an edge from Y to Z means that a symptom y belongs to a disease z . Therefore, exploring the relationships between nodes would allow to determinate which disease z a patient x might have. Fig. 1 illustrates this example.

Furthermore, it is possible to extend the example considered above by adding a new set U whose elements are nodes that correspond to treatments for particular diseases. In this sense, for a patient with a given set of symptoms, it will be possible to determinate suspicious diseases and to suggest possible treatments (see Fig. 2).

4.2 Fuzzy Logic

Human beings have the ability of reasoning under imprecision and make inferences, or come up with approximated answers based on knowledge, which may be inexact or incomplete. Fuzzy logic aims to model those imprecise modes that humans have when reasoning and making decision under uncertainty (Zadeh 1988).

Definition 1 A fuzzy set is built from a reference set called *universe of discourse*. The reference set is never fuzzy. Assuming $U = \{x_1, x_2, \dots, x_n\}$ as the *universe of discourse*, then a fuzzy set A ($A \subset U$) is defined as a set of ordered pairs:

$$\{(x_i, \mu_A(x_i))\} \quad (1)$$

where $x_i \in U, \mu_A : U \rightarrow [0, 1]$ is the membership function of A and $\mu_A(x) \in [0, 1]$ is the degree of membership of x in A .

The applications that fuzzy logic can have are countless; it can be applied to solve problems of pattern recognition, classification, decision-making, optimization, etc.

When it comes to the health domain and graph theory, fuzzy logic concepts could be used to weight or describe the relationships among patients, symptoms and diseases which can be translated as a weighted graph, taking into consideration that measurements in this field can be used in terms of fuzzy concepts (e.g., high, medium, low). Moreover, by means of fuzzy logic it would be possible to represent in a more natural manner the presence of symptoms a person has. For example, a person can present *low-grade* fever, *severe* pain, *mild* bleeding, all characteristics that are almost impossible to represent with traditional logic or binary classification (e.g., yes/no, black/white).

Let us note, that the knowledge-based approach proposed by Bandler and Kohout (1986) and dealing with relational compositions was from the very beginning designed and studied also in the fuzzy variant (Bandler and Kohout 1978, 1986). The fuzzy relational compositions have then been theoretically investigated by many other researchers, such as Baets and Kerre (1993) and Běhounek and Daňková (2009), which has led to several applicably oriented extensions, such as involving generalized quantifiers (Cao et al. 2018) or employing the concept of excluding features/symptoms (Cao et al. 2017).

Apart from the mentioned above, the usage of fuzzy logic to implement health recommender systems has been explored by several authors. Song, Meng and Zhent (2015) attempted to enhance accuracy of medical diagnosis systems by employing and combining techniques and methods of *intuitionistic fuzzy sets*, which are sets with elements, that have degrees of membership and non-membership. From diagnose systems based on intuitionistic fuzzy sets methods, the researchers identified the following weaknesses: imprecision in the relation values between patients and symptoms, historic diagnoses are not taken into account to make new predictions, the determination of the most frequently contracted disease is dependent on the defuzzification method (affecting the performance of the algorithm). Regarding recommender system approaches, the following disadvantages were identified: when the relations are expressed in terms of fuzzy memberships, the accuracy of the diagnosis depends on the encoding method, the precision of results is influenced by the defuzzification method and recommender systems work only, if a training dataset is provided.

Taken the aforementioned into consideration, the authors proposed to combine techniques from intuitionistic fuzzy sets and recommender systems, to appease their weaknesses by introducing new definitions of single-criterion intuitionistic fuzzy sets recommender systems and multi-criterion intuitionistic fuzzy sets recommender systems to extend the definition of recommender systems, a new definition of intuitionistic fuzzy matrix, new similarity degrees of intuitionistic fuzzy matrices

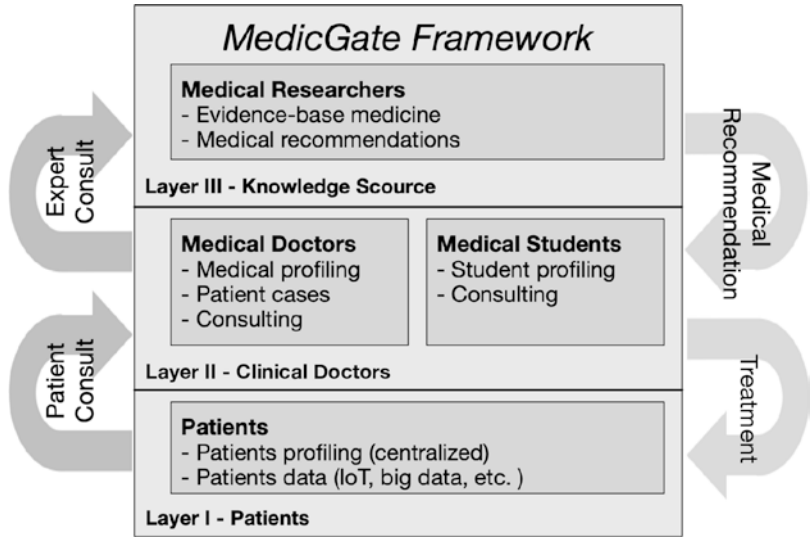


Fig. 3 Knowledge-based framework of *MedicGate*

and an intuitionistic fuzzy collaborative filtering method. As outcomes, the researchers found that the method they develop, intuitionistic fuzzy collaborative filtering method, was capable to make predictions and recommendations with more types of datasets than intuitionistic fuzzy sets and recommender systems are capable of on their own.

Other related initiative was conducted by Xiao (2018), the author proposed the use of hybrid-fuzzy-soft set decision making methods, conceived by integrating the belief entropy, fuzzy preference relations analysis, with Dempster-Shafer theory or evidence theory, introduced by Dempster (1966). This work’s goal was to use the proposed method to improve the accuracy of the recommendations, while reducing the uncertainty’s belief measure. After implementing the algorithm and tested it in a medical diagnosis application, the users were able to conclude that their method is efficient when it comes to improve the decision-making level and obtain a better computational performance. Although the aforementioned studies propose state-of-the-art techniques to improve the recommendation quality of health recommender systems, they are still not tested in a fully operative environment and therefore more evidence is needed to claim that they can work on a real-life scenario.

5 Theoretical Framework

The objective of this project is to detail the steps in specialized research in recommendation systems focused on the area of health, cognitive science and online communities with an anthropocentric approach to improve medical decision-making. To achieve this goal, the theory and practice of social computing will be addressed

through the construction and understanding of health information systems towards the so-called *science of organization* (Chia 1996).

The expected result is the introduction of an intelligent social network providing a tool for teaching and consulting for medical students and physicians, providing digital tools for the health sector, which could be used in different regions. This platform will be at the service to the public and private sectors and to academic communities to take advantage of distributed knowledge or *crowdsourcing*, to generate distributed open collaboration. Fig. 3 presents a general framework for the knowledge acquisition in the system proposed. It includes three layers, the first layer involves patients, the second layer is for physicians and students, and the third layer is dedicated to the knowledge acquisition.

The work presented in this article, optimization of medical decisions, includes only layers II and III of the framework presented in Fig. 3, where users of the second layer can make queries or consultations to the third layer (defined as *knowledge layer*). The knowledge layer used the sources provided by the Cochrane Network. Cochrane uses high-quality information to make health decisions. Cochrane has a network of 11,000 members from more than 130 countries, worldwide, including researchers, health professionals, and patients.⁶ Clinical decision optimization systems might improve the following aspects of medical practice:

- Decrease clinical heterogeneity among physicians in decision-making diagnostic process of individual patients, achieving standardization of medical practice.
- Decrease in the incidence of medical errors in daily practice by standardizing the prescription of treatments.
- Decrease the time spent on search for information about management alternatives for individual patients through the provision of protocols and clinical practice guidelines based on best scientific evidence.
- Development of computer platforms for the registration of medical records that allow improving the quality of data in patient care.
- Development of information and communication technologies between patients and colleagues that allow a more direct and effective communication for shared clinical decision-making.
- Development of universal data platforms for patient mobility among medical specialists or between geographical areas avoiding unnecessary or repeated examinations.
- Development of universal data platforms for decision making in public health and epidemiological surveillance.
- Increase of effective operative research reports for health decision making in emergencies and disasters.

The proposed framework should include a mechanism to evaluate each of aspects mentioned above, which are required in medical practice as key performance indicators.

⁶<https://www.cochrane.org/about-us>

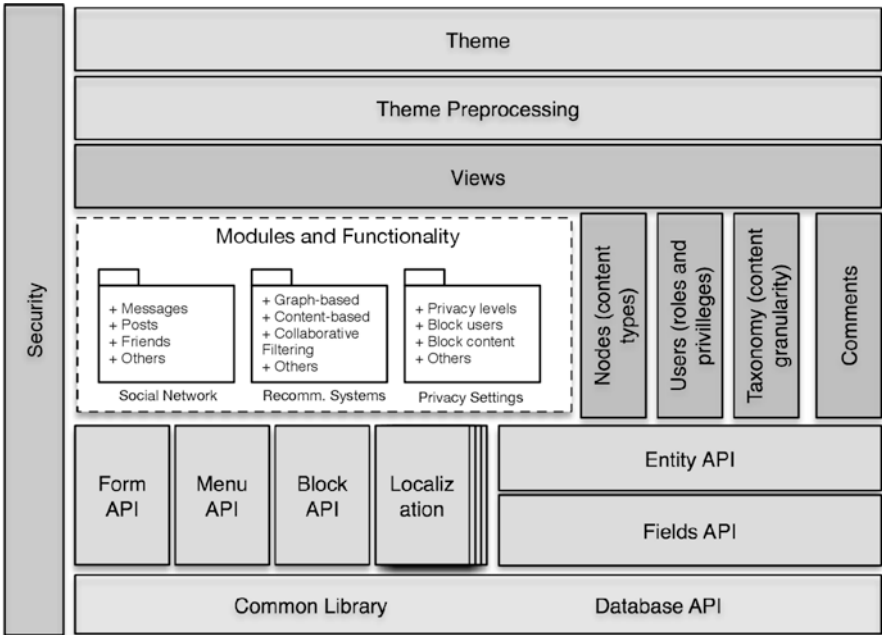


Fig. 4 Architecture of *MedicGate* platform

5.1 System Architecture

In this section, the architecture of the *MedicGate* project, which attempts to optimize medical decision taking, is described in more detail. It uses *Drupal*⁷ framework for the development of the social network features implemented. Drupal is an open-source solution and has been positioned as one of the most-used frameworks, according to the reports produced by Web Technology Surveys (2018). To implement the social network platform, the first step is to select the appropriate framework which can fulfill the requirements proposed for this social network. The Drupal content management system is an open-source software maintained and developed by a community of more than 630,000 users and developers.

The Drupal content management system is a framework with functionalities that can be added as modules, which can be enabled or disabled. It includes more than 22,000 modules developed for different purposes, such as social networking, Semantic Web, workgroups, user management, and forums.

The output of this project will be the introduction of a smart social network which optimizes medical decisions that could be used in different regions around the world as a vehicle for collecting, gathering, and analyzing data, while considering the local contexts. This platform, as well as the generated databases, will be put at the service of different academic communities. Different research communities

⁷<https://drupal.org>

can conduct studies using the datasets that will be provided, including the following: recommender systems, human-computer interaction, social networks, social behavior, physicians, medical researchers, and students. Fig. 4 presents the architecture of *MedicGate*, it includes seven different layers defined as, *data management* (common library and database API), *core blocks* (form API, menu API, block API, localization, entity API and fields API), *user and content management* (nodes, users, taxonomy, comments), *custom modules and functionalities* (social network, recommender system, privacy settings), *views management*, *theme processing*, and *security management*.

The *MedicGate* project is inspired by two of the most important projects in the domain of recommender systems, *ResearchGate*,⁸ a social networking site for scientists and researchers to share papers, ask and answer questions, and find collaborators and *GroupLens*,⁹ which was developed at the University of Minnesota. The latter includes different social platforms for conducting studies and for providing the academic community with databases in different fields such as movies,¹⁰ books¹¹ and geolocalization data.¹² These datasets have been intensively used to conduct studies, test algorithms, and evaluations.

5.2 User Registration Process

As previously stated, *MedicGate* is intended to become an intelligent social network targeted at physicians and students. It is enriched by knowledge provided by medical researchers and since the data stored in the platform is sensitive, it is necessary to perform a validation and a verification of the identity processes before granting access to users. In the case of medical students, a database of enrollments in medicine programs needs to be queried to verify if the potential new user is effectively someone studying at a university. In the case of physicians, it is necessary to verify their activity and affiliation to a registered health institution such as a hospital. Therefore, *MedicGate* shall have access to databases of the public health system to verify such information. Finally, in the case of medical researchers, their status needs to be checked against the records of the Cochraine network.

Once the status of the potential new user is verified, it will be possible to create the user in the platform and to grant access to the services. The sequence diagram shown in Fig. 5 illustrates this process.

⁸<https://www.researchgate.net/>

⁹<https://grouplens.org/>

¹⁰<https://movielens.org/>

¹¹<https://booklens.umn.edu/>

¹²<http://cyclopath.org/>

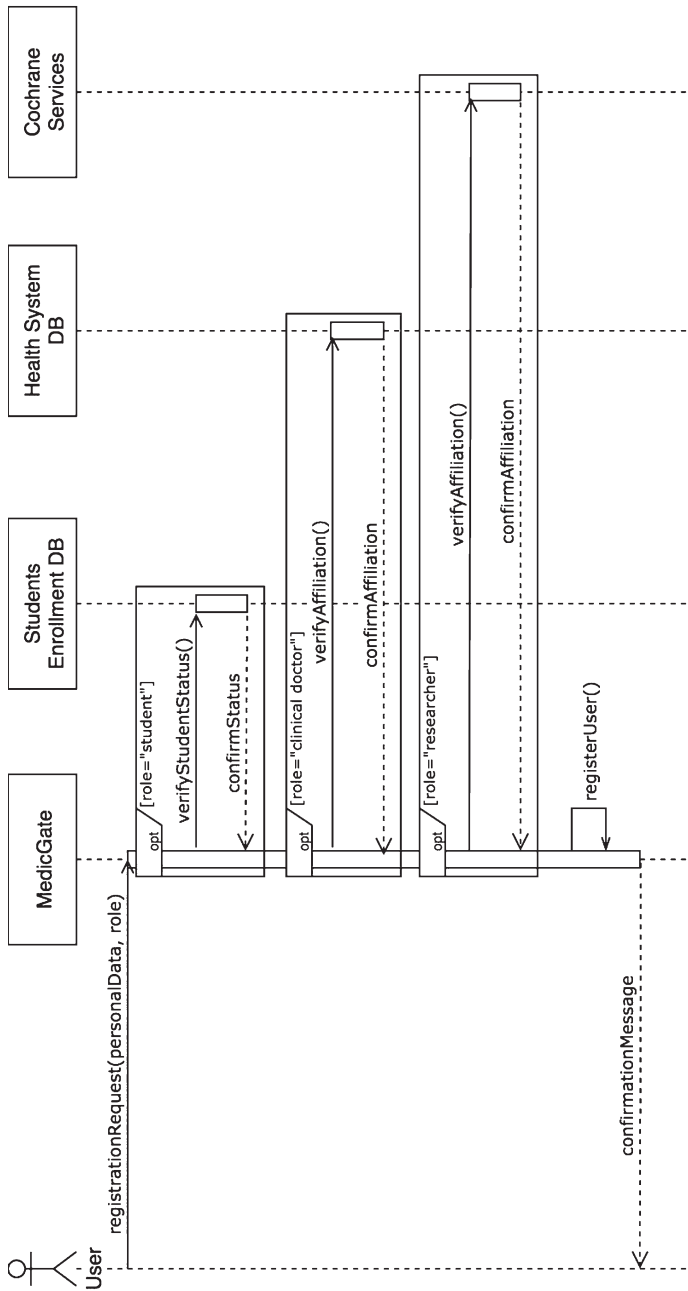


Fig. 5 User registration sequence

Table 1 User roles and capabilities

Capability	Role		
	<i>Student</i>	<i>Physician</i>	<i>Researcher</i>
Define user profile	✓	✓	✓
Search and query	✓	✓	✓
Add publications	✓	✓	✓
Participate in discussions	✓	✓	✓
Manage patients' information		✓	
Enrich knowledge base			✓
Provide medical recommendations			✓

5.3 Users Capabilities

Users of *MedicGate* will have access to different options and functionalities in the platform according to their role. For instance, medical students will be able to define their user profile, adding interests, educational background, and search and query the knowledge base for information regarding diseases, symptoms, possible treatments, etc. They also will be able to add scientific publications to their profile and to participate in discussions with their peers and physicians.

Besides the functionalities stated above, physicians will have access to a module that allows them to manage patients' information (e.g., clinical history, exams results), they also will have access to the recommendation engine to obtain suggestions about diagnosis and treatments when given a patient condition, as well as participate in discussions with peers. In the case of medical researchers, they will be the only ones capable of enriching the knowledge base with new findings for example and to provide medical recommendations, such as new treatments for certain diseases.

It should be pointed out, that the aforementioned constitutes an initial proposal for this project and it could change as *MedicGate* matures. Table 1 provides an overview of the user capabilities at this stage.

5.4 Impact

The impact and use of the platform will depend on the communication strategy used to reach as many users as possible, but also on the support of strategic partners. In addition, web analysis tools should be established for each case study to collect user interactions within the platform, including page visits, bounce rates, page time views, session duration, clicks and geolocations. Each case study can be used as input for academic publications in conferences and journals. The expected social impact would be measured using several indicators, such as the following:

- Number of users who visit the platform
- Number of people who use the information obtained in the platform for decision making

- Number of people who use the information obtained in the platform for the discussion of issues of relevance and support
- A comparison of usability by age, sex and territory
- A segmentation of interests contrasted with variables of age, sex and territory
- A comparative analysis of usability and penetration in the different regions where available

There are two perspectives to scale the project.

- The first one is technological and includes the following: the implementation of mobile applications, the improvement of existing modules, visualizations, comments from users and performance evaluations in the case of user scaling.
- The second perspective is regional, which includes running the platform in different regions and contexts; it implies cooperation with other research groups and stakeholders in the use of the platform (e.g., hospitals, clinics, research groups, and public institutions).

5.5 Social and Ethical Aspects

The development and creation of *MedicGate* and its subsequent development raise the compliance of protection of personal data, as well as bio-ethical principles. During the development of the system and during the case studies stage the following will be taken into account:

Personal Data Protection The privacy of the contact data of the members will be guaranteed in accordance with the regulations of the region, where the project will be developed. This guarantees the rights for the protection of personal data, which includes access and the decision on information and data of this nature as well as its corresponding protection. The collection, processing, distribution or spreading of these data or information will require the authorization of the person concerned or the mandate of the law.

In this sense, compliance according to the laws of the public data registration and any other norm, which regulates the handling of personal data, must be guaranteed. However, it is pertinent to indicate that once the system begins to operate, different types of research will be developed, depending on the needs and challenges posed by the participating institutions and researchers.

Ethical Rules In this project, it is expected to provide freedom of research. However, compliance with bioethics standards will be the responsibility of the institutions linked to the different researches. It is necessary to ensure compliance with the respective national regulations. The information that will be presented in this platform must be accurate and up-to-date. This space is aimed especially for physicians, health professionals, and medical students.

It is important to emphasize that the information contained in this site does not constitute medical advice. Nevertheless, the *MedicGate* project is not responsible for damages resulting from applying the information contained in this platform.

5.6 Implementation Strategies

During the initial phase of the project, it is necessary to identify the potential stakeholders, which will participate in the development of the project, including the health sector (public and private) and the interdisciplinary and transdisciplinary research groups. The project will bring together well-established research groups and networks with interdisciplinary interests in exploring the borders of so-called *collaborative work applications*. The project will be carried out with the support of medical researchers from the Equinoccial Technological University (UTE), Ecuador¹³ and its Biomedical Research Center (CENBIO), and the knowledge-based sources provided by the Cochrane Network.¹⁴

The second phase of this project is dedicated to the design of prototypes. This research plan uses the guidelines for the science of design in information systems (Von Alan et al. 2004). Design science is an information technology research methodology based on problem solving that offers specific guidelines for evaluation and iteration within research projects. It focuses on the development of artifacts (prototypes) with the explicit intention of improving its functional performance. Research in design science generally applies to categories of artifacts that include algorithms, human-computer interfaces, and design methodologies.

The third phase of the project is dedicated to the public release of the latest versions of the tools developed within the Web-based and mobile applications in a real scenario. The platform will be prepared with the initial conditions necessary to be put at the service of society, such as the creation of profiles, topics of interest, visualizations and decision-making.

The last phase of this project will be dedicated to the analysis of the data collected for validation. At this point, patient data and profiles will be filtered and investigated for specific purposes. Different evaluation methods will be carried out. The selection criteria will depend on the specific element, which will be analyzed and may include algorithms, visualizations, human-machine interaction and social impact.

6 Outlook and Conclusions

Recommendation systems are mainly used for eBusiness and eCommerce to increase sales by providing products to customers. Nevertheless, in the health sector, the use of recommender systems has opened additional channels to improve the

¹³<https://www.ute.edu.ec/>

¹⁴<https://www.cochrane.org/>

decision-making of physicians and administrators, which are part of the health system. Additionally, the usage of web-based applications, social networks and virtual communities that provide a huge amount of information, are a promising approach to address the decision-making tasks in the health sector.

Given all these factors, the creation of intelligent mechanisms and tools, such as the one proposed in this work, could allow physicians and people working on the health sector to be better informed and to improve communication channels. By means of cognitive computing, graph theory, fuzzy logic and using adaptive mechanisms that keep the human factor in the process, we expect to help improving human decision-making (Calderón et al. 2017; Cherrez Ojeda et al. 2017).

This work presents a theoretical framework for the development and the implementation of the *MedicGate* project, a smart social network for optimizing medical decisions. The initial studies will focus on Ecuador; however, in the future the project is expected to extend to other regions. It is expected that this project will affect decision making, providing more information and resources on topics of interest, medical cases, advances in technology and latest discoveries, as well as the reinforcement of collaborative work.

Due to the interdisciplinary nature of this project, it is expected that the datasets can be used in different domains of research. The data collected throughout the platform will be available for research purposes to understand and evaluate the behavior of users in different contexts, among which the authors can mention the following: recommender system design, human-machine interaction, social networks, social behavior, health, prevention, treatments, among others. According to the *privacy policy*, the information that users provide within the platform can be used for the following: (a) for the purpose for which it was provided; (b) for academic and research reasons; (c) send emails with our newsletters, to answer questions or comments and to keep our users informed.

Acknowledgements The authors would like to thank the members of Information System Research Group and the Human-IST Institute at the University of Fribourg for contributing with valuable thoughts and comments. We would also like to thank the Swiss Post and Universidad Tecnológica Equinoccial (UTE) for the institutional support to develop this project. Thanks also to Dámaris Intriago-Baldeón from the German Cancer Research Center for her valuable comments and recommendations.

References

- Ahn HS, Kim HJ, Welch HG (2014) Korea's thyroid-cancer "epidemic" – screening and overdiagnosis. *New England Journal of Medicine*, 371:1765–1767.
- Akerlof GA (1978) The market for "lemons": Quality uncertainty and the market mechanism. *Uncertainty in Economics*, pp. 235–251. Elsevier.
- Baets BD, Kerre EE (1993) Fuzzy relational compositions. *Fuzzy Sets and Systems*, 60:109–120.
- Bandler W, Kohout LJ (1986) A survey of fuzzy relational products in their applicability to medicine and clinical psychology. *Knowledge Representation in Medicine and Clinical Behavioural Science*, pp. 107–118. Abacus Press.

- Bandler W, Kohout LJ (1978) Fuzzy relational products and fuzzy implication operators. *Proceedings of the International Workshop on Fuzzy Reasoning Theory and Applications*, pp. 239–244. Queen Mary College, London.
- Běhounek L, Daňková M (2009) Relational compositions in Fuzzy Class Theory. *Fuzzy Sets and Systems*, 160:1005–1036.
- Bekelman JE, Halpern SD, Blankart CR, Bynum JP, Cohen J, Fowler R, Kaasa S, Kwietniewski L, Melberg HO, Onwuteaka-Philipsen B, Oosterveld-Vlug M, Pring A, Schreyögg J, Ulrich CM, Wunsch H, Emanuel EJ (2016) Comparison of site of death, health care utilization, and hospital expenditures for patients dying with cancer in 7 developed countries. *Jama* 315:272–283.
- Braman S (1993) Asthma in the elderly patient. *Clinics in chest medicine*, 14:413–422.
- Brownlee S (2010) *Overtreated: Why too much medicine is making us sicker and poorer*. Bloomsbury Publishing USA.
- Brownlee S, Chalkidou K, Doust J, Elshaug AG, Glasziou P, Heath I, Nagpal S, Saini V, Srivastava D, Chalmers K, Korenstein D (2017). Evidence for overuse of medical services around the world. *The Lancet*, 390:156–168.
- Calderón J, Cherrez A, Ramón GD, Jove OL, Baptist A, Matos E, Maciel BM, Calero E, Sanchez-Borges M, Cherrez S, Simancas-Racines D, Cherrez Ojeda I (2017) Information and communication technology use in asthmatic patients: a cross-sectional study in Latin America. *ERJ open research*, 3: 00005–2017.
- Calero Valdez A, Ziefle M, Verbert K, Felfernig A, Holzinger A (2016) Recommender systems for health informatics: state-of-the-art and future perspectives. *Machine Learning for Health Informatics*, pp. 391–414. Springer.
- Cao N, Holčapek M, Štěpnička M (2018) Extensions of fuzzy relational compositions based on generalized quantifiers. *Fuzzy Sets and Systems*, 339:73–98.
- Cao N, Štěpnička M, Burda M, Dolný A (2017) Excluding features in fuzzy relational compositions. *Expert Systems with Applications*, 81:1–11.
- Cerović L, Dukić N, Horvat T (2012) The agency problem in healthcare and the importance of incentives. 4th International Scientific Conference “Knowledge and Business Challenges of Globalisation in 2012”.
- Cherrez Ojeda I, Calderon J, Jove OL, Guerreros A, Plaza KJ, Cano JA, Vanegas E, Felix M, Mata V, Calero E, Cherrez A, Simancas-Racines D (2017) What kind of information and communication technologies do patients with COPD prefer to use? A cross-sectional study in Latin America. *Chronic respiratory disease*, 1479972317741895.
- Chia R (1996) The problem of reflexivity in organizational research: Towards a postmodern science of organization. *Organization*, 3:31–59.
- Danilova V, Ponomarev A (2017) Hybrid Recommender Systems: The Review of State-of-the-Art Research and Applications. *Proceedings of the 20th Conference of FRUCT Association*.
- Davidson JR, Meltzer-Brody SE (1999) The underrecognition and undertreatment of depression: what is the breadth and depth of the problem? *The Journal of clinical psychiatry*.
- Dayal P, Hort K (2015) Quality of care: what are effective policy options for governments in low and middle income countries to improve and regulate the quality of ambulatory care? Manila: WHO Regional Office for the Western Pacific.
- Deandrea S, Montanari M, Moja L, Apolone G (2008) Prevalence of undertreatment in cancer pain. A review of published literature. *Annals of oncology*, 19:1985–1991.
- Dempster AP (1966) New methods for reasoning towards posterior distributions based on sample data. *The Annals of Mathematical Statistics*, 37:355–374.
- Duan L, Street WN, Xu E (2011) Healthcare information systems: data mining methods in the creation of a clinical recommender system. *Enterprise Information Systems*, 5:169–181.
- Eysenbach G, Jadad AR (2001) Evidence-based patient choice and consumer health informatics in the Internet age. *Journal of medical Internet research*, 3.
- Fernandez-Luque L, Karlsen R, Vognild LK (2009) Challenges and opportunities of using recommender systems for personalized health education, pp. 903–907. MIE.
- Gana AMH, Fitzgerald DA (2018) Question 1: Why do children still die from asthma? *Paediatric respiratory reviews*.

- Glasziou P, Moynihan R, Richards T, Godlee F (2013) Too much medicine; too little care. *BMJ: British Medical Journal* (Online), 347.
- Helman DA, Addeo EJ, Santoso NI, Walters DW, Helman GT (2015) Delivering eHealthcare: A Perspective on Opportunities, Challenges, and Information Gaps. 17th International Conference on E-Health Networking, Application & Services (HealthCom), pp. 153–159. IEEE.
- Hti R, Desarkar M (2018) Personalized Tourist Package Recommendation using Graph Based Approach. Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization, pp. 257–262. ACM.
- Huang Z, Chung W, Ong TH, Chen H (2002) A graph-based recommender system for digital library. Proceedings of the 2nd ACM/IEEE-CS Joint Conference on Digital Libraries, pp. 65–73. ACM.
- Keyhani S, Falk R, Howell EA, Bishop T, Korenstein D (2013) Overuse and systems of care: a systematic review. *Medical care*, 51.
- Lee C, Lee M, Han D, Jung S, Cho J (2008) A framework for personalized healthcare service recommendation. 10th International Conference on E-Health Networking, Applications and Services (HealthCom), pp. 90–95. IEEE.
- Lu J, Wu D, Mao M, Wang W, Zhang G (2015) Recommender system application developments: a survey. *Decision Support Systems*, 74:12–32.
- McCartney M (2013) *The patient paradox*. Pinter & Martin Publishers.
- McDonagh MS, Smith DH, Goddard M (2000) Measuring appropriate use of acute beds: a systematic review of methods and results. *Health policy*, 53:157–184.
- Mezrigui H, Theljani F, Laabidi K (2017) Decision support system for medical diagnosis using a kernel-based approach. *International Conference on Control, Automation and Diagnosis (IC-CAD)*, pp. 303–308. IEEE.
- Mitchell J (1999) From telehealth to e-health: the unstoppable rise of e-health. Department of Communications, Information Technology and the Arts.
- Morichi V, Dell'Aquila G, Trotta F, Belluigi A, Lattanzio F, Cherubini A (2015) Diagnosing and treating depression in older and oldest old. *Current pharmaceutical design*, 21:1690–1698.
- OECD (2016) *Better Ways to Pay for Health Care*. <https://doi.org/10.1787/9789264258211-en>.
- Park DH, Kim HK, Choi IY, Kim JK (2012) A literature review and classification of recommender systems research. *Expert Systems with Applications*, 39:10059–10072.
- Pattaraintakorn P, Zaverucha GM, Cercone N (2007) Web based health recommender system using rough sets, survival analysis and rule-based expert systems. *International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing*, pp. 491–499. Springer.
- Pindyck RS, Rubinfeld DL (2013) *Microeconomics*. Boston: Pearson.
- Robinson I, Webber J, Eifrem E (2013) *Graph databases*. O'Reilly Media, Inc.
- Schwartz AL, Landon BE, Elshaug AG, Chernew ME, McWilliams JM (2014) Measuring low-value care in Medicare. *JAMA internal medicine*, 174:1067–1076.
- Shen WC, Chen JS, Shao YY, Lee KD, Chiou TJ, Sung YC, Rau KM, Yen CJ, Liao YM, Liu TC, Wu MF, Lee MY, Yu MS, Hwang WL, Lai PY, Chang CS, Chou WC, Hsieh RK (2017) Impact of undertreatment of cancer pain with analgesic drugs on patient outcomes: a nationwide survey of outpatient cancer patient care in Taiwan. *Journal of pain and symptom management*, 54:55–65.
- Son J, Kim SB (2017) Content-based filtering for recommendation systems using multiattribute networks. *Expert Systems with Applications*, 89:404–412.
- Song S, Meng Y, Zheng Z (2015) Recommending Hashtags to Forthcoming Tweets in Microblogging. *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 1998–2003. IEEE.
- van Boeckel TP, Gandra S, Ashok A, Caudron Q, Grenfell BT, Levin SA, Laxminarayan R (2014) Global antibiotic consumption 2000 to 2010: an analysis of national pharmaceutical sales data. *The Lancet Infectious Diseases*, 14:742–750.
- van Breukelen-van der Stoep DF, van Zeben D, Klop B, van de Geijn GJM, Janssen HJ, van der Meulen N, De Vries MA, Hazes M, Birnie E, Castro Cabezas M (2016). Marked underdia-

- gnosis and undertreatment of hypertension and hypercholesterolaemia in rheumatoid arthritis. *Rheumatology*; 55:1210–1216.
- Vasilopoulou I, Papakonstantopoulou I, Salavoura K, Laliotou N, Kaditis A, Gemou-Engesaeth V (2015). Underdiagnosis and undertreatment of asthma in children: a tertiary hospital's experience. *Clinical and translational allergy*, 5:P19.
- Von Alan RH, March ST, Park J, Ram S (2004). Design science in information systems research. *MIS quarterly*, 28:75–105.
- Web Technology Surveys (2018) Market share trends for content management systems for websites.
- Wiesner M, Pfeifer D (2014) Health recommender systems: concepts, requirements, technical basics and challenges. *International journal of environmental research and public health*, 11:2580–2607.
- Xiao F (2018) A Hybrid Fuzzy Soft Sets Decision Making Method in Medical Diagnosis. *IEEE Access*, 6:25300–25312.
- Zadeh LA (1988) Fuzzy logic. *Computer*, 21:83–93.

Glossar

Agnostische Statistik Kognitive Anwendungsdomänen sind charakterisiert durch Muster höherer Ordnung (d. h. aus vielen Elementen bestehend). Diese müssen in lernenden Systemen durch statistische Methoden gefunden werden. Von agnostischer Statistik spricht man, wenn keine Annahmen *a priori* über die Form dieser Muster gemacht werden. In dem Fall muss ein lernendes System sämtliche Muster bis zu einer gewissen Ordnung durchsuchen. Da die Zahl aller Muster einer gegebenen Ordnung exponentiell mit dieser Ordnung wächst, ist diese Ordnung bei agnostischer Suche auf kleine Zahlen beschränkt. Kognitive Phänomene (etwa Sprache oder visuelle Szenen) enthalten jedoch Muster sehr hoher Ordnung. Die Effizienz des Lernens im menschlichen Gehirn macht deutlich, dass dieses von sehr spezifischen *a priori* Annahmen zur Struktur dieser Muster hoher Ordnung ausgeht und diese durch Emergenz präferenziell nachbilden und erkennen kann.

AI Application Um ein Ziel in einem bestimmten Bereich (d. h. Anwendungsbe-
reich) zu erreichen, ist eine AI Application erforderlich, die aus verschiedenen AI Capabilities besteht. AI Applications sind dabei oft in einer Geschäftsanwendung integriert (z. B. Chatbots im Kundencenter).

AI Capabilities AI Capabilities beschreiben Fähigkeiten, die ein System besitzen muss, um seinen Zweck vollständig zu erfüllen. Die genaue Kombination und das Niveau der erforderlichen Fähigkeiten sind von Fall zu Fall festzulegen. Sobald die erforderlichen AI Capabilities definiert sind, müssen sie implementiert oder erworben werden, vorzugsweise in einer einfach zu implementierenden Form (sog. AI Application).

Antidisziplinäre Forschung Antidisziplinäre Forschung bezeichnet die Art von Forschung, die ausserhalb konventioneller akademischer Disziplinen durchgeführt wird und somit kein spezifisches Vokabular und keine spezifischen Rahmenbedingungen und Methoden einer Disziplin verwendet.

API API ist die etablierte Abkürzung für Programmierschnittstelle (engl. *application programming interface*) und steht für eine Schnittstelle beziehungsweise einem Programmteil, der von einem System an anderen Programmen zur Anbindung an das System bereitgestellt wird.

Artificial General Intelligence Damit Artificial General Intelligence Realität wird, müsste ein System intelligentes Verhalten in allen Bereichen zeigen (ohne auf einen bestimmten Kontext beschränkt zu sein) und ein hohes Mass an autonomer Selbstkontrolle besitzen. Dies wird auch als eine starke Form der „*wahren*“ künstlichen Intelligenz bezeichnet.

Artificial Narrow Intelligence Artificial Narrow Intelligence steht für eine schwache Form der „*wahren*“ künstlichen Intelligenz. Um diese zu erreichen, müsste das System mindestens einige Fähigkeiten in der Wahrnehmung, im Lernen, im Verallgemeinern, in der Anpassung, im Schlussfolgern und in der Wissensrepräsentation besitzen, und fähig sein, diese innerhalb eines bestimmten Bereichs autonom zu verwenden.

Between-Subject-Design Verschiedene Personen, meist eingeteilt in Treatment- und Kontrollgruppen, testen die experimentellen Bedingungen und werden danach miteinander verglichen. Bei der Treatmentgruppe wird die unabhängige Variable manipuliert, bei der Kontrollgruppe geschieht dies nicht.

Bisociation Bisociation bezieht sich auf die Mischung von Konzepten aus zwei Kontexten oder Kategorien von Objekten, die normalerweise durch die literalen Prozesse des Geistes getrennt betrachtet werden.

Blockchain Eine Blockchain ist eine Kette von Datensätzen, die mittels kryptographischer Verfahren miteinander verbunden sind. Die Blockchain wird dezentral auf einer Vielzahl von Servern (sog. *nodes*) gespeichert und ermöglicht somit eine Grundlage, auf der sich neue Applikationen entwickeln lassen.

Chatbot Chatbots sind Dialogsysteme, die sich mit einem Nutzer in natürlicher Sprache über Textnachrichten unterhalten und eigenständig einfache Aufgaben ausführen können. Sie werden hauptsächlich auf Webseiten oder Messaging-Plattformen eingesetzt.

Citizen Scientist Als Citizen Scientist bezeichnet man Menschen, die nicht als formell ausgebildete Wissenschaftler und ausserhalb der traditionellen wissenschaftlichen Strukturen und Institutionen als eine Art *Amateur-Wissenschaftler* tätig sind. Citizen Scientist können sich beispielsweise über Online-Plattformen in den Forschungsprozess einbringen und auf diese Weise mit *Profi-Wissenschaftlern* zusammenarbeiten. Oft übernehmen sie dabei Aufgaben, die einen sehr hohen Arbeitsaufwand mit sich bringen und gleichzeitig nicht von Computern erledigt werden können. Der hohe Arbeitsaufwand kann durch die grosse Zahl teilnehmender Citizen Scientists dann gelöst werden, wenn sich die Arbeit gut auf viele Personen verteilen lässt.

Cloud Service Ein Cloud Service ist eine Dienstleistung, welche dem Benutzer bei Bedarf vom Server eines Cloud Computing Anbieters über das Internet zur Verfügung gestellt wird. So muss der Benutzer keine eigenen Server vor Ort verwalten und kann die Produkte (Server, Speicherort, Datenbanken, etc.) beziehen, die er gerade braucht.

Code Audits Code Audits sind umfassende Analysen des Quellcodes in einem Programmierprojekt mit dem Ziel, allgemeine Fehler oder Sicherheitsverletzungen zu entdecken, und schliesslich eine fehlerlose Softwareveröffentlichung zu ermöglichen.

Cognitive Services Cognitive Services stehen für die verschiedenen kognitiven Dienste, die Microsoft als einzelne Cloud-basierten APIs anbietet. Diese lassen sich in die Kategorien *Knowledge* (Wissensmanagement), *Language* (Textverarbeitung), *Search* (Suchmaschinen), *Speech* (Sprachverarbeitung) und *Vision* (Bildverarbeitung) unterteilen.

Confidence Score Der Confidence Score beziehungsweise der Vertrauenswert ist ein Schwellenwert [0, 1], der den Mindestwert bestimmt, der akzeptabel ist, um eine Aktion auszulösen. Bei mehreren Vertrauenswerten wird die Aktion ausgelöst, die den höchsten Vertrauenswert aufweist.

Connectomics Connectomics steht für die Produktion und Erforschung von Connectomen – umfassende Karten von Verbindungen innerhalb des Nervensystems eines Organismus (typischerweise eines Gehirns).

Conversational Agent Conversational Agents sind Softwaresysteme, mit denen Nutzer über natürliche Sprache interagieren können. Dazu zählen sowohl die textbasierten Chatbots als auch die populären Sprachassistenten wie Amazon's Alexa oder Apple's Siri.

Data Scientist Ein Data Scientist ist ein IT-affiner analytischer Datenexperte. Er verfügt über die technischen Fähigkeiten, um Daten auszuwerten und Wissen zu extrahieren. Die daraus gewonnenen Erkenntnisse dienen als Entscheidungsgrundlage für die Problemlösung.

Deadlock Deadlock ist die gegenseitige Blockierung von Fahrzeugen an einem Wegabschnitt. Durch das gleichzeitige Antreffen von Fahrzeugen an einer Wegstrecke behindern sich die Fahrzeuge gegenseitig und kein Fahrzeug kann weiterfahren. Dieses Verhalten kann durch mangelnde Steuerung vor allem an Kreuzungen entstehen.

Decision Support System Ein Decision Support System ist ein computergestütztes System, das flexibel und anpassungsfähig ist und Unterstützung bei der Planung und Problemlösung bietet.

Deductive Reasoning Deductive Reasoning beginnt mit einem Regelwerk (das konsistent ist und sich nicht widerspricht) und geht, ausgehend von einem Input, mit logischem Denken weiter, um zu einer Schlussfolgerung zu gelangen, die diese Regeln respektiert.

Deep Learning Deep Learning gehört zum Bereich Machine Learning, in welchem das Programm (meist bestehend aus einem neuronalen Netzwerk) aus Beispielen selbstständig lernt, Muster erkennt und darauf basierend Schlussfolgerungen zieht. Von Deep Learning ist die Rede, wenn sehr viele Zwischenschichten (ca. >15) vorhanden sind.

Design Science Research Design Science Research beschäftigt sich mit der iterativen Gestaltung und Evaluation innovativer Artefakte zur Lösung realer Probleme, die in verschiedenen Anwendungsbereichen beobachtet werden. In der Wirtschaftsinformatik können solche Artefakte Konstrukte, Modelle, Methoden und soziale Innovationen sein.

Digital Society Digital Society umfasst das Konstrukt einer modernen Gesellschaft, in welcher Informations- und Kommunikationstechnologien in alltäglichen Situationen (bspw. im Haus, auf der Arbeit, in der Freizeit) integriert sind und diese auch verwendet werden.

Digitalisierung Im ursprünglichen Sinn steht die Digitalisierung für die Transformation von einer analogen zu einer digitalen Welt. Dabei sollen analoge Informationen, Objekte und Ereignisse zu digitalen Repräsentationen dieser Werte umgewandelt werden.

Direkte vs. indirekte Algorithmen Direkte Algorithmen sind solche, die spezifische Lösungen für spezifische Anwendungen darstellen. Indirekte Algorithmen dagegen formulieren allgemeinere heuristische Methoden, mit denen Lösungen für breitere Anwendungsfelder gefunden werden können. Der Unterschied zwischen beiden Typen von Algorithmen ist graduell. An einem Ende des Spektrums stünde zum Beispiel ein Schachprogramm, das über Züge nach expliziten strategischen Regeln entscheidet, am anderen Ende ein Programm, das lediglich das finale Ziel vorgibt und die Strategie selbst entwickelt. Zu den indirekten Algorithmen zählen evolutionäre Methoden, Reinforcement Learning und Back-propagation of Error.

Diskrete Planung Algorithmen der diskreten Planung stellen die Basis für viele weitere Wegeplanungsalgorithmen dar. Sie werden hauptsächlich in Umgebungen verwendet, die endlich sind. In vielen Fällen unterteilt der Algorithmus die Umgebung in Zellen und weist jeder Zelle einen Status zu. Durch Transitions-Funktionen wird der Status der jeweiligen Zelle geändert und durch eine Zielfunktion der optimale Weg durch die Zellen geplant.

Diskurs Ursprünglich mit der Bedeutung *erörternder Vortrag* oder *hin und her gehendes Gespräch*, stellt der Diskurs in diesem Buch die Gesamtheit aller Aussagen zu einem bestimmten Thema innerhalb einer bestimmten Epoche dar, welche für diesen Zeitraum das in der Sprache aufscheinende Verständnis von Wirklichkeit konstituieren.

Einzelzellhypothese Wenn ein neuronales System eine Entscheidung trifft – etwa über das Vorhandensein oder Nichtvorhandensein eines gewissen Typs von Merkmal oder Objekt in einem Bild – dann gibt es nach der Einzelzellhypothese immer (mindestens) ein Neuron, das auf diese Entscheidung spezialisiert ist. Diese Hypothese beschränkt die heutigen neuronalen Modellsysteme auf vordefinierte Funktionen, erlaubt diesen also keinen kreativen Umgang mit neuen Problemen. Dies ist im Kontrast zu sehen zu Systemen, die neue Probleme durch neue Konstellationen vorhandener Funktionskomponenten, also etwa spontane Emergenz neuer Nervennetze, zu lösen fähig sind.

Emergenz Ein selbstorganisierendes oder zu Emergenz fähiges System besteht aus einer grossen Zahl von (zumeist mikroskopischen) Elementen, die unter dem Einfluss ihrer Wechselwirkungen die Tendenz haben spontan in geordnete Muster zu fallen. Ein typisches Beispiel ist die Bildung von Membranen durch Seifenmoleküle und die Bildung von Schaumgebilden aus diesen. Emergente Phänomene brauchen zu ihrer Beschreibung Begriffe, die nur schwierig oder gar nicht auf die der Beschreibung der Bauelemente zurückzuführen sind. Im Gehirn beruht Emergenz auf der Wechselwirkung neuronaler Verbindungen und ist die Basis für seine kognitive Architektur.

Evidenz-basierte Medizin Evidenz-basierte Medizin bezieht sich auf die gewissenhafte Verwendung der aktuell besten Erkenntnisse (sog. Evidenz od. Fakten) aus der Medizin, um die Entscheidungsfindung über die Gesundheitsversorgung und das Wohlergehen jedes Einzelnen zu optimieren.

Expertensystem Ein Expertensystem ist ein Computersystem, das darauf ausgerichtet ist, dem Nutzer im Kontext passende Handlungsempfehlungen auszusprechen. Für die Generierung von Empfehlungen, verwendet das System unter anderem die Daten des Nutzers und die abgespeicherten Daten in der Wissensdatenbank.

Explorative Kreativität Explorative Kreativität steht für die Fähigkeit, innerhalb eines bestimmten begrifflichen Konzeptes (z. B. Stilrichtung, wissenschaftliche Theorie), auf Entdeckungsreise zu gehen und neue, überraschende Ideen innerhalb des existenten Feldes zu kreieren.

Fahrerlose Transportfahrzeuge Fahrerlose Transportfahrzeuge sind flurgebundene Fördermittel für den innerbetrieblichen Materialtransport. Sie sind automatisch gesteuert und bewegen sich auf vorgegebenen Wegenetzen. Zur Lastaufnahme und – abgabe verfügen sie über aktive oder passive Lastaufnahmemittel (z. B. Hubplattformen). Sie haben einen eigenen Lenkmotor und sind mit Sensorik zur Navigation und des Personenschutzes ausgestattet.

Fahrerlose Transportsysteme Ein Fahrerloses Transportsystem besteht aus Fahrerlosen Transportfahrzeugen und einer Leitsteuerung, mit der Aufgabe den innerbetrieblichen Materialtransport zu organisieren. Durch den automatisierten Materialtransport sollen Transportvorgänge pünktlich und standardisiert sowie Transportschäden und Personalbindung im Transport minimiert werden.

Fields of Application Anwendungsbereiche stellen Aufgaben oder Ziele dar, die den KI-Systemen zum Wohle des Menschen zugeordnet werden können. Jedes Anwendungsgebiet erfordert eine eigene Kombination von High-Level-Fähigkeiten, was zu einer Vielzahl von KI-Technologien führt.

Fuzzy Logik Fuzzy Logik ist eine Erweiterung der binären (booleschen) Logik. Anstelle von $X \rightarrow \{0,1\}$ (wahr oder falsch) werden reelle Zahlen im Intervall $X \rightarrow [0,1]$ (von wahr bis falsch) für die Abbildung einer Variable X verwendet. Sie erlaubt Elementen eine graduelle Zugehörigkeit zu Mengen.

Genetische Algorithmen Genetische Algorithmen stellen evolutionäre Mechanismen mithilfe von mathematischen Algorithmen dar. Inspiriert durch die Evolution, suchen sie iterativ in einem Lösungsraum nach der optimalen Lösung, um das Produkt oder Verfahren zu verbessern. Dabei kommen Operationen wie die Selektion, Mutation und Rekombination zum Zug.

Geometrische Ansätze Geometrische Ansätze versuchen das Problem der langen Berechnungszeiten der Diskreten Planung zu überwinden. Der Fokus dieser Ansätze liegt in der effizienten Unterteilung der Umgebung. Anstatt die Umgebung in gleich grosse Zellen zu unterteilen, werden verschiedene Konfigurationen der Umgebung unternommen, wie zum Beispiel Dreiecke oder Polygone.

Gestaltungsprinzip Gestaltungsprinzipien sind präskriptive Anweisungen, die abstrakt beschreiben, wie eine Klasse von Artefakten gestaltet werden sollte, um

ein bestimmtes Problem zu lösen. Diese Prinzipien entstehen in der Regel im Rahmen eines Forschungsprojekts mit einem gestaltungsorientierten Ansatz.

Graph Theory Graph Theory (deut. Graphentheorie) bezeichnet das Studium von Graphen, das heisst, Strukturen, die dazu dienen, Beziehungen zwischen Entitäten und deren Eigenschaften zu beschreiben.

Hard Computing Hard Computing ist ein Oberbegriff für exakte Modelle, die in der Regel nur mit strukturierten Daten umgehen können und deren Anwendung präzise Antworten generiert. Es bietet den Vorteil von vorhersehbaren und nachvollziehbaren Ergebnissen, die nur unter den zuvor definierten Bedingungen gültig sind und nur bedingt auf die Realität angewendet werden können.

Health Recommender System Ein Health Recommender System ist ein Informationssystem, das Empfehlungen zur Verbesserung des Wohlbefindens der Patienten geben kann und Ärzte bei Entscheidungsprozessen in Bezug auf Diagnose und Behandlung von Patienten unterstützt.

Homeomorphie Gegeben zwei Gebilde, die aus Teilen und deren Beziehungen bestehen, wie zum Beispiel das Bild eines Objekts und eine schematische Beschreibung eines Objekts. Eine Abbildung zwischen den beiden Gebilden, die Teile des einen Gebildes auf Teile derselben Art in dem anderen Gebilde so bezieht, dass Beziehungen in Beziehungen übergehen, nennt man eine Homeomorphie. Homeomorphie herrscht zwischen Objekten oder Vorgängen und den sie definierenden Schemata.

Human Resource Management Human Ressource Management (deut. Personalmanagement) ist der strategische Ansatz für das effektive Management von Mitarbeitern in einem Unternehmen. Diese Abteilung ist zuständig, geeignete Arbeitnehmer zu akquirieren, zu fördern und zu halten, um die Ziele des Unternehmens zu erfüllen.

Ill-structured problems Als *ill-structured problems* (schlecht strukturierte Probleme) bezeichnet man Probleme, die sich dadurch auszeichnen, dass die Ausgangslage diffus ist und dass klare Kriterien zu Lösungsvalidierung fehlen. Die zur Lösung von *ill-structured problems* benötigten Wissensgrundlagen sind unklar oder fehlen und es gibt keine formellen Lösungsregeln. *Ill-structured problems* verschliessen sich also einem antizipierbaren Lösungsweg.

Inductive Reasoning Inductive Reasoning beginnt damit, einen Datensatz zu erforschen und daraus allgemeine Regeln abzuleiten. Die anschliessende Anwendung dieser Regeln auf eine bestimmte Eingabe führt nicht zu einer Proposition, die notwendigerweise eine logische Folge ist. Im Wesentlichen geht es darum, Informationen aus einem grossen Datensatz zu extrahieren, aus dem Wissen gewonnen werden kann, ohne im Voraus zu bestimmen, wonach gesucht wird.

Innovation Innovation steht einerseits für eine Verbesserung oder Weiterentwicklung von Bestehendem (inkrementell) oder andererseits für eine grundlegende Veränderung von Bestehendem oder Entwicklung von etwas ganz Neuem (radikal).

Intelligent Design Intelligent Design steht für die kreationistische Auffassung, dass das Leben oder das Universum nicht zufällig entstanden sein kann und von einer intelligenten Einheit entworfen und erschaffen wurde.

Intent-Matching Intent-Matching ist der Prozess, in welchem der eingegebene Input mit im System bestehenden Input-Output-Paaren abgeglichen wird. Der beste Match, das heisst die grösste Übereinstimmung (ausgedrückt mit Vertrauenswerten), wird in den meisten Fällen wiedergegeben.

Kognitive Architektur Eine kognitive Architektur ist der organisatorische Rahmen, im Gehirn oder im Rechner, innerhalb dessen sich Cognitive Computing abspielt. Dieser Rahmen muss weit genug sein, um alle Arten kognitiver Prozesse zu ermöglichen, andererseits aber im Interesse von Effizienz bereits verengend auf die allgemeinen Züge kognitiver Phänomene zugeschnitten sein.

Kognitiver Digitaler Assistent Ein kognitiver, digitaler Assistent bietet im Bereich Personalmanagement Vorschläge zu Rollenwechsel, Weiterbildung und Umqualifizierung unter Berücksichtigung von Trends am Markt, der Strategie der Unternehmensführung, den Erfahrungen und Talenten der Mitarbeiter.

Kognitives System Computersysteme, die durch die Interaktion mit Datenmengen und menschlichen Eingaben selbsttätig lernen und im begrenzten Rahmen neu programmieren können, um neue und zuvor unbekannte Aufgaben zu bewältigen und den Menschen augmentativ zu unterstützen.

Kollektive Intelligenz Kollektive Intelligenz ist das Produkt einer Zusammenarbeit von Individuen mit unterschiedlichen Intelligenzen und Hintergründen. Entscheidungen, die gemeinsam getroffen werden, sind meist smarter, als solche, die ein Individuum alleine trifft.

Kombinatorische Kreativität Bei der kombinatorischen Kreativität werden innerhalb bekannter und vertrauter Ideen ungewohnte Kombinationen hergestellt, die auf ihre Art und unter Berücksichtigung bestehender Regeln und Normen neue Herangehensweisen oder ungewöhnliche Zugänge zu bestehenden Inhalten aufzeigen.

Kreativität Kreativität wird als ein begriffliches Konstrukt verstanden, welches sich einer klaren Definition entzieht und hierdurch als Vehikel zahlreicher und unterschiedlicher Zuschreibungen benutzt wird, die von der Tradition der Genieästhetik (nur wenige Menschen sind wahrhaft schöpferisch tätig) bis hin zur Logik eines kreativen Imperativs reichen (alle Menschen müssen irgendwie kreativ sein).

Künstliche Kreativität Künstliche Kreativität ist das multidisziplinäre Bestreben (menschliche) Kreativität zu reproduzieren oder zu simulieren. Durch die Indifferenz des Kreativitätsbegriffs ist es ausgewiesenes Ziel auch dieses Forschungszweigs die menschliche Kreativität besser zu verstehen und durch Computersysteme zu fördern.

Künstliches Neuronales Netz Inspiriert durch das menschliche Gehirn, haben künstliche neuronale Netze Neuronen, mit denen sie Informationen aufnehmen, sie verarbeiten und dann als Ergebnis wieder weitergeben. Das einfachste Netzwerk besteht aus einer Eingabeschicht (Input-Layer), einer Zwischenschicht (Hidden-Layer) und einer Ausgabeschicht (Output-Layer).

Legal Tech Legal Tech steht für Software, welche in der juristischen Praxis eingesetzt wird, um Prozesse zu automatisieren und dadurch JuristInnen in ihrer Arbeit, wie beispielsweise Recherchearbeiten, zu unterstützen. Legal Tech geht einher mit einer Transformation der juristischen Geschäftsmodelle, etwa durch das Entstehen von Legal-Tech-Startups.

LUIS (Language Understanding Intelligence Service) LUIS ist eine kognitive Programmierschnittstelle von Microsoft für die maschinelle Sprachverarbeitung. Dieser Dienst basiert auf Machine Learning und ermöglicht eingehende Informationen (z. B. Textnachrichten) seitens der Nutzer zu verarbeiten, sodass passende Antworten (die mit den höchsten Vertrauenswerten) generiert werden können.

Minimum Viable Product (MVP) Ein neues Produkt wird mit ausreichenden Funktionen entwickelt, um den Anforderungen der Erstnutzer zu genügen. Die Weiterentwicklung des Produktes bis zu seiner endgültigen Fertigstellung erfolgt unter Berücksichtigung der Rückmeldungen der Benutzer.

Moravec's Paradoxon Moravec's Paradoxon steht für die Entdeckung, dass im Gegensatz zu dem, was normalerweise angenommen wird, hochrangiges Denken sehr wenige Rechenressourcen und niedrigrangige Wahrnehmungs- und motorische Fähigkeiten eher enorme Rechenleistungen erfordern. Es ist zwar relativ einfach, einem System das Spiel Dame beizubringen, jedoch fast unmöglich, die sensomotorischen Fähigkeiten eines Einjährigen zu vermitteln.

Natural Language Processing Im textbasierten Umkreis, wird bei Natural Language Processing zwischen syntaktischen und semantischen Operationen unterschieden. Die syntaktischen Methoden konzentrieren sich auf die Strukturen des Satzes und die Funktionen der einzelnen Elemente darin und die semantischen Methoden fokussieren sich hingegen auf die Bedeutung dieser Elemente.

Netzwerk-Selbstorganisation Wenn eine gewisse Menge von Neuronen wiederholt aktiv ist, werden dieser Aktivität durch die vorhandenen Verbindungen zeitliche Korrelationen aufgeprägt. Diese Korrelationen wirken durch synaptische Plastizität auf die Stärke der Verbindungen zurück, was wiederum die Statistik der Signale verändert. Die Verbindungen sind in diesem Kreislauf zwei Einflüssen unterworfen – einerseits konkurrieren auf ein Zielneuron konvergierende Verbindungen miteinander, andererseits kooperieren Verbindungen, die von einer Ursprungszelle ausgehen und zur gleichen Zielzelle führen. Diese Wechselwirkungen führen zur Emergenz gewisser Netzstrukturen, die als Basis für Kognition dienen können.

Neuronaler Code Unter neuronalem Code versteht man eine Interpretation der physischen Struktur und Aktivität des Gehirns als Datenformat für kognitive Strukturen und Prozesse. Zusammen mit den Mechanismen der Organisation dieses Datenformats konstituiert der neuronale Code die kognitive Architektur des Gehirns.

Neural Network Als Neural Network (neuronales Netzwerk oder neuronales Netz) bezeichnet man in den Neurowissenschaften eine beliebige Anzahl miteinander verbundener Neuronen, die als Teil eines Nervensystems einen Zusammenhang bilden, der einer bestimmten Funktion dient. In der Informatik, der Informationstechnik und der Robotik werden funktionale Strukturen als künstliches neuronales Netz modelliert und technisch nachgebildet, simuliert und abgewandelt.

Overfitting Overfitting beziehungsweise Überanpassung tritt auf, wenn das Modell das Rauschen der Daten erfasst und folglich zu gut zu den Daten passt. Dies kommt häufig bei komplexeren und grösseren Modellen vor und ist erkennbar bei einer geringen Verzerrung, aber gleichzeitig hoher Varianz.

Pattern Matching Pattern Matching ist ein Mechanismus, der ständig versucht in den verfügbaren Daten Muster zu erkennen und diese mit vordefinierten Mustern abzugleichen. Das heisst, das Ergebnis kann entweder wahr (eine Übereinstimmung) oder falsch (keine Übereinstimmung) sein.

Phenotropic Computing Phenotropic Computing ist das Konzept von Soft- und Hardwaresystemen, die über Oberflächen mit Hilfe von Mustererkennung kommunizieren, im Gegensatz zur traditionellen Kommunikation, die auf Single Point of Contact und Protokolleinhaltung basiert.

Plastizitäts-Stabilitäts-Dilemma Das Plastizitäts-Stabilitäts-Dilemma steht für das zentrale Problem des kontinuierlichen Lernens eine sinnvolle Balance zwischen Stabilität und Plastizität zu finden – also so wenig wie möglich des Erlernens zu verlernen und dennoch den neuen Veränderungen des Prozesses zu folgen.

QnA Maker Der QnA Maker ist ein webbasierter Dienst von Microsoft, der zum einen als Informationsspeicher dient (bzw. Wissensdatenbank, z. B. bestehend aus Frage-Antwort-Paaren) und zum andern als Antwortgenerator (d. h. nach Abgleich eines Frage-Antwort-Paars folgt die bestmögliche Antwort).

Reactive Machines Reaktive Maschinen enthalten Systeme, die in der Lage sind, sofort auf aktuelle Eingaben zu reagieren, ohne notwendigerweise explizit darauf programmiert zu sein. Dadurch können reaktive Systeme auch komplexere oder unscharfe Situationen bewältigen.

Recommender System Recommender System (deut. Empfehlungssystem) ist ein Informationssystem, das darauf abzielt, Artikel zu entdecken, die für Kunden von Interesse sein können. Diese Empfehlung basiert in der Regel auf den früheren Entscheidungen oder Verhaltensweisen der Benutzer und den Entscheidungen anderer ähnlicher Benutzer.

Regular Expression Regular Expressions (abgekürzt Regex) sind eine Folge von Zeichen, die ein Suchmuster definieren und somit als Filterkriterien verwendet werden können. Zum Beispiel findet das Muster `[0-9]{2}` alle Instanzen in einem Text, in dem zwei Ziffern zwischen 0 und 9 aufeinander folgen.

Ricardian Contracts Ein Ricardian Contract stellt einen digitalen Vertrag dar, der die Bedingungen für eine Interaktion zwischen zwei oder mehreren Peers festlegt, der kryptographisch signiert und verifiziert ist. Dieser ist sowohl für den Menschen als auch für die Maschine lesbar und digital signiert.

Schema Ein Schema ist eine auf das Wesentliche beschränkte Darstellung eines Objekts oder Vorgangs und kann als deren abstrakte Definition angesehen werden. Ein Schema besteht aus Beschreibungen von Teilen und deren Beziehungen, wobei diese selbst wieder zusammengesetzte Schemata sein können. Schemata werden seit Kant als das Werkzeug gesehen, mit dem Dinge verstanden werden und sind die Grundlage für Verallgemeinerung zwischen verschiedenen Beispielen derselben Sache.

Semantic Computing Semantic Computing umfasst Technologien, welche die Ableitung, Beschreibung, Integration, Manipulation und Verwendung von Semantik (d. h. Bedeutung, Kontext) für vielerlei Arten von Daten ermöglichen.

Semantic Web Das Semantic Web ist eine Erweiterung des World Wide Web durch Standards, die von vom World Wide Web Consortium (W3C) festgelegt sind.

Diese Standards schaffen ein hohes Mass an Einheitlichkeit in Bezug auf Datenformate und Datentransfer-Protokolle. Damit schafft das Semantic Web eine Art Rahmenvereinbarung, dank der Daten anwendungs-, unternehmens-, oder regionenunabhängig geteilt und weiterverarbeitet werden können.

Sensorik In der Technologie steht Sensorik für die Wissenschaft und Anwendung von Sensoren. Sensoren können Veränderungen in der eingesetzten Umgebung erkennen und Echtzeitdaten zu angehängten Systemen senden, welche wiederum die Daten auswerten und verwenden können.

Simultaneous Localization and Mapping (SLAM) Simultaneous Localization and Mapping stellt ein Problem in der Robotik dar. Der mobile Roboter muss zum einen eine Karte seiner unbekannten Umgebung erstellen oder aktualisieren und zum andern seine Position innerhalb dieser Umgebung schätzen.

Smart City Die effiziente Sammlung und Auswertung stadtbezogener Daten sowie die Koordination ihrer Nutzung mittels Informations- und Kommunikationstechnologien, um den sozialen und ökologischen Lebensraum nachhaltig zu entwickeln, zeichnen eine Smart City aus.

Smart Contract Smart Contracts sind Computerprotokolle. Sie bilden Verträge ab oder unterstützen die Abwicklung eines Vertrags. Dies mit dem Ziel eine höhere Vertragssicherheit und Transparenz zu erlangen und Transaktionskosten zu minimieren. Aktuelle Implementierungen von Smart Contracts sind in erster Linie auf Blockchains zu finden.

„Smart“ Information Systems „Smart“ Information Systems beziehen sich auf Systeme mit Fähigkeiten, die leicht über denen der traditionellen älteren Informationstechnologien liegen. Sie sind in der Lage variable und möglicherweise weniger strukturierte und unvollständige Eingaben zu verarbeiten und darauf basierend Rückschlüsse zu ziehen.

Soft Computing Soft Computing ist im Vergleich zu Hard Computing tolerant gegenüber unsicheren, unvollständigen und vagen Informationen, die typisch für viele praktische Problemstellungen sind. Unter Soft Computing ist der Einsatz approximativer Kalkulationen zu verstehen, die das Ziel verfolgen geeignete (auch wenn nicht exakte) Lösungen für komplexe Problemstellungen zu finden.

Soziales Signal In der zwischenmenschlichen Kommunikation sind soziale Signale implizite Hinweisreize (z. B. Lächeln, Blickkontakt, Stimmelage), aus denen sich soziale Informationen ableiten lassen. Selbst in der Interaktion mit Technologien nehmen Nutzer diese sozialen Signale unterbewusst wahr und reagieren automatisch darauf.

Soziale Reaktion Soziale Reaktionen sind emotionale und kognitive Reaktionen in der Interaktion mit Technologien. Diese Reaktionen werden durch soziale Signale hervorgerufen, die dazu führen, dass Nutzer einer Technologie menschliche Attribute zuschreiben und gelerntes Verhalten aus der zwischenmenschlichen Kommunikation auf diese Technologie anwenden.

Soziotechnische Gestaltung Die soziotechnische Gestaltung zielt darauf ab, sowohl technische als auch soziale Faktoren und Erkenntnisse möglichst gleichberechtigt in den Gestaltungsprozess einer Technologie einzubringen. In der Entwicklung eines Dialogsystems würden beispielsweise Erkenntnisse aus der zwischenmenschlichen Kommunikation berücksichtigt werden.

Strategic Workforce Management Strategic Workforce Management ist dafür zuständig, Talente im Unternehmen zu erkennen, diese zu fördern und den entsprechenden unternehmerischen Herausforderungen zuzuteilen und daher die Zukunftsfähigkeit des Unternehmens zu sichern.

Systemplaner Der Systemplaner ist verantwortlich für die Erstellung eines optimalen Wegenetzes. Dabei muss er unterschiedliche Inputfaktoren, wie die Transportmatrix, das Layout der Umgebung oder Eigenschaften des Fahrerlosen Fahrzeugs berücksichtigen und bewerten.

Ticket Ein Ticket repräsentiert ein strukturiertes Formular, das die wichtigsten Informationen (z. B. Kontakt, Problemstellung) enthält für die Bearbeitung einer Anfrage oder das Lösen eines Problems. Es vereinfacht zudem die Verwaltung der Fälle, da alles zentral dokumentiert werden kann.

Transdisziplinäre Forschung Interdisziplinarität bedeutet, dass mindestens zwei verschiedene, für das Problem relevante, Disziplinen zusammenarbeiten und die Erkenntnisse aus ihrer Disziplin zusammenfügen und gemeinsam das Problem lösen. Bei der Transdisziplinarität wird hierzu noch die Expertise aus der Industrie miteinbezogen, wobei die Problemlösung durch die Zusammenarbeit von Akademie und Praxis erarbeitet wird.

Transformative Kreativität Transformative Kreativität steht für die Fähigkeit, etwas zu denken oder zu kreieren, was auf Grundlage der dem Menschen bekannten konzeptuellen Muster eigentlich nicht gedacht oder kreiert werden könnte. Bestehende Strukturen werden aufgerüttelt und Bestehendes wird durch grundlegend Neues ausgewechselt.

Transportmatrix Die Transportmatrix gibt Aufschluss darüber, zu welchem Zeitpunkt welche Menge von Material von verschiedenen Startstationen zu Zielstationen transportiert werden muss. Aus diesen Informationen können wichtige Anforderungen (z. B. Auslastung einzelner Weg, Notwendigkeit von Alternativwegen) an ein Wegenetz definiert werden.

Uncanny Valley Uncanny Valley beziehungsweise das unheimliche Tal stellt ein beunruhigendes Gefühl dar, welches Menschen erleben, wenn sie mit humanoiden Robotern oder menschenähnlichen Simulationen zu tun haben, die in vielerlei Hinsicht dem Menschen sehr ähneln, aber dennoch nicht ganz überzeugend realistisch sind.

User Experience Die User Experience umfasst die gesamten emotionalen, kognitiven und physischen Erlebnisse, die ein Nutzer in der Interaktion mit einem Produkt oder Dienstleistung eines Unternehmens gemacht hat. Hierbei gilt: positive Erfahrungen führen zu einer guten User Experience.

User Help Desk Der User Help Desk der Informatikabteilung der Schweizerischen Post ist die zentrale Anlaufstelle für Mitarbeitende bei Anfragen oder Problemen in Bereich der Hard- und Software sowie der IT Services. Dieser kann über Telefon, E-Mail und Webformular und in drei Sprachen kontaktiert werden.

Virtual Reality Mittels Computertechnologie wird eine simulierte Umgebung geschaffen. Im Vergleich zu den herkömmlichen Benutzeroberflächen, in welchen der Benutzer die Programme beispielsweise über den Bildschirm abruft, taucht hier der Benutzer in die geschaffene Simulation ein und kann mit der 3D-Welt interagieren.

Web Crawler Ein Web Crawler ist ein Programm oder ein automatisiertes Skript, welches das Internet methodisch, automatisiert und mit vordefinierten Mustern durchsucht und somit die Informationssammlung erleichtert.

Wegenetz Wegenetze sind vorgegebene Routen, auf denen sich Fahrerlose Transportfahrzeuge bewegen. Sie können physisch (z. B. durch Dauermagnete, Farbstreifen) auf dem Boden angebracht oder digital in Leitrechnern hinterlegt sein. Neben baulichen Verordnungen (bspw. Sicherheitsabständen) ist für ein optimales Wegenetz auch die Beachtung der möglichen Auslastung und die Vermeidung von Deadlocks wichtig.

Well-structured problem Als *well-structured problems* (wohl strukturierte Probleme) bezeichnet man Probleme, die sich dadurch auszeichnen, dass die Ausgangslage klar und beschreibbar ist, dass der formelle Rahmen des Problems abgegrenzt werden kann und dass sich die Validität einer Lösung anhand bekannter Kriterien testen lässt. Zudem lassen sich bei *well-structured problems* die notwendigen Wissensgrundlagen annähernd abschliessend erfassen und die zulässigen Transformationsmechanismen von einem intermediären Zustand der Problemlösung in einen anderen (also die formellen Lösungsregeln) sind bekannt.

Wissen Wissen ist im weiteren Sinne epistemologisch zu verstehen und meint also nicht nur die Kenntnis etwa von Fakten und Informationen, sondern bezieht die über wahrgenommene Empfindungen erreichten, nicht in Sprache auszudrückenden Erkenntnisse explizit mit ein.



Organ der Gesellschaft für Informatik
e.V. und mit ihr assoziierter
Organisationen

Haupterausgeber: Thomas Ludwig
Chefredakteur: Peter Pagel

ISSN: 0170-6012 (gedruckte Version)

ISSN: 1432-122X (elektronische Version)



Informatik Spektrum

Hauptaufgabe dieser Zeitschrift ist die Publikation wissenschaftlicher Ergebnisse und aktueller, praktisch verwertbarer Informationen über technische und wissenschaftliche Fortschritte aus allen Bereichen der Informatik und ihrer Anwendungen in Form von Übersichtsartikeln und einführenden Darstellungen sowie Berichten über Projekte und Fallstudien, die zukünftige Trends aufzeigen.

Sie spricht alle Leser an, die sich in neue Sachgebiete der Informatik einarbeiten, sich weiterbilden oder sich einen Überblick verschaffen wollen. Damit wendet sie sich nicht nur an ausgebildete Informatikspezialisten, sondern auch an Praktiker, die neben ihrer Tagesarbeit die wissenschaftliche Entwicklung der Informatik verfolgen, sowie an Studierende an Hochschulen oder Universitäten, die sich Einblick in Aufgaben und Probleme der Praxis verschaffen möchten.