

Christian Gärtner

Smart HRM

Digitale Tools für die Personalarbeit



Springer Gabler

Smart HRM

Christian Gärtner

Smart HRM

Digitale Tools für die Personalarbeit

Christian Gärtner
Professur für Human Resource Management, Arbeitspsychologie und Digitalisierung
der Arbeitswelt
Hochschule München
München, Deutschland

ISBN 978-3-658-29430-4 ISBN 978-3-658-29431-1 (eBook)
<https://doi.org/10.1007/978-3-658-29431-1>

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

Springer Gabler

© 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 Gabler 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

Vorwort

Seit einigen Jahren wird viel über Künstliche Intelligenz, Maschinelles Lernen, Robotic Process Automation und Chatbots gesprochen und geschrieben. Durch die Digitalisierung von Prozessen und die Verschmelzung der physischen mit der digitalen (Arbeits-)Welt generieren wir alle ständig Daten. Klar ist deshalb, dass sich die Arbeit der Personaler durch Analytics und Automatisierung bereits verändert hat und noch weiter verändern wird. Noch offen ist, ob dies dazu führt, dass Personalarbeit den Blick auf das verliert, was die einzelnen Mitarbeitenden und – noch wichtiger – deren Zusammenarbeit ausmacht.

Statt einer Daten- und Technikgläubigkeit zu verfallen oder Angeboten Dritter blind zu vertrauen, sollten HR-Professionals und HR-Studierende¹ verstehen, wie die digitalen Tools funktionieren, um entscheiden zu können, was davon im eigenen Umfeld gewinnbringend eingesetzt werden kann. Nur dann kann Personalarbeit wirklich „smart“ werden. Dazu will ich mit diesem Buch beitragen. Auch wenn in diesem Buch der Fokus auf den harten Fakten und Tools der digitalen Personalarbeit liegt, so ist doch klar, dass in der Personalarbeit auch „Softes“ – Interpretation, Gespür, Hartnäckigkeit und vieles mehr – gefragt ist. Schon allein, weil es das Ideal, Personalarbeit nur auf Basis objektiver Daten zu betreiben, nicht gibt und nicht geben kann. Denn in der Praxis werden Fakten geschaffen (und nicht einfach Ereignisse in Daten abgebildet) und es muss angesichts knapper Ressourcen gehandelt werden, wobei widerstreitende Interessen aufeinandertreffen.

Beim Schreiben eines Textes entstehen immer die Spannungsverhältnisse zwischen Breite (möglichst viele digitale Tools in möglichst allen Personalteilstruk-

¹Zur einfacheren Lesbarkeit wird nur ein grammatisches Geschlecht (oder der Plural) verwendet. Alle personenbezogenen Bezeichnungen sind geschlechtsneutral zu verstehen.

nen) und Tiefe (möglichst genaue Erläuterungen der Technologie und ihrer Auswirkungen) sowie Theorie (Modelle, Forschungsergebnisse) und Praxis (Beispiele von und in Unternehmen), die ich versucht habe, auszubalancieren – so gut es eben im Rahmen eines(!) Buches geht. Parallel dazu entstand ein Herausgeberwerk namens „Smart Human Resource Management“, zu dem mehrheitlich Praktiker beigetragen haben und das somit punktuell in die Tiefe geht, weil den Beispielen mehr Platz eingeräumt wird. Zudem gibt es auf der Website www.smarthrm.de weitere Informationen (z. B. eine Liste mit mehr als 250 HR Start-ups, geclustert nach ihren Angeboten).

Auch die Arbeit an dem vorliegenden Buch war keine Einzelleistung, schon allein, weil ich mich auf viele Vorarbeiten anderer stützen konnte. Eine weitere Unterstützung war Jacqueline Kadlec, die das Manuskript Korrektur gelesen und auch inhaltliche Inputs geliefert hat – vielen Dank dafür!

Frankfurt am Main
Januar 2020

Christian Gärtner

Inhaltsverzeichnis

1	Smart HRM: Digitale Tools für die Personalarbeit	1
2	Smarte oder unmenschliche Personalarbeit?	5
2.1	Big Brother oder Big Botler?	7
2.2	Smart HRM: Hart, aber auch herzlich	11
	Literatur	14
3	Grundlagen: KI, ML, DL, RPA und Co.	17
3.1	Algorithmen und Heuristiken	17
3.2	Big Data	18
3.3	Künstliche Intelligenz (KI)	20
3.4	Machine Learning (ML)	22
3.5	Künstliche Neuronale Netze (KNN)	30
3.6	Analytics: Descriptive, Diagnostic, Predictive und Prescriptive	33
3.7	Automatisierung und Robotic Process Automation (RPA)	37
3.8	Intelligent Process Automation und Chatbots	41
3.9	Virtual und Augmented Reality	44
3.10	Fazit	45
	Literatur	48
4	Digitale Tools für smarte Personalarbeit	51
4.1	Use Cases entlang der HR-Wertschöpfungskette	51
4.2	Personalmarketing & -rekrutierung	54
4.2.1	Sentiment-Analyse des Arbeitgeberimages	54
4.2.2	Robo-Recruiting: Von (Active) Sourcing über Matching bis Selecting	65

4.3	Personalplanung & -einsatz	89
4.3.1	Strategische Personalplanung	89
4.3.2	Personaleinsatz	94
4.4	Performance Management	99
4.4.1	Ziel- und Feedbacksysteme	100
4.4.2	Performance Management mit Workplace Analytics	117
4.5	Personalentwicklung	132
4.5.1	Learning Analytics & Learning Management Systeme: Personalisiertes, adaptives und selbstorganisiertes Lernen	135
4.5.2	Wearables, Augmented und Virtual Reality für personalisiertes und adaptives Lernen	149
4.6	Personalbindung und -freisetzung	160
4.6.1	Abschätzung der Kündigungswahrscheinlichkeit	160
4.6.2	Nachfolgeplanung	172
	Literatur	175
5	Unterstützung und Ersetzung durch RPA	191
5.1	Nutzenversprechen und Anwendungskriterien	192
5.2	Recruiting, On- und Offboarding	197
5.3	Personalplanung: Stammdatenmanagement und Reporting	199
5.4	Personalbetreuung	200
5.5	Gehaltsabrechnung, Reisekostenabrechnungs-Prozess	201
	Literatur	204
6	Ethische und gesetzliche Richtlinien für Smart HR	207
6.1	Zentrale gesetzliche Vorgaben	208
6.1.1	DSGVO und BDSG: Transparenz und Informationspflicht, Zweckbindung und Datenminimierung	211
6.1.2	AGG: Diskriminierung	215
6.2	Ethikrichtlinien	217
6.3	Technische und organisatorische Vorkehrungen zur Lösung ethisch-rechtlicher Probleme	225
	Literatur	229
7	Smart HRM: Wohin geht die Reise?	233
	Literatur	236

Abbildungsverzeichnis

Abb. 2.1	Faktoren, die die Akzeptanz von automatisierten Überwachungs- und Entscheidungssystemen beeinflussen	9
Abb. 3.1	Smiley als Schwarz-Weiß-Pixel, die in Zahlen umgewandelt werden.	22
Abb. 3.2	Multiple Regression, schematisch für mögliche Einflussfaktoren auf Arbeitszufriedenheit	27
Abb. 3.3	Logistische Regression, schematisch für mögliche Einflussfaktoren auf Eigenkündigung.	28
Abb. 3.4	Entscheidungsbaum zur Abschätzung des Kündigungsrisikos (vereinfacht, eigene Erstellung)	29
Abb. 3.5	Künstliches Neuronales Netz zur Bilderkennung durch Kategorisierung von Pixeln (vereinfacht)	32
Abb. 3.6	Deep Learning zur Bilderkennung am Beispiel einer Zahl (schematisch und fiktiv)	32
Abb. 3.7	Analytics-Stufen	33
Abb. 3.8	Vorgehensmodell bei People Analytics.	36
Abb. 3.9	Stufen der Automatisierung	38
Abb. 3.10	Funktionslogik von Robotic Process Automation.	39
Abb. 3.11	Intelligent Process Automation (schematisch)	42
Abb. 4.1	Anwendungsbeispiele entlang der HR-Wertkette.	52
Abb. 4.2	Pre-Processing bei Sentiment-Analysen.	56
Abb. 4.3	Text-Mining mit RapidMiner: Operatoren innerhalb der Funktion „Process Documents“	57
Abb. 4.4	Frage-Antwort-Dialog im Chatbot von Jobpal	72

Abb. 4.5	Treiberanalyse („Impact-Grid“) von functionHR	110
Abb. 4.6	Microsoft Office365 Graph	119
Abb. 4.7	MotionMiners: Hardware-Ausrüstung für die Prozessmessung	126
Abb. 4.8	Entscheidungsbaum zur Kündigungsvorhersage	163
Abb. 4.9	Abschätzung des Kündigungsrisikos und der -kosten (Predict42)	166
Abb. 5.1	Nutzenversprechen von RPA	192
Abb. 5.2	Automatisierter Qualitätskontrollprozess für Payroll-Daten	203

Tabellenverzeichnis

Tab. 3.1	Typische Fragen für Machine-Learning-Algorithmen	26
Tab. 3.2	Eignung ausgewählter ML-Algorithmen	46
Tab. 4.1	Performance Management im Wandel	101
Tab. 5.1	Anwendungskriterien für RPA	194
Tab. 5.2	Scoring-Modell zur Auswahl von Prozessen für RPA	195
Tab. 6.1	Zentrale ethische und rechtliche Fragestellungen beim Einsatz von KI/ADM-Systemen	209

Smart HRM: Digitale Tools für die Personalarbeit

1

Zusammenfassung

Für die Beschäftigung mit Smart HRM – dem intelligenten Einsatz von digitalen Tools zur dateninspirierten und automatisierten Personalarbeit – gibt es eine Reihe von Gründen: Innovationsargument, Produktivitätsargument, Spiegelungsargument, Legitimationsargument, Objektivitäts- bzw. Neutralitätsargument und Erkenntnisargument. Der Einsatz digitaler Tools in der Personalarbeit ist deshalb keine Frage mehr des „Ob“, sondern des „Wo“ und „Wie“: In welchen Teilbereichen der Personalarbeit lohnen sich digitale Tools? Wie funktionieren sie, was kann man mit ihnen machen kann und was geht (noch) nicht?

Smart HRM meint den intelligenten Einsatz von digitalen Tools in der Personalarbeit, insbesondere in den Bereichen **Analytics und Automatisierung**. Digitale Tools umfassen ein breites Sortiment an Instrumenten, die aus Hard- und Software bestehen und zur Lösung personalwirtschaftlicher Probleme genutzt werden. Dass gute Personalarbeit nicht nur aus Tools, sondern auch aus innovativen Ideen, überzeugenden Strategien, durchdachten Konzepten, klaren Prozessen und der professionellen Umsetzung dieser besteht, ist klar und soll nicht bezweifelt werden. In diesem Buch stehen aber die Tools im Vordergrund. Sie sind Hilfsmittel, um die Personalarbeit leichter und besser zu machen. Leichter, weil sie Arbeitsschritte automatisieren und besser, weil mit ihrer Hilfe unternehmensspezifische Daten

analysiert und personalrelevante Fragestellungen so beantwortet werden können, dass sie auf den jeweiligen Kontext zugeschnitten sind. **Die Fragen berühren alle Teilfunktionen des Personalmanagements**, z. B.:

- Wie ist das zu erwartende Angebot auf dem Arbeitsmarkt für bestimmte Job-Profile?
- Wie wird die Arbeitgebermarke wahrgenommen?
- Welche Personalmarketing-Maßnahmen führen dazu, dass wir als Firma die passendsten Bewerber bekommen (z. B.: Wo müssen wir wann Stellenanzeigen schalten, um die meisten Klicks von passenden Kandidaten zu bekommen?)?
- Welche Jobs passen zu einer Kandidatin und sollten ihr deshalb vorgeschlagen werden?
- Welche Bewerber werden später am ehesten erfolgreich sein und welche werden am ehesten wieder kündigen?
- Wie können automatisierte Dialogsysteme (Chatbots) im Recruiting eingesetzt werden?
- Was macht gute Führung aus?
- Wer eignet sich als Führungskraft und sollte befördert werden?
- Was macht erfolgreiche Mitarbeiter oder Teams aus?
- Wie hoch ist der Einfluss guter Führung auf die Arbeitsleistung von Mitarbeitern oder Teams?
- Welchen Einfluss haben Kommunikationsströme auf die Leistung eines Teams?
- Was sind die wichtigsten Treiber der Arbeitszufriedenheit in unserem Unternehmen?
- Wie stark wirken sich das Gehalt, die wöchentliche Arbeitszeit, die Aufstiegschancen oder andere Anreize auf die Arbeitszufriedenheit aus?
- Welche Trainingsinhalte ermöglichen den größten Lernerfolg?
- Wie können Chatbots den Wissenstransfer unterstützen?
- Welche Maßnahmen bereiten ein Talent am ehesten auf den nächsten Karriere-schritt vor?
- Wie wird die zukünftige Arbeitsleistung bei der Nachbesetzung einer Stelle sein?
- Welche Verhaltensweisen beim Lernen mit Online-Inhalten führen zum höchsten Lernerfolg?
- Welche Mitarbeitergruppen verlassen unsere Firma am ehesten (z. B.: Welche Personen- oder Verhaltensmerkmale ermöglichen die beste Vorhersage der Kündigungswahrscheinlichkeit?)?
- Wo im Unternehmen sitzen Schlüsselpersonen, die am meisten für den Wissensaustausch tun?

- Welche Schlüsselpersonen kündigen mit hoher Wahrscheinlichkeit und welche Kosten verursacht das?
- Welche Merkmale sagen am besten vorher, ob Arbeitnehmer, die Elternzeit in Anspruch nehmen, anschließend wieder zurückkehren?
- Welche Personengruppen werden zu welchen Zeitpunkten Fehlzeiten aufweisen?
- Wie hoch sind die Auswirkungen von Fehlzeiten auf die Kundenzufriedenheit?
- Wie können wiederkehrende Tätigkeiten möglichst schnell, kostengünstig und mit gleichbleibender Qualität abgewickelt werden?

Wem diese Liste noch nicht als Anreiz genügt, sich über digitale Tools in der Personalarbeit Gedanken zu machen, der lässt sich vielleicht von den folgenden Argumenten überzeugen. Eine Beschäftigung ist sinnvoll,

- ... weil der Einsatz von digitalen Tools im HR zu innovativen Produkten, Dienstleistungen, Prozessen oder Geschäftsmodellen führen kann. Die Arbeit mit digitalen Tools ist dann – in Anlehnung an das Bonmot Bismarcks – die Kunst des Möglichen, wobei Kunst nicht nur die Kunstfertigkeit meint, sondern auch die Verschiebung der Grenzen des Möglichen (**Innovationsargument**).
- ... weil diejenigen, die digitale Tools anwenden, bessere Ergebnisse oder zumindest erwünschte Wirkungen im Unternehmen erzielen können (**Produktivitätsargument**).
- ... weil, wenn sich der Rest der Organisation der Digitalisierung verschreibt, HR dies spiegeln und ebenfalls den Einsatz digitaler Tools prüfen sollte (**Spiegelungargument**).
- ... weil sich, wenn viele oder alle digitale Tools in der Personalarbeit einsetzen, vor allem jene rechtfertigen müssen, die nicht mit dem Trend gehen. Sie sind unter dem Druck, potenziellen Bewerberinnen, aktuellen Mitarbeitern und anderen Stakeholdern zu erklären, warum sie sich verweigern und warum sie (dennoch) moderne Personalarbeit betreiben (**Legitimationsargument**).
- ... weil sich durch die eingehende Beschäftigung mit den Tools objektiver auf die Möglichkeiten und Grenzen des Einsatzes technologischer Hilfsmittel blicken lässt und man meist erkennt, dass Technologie weder ein allumfassender Heilsbringer noch unnützer Firlefanz ist (**Objektivitäts- bzw. Neutralitätsargument**).
- ... weil nicht ganz klar ist, was gute Personalarbeit oder Zusammenarbeit in Unternehmen wirklich ausmacht – wäre es so, bräuchte es nicht tausende Bücher, Vorträge, Best-Practices-Workshops usw. Angesichts dieser Erkenntnislücke können uns Verfahren aus dem Bereich der Künstlichen Intelligenz lehren, die richtigen Fragen über intelligentes bzw. smartes Handeln zu stellen. Spätestens, wenn wir die Tools anwenden, sind wir gezwungen, gründlich darüber nachzudenken, was wirklich wichtig und richtig ist (**Erkenntnisargument**).

Ziel des vorliegenden Buches ist es zu zeigen, **welche Tools zur mittelbaren oder unmittelbaren Wertsteigerung** in der Personalarbeit eingesetzt werden und **wie sie dies tun**. Um die Funktionsweise zu verstehen, werden nach einem Kapitel mit einer allgemeinen Diskussion von Aspekten einer smarten Personalarbeit (Kap. 2) die wichtigsten Begriffe, Algorithmen und Technologien erläutert (Kap. 3). Auf diese greifen die Tools (Kap. 4) in unterschiedlichsten Kombinationen zurück, weshalb die allgemeine Erläuterung an den Beginn gestellt wurde und dann in den Unterkapiteln auf die entsprechenden Algorithmen bzw. Technologien verwiesen wird. Die Beschreibung ist jedoch keineswegs nur technisch angelegt, sondern auch im Hinblick auf den Aspekt der Wertsteigerung. Zusammengefasst kann HR in dreierlei Hinsicht einen wertschöpfenden Beitrag zum Unternehmen leisten:

1. Indem **neue HR-Produkte** angeboten werden, z. B. eine Echtzeit-Simulation der Kosten und Nutzen von unterschiedlichen Recruiting-Kampagnen und -Kanälen oder eine datenbasierte Analyse, wer tatsächlich Schlüsselpersonen sind, inklusive einer Berechnung der Wahrscheinlichkeit, dass sie die Organisation verlassen.
2. Durch die Erbringung der Personalarbeit in einer **noch nicht dagewesenen Qualität**, weil Roboter keine Fehler machen und 24 Stunden am Tag für Fragen von Bewerbern und Mitarbeitern zur Verfügung stehen.
3. Indem **Kosten eingespart** werden, z. B. durch die Automatisierung standardisierter und repetitiver Prozessschritte. Dass Arbeitsschritte automatisiert werden, führt nicht sofort zur Entlassung von Personal, weil nicht gleich ganze Jobs wegfallen, sondern nur einzelne Tätigkeiten – es bleibt genügend anderes zu tun. Z. B., sich zu überlegen, wie sich durch optimierte Personalsuche und -bindung Kosten noch ganz anders als durch Automatisierung sparen lassen oder Mitarbeiter begeistert werden können.

Eine quantitative **Wirtschaftlichkeitsrechnung muss immer organisations-spezifisch** sein, schließlich unterscheiden sich das Ausmaß des Problems, die Anzahl der auftretenden Fälle, die vorhandene IT-Landschaft, die Implementierungsdauer, die Akzeptanz neuer Technologien etc. von Organisation zu Organisation – und als Resultat dieser Faktoren auch die Kosten und möglichen Gewinne. Aus diesem Grund fokussiere ich mich bei der Beschreibung des Einsatzes von digitalen Tools in der Personalarbeit (Kap. 4) auf die ersten beiden Punkte zur Wertschöpfungssteigerung. Ob aus den Use Cases auch ein positiver Business Case wird, muss jede Organisation selbst herausfinden.

Smarte oder unmenschliche Personalarbeit?

2

Zusammenfassung

Der Einsatz von algorithmenbasierten Entscheidungssystemen ist ein zweischneidiges Schwert: Am Beispiel der TOT-Zeiten bei Amazon wird gezeigt, dass die automatisierte Überwachung und systemseitige Entscheidung einerseits zu Abwehr und Abscheu führt, andererseits aber innovative Lösungen ermöglicht. Der Grat zwischen Big Brother und Big Botler ist schmal. Dafür gibt es vier Gründe, die näher erörtert werden: (1) der Technikeinsatz ist inhaltlich ambivalent; (2) verschiedene Anspruchsgruppen legen in unterschiedlichen Situationen unterschiedliche Akzeptanzkriterien an die Systeme an; (3) die Art und (4) Weise der Einführung digitaler Tools macht einen Unterschied. Darüber hinaus wird diskutiert, für wen und wann sich der Einsatz digitaler Tools eignet, wobei vier Strategien beschrieben werden, durch die auch kleinere Organisationen von Analytics und Automatisierung profitieren können.

Beispiel: TOT-Zeiten bei Amazon

Im April 2019 wurde bekannt, dass Amazon nicht nur die Arbeit von Lagermitarbeitern mit einem Computersystem steuert, sondern diesen auch automatisch kündigt, wenn sie die vorgegebenen Leistungsziele wiederholt nicht erreichen (Lecher 2019). Die Leistung wird vor allem über die Anzahl gepackter Pakete und die Abwesenheit vom Arbeitsplatz bzw. Unterbrechung der Arbeit, z. B. auf-

grund von Pausen oder Toilettengängen, gemessen, die sogenannte „time off task“ (abgekürzt: TOT). Beispielsweise versendet das System automatisch Warnungen bei zu hohen TOT-Zeiten und die 5 % der Belegschaft mit den niedrigsten Leistungsdaten werden direkt für ein Trainingsprogramm angemeldet – oder ihnen wird bei wiederholter Zielerfahrung gekündigt. Amazon hat bestanden, dass ein Bot die Kündigung ausstelle, und betont, dass immer noch Führungskräfte das letzte Wort hätten. Allerdings wurde ein Brief einer Amazon-Anwältin an das *National Labor Relations Board* öffentlich, in dem darauf verwiesen wird, dass allein in einem Lager in Baltimore im Zeitraum August 2017 bis September 2018 ca. 300 Angestellten automatisch gekündigt worden sei. Bei Amazon scheinen Entscheidungen, die klassischerweise von Führungskräften getroffen wurden (z. B. Arbeit verteilen, Arbeitsleistung bewerten und Kündigungen aussprechen), nun durch ein algorithmenbasiertes Entscheidungssystem ersetzt worden zu sein.

Was an diesem Beispiel verdient den Namen „Smart HRM“, ist also clever, schlau, intelligent oder klug? Bevor diese Frage beantwortet werden kann, muss differenziert werden, wie sich sogenannte **algorithmenbasierte Entscheidungssysteme (ADM = Algorithmic Decision Making)** auf Arbeitsabläufe auswirken können. Dabei gibt es vier Varianten: (1) sie können die menschliche Arbeitskraft unterstützen, (2) sie substituieren, (3) sie zum Objekt degradieren und/oder (4) innovative Abläufe und Ergebnisse ermöglichen. Im Folgenden werden die vier Aspekte genauer erläutert.

Zu (1): Offensichtlich steuert und optimiert Amazon die Arbeit in den Versandlagern datenbasiert und weitgehend automatisiert. Glaubt man der Amazon-Darstellung, so werden die Führungskräfte nur in ihrer **Tätigkeit unterstützt**, weil ihnen Daten über die Produktivität zur Verfügung gestellt werden.

Zu (2): Wenn Amazon-Lagermitarbeiter auf Basis von Datenanalysen sowie Regeln zur Arbeitsverteilung und Leistungsmessung automatisch verplant, bewertet und entlassen werden, dann **substituiert** das ADM-System Aufgaben von Vorgesetzten.

Zu (3): Die TOT-Kennzahl ist oft kritisiert worden, weil sie Mitarbeiter dazu nötige, sich notwendige Pausen zu verkneifen (Lecher 2019). Amazon behandle die Lagermitarbeiter wie Ressourcen bzw. Objekte, die für die Erreichung wirtschaftlicher Ziele optimiert werden. In den Worten Neubergers (1990): Der **Mensch ist nicht Mittelpunkt, sondern Mittel. Punkt.**

Zu (4): Weniger offensichtlich ist, dass es genau diese amoralischen Praktiken sind, die jene agile Warenlieferungen erst ermöglichen, die Millionen von Amazon-Kunden schätzen und den Kern von Amazons **innovativem Geschäftsmodell** bilden. Perfektioniert im wirtschaftlichen Sinne ist die Logistik erst durch das

optimale Zusammenspiel aus datenbasierter und automatisierter Leistungssteuerung einerseits und Behandlung der Mitarbeiter als bloßes Produktionsmittel andererseits.

Gerade die Punkte 2 und 3 stehen der positiven Konnotation, mit der das Wörtchen „smart“ im Sinne von clever, schlau, intelligent, usw. verbunden ist, entgegen – zumindest aus einer humanistischen Perspektive bzw. wenn die Auffassung vertreten wird, dass Unternehmen mehr bieten sollten, als die optimierte Bearbeitung von Aufgaben unter rein wirtschaftlichen Gesichtspunkten. Lässt man diese normative Forderung außen vor, erschließt sich, dass Substitution (Punkt 2) und Unterwerfung unter die Zwänge des Produktionsprozesses (Punkt 3) notwendig sind, um profitorientierte Organisationen an Veränderungen anzupassen und ihren Fortbestand auf Dauer zu sichern. Smartes Handeln umfasst nicht nur die positiv-konnotierten Facetten wie **intelligent und clever, sondern auch raffiniert, gerissen und scharfsinnig** – was im Übrigen weitere Übersetzungen von „smart“ sind. Im Unternehmenskontext ist es ebenso scharf wie sinnig, sowohl über die Ersetzung von Managern und Mitarbeitern nachzudenken als auch über deren Rolle als Mittel zur Erfüllung der Unternehmensziele.

Um über den Einzelfall hinaus beurteilen zu können, ob datenbasierte und automatische Entscheidungen bzw. Überwachung smarte oder unmenschliche Ergebnisse hervorbringen, hilft ein Blick auf einschlägige Forschungsergebnisse. Diese werden in den Unterkapiteln zu den jeweiligen Personalteilfunktionen herangezogen. Im Folgenden soll ein allgemeiner Überblick zum Wohl und Wehe von ADM gegeben werden.

2.1 Big Brother oder Big Botler?

Es gibt eine Reihe von wissenschaftlichen Erkenntnissen zu den Auswirkungen elektronischer Überwachung und automatisierten Personalentscheidungen, die sich wie folgt grob zusammenfassen lassen: Bei den Betroffenen stoßen solche Systeme meist auf eine **geringe Akzeptanz und führen zu kognitiv-emotionalen Belastungen**. Beispielsweise hat Backhaus (2019, S. 8 ff.) in seiner Auswertung von 85 Studien gezeigt, dass die negativen Effekte einer elektronischen Überwachung eher gering sind, wobei es v. a. negative Auswirkungen auf das subjektive Empfinden gibt (z. B. Stress, wahrgenommene Kontrolle, Zufriedenheit). Gleichzeitig steigen aber **Motivation, die Arbeitsleistung und Entscheidungsqualität**. Manche Studien berichten auch von einer gestiegenen Mitarbeiterzufriedenheit, wenn und weil die Mitarbeiter in den datenanalytisch informierten Entscheidungen den Versuch sehen, die Arbeitsbedingungen möglichst **objektiv und optimal** zu

gestalten (Strohmeier et al. 2016, S. 843; s. a. Cheng und Hackett 2019; Tambe et al. 2019). Kritiker sehen in der dauerhaften Messung am Arbeitsplatz (als Verlängerung der „quantified self“-Bewegung im privaten oder halb-öffentlichen Raum) einen **Rückschritt in tayloristische Zeiten** (z. B. Moore und Robinson 2016; Lemov 2018). Befürworter sehen darin die Möglichkeit, eine Physik des Sozialen zu entwerfen (z. B. Pentland 2012, 2014). Letztlich malen manche ein sehr positives Bild der „neuen Arbeitswelt“, weil sich einzelne Talente aufgrund ihrer Kompetenzen und neuer digitaler Services (von mobiler Weiterbildung bis zu Cognitive Personal Assistants) individueller entwickeln können, während andere das düstere Szenario einer weit umgreifenden Ausbeutung und eines globalen Pekariats sehen (Spreitzer et al. 2017).

Genau wie es von wissenschaftlicher Seite unterschiedliche Positionen gibt, sind auch auf Seiten der Praktiker die Meinungen über Wohl und Wehe digitaler Tools im HR-Bereich geteilt. Bezeichnenderweise trägt der Statusbericht des Bundesverbandes der Personalmanager (BPM 2019) den Titel: „**Zwischen Euphorie und Skepsis: KI in der Personalarbeit**“. Einerseits geht die Befürchtung um, der Mensch stehe nicht mehr im Mittelpunkt und müsse den Verlust des Arbeitsplatzes fürchten. Auch gebe es keine Fairness und kein Verantwortungsbewusstsein für (Management-)Entscheidungen mehr. Andererseits sehen die Praktiker viele Vorteile, angefangen bei kostengünstigeren Prozessen, die zudem Compliance-konform und mit hoher Qualität ablaufen, über neue Produkte bis hin zu einer verbesserten Candidate Experience. Was sowohl in der Studie des BPM als auch in der Umfrage der Deutschen Gesellschaft für Personalführung (DGFP 2019) auf-taucht: selbst, wer überwiegend positive Effekte erwartet, setzt keine Künstliche Intelligenz ein – insgesamt bejahen nur 3 %, dass KI bereits in HR eingesetzt würde (DGFP 2019, S. 7). Dazu passt die Befragung von knapp 200 Unternehmen der Top 1000 und Top 300 IT-Unternehmen im Rahmen der „Recruiting Trends“-Studie: ca. zwei Drittel meinen, dass eine automatisierte Vorauswahl von Bewerbungen eine diskriminierungsfreie Bewerberauswahl fördert – aber nur knapp 6 % machen davon Gebrauch (Laumer et al. 2019, S. 13 f.).

Es gibt **vier Gründe für die gemischte Daten- und Erkenntnislage** über die Vor- und Nachteile beim Einsatz digitaler Tools: (1) der Technikeinsatz ist inhaltlich ambivalent; (2) verschiedene Anspruchsgruppen legen in unterschiedlichen Situationen unterschiedliche Akzeptanzkriterien an die Systeme an; (3) die Art und (4) Weise der Einführung macht einen Unterschied (siehe Abb. 2.1).

1. Zunächst ist Überwachung genau wie automatisierte Entscheidungsfindung ein zweischneidiges Schwert: beides kann sowohl zum Schutz der Mitarbeiter als auch zur repressiven Kontrolle und ethisch fragwürdigen Optimierung

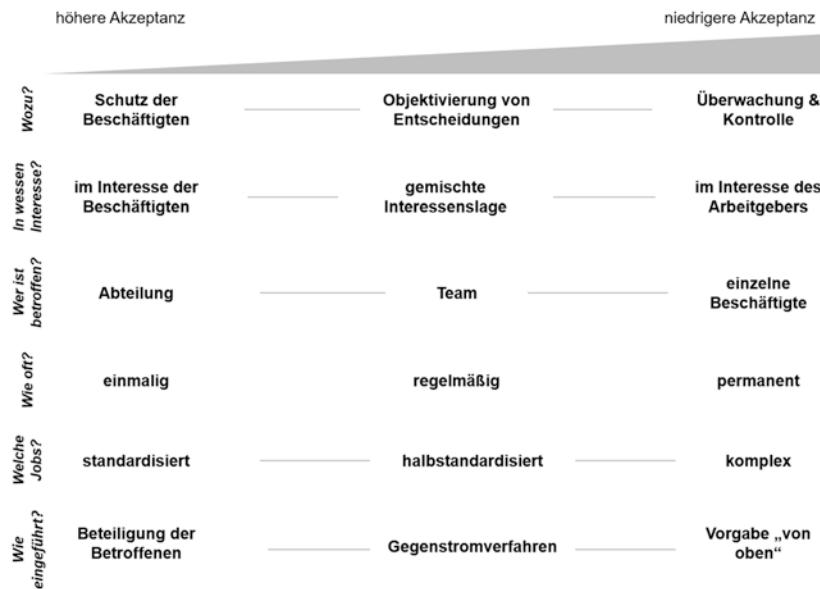


Abb. 2.1 Faktoren, die die Akzeptanz von automatisierten Überwachungs- und Entscheidungssystemen beeinflussen

eingesetzt werden. Positiv gewendet kann die Fürsorgepflicht des Arbeitgebers unterstützt werden, wenn stressauslösende Faktoren, ergonomische Fehlbelastungen oder Gefahren am Arbeitsplatz entdeckt und entfernt oder zumindest kompensiert werden (Backhaus 2019, S. 4). Die Arbeitnehmerseite kann der digitalen Überwachung und Steuerung Positives abgewinnen, wenn eine Objektivierung der Leistungserfassung bzw. -bewertung und personalbezogener Entscheidungen zu erwarten ist, beispielsweise indem Arbeitszeiten genauer erhoben werden (weil E-Mails, die abends geschrieben werden, nun auch als Arbeitszeit erfasst werden) oder subjektive Verzerrungen der Vorgesetzten bei Bewertung und Beförderung reduziert werden (Ajunwa 2019; Knobloch und Hustedt 2019).

2. Arbeitgeber und Arbeitnehmer sind nur zwei der zahlreichen Gruppen, die ein Interesse am Einsatz von ADM-Systemen bzw. deren Verhinderung haben: Allein im Bereich Personalauswahl müssten die Ansprüche von Kandidaten, Recruitern, Psychologen und Computerspezialisten geprüft werden, weil sie solche Systeme aller Voraussicht nach aus unterschiedlichen Gründen akzeptieren oder ablehnen (Liem et al. 2018). Psychologen achten z. B. eher auf die pro-

- gnostische Validität von Auswahlverfahren, während Recruiter auch die Kosten und einfache Handhabbarkeit berücksichtigen müssen. Kandidaten wiederum gewichten die automatische und schnelle Auswertung geringer als Recruiter und wollen dafür einen möglichst einfachen Zugang, z. B. zu einem Auswahltest. Recruiter hätten wahrscheinlich gern ein Verfahren, bei dem sie den Ablauf und das Ergebnis gut kontrollieren können (z. B. über einen automatisierten Test), was bei den Kandidaten aber das Gefühl hervorrufen kann, dem ‚System‘ ausgeliefert zu sein. Die Ansprüche variieren mit der Entscheidungssituation, z. B. je nachdem, ob die System-Nutzer viel oder wenig verlieren bzw. gewinnen können.
3. Mit den Interessen verbunden ist die Art bzw. Ausprägung der Überwachung und Entscheidungsautomatisierung. Es macht einen Unterschied, ob dies permanent oder nur stichpunktartig geschieht, ob alle oder nur einzelne betroffen sind, und ob jeder einzelne Arbeitsschritt im Detail oder nur Arbeitsergebnisse automatisch gemessen, bewertet bzw. entschieden werden (Backhaus 2019, S. 4 f.). Des Weiteren spielt es eine Rolle, ob die Beteiligten über einen längeren Zeitraum eine Beziehung aufbauen konnten (z. B. im Fall langjähriger Mitarbeiter-Vorgesetzten-Konstellationen) oder sich nur flüchtig kennen (z. B. im Fall eines Personalauswahlverfahrens oder in Arbeitsumgebungen, in denen hohe Fluktuation herrscht). Letztlich verweisen diese Faktoren auf die Charakteristika der Jobs, die automatisiert überwacht und gesteuert werden: Meist sind es relativ standardisierte Büro- und Dienstleistungstätigkeiten, z. B. in Call-Centern (mit typischerweise hoher Fluktuation) oder bei Crowd- und Clickworkern, bei denen automatisierte Überwachungs- und ADM-Systeme eingesetzt werden. Infolgedessen adressieren die meisten Studien auch nur diese Jobs, weshalb über die Auswirkungen bei wissensintensiven, komplexen Tätigkeiten mit unklarer Rollenverteilungen und Weisungsbefugnissen kaum gesicherte Erkenntnisse vorliegen.
 4. Neben der inhaltlichen Ambivalenz und der jeweils unterschiedlichen Beurteilung durch verschiedene Stakeholdergruppen ist auch die Einführung von Überwachungs- und ADM-Systemen differenziert zu beurteilen. So kann professionelles Change Management, das die Beschäftigten bei der Einführung des Systems beteiligt und sich um eine überzeugende Begründung der Einführung bemüht, negative Effekte kompensieren helfen (Backhaus 2019, S. 10; siehe auch DIN EN ISO 9241-210:2010).

Überwachung und automatisierte Entscheidungsfindung sind bei genauer Be trachtung also **mehrschneidige Schwerter**: Sie können nicht nur zum Schutz der Mitarbeiter oder der repressiven Kontrolle und ethisch-moralisch fragwürdigen Op

timierung eingesetzt werden, sondern auch in Ergänzung zu anderen verhaltenssteuernden Einrichtungen, die es in Organisationen nun einmal gibt (z. B. Arbeitsgesetze, Betriebsvereinbarungen, kulturelle Normen) und die das Wohl der Vielen auch zu Ungunsten des Einzelnen im Blick haben. Mit anderen Worten: Die Verhaltenssteuerung durch ADM kann auch Gutes bewirken, *indem* sie die Handlungsfreiheiten des Einzelnen einschränkt. Es ist wie beim Einsatz von Algorithmen in Autos: Die elektronische Abriegelung beim Überschreiten der zulässigen Höchstgeschwindigkeit mag eine Fahrerin als Einschränkung erleben, sie dient aber allen. Bei Alkoholzündschlossperren profitiert nicht nur die Allgemeinheit (die Trunkenheitsfahrt wird verhindert), sondern auch der Fahrer, und zwar *indem* seine Handlungsmöglichkeiten eingeschränkt werden, weil er nicht mehr mit dem Auto losfahren kann und dadurch aber seine Unversehrtheit sichert. Zur Verhaltenssteuerung hinzu kommt das Spannungsfeld zwischen Datensammlung und Datenschutz. Einerseits erfassen Arbeitgeber immer mehr Daten über ihre Mitarbeitenden (Personalstammdaten, Leistungsdaten, Gewohnheiten, Meinungsäußerungen, etc.), andererseits ist die Sensibilität für die informationelle Selbstbestimmung spätestens mit Verabschiedung der europäischen Datenschutzgrundverordnung gestiegen. Die wichtigen Fragen, ob und wie der Einsatz solcher Technologien ethisch, gesellschaftlich und rechtlich zu rechtfertigen ist, werden später erläutert (Kap. 6).

2.2 Smart HRM: Hart, aber auch herzlich

Für wen eignen sich die digitalen Tools? Es stimmt, dass ADM-Systeme, die Text-, Sprach- oder Bilddaten auswerten, **tausende Datensätze** benötigen, um überhaupt Muster erkennen und Empfehlungen geben zu können. Dabei ist die **Aufbereitung der Daten aufwändig**, da sie oft aus unterschiedlichen Quellen zusammengesucht, harmonisiert und bereinigt werden müssen (Schätzungen gehen davon aus, dass 85–95 % der Projektdauer für die Datenaufbereitung verwendet werden). Entsprechende Anwendungen eignen sich deshalb vor allem für große Organisationen – das denken zumindest viele. Es gibt aber vier Strategien, mit denen **auch kleinere Organisationen** von Analytics und Automatisierung profitieren können:

1. Man erkennt, dass für manche Problemstellungen weniger interne, sondern externe Daten relevant sind (z. B. Arbeitsmarktstatistiken und Stellenanzeigen bei Jobbörsen für die strategische Personalplanung oder Einträge auf Arbeitgeberbewertungsportalen sowie-Social Media-Posts zur Analyse des Arbeitgeberimages).

2. Man beschränkt die Analyse auf wenige unabhängige Variablen (z. B. 5) und setzt einfache Algorithmen ein (z. B. Regressionen). Dann liefern schon ein paar hundert Datenpunkte Ergebnisse mit statistischer Erklärungskraft (die Mindeststichprobengröße hängt vor allem von der Anzahl unabhängiger Variablen, deren Korrelation untereinander und der gewünschten Effektstärke ab; Maxwell 2000, S. 442 f.; s. a. Ratner 2017, S. 7 ff.).
3. Man greift auf wissenschaftliche Studien zum gleichen Problem zurück oder durchsucht veröffentlichte Daten bzw. Datenbanken (z. B. DBpedia, O*NET, Google Dataset Search, OpenML, UCI Machine Learning Repository).
4. Man verlässt sich darauf, dass die Daten und Algorithmen, auf denen die ADM-Systeme der externen Anbieter basieren, auch für den eigenen Kontext passen.

Doch Vorsicht: Wer „das Personal“ nur unter dem Brennglas der Analyse von Datensätzen einzelner (oder einer anonymen Vergleichsmasse) betrachtet, über sieht, dass Organisationen nicht aus der Summe vieler Einzelner bestehen, sondern dass die Mitarbeitenden ihren wettbewerbsstiftenden Mehrwert erst dann entfalten, wenn tatsächlich alle mitarbeiten und miteinander arbeiten. Es ist dieses Zusammenspiel der Mitarbeiter, die Theorien wie der Resource-based View oder der Relational View vor Augen haben, wenn vom Personal als strategisch wertvoller Resource gesprochen wird (Nyberg und Wright 2015). Die treffen in komplexen Entscheidungssituationen aufeinander. Allein die Tatsache, dass Personalentscheidungen einer oder mehrerer face-to-face-Gespräche bedürfen, führt zu einer hohen sozialpsychologischen Komplexität. Die physische Ko-Präsenz nötigt den Anwesenden Verhaltensweisen auf, die sich **kaum logisch, sondern nur psycho-logisch** verstehen lassen. Deshalb spielen beim Personalmanagement auch Faktoren wie Glaubwürdigkeit, Authentizität, Humor, Ironie und Emotionen eine wichtige Rolle (Gärtner 2017).

Diese Faktoren (automatisch) zu erfassen bzw. aus Daten herauszulesen schaffen auch Technologien des sogenannten „**Affective Computing**“ bzw. „Emotional Decoding“ nicht, trotz der jahrzehntelangen Arbeit an ihrer Entschlüsselung durch Wissenschaft und Wirtschaft (Picard 2019). Das liegt allein schon daran, dass bei der Emotionserkennung bisher nur auf die Sprache (z. B. IBM Watson Tone Analyzer) oder in die Gesichter (z. B. Affectiva, Pyramics) geschaut wird, menschliche Emotionen aber auch über Hände und Arme ausgedrückt werden (Gestik), sich in der Körperhaltung und -bewegung (Proxemik) zeigt und in der Stimmmodulation (Prosodik) niederschlägt. Genau an dieser **Multimodalität** arbeiten allerdings viele Forscher und die entsprechenden Tools machen Fortschritte (siehe Oviatt et al. 2019) – allerdings gibt es (noch) keine ausgereiften für die Personalarbeit.

Über die Multimodalität hinaus sind vor allem soziale Aspekte wie die Einbettung in eine Handlungssituation und Gemeinschaft oder Gesellschaft zu beachten, weil diese Erwartungen, Regeln und Sanktionen für den Emotionsausdruck und die Emotionsdechiffrierung bereithalten und durchsetzen. Diese Aspekte werden bisher von den Tools nicht berücksichtigt.

Statt Entscheidungen nur datengetrieben vorzunehmen und zu rechtfertigen, werden **individuelle Kreativität und Empathie noch lange Zeit relevant bleiben** – insbesondere bei komplexen Aufgaben, in denen passend zur jeweiligen Situation (re-)agiert werden muss. Davon gibt es bei der Personalarbeit jede Menge: HR-Professionals müssen kreative Personalmarketing-Kampagnen entwickeln und auch gegen Widerstand hartnäckig umsetzen. Sie müssen mit Kollegen aus unterschiedlichen Bereichen umsichtig abstimmen, welche Anforderungen wirklich gebraucht werden. Sie müssen bei Auswahlgesprächen die versteckten Interessen der Kandidaten entdecken und herausfinden, wie sie den Top-Talenten den Arbeitsplatz schmackhaft machen können. Sie müssen zusammen mit den Mitarbeitenden über nächste Karriereschritte nachdenken, passende Trainings finden und darüber entscheiden. Sie müssen improvisieren, wenn die IT versagt oder Daten nicht korrekt sind, usw. Die Verben und Adjektive zeigen an, dass der Alltag im (Personal-) Management von soften Einflussfaktoren geprägt ist: hartnäckig umsetzen, umsichtig abstimmen, entdecken, schmackhaft machen, etc. Tools können hier bestenfalls hilfreiche Informationen liefern, die dann im sozialen Miteinander eingesetzt werden. Nicht nur deshalb sind in der Personalarbeit **neben harten Daten auch das Herz, Gespür, Kreativität, Intuition, Mut und Entscheidungsfreude** gefragt. Dies ist auch der Fall, weil es das Ideal, Personalarbeit nur auf Basis objektiver Daten zu betreiben, nicht gibt und nicht geben kann, denn in der Praxis treffen widerstreitende Interessen aufeinander und es werden **Fakten geschaffen, nicht einfach in Datensätzen abgebildet** (Gärtner 2007, S. 30 ff.; Neuberger 2006). Aber selbst wenn man davon ausgeht, dass die Daten die Wirklichkeit abbilden, so bleibt das Problem, dass gesellschaftliche Verzerrungen abgebildet werden, die als sozial unerwünscht gelten (Abschn. 4.2.2.4). Letztlich gibt es in der Forschung unterschiedliche Ansätze, Wissen, Intelligenz und Weisheit zu fassen, aber aus der Kognitions- und Neurowissenschaft, der Robotik sowie der Philosophie des Geistes wissen wir: diese Leistungen können nicht auf die Verarbeitung von Informationen im Geist oder Computer reduziert werden, sondern sie sind an den menschlichen Körper gebunden (siehe Gärtner 2007, 2013).

Trotz aller Einschränkungen und Kritikpunkte, die für digitale Tools der Personalarbeit gelten, ist nicht abzusehen, wann und warum der Trend enden sollte. Ein Grund dafür sind die **Nachteile, die menschliche Entscheidungen haben: sie sind alles andere als rational**, weil in ihnen Vorurteile, kognitive Verzerrungen,

physische Limitationen, tradierte Regeln, Interessen und Machtkämpfe genauso eine Rolle spielen wie Moral, Ethik und Religion (Neuberger 2006). Es lohnt sich also, die neuen Möglichkeiten zu erörtern, nicht nur, um auf Höhe der Zeit zu bleiben, sondern um Lösungen zu finden, die mehr Vor- als Nachteile haben. Die begründete Vermutung ist, dass solche Lösungen die Anwendung digitaler Tools beinhaltet.

Literatur

- Ajunwa, I. (2019). Algorithms at work: Productivity monitoring applications and wearable technology as the new data-centric research agenda for employment and labor law. *Saint Louis University Law Journal*, 63, 21–53.
- Backhaus, N. (2019). Kontextsensitive Assistenzsysteme und Überwachung am Arbeitsplatz: Ein meta-analytisches Review zur Auswirkung elektronischer Überwachung auf Beschäftigte. *Zeitschrift für Arbeitswissenschaft*, 73(1), 2–22.
- BPM. (2019). Zwischen Euphorie und Skepsis: KI in der Personalarbeit. https://www.bpm.de/sites/default/files/bpm-service-24-ki_in_der_personalarbeit.pdf. Zugegriffen am 09.03.2019.
- Cheng, M. M., & Hackett, R. D. (2019). A critical review of algorithms in HRM: Definition, theory, and practice. *Human Resource Management Review (online first)*.
- DGFP. (2019). *Künstliche Intelligenz in HR*. https://www.dgfp.de/fileadmin/user_upload/DGFP_e.V/Medien/Publikationen/2019/Befragung-KI-in-HR_September-2019.pdf. Zugegriffen am 20.10.2019.
- Gärtner, C. (2007). *Innovationsmanagement als soziale Praxis: grundlagentheoretische Vorarbeiten zu einer Organisationstheorie des Neuen*. München/Mering: Rainer Hampp.
- Gärtner, C. (2013). Cognition, knowing and learning in the flesh: Six views on embodied knowing in organization studies. *Scandinavian Journal of Management*, 29(4), 338–352.
- Gärtner, C. (2017). Mensch oder Maschine: Wer trifft die besseren Personalentscheidungen? *HR Performance*, 2, 14–17.
- Knobloch, T., & Hustedt, C. (2019). *Der maschinelle Weg zum passenden Personal*. Gütersloh: Bertelsmann Stiftung.
- Laumer, S., Weitzel, T., & Luzar, K. (2019). Robo-Recruiting: Status quo und Herausforderungen für die KI in der Personalgewinnung. *Personal Quarterly*, 71(3), 10–15.
- Lecher, C. (2019). How Amazon automatically tracks and fires warehouse workers for ‚productivity‘. <https://www.theverge.com/2019/4/25/18516004/amazon-warehouse-fulfillment-centers-productivity-firing-terminations>. Zugegriffen am 01.07.2019.
- Lemov, R. (2018). Hawthorne’s renewal: Quantified total self. In P. Moore, M. Upchurch & X. Whittaker (Hrsg.), *Humans and machines at work* (S. 181–202). Cham: Palgrave Macmillan.
- Liem, C. C., Langer, M., Demetriou, A., Hiemstra, A. M., Wicaksana, A. S., Born, M. P., & König, C. J. (2018). Psychology meets machine learning: Interdisciplinary perspectives on algorithmic job candidate screening. In H. J. Escalante, S. Escalera, I. Guyon, X. Baró,

- Y. GüclüTürk, U. Güclü & M. van Gerven (Hrsg.), *Explainable and interpretable models in computer vision and machine learning* (S. 197–253). Cham: Springer.
- Maxwell, S. E. (2000). Sample size and multiple regression analysis. *Psychological Methods*, 5(4), 434–458.
- Moore, P., & Robinson, A. (2016). The quantified self: What counts in the neoliberal workplace. *New Media & Society*, 18(11), 2774–2792.
- Neuberger, O. (1990). Der Mensch ist Mittelpunkt. Der Mensch ist Mittel. Punkt. 8 Thesen zum Personalwesen. *Personalführung*, S. 3–10.
- Neuberger, O. (2006). *Mikropolitik und Moral in Organisationen*. Stuttgart: Lucius & Lucius.
- Nyberg, A. J., & Wright, P. M. (2015). 50 years of human capital research: Assessing what we know, exploring where we go. *The Academy of Management Perspectives*, 29(3), 287–295.
- Oviatt, S., Schuller, B., Cohen, P. R., Sonntag, D., Potamianos, G., & Krüger, A. (Hrsg.). (2019). *The handbook of multimodal-multisensor interfaces: Language processing, software, commercialization, and emerging directions – Volume 3*. New York: ACM/Morgan & Claypool.
- Pentland, A. (2012). The new science of building great teams. *Harvard Business Review*, 90(4), 60–69.
- Pentland, A. (2014). *Social physics: How good ideas spread-the lessons from a new science*. London: Penguin.
- Picard, R. (2019). Affective computing, emotion, privacy, and health. Artificial Intelligence Podcast. <https://www.youtube.com/watch?v=kq0VO1FqE6I>. Zugriffen am 05.07.2019.
- Ratner, B. (2017). *Statistical and machine-learning data mining: Techniques for better predictive modeling and analysis of big data* (3. Aufl.). Boca: Chapman & Hall/CRC.
- Spreitzer, G. M., Cameron, L., & Garrett, L. (2017). Alternative work arrangements: Two images of the new world of work. *Annual Review of Organizational Psychology and Organizational Behavior*, 4, 473–499.
- Strohmeier, S., Majstorovic, D., Piazza, F., & Theres, C. (2016). Smart HRM – das „Internet der Dinge“ im Personalmanagement. *HMD Praxis der Wirtschaftsinformatik*, 53(6), 838–850.
- Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial intelligence in human resources management: Challenges and a path forward. *California Management Review*, 61(4), 15–42.



Grundlagen: KI, ML, DL, RPA und Co.

3

Zusammenfassung

Da im Umfeld der digitalen Tools für die Personalarbeit sehr viele Begriffe verwendet werden, die Verwirrung stiften können, werden in diesem Kapitel die wichtigsten eingängig erläutert: Algorithmen und Heuristiken, Big Data, Künstliche Intelligenz, Machine Learning, Künstliche Neuronale Netze, Analytics (Descriptive, Diagnostic, Predictive und Prescriptive), Automatisierung und Robotic Process Automation, Intelligent Process Automation und Chatbots, Augmented und Virtual Reality. Insbesondere werden zentrale Machine-Learning-Algorithmen vorgestellt und ihre Anwendungsbedingungen sowie -grenzen benannt. Damit werden die Grundlagen gelegt, um besser verstehen zu können, was digitale Tools in der Personalarbeit können und was (noch) nicht.

3.1 Algorithmen und Heuristiken

Algorithmen sind Schritt-für-Schritt-Anleitungen, um ein mathematisch beschreibbares Problem in Einzelschritten zu lösen, d. h. sie beinhalten eine endliche Befehlsfolge zur Transformation eines Inputs mit abschließender Ausgabe eines Ergebnisses (Output). Als solches ist der Begriff des Algorithmus ein sehr alter und wird für vielerlei Handlungsvorschriften in unterschiedlichen Lebensbereichen benutzt: von Kochrezepten, die beispielsweise das Kuchenbacken in kleine Schritte

zerlegen und so Handlungen anleiten, über Vorschriften des Rechts (Gesetze), die ja auch menschliches Verhalten steuern sollen, bis hin zu Computeralgorithmen. Ihnen allen ist gemein, dass sie eine Verhaltenssteuerung bewirken (sollen). Was Computeralgorithmen von Kochrezepten und Gesetzen unterscheidet, ist, dass sie Probleme mathematisch exakt beschreiben, durch eine bestimmte Programmiersprache repräsentiert sind und der Programmcode auf einem Computersystem ausgeführt wird (Martini 2019, S. 17 f.). Außerdem liefert ein Algorithmus immer das gleiche Ergebnis, was beim Kochen nach Rezept eher selten der Fall ist. Algorithmen allein reichen für ADM nicht aus; es braucht ein **System** aus Hard- und Software (Prozessoren, Speicher, Programmiersprache und Daten). Deshalb wird auch immer wieder von **ADM-Systemen** gesprochen (und da Daten nicht einfach da sind, sondern in einem gesellschaftlichen Zusammenhang erzeugt wurden und erhoben werden, kann man zum System auch soziale Faktoren mit hinzuzählen; siehe Kap. 6). Die ADM-Systeme können deterministisch oder probabilistisch sein, d. h. es gibt entweder genau ein sicheres Ergebnis als Ausgabe oder eine Reihe von Ergebnissen, die alle unterschiedliche Wahrscheinlichkeiten haben. Im Bereich der Künstlichen Intelligenz sind heute Verfahren vorherrschend, die beides vereinen und so strenges Schließen mit unsicherem Wissen ermöglichen (Russell und Norvig 2013, S. 50). Es wäre also irreführend, würde man ADM-Systeme mit Algorithmen gleichsetzen, denn sie bestehen aus mehr als einem Algorithmus.

3.2 Big Data

Eine Folge der Verbreitung digitaler Technologien ist, dass die Menge an Daten immer größer wird, wofür sich der Begriff „**Big Data**“ etabliert hat. Als Endverbraucher generieren wir Daten vor allem durch die Nutzung sozialer Medien wie Facebook, Instagram oder YouTube und den Einsatz unseres Smartphones (z. B. für die Navigation beim Autofahren oder das Bezahlen per App im Supermarkt). Als Organisationsmitglieder produzieren wir allerdings auch Datenmengen in Gigabyte-Höhe, z. B. indem wir Dokumente und Präsentationen anlegen, unsere Personalstammdaten und Mitarbeiterbewertungen speichern, im Firmenintranet etwas Posten, uns in Enterprise Resource Planning-Systeme einloggen und darin Prozesse bearbeiten, etc. Diese Menge an (un-)strukturierten Dokumenten (u. a. Texte, Tabellen, Präsentationen, Videos), Weblogdaten, Prozesslogdaten, Stamm-, Verhaltens- und Interaktionsdaten fallen unter den Begriff Big Data. Sie sind vor allem durch ihre schiere Menge (**Volume**) gekennzeichnet, das heißt man spricht von Datenmengen in der Größenordnung von Terabytes (Strohmeier 2017). Dabei ist die Art der Daten sehr unterschiedlich (**Variety**), von strukturierten Daten, die eine vorgegebene oder erwartbare Form haben (z. B. Personalstammdaten) bis zu un-

strukturierten Daten (z. B. Videobewerbungen). Zusätzlich werden sie sehr schnell generiert und weiterverarbeitet (**Velocity**). Sofern sie objektiv durch Technologien erhoben wurden (z. B. bei Logdaten) ist die Frage nach ihrem Wahrheitswert und ihrer Genauigkeit (**Veracity**) einfach zu beantworten. Schwieriger wird dies bei nutzergenerierten Daten (z. B. Posts im Intranet). Letztlich sollen die Daten für verbesserte Formen der Wertschöpfung genutzt werden (**Value**) (Gandomi und Haider 2015, S. 138 f.).

Bei den Datenarten gibt es unterschiedliche Typen und ein Ereignis (z. B. die Kündigung eines Mitarbeiters) lässt sich durch mehr als einen Datentyp repräsentieren bzw. vorhersagen (Scholz 2017, S. 14 f.):

- primäre Verhaltensdaten: Daten über vergangenes Verhalten der gleichen Art, wie dasjenige, das vorhergesagt werden soll (z. B. für die Kündigungswahrscheinlichkeit: Anzahl und Frequenz früherer Arbeitgeberwechsel)
- sekundäre Verhaltensdaten: Daten über vergangenes Verhalten, das dem vorherzusagenden ähnlich ist (z. B. Anzahl von Wohnortwechseln, Beschwerden über die Arbeitsbedingungen)
- sozio-demografische Daten: Daten über den persönlichen und sozialen Status (z. B. Alter, Anzahl an Berufsjahren, Erwerbsstatus des Partners)
- grafische Daten: Daten über Orte und Distanzen (z. B. Entfernung zwischen Wohn- und Arbeitsort)
- Einstellungen: Daten über die Einstellung bzw. Meinung zu bestimmten Sachverhalten oder die Gefühlslage (z. B. Angaben zur Wechselbereitschaft im XING- bzw. LinkedIn-Profil, Aussagen zur Arbeitszufriedenheit in Mitarbeiterumfragen, im Intranet oder sozialen Medien)
- Netzwerkdaten: Daten über das (organisationale) Beziehungsgeflecht, in dem sich eine Person befindet (z. B. Anzahl und Frequenz an E-Mails oder Meetings mit Kollegen, Führungskräften oder Kunden)

Bei der Analyse von Big Data gibt es einige technische, methodische und rechtlich-ethische Herausforderungen (Garcia-Arroyo und Osca 2019, S. 8 ff.). Zu den technischen gehören Probleme beim Zugriff, der Aufbereitung und Verarbeitung der Daten. Auf rechtlich-ethische Fragestellungen wird später eingegangen (siehe Kap. 6). Die methodischen Herausforderungen sind hauptsächlich die folgenden: (1) die große Anzahl von analysierten Daten legt unberechtigte Korrelationen nahe; (2) da es praktisch unmöglich ist, keine signifikanten Beziehungen in einem sehr großen Stichprobenumfang zu finden, müssen andere Indikatoren wie Effektgröße, der Korrelationskoeffizient oder Konfidenzintervalle herangezogen werden; (3) die Qualität der Daten und Erhebungsverfahren muss gewährleistet sein (auch bei Soft- bzw. Hardware-basierten Daten muss die Reliabilität und

Validität nachgewiesen werden); und (4) die Geschwindigkeit und Unmittelbarkeit, mit der die Daten vorliegen, macht sie anfällig für Schwankungen. Diese Herausforderungen gelten im Allgemeinen auch für die Methoden aus dem Bereich der Künstlichen Intelligenz bzw. des Machine Learnings, deren Anwendung für die Nutzbarmachung von Big Data notwendig ist.

3.3 Künstliche Intelligenz (KI)

Künstliche Intelligenz ist der **Oberbegriff für alle Maschinen** (Computer/Rechner, Roboter, ADM-Systeme), die über Fähigkeiten verfügen, **die dem menschlichen Verhalten beim zielgerichteten Lösen von Problemen ähneln**. Dazu gehört mindestens, dass ein intelligenter ‚Agent‘ Informationen wahrnehmen und verarbeiten sowie entsprechend handeln kann, und dass der Agent lernt, also Informationen über die Welt nicht nur repräsentiert und speichert, sondern auch verändert (Russell und Norvig 2013, S. 22; Hutson 2017). Die wohl größte Bekanntheit haben Agenten wie Amazons Alexa, Apples Siri oder Googles Duplex erlangt, wenngleich KI nicht nur zur Verarbeitung von Sprache eingesetzt wird. Unzählige digitale Dienstleistungen beinhalten KI, um Empfehlungen auszusprechen (z. B. Routenvorschläge, Konsumartikel, News und Videos), Bilder, Texte oder Objekte zu erkennen (z. B. Identifikation von Fingerabdrücken, Gesichtern, Straßenschildern, Hindernissen, Hautkrebs, Schädlingen auf Pflanzen), den Fluss von Dingen und Daten zu organisieren (z. B. Daten- und Informationssuche im Internet, Energiemanagement, Finanztransaktionen) und vieles mehr. Neben der bloßen Erkennung kam vor einigen Jahren noch die Generierung von Sprach-, Bild- und Videodaten hinzu, wodurch auch die Fälschung von Bild- und Tondokumenten („Deep Fakes“) möglich wird (siehe Goodfellow et al. 2018, S. 1 ff., 788 ff.). Beispiele wie diese bilden den Nährboden der aktuellen Diskussion um KI, weil dort beachtliche Fortschritte bei der Verarbeitung riesiger unstrukturierter Datenmengen gemacht wurden.

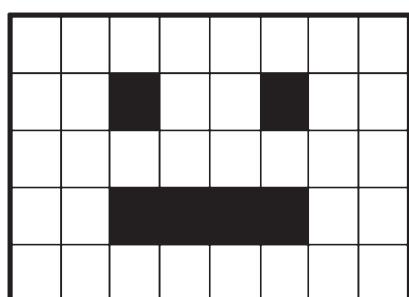
Trotz aller Fortschritte sind alle aktuell verfügbaren **Maschinen nur auf spezifische Ziele bzw. Anwendungsprobleme hin ausgerichtet** (z. B. Bilder erkennen oder Texte übersetzen oder Computerspiele spielen, bei denen das Ziel eindeutig ist). Sie sind insofern „Fachidioten“. Eine generelle künstliche Intelligenz (bzw. Artificial General Intelligence, AGI) müsste Probleme kontextübergreifend lösen können, also z. B. Lesen und logisch Schlussfolgern, kreativ Malen, Schach spielen und Go spielen und Starcraft spielen, usw. Dafür müssten Maschinen nicht nur über unterschiedliche Aufgabenfelder hinweg lernen können, sondern auch noch

über Empathie, Selbstreflektion und Bewusstsein verfügen, um Ziele auszuwählen oder zu generieren oder sie zu ändern. All dies ist bisher nicht der Fall und ob Maschinen dazu überhaupt jemals im Stande sein werden, ist hoch umstritten (Gärtner 2007, S. 149 ff.; Russell und Norvig 2013, S. 1182 ff.). Manchmal finden sich in diesem Zusammenhang noch die Bezeichnungen „**schwache KI**“ und „**starke KI**“. Ursprünglich war damit die Unterscheidung gemeint, ob eine KI nur so tut, als könnte sie denken und fühlen (schwache KI), oder ob dies tatsächlich der Fall ist (starke KI) (Russell und Norvig 2013, S. 1176 ff.). Heute werden diese philosophischen Begriffe insbesondere von Ingenieuren und Softwareentwicklern eher unter dem Leistungsaspekt gesehen: Je nachdem wie gut ein KI-System in der Lage ist, menschenähnliche Leistungen auf einem Gebiet zu erbringen, ist es ein starkes oder schwaches System – von Qualitäten des menschlichen Geistes oder Bewusstseins wie Einsicht, Verständnis von Gründen und Ursachen, Kreativität, Fühlen, Empathie etc. ist hier nicht die Rede. Genau diese Aspekte werden allerdings oft genannt, wenn es darum geht menschliche von maschineller Intelligenz zu unterscheiden: Maschinen lösen Aufgaben, indem sie Muster suchen und anwenden, ohne aber den Sinn hinter den Mustern zu verstehen – dafür braucht es menschliches Denken und Fühlen. Nur dieses kann Ursache-Wirkungszusammenhänge in der Interaktion mit der Umwelt aufstellen, austesten, abändern, erinnern und auf neue Situationen übertragen (Gärtner 2007, S. 175ff., 2017; Marcus und Davis 2019).

KI umfasst eine Reihe von Teildisziplinen und -methoden, u. a. stochastische Verfahren (Machine-Learning, genetische oder evolutionäre Algorithmen) sowie regelbasierte Expertensysteme und Wissensrepräsentationen, die schon vor Jahrzehnten entwickelt wurden (Russell und Norvig 2013, S. 50, 518 ff., 551). Letztere kombinieren kleinere Wissenseinheiten von Experten auf Basis von vorgegebenen Regeln, um Probleme in einem begrenzten Aufgabengebiet zu lösen. Innerhalb dieses Aufgabengebiets (z. B. Rekrutierung von Softwareentwicklern) muss dem Expertensystem ein sogenanntes Domänenwissen zur Verfügung gestellt werden (z. B. welche objektorientierten Programmiersprachen oder Entwicklungsumgebungen für KI-Anwendungen es gibt), damit es Entscheidungen treffen kann (z. B. Einladen einer Entwicklerin, wenn sie mindestens drei Sprachen beherrscht). Ursprünglich beruhten diese Techniken nicht auf Wahrscheinlichkeitsberechnungen und konnten ihre Regeln zum Schlussfolgern nicht ändern. Heutzutage verwenden sie auch probabilistische Techniken und werden mit Machine-Learning-Algorithmen kombiniert (Russell und Norvig 2013, S. 640 f.). Wenn in der Öffentlichkeit von „der KI“ oder „dem KI-System“ gesprochen wird, dann stecken dahinter meist Modelle, die Machine-Learning-Algorithmen einsetzen.

3.4 Machine Learning (ML)

Machine Learning (ML) ist ein Sammelbegriff für Methoden bzw. Algorithmen, die Wissen aus Daten generieren, indem **Muster in den Daten erkannt und für Vorhersagen genutzt** werden. Vorhersagt wird ein Ereignis, wie z. B. die Kündigung eines Mitarbeiters, indem Vergangenheitsdaten über frühere Kündigungen auf Regelmäßigkeiten in den Merkmalsausprägungen hin geprüft werden und in ein statistisches Modell eingehen. Diese Modelle versuchen dann die Merkmalskombinationen herauszufinden, die die richtigen sind, also die gesuchten Werte gut vorhersagen. Für die Vorhersage von Kündigungen könnten z. B. die Merkmale Gehalt, Zufriedenheit mit der Vorgesetzten oder Entfernung bis zur Arbeit größer als 100 km hilfreich sein. Dazu müssen ML-Algorithmen Regeln zur Abschätzung der Daten „lernen“ – Lernen meint im ML-Umfeld also, dass in (großen) Datensätzen Muster gefunden werden, die mit dem zu bewertenden oder zu prognostizierenden Ereignis zusammenhängen. Welcher Art diese Regeln sind und wie gut das Modell die Daten abschätzen kann, hängt davon ab, welcher ML-Algorithmus benutzt wird und welche Daten als Input vorliegen (z. B. Regressionen für lineare Zusammenhänge und stetige Daten oder baumartige Strukturen für nicht-lineare Zusammenhänge und diskrete Daten). Gemeinsam ist den ML-Algorithmen, dass sie numerische Werte benötigen, das heißt Bilder, Texte oder auch Videos müssen in Zahlen umgewandelt werden, sonst kann nichts berechnet und abgeschätzt werden (siehe Abb. 3.1, die einen Smiley als schwarze und weiße Pixel bzw. als Zahlenwerte zeigt: 1 für schwarz und 0 für weiß). Beim Lernen der Regeln werden drei Haupttypen unterschieden (Russell und Norvig 2013, S. 811): überwachtes Lernen (supervised learning), nicht überwachtes Lernen (unsupervised learning) und Verstärkungslernen (reinforcement learning).



0	0	0	0	0	0	0	0	0
0	0	1	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0
0	0	1	1	1	1	0	0	0
0	0	0	0	0	0	0	0	0

Abb. 3.1 Smiley als Schwarz-Weiß-Pixel, die in Zahlen umgewandelt werden

Beim **überwachten Lernen** kennt man die Input-Output-Paare eines Datensatzes (im Fachjargon: der Datensatz ist annotiert bzw. die Ziel-/Target-Variable ist bekannt). Das heißt, man weiß z. B. bei Bildern, ob sie eine Katze zeigen oder nicht, und bei Mitarbeitern, ob sie gekündigt haben oder nicht. Es soll dann ein statistisches Modell gefunden werden, dass die Eingaben auf die Ausgaben abbildet (Russell und Norvig 2013, S. 811). Die Aufgabe des ML-Modells ist es also, Merkmale im Datensatz so zu kombinieren, dass bereits bekannte Datenpaare gut abgeschätzt werden und der Ausgabewert für neu hinzukommende Daten vorhergesagt wird. Lernen meint dann die Verbesserung der Prognosegüte im Zeitverlauf. Im Beispiel geht es um Klassifizierungsprobleme: Liegt ein Katzenbild bzw. eine Mitarbeiterkündigung vor oder nicht? Das Lernproblem könnte auch aus mehreren Klassen oder einer Zahl bestehen, z. B. die morgige Temperatur, womit das Lernproblem eine Regression wäre (Russell und Norvig 2013, S. 812).

Um ein Problem zu lösen, wird ein vorhandener Datensatz (z. B. alle Kündigungen der letzten Jahre) zunächst in Trainings- und Testdaten geteilt (häufig im Verhältnis von 70–80 % Trainingsdaten zu 20–30 % Testdaten; Siegel 2013, S. 125). In den **Trainingsdaten sollen nun Muster erkannt werden**, was nichts anderes bedeutet, als dass ein Algorithmus gesucht wird, der die Input-Output-Paare gut nachbildet. Z. B. könnte der ML-Algorithmus die Regel ableiten, dass Mitarbeitende dann kündigten, wenn sie mehr als 100 km zur Arbeit fahren mussten und ihre Vorgesetzten schlechte Bewertungen bekamen. Der Algorithmus wendet dann diese, während des Trainings gelernte Regel auf die Testdaten an. Unter der Annahme, dass neue Daten dem gleichen Muster folgen. Im Beispiel würde der Algorithmus anhand der beiden Merkmale jene Mitarbeiter identifizieren, die eher kündigen. So kann ein ML-Algorithmus neue Inputdaten verarbeiten, d. h. Schätzwerte für den zum Input (Daten über neue Mitarbeiter) zugehörigen Output (Kündigung) angeben. Mit großer Wahrscheinlichkeit werden mit dieser Regel aber nicht alle Output-Daten richtig vorhergesagt (weder im Trainings- noch im Testdatensatz und auch nicht bei ganz unbekannten Daten), weil beispielsweise auch Mitarbeitende kündigten, die mit ihrer Vorgesetzten zufrieden sind, und Mitarbeitende nicht kündigten, obwohl sie unzufrieden sind. Immerhin kann aber geprüft werden, wie viele der Testdaten richtig bzw. falsch vorhergesagt wurden, denn diese sind ja bekannt.

Beim **nicht überwachten Lernen** sollen Muster in Datensätzen auch ohne Wissen über den korrekten Output erlernt werden (der Datensatz ist nicht annotiert bzw. gelabelt). Das heißt, es ist nicht bekannt, welche Bilder eine Katze zeigen oder welche Mitarbeiter gekündigt haben. Die Algorithmen suchen nach Mustern in Datensätzen durch Beobachtung – das kann dazu führen, dass die Merkmale,

nach denen sortiert oder geschätzt wird, keinen Sinn ergeben. Die häufigste Aufgabe für nicht überwachtes Lernen ist das Clustern bzw. die Assoziationsanalyse. Bei beiden geht es darum, Ähnlichkeiten zwischen Datenpunkten im Datensatz zu identifizieren, ohne dass vorher Hypothesen über solche Muster getroffen werden. Beispielsweise werden beim Clustern die Daten so in Klassen gruppiert, dass sich Daten innerhalb einer Klasse/eines Clusters stark ähneln, während sie zwischen den Clustern sehr unähnlich sind. Solche Analysen sind seit Jahrzehnten im Marketing verbreitet und werden unter anderem zur Warenkorb- und Umsatzanalyse eingesetzt. Beispielsweise könnte ein ML-Algorithmus für Taxi-Bestellungen allmählich ein Konzept von „guten Verkehrstagen“ und „schlechten Verkehrstagen“ anhand der Auslastungsdaten erlernen, ohne jemals benannte Beispiele dafür von einem Überwacher bekommen zu haben (Russell und Norvig 2013, S. 811).

Beim **Verstärkungslernen** basiert der Lernprozess des Algorithmus auf einem Prinzip, das man auch in der Psychologie und Ökonomie: Ein Akteur wählt jene Handlung aus, die sein Wohlbefinden oder seinen Nutzen erhöht und tut dies immer wieder solange er Belohnungen erhält bzw. keine Bestrafung). Beim Verstärkungslernen ist es ähnlich: Der Algorithmus, auch Agent genannt, versucht eine Nutzenfunktion zu maximieren, indem er Belohnungen maximiert. Um eine Aufgabe zu lernen, nimmt der Algorithmus zunächst eine zufällige Abschätzung vor, welche Aktion zu einer Belohnung führt. Dann führt er diese Aktion aus, beobachtet die tatsächliche Belohnung und passt seine Vorhersage entsprechend des Fehlerwerts an (Botvinick et al. 2019, S. 417). Beispielsweise könnte das Fehlen eines Trinkgeldes am Ende einer Fahrt für den Taxi-Algorithmus ein Hinweis darauf sein, dass etwas falsch gelaufen ist – nun muss der Algorithmus noch herausfinden, welche der Aktionen vor der Verstärkung am meisten dafür verantwortlich waren, z. B. Vorfahrtsregeln missachtet oder Autobahnen gemieden zu haben (Russell und Norvig 2013, S. 811). Dieser Vorgang wird millionenfach (oder noch öfter) wiederholt, bis der Vorhersagefehler des Algorithmus gegen Null tendiert, weil er dann gelernt hat, welche Aktionen er durchführen muss, um die Belohnung zu maximieren und damit seine Aufgabe zu erfüllen. Das entscheidende ist, dass die Algorithmenentwickler beim Reinforcement-Learning **keine Vorgaben oder Korrekturen bei den Input-Output-Paaren** mehr machen müssen: Der Algorithmus lernt durch eigene Erfahrungen. Tiefe Reinforcement-Modelle haben eindrucksvolle Ergebnisse beim Spielen von Computerspielen, Go und Poker gezeigt, weil sie die menschlichen Profis schlagen konnten (Botvinick et al. 2019, S. 408).

In der Praxis liegt oft eine **Mischung** vor (**halb überwachtes Lernen**). Meist gibt es ein paar benannte Input-Output-Paare, aber es ist nicht klar, ob diese wahr-

heitsgemäß zugeordnet wurden, und es soll zudem aus einer viel größeren Menge an nicht benannten Beispielen ein Problem gelöst werden. Außerdem lassen sich die **Verfahren kombinieren**, z. B. kann auf das Clustern von Daten (nicht überwachtes Lernen) noch die Anwendung eines Entscheidungsbaumes (überwachtes Lernen) folgen (Russell und Norvig 2013, S. 953 f.). Ähnlich ist es beim Reinforcement-Learning: Da hier sehr viele Daten und Durchläufe gebraucht werden, ist es ökonomisch oft ineffizient und langsam, weshalb es mit Techniken der Wissensrepräsentation kombiniert wird, um einmal gemachte Erfahrungen zu speichern und als Referenzpunkt für die nächste Aktion zu setzen (Botvinick et al. 2019, S. 411). Letztlich liegt es an der verfolgten Fragestellung, welche Algorithmen in das gesamte ML-Modell eingehen. So verwendet man beispielsweise bei sensiblen Problemen – wenn also die Ergebnisse große Auswirkungen auf Leib und Leben (oder Geld) haben – eher ML-Algorithmen aus dem Bereich des überwachten Lernens, weil man bei diesen besser nachvollziehen kann, wie ein Ergebnis entstanden ist.

Wie bereits erwähnt, hängt die Frage, welcher ML-Algorithmus und welches Lernverfahren angewandt wird, davon ab, **welches Problem** gelöst werden soll (z. B. Klassifikations- oder Regressionsproblem) und **welche Daten** vorliegen. Dabei lassen sich die allermeisten Fragen, die von ML-Algorithmen gut bearbeitet werden können, einem von sechs Typen zuordnen (siehe Tab. 3.1), wobei ein ML-Modell auch aus mehr als einem Algorithmus bestehen kann. Weil es für ein Problem viele geeignete ML-Algorithmen gibt, muss die **Vorhersagegüte mehrerer Algorithmen verglichen** werden. Für die Vorhersagegüte gibt es typischerweise mehrere Kennzahlen, die abhängig vom Lernproblem sind. Z. B. werden bei Klassifizierungsproblemen meist die Werte *accuracy* (richtig zugeordnete Daten), *precision* (richtig positiv zugeordnete Daten dividiert durch die Summe aus falsch und richtig positiven) und *recall* (richtig Positive dividiert durch die Summe aus richtig Positiven und falsch Negativen) angegeben. Allgemeiner geht es um die Passung bzw. den Fit eines Modells, beschrieben als Unterschied zwischen beobachteten Werten und den vom Modell vorhergesagten Werten, oder wie robust bzw. stabil das Modell bei neuen Daten ist. Die Gefahr, dass falsche Schlüsse aus neuen Daten, die nicht dem alten Muster folgen, gezogen werden, wird auch als Überanpassung des statistischen Modells an die Trainingsdaten bezeichnet (siehe Provost und Fawcett 2017, S. 145 ff.; Goodfellow et al. 2018, S. 121 ff.). Allerdings spielen in der Praxis **neben statistischen Gütwerten auch pragmatische Kriterien** wie „einfach zu verstehen und anzuwenden“, „niedrige Kosten bei der Berechnung und Implementierung“ eine Rolle. Deshalb gibt Siegel (2013, S. 128) allen ML-Experten folgenden Rat: „If you care about your model, **give it a KISS: ,Keep it simple, stupid!**“

Tab. 3.1 Typische Fragen für Machine-Learning-Algorithmen

Frage-/Problemtypus	Algorithmen/-klassen
1. Gehört das Ereignis/der Datenpunkt zu A oder B?	Klassifikationsalgorithmen (z. B. logistische Regression, Support Vector Machines, Entscheidungsbäume (Decision Trees und Random Forests), Extreme Gradient Boosting, k-Nearest Neighbor, Naïve Bayes, Linear Discriminant Analysis, Künstliche Neuronale Netze)
2. Ist das Ereignis/ein Datenpunkt ungewöhnlich?	Ausreißer-Analysen mit Zeitreihen- (z. B. Locally Estimated Scatterplot Smoothing), Klassifikationsalgorithmen (Isolation Forest) oder Clusteralgorithmen (z. B. Density Based Spatial Clustering of Applications with Noise [DBSCAN])
3. Wie ähnlich sind sich Ereignisse/Datenpunkte und was ist ihre Strukturierungslogik?	Clusteralgorithmen (z. B. DBSCAN, k-Means, Balanced Iterative Reducing and Clustering using Hierarchies [BIRCH])
4. Wie viel von einem Ereignis/Datenpunkt wird durch andere Faktoren bewirkt?	Regression, Assoziationsanalyse (z. B. multivariate Regression), Bayes'sche Netzwerke
5. Was wird als nächstes passieren bzw. was sollte ich als nächstes tun?	Ereigniszeitanalyse (z. B. Cox-Regression) bzw. Verstärkungslernen (z. B. Generative Adversarial Networks), Bayes'sche Netzwerke
6. Wie können fehlende Inputdaten abgeschätzt werden?	Imputationsalgorithmen bzw. Missing-Data-Techniques (z. B. Mittelwert-Ersetzung, Last Observation Carried Forward)

ML-Algorithmen eignen sich für bestimmte Fragestellungen (eigene Erstellung)

Im Folgenden sollen ein paar weit verbreitete ML-Algorithmen vorgestellt werden: die multivariate und die logistische Regression, Entscheidungsbäume und neuronale Netze.

Mit der multivariaten bzw. **multiplen Regressionsanalyse** lässt sich herausfinden, ob ein Zusammenhang zwischen mehreren unabhängigen und einer abhängigen Variable besteht, z. B. wie stark die Mitarbeiterzufriedenheit (hier: abhängige Variable) von der Zufriedenheit mit dem Führungsstil der Vorgesetzten, der Anzahl an Arbeitsstunden pro Woche, dem Gehalt oder weiteren Variablen beeinflusst wird. Die lineare Regression geht davon aus, dass die Einflüsse der unabhängigen Variablen X_1 bis X_k auf die abhängige Variable Y additiv verknüpft sind (dazu kommt noch eine Zufallskomponente ϵ). Die Korrelationskoeffizienten β geben dabei an, um wie viel sich Y verändert, wenn sich das zu β gehörige X um eine Einheit ändert. In Abb. 3.2 ist dies für die Arbeitszufriedenheit schematisch

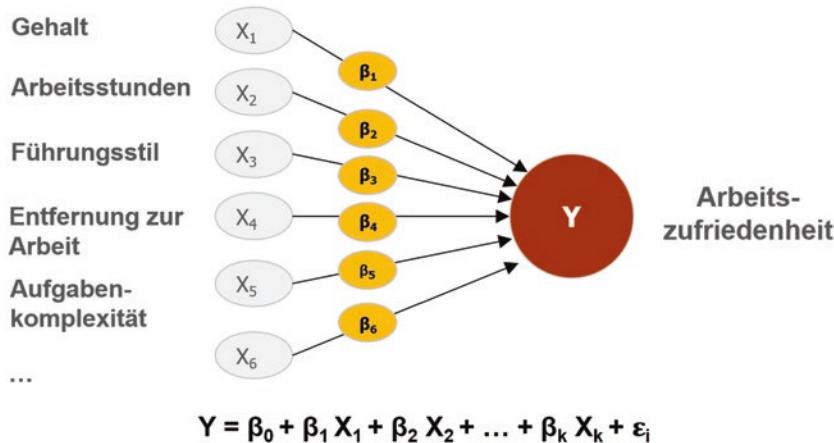


Abb. 3.2 Multiple Regression, schematisch für mögliche Einflussfaktoren auf Arbeitszufriedenheit

dargestellt. Beispielsweise könnte sich eine Gehaltserhöhung um eine Einheit in einer positiven Veränderung der Arbeitszufriedenheit niederschlagen, die dafür als metrischer und intervallskalierter Wert vorliegen muss (z. B. auf einer Skala von 1–5). Es lassen sich natürlich auch andere Variablen einsetzen (z. B. Arbeitsleistung, Kundenzufriedenheit, Fehlzeiten, Trainingsinhalte, Führungsverhaltensweisen), sodass beliebig viele Wirkungsverhältnisse analysiert werden können.

Mit Hilfe einer **logistischen Regression** lassen sich Klassifikationsprobleme lösen. Der Vorteil bei einer logistischen Regression ist, dass Datenpunkte nicht nur in Klassen eingesortiert werden, sondern dass zusätzlich die Wahrscheinlichkeiten für die Klassenzugehörigkeiten angegeben werden (Russell und Norvig 2013, S. 844). Das geht bei einer binären logistischen Regression nur für abhängige Variablen, die zwei Werte annehmen (also in eine von zwei Klassen fallen), dafür dürfen die unabhängigen Variablen unterschiedlicher Art sein (metrisch, kategorial) (Backhaus et al. 2016, S. 286). Dann kann z. B. abgeschätzt werden, wie stark die Kündigungswahrscheinlichkeit von der zurückgelegten Distanz zum Arbeitsplatz, der Zeitspanne seit der letzten Beförderung oder der Zufriedenheit mit dem Führungsstil des Vorgesetzten beeinflusst wird (siehe Abb. 3.3). Die Wahrscheinlichkeit, dass Mitarbeiter kündigen, ist die abhängige binäre Variable, die nur in eine von zwei Kategorien fallen kann: gehen oder bleiben. Dies wird mathematisch über die logarithmierte Chance ausgedrückt, dass ein bestimmtes Ereignis

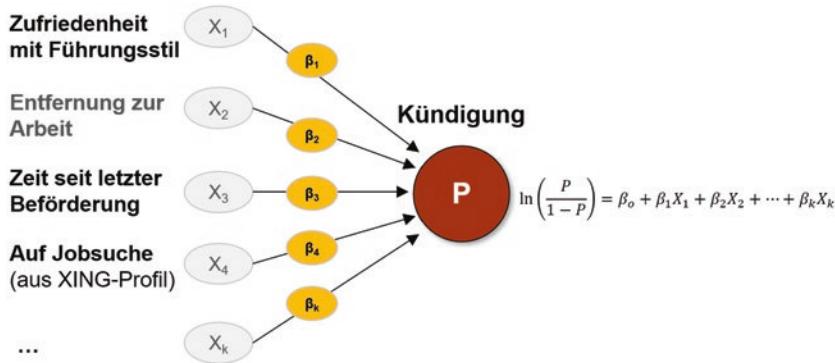


Abb. 3.3 Logistische Regression, schematisch für mögliche Einflussfaktoren auf Eigenkündigung

(z. B. Kündigung) eintritt (Provost und Fawcett 2017, S. 128). Es werden also anders als bei der linearen Regression nicht die Werte der abhängigen Variablen y vorhergesagt, sondern die Eintrittswahrscheinlichkeit, dass $y = 1$ ist. Weil der Algorithmus auf logarithmierten Werten basiert, dürfen die β -Koeffizienten nicht als lineare Faktoren aufgefasst werden (wie bei der linearen Regression, wo die Änderung von x um eine Einheit zu einer Änderung von y um den Faktor β führt). Meist werden deshalb Effektkoeffizienten angegeben, bei denen Werte kleiner als 1 einen negativen Einfluss und Werte größer als 1 einen positiven Einfluss auf die Odds, d. h. das Verhältnis von $P(Y=1)$ zur Gegenwahrscheinlichkeit $1-P(Y=1)$, haben (Backhaus et al. 2016, S. 308).

Die Abschätzung von Klassenzugehörigkeiten lässt sich auch mit anderen ML-Algorithmen vornehmen, z. B. Entscheidungsbäumen (Classification Decision Trees) und Gradient Tree Boosting, Support Vector Machines (hier wird der Datensatz über „Stützvektoren“ – Support Vectors – so in Klassenaufgeteilt, dass um die Klassengrenzen herum ein möglichst breiter Bereich ohne Daten bleibt) oder Künstlichen Neuronalen Netzen (Goodfellow et al. 2018, S. 155 ff.). **Entscheidungsbäume sind eine eingängige Klassifikationsmethode**, weil die aufeinanderfolgenden Schritte übersichtlich visualisiert werden: Genau wie bei einem Baum gibt es eine Wurzel (eine Klasse), von der aus Schritt für Schritt Äste und Blätter (weitere Klassen) abzweigen, wobei jedem neuen Zweig eine Entscheidungsregel zugrunde liegt. Soll beispielsweise abgeschätzt werden, welche Klassen von Mitarbeitern eher kündigen werden, so könnte das Verfahren zu-

nächst das Merkmal „Entfernung zur Arbeit“ prüfen und eine Entscheidungsregel anwenden (mehr oder weniger als 100 km), um zu einer ersten Einteilung in Klassen zu kommen. Dann könnten weitere Fragen gestellt werden, z. B. ob die letzte Beförderung länger als 3 Jahre her ist und ob Zufriedenheit mit dem Führungsstil des Vorgesetzten vorliegt. Je nach Antwort wird eine Entscheidung über die Klassenzugehörigkeit eines Datenpunktes getroffen (siehe Abb. 3.4 mit fiktiven Werten). Beim Durchqueren des Baumes von der Wurzel bis zum Blatt muss also eine Frage nach der anderen beantwortet werden, um das Klassifizierungsproblem zu lösen. Das klingt einfach und ist visuell gut nachvollziehbar – aber entscheidend ist, welche Fragen an welcher Stelle gestellt werden und wie viele insgesamt. Das ist Aufgabe des überwachten Lernens beim Training eines Entscheidungsbaums, wobei ein einfaches Prinzip angewendet wird: immer das wichtigste Merkmal zuerst testen, das heißt, die Frage zuerst zu stellen, die für die Klassifizierung eines Beispiels den größten Unterschied verursacht (Russell und Norvig 2013, S. 817). Im Beispiel: Erlaubt das Merkmal „Entfernung zur Arbeit“ eine bessere Vorhersage der Kündigung oder die „Zeitspanne seit der letzten Beförderung“ oder die „Zufriedenheit mit dem Führungsstil“? Welches es tatsächlich ist, hängt natürlich von den vorliegenden Daten ab und kann zwischen Unternehmen variieren (aber genaue diese Spezifizierung ist eine Stärke von HR Analytics).

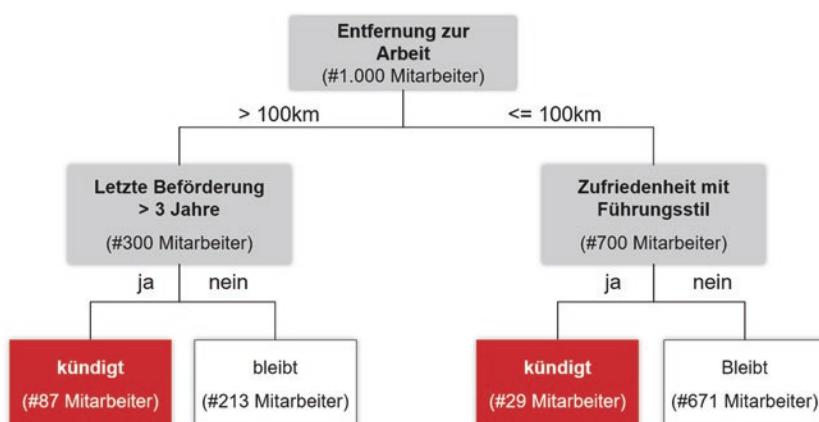


Abb. 3.4 Entscheidungsbaum zur Abschätzung des Kündigungsrisikos (vereinfacht, eigene Erstellung)

Mit dem **Random-Forest-Verfahren** werden viele unterschiedliche Entscheidungsbäume generiert, sodass viele Entscheidungsregeln und damit Klassenbildungen durchprobiert und jene ausgewählt werden können, die die Verteilung der Daten am besten vorhersagen (z. B. könnte ein Entscheidungsbäum auch das Gehalt berücksichtigen und zum Ergebnis kommen, dass damit die Klasse an Mitarbeitern, die kündigen, gut vorhergesagt werden kann). Entscheidungsbäume können sehr viel komplexer werden und auch stetige (statt nur diskrete) Variablen abschätzen (sogenannte Regression Trees) (Russell und Norvig 2013, S. 878).

Verfahren wie die multiple und logistische Regression oder Entscheidungsbäume sind schon seit Jahrzehnten bekannt, aber erst durch das **Zusammenspiel von neuer Hardware** (z. B. Graphics Processing Units), **Massen an (un-)strukturierten Daten** (z. B. Texte, Videos) und weiteren **Verarbeitungsprozessen** (z. B. Batch Normalisierung, Convolutional Layers) sind jene Algorithmen entstanden, die heute Anwendungen wie Text-, Bild- und Spracherkennung oder Empfehlungen möglich machen. Dazu gehören Künstliche Neuronale Netze, deren Grundprinzipien auch schon vor über 50 Jahren beschrieben wurden.

3.5 Künstliche Neuronale Netze (KNN)

Künstliche Neuronale Netze (KNN) bestehen aus **Schichten von verknüpften Algorithmen**, die die Aktivitäten von Neuronen im Neokortex nachahmen sollen. Deshalb gibt es bei KNNs miteinander verbundene Knoten – quasi die ‚Neuronen‘ des künstlichen Gehirns, welche aber nichts weiter als einfache Rechenregeln sind. Die Knoten sind über mehrere Schichten verteilt und berechnen in Abhängigkeit von ihrem aktuellen ‚Aktivierungszustand‘, dem erhaltenen Input und einem Gewichtungsfaktor einen Ausgabewert. Im einfachsten Fall nimmt ein KNN nur eine Reihe an gewichteten Summationen vor, wobei die Gewichtungsfaktoren während des Trainings des KNNs mit vielen Daten angepasst werden, indem entweder die Abweichungen des ausgegebenen Rechenergebnisses vom tatsächlichen Wert minimiert oder Rechenwege, die zu richtigen bzw. erwünschten Ergebnissen führen, verstärkt werden. Diese Anpassungen werden als **Lernen des Netzwerks** bezeichnet. Besteht ein KNN aus vielen Schichten, spricht man von einem ‚tiefen‘ Netz oder auch **Deep Learning** (DL). Es gibt eine ganze Reihe von KNN, die sich insbesondere durch die Art der miteinander verknüpften Daten und Algorithmen unterscheiden. Z. B., ob die Input- und Output-Daten zeitlich oder räumlich strukturiert sind (Ton- oder Bilddaten), welcher Aktivierungsfunktion gewählt wurde (z. B. Rectified Linear Units, Tanh oder

Sigmoid) und welches Vorgehen beim Anpassen der Gewichtungen (Backpropagation). Am weitesten verbreitet sind Convolutional Networks (CNN) für die Bilderkennung sowie Recurrent Neural Networks (RNN) für Sprache und deren Unterart, die Long Short Term Memory-(LSTM) Netzwerke, mit deren Hilfe die Text- und Sprachverarbeitung (Natural Language Processing, NLP) deutlich verbessert wurde (Goodfellow et al. 2018, S. 513 ff.). Eine neuere Anwendung sind Generative Adversarial Networks (GAN), bei denen zwei Netzwerke gegeneinander „antreten“: Eines, das Daten generiert (Generator-Netz), und ein zweites (Diskriminator-Netz), das untersucht, ob eine vorliegende Stichprobe aus den (echten) Trainingsdaten kommt oder vom Generator-Netz (künstlich) erzeugt wurde (Goodfellow et al. 2018, S. 788 f.). Mit GANs lassen sich Dokumente herstellen (z. B. Bilder, Videos), die zwar real aussehen, aber Fälschungen sind (Deep Fakes). Diese Möglichkeit löst große Sicherheitsbedenken aus, man denke nur an gefälschte Reden von Politikern und gefälschte Telefon- oder Videoanrufe von Personen, denen man vertraut und denen man deshalb auch vertrauliche Daten offenbart (z. B. Passwörter).

Moderne DL-Netze haben dutzende Schichten mit tausenden von Knoten, verwenden nicht nur einfache Summenbildung, sondern auch nicht-lineare Funktionen (z. B. quadratische oder logistische Sigmoid-Funktionen), und **ihr hierarchischer Aufbau erlaubt es Aufgaben, die in kleinere Teilaufgaben zerlegt werden können, zu lösen** (Goodfellow et al. 2018, S. 185 ff.). Z. B. bestehen Bilder aus einer Ansammlung von Pixeln, deren Werte den Input-Vektor darstellen (z. B. Helligkeitswerte). Beim Durchlaufen der Schichten des KNNs werden diese Werte durch mathematische Funktionen (z. B. gewichtete Summenbildung, logarithmische Umformung) umgerechnet und nach und nach transformiert, bis der Output-Vektor in der Ausgabeschicht erscheint (siehe Abb. 3.5).

Jede Schicht extrahiert immer abstraktere Merkmale und löst damit das Gesamtproblem, indem Teilaufgaben bearbeitet werden. Während in der ersten Schicht anhand der Helligkeitswerte der Pixel Kanten bestimmt werden, erfolgt in der zweiten Schicht die Suche nach Konturen und Ecken durch die Aneinanderreihung von Kanten. In der dritten Schicht werden zusammenhängende Objektteile erkannt und so weiter, bis dann in der Output-Schicht eine Reihe von identifizierten Objekten ausgegeben werden (Goodfellow et al. 2018, S. 7). Diese **Ausgabe ist mit Wahrscheinlichkeitsangaben versehen**, wie sicher ein Objekt als solches erkannt wurde (z. B. eine Zahl, siehe Abb. 3.6). Häufig wird dazu in der letzten Schicht eine sogenannte Softmax-Funktion verwendet (eine Erweiterung der logistischen Regression), die die Wahrscheinlichkeit der Eingabe zu jeder der vorhandenen Klassen berechnet und die

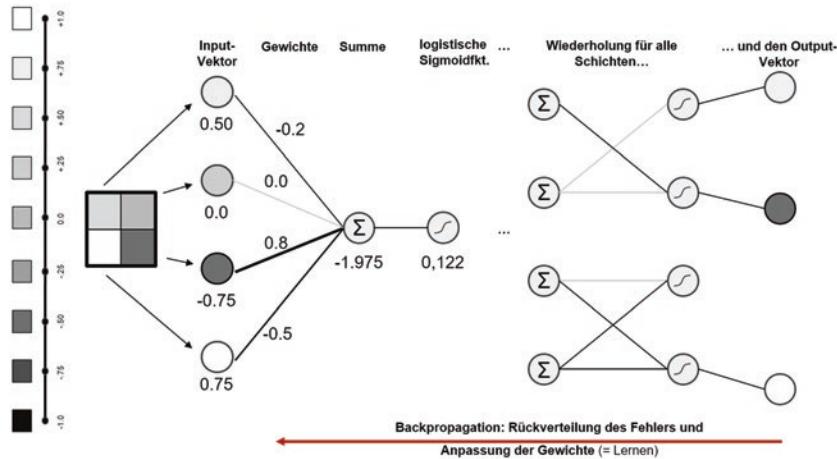


Abb. 3.5 Künstliches Neuronales Netz zur Bilderkennung durch Kategorisierung von Pixeln (vereinfacht)

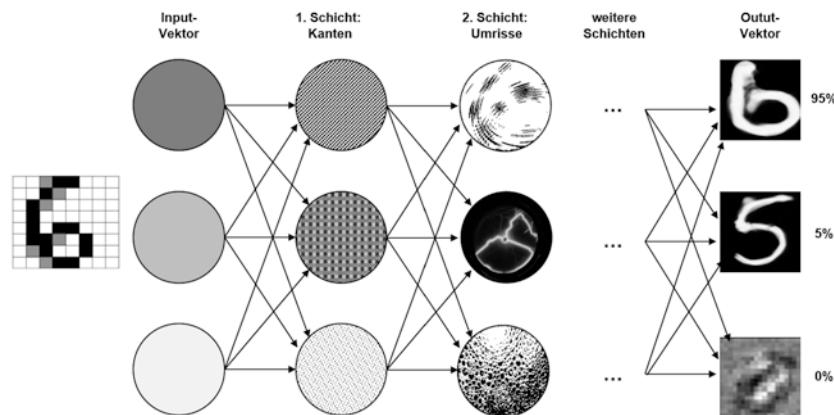


Abb. 3.6 Deep Learning zur Bilderkennung am Beispiel einer Zahl (schematisch und fiktiv)

Klasse mit der höchsten Wahrscheinlichkeit als vorhergesagte Klasse auswählt (Goodfellow et al. 2018, S. 203). Mit KNN lassen sich sowohl Klassifizierungs- als auch Prognoseprobleme bearbeiten (Backhaus et al. 2016, S. 604). Dazu gehört die Bild- bzw. Spracherkennung genauso wie Regressionen oder Zeitreihenanalysen.

3.6 Analytics: Descriptive, Diagnostic, Predictive und Prescriptive

Die Idee, Personalarbeit auf Basis von Zahlen bzw. Messgrößen und wissenschaftlicher Evidenz zu gestalten, ist nicht neu (z. B. Fitz-enz 1984; Lawler et al. 2004). Im Laufe der Jahrzehnte wurden viele sogenannte Key Performance Indicators entwickelt, mit deren Hilfe vor allem die Effektivität und Effizienz der Personalprozesse gemessen wurde (z. B. Personalkostenquote, Zeitdauer zwischen Ausschreibung einer Stelle und Einstellung einer Kandidatin). Solche Kennzahlen geben nur Antworten auf die Frage „Was ist geschehen?“ („Descriptive Analytics“) (siehe Abb. 3.7). Sie können nur der Anfang von weiteren Nachforschungen zu möglichen Ursachen sein (z. B. Warum sind die Personalkosten so, wie sie sind? Warum musste eine Stelle zentral ausgeschrieben werden?). Für diese müssen dann weitere Informationen aus Benchmarks, Umfragewerten, Interviews, etc. hinzugezogen werden („Diagnostic Analytics“: Warum ist es geschehen?). Auch dies ist seit vielen Jahren fester Bestandteil guter Personalarbeit.

Etwas neuer ist der Ansatz, nicht nur Kennzahlen zu erheben, sondern Wirkungsverhältnisse zu analysieren und Vorhersagen zu treffen. Mit „Predictive Analytics“ soll die Frage „Was könnte geschehen?“ beantwortet werden (für Beispielefragen siehe Kap. 1). Unter dem Label People Analytics (auch HR oder Workplace Analytics) sollen Ereignisse vorhergesagt werden, die einen Bezug zum Personal (z. B. deren Einstellungen, Fähigkeiten, Arbeitsleistungsdaten), den Personalprozessen, dem Arbeitsplatz und/oder der Unternehmensleistung haben (Marler und Boudreau 2017). Im Unterschied zur deskriptiven Analytik geht es nicht nur um die Aufbereitung von Vergangenheitsdaten in Reports, Dashboards,



Abb. 3.7 Analytics-Stufen

KPI-Cockpits und/oder Soll-Ist-Abgleichen, sondern um die Prognose zukünftiger Ereignisse auf der Basis von Mustern in einem Datensatz, die über mathematisch-stochastische Verfahren wie Regressionsanalysen, Clusteranalysen, Entscheidungsbäume oder KNN identifiziert werden. Dabei geht es immer um Wahrscheinlichkeiten, mit denen zukünftige Ereignisse (z. B. Kündigungen von Schlüsselpersonen) eintreten könnten – und diese Wahrscheinlichkeiten müssen auf Basis von Vergangenheits- oder Echtzeitdaten geschätzt werden (letzteres ist im HR-Umfeld noch selten der Fall). Da bei Predictive Analytics nach Mustern in den Daten gesucht bzw. „geschrifft“ wird, spricht man auch von Data Mining. Während beim generellen Data Mining meist auf strukturierte Daten zurückgegriffen wird (z. B. Personalstammdaten, Gehälter oder Kompetenzbewertungen), liegen dem Text Mining unstrukturierte zugrunde (z. B. Bewerbungsschreiben, Ausschreibungen auf Online-Stellenbörsen) und beim Process Mining werden Systemlogdaten ausgewertet (z. B. Dauer der Erfassung eines neuen Mitarbeiters im System). Für Predictive Analytics im HR-Umfeld werden also heute mehr und vielfältigere Daten ausgewertet als das früher möglich war, insbesondere was das Verhalten der Mitarbeitenden angeht (z. B. Tastaturanschläge, Öffnungs- und Überarbeitsraten von Dokumenten, Posts, Anzahl und Dauer von Meetings, Bewegungs-, Kommunikations- und Interaktionsdaten).

Die letzte Stufe der Entwicklung bei den Herangehensweisen an die datenbasierte Personalarbeit bildet „**Prescriptive Analytics**“, d. h. „Was soll geschehen?“. Im Bestfalle haben die vorgeschalteten Stufen eindeutige Erkenntnisse über Ursache-Wirkungszusammenhänge geliefert, sodass nun Entscheidungsregeln vorgegeben werden können. Weil die Ergebnisse nur auf (noch nicht falsifizierten) Hypothesen und Wahrscheinlichkeiten basieren, es also keine 100 %-ige Sicherheit gibt, bildet man bei den Vorschriften unterschiedliche Szenarien und beschreibt die Risiken, die bei Nichteintreten des Vorhergesagten und Vorgeschriftenen entstehen.

Im Analytics-Umfeld herrscht nahezu eine **babylonische Sprachverwirrung**: Neben HR Analytics, People Analytics, Workplace Analytics und Human Capital Analytics gibt es ähnliche Labels auch für einige Unterdisziplinen des HR-Managements, z. B. Talent Analytics oder Learning & Development Analytics. Dass es so viele Begriffe gibt, liegt nicht zuletzt daran, dass datengetriebene HR-Arbeit spätestens durch Davenport et al. (2010) in den Stand eines Wettbewerbsfaktors erhoben wurde, weshalb kaum eine Personalmanagement-Konferenz oder Personalzeitschrift ohne das Thema auskommt. Dabei fallen oft Begriffe wie KI, DL oder ML. Dass auch mit einfacheren und etablierten statistischen Verfahren (siehe Abschn. 3.4) bereits Predictive Analytics im Personalbereich durchgeführt werden können, wird nicht selten verschwiegen. Sofern es in den Vorträgen und Artikeln nur um Kosten- oder Zeitaspekte und nicht auch um die **Qualität des**

Outputs der Personalarbeit geht, greifen die Analysen zu kurz, wie auch immer sie benannt sind. Das wird klar, wenn man sich vor Augen führt, was ein gutes Restaurant ausmacht. Wenn wir Restaurants nur danach bemessen würden, wie schnell wir bedient werden und wie billig es ist, würden wir nur bei Fast-Food-Ketten essen. Je wichtiger Personalarbeit für die Wettbewerbsfähigkeit eines Unternehmens ist, desto mehr ausgefallene Gerichte sollte HR anbieten – bis hin zur Sterngastronomie.

Ein Teil der Verwirrung bezieht sich auf die Frage, inwieweit durch HR Analytics **nur Korrelationen oder auch verursachende Wirkungsbeziehungen** aufgedeckt werden können. Welche Art von Zusammenhang proklamiert werden kann, liegt nicht am verwendeten Begriff (z. B. Predictive oder Prescriptive), sondern am jeweiligen methodischen Vorgehen. So müssen bei der Analyse drei Bedingungen erfüllt sein, um von einem kausalen Einfluss einer Variable X (z. B. einer Gesundheitsmaßnahme) auf eine Variable Y (z. B. dem Krankenstand in der Belegschaft) sprechen zu können (Cheng 2017, S. 5):

1. X muss zeitlich vor Y liegen, also die Maßnahme muss der Periode, in der die Krankheitstage gemessen werden, vorangehen.
2. X und Y müssen korrelieren, das heißt der Krankenstand und die Gesundheitsmaßnahme müssen miteinander in einem Zusammenhang stehen, der nicht rein zufällig auf die Stichprobenauswahl zurückzuführen ist.
3. Alternative Erklärungen für den Zusammenhang von X und Y müssen ausgeschlossen werden (inklusive der umgekehrten Kausalität, also, dass die Gesundheitsmaßnahme nicht ursächlich für den sich veränderten Krankenstand war, sondern als Folge dessen eingeführt wurde).

Um kausale Zusammenhänge zu analysieren, sollten im Idealfall randomisierte Feldexperimente durchgeführt werden, in denen die Beschäftigten zufällig einer Gruppe, die an der Gesundheitsmaßnahme teilnimmt (Treatmentgruppe) und einer Kontrollgruppe, die nicht teilnimmt, zugeordnet werden. Durch den Vergleich der Krankheitstage zwischen Treatment- und Kontrollgruppe können Rückschlüsse auf den Effekt der Maßnahme getroffen werden. Solche **Experimente sind in der Praxis kaum umsetzbar**, sowohl aus pragmatischen Gründen (wie kann man die Gruppen ohne größeren Aufwand teilen und erfolgt die Zuordnung wirklich zufällig und freiwillig?) als auch aus ethisch-moralischen Erwägungen heraus (wie ließe sich rechtfertigen, dass eine Gruppe nicht an einer Gesundheitsmaßnahme teilnehmen darf?). Um trotz fehlender Randomisierung einen kausalen Rückschluss vornehmen zu können, bieten sich – bestimmte Annahmen vorausgesetzt – eine Reihe von statistischen Verfahren wie Difference-in-Difference Schätzungen

oder Regression-Diskontinuitäts-Analysen an, um Ursache-Wirkungsbeziehungen zu untersuchen (Cheng 2017, S. 5 f.). Weit verbreitet – z. B. in digitalen Sprachassistenten – sind auch Bayes'sche Netzwerke. Werden diese auf Basis einer Theorie oder Hypothesen über Ursachen und Wirkungen konstruiert, können auch kausale Zusammenhänge in Daten identifiziert werden (Pearl & Mackenzie 2019, S. 128 ff.).

Unabhängig von der Frage, ob Kausalität oder nur Korrelationen identifiziert werden können, gibt es ein mehr oder weniger standardisiertes **Vorgehen bei HR Analytics** (siehe Lawler et al. 2004; Marler und Boudreau 2017; Mühlbauer et al. 2018). Im Kern sind es neun Schritte, die bei People-Analytics-Projekten zu beachten sind (siehe Abb. 3.8): Am Beginn steht immer die **Problemdefinition** mit der Frage, warum eine etwaige Analyse relevant für das Geschäft ist. Dies mündet in eine Liste an **Hypothesen**, wie das Problem adressiert werden kann. Die Hypothesen sollten dabei ergebnisorientiert sein, d. h. man muss sich vorab Gedanken machen, welche Daten man braucht, um das Problem zu lösen, welche Daten gesammelt werden können bzw. dürfen und ob genügend Zeit und Personal für die Auswertung zur Verfügung steht. Dieser Schritt erfordert Erfahrung und Vorstellungskraft (über den Projektlauf und erwartbare Ergebnisse), kann aber davor bewahren unnötig Ressourcen in ein Projekt zu stecken, das dann keine Resultate bringt. Erst danach sollte die **Datensammlung** beginnen, und zwar ebenso hypothesesbezogen wie datenschutzkonform. Die darauffolgende **Datenaufberei-**



Abb. 3.8 Vorgehensmodell bei People Analytics

tung nimmt in der Praxis meist den größten Anteil ein (ca. 85–95 % der Projekt-dauer; Abschn. 2.2). An die Datenqualität können je nach Projekt unterschiedliche Anforderungen gestellt werden, meist sollten sie aber akkurat, konsistent, vollständig (so weit wie möglich) und anonymisiert (so weit wie nötig) sein. Im Vergleich zur Aufbereitung geht die **Analyse der Daten** sehr schnell. Zu beachten ist, dass unterschiedliche Modelle gerechnet und auf ihre Performanz hin geprüft werden (z. B. wie viele falsch positive und falsch negative Vorhersagen angenommen werden müssen). Auch bei Algorithmen gilt: Viele Wege führen nach Rom (beispielsweise gibt es eine Vielzahl an Klassifikationsalgorithmen, mit jeweiligen Vor- und Nachteilen). Für alle Analysen gilt allerdings: Die Maßnahmen für eine Lösung stecken nicht in den Daten allein. Anwender müssen **Verbesserungsmaßnahmen erst noch ableiten bzw. entwickeln**. Kennzahlen, Dashboards und statistische Auswertungen können nur Hinweise geben, wo Probleme sind und Lösungen gesucht werden können. Gefunden, ausgearbeitet und umgesetzt werden müssen die Maßnahmen aber noch von Menschen. Und bevor Menschen dies tun, wollen sie überzeugt werden, mit Argumenten, aber auch mit zielgruppenadäquaten **Visualisierungen** (Diagramme, Bilder, Metaphern) und einer zum Kontext passenden **Geschichte**. Das erhöht die Wahrscheinlichkeit der Umsetzung (Knaflic 2015, S. 19, 253 f.). Für die **Umsetzung** selbst kann man sich je nach Projektumfang professionelle Change Management-Unterstützung holen. Diese wird, wenn sie evidenzbasiert arbeitet, sowohl auf harte Faktoren (Ziele, Strukturen, Anreize und Belohnungen, etc.) als auch auf die zwischenmenschliche Komponente achten (Motivation, Kompetenz-erleben, Fairness, etc.). Dass diese Elemente in bestimmten Kontexten erfolgswirksam sind, ist recht gut erforscht (Stouten et al. 2018; Gärtner 2019). Im letzten Schritt gilt es zu **evaluieren**, ob die Analyse und entwickelten Maßnahmen zur erwarteten Problemlösung geführt haben oder ob sie zu **optimieren** sind.

3.7 Automatisierung und Robotic Process Automation (RPA)

Automatisierung meint, dass ein Prozess unter festgelegten Bedingungen ohne menschliches Eingreifen abläuft (DIN IEC 60050-351). Ein Prozess beinhaltet eine Abfolge von Arbeitsschritten (z. B. Fließbandfertigung, Versand von Werbe-E-Mails, Freigabe von Kundenaufträgen nach Kreditwürdigkeitsprüfung).

Seit Frederick W. Taylor und Henry Ford Anfang des 20. Jahrhunderts mit Standardisierung und Automatisierung die Produktion von Gütern nachhaltig veränderten, verfolgt Automatisierung letztlich immer die gleichen Ziele:

- Erhöhung der Effizienz (erhöhter Output bei gesenkten Kosten),
- Verkürzung der Bearbeitungszeiten,
- Erleichterung der menschlichen Arbeit,
- Erhöhung der Qualität bzw. Verringerung der Fehler und Ausschussraten

Dennoch ist Automatisierung nicht gleich Automatisierung. Es gibt **Stufen der Automatisierung**, ähnlich wie bei autonom fahrenden Autos von unterschiedlichen Leveln gesprochen wird (z. B. Level 1: Fahrassistenzsysteme, Level 5: Autonomes Fahren). Unterschieden wird, welche Anteile der Aufgabenerfüllung bzw. des Entscheidens vom Menschen und welche automatisiert von Maschinen übernommen werden (siehe Abb. 3.9). Auf der niedrigsten Stufe (manuelles System) nimmt der Mensch (Benutzer) dauerhaft alle anfallenden Aufgaben und Entscheidungen zur Zielerreichung selbst wahr, ohne dass ein Softwaresystem vorhanden oder involviert ist (Welsch et al. 2018, S. 369). Auf der ersten Automatisierungsstufe unterstützt das System hingegen den Menschen bereits bei der Ausführung definierter Aufgaben, während der Mensch weiterhin die Kontrolle hat und die Entscheidungen selbst trifft. Auf der zweiten Stufe kommt hinzu, dass Systeme nicht nur mehr Aufgaben übernehmen, sondern einfache Funktionen automatisch und gemäß einem fixierten Ziel ausführen, d. h. ohne Auslösung durch den Benutzer (autonome Assistenz, Braun 2018, S. 12 f.). In Stufe 3 werden dann unter Nutzung von ML-Verfahren Wege zur Zielerreichung angepasst. Z. B. würde ein Chatbot anhand vergangener Interaktionen die (wahrscheinlichen) Intentionen des Nutzers vorhersagen und seinen Output entsprechend anpassen. Der Mensch agiert als Überwacher, der bei Bedarf (re-)agiert (Welsch et al. 2018, S. 369; Braun 2018,

Was?	Wer?	Mensch		Maschine	
Stufe	0 (manuelles System)	1 (assistiertes Entscheiden)	2 (autonome Assistenz)	3 (lernende Assistenz)	4 (autonomes System)
Dauerhafte und vollständige Ausführung aller anfallenden Aufgaben und Entscheidungen	Übrige Aspekte der Aufgaben und Entscheidungen, Zielvorgabe	Übrige Aspekte der Aufgaben und Entscheidungen, Zielvorgabe	Überwachung, (Re-)Aktion bei Bedarf alle Aspekte der Aufgaben werden abgearbeitet und die Wege zur Zielerreichung angepasst (Lernen)	Dauerhafte und vollständige Ausführung aller anfallenden Aufgaben und Entscheidungen, Ziele werden selbst gesetzt	
Definierte Aufgaben (ein System)	Definierte Aufgaben mit fixiertem Ziel werden automatisch ausgeführt				

Abb. 3.9 Stufen der Automatisierung (eigene Erstellung in Anlehnung an Braun 2018; Welsch et al. 2018)

S. 13). Die in diesem Buch dargestellten digitalen Tools liegen auf den beiden Studien 2 und 3. Bei der nächsten Stufe 4 werden alle Aspekte der Aufgaben und Entscheidungen von der Maschine übernommen und sie passt sogar ihre Ziele an (autonomes System). Dies wäre jener Zeitpunkt, vor dem einige prominente Personen wie Elon Musk oder Stephen Hawking (medienwirksam) gewarnt haben: eine KI könnte dann Ziele entwickeln, für deren Verwirklichung Menschen „geopfert“ werden müssten, z. B. wäre der Schutz des Klimas einfacher, wenn es keine Menschen mehr gäbe (siehe dazu [Tegmark 2017](#)).

Im Bereich der Informations- und Kommunikationstechnologie gibt es zwei traditionelle Ansätze der Prozessautomatisierung: die Umsetzung von Prozessen durch standardisierte Anwendungssoftware, wie z. B. Enterprise Ressource Planning (ERP) oder Customer Relationship Management (CRM)-Systeme, und die Automatisierung von eigens modellierten Prozessen durch Business-Process-Management-Systeme (BPMS) (Czarnecki und Auth [2018](#), S. 116; Czarnecki et al. [2019](#)). Diese Systeme automatisieren im Wesentlichen wiederkehrende Schritte, denen einfache Entscheidungsregeln zugrunde liegen. Bei deren Abarbeitung sind noch menschliche Zu-Taten erforderlich, z. B. Eingabedaten vorbereiten und in Formularfelder eingeben; die Software verarbeitet die Eingabedaten dann automatisch. Um diese zusätzlichen manuellen Schritte geht es bei Robotic Process Automation (RPA).

Mit RPA werden manuelle Tätigkeiten am Computer durch Softwareroboter automatisiert und dadurch ersetzt (siehe Abb. 3.10). Dabei kommuniziert der Softwareroboter über die Präsentationsschicht mit darunterliegenden Systemen,

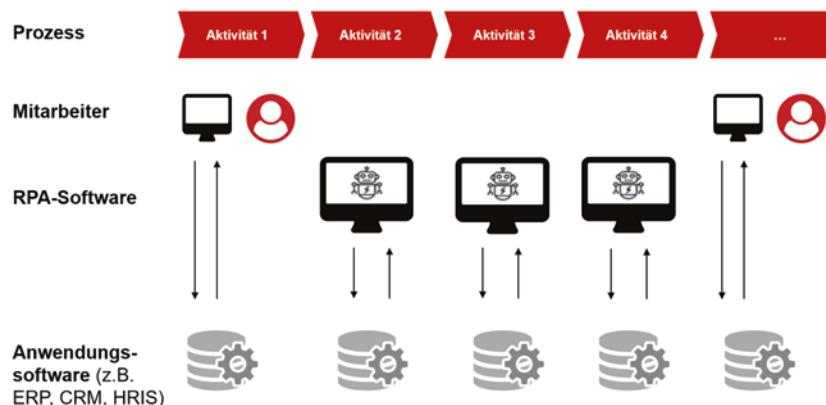


Abb. 3.10 Funktionslogik von Robotic Process Automation (eigene Erstellung)

sodass die bereits implementierte Anwendungssoftware (ERP, CRM, HR-Informationssystem, etc.) nicht geändert werden muss. Die Datenverarbeitung verläuft entlang der Prozessaktivitäten so wie vorher durch den Mitarbeiter (Czarnecki und Auth 2018, S. 116). Der Software-Bot agiert quasi als virtuelle Arbeitskraft, d. h. ihm ist ein Passwort zugeordnet, mit dem er Lese- und Schreibberechtigungen hat. So kann er mittels einer virtuellen Maus und einer virtuellen Tastatur Programme öffnen und z. B. E-Mails bearbeiten, Daten in Masken eintragen und aus unterschiedlichen Systemen integrieren, Berechnungen ausführen, Wenn-Dann-Regeln befolgen, auf soziale Medien zugreifen u. v. m. (Scheer 2017, S. 35). Ausgestattet mit diesem Funktionsumfang, können Bots dann Daten für Reports aus unterschiedlichen Systemen ziehen und den Report erstellen oder im zentralen HR-System Daten einer neuen Mitarbeiterin anlegen und diese Informationen dann über das E-Mail-System an weitere Stellen im Unternehmen verteilen (z. B. die IT-Abteilung, die dann für die neue Kollegin einen PC konfiguriert oder Zugangskarten kodiert). Dazu ist für jede Interaktion eine Anmeldung im entsprechenden Anwendungssystem notwendig, weshalb die Anzahl der serverseitig bearbeitbaren Client-Anfragen eine Restriktion für die Skalierbarkeit eines RPA-Systems bildet. Die Skalierbarkeit ist ebenfalls limitiert, wenn RPA im so genannten „attended mode“ läuft, weil dann ein Mitarbeiter den Software-Roboter auf seinem Gerät extra händisch starten muss (im „unattended mode“ läuft RPA auf virtuellen Maschinen im Backend tatsächlich automatisch). Unabhängig von der Art der Durchführung ist die Verwaltung der Zugangsdaten ein sicherheitsrelevantes Thema (Czarnecki et al. 2019, S. 799).

Da ein Softwareroboter nur jeweils ein Set an Verarbeitungsregeln und -parametern abbilden kann, ist der Einsatz von **RPA nur bei strukturiertem Dateninput mit regelbasierten Entscheidungen** technisch sinnvoll – ökonomisch vorteilhaft wird der Einsatz dann, wenn die Prozessschritte in hoher Frequenz bzw. einem hohen Volumen auftreten, keine Fehler gemacht werden dürfen und menschliche Arbeitskräfte nicht gleichzeitig billiger wären (Lacity und Willcocks 2016a, S. 43; Czarnecki und Auth 2018). In Unternehmen betrifft dies einige standardisierte Routineaufgaben, z. B. die Erfassung der Personalstammdaten (durch Übertragen von einer papierbasierten Bewerbung oder aus einer eigenen Bewerbersoftware in das zentrale HR-Informationssystem) oder die Erstellung von regelmäßigen Reports mit Kennzahlen über die Belegschaft (siehe Kap. 5).

Die Angaben zu **Kosteneinsparungen** durch RPA differieren stark und reichen von **20 bis 80 %** (Capgemini Consulting 2016; EY 2016a, b; Lacity und Willcocks 2016a, b; Scheer 2017; Gartner 2018). Die Unterschiede ergeben sich zunächst aus den Volumina der jeweiligen Prozesse sowie den Personalkosten eines Unternehmens – und welche Kosten genau verglichen wurden (z. B. nur die Lizenzen für

einen Software-Bot, die bei ca. 20 bis 35 % eines Mitarbeiters liegen sollen, oder auch eine Umlage der Implementierungskosten). Darüber hinaus hängen die Kosteneinsparungen von der Komplexität und Qualität des Prozesses ab, der automatisiert werden soll. Komplexe und nicht-standardisierte Prozesse erlauben kaum Kosteneinsparungen durch RPA. Recht ausführlich dokumentiert ist die Einführung von RPA für Back-Office-Tätigkeiten bei Telefónica UK, wo die Anzahl an Full-time Equivalents auf ca. ein Drittel gesenkt werden konnte und ein Return on Investment von 650–800 % über drei Jahre hinweg generiert werden konnte (Lacity et al. 2015; Lacity und Willcocks 2016b). Viele RPA-Protagonisten betonen allerdings immer wieder, dass es nicht nur um Kosteneinsparungen durch Personalreduktion gehen kann und soll, sondern durch RPA **strategische Vorteile** geschaffen werden können, wenn Mitarbeiter von belastenden Aufgaben befreit werden und Nutzer einen Mehrwert durch die Automatisierung erleben (Capgemini Consulting 2016; Lacity und Willcocks 2018).

3.8 Intelligent Process Automation und Chatbots

Für Prozesse, in denen unstrukturierte Daten (z. B. E-Mail-Texte, Audio, Bilder) verarbeitet werden, muss RPA mit ML-Algorithmen, insbesondere KNN (siehe Abschn. 3.5), verbunden werden, um Sprache zu verstehen, Bilder zu erkennen und Texte zu interpretieren. Mit diesem sogenannten **Intelligent Process Automation (IPA)** können **unstrukturierte Aufgaben** bearbeitet werden, was auch mit dem Begriff Cognitive Robotics Process Automation umschrieben wird (Lacity und Willcocks 2016a, S. 43; Masood und Hashmi 2019, S. 225). Mit diesen Funktionen werden die kognitiven Bots flexibler, weil sie auf Veränderungen der Systemumgebung reagieren können. So wird es beispielsweise durch die Bildverarbeitung möglich, dass Veränderungen in der Benutzeroberfläche einer Anwendung erkannt werden: Wenn die Personalnummer nicht mehr in das erste sondern dritte Feld einer Maske kommt, dann muss dies – anders als bei RPA – nicht in einer Regel hinterlegt werden, sondern kognitive Bots erkennen dies selbst. Außerdem lassen sich mit IPA nicht nur Nutzerfragen aufnehmen, Veränderungen in den Anwendungssystemen anstoßen und Ergebnisse zurückmelden, sondern der kognitive Bot lernt auf Basis früherer Anfragen und kann so bessere Ergebnisse ausspielen (siehe Abb. 3.11). Z. B. können text- oder sprachbasierte Urlaubsanfragen geprüft und der Urlaub im entsprechenden HR-System automatisch eingetragen werden (Masood und Hashmi 2019, S. 227 ff.). Im Stufensystem der Automatisierung (siehe Abb. 3.9) entspricht dies der Stufe 3. Bei IPA ist es nicht nur das Ziel, die Kosten zu reduzieren, sondern auch **zusätzlichen Nutzen** zu erzeugen, indem

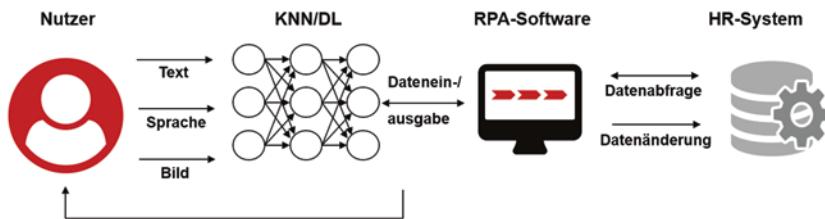


Abb. 3.11 Intelligent Process Automation (schematisch)

Leistungen angeboten werden, die es so entweder noch nicht gab oder nicht zu diesem Qualitätsniveau (Czarnecki et al. 2019, S. 804 f.).

Eine weit verbreitete Form von IPA sind **Chatbots** (oder: Dialogsysteme), die der breiten Masse durch Apples Siri, Amazons Alexa, Microsofts Cortana oder Googles Duplex bekannt sind. Grundidee ist, dass ein Dialog zwischen Nutzer und Chatbot zustande kommt, indem der Nutzer Eingaben macht (z. B. in Form einer Frage) und der Chatbot mit Ausgaben antwortet (z. B. in Form einer inhaltlichen Antwort, einer Gegenfrage zum besseren Verständnis oder einer Weiterleitung zu einem Menschen). Chatbots können entweder geschriebene oder gesprochene Sprache verarbeiten und dabei entweder einfach **Antworten regelbasiert abrufen (retrieval)** oder sie generieren (**generative models**) (Lokman und Ameedeen 2019, S. 1013; Czarnecki et al. 2019).

Regelbasiert kann in der einfachsten Ausprägung bedeuten, dass der Chatbot vorstrukturierte Frage-und-Antwort-Schemata abarbeitet, d. h. der Nutzer kann nur für bestimmte Fragen auf vorab definierte Antwortoptionen klicken. Freitexteingaben gibt es in der hochstrukturierten Form nicht (weshalb strittig ist, ob in diesem Fall überhaupt von IPA gesprochen werden kann oder es lediglich RPA ist). Etwas komplizierter wird es, wenn Fragen offen gestellt werden können und Antworten aus einem Fundus anhand der Wahrscheinlichkeit, dass sie zur Frage passen, ausgewählt werden. Eine einfache Matching-Regel wäre: Gebe die Antwort aus der Datenbank aus, bei der die dazugehörige Frage die höchste Übereinstimmung mit den Zeichen des eingegebenen Textes hat. Verglichen werden also nur die Anzahl übereinstimmender Zeichen zwischen Eingabe und gespeicherten Datensätzen. Da Dialoge normalerweise aus mehr als bloß einem Frage-Antwort-Paar bestehen, muss ein Chatbot den Dialog-Fortschritt speichern, um sich zu merken, was der Benutzer bereits mitgeteilt hat und in welchem Kontext die Äußerungen zu verstehen sind. Viele Chatbots bilden dies in sogenannten **Dialogbäumen** ab, die dem Chatbot Regeln vorgeben, in welcher Reihenfolge die vorgefertigten Antworten (normalerweise) kommen müssen (Stucki et al. 2018, S. 734).

Generierend meint, dass Antworten immer wieder neu generiert werden, was das „Lernen“ anhand von Trainings- und Testdaten voraussetzt. Die dafür eingesetzten ML-Verfahren variieren. Aktuell sind es vor allem RNN und LSTM, weil damit sequenzielle Inputdaten gut verarbeitet werden können (Lokman und Ameedeen 2019, S. 1015). Genau solche liegen bei Text oder Sprache vor: Wörter vom Anfang eines Satzes oder aus dem Satz davor müssen „erinnert“ werden können, um den Sinnzusammenhang zu verstehen. Neben diesen Unterscheidungskriterien gibt es weitere Aspekte wie die Frage, ob der Chatbot nur Wissen in einem bestimmten Feld hat oder offen für alle Wissensgebiete ist (was natürlich ungleich schwieriger ist; Lokman und Ameedeen 2019, S. 1013). Während Siri, Alexa und Co. offen konzipiert sind, müssen Chatbots im HR keine Informationen über das Wetter, die Müllabfuhrtermine oder das Kinoprogramm parat haben – es ist schon **aufwändig genug HR-bezogenes Domänenwissen abzubilden** (z. B. Fragen zum Einstellungsprozess oder wie die Bonusrunde abläuft), weil hierbei eine Vielzahl an Konzepten berücksichtigt werden muss (z. B. welche Rollen/Personen am Einstellungsprozess beteiligt sind, welche Auswahlverfahren zur Anwendung kommen, wie personenbezogene Daten gespeichert und weiterverarbeitet werden, usw.).

Chatbots, die nur regelbasierte Abrufe tätigen, machen keine Verständnisfehler, aber ihr Einsatzspektrum ist begrenzt. Bei generativen Chatbots ist die große Herausforderung, dass sie die **Absicht (Intent) hinter der Text- bzw. Spracheingabe erkennen** müssen. Beispielsweise müssen sie identifizieren, ob die Nutzerin nach einer bestimmten Stelle oder nach allgemeinen Informationen zum Arbeitsplatz sucht. Das kann über einfache Matching-Regeln geschehen (z. B. den Vergleich der Zeichenanzahl, siehe oben) oder wiederum über ML-Algorithmen zur Klassifikation. Bei SAP wird beispielsweise die Benutzerintention hinter einer Aussage gegen eine vordefinierte Liste von möglichen Absichten geprüft und klassifiziert. Zusätzliches Hintergrundwissen wie domänen spezifische Dokumente oder semantisch gleichwertige Trainingseingaben helfen dem KNN ein robustes Intent-Matching-Model aufzubauen (Leukert et al. 2019, S. 58). Vor dem Intent-Matching müssen lernende Chatbots aber zunächst die sprachlich verfassten Eingaben der Nutzer in maschinenlesbare umwandeln. Es müssen nicht nur Satzzeichen entfernt, Groß- und Kleinschreibung einheitlich gemacht und Tippfehler korrigiert (bzw. ignoriert) werden, sondern der unstrukturierte Text muss in numerische Werte übertragen werden, damit statistische Verfahren überhaupt angewendet werden können. Bei der Aufbereitung der Daten kommen eine Reihe von Verfahren des **Text Minings bzw. Natural Language Processing** zum Einsatz (z. B. Tokenizing, Stemming, Part-of-Speech; siehe Abschn. 4.2.1.2).

Aktuelle Chatbots sind meist geschlossene Hybride, d. h. sie sind auf ein oder wenige Wissensgebiete spezialisiert, wobei die Antworten sowohl regelbasiert abgerufen als auch generiert werden, wofür wiederum unterschiedliche ML-Verfahren eingesetzt werden (Lokman und Ameedeen 2019). Ebenso zahlreich wie die eingesetzten Verfahren sind die Anbieter: Von kleineren Firmen wie Jobpal oder Meya über die großen Konzerne wie Amazon, Google, IBM und Microsoft bis hin zur Open Source-Software von wit.ai und RASA gibt es verschiedene Chatbot-Services, die getrennt oder kombiniert Anwendung finden. Z. B. bietet Chatfuel ein Framework für die Botentwicklung ohne große Programmierkenntnisse, ähnliches bietet Microsoft mit dem QnA-Maker und der Sprachverarbeitungsumgebung LUIS, die vordefinierte Domänenmodelle, einschließlich vor-klassifizierter Nutzer-Absichten und Äußerungen enthält. Trotz ihrer Unterschiede haben die Systeme auch Gemeinsamkeiten im Aufbau: Dialogsysteme brauchen ein Modul für die automatische Spracherkennung (Analyse), das Verständnis der Spracheingabe (Interpretation) und die Sprachsynthese (Generierung) (Jokinen und McTear 2010, S. 32). Der Ablauf ist dann wie folgt: Ein Nutzer stellt eine Anfrage über eine text- bzw. sprachbasierte Schnittstelle, die von der Spracherkennung zunächst geparsst und dann durch das Sprachverständnis-Modul bzw. die Conversation Engine interpretiert wird. Aus den resultierenden Daten wird abschließend ein Antworttext generiert und an den Benutzer übermittelt (Czarnecki et al. 2019, S. 802).

3.9 Virtual und Augmented Reality

Virtual Reality (VR) ist eine **umfassende Simulation**, die für Menschen nicht mehr von der echten Realität zu unterscheiden ist, indem 3D-Inhalte in Echtzeit dargestellt werden und ihre Manipulation möglich ist. Diese Simulation wird aus einem System von geeigneter Software und Hardware erzeugt (Dörner et al. 2019a, S. 7, 12 f.). Bei **Augmented Reality** (AR) wird die Realität **durch virtuelle Inhalte ergänzt** und es kann mit ihnen in Echtzeit interagiert werden (Dörner et al. 2019a, S. 20). Beide Technologien nutzen Wearables, wie z. B. Brillen und Handschuhe oder auch Smartphones und Tablets, um nicht-vorhandene Informationen oder Gegenstände in die echte Umgebung einzublenden. Zusätzlich werden 3D-Interaktionsgeräte eingesetzt, die die Position des Nutzers im Raum verfolgen und Körperbewegungen erkennen, sodass mit der Umgebung interagiert werden kann. Mit jeder Bewegung des Nutzers wird die 3D-Umgebung automatisch aus seiner Perspektive neu generiert (Dörner et al. 2019a, S. 12 f.). Ein Ziel ist, den Nutzern das Gefühl des vollständigen Eintauchens (**Immersion**) in die virtuelle Welt zu ermöglichen (Dörner und Steinicke 2019, S. 56). Dann wird die virtuelle Um-

bung als real empfunden. Erreicht wird dies, indem ihnen multimodale Stimulatio-
nen rückgespielt werden, beispielsweise durch Objekte, die vibrieren, wenn man
sie berührt, wobei diese Vibration über den Handschuh übertragen wird. Dies ist
nur ein einfaches Beispiel für die **vielfältigen wahrnehmungspsychologischen
Prozesse** (insbesondere bei den visuellen, akustischen und haptischen Aspekten
menschlicher Wahrnehmung), die bei der Gestaltung effektiver virtueller Welten zu
berücksichtigen sind (siehe Dörner und Steinicke 2019, S. 46 ff.). Wegen dieser
personenbezogenen Prozesse aber auch wegen der unterschiedlichen technischen
Geräte, die zum Einsatz kommen, ist Immersion immer ein graduelles Merkmal
(Dörner et al. 2019a, S. 14).

VR und AR werden in vielen Feldern eingesetzt, unter anderem im Freizeit-
bereich (Videospiele, Freizeitparks), beim Produktdesign (virtuelle Modelle), für
Schulungen und interaktive Anleitungen (Bedienung von Maschinen), bei Raum-
planungen (Bürodesign, Innenarchitektur) und virtuellen Bürorundgängen oder im
Tourismus (Reiseort, Sehenswürdigkeiten) (siehe Jäger und Petry 2018, S. 48 ff.;
Dörner et al. 2019a). Für den Einsatz in Unternehmen haben sich vor allem Daten-
brillen, Handschuhe und Tablets etabliert. Beispielsweise setzt Thyssenkrupp zur
Wartung seiner Aufzüge und im Vertrieb die HoloLens-Technologie von Microsoft
ein (Thyssenkrupp 2016; Dilba 2018). Die Brille zeigt dem Servicetechniker be-
reits vor seinem Einsatz die spezifischen Kenndaten eines Aufzugs an und vor Ort
alle relevanten technischen Informationen. Die Wartung kann damit nicht nur
schneller, sondern auch sicherer erledigt werden, weil die Hände zum Arbeiten frei
bleiben, während die Informationen eingespielt werden. Bei fehlerhafter Bedie-
nung einer Maschine kann in der Brille oder auf einem Tablet ein rotes Signal er-
scheinen, das den Bediener zu einer Korrektur auffordert und auch Empfehlungen
für die richtige Handhabung anzeigt.

3.10 Fazit

KI und ML sind Oberbegriffe für Algorithmen und Systeme (Algorithmen +
Technik + Daten), mit deren Hilfe komplexe Probleme zielgerichtet gelöst werden
können. Dafür braucht es unterschiedliche Fähigkeiten: Erkennen, Verstehen (Be-
deutungen identifizieren), Erinnern, Lernen, Entscheidungen treffen, ausführen
und dabei mit anderen Dingen und Menschen interagieren. Für diese braucht es
wiederum weitere Fähigkeiten wie visuelle und emotionale Wahrnehmung, Wis-
sensrepräsentation, logisches Schlussfolgern, Mustererkennung, Text- und Sprach-
verarbeitung, etc. Selbst die besten **Maschinen sind nur für bestimmte (Teil)-
Aufgaben effektiv einsetzbar** – sie sind quasi Fachidioten und eignen sich nur bei
bestimmten Datenstrukturen (siehe Tab. 3.2).

Tab. 3.2 Eignung ausgewählter ML-Algorithmen

ML-Algorithmus	geeignet ...	nicht geeignet ...
Lineare Regressionen	... um auch aus kleinen Datensätzen Beziehungen zu extrahieren	... um nicht-lineare Zusammenhänge zu analysieren/modellieren und wenn die unabhängigen Variablen zusammenhängen (Multikollinearität)
Logistische Regression	... um die Wahrscheinlichkeit, dass ein Ereignis (nicht) eintritt schnell zu modellieren	... wenn Ausreißer oder Multikollinearität vorliegen oder Daten im Zeitverlauf repräsentiert werden sollen
Decision Trees	... um nicht-lineare Zusammenhänge unterschiedlichster Datentypen zu analysieren/modellieren und eingängig zu visualisieren	... wenn sich die Daten auch nur leicht verändern, weil dies zu ganz anderen Entscheidungsbaummodellen führt (Decision Trees sind wenig „robust“)
Convolutional Neural Networks(CNN)	... um unstrukturierte, große Datenmengen zu analysieren, die eine räumliche Beziehung haben (z. B. Bild-/Objekterkennung), oder wenn grundsätzlich Muster (auch nicht-lineare) aufgedeckt werden sollen	... wenn unterschiedliche Aufgaben gelöst werden sollen (sehr spezifisch trainiert: overfitting), Input-Daten widersprüchlich sind oder verändert werden und der Rechenweg nachvollziehbar sein muss (keine Black Box)
Recurrent Neural Networks (RNN)	... um unstrukturierte Daten, die eine zeitliche Beziehung haben (z. B. Text- & Sprache), zu analysieren oder wenn grundsätzlich Muster (auch nicht-lineare) aufgedeckt werden sollen	... siehe CNN; außerdem nicht geeignet bei strukturierten Daten („Tabellenform“) und Bilddaten

Viele ML-Algorithmen eignen sich nur für bestimmte Fragestellungen und bestimmte Verteilungen der Variablen im Datensatz (eigene Erstellung)

Ungeachtet der Einschränkungen sind die Ergebnisse, die bei der Bild-, Text-, Sprach- und der Mustererkennung im Allgemeinen in den letzten Jahrzehnten erreicht wurden, beeindruckend – weil sie die besten menschlichen Leistungen übertrumpft haben. **Da, wo Maschinen gut sind, sind menschliche Problemlöser meist schwach:** wir können kaum den Überblick über große Datenmengen behalten, wir tun uns schwer mit nicht-linearen Wirkungszusammenhängen,

insbesondere, wenn diese in mehrdimensionalen Räumen stattfinden (schon wenn Zeit als zusätzliche Dimension hinzukommt, können wir manche Zusammenhänge nicht mehr erkennen oder glauben welche zu erkennen, wo gar keine sind) und wir machen Fehler beim logischen Schlussfolgern. Wir sind ungenau, subjektiv, emotional und handeln auf Basis unserer Erfahrungen und Interessen. Diese Schwächen können zu Stärken werden, z. B. in sich ständig ändernden bzw. völlig neuen Situationen oder wenn Einfühlungsvermögen gefragt ist.

Von einer allgemeinen oder starken Künstlichen Intelligenz sind wir aber weit entfernt und es gibt starke Argumente dafür, dass dies noch lange so bleiben wird. Erstens, weil bisherige Verfahren sehr viele Daten zum Trainieren bzw. die Lösung eines Problems brauchen und diese nicht immer vorhanden sind. Zweitens, weil die Lösungsfunktionen einer KI aus einem Bereich nicht einfach auf andere übertragen werden können (fehlende Generalisierbarkeit). Drittens, weil sie keinen Körper – mitsamt seinen sensomotorischen Fähigkeiten, seiner Vergänglichkeit und einem (Über-)Lebensdrang – haben. Menschliche Intelligenz ist unter anderem durch diese drei Aspekte gekennzeichnet. Wir können auch aus nur einer Situation etwas lernen (wir langen normalerweise nur einmal auf die heiße Herdplatte), wir können unsere Lösungen generalisieren (Verhaltensweisen wie do-ut-des oder tit-for-tat setzen wir erfolgreich bei Freunden, Familienmitgliedern, Kollegen, Chefs und sogar politischen oder militärischen Konflikten ein), wir sind uns unserer Vergänglichkeit bewusst und achten daher auf unsere Gesundheit, schätzen Werte wie Fairness oder friedvolles Miteinander. Bei all diesen Vorzügen darf nicht vergessen werden: wir lernen nicht immer das Richtige, unterliegen kognitiven Verzerrungen sowie sozialerwünschtem Verhalten – und wir achten längst nicht immer auf unsere Gesundheit, zerstören unseren Lebensraum und greifen unsere Mitmenschen an. **Die Stärken der Maschinen können unsere Schwächen kompensieren – und vice versa.** Dies ist der tiefere Grund, warum man sich mit den digitalen Tools für die Personalarbeit beschäftigen sollte.

Sprache ist unser Mittel zur Welt – und mit VR-Technologien werden **aus Dialogsystemen virtuelle Assistenten**, die im Bestfall unsere Absichten verstehen und automatisch Prozesse zur Bearbeitung oder Lösung unseres Anliegens anstoßen. Auch hier gilt, dass die Erwartungen aktuell die technologischen Möglichkeiten übersteigen. Sprache und soziale Interaktion sind komplexe Phänomene, die sich kaum in probabilistische Systeme und noch weniger in feste Regeln pressen lassen. Dennoch: Es gibt Anwendungsfälle und sie werden immer mehr. Ein weiterer Grund, sich mit digitalen Tools zu beschäftigen, ist, durch die Auseinandersetzung realistische Erwartungen zu entwickeln.

Literatur

- Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. (2016). *Multivariate Analysemethoden – Eine anwendungsorientierte Einführung* (14., überarb. u. akt. Aufl.). Berlin/Heidelberg: Springer.
- Botvinick, M., Ritter, S., Wang, J. X., Kurth-Nelson, Z., Blundell, C., & Hassabis, D. (2019). Reinforcement learning, fast and slow. *Trends in Cognitive Sciences*, 23(5), 408–422.
- Braun, M. (2018). Digitale Assistenzsysteme. *Industrie 4.0 Management*, 34(4), 11–14.
- Capgemini Consulting. (2016). Robotic process automation – Robots conquer business processes in back offices. <https://www.capgemini.com/consulting-de/wp-content/uploads/sites/32/2017/08/robotic-process-automation-study.pdf>. Zugegriffen am 30.08.2017.
- Cheng, M. (2017). Causal modeling in HR Analytics: A practical guide to models, pitfalls, and suggestions. *Academy of Management Annual Meeting Proceedings*, 1, 1–6.
- Czarnecki, C., & Auth, G. (2018). Prozessdigitalisierung durch Robotic Process Automation. In T. Barton, C. Müller & C. Seel (Hrsg.), *Digitalisierung in Unternehmen* (S. 113–131). Wiesbaden: Springer.
- Czarnecki, C., Bensberg, F., & Auth, G. (2019). Die Rolle von Softwarerobotern für die zukünftige Arbeitswelt. *HMD Praxis der Wirtschaftsinformatik*, 56(4), 795–808.
- Davenport, T. H., Harris, J., & Shapiro, J. (2010). Competing on talent analytics. *Harvard Business Review*, 88(10), 52–58.
- Dilba, D. (2018). Deal in Sicht. *brand eins*, 7, 90–93.
- Dörner, R., & Steinicke, F. (2019). Wahrnehmungsaspekt von VR. In R. Dörner, W. Broll, B. Jung & P. Grimm (Hrsg.), *Virtual und Augmented Reality (VR/AR): Grundlagen und Methoden der Virtuellen und Augmentierten Realität* (2. Aufl., S. 43–78). Berlin/Heidelberg: Springer.
- Dörner, R., Kuhlen, T. W., Matthys, G., et al. (2019a). Fallbeispiele für VR/AR. In R. Dörner, W. Broll, B. Jung & P. Grimm (Hrsg.), *Virtual und Augmented Reality (VR/AR): Grundlagen und Methoden der Virtuellen und Augmentierten Realität* (2. Aufl., S. 357–392). Berlin/Heidelberg: Springer.
- EY. (2016a). Ernst & Young: Jobstudie 2016. [https://www.ey.com/Publication/vwLUAssets-SPI/EY_Jobstudie_2016/\\$FILE/EY-Jobstudie-2016-Karriere-und-Wechselbereitschaft.pdf](https://www.ey.com/Publication/vwLUAssets-SPI/EY_Jobstudie_2016/$FILE/EY-Jobstudie-2016-Karriere-und-Wechselbereitschaft.pdf). Zugegriffen am 01.07.2017.
- EY. (2016b). *Robotic process automation for HR & payroll* (White paper).
- Fitz-enz, J. (1984). *How to measure human resources management*. New York: McGraw-Hill.
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144.
- Gartner. (2018). Manage robotic process automation. <https://www.gartner.com/en/finance/trends/robotic-process-automation>. Zugegriffen am 26.08.2019.
- Gärtner, C. (2007). *Innovationsmanagement als soziale Praxis: grundlagentheoretische Vorarbeiten zu einer Organisationstheorie des Neuen*. München/Mering: Rainer Hampp.
- Gärtner, C. (2017). Mensch oder Maschine: Wer trifft die besseren Personalentscheidungen? *HR Performance*, 2(2017), 14–17.
- Gärtner, C. (2019). „10 + x Gebote“ für ein erfolgreiches Change Management. *change-ment*, 1, 38–39.
- Goodfellow, I., Bengio, Y., & Courville, A. (2018). *Deep learning: Das umfassende Handbuch*. Frechen: mitp.

- Hutson, A. (2017). AI glossary: Artificial intelligence, in so many words. *Science*, 357(6346), 19.
- Jäger, W., & Petry, T. (2018). Digital HR – Ein Überblick Petry. In T. Petry & W. Jäger (Hrsg.), *Digital HR: Smarte und agile Systeme, Prozesse und Strukturen im Personalmanagement* (S. 27–99). Freiburg: Haufe.
- Jokinen, K., & McTear, M. (2010). *Spoken dialogue systems synthesis lectures on human language technologies*. San Rafael: Morgan & Claypool Publishers.
- Knaflic, C. N. (2015). *Storytelling with data: A data visualization guide for business professionals*. Hoboken: Wiley.
- Lacity, M., & Willcocks, L. P. (2016a). A new approach for automating services. *MIT Sloan Management Review*, 58, 40–49.
- Lacity, M., & Willcocks, L. P. (2016b). Robotic process automation at Telefónica O2. *MIS Quarterly Executive*, 15, 21–35.
- Lacity, M., & Willcocks, L. P. (2018). *Robotic process and cognitive automation: The next phase*. Ashford: SB Publishing.
- Lacity, M., Willcocks, L. P., & Craig, A. (2015). *Robotic process automation at Telefónica O2* (The outsourcing unit working research paper series). London School of Economics and Political Science, Paper 15/02.
- Lawler, E. E., III, Levenson, A. R., & Boudreau, J. W. (2004). HR metrics and analytics: Use and impact. *People and Strategy*, 27(4), 27–35.
- Leukert, B., Müller, J., & Noga, M. (2019). Das intelligente Unternehmen: Maschinelles Lernen mit SAP zielgerichtet einsetzen. In P. Buxmann & H. Schmidt (Hrsg.), *Künstliche Intelligenz* (S. 41–62). Wiesbaden: Springer Gabler.
- Lokman, A. S., & Ameedeen, M. A. (2019). Modern chatbot systems: A technical review. In K. Arai, R. Bhatia & S. Kapoor (Hrsg.), *Proceedings of the future technologies conference 2018. Advances in intelligent systems and computing* (Bd. 881, S. 1012–1023). Cham: Springer.
- Marcus, G., & Davis, E. (2019). *Rebooting AI: building artificial intelligence we can trust*. New York: Pantheon.
- Marler, J. H., & Boudreau, J. W. (2017). An evidence-based review of HR analytics. *The International Journal of Human Resource Management*, 28(1), 3–26.
- Martini, M. (2019). *Blackbox Algorithmus – Grundfragen einer Regulierung Künstlicher Intelligenz*. Berlin/Heidelberg: Springer.
- Masood, A., & Hashmi, A. (2019). Cognitive robotics process automation: Automate this! In A. Masood & A. Hashmi (Hrsg.), *Cognitive computing recipes* (S. 225–287). Berkeley: Apress.
- Mühlbauer, D., Huff, J., & Süß, J. (2018). People Analytics und Arbeit 4.0. In S. Werther & L. Bruckner (Hrsg.), *Arbeit 4.0 aktiv gestalten* (S. 107–132). Berlin/Wiesbaden: Springer.
- Pearl, J., & Mackenzie, D. (2019). *The book of Why: The new science of cause and effect*. London: Penguin.
- Provost, F., & Fawcett, T. (2017). *Data Science für Unternehmen: Data Mining und datenanalytisches Denken praktisch anwenden*. Frechen: mitp.
- Russell, S., & Norvig, P. (2013). *Künstliche Intelligenz – Ein moderner Ansatz* (3., akt. Aufl.). München: Pearson.
- Scheer, A.-W. (2017). *Performancesteigerung durch Automatisierung von Geschäftsprozessen* (Whitepaper). Saarbrücken: Scheer Holding.

-
- Scholz, T. M. (2017). *Big data in organizations and the role of human resource management: A complex systems theory-based conceptualization*. Frankfurt a. M.: Peter Lang.
- Siegel, E. (2013). *Predictive analytics: The power to predict who will click, buy, lie, or die*. Hoboken: Wiley.
- Stouten, J., Rousseau, D. M., & De Cremer, D. (2018). Successful organizational change: Integrating the management practice and scholarly literatures. *Academy of Management Annals*, 12(2), 752–788.
- Strohmeier, S. (2017). Big HR Data – Konzept zwischen Akzeptanz und Ablehnung. In W. Jochmann, I. Böckenholt & S. Diestel (Hrsg.), *HR-Exzellenz: Innovative Ansätze in Leadership und Transformation* (S. 339–355). Wiesbaden: Springer Gabler.
- 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.
- Tegmark, M. (2017). *Leben 3.0: Mensch sein im Zeitalter Künstlicher Intelligenz*. Berlin: Ullstein.
- Thyssenkrupp. (2016). Thyssenkrupp treibt Digitalisierung des weltweiten Aufzugsservice weiter voran. <http://www.thyssenkrupp-elevator.com/de/newsroom/press-releases-20928.html>. Zugegriffen am 29.11.2019.
- Welsch, A., Eitle, V., & Buxmann, P. (2018). Maschinelles Lernen. *HMD Praxis der Wirtschaftsinformatik*, 55(2), 366–382.

Digitale Tools für smarte Personalarbeit

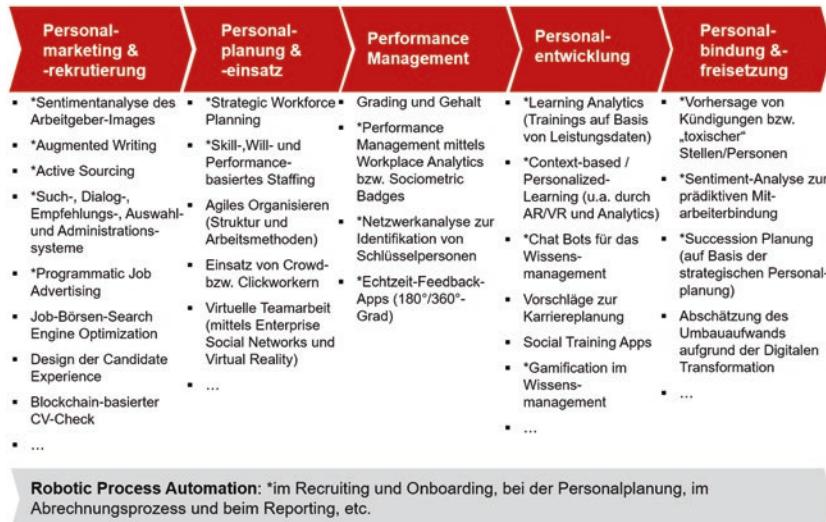
4

Zusammenfassung

Das Kapitel stellt zahlreiche Anwendungsfälle für digitale Tools in der Personalarbeit dar. Entlang der HR-Wertschöpfungskette wird jeweils erörtert, wie Analytics und Automatisierung zu innovativen HR-Produkten und Dienstleistungen beitragen können. Der Einsatz unterschiedlicher digitaler Tools wird für fünf Personalteifunktionen diskutiert (Personalmarketing und -rekrutierung, Personalplanung und -einsatz, Performance Management, Personalentwicklung sowie Personalbindung und -freisetzung), wobei jeweils vorgestellt wird, worum es geht, wie es geht, was man noch machen kann und was (noch) nicht geht. Neben konkreten Unternehmensbeispielen werden Ergebnisse aktueller wissenschaftlicher Studien herangezogen, um die Möglichkeiten und Grenzen praxisnah und fundiert ausloten zu können. Unter anderem werden der Einsatz von Sentiment- und Text-Mining-Analysen, Matching- und Empfehlungsalgorithmen, Chatbots sowie Augmented und Virtual Reality erläutert.

4.1 Use Cases entlang der HR-Wertschöpfungskette

Datenanalytische Verfahren und intelligente Prozessautomatisierung ermöglichen **neue Produkte und Dienstleistungen entlang der gesamten HR-Wertschöpfungskette**, vom Finden bis zum Binden des Personals (siehe Abb. 4.1). In manchen Anwendungsfeldern, wie z. B. der text- oder sprachba-



* Diese Use Cases werden nachfolgend beschrieben. Für weitere Anwendungsbeispiele und Erläuterungen siehe Gartner (2020).

Abb. 4.1 Anwendungsbeispiele entlang der HR-Wertkette

sierten Interaktion mit Mitarbeitenden, treten Analytics und Automatisierung in Kombination auf. Deshalb werden die folgenden ausgewählten Beispiele entlang der Wertkette diskutiert und darauf verwiesen, wie Analytics und Automatisierung zusammenwirken. Manchmal ist dies nur oberflächlich möglich, weil die Verarbeitungsprozesse in Tools eingebaut sind und die Anbieter die dahinter liegenden Algorithmen und Automatisierungsverfahren nicht offenlegen. Dies ist vor allem bei den großen Anbietern von integrierten Softwarepaketen der Fall (Cornerstone, Oracle PeopleSoft, SAP Successfactors, Workday, Visier Workforce Analytics, usw.). Daneben gibt es viele HR Startups, die nicht selten nur auf eine Personalteilfunktion spezialisiert sind (für eine Liste von HR Startups siehe www.smarthrm.de). Dass nicht alle Anbieter am Markt bestehen bleiben werden, versteht sich eigentlich von selbst, wird aber noch deutlicher, wenn man bedenkt, dass selbst Google sein Recruiting-Angebot namens Hire im September 2020 einstellt. Für die statistischen Analysen allein (ohne darüber hinausgehende Prozessautomatisierung) werden oft weithin bekannte Tools eingesetzt: SPSS, R, SAS, Tableau und Excel – letzteres rangierte bei einer Umfrage unter 340 HR-Experten sogar auf Platz 1 der Tools für prädiktive Analysen (Lunsford und Phillips 2018, S. 10). Im Personalumfeld weniger bekannt sind Programme zur Datenanalyse wie Alteryx, RapidMiner oder MATLAB und Tools zur Datenaufbereitung wie z. B. OpenRefine, WinPure oder Trifacta

Wrangler. Nicht nur der Funktionsumfang variiert zwischen diesen Tools, auch die Nutzerfreundlichkeit: Altery und RapidMinder bieten beispielsweise sehr viele Funktionen, die zudem über eine grafische Benutzeroberfläche und „Drag & Drop“ verwendet werden können. Während diese Instrumenten die Arbeit von Data Science-Expertenoft erleichtern, verwenden RPA-Spezialisten oft die Softwarepakete von UiPath, BluePrism, Automation Anywhere oder Kofax, um nur einige zu nennen. Designer von AR/VR-Angeboten müssen nicht nur entsprechende Softwarewerkzeuge beherrschen (z. B. Game-Engines wie Unity oder Unreal Engine und AR-Frameworks wie ARCore oder ARKit), sondern auch die Hardware-Komponenten (z. B. Headsets wie Oculus Rift oder HTC Vive) sowie die Schnittstellen und Konvertierungswerkzeuge, um die unterschiedlichen Elemente in einem System zum Funktionieren zu bringen. Bei der Auswahl all dieser Tools, sollten Anwender nicht nur auf die Funktionalität, Nutzerfreundlichkeit, Leistungsfähigkeit und Kosten (insbesondere Implementierungs- und Lizenzierungskosten) achten, sondern auch auf ihre eigenen Fähigkeiten sowie Unterstützungsangebote (angefangen von Tutorials und Schulungen bis hin zu einer aktiven User-Community).

Die Auswahl der Beispiele basiert auf der Einschätzung des Autors, welche **Anwendungen (Use Cases) sich entweder praktisch oder wissenschaftlich bewährt** haben. Dennoch bleibt es eine subjektive Auswahl, die auch nicht alle Anwendungsfälle umfasst. Beispielsweise wird nicht auf den Einsatz von Virtual und Augmented Reality-Anwendungen im Personalmarketing und Recruiting eingegangen, obwohl viele große Organisationen – angefangen von der British Army bis zu Bayer – davon Gebrauch machen. Deutlich gemacht werden muss, dass es sich tatsächlich um Use Cases und **nicht Business Cases** handelt, d. h. dass keine Wirtschaftlichkeitsbetrachtung erfolgt, unter anderem, weil diese in der Praxis auch nicht immer vorgenommen wird. Deswegen ist mit den Ausführungen auch keine Empfehlung verbunden, schon gar nicht für einen der Anbieter. Eher lautet die Empfehlung, die anstehenden **Herausforderungen im Lichte der zur Verfügung stehenden Tools zu priorisieren**, dann entsprechende Auswahlentscheidungen zu treffen und das Projekt als Lernerfahrung zu begreifen und zu positionieren. Ein solches Vorgehen berücksichtigt sowohl die betrieblichen Gegebenheiten als auch die marktseitigen Gelegenheiten, die neue Tools bieten und durch die unter Umständen Fragen aufgeworfen werden, an die vorher noch gar nicht zu denken war. Eine weitere Empfehlung schließt sich dem an: Nicht zu viele unterschiedliche Tools benutzen. Es hat sich in der Praxis immer wieder gezeigt, dass mehrere unterschiedliche IT-Systeme mehr Probleme schaffen als ein integriertes – nicht nur technische (Bondarouk et al. 2017), sondern z. B. auch mehr „Technostress“ für die HR-Abteilung (Florkowski 2019).

Deutlich wird, dass es vor allem viele Angebote im Recruiting-Umfeld gibt, was vermutlich daran liegt, dass hier auf viele Daten zurückgegriffen werden kann und keine großen Datenschutzprobleme bestehen, da die Daten bei der Bewerbung oder in sozialen bzw. beruflichen Netzwerken freiwillig geteilt werden. Außerdem wiederholen sich die Aufgaben, für die Algorithmen während des Recruiting-Prozesses eingesetzt werden, in späteren Phasen der Wertkette: Immer wieder geht es um das **automatisierte Suchen, Vergleichen (Matching) und Vorschlagen**, z. B. von neuen Stellen, Teamkollegen, Lerninhalten oder Vorgehensweisen.

4.2 Personalmarketing & -rekrutierung

Ziel des Personalmarketings ist es, dass potenzielle Mitarbeitende das Unternehmen als attraktiven Arbeitgeber wahrnehmen. Dies kann erreicht werden, indem die Bekanntheit des Unternehmens unter den potenziellen Mitarbeitenden gesteigert wird und dessen Alleinstellungsmerkmale als Arbeitgeber deutlich gemacht werden. Über diese **Profilierungs- und Akquisitionsfunktion** hinaus kann Personalmarketing auch die **Motivierung** der aktuellen Mitarbeitenden umfassen, sodass sich sogar eine Bindungswirkung aus den Personalmarketingaufgaben ableiten lässt (Scholz & Scholz 2019, S. 143 ff.). Diesem Aspekt wird im Folgenden nicht weiter nachgegangen.

Mit Personalrekrutierung ist sowohl die bedarfs- & zielgruppenorientierte Ansprache als auch die Auswahl von Bewerbern gemeint. Im Folgenden stehen Fragen der Arbeitgeberattraktivität und Selektion im Vordergrund.

4.2.1 Sentiment-Analyse des Arbeitgeberimages

4.2.1.1 Worum es geht

Sentiment-Analysen (**Stimmungsanalysen**) zielen darauf ab, herauszufinden, welche Meinung oder emotionale Einstellung Einzelne bzw. eine Gruppe zu einem bestimmten Thema, Produkt oder einem anderen Bezugsobjekt haben. Klassischerweise finden Sentiment-Analysen bei der **automatisierten Auswertung von Posts in sozialen Netzwerken** statt, um herauszufinden, ob die Gefühlslage zu einem Thema positiv, negativ oder neutral ist. Nach diesem Muster werden auch Kundenfeedbacks und Online Reviews (z. B. zu Konsumprodukten oder Filmen) seit Jahren ausgewertet (Liu 2015). Zusätzlich zur Analyse lassen sich solche Ergebnisse auch für Prognosen einsetzen, beispielsweise um mögliche Reaktionen auf ein neues Produkt zu antizipieren (Liu 2015, S. 3). Wie so oft, kommen Ideen und Tools aus dem Marketing irgendwann im HR-Umfeld an. Im Fall von Sentiment-Analysen werden statt der Markenwahrnehmung oder Produkt-Reviews nun eben **Bewertungen von Arbeitgebern** oder der **Stimmung am Arbeitsplatz** ana-

lysiert. In Zeiten, in denen Talente knapp sind, kann mit Hilfe von Sentiment-Analysen herausgefunden werden, wie die Wahrnehmung des Arbeitgebers- bzw. Arbeitsplatzes ist und welche Aspekte dazu beitragen, warum sie gut oder schlecht ist. Nutzen lassen sich solche Informationen im **Personalmarketing**, sowohl für eine angepasste Ansprache der Zielgruppen als auch die Kommunikation des Employer Brandings. Da zudem mit der Verbreitung von Social Media letztlich alle Mitarbeiter durch ihre Posts die **Reputation des Unternehmens bei Kunden, Kollegen und Bewerbern** beeinflussen können, ist es noch wichtiger geworden, herauszufinden, welche Einstellung und Emotionen die Mitarbeiter dem Unternehmen gegenüber haben. Dabei ist auch daran zu denken, dass es eine **Rückkopplung der Unternehmensreputation auf das Mitarbeiterverhalten** geben kann (Ivens 2018, S. 59 ff.).

4.2.1.2 Wie es geht

Die Datensammlung erfolgt meist über einen Crawler, der soziale Medien und Portale ausliest (z. B. Twitter, Facebook, Glassdoor, kununu), aber auch firmeninterne Daten wie Mitarbeiterumfragen, E-Mails, Posts im Intranet oder sogar Sprach- und Videoaufzeichnungen können grundsätzlich in das anschließende Text Mining eingehen. Hier werden dann **kognitive Einstellungen** (z. B. Beurteilungen, Bewertungen) und/oder **emotionale Reaktionen** (z. B. Ärger, Hass) ausgewertet. Im Falle einer Sentiment-Analyse des Arbeitgeberimages wird untersucht, welche Meinungen am Arbeitsmarkt über ein bestimmtes Unternehmen vorherrschen, ob es also eher positiv, neutral oder negativ wahrgenommen wird.

Nachdem die Daten eingesammelt wurden, werden die Texte in kleinere Bestandteile zerlegt, normalerweise in Sätze und Wörter. Diesen qualitativen Merkmalen werden Zahlenwerte zugeordnet, sodass nicht mehr nominal skalierte Daten, sondern ordinal skalierte vorliegen, die sich in eine Reihenfolge bringen lassen. Um aus Texten die Stimmung herauszulesen, gibt es letztlich drei Vorgehensweisen: Entweder man benutzt **Lexika**, in denen die Sentiments vieler Wörter als Zahlenwerte definiert sind, oder man setzt **ML-Verfahren** ein, die mit einem historischen Datensatz trainiert wurden – oder man wählt eine **Kombination** aus beidem. Solche hybriden Ansätze haben sich mittlerweile als Standard etabliert, insbesondere, wenn die Sentiment-Analyse nicht nur auf Textdaten, sondern auch auf Bilder oder Videos angewendet wird. Bei solchen multimodalen Sentiment-Analysen setzen Forscher neben Lexika vor allem CNN, LSTM und Klassifikationsalgorithmen, insbesondere Bayes-Netze (die mit bedingten Wahrscheinlichkeiten arbeiten) und Support Vector Machines, ein (Shi et al. 2019, S. 1246, 1265). Viele kommerziell tätige Anbieter wie IBM (Watson), Microsoft, SAP oder Adobe, aber auch kleinere Anbieter wie RapidMiner, offerieren Softwarepakete, mit deren Hilfe Online-Reviews, Social-Media-Beiträge, interne Dokumente und offizielle Publikationen analysiert werden können. Meist lassen sich

mit diesen Tools die eingelesenen Texte danach auswerten, in welchem (Online-) Kanal, in welcher Region (Sprache) und in welcher Zeitspanne positiv, neutral oder negativ über den Arbeitgeber gesprochen wurde, indem die in den Texten verwendeten Begriffe entsprechend kategorisiert werden. Für die Kategorisierung greifen die Tools auf Sentimentwörterbücher und ML-Modelle zurück.

Sentimentwörterbücher sind Listen, in denen tausenden von Begriffen jeweils eine Gewichtung zwischen –1 (negativ) bis +1 (positiv) zugeordnet wurde (in manchen Lexika zusätzlich noch, ob ein Begriff verstärkend, abschwächend oder bedeutungsumkehrend wirkt). Einige Universitäten haben (deutschsprachige) Wörterbücher zur Verfügung gestellt, z. B. „SentimentWortschatz“ (Universität Leipzig), „German Polarity Clues“ (Universität Bielefeld) oder der „Multi-layer Reference Corpus for German Sentiment“ (Universität Mannheim). Eine weit verbreitete Liste ist SentimentWortschatz (SentiWS), die insgesamt 1644 positive und 1827 negative Wortformen (inkl. Flexionsvarianten sind es insgesamt über 31.000 Einträge) enthält, wobei nicht nur Adjektive und Adverbien, sondern auch Nomen und Verben die Träger von Sentiments sind (Remus et al. 2010; SentiWS wurde 2018 aktualisiert). Beispielsweise ist dem Adjektiv „gut“ ein Wert von 0,3716 zugeordnet, dem Nomen „Desinteresse“ ein Wert von –0,3348. Die Gewichtungen helfen dabei zu erkennen, dass der Satz „Das Arbeitsklima ist gut“ weniger positiv ist als der Satz „Das Arbeitsklima ist super“ („super“ hat einen Gewichtungsfaktor von 0,5012). Bei der Analyse werden die Texte auf die **Anzahl und Sentiment-Gewichtung der vor-kommenden Begriffe** hin ausgewertet, sodass beispielsweise positive Gesamtwerte zu einer Klassifizierung als positive Stimmung führen. Zusätzlich müssen noch verstärkende und schwächende Begriffe (z. B. sehr, wenig) und Verneinungen über weitere Wortlisten berücksichtigt werden.

Bei Sentiment-Analysen, die **rein mit ML-Algorithmen** durchgeführt werden, kommen keine Lexika zum Einsatz. Stattdessen lernen die Systeme zunächst ein Modell auf Basis von Beispielinstanzen (Trainingsdaten), deren Sentiments vorab manuell bestimmt wurden. Mit dem Modell können dann die Sentiments von neuen, zuvor nicht klassifizierten Textbeispielen vorhergesagt werden. Anschließend wird die Genauigkeit der Vorhersage anhand einer Testmenge evaluiert. Bevor



Abb. 4.2 Pre-Processing bei Sentiment-Analysen

Texte (auf Sentiments hin) analysiert werden können, muss der Textkorpus bearbeitet werden. In diesem sogenannten und meist sehr aufwändigem **Pre-Processing** muss der Text umgewandelt und reduziert werden (siehe Abb. 4.2).

Beim Tokenizing wird der Text in Bausteine aufgeteilt, z. B. wortweise oder satzweise. Dann müssen irrelevante Wörter und Zeichen herausgefiltert werden. Erst danach erfolgt die Auswertung der (**relativen**) **Häufigkeit von positiven, negativen und neutralen Wörtern bzw. Wortstämmen** und den damit verbundenen Sentiment-Kategorien. Da ML-Algorithmen wie Naïve Bayes oder Support Vector Machines nur numerische Werte verarbeiten können, müssen die Texte noch in Zahlen umgewandelt werden. Je nach Datengrundlage und Erkenntnisinteresse muss während des Analyseprozesses an unterschiedlichen Stellen **mit unterschiedlichen Parameter-Einstellungen gefiltert, reduziert und ausgezählt** werden. Deshalb wird im Folgenden nur ein idealtypisches Vorgehen einer Sentiment-Analyse mit RapidMiner beschrieben.

Beispiel: Sentiment-Analyse mit RapidMiner

RapidMiner ist eine ursprünglich von der TU Dortmund entwickelte Software für maschinelles Lernen und Data-Mining. Das Tool hält eine ganze Reihe von Datenanalyseverfahren bereit, angefangen bei einfachen Regressionen über Entscheidungsbäume bis hin zu künstlichen neuronalen Netzen (siehe Abschn. 3.4 und 3.5). Diese und weitere Hilfsmittel, wie z. B. Wörterbücher, können über eine graphische Benutzeroberfläche zu Datenanalysemodellen zusammengebaut werden.

Zunächst müssen die **Daten geladen und strukturiert** werden. Strukturieren meint sowohl die Umwandlung von nominalen Daten in numerische als auch die Anwendung unterschiedlicher Filter und Wort(stamm)reduktionen. Für das Einlesen (Operatoren „Retrieve“ und „Read ...“) aller Dokumente eines Ordners gibt es

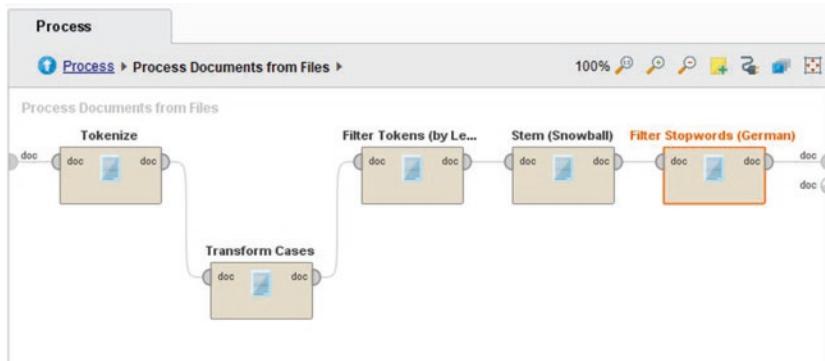


Abb. 4.3 Text-Mining mit RapidMiner: Operatoren innerhalb der Funktion „Process Documents“

eine Loop-Funktion, welche die Arbeit vereinfacht, wenn die Daten in unterschiedlichen Dateien gespeichert sind. Mit dem Operator „**Process Documents** (from Data)“ werden die Texte strukturiert. Dafür beinhaltet der Operator weitere Operatoren, in diesem Fall „Tokenize“, „Transform Cases“, „Filter Tokens by Length“, „Stem (Snowball)“ und „Filter Stopwords (German)“ (siehe Abb. 4.3).

Mit dem Operator „**Tokenize**“ wird jedes Dokument in durch Leer- und Sonderzeichen getrennte Wörter aufgeteilt (sogenannte Tokens), wobei auch andere Kriterien zur Auftrennung gewählt werden können (z. B. nach einem Doppelpunkt oder anderen Zeichen). Zudem können die Tokens reduziert werden, indem nur eine bestimmte Zeichenanzahl zugelassen wird (z. B. mehr als 3 und weniger als 30). Die nun extrahierten Merkmale werden vom Operator „**Transform Cases**“ alle in Kleinschreibung umgewandelt. Dieser Schritt ist nicht notwendig und sollte ausgelassen werden, falls ein Lexikon eingebunden wird, das Klein- & Großschreibung berücksichtigt, weil sonst keine Übereinstimmungen gefunden werden. Wenn aber die Daten z. B. aus Social-Media-Kanälen stammen, dann ist er hilfreich, weil dort nicht alle Nutzer die Groß- und Kleinschreibung beachten und sonst gleiche Wörter nur aufgrund der falschen Schreibweise unterschieden würden. Ebenso sinnvoll ist es, die Anzahl der Wörter im nächsten Schritt zu reduzieren, indem sogenannte **Stoppwörter** herausgefiltert werden (Operator „Filter Stopwords“), weil diese zwar oft vorkommen, aber keine Stimmung transportieren. In der deutschen Sprache sind dies z. B. unbestimmte Artikel, Präpositionen, Konjunktionen und Hilfsverben, aber auch Satzzeichen. In RapidMiner gibt es vordefinierte Stopwortlisten. Allerdings kann es geboten sein, eigene Stoppwörter hinzuzufügen. Z. B. könnte bei einer Sentiment-Analyse des Arbeitgeberimages der ML-Algorithmus aus den Kommentaren lernen, dass der Firmenname die beste Vorhersage für ein positives (bzw. negatives) Sentiment liefert – und dann jedes Mal, wenn die Firma in einer Bewertung genannt wird, direkt darauf schließen, dass es sich um eine positive (bzw. negative) Meinungsäußerung handelt. Dies wäre ein Fall der sogenannten Überanpassung (siehe Abschn. 3.4), welche dadurch vermieden wird, dass der Firmenname der Stopwortliste hinzugefügt und somit während der Analyse herausgefiltert wird. Ähnlich verhält es sich, wenn Dokumente per Webcrawler eingelesen wurden. Dann kann es sein, dass URLs und Links entfernt werden müssen, damit diese das Ergebnis nicht beeinflussen. Neben der Filterung nach Tokenlänge und Stoppwörtern sollte eine **Stammformreduktion** („stemming“) vorgenommen werden, die konjugierte und deklinierte Worte auf ihren Wortstamm zurückführt. Es werden z. B. Pluralendungen aber auch längere Wortteile gekürzt (z. B. wird „spielerisch“ zu „spiel“ gekürzt). RapidMiner bietet verschiedene Operatoren für das „**Stemming**“ (hier: „**Stem-Snowball**“, weil dieser auch für die deutsche Sprache eingestellt werden kann). Beim **Pruning** werden Wort(teil)e, die selten vorkommen, entfernt. Bei Sentiment-Analysen kann es passieren, dass Worte zwar gleich über die Katego-

rien positiv, negativ und neutral verteilt sind, aber vom Algorithmus unverhältnismäßig oft einer Kategorie zugeordnet werden (z. B. der positiven). Dann muss die Grenze für das Pruning, also bei welchem Prozentsatz an Textbeispielen ein Wort(teil) vorkommen muss, manuell bestimmt werden (z. B. 2 %, d. h. alle Wörter, die in weniger als 2 % der Texte vorkommen, werden aus der Analyse herausgenommen) (Abel et al. 2017).

Mit dem nächsten Schritt – **Part-of-Speech-Tagging** (Zuweisung der Wortart) – beginnt die syntaktische Textanalyse: Über den Operator „Filter Tokens (by POS Tags)“ können Wortarten, die normalerweise kein Sentiment transportieren, ausgeschlossen werden (z. B. Konjunktionen, Präpositionen, Pronomen). Diese Aufgabe übernimmt normalerweise ein POS-Tagger, der vorher über eine Beispielmenge mit bereits annotierten Texten angelernt wurde. Zwar gibt es solche annotierten Korpusse für verschiedene Sprachen, aber für die deutsche Sprache bietet RapidMiner nur unzureichende Unterstützung. Gleches gilt für die Analyse doppeldeutiger Wörter. Auch Abkürzungen müssen ausgeschrieben werden, da sie ansonsten nicht erkannt werden.

Erst nach diesen zahlreichen Schritten wird das **ML-Modell** trainiert, auf Testdaten angewendet und die Performance bestimmt. Da es um eine Klassifizierungsaufgabe geht, kommen entsprechende Algorithmen zum Einsatz, angefangen bei einfacheren Verfahren wie Naïve Bayes, logistische Regression oder Support Vector Machine (eine SVM sucht „Stützvektoren“ – Support Vectors –, die den Datensatz so aufteilen, dass um die Klassengrenzen ein möglichst breiter Bereich ohne Daten bleibt) bis zu komplexeren künstlichen Netzwerken wie LSTM.

Dem eben beschriebenen Verfahren sehr ähnlich ist das Vorgehen von Abel et al. (2017), die eine **Sentiment-Analyse von Arbeitgeberbewertungen** vorgenommen haben. Sie verwendeten dafür Daten von kununu und trainierten unterschiedliche Klassifikationsalgorithmen, um zu erforschen, welcher die besten Ergebnisse liefert. Auf Basis ihrer Daten kommen sie zu dem Schluss, dass Support Vector Machines (SVM) am besten abschneiden, weil sie akurater waren als Naïve Bayes und Entscheidungsbaum-Algorithmen. Ein weiterer Vergleichsmaßstab war die lexikonbasierte Sentiment-Analyse namens SentiStrength. Sie lieferte nur marginal schlechtere Ergebnisse als das SVM-Modell (Abel et al. 2017, S. 44 ff.). Drei Punkte gilt es allerdings zu dem Vorgehen anzumerken. Erstens sind **SVMs prinzipiell nur für die Kategorisierung in zwei Klassen** gedacht, weshalb neutrale Texte entweder als positiv oder negativ klassifiziert werden. Um dennoch mehrere Klassen zu erhalten, können mehrere SVM-Modelle kombiniert werden. Alternativ bzw. zusätzlich hätten auch KNN bzw. hybride Verfahren getestet werden können. Zweitens wurde für das Training die Gesamtbewertung – 1 bis 5 Sterne – als Indikator für ein positives bzw. negatives Sentiment herangezogen. Das spart manuellen Aufwand beim An-

notieren. Gleichzeitig wurden für den Trainingsdatensatz alle Bewertungen zwischen 2,5 und 4 Sternen herausgenommen, um positive eindeutig von negativen unterscheiden zu können (Abel et al. 2017, S. 43). Damit geht natürlich auch Information verloren. Drittens wurde keine Topic-Analyse vorgenommen, das heißt, es bleibt im Unklaren, welche Themen (topics) zu den jeweiligen Sentiments geführt haben oder welche Themen wichtig sind.

Herauszufinden, welche Themen potenziellen Mitarbeitenden wichtig sind, ist zentral für die Arbeit im Personalmarketing. Hinweise darauf können Text-Mining-Verfahren auf Basis von Daten über Arbeitgeber liefern, die ähnlich ablaufen wie Sentiment-Analysen. Beispielsweise wurden in einer Studie knapp 15.000 **Reviews auf Glassdoor**, einem Portal zur Bewertung von Arbeitgebern, mit Hilfe von IBM Watson ausgewertet (Dabirian et al. 2019). Ziel war es, herauszufinden, was IT-Experten an ihren aktuellen oder ehemaligen Arbeitgebern (u. a. Facebook, Google, SAP, XEROX) schätzen bzw. bemängeln. IBM Watson basiert auf dem Training eines tiefen künstlichen Netzes (siehe Kap. 3) und klassifiziert z. B. eine Bewertung „free food, free trainings“ als „free food“ und „free trainings“, während die Bewertung „money money money money money“ nur als „money“ erkannt wird (Dabirian et al. 2019, S. 84). Diesen Output mussten die Forscher anhand eines Wörterbuchs mit Begriffen, die beim Employer Branding als Wertversprechen (Employer Value Proposition) gebraucht werden, abgleichen. Beispielsweise wurde „money“ dem ökonomischen Wertversprechen zugeordnet und „free training“ dem Entwicklungsversprechen. So erhält man zunächst eine Liste an Dingen, die IT-Experten am meisten bemängeln oder eben schätzen (insgesamt acht: von ökonomischen Vorteilen über Work-Life-Balance bis hin zur Möglichkeit, mit neuester Hard- und Software arbeiten zu können). Außerdem kann ausgewertet werden, welche Arbeitgeber diese relevanten Werte wie gut erfüllen. Im Ergebnis sieht das Personalmarketing, in welchen Bereichen noch Nachholbedarf besteht oder wo man als Arbeitgeber sogar zu viel des Guten anbietet. Das ist gerade in Zeiten, in denen das Angebot an IT-Experten deutlich geringer ist als die Nachfrage, hilfreich für die **Optimierung der Personalbeschaffung**.

4.2.1.3 Was man noch machen kann

Sentiment-Analysen können für alle Personalthemen eingesetzt werden, für die Meinungsäußerungen relevant sind und bei denen große Textdatenmengen vorliegen (ab ca. 1200 Beispielen; Abel et al. 2017, S. 42). Neben dem Personalmarketing-Bereich ist dies am ehesten bei generellen Mitarbeiterumfragen der Fall, aber auch **themenspezifische Stimmungsbarometer** können genügend Daten bieten, angefangen bei der Personalbetreuung und der Qualität von Trainings bzw. der Administration von Weiterbildungsangeboten bis hin zu Pulse Checks bei Veränderungsprojekten, in denen (das Auf und Ab von) **Emotionen** abgefragt werden (siehe auch Stone und Chatagny 2019). Beispielsweise hat der Siemens Global Learning Campus mit einer Sentiment-Analyse kritische und positive Kommentare aus dem **Feedback zu Trainings** herausgefiltert

und diese an die entsprechenden Verantwortlichen weitergeleitet, um darauf aufbauend die Inhalte und Abläufe zu verbessern (Liebert und Talg 2018, S. 204). Neben Umfragedaten und Bewertungen können **E-Mails und Posts im Intranet oder in Social-Media-Kanälen ausgewertet werden** (z. B. Gloor et al. 2017; Gelbard et al. 2018; Wen et al. 2019). Klar ist, dass hier der Datenschutz eine große Rolle spielt und Anonymisierung Pflicht ist. Da es bei den meisten Sentiment-Analysen nur darauf ankommt, die Stimmung einer Gruppe als Ganzes und nicht einzelner Mitarbeiter zu erfahren, kann eine Anonymisierung der Daten zugesichert werden (so, wie es bei klassischen Mitarbeiterumfragen auch der Fall ist). Dass die Algorithmen grundsätzlich funktionieren, wurde am Beispiel von Enron gezeigt (siehe Beispiel-Kasten).

Beispiel: Sentiment-Analyse auf Basis von E-Mails bei Enron

Über Enron, jenes Unternehmen, das Anfang der 2000er aufgrund von Bilanzfälschungen einen der größten Unternehmensskandale der Wirtschaftsgeschichte verursachte, gibt es viele interne Datensätze, die nach dem Bankrott öffentlich wurden. Für die Stimmungsanalyse wurden ca. 250.000 E-Mails von 151 Mitarbeitern aus dem Zeitraum von Mitte 2000 bis Anfang 2002 ausgewertet (Gelbard et al. 2018). Ziel war es, aus den E-Mails herauszufiltern, wie zufrieden und motiviert die Mitarbeitenden waren. Die automatische Kategorisierung auf Basis eines vorhandenen Sentiment-Lexikons lieferte zunächst wenige korrekte Einsortierungen in positive, negative und neutrale Stimmungen, weil der Datensatz zu unausgeglichen war und zu viele E-Mails als neutral eingestuft wurden. Durch eine händische Vergabe der Labels konnte der Algorithmus (Naïve Bayes Klassifizierer) verbessert werden (Gelbard et al. 2018, S. 9). Die Auswertung erfolgte entlang kritischer Ereignisse (z. B. Ernennung eines neuen CEOs, Beginn der Untersuchungen durch die Finanzaufsicht). So konnte gezeigt werden, wie sich die Stimmung verändert, dass z. B. Mitarbeiter auf die Bekanntgabe des neuen CEOs, Jeffrey Skilling, positiv reagierten – anscheinend also Hoffnung schöpften –, während die erneute Ernennung von Kenneth Lay zum CEO ein halbes Jahr später keine positive Stimmung (Hoffnung) mehr verbreiten konnte (Gelbard et al. 2018, S. 10 f.).

Es gibt einige Anbieter, die Sentiment-Analysen als Dienstleistung anbieten oder schon vorgefertigte Analysen präsentieren. Beispielsweise werden auf der Website des Projekts „Culture500“ der MIT Sloan Management Review **Auswertungen zur Kultur** großer Firmen präsentiert. Als Datenbasis dienen die Beurteilungen auf dem Arbeitgeberbewertungsportal Glassdoor. Neben Sentiment-Analysen gibt es auch andere Text-Mining-Verfahren, die sich zur Analyse von Organisationskulturen eignen. Insbesondere bietet sich der Einsatz von **Natural Language Processing** an, weil damit nicht nur einzelne Wörter, sondern längere

Textpassagen automatisiert ausgewertet werden können (entweder durch Verwendung von Wörterbüchern und/oder Machine-Learning-Algorithmen). Solche automatisierten Textauswertungen zeigen sowohl eine hohe inhaltliche und externe Validität als auch eine gute Vorhersagegüte in Bezug auf Konstrukte, die klassischerweise mit Organisationskultur in Verbindung gebracht werden, wie z. B. Profitabilität oder Marktwert (Pandey und Pandey 2019). Auch wenn Text-Mining-Verfahren den sozialen Kontext (z. B. Was bedeutet ‚Kundenorientierung‘ in einer bestimmten Situation und von wem geht diese Bedeutungszuschreibung aus?) nicht berücksichtigen können, so bieten sie dennoch eine Alternative zu klassischen Fragebögen, weil der Aufwand für wiederholte Analysen geringer ist und große Textmengen analysiert werden können (z. B. Geschäftsberichte, Aktionsbriefe, Rundschreiben des Vorstands, Posts auf Social bzw. Business-Plattformen).

Sentiment-Analysen wurden auch erfolgreich zur **Auswertung von Leistungsbeurteilungen** eingesetzt (Speer 2018). Da bei Performance Appraisals regelmäßig Kommentare zu den bewerteten Dimensionen abgegeben werden, liegen hier viele Texte vor, die zudem oft ein klares Sentiment beinhalten (z. B. sehr gut, verbesslungsfähig, tut sich schwer). Schon mit einer einfachen Sentiment-Analyse auf Basis einer logistischen Regression, die nur zwischen positiven und negativen Kommentaren bei Leistungsbeurteilungen unterscheidet, lässt sich die numerische Gesamtbeurteilung gut vorhersagen (Speer 2018, S. 312 ff.). Zudem erklärt die Sentiment-Analyse sowohl die Varianz von zukünftigen Leistungsbewertungen und unfreiwilliger Fluktuation als auch die von Beförderungen und Gehaltserhöhungen (Speer 2018, S. 315 ff.). Wenn schon relativ einfache Verfahren zu guten Ergebnissen führen, kann überlegt werden, auf numerische Ratings zu verzichten und Leistung nur auf Basis der Auswertung von schriftlichen Kommentaren zu bewerten.

Während Sentiment-Analysen von Texten schon seit den 1970er-Jahren vorgenommen werden, ist die **Stimmungsanalyse von Bildern, Ton- und Videodaten** ein vergleichsweise junges Feld, das aber angesichts des Trends zu Fotos und Videos im Internet immer bedeutsamer wird (Poria et al. 2018). Für diese multimodale Sentiment-Analyse kommen unterschiedliche KNN-Verfahren zum Einsatz. Beispielsweise konnten Poria et al. (2017) mit Hilfe eines KNN unterschiedliche Emotionen in Videos identifizieren. Während aus wissenschaftlicher Sicht einiges für diese Verfahren spricht, ist ihre Verbreitung in der Praxis noch sehr gering.

4.2.1.4 Was (noch) nicht geht

Die folgenden Schwachpunkte gelten nicht für alle Anbieter von bzw. Modelle für Sentiment-Analysen, da insbesondere Deep-Learning-Verfahren – und hier: Generative Adversarial Networks (GANs) – ein paar Probleme minimieren (siehe

Prabha und Srikanth 2019; Han et al. 2019). Dennoch sich die folgenden Herausforderungen noch nicht vollständig gelöst.

Die **Komplexität und Kontextsensitivität sprachlicher Äußerungen sind nur ansatzweise abgedeckt**. Viele Sentiment-Analysen kategorisieren nur in positive und negative Einstellungen, was natürlich viel zu grob ist, um **semantische Nuancen oder den eigentlichen Sinn** hinter dem Gesagten zu erfassen. Selbst weiterentwickelte Verfahren sortieren oft nur in wenige Kategorien ein (z. B. frustriert, wütend oder glücklich). Schwierig zu analysieren sind Negationen („nicht ... brillant“) und sogenannte Valenzverschiebungen („valence shifters“, Liu 2015, S. 59), bei denen das Gemeinte das Gegenteil des Gesagten/Geschriebenen ist. Ironie, Sarkasmus und rhetorische Fragen führen zu solchen Verschiebungen in der Wertung dessen, was gesagt wird (Liu 2015, S. 80 ff.). So könnte beispielsweise das Feedback eines Mitarbeiters über das HR-Call Center wie folgt lauten: „Wow, schon nach 30 Minuten in der Warteschleife wurde meine Anfrage von HR beantwortet“. Hier muss der Algorithmus die Ironie erkennen und darf den Satz nicht als positiv werten, obwohl das Wort „Wow“ vorkommt (solche Algorithmen gibt es, siehe Liu 2015, S. 83 f., aber sie müssen natürlich auch in die Sentiment-Analyse-Systeme integriert werden). Selbst ganz normale Fragen und Konditionalsätze können Probleme bei der automatischen Analyse bereiten, wenn sie auf Alltagswissen aufbauen. Z. B. ist Menschen sofort klar, dass bei dem Satz „Der Pokal passt nicht in den Koffer, weil er zu groß ist“ mit „er“ der Pokal gemeint ist und nicht der Koffer, obwohl dies grammatisch möglich wäre. Wir wissen das, weil wir bereits als Kleinkinder ein Verständnis über einige Zusammenhänge der physischen Welt erworben haben, z. B., dass kleine Dinge in größere passen können, aber nicht umgekehrt (siehe Gärtner 2013). In all diesen Fällen können nicht einfach Wörter aus einer Liste genommen und anhand ihrer Konnotationen klassifiziert werden.

Die zweite große Herausforderung ist die **Kontextsensitivität sprachlicher Äußerungen**. Einerseits gibt es Wörter, die trotz gleicher Form Unterschiedliches bedeuten (Homonyme und Polyseme), und der gemeinte Sinn sich erst aus dem Zusammenhang mit anderen Begriffen bzw. dem ganzen Satz ergibt. Ein Beispiel aus dem Personalkontext ist das Wort „Leiter“, bei dem das grammatischen Geschlecht entscheidend ist, schließlich ist „die Leiter“ nicht „der Leiter“ (einer Abteilung). Für viele Fragestellungen ist es zudem wichtig zu wissen, auf wen oder was sich ein Sentiment bezieht, um beispielsweise nicht nur zu erfahren, ob ein Arbeitgeber als negativ wahrgenommen wird, sondern woran das liegt (Führungskräfte, Gehalt, Arbeitszeiten, etc.). Sentiment-Modelle müssen deshalb mehr als nur einzelne Worte berücksichtigen (und gegebenenfalls weitere situative Merkmale wie den Zeitpunkt der Äußerung, Emotion und Emotionsträger; Liu 2015, S. 36 ff.). Dabei können selbst scheinbar eindeutige Wörter wie „groß“ je nach Situation Positives oder Negatives meinen: Im Kontext eines Fernseherkaufs mag „groß“ für viele Konsumenten gut sein, im Zusammenhang mit dem Weiterbil-

dungsbedarf der Belegschaft wäre „groß“ eher negativ einzustufen. Sogenannte Recurrent Neural Networks und Long Short Term Networks (siehe Abschn. 3.5) sind hier probate ML-Verfahren, weil sie Tokens der Reihe nach aufnehmen und die Abfolge speichern können. So können zumindest Teile des Kontexts berücksichtigt werden, soweit sie über Beziehungen zwischen Textfragmenten repräsentiert sind. Neben dieser allgemeinen Kontextsensitivität funktionieren die Systeme meist nur in einem spezifischen Berufsfeld, auf das sie trainiert wurden: beispielsweise verwenden IT-Expertinnen andere Begriffe als Facharbeiterinnen oder Auszubildende, weil sie unterschiedliche Werte betonen und andere Wörter benutzen. Dieses Domänenwissen muss für die Textanalyse hinterlegt sein.

Beim Trainieren der Klassifikationsalgorithmen stellt sich oft heraus, dass es **Probleme mit den Datensätzen** gibt. Diese können z. B. unvollständig oder nicht repräsentativ und strukturell verzerrt sein. Z. B. können bei Daten aus Arbeitgeberbewertungsportalen Angaben zu Gesundheitsleistungen oder Gehaltszulagen fehlen oder sie variieren extrem zwischen Organisationseinheiten (z. B. Vertrieb vs. Controlling), sodass sie kaum das gesamte Bild widerspiegeln. Meist sind die Datensätze nicht ausbalanciert, d. h., dass die Anzahl an positiv und negativ gekennzeichneten Textdaten nicht gleich ist. Wenn zu viele Datensätze als neutral kategorisiert werden, lässt sich die Frage, ob die Stimmung positiv oder negativ ist, kaum beantworten. Es gibt zwar Methoden, um mit nicht-ausbalancierten Datensätzen arbeiten zu können, aber eines sollte nach all diesen Schwachpunkten klar sein: Auf eine rein automatische Analyse von unstrukturierten Text- oder Sprachdaten können sich HR-Professionals nicht verlassen.

Die situativen Auslöser des Sentiments sind nur näherungsweise bekannt. Was eine Mitarbeiterin zu einer positiven oder negativen Äußerung veranlasst – intrapersonale und externe Ereignisse – kann in manchen Fällen abgeschätzt werden, z. B. wenn Texte einen Zeitstempel haben und dieser einem Ereignis zugeordnet werden kann (bspw. E-Mails, die direkt vor und nach einem skandalösen Zeitungsbericht über den Arbeitgeber verschickt werden). Da emotionale Reaktionen aber aus einer Wechselwirkung zwischen externem Ereignis und internen Zuständen, persönlichen Fähigkeiten, sozialer Einbettung und erlernten Reaktionen auf bestimmte Ereignisse entstehen und sich dynamisch entwickeln (siehe Küpers und Weibler 2005), bleibt die Abschätzung ungenau.

Angesichts der deutlichen Limitationen drängt sich die Frage auf, ob sich der **Aufwand** für solche Sentiment-Analysen überhaupt lohnt. Das hängt einerseits davon ab, wie groß das zu lösende Problem ist: wer händeringend IT-Experten sucht, wird froh um Erkenntnisse über deren Motivlage sein, und wer in der Zufriedenheit der Belegschaft einen Hebel für deren Engagement und Leistung sieht,

wird wohl eher in Stimmungsanalysen investieren. Wenn dann noch der monetäre Aspekt berücksichtigt wird, lässt sich zum Use Case ein Business Case berechnen – der je nach Annahmen schnell positiv wird (siehe Mühlbauer et al. 2018). Andererseits sind die verfügbaren Verfahren von sehr unterschiedlicher Güte, was auch bedeutet, dass es eine Reihe gut erforschter Algorithmen und Modelle gibt (siehe Liu 2015).

4.2.2 Robo-Recruiting: Von (Active) Sourcing über Matching bis Selecting

4.2.2.1 Worum es geht

Die Kernaufgabe im Recruiting lautet seit Jahrzehnten gleich: Finde für eine vakante Stelle jene Arbeitskraft, die am besten geeignet ist und unterbreite ihr ein Angebot. Es geht nicht um die beste, sondern die passendste Person – und deshalb müssen Recruiter nach dem höchsten Job-Person-Fit suchen, d. h. der größtmöglichen Übereinstimmung zwischen Anforderungen des Jobs und Eigenschaften bzw. Fähigkeiten der Kandidaten. Seit einigen Jahren gibt es für diesen Prozess automatisierte **Such-, Empfehlungs-, Dialog- und Auswahlsysteme** (Lochner und Preuß 2018; Laumer et al. 2019). Diese Systeme sind oft Teil der integrierten Softwaresuiten großer Anbieter wie SAP, IBM, Oracle/Taleo oder Workday, die zudem noch die Administration rund um den Bewerbungsprozess übernehmen: Jobprofile erstellen, Karriereseiten anlegen, Stellenausschreibungen schalten (mittlerweile auch Video-Stellenanzeigen) sowie eingehende Bewerbungen filtern, Termine vereinbaren und den Kontakt halten. Vieles davon können auch die Tools kleinerer Firmen wie Prescreen, TribePad oder rexx.

Zwar hat sich die grundsätzliche Aufgabe des Suchens und „Matchens“ nicht geändert, aber die Datenlage ist heute eine andere als noch vor zehn Jahren. Es kommt zu einer **Inflation der Jobbezeichnungen und Kompetenzen**: Neben eher nüchternen Berufen wie Data Scientists und Web oder Java Developern werden gerade in der Digitalszene Innovation Evangelists und Digital Prophets genauso gesucht wie Software Ninjaneers, Retail Jedis oder Happiness Manager und Chief Heart Officers. Angesichts dieser Vielfalt wird die Suche nach einer Übereinstimmung schwierig, wenn die Systeme nur die Eingabe vordefinierter Suchkriterien bzw. Schlüsselwörter und Boolescher Operatoren erlauben. Wenn die Jobwelt komplex, mehrdeutig und dynamisch ist, veralten vorab definierte Kriterien schnell bzw. führen zu unzureichenden Ergebnissen und es Bedarf der manuellen Pflege der Job- und Kompetenzbezeichnungen – oder man setzt **selbstlernende Systeme** ein.

4.2.2.2 Wie es geht

Suchsysteme

folgen Links auf Internetseiten (**Crawling**), sammeln die darin enthaltenen Informationen (**Scraping**) und analysieren sie (**Parsing**). So können z. B. Name, Berufserfahrung und Jobwünsche auf der Profilseite einer Person in Business-Netzwerken wie XING, LinkedIn oder Github ausgelesen werden. Crawling und Scraping funktionieren sowohl mit Zusatzprogrammen für Browser wie Firefox oder Chrome als auch über eine Application Programming Interface (API), wie im Falle der Business-Netzwerke, die über eine Schnittstelle den Zugriff auf freigegebene Nutzerdaten ermöglichen. Diese Netzwerke offerieren zudem eigene (webbasierte) Anwendungen und besondere Mitgliedschaften, die Recruiter nutzen können, um nach Kandidaten zu suchen (siehe Gärtner 2018a). Daneben gibt es zahlreiche weitere Anbieter, wie z. B. TalentBin (von Monster übernommen) oder Talentwunder, die Profile von Nutzern auf Webseiten, v. a. sozialen Netzwerken, durchsuchen (siehe Beispiel-Kasten). Die gesammelten **Daten werden mit Hilfe von Text-Mining-Algorithmen** ausgewertet. Eine relativ einfache Textklassifikation nimmt z. B. die Latent Dirichlet Allocation (LDA) vor: Unter der Annahme, dass in Texten mehrere Themen (Topics) auftauchen, werden Wörter hinsichtlich ihrer Wahrscheinlichkeit, zu einem Thema zu gehören, klassifiziert (Blei et al. 2003). Z. B. wären in einer Stellenbeschreibung (für Softwareentwickler) Wörter wie Java, Python oder R wohl dem Thema „Programmiersprachen“ aber auch „Anforderungen“ oder „Qualifikationen“ zuzuordnen. Eine Topic-Analyse mit LDA liefert allerdings nur akkurate Ergebnisse, wenn die Reihenfolge der Wörter keine Rolle spielt. Um semantische Verbindungen zwischen Wörtern herzustellen, werden Wort-Netzwerke verwendet, bei denen die Begriffe als Knoten dargestellt und ihre Beziehungen zueinander definiert werden (siehe z. B. die schon vor Jahrzehnten angelegten Datenbanken WordNet und EuroWordNet).

Beispiel: Active Sourcing mit Talentwunder (<https://talentwunder.com/de/>)

Talentwunder bietet Recruitern die Möglichkeit mit Hilfe einer Suchmaschine passende Kandidaten aus über 1,7 Mrd. Profilen zu suchen (Dittes 2018). Die Daten werden aus den zugänglichen Angaben, die Kandidaten in sozialen Netzwerken gemacht haben, gecrawlt und gescrappt. Insgesamt durchsucht die Talentwunder-Software über 75 Online-Netzwerke, u. a. XING, LinkedIn, GitHub, Behance, GitHub, ResearchGate oder Meetup.

Recruiter geben die Anforderungen an einen Kandidaten für eine von ihnen zu besetzende Stelle in eine Suchmaske ein (basierend auf Booleschen Opera-

toren) und erhalten eine Liste mit potenziell passenden Kandidaten. Einzelne Kandidaten aus dieser Trefferliste können markiert und in einer Favoritenliste verwaltet werden. Die Auswahl der Kandidaten beruht darauf, dass deren Nutzerprofile nach der Einschätzung der Talentwunder-Software den Suchkriterien des Kunden entsprechen. Zusätzlich wird die Bereitschaft zum Arbeitsplatzwechsel und zum Umzug berechnet, basierend auf der Auswertung von Daten wie vergangenen Arbeitsplatzwechseln, Profilupdates, der Entfernung zwischen Wohn- und Arbeitsort sowie einer Text-Mining-Analyse des Feldes „offen für ...“ (welches auf Plattformen wie XING oder LinkedIn ausgefüllt werden kann). Beispielsweise wird ausgewertet, wie lange ein Kandidat im Vergleich zu anderen Personen aus der gleichen Branche im aktuellen Job ist, wann das Profil zuletzt aktualisiert wurde und welche Informationen hinzugefügt wurden (Dittes 2018, S. 219). Wer z. B. ein neues Profilfoto hochgeladen hat, war wahrscheinlich vorher beim Fotografen und dies wiederum, weil eine hohe Wechselwilligkeit vorhanden ist. Für Recruiter sind die Informationen zu Wechsel- und Umzugsbereitschaft interessant, denn je niedriger (höher) die Bereitschaft ist, desto teurer (billiger) wird es, einen Kandidaten abzuwerben. Deshalb steckt in der Trefferliste letztlich auch eine Empfehlung.

Allerdings weist Talentwunder in den AGBs darauf hin, dass weder die Richtigkeit, Vollständigkeit oder Aktualität der durchsuchten oder angezeigten Profile gewährleistet werden können, noch die tatsächliche Eignung oder das tatsächliche Interesse (einschließlich Wechsel- oder Umzugswilligkeit) der potenziellen Kandidaten.

Empfehlungssysteme (Recommender)

schlagen den Mitarbeitern im Recruiting passende Kandidaten für eine Stelle vor oder, wenn die Suchrichtung umgekehrt ist, den Kandidaten einen passenden Job. Welche Algorithmen genutzt werden, variiert zwischen den Software-Anbietern, wie z. B. Textkernel (u. a. in SAP SuccessFactors integriert, siehe Beispiel-Kasten), Candidatezip (u. a. in Salesforce integriert), Cobrainer, Talents Connect, Wollmilchsau, Joinvision oder rexx. Dies nicht zuletzt, weil die Matching & Empfehlungsaufgabe von mehreren ML-Verfahren gelöst werden kann: von Bayes-Netzen über Support Vector Machines bis hin zu KNN/DL gibt es zahlreiche Möglichkeiten (Portugal et al. 2018). Entscheidend ist, dass die eingesetzten Algorithmen nicht einfach nur Wörter miteinander vergleichen, sondern **Satzstrukturen und Bedeutungen beachten**, also Syntax und Semantik. Nur so erkennt auch eine Maschine, was uns Menschen völlig klar ist: dass es etwas anderes ist, „Geschäftsführer zu sein“ oder „an den Geschäftsführer zu berichten“, obwohl beide Male das Schlagwort „Geschäftsführer“ auftaucht. Smartes Matching vergleicht also nicht ein-

fach nur einzelne Schlagworte oder angeklickte Kategorien, sondern interpretiert Text. Dafür braucht es Techniken zur Wissensrepräsentation, wie z. B. semantische Wortnetze (Ontologien), die nicht nur Synonyme, sondern auch Fachbegriffe und deren Beziehungen zueinander sowie Wörter aus anderen Sprachen richtig zuzuordnen (z. B. in der Form <Bereichsleiter> <berichtet an> <Geschäftsführer> <ist Teil von> <Management Board>).

Beispiel: Textbasiertes Recruiting mit Textkernel (<https://www.textkernel.com/de>)

Textkernel bietet eine semantische Suche, mit deren Hilfe Recruiter in firmeneigenen Personaldatenbanken und externen sozialen Netzwerken nach Kandidaten für einen Job suchen können. Als Ergebnis wird ein Ranking von passenden Personen ausgespielt. Dazu müssen Texte – die Stellenanzeige und der Lebenslauf – ausgelesen und verglichen werden. Der Fokus liegt also auf der Verarbeitung und dem Matching unstrukturierter Texte (CVs und Stellenbeschreibungen). Zudem ist HR-spezifisches Wissen über typische Karrierepfade gefragt, sodass einer Person, die bereits Junior Marketing Manager ist, nicht noch eine weitere Junior-Position vorgeschlagen wird.

Um Begriffe in ihrer Bedeutung und ihrem Zusammenhang zu verstehen, arbeitet Textkernel mit einer Kombination aus String-Matching-Algorithmen, DL/LSTM und eigenen, anpassbaren Worttaxonomien sowie Wortnetzen, die Jobtitel, Synonyme, Ausbildungsniveaus und Kompetenzen in verschiedenen Sprachen als Wissensdatenbank bereithalten (Neculoiu et al. 2016; <https://www.textkernel.com/category/technology/>). Damit lässt sich aus Lebensläufen – deren Daten unstrukturiert sind, weil CVs immer etwas anders aufgebaut und geschrieben sind – herauslesen, dass eine Kandidatin, die „Software Architekt Java/J2EE“ als aktuelle Jobbezeichnung angibt, auch den Job als „Java-Entwicklerin“ erfüllen kann. Weitere Qualifikationen, Funktionserfahrung, Geburtsdatum, Fahrstrecke zur Arbeitsstelle, etc. können ebenfalls relativ exakt extrahiert werden, wenn auch nicht perfekt. Dann prüft die Software, inwieweit Kandidaten- und Stellenprofil übereinstimmen und gibt eine Empfehlung ab (Matching-Wert). Beim Ranking der Kandidaten werden unterschiedliche Gewichtungen der Anforderungen beachtet. Beispielsweise ist bei der Suche nach Call Center Agents der Wohnort eines Kandidaten wichtig, während er bei SAP Consultants nachrangig ist, aber Kompetenzen wichtiger sind.

Textkernel verwendet mehrere Technologien und Algorithmen, um (papierbasierte) CVs und Stellenanzeigen auswertbar zu machen: angefangen mit Optical Character Recognition (OCR) und Parsing, um eingescannte Texte in computerverwertbare Zeichenfolgen umzuwandeln, über Fuzzy-Text-Matching, um

auch schlecht gescannte Texte richtig zu entziffern (es werden auch Zeichenketten erkannt, die nicht gleich, aber ähnlich sind), über Wort- und Wissensdatenbanken bis hin zum Matching (Learning to Rank- und Klassifikationsalgorithmen). Um sicherzugehen, dass möglichst viele Jobs und Kompetenzen erkannt und richtig zugeordnet werden, bindet Textkernel externe Taxonomien und Wissens- bzw. Wortdatenbanken ein (z. B. ISCO, O*NET, DBpedia), ermöglicht aber auch eine firmenspezifische Anpassung. Das gesamte Such- und Matching-System beruht also sowohl auf Machine Learning als auch Wissensrepräsentationen mit manueller Anpassung.

Wie akkurat die Empfehlungen sind, lässt sich selten wissenschaftlich prüfen, weil die Firmen ihre Algorithmen nicht verraten und nicht an vergleichenden Tests teilnehmen. Für das LinkedIn-Netzwerk liegen Studien vor, die zeigen konnten, dass Generalized Linear Mixed Models (GLMix) – letztlich eine verfeinerte logistische Regression – zu einer **höheren Passung zwischen eingeblender Stellenanzeige und Kandidatenpräferenzen** führt, weil die Nutzer 20–40 % öfter auf den „Apply Now“-Button klickten (Zhang et al. 2016).

In einer anderen Studie kam ein KNN zur Spracherkennung zum Einsatz, um Lebensläufe automatisch auszuwerten und basierend darauf ein Scoring bzw. eine Rangliste der Bewerber zu erstellen (Campion et al. 2016). Einerseits wurden die Lebensläufe anhand von sechs Kompetenzen durch Menschen bewertet, andererseits der Algorithmus auf die Sortierung entlang dieser Kompetenzen trainiert. Dabei waren die Ratings der Maschine mit denen der menschlichen Rater vergleichbar gut, bezogen auf zwei wichtige Kriterien in der Personalauswahl, nämlich Reliabilität sowie (konvergente und diskriminante) Validität (Campion et al. 2016, S. 966, 970).

Ein weiteres Ergebnis aus dem Rekrutierungsumfeld kommt von Chalfin et al. (2016), die untersuchen, woran man frühzeitig erkennen kann, dass Polizisten gewalttätig werden. Sie konnten zeigen, dass ihr Klassifikationsalgorithmus (Gradient Boosting) zu besseren Vorhersagen kommt als die aktuell bei der Polizei von Philadelphia verwendete Methode.

Während der Einsatz von solchen Recommender-Systemen kaum Widerstand hervorruft und bei ca. 10 % der Top-1000-Unternehmen bereits erfolgt (Laumer et al. 2019, S. 12), ist der umgekehrte Fall strittiger: Wenn ein System auf Basis des automatisierten Abgleichs zwischen Stellenanforderungen und Qualifikationsprofil manche Kandidaten empfiehlt und andere nicht, dann ist dies mindestens eine Vorauswahl (wenn die Bewerber dem Recruiter zur weiteren Bearbeitung vorgeschlagen werden) oder gar eine finale Auswahl (wenn direkt und automatisch Absagen oder Einladungen verschickt werden). Auch wenn dies hart und sehr wenig

herzlich klingt: Laut einer aktuellen Studie gehen sieben von zehn der Top-1000-Unternehmen davon aus, dass Systeme zur automatisierten Vorauswahl von Bewerbern in Zukunft immer häufiger zum Einsatz kommen werden (Laumer et al. 2019, S. 13).

Digitale Dialogsysteme (Chatbots)

beantworten automatisiert bestimmte Fragen von Kandidaten zu offenen Stellen, Karrierethemen oder allgemein zum Unternehmen. Chatbots können sowohl externe Eingaben (z. B. Kundenanfragen zu einem Produktpreis) als auch interne (z. B. Mitarbeiterfragen zum Resturlaub) in natürlicher Sprache bearbeiten (siehe Abschn. 3.8).

Zwei Unterscheidungen gilt es dabei zu beachten. Einerseits muss darauf geachtet werden, ob von **text- oder sprachbasierten Bots** die Rede ist. Erstere funktionieren in der Praxis relativ zuverlässig. Beispielsweise tritt der Chatbot „Mo“ der Jobsuche-Plattform MoBerries anhand von vorgegebenen Entscheidungsbäumen in einen Dialog mit dem Nutzer und sobald Mo genügend Informationen bekommen hat, stellt Mo eine Auflistung relevanter Ergebnisse zusammen. Sprachbasierte Bots kommen schnell an ihre Grenzen, wenn starker Dialekt gesprochen wird, die Umgebungsgeräusche zu laut sind, das Anliegen auf mehrere Nachrichten verteilt wird oder komplizierte Formulierungen verwendet werden (vgl. Hill et al. 2015; Stucki et al. 2018).

Andererseits gibt es **(regelbasierte) Auskunftssysteme und (KI-basierte) Dialogsysteme**. Erstere arbeiten vorstrukturierte Frage-und-Antwort-Schemata ab, können also nur so agieren wie es ihre Regeln und ihre Informationsdatenbank, auf die sie zurückgreifen, erlauben. Dies limitiert nicht nur die möglichen Fragen, die der Chatbot beantworten kann, sondern auch die Antworten, die zwar in der Formulierung variieren können, aber sonst immer gleich ausfallen, egal in welcher Situation (z. B. emotionalen Stimmung) sich der menschliche Interaktionspartner befindet. KI-basierte Chatbots können auf Basis von Erfahrungen lernen und ihr Frage-Antwort-Verhalten anpassen. Werden beispielsweise KNN eingesetzt, trainiert man den Bot/Algorithmus mit Hilfe eines Trainingsdatensatzes, sodass dieser Muster in dem bekannten Datensatz erkennt und diese anschließend auf unbekannte Datensätze derselben Art (d. h. Text oder Sprache) anwendet. Da die Antwortoptionen auf Wahrscheinlichkeitsrechnungen beruhen und daher Falschantworten möglich sind, bauen Unternehmen Schwellwerte als Sicherheiten ein: Unterschreitet der Wert, mit dem das System wahrscheinlich richtig liegt, die gewünschte Prozentzahl, wird die Anfrage an einen Mitarbeiter weitergegeben.

Aktuell bieten nur 2,8 % der Top-1000 Unternehmen **Chatbots im Recruiting** an, auch wenn viele davon ausgehen, dass Chatbots in Zukunft immer häufiger zum Einsatz

kommen werden – schon allein, weil vor allem jüngere Kandidaten solche Systeme gerne nutzen würden (Laumer et al. 2019, S. 11). Beispiele sind die Deutsche Telekom, die Allianz und Sixt, die alle einen textbasierten Chatbot über den Facebook – Messenger betreiben. In der Beispiel-Box ist der Einsatz der Software von Jobpal bei der Deutschen Telekom beschrieben, ähnlich funktioniert der Chatbot Mo von MoBerries. International bekannt ist der Chatbot Mya, der auf mehreren Plattformen (WhatsApp, Facebook, Skype und LinkedIn) läuft, Kandidatenfragen beantworten und Informationen für die spätere Auswahl sammeln kann. Ebenfalls auf dem Facebook-Messenger läuft der Chatbot, den Bayer im Recruiting einsetzt. Wie bei vielen anderen, geht es Bayer vor allem darum, dass die Informationen aus vielen unterschiedlichen Kanälen ineinanderfließen (Multichannel-Fähigkeit). Der Dialog mit den Zielgruppen muss deshalb über die Bayer-Karriere-App „Bobby2go“ genauso wie über den Bayer-eigenen Alexa-Skill stattfinden – und zwar im Bestfalle so, dass das jeweilige System den Diskussionsstand aus den anderen Kanälen kennt, weil sonst der Nutzer alle Informationen von Neuem eingeben muss. Nebenbei haben auch die humanoiden Roboter von Bayer ein Sprachmodul, sodass sie bei Karrieremessen auf die Fragen der Interessenten zum Thema „Karriere und Einstiegsmöglichkeiten“ antworten können. Die technische Grundlage dafür liefern die Microsoft-Bot-Services aus der Azure Cloud, allen voran der QnA-Maker und LUIS (Microsoft 2019a).

Beispiel: Deutsche Telekom nutzt Recruiting Chatbot von Jobpal
(<https://jobpal.ai/de/>)

Ende 2016 startete die Deutsche Telekom ein Projekt im Rahmen ihres Start-up-Accelerators hub:raum und setzte den textbasierten Chatbot von Jobpal für Standardanfragen (Frequently Asked Questions) ein (siehe Abb. 4.4). Ein Ergebnis war, dass sowohl die Anzahl an Bewerbungen als auch deren Qualität gesteigert werden konnte, was die Projektverantwortliche vor allem darauf zurück führt, dass 62 % der Besuche auf den Stellenseiten eines Unternehmens außerhalb der Arbeitszeit des Recruiting-Teams stattfinden: Wenn potenzielle Kandidaten nach Antworten suchen (die aus ihnen Bewerber machen könnten), ist keiner da, der sie geben könnte (Ott 2017). Dass die Reaktion auf Fragen der Kandidaten verbessert wurde, ist wenig überraschend, denn schließlich sind Chatbots 24 Stunden am Tag online und antworten sofort – sofern sie so trainiert sind, dass sie die Fragen verstehen und passende Antworten abgespeichert sind (wie z. B. für Fragen nach dem Gehalt oder Jobanforderungen). Fragen, die der Chatbot noch nicht beantworten kann, werden an einen menschlichen Recruiter weitergeleitet und die Antworten für die Zukunft gespeichert. Etwa 70 % der Fragen werden vom ADM-System (einer Kombination aus logistischer Regression und NLP) mit einer so hohen Wahrscheinlichkeit einer Antwort zugeordnet, dass dem System eine automatische Antwort erlaubt wird – die restli-

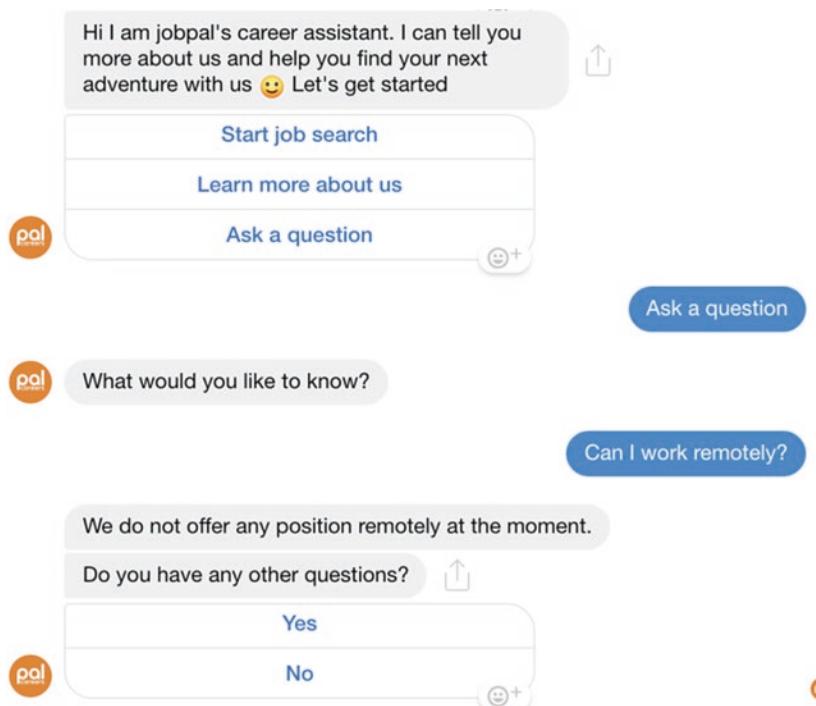


Abb. 4.4 Frage-Antwort-Dialog im Chatbot von Jobpal (Kuksenok und Martyniv 2019, S. 4)

chen 30 % der Anfragen werden an einen Menschen weitergeleitet (Kuksenok und Martyniv 2019). Mit anderen Worten: bei ca. einem Drittel der Fragen wird nicht erkannt, welche Absicht (Intent) der Fragesteller verfolgte. Und selbst wenn Frage und Antwort richtig klassifiziert werden, können nur vorgefertigte Antworten ausgegeben werden. In der Praxis zeigt sich dann immer wieder, wie viele unerwartete Fragen Kandidaten haben, was auch zu Überraschungen auf Seiten der HR-Experten der Telekom führte (Ott 2017). So mögen Fragen zum Gehalt und den Stellenanforderungen vorhersehbar sein, schwieriger wird es bei Detailfragen, z. B. zu welchen Konditionen ein Firmenfahrrad gestellt wird, oder bei sehr breiten und daher mehrdeutigen Fragen, z. B. wie es ist bei der Firma zu arbeiten (es gibt auch Fragen, die den Chatbot selbst betreffen, was er/sie so macht, denkt und fühlt). Dabei geben die HR-Mitarbeiter die Antworten vor und überarbeiten sie (der Chatbot generiert die Antworten also nicht selbst

aus einem Wort- bzw. Wissensnetz). Ein interessanter Nebeneffekt: Da die Antworten letztlich von Mitarbeitern kommen, spiegeln sie zumindest teilweise die Tonalität und die Umgangsformen des Teams wider, was dem Chatbot einen „human touch“ verleiht (Ott 2017).

Auswahlsysteme

sind digitale **Instrumente zur Beurteilung und Vorhersage beruflich relevanter Variablen** zur Abschätzung der Eignung (Diercks 2020, S. 88). Es gibt sie in unterschiedlicher Ausprägung: angefangen bei der automatisierten Auswertung von Lebensläufen und webbasierten Persönlichkeits- oder Kreativitätstest, über Online-Spiele bis hin zur Selektion auf Basis von Sprachproben oder Videointerviews. Weltweit treten immer wieder neue Anbieter mit Tools zur (Vor-)Auswahl von Bewerbern auf den Markt (z. B. Cyquest, HireVue, Precire, Pymetrics, SEEDLINK, VCV, 100 Worte).

Bei **Online-Assessments** handelt es sich um Personalauswahlinstrumente, die über das Internet mit dem Ziel der Selbst- oder Fremdselektion durchgeführt werden. Für die Selbstselektion, also die Beantwortung der Frage, ob sich eine Kandidatin überhaupt bei einem Unternehmen bewerben möchte, kommen **Intelligenztest, Persönlichkeitsinventare oder Berufsorientierungsspiele** zum Einsatz (Diercks 2020, S. 84; Warszta et al. 2020, S. 34 f.; siehe Beispiel-Kasten). Bei der Fremdselektion geht es aus Sicht der Unternehmen darum, die geeignete Kandidatin zu identifizieren, und zwar mit möglichst wenig Streuverlust, niedrigen Kosten und kurzen Bewerbungszeiten. Sowohl bei der Selbst- als auch der Fremdselektion werden seit ca. 2010 vermehrt Spiele oder spielerische Elemente eingesetzt, weshalb auch von **Game-Based-Assessments** oder **Recrutainment** gesprochen wird (Diercks und Kupka 2014). Die Begriffe dürfen aber nicht darüber hinwegtäuschen, dass es sich auch bei den Spielen um Tests handelt. Deshalb sollten die eingesetzten Instrumente auch die in der DIN 33430 zur berufsbezogenen Eignungsdiagnostik formulierten Anforderungen erfüllen, d. h. neben Objektivität, Reliabilität und Validität auch eine theoretische Fundierung und einen Bezug zu den Anforderungen aufweisen sowie angemessen sein (Diercks 2020, S. 94).

Beispiel: Einsatz von CYQUEST bei Airbus zum Online-Assessment (<https://www.cyquest.net/portfolio-item/airbus-online-test/>)

Das Online-Assessment bei Airbus besteht aus elf Leistungs- und Wissenstests für insgesamt 29 verschiedene Ausbildungsberufe und duale Studiengänge. Die unterschiedlichen Berufsgruppen erhalten individuelle Testabläufe, sodass die Eignung berufsspezifisch erfasst wird. Getestet werden unter anderem die kognitive Leistungsfähigkeit, indem z. B. Jahreszahlen fortgeführt, Diagramme analysiert oder ein Mathematiktest absolviert werden müssen. Im Bereich Wis-

sensverfahren werden berufsrelevante (Vor-)Kenntnisse, z. B. ein mechanisch-technisches Grundverständnis oder räumliches Vorstellungsvermögen geprüft.

Neben den Tests nutzt Airbus den Kontakt zu den Kandidaten, um Informationen über das Unternehmen zu vermitteln. Eine der sogenannten „Personal-marketinginseln“ besteht aus einem interaktiven Suchbild, in dem die Kandidaten z. B. erfahren können, dass Auszubildende und dual Studierende im Rahmen eines zweiwöchigen Erlebnislehrgangs auf Juist selbstständig ein Segelflugzeug steuern und landen dürfen. Mit solchen Informationen möchte sich Airbus als attraktiver Arbeitgeber präsentieren und Einfluss auf die Selbstselektion nehmen. Das Online-Assessment ist deshalb sowohl ein Test- als auch ein Personalmarketinginstrument.

Während Spiele bzw. Tests in Deutschland und weltweit weitgehend akzeptiert sind (Diercks 2020, S. 81 f., 90 ff.), gibt es **Sprach- und Videoanalysen** erst seit wenigen Jahren – und sie stehen vor allem in Deutschland in der Kritik (siehe Abschn. 4.2.2.4). Zum Einsatz kommt die Analyse von Sprach- und Videodaten bei Unternehmen, z. B. Hilton Hotels, L'Oréal (siehe Beispiel-Kasten), Talanx, Tesco oder Unilever (siehe Beispiel-Kasten). Entweder werden Videoabschnitte direkt indexiert bzw. Bilddaten analysiert oder die gesprochene Sprache wird in Text umgewandelt, der dann wie beim Text Mining von Lebensläufen ausgewertet wird. Teil der Auswertung können **Persönlichkeitsanalysen** sein, wie sie etwa von TribePad, IBM Watson Personality Insights, Precire oder 100 Worte vorgenommen werden. HireVue und TribePad werten neben Sprach- auch noch Bilddaten aus und vergleichen Tonalität, Wortmuster und Mimik (z. B. Wimpernschläge, Sprünge der Augenbewegungen, das Starren auf einen bestimmten Punkt) der Bewerber mit Mitarbeitern, die als High Performer gelten (Sánchez-Monedero und Dencik 2019, S. 12). Meist werden aber nur Sprachproben in Text umgewandelt und dieser wird auf Muster und Zusammenhänge mit bekannten Persönlichkeitsmerkmalen (z. B. den Big Five) oder anderen berufsrelevanten Merkmalen (z. B. Durchsetzungsvermögen, Belastbarkeit) hin geprüft. IBM braucht beispielsweise für die Analyse mit Watson Personality Insights mindestens 3500 Wörter (Mahmud 2015), bei Precire sind es Sprachaufnahmen von maximal 25 Minuten mit einfachen Fragen zum bisherigen Leben der Bewerber (Schmidt-Atzert et al. 2019).

Beispiel: Einsatz der Sprachanalyse von SEEDLINK bei L'Oréal (<https://www.seedlinktech.com/>)

L'Oréal setzt das NLP-System von SEEDLINK nach positiven Ergebnissen in China, Spanien und UK seit November 2018 auch in Deutschland für das Recruit-

ting von Praktikanten ein (ca. 200 pro Jahr in Deutschland; Queb 2019). Nach dem Hochladen der Bewerbungsunterlagen werden die Kandidaten zu SEED-LINK weitergeleitet und gebeten drei kurze Fragen zu Kompetenzen schriftlich zu beantworten. Eine Frage lautet z. B.: „Beschreib bitte eine Situation, in der du in kürzester Zeit etwas Neues lernen musstest. Wie bist du dabei vorgegangen? Was war das Ergebnis? Wie könntest du das Gelernte bei L'Oréal einbringen?“ (Queb 2019). Die Auswertung zeigt, wie wahrscheinlich die Kandidaten ein sehr gutes Praktikum absolvieren werden und wie gut sie zu L'Oréal passen. SEED-LINK offeriert zwei Softwarebausteine, einmal für die Vorhersage des Werte-Matchings und einmal für den Job-Fit. Auf dieser Basis entscheidet ein Recruiter, ob ein Kandidat zum persönlichen Gespräch eingeladen wird.

Das KNN/NLP-System wurde mit Datensätzen von 74 vergangenen Praktikantenbewerbungen bei L'Oréal trainiert, was ca. 40.000 Datenpunkte ergab (Queb 2019). Es werden nicht nur Worte verglichen, sondern Beziehungsmuster zwischen Wörtern in Relation mit den Kriterien für einen passenden Kandidaten gebracht. Im Rahmen des Trainierens der KI werden entsprechend des Feedbacks aus den Fachbereichen die Gewichtungen des KNN/NLP-Systems angepasst (supervised learning). Laut L'Oréal erzielt das System eine prognostische Validität von über .70, was sehr hoch ist. Zukünftig will L'Oréal die Ergebnisse mit größeren Datenmengen validieren. Zudem könnte auch für Deutschland die Integration mit dem Chatbot Mya erfolgen, der bereits in Großbritannien, den USA und Frankreich eingesetzt wird.

Beispiel: Unilever nutzt verschiedene Technologien zur Automatisierung des Recruiting-Prozesses

Die Firma Unilever wickelt den Recruiting-Prozess für ihr Graduate- bzw. Future-Leaders-Programm von der Bewerbung bis zur Auswahlentscheidung über unterschiedliche digitale Tools ab – und zwar in über 60 Ländern (Feloni 2017). Nachdem Bewerber ein Online-Formular ausgefüllt haben bzw. die Daten aus ihrem LinkedIn-Profil automatisch haben einlesen lassen, absolvieren sie bis zu zwölf verschiedene Online-Spiele, Persönlichkeits- und Kompetenztests. Beispielsweise wird im Ballon-Spiel die Risikoneigung getestet: Die Bewerber sollen per Mausklick einen Ballon aufpumpen, wobei jeder Klick dem Spieler 5 Cent einbringt. Nach jedem Pumpvorgang kann das Spiel fortgesetzt oder beendet und die eingespielte Summe behalten werden. Pumpt ein Spieler zu oft, platzt der Ballon und das Geld ist weg. Die Ergebnisse der Spiele bzw. Tests, die von Pymetrics programmiert wurden, werden mit jenen von bereits eingestellten, erfolgreichen Mitarbeitern verglichen, um ähnliche Profile

zu identifizieren (Feloni 2017; Unilever 2018). Anschließend werden ausgewählte Bewerber um ein Videointerview gebeten, das von der HireVue-Software automatisch analysiert wird. Im Zentrum der Auswertung stehen folgende Fragen (Feloni 2017; Unilever 2018): Wie oft wurde der Browser verlassen? Wurde mit copy-paste gearbeitet? Wie waren Tonalität, Wortmuster und Mimik (bestimmte Begriffe, Intonation, Sprechgeschwindigkeit, etc.)? Ziel ist es, die Nervosität, Stimmung und Verhaltensmuster der Bewerber zu bestimmen, wiederum im Vergleich mit Mitarbeitern, die als High Performer gelten. Wer auch diese Runde übersteht, wird im letzten Schritt in das firmeneigene Discovery Centre eingeladen, um dort im Rahmen von Simulationen jobspezifische Aufgaben, Rollen und Prozesse durchzuspielen (Tallgauer et al. 2020, S. 31).

Neben der inhaltlichen Passung achtet Unilever nach eigenen Aussagen darauf, dass die Recruiting-Entscheidungen diskriminierungsfrei sind. Das bezieht sich nicht nur auf das Geschlecht, sondern es wird zumindest in den USA auch darauf geachtet, dass die Bewerber von unterschiedlichen Universitäten kommen und der sozioökonomische Status ausgeglichen ist (Feloni 2017). Zu diesen Kriterien liegen keine genauen Daten vor.

Laut Unilever können durch die Automatisierung global ca. 100.000 Stunden an Interviewzeit und ca. \$ 1Mio. an Recruiting-Kosten pro Jahr eingespart werden (Booth 2019). Die Höhe des eingesparten Betrags hängt von der Anzahl an Bewerbungen ab (zurzeit ca. 250.000 pro Jahr). Zugleich sei die Zeit zwischen Ansprache und Einstellung von Bewerbern von vier Monaten auf vier Wochen verkürzt worden – bei hoher Zufriedenheit mit dem Recruiting-Prozess, wie eine Umfrage unter 25.000 Bewerbern ergab (Feloni 2017).

In Deutschland und Europa wird **nur die Empfehlung automatisiert, nicht die tatsächliche Auswahlentscheidung**, weil dies nicht konform mit den Datenschutzverordnungen wäre (siehe Kap. 6). Dennoch nehmen die existierenden Systeme Recruitern einige Aufgaben und damit auch Entscheidungen im Bewerbungsprozess ab, woraus sich zwei Fragen ergeben: **Ist die maschinelle Auswahl besser als die menschliche? Sollen Unternehmen Maschinen oder Menschen die Auswahlentscheidung überlassen?** Die erste Frage kann mit empirischen Studien beantwortet werden – vorausgesetzt, es existiert ein geteiltes Verständnis darüber, was „besser“ bedeutet. Neben den klassischen Gütekriterien wie Validität, Objektivität und Reliabilität werden oft weitere Aspekte wie Fairness, Transparenz, Nachvollziehbarkeit, Kontrollierbarkeit, Nutzerakzeptanz und Kosten mit ins Kalkül gezogen, um zu entscheiden, welches Vorgehen „besser“ ist. Solche Aspekte berühren bereits Wertvorstellungen, weshalb zwar eingefleischte Empiriker nach Näherungsgrößen für Werte suchen, allen anderen aber klar ist, dass eine rein em-

pirische Betrachtung zu kurz greift. Die Grenzen werden spätestens bei der Beantwortung der zweiten Frage noch deutlich, denn sie ist eine normative und berührt deshalb ethisch-moralische Aspekte (Was soll/darf nicht sein?). Im Folgenden wird nur auf die erste Frage rekurriert, die ethische Dimension wird später diskutiert (Kap. 6).

Zunächst lässt sich argumentieren, dass die maschinelle Auswahl **reliabel, objektiv und ohne Vorurteile** verläuft (Diercks 2020, S. 97; Dudler 2020, S. 106; Raghavan et al. 2019). Die Idee dahinter: Anders als bei menschlichen Entscheidern läuft die Personalauswahl durch Algorithmen für alle Testteilnehmer immer wieder gleich ab. Es gibt keine unbewussten Verzerrungen, Vorurteile oder Diskriminierung aufgrund von Geschlecht, Hautfarbe, Herkunft, Sprache, Alter etc. Denn einerseits ist ein hoher Grad an Anonymisierung gewährleistet und andererseits werden alle Bewerber anhand der gleichen Regeln oder des gleichen statistischen Modells evaluiert, sodass alle die gleichen Chancen haben (was nicht gleichzusetzen ist mit einer Gleichverteilung über bestimmte Bewerbergruppen *nach* der Auswahl; Raghavan et al. 2019, S. 4). Diese Argumentation wird von vielen Anbietern digitaler Tools von Eignungs- bzw. Leistungs- und Persönlichkeitstests vertreten.

HireVue ist beispielsweise einer der vielen Anbieter, die behaupten, **Verzerrungen bei Personalauswahlentscheidungen zu reduzieren** (Raghavan et al. 2019, S. 7 f.; Sanchez-Monadero et al. 2019, S. 1). Dabei erfolgt die Bias-Erkennung durch Messung des Anteils von geschützten Minderheiten. Gemäß US-Regularien (Uniform Guidelines on Employment Selection Procedures) darf die Selektionsrate (der Anteil ausgewählter Bewerber an allen Bewerbern) von Gruppen mit bestimmter Hautfarbe, Geschlecht, Alter, sexueller Orientierung etc. 4/5 der Gruppe mit der höchsten Selektionsrate nicht unterschreiten. Bei diesem sogenannten „Adverse Impact Ratio“ wird also der Anteil der erfolgreichen Bewerber einer benachteiligten Gruppe ins Verhältnis zum Anteil der erfolgreichen Bewerber einer bevorteilten Gruppe gesetzt. Werden beispielsweise von 80 Bewerbern mit heller Hautfarbe 8 ausgewählt und bei 40 Bewerbern mit dunkler Hautfarbe nur 2, dann ist das Verhältnis $(2/40)/(8/80) = 0,5$ bzw. 50 %. Damit beträgt der Anteil an ausgewählten Personen mit dunkler Hautfarbe weniger als 4/5 oder 80 % und das Auswahlverfahren gilt als diskriminierend („Adverse Impact“). Diese Ungleichverteilung kann rechtliche und wirtschaftliche Konsequenzen für den Arbeitgeber haben (Raghavan et al. 2019, S. 5). Um demografische Parität zu erreichen und Verzerrungen zu minimieren, wendet HireVue zwei Strategien an (Sanchez-Monadero et al. 2019, S. 4 f.): Die erste Strategie besteht in der Beseitigung von Indikatoren, die sich wissentlich nachteilig auf geschützte Gruppen (Rasse, Geschlecht, Alter, etc.) auswirken und gegen die 4/5-Regel verstößen.

Z. B. wurde herausgefunden, dass Leistungsträger im technischen Support oft langsamer sprechen, eine Eigenschaft, die auch mit dem männlichen Geschlecht korreliert und damit Männer bevorzugt als geeignet klassifiziert. HireVue eliminiert diese Merkmale aus dem Modell, trainiert und prüft es so lange, bis keine Merkmale mehr enthalten sind, die zu verzerrten Ergebnissen führen. Die zweite Strategie bezieht sich auf die Berechnung des Fehlers, den ML-Algorithmen machen, wenn sie sich den Daten anpassen. Vereinfacht formuliert: Statt die Summe über alle (quadrierten) Abweichungen insgesamt zu berechnen, werden die Abweichungen für jede geschützte Gruppe separat berechnet. Zusätzlich wird ein Strafterm addiert, wenn gegen die 4/5-Regel verstoßen wird. Der ML-Algorithmus ist dann für alle Gruppen ähnlich fehleranfällig (bzw. akkurat), sodass Fairness zwischen den Gruppen hergestellt wird. Auch wenn diese Doppelstrategie sich nur auf US-amerikanische Vorgaben zur Gleichverteilung zwischen geschützten Gruppen bezieht, zeigt das Vorgehen, dass der Bias einer Maschine statistisch erfasst und gegengesteuert werden kann – mehr als dies bei menschlichen Verzerrungen möglich ist (Laumer et al. 2019, S. 14).

Neben Vorteilen in Bezug auf die Objektivität, gibt es auch Hinweise, dass die prädiktive Validität, also die **Vorhersagegüte der maschinellen Auswahl in manchen Situationen besser** ist. Allerdings muss vorab darauf hingewiesen werden, dass der (Arbeits-)Kontext, in dem die Daten erhoben und die Vergleichstests durchgeführt wurden, eine große Rolle spielt. Zudem erfolgt die Validierung der neuen Tools meist über Korrelationen mit bereits etablierten Verfahren, d. h. es wird beispielsweise gezeigt, dass ein digitales Tool zur Messung von Persönlichkeitseigenschaften positiv mit den Ergebnissen eines klassischen Persönlichkeitstests (z. B. Ten Item Personality Inventory) korreliert. Nun weiß man, dass Persönlichkeitseigenschaften positiv mit Berufserfolg korrelieren, wenn auch nur mit geringen Korrelationskoeffizienten, die deutlich weniger an Varianz des Berufserfolgs aufklären als beispielsweise Intelligenztests (Schmidt und Hunter 1998). Natürlich gibt es auch digitalisierte Intelligenztests, aber oft werden eben andere Aspekte einer potenziellen Eignung gemessen. Und hierbei wird der Umweg, den digitale Tools über die zusätzliche Korrelation zu klassischen Instrumenten gehen, problematisch: Wenn beispielsweise ein digitales Auswahltool 25 % der Varianz einer Persönlichkeitseigenschaft aufklärt und diese wiederum 10 % der Varianz des Berufserfolgs, ist unbekannt, ob das digitale Tool wirklich diese 10 % Varianz erfasst, die mit dem Berufserfolg zusammenhängen (Warszta et al. 2020, S. 35). Angesichts der Abhängigkeit vom Kontext und der verwendeten Tests, ist eine allgemeine Aussage zur Überlegenheit von menschlicher oder maschineller Auswahlentscheidungen nicht sinnvoll. Stattdessen möchte ich zwei konkrete Studien kurz vorstellen.

Hoffman et al. (2017) untersuchten die Profile von ca. 266.000 Neueinstellungen und 400.000 Bewerbern bei 15 Firmen, um herauszufinden, welche Faktoren dazu führen, dass die Mitarbeiter im Niedrigqualifikationssektor länger im Job bleiben (Fluktuation ist ein großes Problem: im Schnitt bleiben Mitarbeiter nur 203 Tage). Im Rahmen eines Fragebogens wurden Daten zu Fähigkeiten, Persönlichkeit, Einschätzung von Arbeitszenarien, etc. erhoben. Ein Matching-Algorithmus glich die Ergebnisse mit dem Profil derjenigen, die länger in den Unternehmen blieben, ab und klassifizierte die Bewerber in drei Kategorien: grün (high potential candidate, d. h. einstellen), gelb (moderate) und rot (low, d. h. nicht einstellen). Die menschlichen Recruiter mussten diesen Vorschlägen, wer eingestellt werden sollte, nicht folgen, d. h. sie konnten einen gelb markierten einstellen, obwohl gleichzeitig die Bewerbung eines grün markierten vorlag (oder, wenn ein roter eingestellt wurde, während grüne oder gelbe es nicht wurden). Die Mitarbeiter, die gemäß der Algorithmus-Empfehlung eingestellt wurden, blieben 11–20 % länger (abhängig von der Kategorie). In diesem Fall **treffen Menschen schlechtere Entscheidungen**, wenn sie sich über die Maschine hinwegsetzen. Dieses Ergebnis bleibt bestehen, wenn neben der Verweildauer noch Produktivität der Mitarbeiter als weiteres Kriterium hinzugezogen wird: Personaler haben kein ‚Näschen‘ für Leute, die mehr leisten, womit eine kürzere Firmenzugehörigkeit wirtschaftlich kompensiert werden könnte (Hoffman et al. 2017, S. 795). Auf die gesamte Arbeits- und Firmenleistung hat die verlängerte Personalbindung gerade bei hohen Flukturationsraten große Auswirkungen, weil sie die Produktivität hochhält (z. B. keine Ausfall- bzw. Anlernzeiten der alten bzw. neuen Mitarbeiter).

Während die eben beschriebene Studie im Kontext einer Branche mit einer hohen Anzahl an eher niedrig qualifizierten Bewerbern stattfand, untersucht eine zweite **Studie hoch qualifizierte Software-Entwickler** (Cowgill 2017). Zudem wurden nicht nur fachliche Anforderungen und Fähigkeiten gemacht, sondern auch weiche Faktoren wie die kulturelle Passung und soziale Kompetenzen wie Verhandlungsfähigkeit und Überzeugungskompetenz. Dies spielt einerseits in der Softwareentwicklung eine Rolle, weil hier in Teams gearbeitet wird, nicht selten die Entwickler(teams) aber auch gegeneinander arbeiten, was die soziale Komplexität erhöht und Mitarbeiter dies erkennen und damit umgehen müssen, um erfolgreich zu sein. Andererseits sind soziale Kompetenzen relevant, weil viele Forscher argumentieren, dass in diesem Bereich menschliche Recruiter den Maschinen überlegen sind (Cowgill 2017, S. 5). In der Studie geht es um die Entscheidung, ob ein Bewerber auf Basis des Lebenslaufscreenings zum Interview mit den möglichen zukünftigen Kollegen eingeladen wird – und ob die menschlichen Entscheider hierbei eine höhere Erfolgsquote verzeichnen als das ADM-System (beide bewerten die gleichen Lebensläufe). Eingesetzt wurden gängige Text Mining- und ML-Algorithmen, die neben den for-

malen Qualifikationen (z. B. Studienabschlüsse) und der Berufserfahrung noch Wortkombinationen im Anschreiben analysierten (z. B. deuten Verben wie „unterstützen“ und „dienen“ auf einen anderen Führungsstil hin als „anweisen“; Cowgill 2017, S. 9). Im Ergebnis hatten die vom **ADM-System ausgewählten Bewerber eine 17 % höhere Wahrscheinlichkeit das Interview mit den Kollegen erfolgreich zu absolvieren**. Außerdem war es eher wahrscheinlich, dass sie ein Jobangebot annehmen und sie waren später auch die produktiveren Mitarbeiter. Dabei war die algorithmenbasierte Entscheidung sogar auch hinsichtlich der weichen Faktoren besser, d. h. die so selektierten Bewerber zeigten später eine höhere Arbeitsleistung (Cowgill 2017, S. 23, 36). Genau wie in der Studie von Hofmann et al. sprechen die Ergebnisse nicht dafür, dass menschliche Recruiter ein besseres Gespür oder Näschen für gute Kandidaten hätten. Beide Studien sind allerdings nur bedingt aussagekräftig. In der ersten wurden nur schriftliche Testergebnisse von Niedrigqualifizierten untersucht und der Algorithmus wurde nicht offengelegt. Daher können die Schlussfolgerungen nicht einfach auf andere Firmen oder interaktivere Auswahlformen übertragen werden. Die Studie von Cowgill ist nur ein vorläufiges Manuskript und untersucht nur eine Jobgruppe mit einfachen statistischen Mitteln, weshalb die Aussagekraft eingeschränkt ist.

In der Praxis wird die Frage, ob maschinelle Personalauswahl „besser“ ist als die menschliche, auch anhand wirtschaftlicher Faktoren beantwortet. **Online-Assessments bieten hier einige Vorteile** (Geister und Rastetter 2009, S. 12), u. a. können sie von jedem Ort der Welt rund um die Uhr absolviert werden, solange ein Internetanschluss besteht. Die Auswertung geschieht automatisiert, was für den Arbeitgeber ökonomisch vorteilhaft ist (Kosten und Zeitaufwand sind niedrig, nachdem die Systeme einmal entwickelt wurden). Außerdem passen sie in das digitale Zeitalter, weshalb man gerade junge Zielgruppen erreichen und sich das Image eines „modernen“ Arbeitgebers geben kann.

Den Vorteilen stehen einige **Nachteile von Online-Assessments** gegenüber (Geister und Rastetter 2009, S. 12). Zunächst sind die hohen Anforderungen an die Gebrauchstauglichkeit der Systeme zu nennen, die auf allen Geräten und Browsern gleich funktionieren müssen, um objektive und reliable Ergebnisse zu generieren. Wer auf einem kleinen Smartphone-Display erst einmal scrollen muss, um eine Testfrage ganz zu lesen, hat schon im Vergleich zur Beantwortung am PC-Bildschirm Zeit verloren. Der Aufbau und das Design solcher Assessments sind deshalb mit einem hohen Aufwand verbunden. Da die Tests immer gleich ablaufen, können sie trainiert werden – oder die Kandidaten holen sich Unterstützung, was die Testergebnisse verfälscht. Außerdem reduzieren Online-Assessments den persönlichen Kontakt zwischen Bewerber und Recruiter.

4.2.2.3 Was man noch machen kann

Seit Jahren schwappen Methoden und Tools aus dem Marketing in die Personalarbeit, in letzter Zeit sind es Begriffe wie Candidate Journey und Candidate Relationship Management. Dem Kunden entspricht ein Kandidat oder Interessent und

statt eines Produktes geht es um den passenden Job. Bei der **Candidate Journey** sollen deshalb die unterschiedlichen Berührungspunkte (Touchpoints), die Kandidaten mit einem Unternehmen haben können, aktiv gestaltet und gesteuert werden. Das heißt vor allem, dass bei den **Touchpoints immer wieder ein Matching** zwischen der Job-Beschreibung und den Interessen bzw. Kompetenzen der Kandidaten stattfinden und eine **automatisierte Ansprache** vorgenommen werden muss. Die Touchpoints können vielfältig sein: von Werbung auf Social-Media-Plattformen wie Facebook und Instagram über Stellenanzeigen bei Job-Börsen oder der Firmenhomepage bis hin zu physischen Events. Dadurch lernen sich Unternehmen und Kandidaten besser kennen, wobei die Firmen die Berührungspunkte nutzen, um die Kandidaten im Verlauf der Reise immer gezielter zu kontaktieren. Am Ende soll die Candidate Journey in eine Bewerbung münden (Nebenziele: mehr Besucher – neue und wiederkehrende Visitors – auf der Karriereseite, längere Verweildauer auf der Seite, Wahrnehmung als moderner Arbeitgeber). Am Anfang steht aber zunächst die Definition von Zielpersonen (oft Persona genannt) aus Unternehmenssicht. Hierbei gilt es herauszufinden, über welche Kanäle bzw. Webseiten man die Persona mit welchen Themen am besten erreichen kann. Hört die Persona z. B. gern Ed Sheeran, so könnten Firmen ihre Werbung auf Youtube oder Spotify einblenden, wenn die Persona einen Sheeran-Song aufruft. Bei einem solchen **Targeting** sind Kanal, Inhalt und Anlass auf die Persona – genauer deren Eigenschaften, Interessen, Erwartungen, Pain-Points und ihr digitales Nutzerverhalten – abgestimmt. Übergeordnetes Ziel ist, dass die Kandidatin auf die Karrierewebsite des Unternehmens oder eine (im Bestfalle personalisierte) Landing Page geht. Sollte sie dort abspringen, kann sie bei einem anderen Touchpoint wieder kontaktiert werden, z. B. beim Lesen eines Artikels über Wirtschaftsspionage oder beim Shoppen nach bestimmten Produkten (je nachdem, welche Themen, Medien und Kontaktpunkte in der Persona definiert wurden).

Auch das Schlagwort **Candidate Relationship Management** stammt aus dem Marketing. Gemeint ist damit die langfristige Pflege von Kontakten zu Kandidaten und Bewerbern über Talent-Pools, regelmäßige Benachrichtigungen über Firmen- und Branchennews, Mitarbeiter-werben-Mitarbeiter-Konzepte, etc. Solche Optionen, die auf Basis von Matching-Verfahren funktionieren, bieten nahezu alle integrierten HR-Softwarepakete aber auch Spezialanbieter (z. B. XING eRecruiting, Talentry). Unter dem Stichwort Advocacy-Marketing sollen nicht nur bestehende Mitarbeiter zu Recruitern werden, indem sie Jobs weiterempfehlen, sondern es soll ein weiteres Matching hergestellt werden, und zwar zwischen Inhalten aus dem Arbeitsumfeld und Jobs. Z. B. soll eine Stellenanzeige mit einer Erfolgsgeschichte aus einem vergangenen Kundenprojekt, dessen Inhalte zur Stellenbeschreibung passen, verknüpft in sozialen Medien ausgespielt werden können.

Ein neues Schlagwort ist **Programmatic Job Advertising**, d. h. der automatische Ein- und Verkauf von Werbeflächen für Stellenanzeigen, wobei die Nutzer die

Stellenanzeige basierend auf ihrem Klickverhalten ausgespielt bekommen. Beiersdorf nutzt dies, damit z. B. Verfahrensingenieure die Jobs angezeigt bekommen, die sie interessieren, wozu auch die Aussteuerung über die richtigen Kanäle und eine Anpassung der Motive gehört (Wittenhagen 2019). Die Software (z. B. Talent-Bait, PERSY von GermanPersonnel oder der Jobspreeder von Wollmilchsau) sagt auf Basis historischer Daten (z. B. Profildaten, Webseiten-Besuche, Klickmuster) und ML-Algorithmen vorher, wie und wo Kampagnen (z. B. Suche nach Softwareentwicklern) am besten ausgespielt werden sollten und überwacht den Erfolg der Anzeigen in den ausgewählten Kanälen (z. B. Klickraten, Bewerberzahlen). Neben demografischem und geografischem Targeting gibt es das Kontext-Targeting, bei dem Text-Mining-Verfahren eingesetzt werden, die Stellenanzeigen anhand von Keywords in Blogs, Online-Zeitungsartikeln, etc. ausspielen. Liest beispielsweise eine Nutzerin auf einem IT-Blog einen Artikel über Java-Programmierung, dann kommen wahrscheinlich Schlagwörter wie Java, J2EE, JVM und OOP vor. Diese erkennt die Software, matcht sie mit Stellenanzeigen und spielt die passende aus (Quast et al. 2018). Ziel ist, dass das Tool die Anzeigen automatisch so aussteuert, dass sie häufiger auf Werbeplätzen angezeigt werden, von denen aus besonders viele Besucher auf die Stellenanzeige klicken oder der Preis pro Klick besonders niedrig ist oder der Inhalt bzw. Kontext sehr passend zur Zielgruppe ist.

Ähnlich verhält es sich mit **dynamischen Stellenanzeigen**, die je nach Kandidat anders aussehen bis hin zu Echtzeitanpassungen aufgrund des aktuellen User-Verhaltens im Internet. Anbieter wie LinkedIn personalisieren und dynamisieren die Anzeigen auf Basis der Profileinstellungen, andere wie Wollmilchsau ziehen die Nutzerdaten über Cookies. Stellenanzeigen.de hält beispielsweise bis zu 50 verschiedene Varianten für jedes Inserat vor, um für jeden Bewerber die am besten passendste auszuspielen.

Unter dem Stichwort **Augmented Writing** werden Text-Mining- und ML-Algorithmen eingesetzt, um Texte wie Stellenausschreibungen oder Arbeitgeberprofile interessanter zu machen. Das bieten beispielsweise Textio, tapRecruit oder die 100 Worte GmbH an: Basierend auf der Auswertung von Millionen von Stellenangeboten und Rekrutierungsmails, sollen die Algorithmen Text-Muster entdecken, die zu höheren Rücklaufquoten und mehr qualifizierten Bewerbern führen. Dann macht die Software entsprechende Vorschläge für die Formulierung von Stellenanzeigen oder E-Mail-Ansprachen – und zwar in Echtzeit (zukünftig auch in Bezug auf die visuelle Gestaltung). Augmented Writing kann auch potenziell **diskriminierende Elemente entfernen** und beispielsweise Anzeigen so aufbereiten, dass sich mehr Frauen bewerben werden, wovon z. B. der Pharmakonzern Johnson & Johnson profitiert hat (Klahre 2017).

Manche Tools gehen noch einen Schritt weiter entlang des Recruiting-Prozesses und unterstützen bei der **Individualisierung des Jobangebots**, einerseits in Bezug auf das Gehalt, andererseits bezüglich weiterer Angebote, die den Bewerber anreizen sollen zu unterschreiben (z. B. Corporate Benefits, Regelungen zu Arbeitszeit und -ort). Oracle und PeopleStrong versprechen beispielsweise, dass mit ihrer Recruiting-Software abgeschätzt werden kann, wie wahrscheinlich es ist, dass ein Angebot angenommen wird (Sánchez-Monedero und Dencik 2019, S. 15). Letztlich wird hier ein Klassifikationsalgorithmus mit Vergangenheitsdaten trainiert, nur, dass nicht die Wechselbereitschaft, sondern eine ‚Unterschriftenbereitschaft‘ abgeschätzt wird. Das funktioniert natürlich nur, wenn interne oder externe Daten aus vergleichbaren Bewerberrunden vorliegen – und zwar über Gehaltsdaten hinaus.

Neben Selbsttests und Sprach- bzw. Videoanalysen gibt es **automatisierte Persönlichkeitsanalysen auf Basis von Social-Media-Daten**. Man muss nicht gleich an den Facebook-Cambridge Analytica-Skandal denken – Persönlichkeitsprognosen gehen auch mit Daten der XING-Plattform (Buettner 2017). Beispielsweise korreliert die Anzahl an Profilbildern positiv mit Extraversion, ebenso die Loginhäufigkeit, während die Anzahl an Blogeinträgen und angegebenen Interessen mit Offenheit korreliert. Der prädiktive Nutzen (Erhöhung der Treffergenauigkeit) des benutzten Entscheidungsbaum-Algorithmus beträgt zwischen 31,4 und 46,2 % (Buettner 2017, S. 25). Der Unterschied zu den anderen Verfahren ist, dass hier **Kandidaten ohne ihr Wissen und Zutun bewertet** werden – ein Vorgehen, das aus ethischer Sicht zu kritisieren ist (siehe Kap. 6).

Ebenfalls Teil des Bewerbungsprozesses ist es, **Absagen an nicht genommene Bewerber zu verschicken**. Die meisten Application Tracking Systems (ATS) unterstützen den automatisierten Versand, nachdem ein menschlicher Recruiter die Entscheidung getroffen hat. Arbeitgeber sollten allerdings darüber nachdenken, welches Signal sie mit einer solchen automatisierten Absage senden – gerade in engen Arbeitsmärkten, weil es hier wahrscheinlich ist, dass sich alle beteiligten Parteien später noch einmal begegnen und zusammenarbeiten wollen (oder müssen). Zudem sind standardisierte Absagen nicht smart, wenn sie nicht datengestützt auf den Einzelfall eingehen. Manche Tools, z. B. Pymetrics, unterbreiten den nicht genommenen Bewerbern immerhin **Vorschläge zur Karriereplanung**, d. h. wie sie ihre Fähigkeiten und Qualifikationen im Hinblick auf zukünftige Positionen weiterentwickeln sollten und welche anderen Stellen interessant wären (Sánchez-Monedero und Dencik 2019, S. 15). Hier sind ganz offensichtlich die gleichen Recommender-Algorithmen am Werk wie bei der Vorauswahl von Kandidaten. Inwieweit **Chatbots für Absage- und Feedbackgespräche** zukünftig zum Einsatz kommen (wie z. B. Mya es bereits heute können soll), muss sich noch zeigen und wird sicher von Stelle zu Stelle verschieden sein. Die Gegenargumente

(unpersönlich, unmenschlich und unwürdig) wiegen schwer – und oftmals wiegen sie die Vorzüge (Gleichbehandlung aller unter Beachtung rechtlicher Vorgaben, kostengünstig, zeitliche und emotionale Entlastung für Recruiter, kein Zeitdruck bei Beantwortung der Fragen) auch auf.

4.2.2.4 Was (noch) nicht geht

Ein Argument für ADM-Systeme im Recruiting ist, dass diese objektiver bewerten als menschliche Recruiter, weil sie keinen der vielen kognitiven Verzerrungen unterliegen, die bei menschlichen Entscheidern immer wieder festgestellt wurden (siehe dazu Kanning 2015, S. 138 ff.). Klar ist: Beim ADM lassen sich kritische Variablen wie Geschlecht, Hautfarbe, religiöse Orientierung, soziomaterieller Status, etc. explizit ausschließen. Jedoch: **Es bleibt die Gefahr der Verzerrung** (bias), weil auch andere Daten Aufschluss über diese Variablen geben und die Trainingsdaten an sich verzerrt sein können. Beispielsweise lässt sich vom Vornamen eines Bewerbers meist auf das Geschlecht schließen und der Wohnort lässt Rückschlüsse auf den soziomateriellen Status zu. Da ML-Algorithmen nach Mustern suchen, **können Ungleichheiten reproduziert werden**, wenn die Trainingsdaten strukturelle Verzerrungen aufweisen und diese nicht korrigiert werden. Wenn es z. B. in der Vergangenheit nur männliche Weiße in das Top-Management geschafft haben, könnte das der Algorithmus als erfolgreiches Muster identifizieren. Dies sind nicht nur theoretische Gefahren, wie Vorfälle bei Amazon, Google u. a. zeigen, wo die verwendeten ML-Algorithmen nach Geschlecht bzw. Rasse diskriminierten (Norten 2014; Vincent 2018; s.a. Gärtner 2017). Ob die landläufige Meinung von „bias in, bias out“ auf ein ML-Modell tatsächlich zutrifft, hängt entscheidend von der Form der Verzerrung und dem Training ab (Rambachan und Roth 2019, S. 2, 12). Wichtig für ein tieferes Verständnis des Problems ist, dass es unterschiedliche Verzerrungsmöglichkeiten gibt. Z. B., wenn die Anzahl an Datenpunkten für verschiedene Klassen nicht gleich ist, etwa weil deutlich mehr männliche Weiße in der Klasse „erfolgreiche Mitarbeiter“ sind, oder die Fehlerraten des Modells zwischen den Klassen variieren (Sanchez-Monedero et al. 2019, S. 2). Deshalb setzt jede Form von ADM im Recruiting-Umfeld eine fundierte Kenntnis darüber voraus, wie die Daten strukturiert sind, worin sich geeignete Kandidaten bzw. Bewerber von nicht geeigneten unterscheiden und wie die Algorithmen funktionieren. Sich nur darauf zu verlassen, dass die Algorithmen adäquate Zusammenhänge erkennen, verkennt die technologisch-methodische Vorgehensweise bei ML. Darauf verweisen auch die Anbieter selbst. Beispielsweise geben die Macher des Chatbots Mya keine Garantie dafür, dass Mya eine korrekte Auswahl trifft. Positiv gewendet könnte man auch sagen, dass die ADM-Systeme der Gesellschaft einen Spiegel vorhalten, weil sie die Voreingenommenheit offenlegen, die sich in

die Daten eingeschrieben haben. So soll Mya beispielsweise Muster in den Fragen und Verhaltensweisen entdecken können, die ohne das System unerkannt geblieben wären. Eines dieser Muster ist, dass Bewerber, die während eines Interviews auf Gehaltsfragen pochen, mit höherer Wahrscheinlichkeit die Firma wieder verlassen als andere (Sánchez-Monadero und Dencik 2019).

Verzerrte Daten sind nur ein Teil des Problems, das bei der Suche nach Mustern in großen Datensätzen auftaucht. Insbesondere bei der Verwendung von KNN/DL kommt es zu einem sogenannten **black boxing**, das heißt, es ist nicht klar, wie genau die eingesetzten Algorithmen die Daten verarbeiten und letztlich zu Entscheidungen kommen (Castelvecchi 2016). Das liegt einerseits an der Vielzahl an Variablen und vorgenommenen Rechenoperationen, andererseits aber auch daran, dass die Logik eines KNN nicht der menschlichen Logik folgt: Ein KNN für die Bilderkennung zerlegt beispielsweise Zahlen nicht notwendigerweise in Striche oder (Halb-)Kreise, sondern in (zunächst) undefinierte Formen, deren Sinnhaftigkeit menschliche Bearbeiter nicht verstehen (Burrell 2016, S. 6). Wissenschaftler suchen unter den Begriffen „**Explainable AI**“ bzw. „**Interpretable ML**“ nach Lösungen dafür, wie die interne Logik eines ADM-Systems für Menschen nachvollziehbar gemacht werden kann und entwickeln entsprechende Tests (siehe dazu Guidotti et al. 2018). Allerdings sind die Methoden zum Öffnen der Black Box aufwändig. Aus dem Personalbereich hat z. B. Pymetrics ein Tool namens Audit-AI, das potenzielle Diskriminierungen anzeigt, auf GitHub veröffentlicht und auch HireVue gibt Einblick in die statistischen Vorkehrungen, mit deren Hilfe Diskriminierung minimiert werden soll (sowohl auf der Homepage als auch in den öffentlich einsehbaren Patentanmeldungen). Insgesamt sind die Informationen, die Externe über die Funktionsweise der Tools erhalten, aber nicht vollumfassend. Unabhängig davon, ist **noch nicht geklärt, was unter Diskriminierung genau zu verstehen ist** (z. B. im Verhältnis zu inhaltlicher und prozeduraler Fairness oder aufgrund unterschiedlicher Gesetzgebungen in den USA und Europa), und wie sie numerisch abgebildet bzw. vermieden werden kann (siehe dazu Sanchez-Monadero et al. 2019, Kap. 6).

Für Bewerber, aber auch Werbende, gibt es eine **weitere Black Box: die Algorithmen von Drittanbietern**, z. B. sozialen Netzwerken oder Job-Plattformen, welche gerade für Recruiting eine wichtige Rolle spielen. Eine aktuelle Studie hat Anzeige- und Klickraten aus über 191 Ländern und von mehreren Plattformen (u. a. Facebook, Instagram, Google Display Network) ausgewertet, um herauszufinden, ob Frauen dort mathematisch-naturwissenschaftliche Jobs seltener angezeigt werden. Dies war der Fall – konnte aber nicht auf Verzerrungen im Datensatz zurückgeführt werden, sondern auf den Preismechanismus der sozialen Netzwerke, der das Ausspielen einer Anzeige für Frauen höher bepreist als für Männer (Lambrecht und Tucker 2019).

Sind die autonomen Entscheidungen eines ADM-Systems nicht nachvollziehbar, so ist dies nicht nur unbefriedigend und erscheint bei Personalfragen als kaum

legitim, es ist auch in legaler Hinsicht kritisch. Einerseits stellen sich **Datenschutzfragen**, insbesondere nach Art. 22 DSGVO. Die Vorschrift verbietet, dass Personen einer automatisierten Entscheidung unterworfen werden dürfen, die dann rechtliche Wirkung entfaltet. Dies wäre beispielsweise bei einem automatisierten Test oder Chatbot der Fall, wenn die Auswahlentscheidung direkt zu einem Arbeitsvertrag führen würde. Die Konsequenz dieses Verbots: Die Bewertung eines Bewerbers durch ein ADM-System darf **nur zur Vorbereitung der Entscheidung durch einen Menschen** dienen. Ein bloßes Scoring oder Profiling zur Entscheidungsvorbereitung fällt aber nicht unter Art. 22 Abs. 1 DSGVO (Blum und Kainer 2019, S. 26; Huff und Götz 2019, S. 75 f.).

Zudem gibt es **Haftungsfragen** (Martini 2019, S. 274 ff.): Was passiert beispielsweise, wenn ein Chatbot falsche Informationen weitergibt, durch die beim Mitarbeiter ein Schaden entsteht? Was passiert, wenn ein Bot gehackt wird und Unbefugte an vertrauliche Daten gelangen? Ein ADM-System ist keine eigenständige rechtliche Person und kann daher nicht haftbar gemacht werden (wohl aber der Arbeitgeber, der dann u. U. den Hersteller der Software in Regress nehmen kann). Dass die **Handlungen von Maschinen nicht jenen von Menschen gleichgesetzt werden können**, ist bisher kaum bezweifelt worden. Zum menschlichen Handeln gehören sowohl die Möglichkeit zum Andershandeln (andere/neue Handlungsweisen oder Ziele) als auch das Angeben von Gründen und nicht bloß (naturwissenschaftlichen) Ursachen (Habermas 2004, S. 873 f.). Diese Kriterien müssten vollkommen autonome Entscheidungssysteme erfüllen, um diesen Namen zu verdienen (siehe Abschn. 3.7). Verfügbare ADM-Systeme können aber bestenfalls die Regeln ihres Verhaltens verändern, sich aber nicht neue bzw. andere Ziele setzen und auch keine Gründe dafür angeben.

Unabhängig von der Datenschutz- und Haftungsproblematik ist die Frage, wie weit die **Akzeptanz von automatisierten Entscheidungen** geht. In Umfragen wird immer wieder thematisiert, dass Bots nicht zu einführender, empathischer Kommunikation in der Lage wären und damit auch keine Wertschätzung ausdrücken können (Laumer et al. 2019, S. 14; Verhoeven 2020, S. 125). Wie bereits erwähnt wird an Emotionalität und Multimodalität gearbeitet (Abschn. 2.2) und es gibt Hinweise, dass Menschen zum Anthropomorphisieren neigen, also Maschinen menschliche Züge zusprechen (man denke daran, wie oft man schon seinen Computer verflucht hat und wie sehr manche ihr Auto lieben; siehe auch die Studien zum „Uncanny Valley“). Aber gerade bei lebenswichtigen Entscheidungen wie der Wahl eines Jobs (auch wenn man ausgewählt wird) bevorzugen die meisten aktuell menschliche Selektierer.

Die **Kritik an automatisierten Auswahlsystemen** richtet sich vor allem an jene Anbieter, die auf Basis von **Sprach- oder Videodaten auf die Persönlichkeit**

und Fähigkeiten der Bewerber zurückzuschließen (z. B. HireVue, Precire, SEED-LINK). Zunächst wird kritisiert, dass das methodische Vorgehen und die verwendeten Algorithmen nicht umfassend offengelegt sind (Schmidt-Atzert et al. 2019, S. 19). Dass diese als Firmengeheimnis gelten, ist zwar nachvollziehbar, aber dennoch verlangen nicht nur Wissenschaftler, sondern auch Praktiker nach Informationen über die Wirkungsweise, um eine fundierte (Kauf-)Entscheidung zu treffen. Es kann allerdings bezweifelt werden, ob eine bloße Offenlegung schon ausreicht, um wirklich zu verstehen, was genau passiert und wie valide dies ist (siehe Kap. 6).

Schwerwiegender ist deshalb, dass die **inkrementelle Validität nicht nachgewiesen** ist (Schmidt-Atzert et al. 2019, S. 20). Es ist also unklar, ob die Tools überhaupt zu besseren Vorhersagen führen als ein klassischer Fragebogen zur Erfassung von Persönlichkeitsmerkmalen. Aber selbst wenn dies gelänge, muss beachtet werden, dass Persönlichkeitsmerkmale eine **geringe Vorhersagegüte (prädiktive Validität) für den Berufserfolg** zeigen: Meta-Studien kommen zu dem Ergebnis, dass, je nach Persönlichkeitsmerkmal, die prädiktive Validität zwischen 0,8 (Verträglichkeit) und 0,22 (Gewissenhaftigkeit) liegt. Zum Vergleich: Intelligenztests schneiden deutlich besser ab, weil sie eine Vorhersagegüte für die Arbeitsleistung von 0,39 für niedrig-qualifizierte Jobs und 0,74 für Führungskräfte haben (Schmidt et al. 2016, S. 13, 65). Auch hinsichtlich weiterer Validitätsmaße, z. B. inhaltliche und diskriminante Validität, müssen ML-basierte Persönlichkeitsdiagnosen untersucht werden (Bleidorn und Hopwood 2019, S. 194 ff.).

Ein weiteres großes Problem besteht, wenn die **Stichprobengröße im Vergleich zu den analysierten Variablen zu klein** ist. Bei Precire wurden 5201 Personen untersucht, aber 564.367 Variablen (Features) analysiert, was für die Erstellung eines Vorhersagemodells für psychologische Merkmale sehr ungünstig ist, denn dann ist es nahezu unvermeidbar, dass zufällig im Datensatz vorhandene Konstellationen von Features ausgebeutet werden (Schmidt-Atzert et al. 2019, S. 20). Im Lichte dieser Betrachtung erscheint das Verhältnis des von L'Oréal eingesetzten Systems noch ungünstiger: 74 Personen bei 39.672 Variablen (siehe Beispiel-Kasten oben, Queb 2019). In Bezug auf das automatisierte Auswerten von Bewerbungsunterlagen ist zu bedenken, dass es bisher keine Befunde gibt, dass das Screening von Lebensläufen durch menschliche Recruiter zu validen Vorhersagen des Berufserfolgs führt (Kanning 2015). Mittlerweile bewirbt Precire sein Produkt auf der Homepage nicht mehr im Hinblick auf die Analyse von Persönlichkeitseigenschaften bzw. zur Eignungsdiagnostik, sondern mit Fokus auf die Analyse der Wirkung von Sprache im beruflichen Kontext.

Der **Kontext, aus dem die Trainingsdaten stammen**, ist – wie bei allen maschinellen Lernverfahren – zentral für das, was das System lernt. So hat beispielsweise ein Text-Mining-Algorithmus, der anhand von professionellen Texten trai-

niert wurde (z. B. Zeitungsartikel), etwas Anderes gelernt als der gleiche ML-Algorithmus, der mit fehlerbehafteten Texten in Bezug auf Grammatik und/oder Rechtschreibung (z. B. Posts auf Twitter) trainiert wurde. Die jeweiligen ADM-Systeme eignen sich deshalb immer nur gut, wenn die gleichen Text-Genre verwendet werden.

Sprache allein ist schon komplex und noch komplexer sind multimodale Interaktionen (die Daten für alle Sinne bieten), weil sie in den **sozio-kulturellen Hintergrund** eingebettet sind und auf **situative Anforderungen** reagieren (siehe Oviatt et al. 2019). In der heutigen Arbeitswelt braucht es zudem Fähigkeiten wie kreatives Denken, Überzeugungskraft und **viele weitere Kompetenzen** (siehe Van Laar et al. 2017). All dies lässt sich mit den bisherigen ADM-Systemen kaum abbilden, da diese mit stark vereinfachten Annahmen und Daten arbeiten. Beispielsweise analysieren Anbieter von Augmented Writing-Software wie 100 Worte den Gender-Bias dadurch, dass sie zwischen femininen Wörtern (z. B. aneinander, an-sprechend, angenehm) und maskulinen Wörtern (abenteuerlich, aktiv, ambitioniert) unterscheiden (Spitzer et al. o. J.). Daran zeigt sich zunächst, welchen tradierten Rollenvorstellungen die Programmierer folgen. Und es muss kritisch gefragt werden, ob sich **Menschen und Jobs in Kategorien einteilen** lassen und, wenn ja, wie viele es bräuchte. Die Annahme, dass es mehr als zwei sind, erscheint plausibel. Man denke nur an den kleinen, aber folgeschweren Unterschied bei den Job-Anforderungen „kennen“ und „können“: Letzteres verweist auf einen höheren Kompetenzgrad (Experte statt Anfänger) in den Anforderungen, weshalb der Text Mining-Algorithmus in der Lage sein sollte, die unterschiedlichen Gewichtungen der Anforderungen in einer Stellenanzeige zu erkennen.

Ebenfalls kritisch betrachtet werden Anbieter von Online-Spielen (z. B. HireVue, Knack). Neben der Frage, für welche Zielgruppen sich solche Spiele eignen (man denke an Auszubildende vs. Manager), ist vor allem zu fragen, inwiefern die **Spiele die realen Anforderungen der Stelle abdecken**. Außerdem wird vom Spielverhalten auf auswahlrelevante Personenmerkmale geschlossen, was aus eignungsdiagnostischer Sicht zu kritisieren ist.

Eine weitere Frage ist, ob das Virtuelle dem Analogen entspricht, unabhängig davon, ob Verhalten oder Eigenschaften gemessen werden sollen. Erste Antworten liefert eine Studie, in der 245 Probanden daraufhin untersucht wurden, ob sie einem **Chatbot gegenüber andere Persönlichkeitsmerkmale präsentieren als Menschen** (Mou und Xu 2017). Im Ergebnis zeigten die Chatbot-Nutzer niedrigere Ausprägungen entlang der „Big Five“-Persönlichkeitsmerkmale, d. h. sie waren weniger offen für Neues, weniger extrovertiert, gewissenhaft und verträglich (Mou und Xu 2017, S. 436 ff.)

Ein Grundsatzproblem ist, dass bei vielen Fragen des Personalmanagements die **Variablen, insbesondere Erfolg, nicht klar definiert** sind: Woran bemisst man erfolgreiche Kandidaten, Bewerber, Mitarbeiter (z. B. Zeit bis zur nächsten Beförderung, Gehalt, Bewertung der Leistung gemäß Zielvereinbarung, Mitarbeitergespräch oder 360°-Feedback)? Was ist eine erfolgreiche Personalmarketing-Kampagne (z. B. Klicks auf die Anzeige, Anzahl oder Qualität eingegangener Bewerbungen, Kosten pro Klick oder Bewerbung)? Anders als im Marketing oder Finanzbereich konnte die HR-Profession sich noch nicht auf Standards und Konventionen einigen. Dieses grundsätzliche Problem betrifft aber auch Verfahren, die nicht auf Daten und Algorithmen basieren.

4.3 Personalplanung & -einsatz

Sind geeignete Kandidaten einmal gefunden und eingestellt, müssen sie für die anfallenden Aufgaben und entsprechend ihrer Fähigkeiten eingesetzt werden. Diese Grundaufgabe der Personal(einsatz)planung kann um zusätzliche Anforderungen erweitert werden, insbesondere hinsichtlich weiterer Persönlichkeits- und Verhaltensmerkmale (z. B. Bedürfnisse und Motive, Leistungsvermögen) und entlang der Zeitachse (strategische Personalplanung).

4.3.1 Strategische Personalplanung

4.3.1.1 Worum es geht

Im Rahmen strategischer Planungsprozesse ist HR gefordert, zu analysieren, wie viel Personal zu welcher Zeit an welchem Ort mit welcher Qualifikation gebraucht werden wird, um davon abgeleitet entsprechende Maßnahmen in der Rekrutierung, Personalentwicklung, Nachfolgeplanung oder Freisetzung anzustossen. Bildlich gesprochen braucht HR dazu sowohl ein **Mikroskop** als auch ein **Teleskop**, um vorhandene Daten genau analysieren und zukünftige Entwicklungen prognostizieren bzw. Szenarien für unterschiedliche Zukünfte aufstellen zu können.

4.3.1.2 Wie es geht

Die einfache strategische Personalplanung berechnet den (Brutto-)Personalbedarf durch Addition des aktuellen Bestands mit dem erwarteten Ersatz- bzw. Zusatzbedarf sowie dem Saldo aus Ab- und Zugängen (Scholz und Scholz 2019, S. 93 f.). Viele Organisationen verlassen sich hierfür schon lange nicht mehr nur auf intuitive Schätzverfahren, sondern setzen **Trend- bzw. Zeitreihenanalysen**, Bayes-

Netzwerke oder Markov-Ketten ein, um auf Basis vergangener Personalbewegungen zukünftige Entwicklungen zu simulieren (siehe Beispiel-Kasten). Viele der seit Jahrzehnten erhältlichen Software-Pakete ermöglichen dies und erlauben zudem eine Verknüpfung dieser Daten mit den dazugehörigen Personalkosten (Strohmeier 2015). Mit PYTHIA stellt die „Initiative Neue Qualität für Arbeit“ (INQA) sogar ein kostenfreies Tool bereit.

Beispiel: KI-basierte Personalbedarfsplanung für die Barmherzigen Brüder

Der Krankenhausverbund „Barmherzige Brüder Österreich“ hat 2018 gemeinsam mit der Firma Solgenium eine Lösung für Personalplanung im Krankenhaus in Linz entwickelt. Das System basiert auf der Software CARA, die mit KI/ML-Methoden arbeitet. Damit soll eine leistungs- und bedarfsorientierte sowie transparente und erweiterbare Personalbedarfsberechnung für verschiedene Berufsgruppen (Pflegekräfte, gehobener Dienst für Gesundheits- und Krankenpflege, etc.) möglich werden (Diensthuber und Gromer 2018, S. 5, 12).

Als Datenbasis dienten bereits verfügbare klinische Routinedaten (Pflegeleistungs- und Personalberechnungsdaten, Schnitt/Naht-Zeiten, etc.) sowie eigens an den Pilotstationen erhobene Daten zu Umfang, Häufigkeit und Dauer von Tätigkeiten (Diensthuber und Gromer 2018, S. 16, 18). Eine Zielsetzung war, die zukünftige Verteilung von patientennahen und patientenfernen Pflegetätigkeiten sowie die ausbildungsrelevanten Inhalte abzubilden, um das zukünftige Tätigkeitsspektrum und die notwendigen Kompetenzen zu definieren. Mit Hilfe von Bayes-Netzen, die eine Abfolge von bedingten Wahrscheinlichkeiten in Form eines Graphen repräsentieren, wurden fachbereichsspezifische Teamzusammensetzungen modelliert, damit die richtige Anzahl an Health Professionals mit den richtigen Kompetenzen zur richtigen Zeit am richtigen Ort zur Verfügung steht (Diensthuber und Gromer 2018, S. 22 ff.). Damit lassen sich auch Szenarien zum Skill & Grade-Mix erstellen, die sowohl bereichs- und berufsguppenabhängige Unterschiede berücksichtigen als auch lernend sind, das heißt, die Berechnungsgenauigkeit steigt durch die kontinuierliche Integration von Daten (Diensthuber und Gromer 2018, S. 31 ff.).

Wichtig ist zudem, dass bisher unbekannte Berechnungsparameter, wie z. B. neue gesetzliche Vorgaben und auch neue Datenpunkte (z. B. durch die elektronische Erfassung von Pflegeleistungen, elektronische Fieberkurven oder externe Datenquellen), berücksichtigt werden können (Diensthuber und Gromer 2018, S. 49).

Wie smart die mit den Tools gemachten Prognosen sind, hängt weniger von der IT ab, sondern vielmehr von der **Art und Vielzahl der berücksichtigten Mitarbeiter**.

beitergruppen und der Einschätzung der zukünftigen Unternehmensentwicklung. Die Frage, für welche Mitarbeitergruppen geplant wird, stellt sich nicht nur für Mitarbeitende mit einem klassischen Arbeitsvertrag (siehe das Beispiel der Barmherzigen Brüder), sondern auch für solche mit einem Dienst-, Werk- oder Überlassungsvertrag: Inwieweit sollen Berater, Freelancer, Clickworker und Leiharbeitnehmer in der Planung berücksichtigt werden? Die Antwort darauf lautet: Es kommt auf die Unternehmenssituation und strategische Relevanz der jeweiligen Gruppe an. Weil die Frage situationsspezifisch beantwortet werden muss, wird sie im Folgenden nicht weiter behandelt. Bei der Einschätzung zukünftiger Entwicklungen lässt sich die **Prognosegüte durch zwei Maßnahmen schärfen:** (1) indem externe Daten hinzugezogen werden und (2) indem auch qualitative Aspekte beachtet werden.

Anbieter wie HR Forecast ermöglichen es, dass **externe Daten** mit in der Planung berücksichtigt werden, beispielsweise Daten über den Arbeitsmarkt (z. B. von der Bundesagentur für Arbeit oder aus Zeitungsberichten), die wirtschaftliche Entwicklung (z. B. Prognosen von Wirtschaftsinstituten), prognostizierte Studienabgangsquoten und Daten über ausgeschriebene Stellen bei Jobbörsen wie Monster oder StepStone (Vetter und Zacharides 2016; Vetter 2018). Dies sind Indikatoren dafür, wie einfach oder schwierig es werden kann, das benötigte Personal auch zu gewinnen: Gibt es in der Postleitzahlregion der Firma viele Arbeitssuchende, (Berufs-)Schul- bzw. Studienabsolventen und sind wenige ähnliche Stellen ausgeschrieben, so wird dies leichter fallen, als wenn Stellenanzeigen schon seit Monaten online sind und Vollbeschäftigung in einer Region herrscht. Berücksichtigt man zusätzlich die Gehälter für die gesuchten Jobs, können Personaler sehen, wo Engpassstellen zu erwarten sind und sie können ihre Zeit und Aufmerksamkeit entsprechend fokussieren. Manche dieser Daten konnten Unternehmen zwar schon vor Jahren berücksichtigen (siehe Strohmeier 2015, S. 29 f.), aber erst durch die breite Digitalisierung von Stellenanzeigen, die Verwendung von Crawlern und Schnittstellenöffnungen können sie einfacher einkalkuliert werden und sind repräsentativer. Durch den Einsatz von **Clusteranalysen** für die Einteilung des Personals in Gruppen können zudem Fehler bei der Überprüfung der aktuellen Ist-Personalstruktur vermieden bzw. Erkenntnisse über zuvor nicht wahrgenommene Zusammenhänge, z. B. zwischen Betriebszugehörigkeit, Abteilung, Jobfamilien, Gehalt, etc., gewonnen werden (Strohmeier 2015, S. 36 f.).

Die Idee, die Qualifikationen und Kompetenzen oder gar Motivation und Potenzial als **qualitative Daten** über das Personal bei der strategischen Planung zu berücksichtigen, gibt es schon lange. In der Praxis sind die Skill-Datenbanken allerdings oft schlecht gepflegt und/oder nicht mit dem HR-Kernsystem verbunden.

Zudem wird die Analyse relativ komplex, wenn viele Kompetenzen mit den jeweiligen Qualifikationsleveln und entlang von Entwicklungspfaden pro Jobfamilie berücksichtigt werden (Berendes und Werner 2015, S. 268 f.). Um im Bild mit dem Mikroskop und dem Teleskop zu bleiben: Ein **hochauflösender Blick in die Belegschaftsstruktur** ermöglicht ausdifferenzierte Prognosen. Es lässt sich beispielsweise analysieren, wie viele extern beschaffte Mitarbeitende es auf Senior-Level an einem Standort über einen bestimmten Zeitraum gab und was diese im Vergleich zu Seniors kosten, die intern Karriere gemacht haben. Diese Drill-downs fallen noch in den Bereich Descriptive und Diagnostic Analytics, nimmt man aber die Abschätzung z. B. des kostenoptimalen Mixes an extern beschafften und intern entwickelten Mitarbeitern hinzu oder die Korrelation von internen/externen zur Fluktuation, bewegt man sich im Predictive-Umfeld.

Die bereits beschriebenen **Text-Mining-Systeme und KNN** können teilweise bei der Bewältigung der Komplexität helfen. Einerseits, indem sie dabei unterstützen, die Angaben in den **unternehmensinternen Skill-Datenbanken** mit jenen **externer Plattformen** (z. B. XING, LinkedIn, Researchgate) automatisch zu vergleichen und Lücken zu füllen. Auf diesen Plattformen gibt es zusätzlich zu den Selbsteinschätzungen noch Bestätigungen bzw. Endorsements von ehemaligen Kunden oder Kollegen. Diese und weitere Indikatoren (z. B. Veröffentlichungen, Patentanmeldungen, Mitarbeiterentwicklungsgespräche) können als zusätzliche Belege für die tatsächlich vorhandenen Kompetenzen herangezogen werden (Lübuda und Fleischmann 2018, S. 240 f.). Andererseits, indem sie die Auswertung beschleunigen und nutzerfreundlich gestalten, was sowohl Freitextsuchen (statt Boolescher Operatoren, Filter oder SQL-Abfragen) als auch die grafische Aufbereitung beinhaltet. Sowohl die großen Anbieter wie SAP SuccessFactors als auch kleinere wie HR Forecast oder Dynaplan bieten beispielsweise Risiko- und Engpassanalysen für Kompetenzen und visualisieren diese in Form von Heat Maps, Treiberbäumen, etc. Im Idealfall soll damit zusätzlich berechnet werden können, wie sich die Personalplanung auf finanzielle Kennzahlen auswirkt.

4.3.1.3 Was man noch machen kann

Weitere Variablen, die bei der Planung berücksichtigt werden können, sind die karrierebezogenen Entwicklungen von Mitarbeitern, sowohl vertikal (Beförderungen) als auch horizontal (Jobwechsel). Unter Einsatz diverser ML-Algorithmen (u. a. Clustering und Entscheidungsbäume) konnten Squires und Kollegen (2017) für Wissensarbeiter in einer Bank einige Muster aufdecken, z. B. zwischen **Karrierepfaden** und Abteilungszugehörigkeit, Jobwechseln, Beförderungen und Leistung (analysiert wurden 2804 Jobsequenzen). Solche Auswertungen können dabei helfen zu erkennen, wer schnell Karriere macht, ob abteilungsüberschreitende Karrieren zu Beförderungen führen und ob überhaupt beim Trend zu „Boundaryless Ca-

reers“ mitgemacht werden sollte, ob also Arbeitnehmer und -geber davon profitieren.

Neben quantitativen und qualitativen Aspekten ließe sich der generierte **Wert des Personals, das Humankapital**, berechnen. Solch eine Messung ermöglicht z. B. die „Saarbrücker Formel“ (Scholz & Scholz 2019, S. 133 f.): Die Formel stellt das Wissen und die Motivation der Mitarbeitenden in den Mittelpunkt und berechnet das Humankapital anhand von Faktoren wie Leistungsbereitschaft, Fluktuation (negativer Einfluss) und Personalentwicklung, die das wertschöpfungsrelevante Wissen auf dem aktuellen Stand halten soll. Humankapitalberechnungen haben sich bisher in der Praxis nicht etabliert, auch wenn es seit 2018 eine ISO-Richtlinie dazu gibt (ISO-30414).

4.3.1.4 Was (noch) nicht geht

Oft liegen in der Praxis nur **unzureichende Daten über das Können** der Belegschaft vor: zu alt sind die Angaben, zu unterschiedlich die verwendeten Begriffe, zu viele Angaben beruhen auf Selbsteinschätzungen, was sowohl zu Über- als auch Unterschätzung führt. Als Kontrollmechanismen könnten (360°-)Feedbacks, Endorsements (Bestätigungen) oder Leistungsbewertungen dienen. Diese werden aber ebenso selten berücksichtigt wie Faktoren, die das **Wollen bzw. die Motivation** abbilden (siehe aber Abschn. 4.3.2). Eine richtige Hilfe wäre strategische Personalplanung allerdings dann, wenn nicht nur Personen an die richtige Stelle gesetzt werden, sondern deren Leistungspotenzial genau dort abgeschöpft werden kann. Dazu müssen Handlungstreiber wie das Wollen und Können akkurat berücksichtigt werden.

Viele der gängigen Textanalyse- und Klassifikationsalgorithmen sortieren die Daten in wenige oder gar dichotome Kategorien, d. h. entweder/oder bzw. 0 vs. 1, was der Komplexität des Problems nicht immer angemessen ist. Gerade bei der qualitativen strategischen Personalplanung zeigt sich: **Managementaufgaben sind vielschichtig und unsicher**. Wer kann schon genau sagen, welche Kompetenzen mit welchem Qualifizierungsniveau Mitarbeiter zukünftig (nicht mehr) benötigen, wenn neue Produkte am Markt eingeführt (und alte eingestellt) werden? Oder wie viele Trainings benötigt werden, um neue Fähigkeiten zu erlernen bzw. ein höheres Niveau zu erreichen? Oder welche Markteinflüsse sich wie auf die Wertschöpfungskette auswirken? Die Antworten darauf müssen unscharf bleiben. Manche Algorithmen (z. B. KNN, Fuzzy Clustering, Evolutionäre Algorithmen, Bayes-Netze) beachten dies, indem sie nicht dichotomisieren, sondern Wahrscheinlichkeiten zwischen 0 und 1 angeben – aber die Daten, mit denen sie trainiert werden, bleiben ungenau und unsicher.

4.3.2 Personaleinsatz

4.3.2.1 Worum es geht

Nach der strategischen Planung muss der operative Personaleinsatz gemanagt werden: Wer arbeitet wann mit wem an welcher Aufgabe zusammen? Neben der Bedarfsplanung, die das Arbeitsaufkommen aggregiert, gibt es die Schicht- und Dienstplanung, in denen geeignete Schichtdienste generiert und die Mitarbeiter eingeteilt werden. Ziel und Hintergrund eines möglichst intelligenten Personaleinsatzes ist es, Personalkosten durch die Vermeidung teurer Überstunden und Leerlaufzeiten zu reduzieren, gleichzeitig die Mitarbeitermotivation hoch zu halten und die Zufriedenheit der Kunden zu steigern – ohne dabei rechtliche Vorgaben oder ethisch-moralische Prinzipien zu verletzen.

4.3.2.2 Wie es geht

In der Praxis erfolgt die kurz- bis mittelfristige **Einsatzplanung oft über MS Excel, per Telefon, direkter Ansprache oder Social-Messenger** (laut einer Umfrage der Working Time Alliance unter 214 Unternehmen arbeiten 26 % mit Tabellenkalkulationsprogrammen und 24 % mit Zetteln, Listen oder Magnettafeln; WTA 2019, S. 17). Damit lässt sich zwar schnell auf Ereignisse reagieren, die zu Planänderungen führen (z. B. Mitarbeiter werden krank oder kündigen, ein unerwarteter Kundenauftrag verlangt nach Zusatzschichten), es führt aber nicht selten zu Konfusion und Korrekturen, die manuell vorgenommen werden müssen – sie können sogar gegen Betriebsvereinbarungen oder Datenschutzgesetze verstossen, wenn z. B. Dienstpläne mit persönlichen Daten wie Namen, Adresse und Telefonnummer über Social-Messenger-Gruppen geteilt werden. Diese **Intransparenz und Ineffizienz** abzuschaffen und gleichzeitig die **Bedürfnisse der Mitarbeiter** zu berücksichtigen, haben sich einige Firmen zum Ziel gesetzt, z. B. ARGOS, Atoss, GFOSS, Inform, Mindsquare, Quinyx oder Vote2Work (siehe Beispiel-Kasten).

Auf Basis von **Vergangenheitsdaten** über Aufträge, zu produzierende Stückzahlen, eingehende Beschwerde- oder Kundenanfragen, Kassentransaktionen, etc. lässt sich die anstehende Arbeit prognostizieren, im Bestfall in Bezug auf das anfallende Volumen sowie in zeitlicher, örtlicher und fachlicher Hinsicht. Je nach Branche funktioniert das unterschiedlich gut: Im Einzelhandel gibt es meist sehr fein granulierte Daten aufgrund der Kassentransaktionen, womit der saisonale, wöchentliche und tägliche Personalbedarf abgeschätzt werden kann. Ähnlich einfach ist die Abschätzung bei Call Centern, in denen dann Hochbetrieb herrscht, wenn die Kunden Arbeitspausen haben (also morgens, mittags und abends). An Flughäfen kann der Bedarf an Wartungspersonal anhand der absolvierten Flugstunden

und des Instandhaltungsprogrammes eines Flugzeugs bedarfsorientiert geplant werden. Eine Methode sind Bayes-Netze, die eine Abfolge von bedingten Wahrscheinlichkeiten modellieren, wobei Vergangenheitswerte z. B. über Flugzeug-Instandhaltungszyklen und Fehlerwerte vergangener Personalplanungen als Vorrücken in das Bayes-Netz eingehen (Wagner 2009, S. 98 ff.). Aus solchen und weiteren Daten können Softwarepakete die bedingten Wahrscheinlichkeiten lernen und die optimale Struktur des Bayes-Netzes vorschlagen.

Viele der verwendeten Verfahren sind aber eher Vorstufen von KI (Cassens-Röhrig 2019), weil sie regelbasierte, nichtlernende Algorithmen verwenden, um **Optimierungsprobleme** auf Ebene der Gesamtbedarfs- und Schichtplanung zu lösen. Z. B. generieren Konstruktionsheuristiken, genetische oder evolutionäre Algorithmen Lösungsvorschläge für das Erreichen von Zielen unter Einhaltung von Nebenbedingungen (Kutschka und Herbers 2016). In den Regeln können – ob gewollt oder nicht – Diskriminierungen einprogrammiert sein. Wenn beispielsweise gewerkschaftlich organisierte Mitarbeiter Zuschläge für Nacht- und Sonntagsarbeit erhalten (Nichtorganisierte aber nicht), wird ein genetischer Algorithmus mit dem Ziel der Kostenminimierung die günstigeren Nichtorganisierten systematisch den unbeliebten Nacht- und Sonntagsschichten zuordnen und so diese Gruppe diskriminieren (Strohmeier 2019, S. 6). Der Komplexitätsgrad der Einsatzplanung hängt deshalb vor allem davon ab, für wie viele unterschiedliche Mitarbeitergruppen geplant wird. Beim sogenannten **Contingent Workforce Management** werden neben den klassischen Vollzeitarbeitnehmern auch solche mit atypischen Beschäftigungsverhältnissen berücksichtigt (Teilzeit, Berater, Freelancer, Leiharbeitnehmer, etc.). Welche Mitarbeiter zum Einsatz kommen, wirkt sich insbesondere auf die Personalkosten aus.

Beispiel: Personaleinsatzplanung mit Vote2Work
(<https://www.vote2work.com>)

Vote2Work unterstützt Kunden beim Aufbau einer flexiblen Personaleinsatzplanung, an der Mitarbeiter teilhaben. Mit ihrem Software-as-a-Service-Angebot können Arbeitgeber – unter Einbeziehung des Betriebsrats und der Mitarbeiter – Antworten auf vier Fragen finden: **Wer kann? Wer darf? Wer soll? Wer möchte?** Auf Basis eines mit dem Betriebsrat abgestimmten Regelwerks werden automatisch die für einen Einsatz am besten geeigneten Mitarbeiter anhand eines Scoring-Modells vorgeschlagen, das u. a. Qualifikationen (Können), vorgeschriebene Ruhepausen (Dürfen), Zeitkontenstände (Sollen) und erwünschte Arbeitszeiten (Wollen) beinhaltet. Mitarbeiter können z. B. Zeitfenster generell blocken (wenn z. B. Kinder von der Schule abgeholt werden müssen) oder Anfragen sogar ablehnen.

Vote2Work, Gewinner des HR Start-up Awards 2019, verspricht Vorteile für Arbeitgeber, Arbeitnehmer und den Betriebsrat. Erstere können schnell und einfach die Einsätze optimal planen. Zweitere verbessern ihre Work-Life-Balance und erfahren eine höhere Wertschätzung, weil sie (in Grenzen) die Einsatzplanung mitbestimmen dürfen. Letztere sehen, was an Überstunden bzw. Mehrarbeit ansteht und können ohne großen Aufwand über die Vorschläge entscheiden. Insgesamt sollen so die Kosten für Mehrarbeit reduziert werden, da über die intelligente Prioritätensteuerung nur Mitarbeiter mit optimalem Zeitkonto vorgeschlagen werden.

Während Vote2Work sich vor allem auf das produzierende Gewerbe fokussiert, sind die Softwarepakete von Can Do, Project Pad und iCombine auf die Projektarbeit ausgerichtet. Die Anbieter versprechen eine **skillbasierte Personaleinsatzplanung für Projekte**. Can Do ermöglicht es beispielsweise Kompetenzen und Kapazitäten projekt-, abteilungs- und standortübergreifend zu vergleichen und verarbeitet bei der Ressourcenplanung auch unscharfe Zeit- und Mengenangaben (z. B. 50–80 Personentage bis Ende KW 38), wofür der Algorithmus – ein regelbasiertes Expertensystem – Echtzeitinformationen über Projekteigenschaften wie Arbeitspakete und Projektphasen berücksichtigt. Bei iCombine können (Personal-) Manager alle Mitarbeiter nach Fähigkeiten und Verfügbarkeiten filtern und sie zu Projekten auf eine bestimmte Rolle einladen. Feedback- und Auswertungsmöglichkeiten über den Planungsstatus, gebuchte Stunden, Mitarbeiter und Rollen sollen bessere Personaleinsatzentscheidungen ermöglichen, weil Überbelastungen und Unterauslastungen sichtbar und messbar werden. Da es beim internen Personaleinsatz genau wie beim Recruiting darum geht, die richtigen Mitarbeiter für eine Stelle zu finden, sind ähnliche Algorithmen zum Parsen, Matchen und Empfehlen im Einsatz (siehe Abschn. 4.2.2).

Mit dem Algorithmus, den Kleinberg und Raghu (2018) auf Basis submodulärer Funktionen entwickelt haben, lässt sich das Potenzial von Mitarbeitern testen, um auf Basis dieser Ergebnisse **Teams optimal zusammenzustellen**, zumindest was deren Leistung angeht. Interessant ist das deshalb, weil Testergebnisse von Einzelpersonen normalerweise keine guten Vorhersagen für die Leistung des Teams, in dem sie eingesetzt werden, erlauben. In diesem Fall konnte der Test die spätere Teamleistung jedoch gut vorhersagen, was wichtig für die Teamzusammensetzung in der Praxis ist, denn bisher weiß man meist erst hinterher, ob eine Personenkonstellation auch als Team gute Leistungen erzielt.

Ein ganz anderes System ist I-Competence, das unterschiedliche KNN und ML-Algorithmen vergleicht, wie gut sie Kompetenzlücken bei Managern von Softwareentwicklungsteams identifizieren und geeignete Maßnahmen für die bessere

Teamzusammenstellung vorschlagen können (Colomo-Palacios et al. 2014). Die Vorhersagegenauigkeit liegt bei 93,23 %, was Projekt- und Personalmanagern zeigen soll, wie gut I-Competence funktioniert. Eine Nutzung dieses von Wissenschaftlern beschriebenen Systems in der Praxis ist mir nicht bekannt.

In Bezug auf die Arbeitsleistung konnten Gombolay et al. (2015) zeigen, dass Algorithmen zur Planungsoptimierung die **Zusammenarbeit zwischen Mensch und Roboter** effizienter machen und zwar vor allem dann, wenn die Maschine die Aufgabenverteilung vorgibt und der menschliche Mitarbeiter entsprechend handelt.

4.3.2.3 Was man noch machen kann

Eine Herausforderung ist die **Verknüpfung der operativen Personaleinsatzplanung mit der strategischen** Planung, die zwar immer wieder gefordert wird (Furkel 2018, S. 35 f.), aber Probleme aufwirft, weil die Planungsgrößen sich unterscheiden: Während die strategische Planung meist auf Ebene von Jobfamilien geschieht, gibt es bei der operativen einen Trend zur Individualisierung, d. h. die Bedürfnisse jedes Einzelnen sollen berücksichtigt werden. Das beginnt bei den Qualifikationen, geht über die persönlichen Lebensumstände (z. B. Kinder, pflegebedürftige Angehörige, Anfahrtsweg zur Arbeit) bis hin zum persönlichen Biorythmus (Zimmermann und Romeiser 2019, S. 8).

Insbesondere in Branchen, in denen Fachkräfte knapp sind, soll eine individualisierte und flexible Einsatzplanung ein **Element des Employer Branding und der Personalbindung** sein, weil sie einen Anreiz für die Wahl des Arbeitgebers bietet. Schon seit Jahrzehnten gibt es flexible Formen des Arbeitens in Bezug auf Zeit und Ort (z. B. Gleitzeit, Vertrauensarbeitszeit, Telearbeit, Homeoffice). In den letzten Jahren sind unter dem Schlagwort „New Work“ zwei weitere Aspekte hinzugekommen: die inhaltliche Arbeitsaufteilung (Jobsharing) und das soziale Mit einander. Mit wem gearbeitet wird, ist gerade in wissensintensiven Jobs entscheidend, weil oft interdependente Aufgaben in Teams bearbeitet werden. Nicht nur bei der Wahl des Arbeitgebers, sondern auch bei der **Wahl eines Projekts** können soziale Dimensionen relevant werden, z. B. Alter, Geschlecht, Familienstatus, Herkunft, etc. Anbieter wie Tandem employ (siehe Beispiel-Kasten) offerieren Software-Lösungen, die Menschen im Unternehmen in Teams (oder Tandems) zusammenbringen, und zwar so, dass sie sich in den jeweiligen Projekten oder Entwicklungsphasen gegenseitig ergänzen und unterstützen können.

Beispiel: Matching mit Tandem employ (<https://www.tandememploy.com/de>)

Die Tandem employ Software verbindet Mitarbeiter innerhalb von Unternehmen bei verschiedenen Anlässen und Zielsetzungen: von Coffee Dates und (Reverse) Mentoring über Jobrotation bis hin zur Zusammenarbeit in Projekten oder dem

Teilen von Jobs (Jobsharing). Manche nennen es deshalb das Business-Tinder: Ziel ist nicht die Partnersuche fürs Leben, sondern für die Arbeit. Und dafür muss es auch nicht gleich fürs Leben sein. Eher sollen neue und flexible Formen der (Zusammen-)Arbeit geschaffen, Wissenstransfer ermöglicht und Silostrukturen geöffnet werden, um das Unternehmen innovativer und wettbewerbsfähiger zu machen – und die Arbeitgeberattraktivität zu steigern. Wer sich beispielsweise einen Job teilen möchte (wie die beiden Gründerinnen von Tandememploy es tun) oder eine erfahrene Kollegin als Mentorin finden möchte, dem hilft die Software die passende Person zu finden.

Das Zusammenbringen funktioniert im Prinzip wie die Matching-Algorithmen beim Recruiting (siehe Abschn. 4.2.2). Die Mitarbeiter müssen sich zunächst auf der Tandememploy-Plattform registrieren und ein – anonymes – Profil mit ihren Fähigkeiten und (Lern-)Interessen erstellen (z. B. welche Kompetenzen sie im Rahmen einer Jobrotation erlernen möchten). Der Tandememploy-Matching-Algorithmus schlägt daraufhin automatisch passende Kollegen oder Angebote vor und die entsprechenden Mitarbeiter können sich untereinander vernetzen. Die Personalabteilung muss nicht involviert werden.

Ein weiterer Trend ist die **Verknüpfung von (Echtzeit-)Daten** über die Tätigkeit (z. B. Ort, Dauer, Schwierigkeitsgrad, Mehrwert, Erledigungsstatus) mit Eigenschaften des Personals (Können, Wollen), dem Beschäftigungssystem (Dürfen und Sollen [z. B. Arbeitszeitgesetz, Regelungen zu Telearbeit, Urlaub und Vertretungen]) und Echtzeitdaten über interne sowie externe Ereignisse (Produktionsstillstand, Wetter, Staus, Sportereignisse, Werbeaktionen und Reaktionen darauf in den sozialen Medien). Letzteres ist noch eher selten verbreitet, aber Anbieter wie Meteolytix, GFOSS oder Quinyx bedienen den Handel, die Gastronomie oder die Baubranche z. B. mit Echtzeitdaten über Wetter und Sportereignisse.

4.3.2.4 Was (noch) nicht geht

Am ehesten gelingt die Verknüpfung mit Kostenfaktoren, Gesetzen und Betriebsvereinbarung, weil diese relativ stabil und einfach in Regeln zu übersetzen sind. Haben Firmen allerdings sehr viele Regelungen zu Arbeitszeit, -menge und -ort (z. B. unterschiedliche Beschäftigungsverhältnisse, kurzfristig einzuplanende Schichtwechsel, Gleitzeiten, Telearbeit) und wollen nicht nur schichtbezogene Einsatzplanung, sondern eine stunden- oder uhrzeitbezogene, verkompliziert das die Planung. Muss dann noch situativ reagiert werden, weil beispielsweise einem Kunden eine ad hoc-Zusage am Telefon gemacht werden muss, erhöht sich die Wahrscheinlichkeit von **unauflösbarer Interessengegensätzen**: Die Arbeitgeber

wollen etwas anderes als Arbeitnehmer und selbst unter Kollegen kann es zu Konflikten kommen, wenn entschieden werden muss, wer wann und wo zum Einsatz kommen soll.

Außerdem sind Daten über Kosten, Kompetenzen, Motivation und Aufgaben meist in unterschiedlichen Systemen gespeichert (z. B. Produktionsplanung, Skill-Datenbank, Talent Management, Ticketing-System), was eine **integrative Analyse erschwert**. Manche Daten sind auch überhaupt nicht vorhanden. Beispielsweise ist unklar, wie der Schwierigkeitsgrad und der Mehrwert einer Aufgabe objektiv abgebildet werden kann. Ein Weg wäre es, die Aufgaben aus der Unternehmensstrategie und dem Geschäftsmodell abzuleiten, um Tätigkeiten als mehr oder weniger (strategisch) wertvoll zu kennzeichnen. Gleichzeitig müsste der Aufgabenkatalog anpassbar sein, sodass beispielsweise in volatilen Umwelten auch Mitarbeiter neue strategierelevante Aufgaben definieren können.

Zukünftig sollen **Chatbots** eingesetzt werden, d. h. Nutzer können dann in natürlicher Sprache mit der Personaleinsatzplanungssoftware interagieren. Schichtänderungswünsche, Krankmeldungen oder Urlaubsanträge müssen dann nicht in ein Formular eingetippt werden, sondern die Anfrage wird per sprachgesteuertem Assistenten übertragen und vollautomatisch erledigt (Furkel 2018, S. 35; Cassens-Röhrlig 2019, S. 77). Noch ist das Zukunftsmusik, weil die entsprechenden Technologien (KNN/NLP, RPA) noch nicht in die (regelbasierten) Systeme integriert sind – aber es gibt sie, weshalb sprachgesteuerte Einsatzplanung im Bereich des Möglichen liegt.

Wenn Maschinen über den Personaleinsatz entscheiden, kann es zu **Widerständen und Demotivation** seitens der Belegschaft kommen. Fühlen sich die Mitarbeiter in ihrer Handlungsfreiheit (Autonomie) eingeschränkt und kontrolliert, so mindert dies ihre intrinsische Motivation (Arnaud und Chandon 2013). Dies ist zwar keine technologische Grenze, aber durch die Technologie induziert.

4.4 Performance Management

Performance Management, das heißt die Bewertung und zielgerichtete Steuerung von Leistung, ist ein **zentrales Element wirtschaftlichen Handelns** in einer marktlich organisierten Gesellschaft. Organisationen, die Wettbewerbsvorteile generieren und langfristig überleben wollen, werden die Leistung ihrer Belegschaft bewerten und Anreize setzen, damit die Beiträge im Sinne der Unternehmung geleistet werden. Dies ist eine grundlegende Annahme der Anreiz-Beitragstheorie und die Idee hat ihre empirische Entsprechung im Laufe der Wirtschaftsgeschichte immer wieder gefunden (March und Simon 1958; Nienhüser 2018). Aus Unter-

nehmensperspektive ist die zentrale Frage, ob für die Kosten, die die Anreize verursachen, ein mindestens genau so hoher Nutzen gezogen werden kann. In der Praxis wird diese Frage **meist nur mit Blick auf die Leistung des einzelnen Mitarbeiter*innen** beantwortet und dafür werden Daten gesammelt, die in einem Zusammenhang mit guter oder schlechter Leistung stehen, z. B. Feedbacks bzw. Bewertungen (Abschn. 4.4.1). Dann werden leistungsschwache Mitarbeiter (Low Performer) und leistungsstarke (High Performer) identifiziert. Die rechtliche Grundlage dieser Einteilung ergibt sich aus der vertraglich geschuldeten Sollleistung in qualitativer oder quantitativer Hinsicht, wobei Minderleistungen sowohl verhaltens-, personen- als auch betriebsbedingte Gründe haben können (Kraus 2018, S. 701). Welcher Grund im konkreten Fall tatsächlich vorliegt, ist aus einer Steuerungs- und Optimierungsperspektive heraus natürlich hoch relevant – und Workplace Analytics können hier Wirkungszusammenhänge aufdecken (Abschn. 4.4.2). Neben der rechtlichen Grundlage kann die Etablierung von Leistungsbeurteilungssystemen auch aus einer anderen Perspektive heraus begründet werden: Da **Leistung ohnehin bewertet** werden wird, gebietet es sich aus Gründen der Fairness, Transparenz und Objektivität ein Regelwerk zu etablieren, damit es nicht zu verdeckten Beurteilungen kommt, sondern Abweichungen oder Regelverstöße sichtbar und einklagbar werden (Goler et al. 2016).

Performance Management umfasst mehrere Teillbereiche: Angefangen von der Definition der zu messenden Aktivitäten, Prozesse oder Strukturen (inklusive geeigneter Messgrößen), über die Abweichungsanalyse bis hin zum Feedback- & Steuerungsprozess, der auch zu einer Veränderung der Ziele, Aktivitäten, Prozesse oder Strukturen führen kann (siehe Kleindienst 2017, S. 33 ff.). Die Ausprägungen dieser **Performance-Management-Elemente haben sich im Verlauf der Jahre gewandelt** (siehe Tab. 4.1). Heutzutage wird mehr Wert auf eine hohe Feedback-Frequenz, offene Bewertungen und die Einbindung von Kollegen als Beurteilende gelegt als noch vor ein paar Jahren (Ledford et al. 2016, S. 255). Im Folgenden werden die Aspekte der Leistungsbeurteilung und -messung sowie Inputquellen im Fokus stehen.

4.4.1 Ziel- und Feedbacksysteme

4.4.1.1 Worum es geht

Ziele bilden seit Jahrzehnten eine wichtige Basis für die Leistungsbeurteilung. Genau wie bei Performance Management im Allgemeinen, geht es auch bei der Einführung von Zielsystemen um mehr als nur einen Zweck. Beim klassischen **Management by Objectives** (MbO) geht es vor allem um eine koordinierte

Tab. 4.1 Performance Management im Wandel

	Traditionelles PM	Modernes PM	Postmodernes PM
Zeitliche Einordnung Aufreten	1950–heute	1995–heute	2010–heute
Häufigkeit der Beurteilung	Jährlich	Jährlich oder halbjährlich	Monatlich oder noch häufiger
Basis der Leistungsbeurteilung	Zielerreichung, Eigenschaften	Kaskadierte Ziele, Kompetenzen	Kaskadierte Ziele, Kompetenzen
Zielobjekt der Leistungsbeurteilung	Individuum (Mitarbeitende)	Meistens Individuen, teilweise Teams	Meistens Individuen (Mitarbeitende, Externe), vermehrt Teams
Beurteilungsmetrik	Komplexes Rating (mehrere Dimensionen und Kennzahlen)	Einfaches Rating	Ohne Rating (keine numerische Bewertung)
Input von Kollegen und anderen	Nicht vorhanden	Über 360°-Feedback	Peer-to-peer und crowd-sourced Feedback
Methode der Vergütungsbestimmung	Vorgesetzte bestimmen Vergütung auf Basis des Ratings und/oder vorgegebener Verteilungen	Kalibrierungsrunde und Formeln	Unterschiedlich (Kritik an Formeln, Ratings und Kalibrierungsrunden)
Leistungs- vs. Entwicklungsfokus	Beides	Beides	Betonung der (zukünftigen) Entwicklung

Veränderung des Performance Managements (PM) im Laufe der Zeit (in Anlehnung an Ledford et al. 2016, S. 254)

Steuerung der vielen Aktivitäten in einem Unternehmen. Um dies zu erreichen, sollen Teilziele aus den übergeordneten Unternehmenszielen abgeleitet und von den Managern bzw. Mitarbeitenden realisiert werden (Drucker 2005 [ursprünglich 1954]). Ein Nebeneffekt ist laut Drucker (2005, S. 149 f.), dass sich die Organisationsmitglieder anhand der Rückmeldung über den Grad der Zielerreichung selbst kontrollieren: **Management durch Beherrschung soll durch Management durch Selbstkontrolle** ersetzt werden. Allerdings macht Drucker (2005, S. 147) auch klar, dass die jeweils übergeordneten Stellen das Recht haben, die Zielsetzungen zu genehmigen oder abzulehnen. Selbstkontrolle meint daher nicht Selbstbestimmung.

Es geht auch darum, dass die digitale Ermöglichung permanenten Feedbacks als **Instrument der Überwachung** gelten kann bzw. als Verwirklichung dessen, was Ortmann (1984, S. 142) schon vor 35 Jahren als „elektronisches Panopticon“ betrieblicher Personalinformationssysteme beschrieben hat: als Technik zur Überwachung aller Arbeitsbereiche, weil mit ihr alles eingesehen werden kann, ohne dass sie selbst einsehbar wäre. In diesem Zusammenhang ist die (In-)Transparenz von Algorithmen genauso zu diskutieren wie die Überlegung, dass zwar alle (rück-)fragen und (rück-)melden können, was andere getan haben, damit aber nicht gesagt ist, dass auch wirklich alle alles sehen und alles sagen (Neuberger 2000) (siehe Abschn. 4.4.1.4).

In den letzten Jahren wurde viel über die Ausgestaltung von Ziel- und Feedbacksystemen diskutiert (z. B. Cappelli und Tavis 2016; Deloitte 2017; Ledford et al. 2016; Levy et al. 2017). Ein Dauerbrenner ist die Frage, wie aus (top-down) Zielvorgaben tatsächliche **Zielvereinbarungen** (auf Augenhöhe) werden, die sowohl die Bedürfnisse der Mitarbeitenden als auch die Anforderungen der Unternehmen berücksichtigen. Theoretisch sind sich viele Wissenschaftler und Praktiker einig, dass Vereinbarungen besser zum aktuellen Zeitgeist der partizipativeren Unternehmensführung passen. Praktisch sind Top-down-Vorgaben dennoch dominierend, wenn auch manchmal durch den Begriff „Ziele kaskadieren“ kaschiert (Ledford et al. 2016, S. 256; Levy et al. 2017, S. 158; Ehl und Schumann 2019, S. 54). Zudem sollen die Rückmeldungen über den Zielerreichungsgrad nicht nur ein- oder zweimal im Jahr, sondern öfter oder gar permanent erfolgen, um schneller gegensteuern zu können. Dadurch sollen Ziel- und Feedbacksysteme eine erhöhte **Agilität der Organisation bei gleichzeitiger Wertschätzung** der Mitarbeiter ermöglichen (Cappelli und Tavis 2016, S. 61; Werther 2018, S. 28). Erreicht werden soll dies z. B. durch zeitnahe regelmäßiges Feedback, die vermehrte Messung von entwicklungsorientierten Faktoren (z. B. dem Leistungspotenzial statt der aktuellen Leistung) und flexiblere Zielsysteme, z. B. der OKR-Methodik (Objectives & Key Results).

Die Diskussion über notwendige Anpassungen der Ziel- und Feedbacksysteme wird letztlich durch **drei Veränderungen in der Arbeitswelt** angetrieben (Gärtner 2016, S. 14 f.; Cappelli und Tavis 2016, S. 63 ff.):

1. In dynamischen Umwelten sind unterjährige Zielanpassungen nicht die Ausnahme, sondern die Regel, weshalb eine höhere Frequenz der Zielsetzung und Beurteilung gefordert wird.
2. Mit komplexer werdenden Tätigkeiten hat sich das Zurechnungsproblem verschärft: Ob das Ergebnis stimmt oder die Ziele unter- bzw. übererfüllt wurden, kann immer weniger auf den einzelnen Mitarbeiter zurückgerechnet werden, sondern ist eine Teamleistung.

3. Die Überlegung, dass Leistungsbeurteilungen besser ohne jeglichen Hintergedanken an Geld stattfinden sollte, führt zu einer Entkopplung von Beurteilung und Vergütung. Damit senden Firmen die implizite Botschaft „Wir vertrauen den Mitarbeitenden, dass sie auch ohne einen zusätzlichen monetären Anreiz ihre beste Leistung abliefern werden“.

In der Praxis zeigt sich zudem noch ein anderes Phänomen: Die Leistungsbeurteilungen sind häufig zu gut(-willig) und die Personalabteilungen beschweren sich, dass die Führungskräfte entweder keine schlechte Stimmung aufkommen lassen wollen oder nicht fähig sind, klare Differenzierungen und mitunter harte Bewertungen vorzunehmen (Cappelli und Tavis 2016, S. 66). Dann werden die Leistungsbewertung und Vergütungsverteilung schnell als unfair erlebt. Dass ein Performance-Management-System also nur so gut wie seine praktische Nutzung ist, wird später noch einmal thematisiert (Abschn. 4.4.1.4).

4.4.1.2 Wie es geht

Die einfachste Variante, das Performance Management durch digitale Tools zu unterstützen, ist, dass Ziel- und Feedbacksysteme **lediglich inhaltlich angepasst** werden und diese neuen Inhalte von bestehenden Tools abgebildet werden. Dafür gibt es eine nahezu unüberschaubare Menge an Tools, angefangen bei den umfassenden Lösungen der großen Softwareanbieter (z. B. Oracle, Saba, SAP, Workday) bis hin zu kleineren (und teilweise auf Feedback spezialisierten) Anbietern (z. B. 15Five, BambooHR, Culture Amp, engagedly, HR Instruments, Impraise, Primalogik, PurelyHR, Reflektive). Klassischerweise standen Führungsfunktionen (z. B. Delegation, Information), Arbeitsergebnisse (z. B. Umsatz, Anteil positiver Qualitätstests) und Persönlichkeitsmerkmale (z. B. Offenheit, Gewissenhaftigkeit) im Zentrum der Bewertungen bzw. Rückmeldungen. Diese individuell und abteilungsintern ausgerichteten Aspekte werden (zumindest teilweise) durch **team- und kundenbasierte Kriterien** (z. B. Projektabschlüsse, Weiterempfehlungsrate) ergänzt und manchmal auch ersetzt (Cappelli und Tavis 2016, S. 64; Gärtner 2016, S. 15; Ledford et al. 2016, S. 258). Beim Echtzeit-Feedback (Instant Feedback) per App sind meist Kategorien vorgegeben, auf die dann nur noch geklickt werden muss: Neben einem einfachen „Danke“ gibt es beispielsweise auch „Super Job“, „Spitzenidee“, „Rettung in letzter Sekunde“, „Kunden zum König gemacht“ oder „Tolle Zusammenarbeit“. Außerdem spielt neben tatsächlich erbrachter Leistung auch vermehrt das **Leistungspotenzial** – abgebildet über (Meta-)Kompetenzen oder Haltungen (z. B. Ergebnis- oder Kundenorientierung, Unternehmertum) – eine Rolle (Cappelli und Tavis 2016, S. 63 f.).

Neben der inhaltlichen Anpassung verändert sich bei modernen Performance-Management-Systemen auch die **Methodik: zusätzlich zur Bewertung durch die**

Vorgesetzte, bewerten auch Kollegen. Die Begründung für den vermehrten Einsatz von horizontalem Peer-to-Peer-Foodback liegt in den heutzutage stärker vernetzten Arbeitsprozessen, bei denen die Leistung bzw. das Ergebnis eher einem Team als einem Einzelnen zuzuschreiben ist (Ledford et al. 2016, S. 258, siehe auch Busch und von der Oelsnitz 2016). Die Annahme ist, dass Mitarbeitende in ähnlichen Rollen die Leistung ihrer direkten Kollegen besser einschätzen können als der Vorgesetzte, weil der seltener die tagtägliche Arbeit beobachten und die (komplexe) Leistung inhaltlich nicht beurteilen kann. Es wird allerdings von manchen Forschern bezweifelt, dass Kollegen (nicht nur formal) qualifiziert sind ein rein vernunft- und faktenbasiertes, nicht subjektiv gefärbtes Urteil abzugeben (Staab und Geschke 2019, S. 16). Letztlich muss das Dilemma entschieden werden, ob man eher auf die verzerrten Urteile einer Vorgesetzten oder auf die der Kollegen setzt. Dabei ist zu bedenken: Mehr Personen sehen nicht besser, sondern vor allem anderes (Neuberger 2000, S. 19).

Ein wesentliches Merkmal aktueller Ziel- und Feedbacksysteme ist, dass die Anpassung der Ziele und Schlüsselergebnisse an geänderte Anforderungen mit einer **höheren Frequenz** als früher – bis hin zum Instant Feedback – erfolgt. Die zurzeit wohl populärste Methode ist **Objectives & Key Results (OKR)**, die viele Unternehmen als Ersatz oder Ergänzung des klassischen Ziel- und Feedbacksystems sehen (z. B. bei Unitymedia; Ehl und Schumann 2019). Wie viele aktuell populäre Management-Methoden kommt der OKR-Ansatz aus dem Silicon Valley und unterscheidet sich vom klassischen MbO in einigen Details (Doerr 2018, S. 33 ff.): Meist werden 3–4 Ziele (Objectives) mit jeweils 2–5 Schlüsselergebnissen (Key Results) pro Quartal vereinbart und in monatlichen oder quartalsweisen Reviews auf ihren Erreichungsgrad hin überprüft. Dabei sollen die Ziele aus strategischen Zielen abgeleitet, abteilungsübergreifend (bei der Arbeit in cross-functional Teams) und sehr ambitioniert sein. Die Ergebnisse beschreiben konkret das Erreichte auf dem Weg zur Zielerreichung. Sie werden zur Messung und Bewertung herangezogen („measure what matters“), allerdings typischerweise ohne Kopplung an Entlohnungsbestandteile. Sowohl die höhere Frequenz als auch die fast schon unrealistische Erreichbarkeit sind bei OKRs gewollt, während dies im MbO-Ansatz eher nicht der Fall ist. Ein weiterer Unterschied zu MbO ist, dass OKRs zu ca. 50 % vom Mitarbeiter und nicht der Führungskraft kommen und öffentlich einsehbar sein sollen. Die OKR-Methode muss nicht, aber sie kann mit digitalen Tools unterstützt werden, von denen Workpath nur eines von vielen ist (siehe Beispiel-Kasten). Ein wesentlicher Vorteil der digitalen Umsetzung ist, dass Übersichten und Auswertungen einfacher erstellt werden können, sodass eine **höhere Transparenz** über Ziele, Ergebnisse und Tätigkeiten zur Erfüllung dieser sowohl innerhalb als auch zwischen Teams entsteht.

Beispiel: OKRs mit Workpath (<https://www.workpath.com/>)

Workpath ermöglicht die Abbildung von OKRs, Check-ins und Feedback in einer eigenen Applikation. Diese lässt sich in gängige HR-Softwaresysteme von SAP, Oracle, Workday, Personio oder BambooHR integrieren und sich mit Kommunikationstools wie Outlook, Slack und GMail verbinden.

OKRs können in der App beschrieben, einem Owner zugeordnet und der Fortschritt im Vergleich zum Start und Ziel angezeigt werden. Zudem gibt es ein Feld für Kommentare und Updates. Die Check-in Meetings sind strukturierte Teambesprechungen, in denen die Teammitglieder, die an gemeinsamen Zielen arbeiten, sich in Bezug auf ihre operativen Aufgaben und strategischen Prioritäten abstimmen. Im OKR-Zyklus finden sie wöchentlich oder zweiwöchentlich zwischen Kick-off und Review statt. In diesen Meetings reflektieren die Teams, woran sie in der letzten Woche gearbeitet haben und was sie in der nächsten Phase tun wollen. Über die App sind die Themen für die Diskussion bereits vorstrukturiert: Fokus der vergangenen Woche, Fokus der kommenden Woche sowie Gelerntes, Risiken oder Blockaden. Die Agenda ist damit vorgegeben und die Aussagen der Teammitglieder werden direkt in der App dokumentiert und können von dort aus in andere Systeme exportiert werden. Sowohl durch die Transparenz über OKRs als auch die Check-ins sollen Abhängigkeiten in und zwischen Teams frühzeitig erkannt und koordiniert werden können. Dazu soll auch die Feedback-Funktion beitragen. Einerseits ist vorgesehen, dass sich Kollegen untereinander Feedback in Form von „recognition“ (Sternchen, Likes) geben, und andererseits können die Top-Fähigkeiten der Mitarbeiter angezeigt werden.

Workpath verspricht, dass eine gute digitale Abbildung der OKR-Methode Hierarchiedenken und Silos überwindet, ohne dass Chaos entsteht, sondern Netzwerk und Autonomie. Dass dies allein durch eine App geschieht, ist zu bezweifeln. Wenn schon kein Netzwerk durch die App entsteht, so lässt es sich doch immerhin visualisieren. Durch die digitale Erfassung kann schnell ein relativ hohes Maß an Transparenz über Verknüpfungen zwischen OKRs einerseits und Personen andererseits hergestellt werden.

Weder klassische noch aktuell populäre Methoden wie OKR oder 4DX (4 Disciplines of Execution; McCchesney et al. 2012) beinhalten notwendigerweise ML-Algorithmen oder einen automatisierten Entscheidungsprozess. Allerdings zeigt das Eingangsbeispiel des Leistungsbewertungssystems bei Amazon (Kap. 2), dass eine solche Kopplung möglich ist und beim (Nicht-)Erreichen von Zielen oder Ergebnissen Konsequenzen automatisch erfolgen. Positiv gewendet, können auf Basis der Zielerreichung bzw. des (stagnierenden) Fortschritts **Empfehlungen mit Hilfe von Klassifikationsalgorithmen** ausgesprochen werden, z. B. hinsichtlich

individueller Entwicklungspläne oder Lernformate (siehe dazu Abschn. 4.5.1). Liegen digital protokollierte Reviews vor, können mit Hilfe von **Text-Mining und Cluster-Algorithmen Analysen zu Blockaden, Risiken oder negativen Stimmungsausserungen** durchgeführt werden. Nach der Auswertung dieser qualitativen Äußerungen wären automatisierte Rückmeldungen an die Teams in mehr oder weniger Echtzeit denkbar. Noch befinden sich diese Anwendungsszenarien aber in der Konzeptphase und **in der Praxis dominieren deskriptive Analysen**, die aber keine Wirkungszusammenhänge aufdecken können (Werther und Woschée 2018, S. 234). Es gibt keine großzahlig empirische Evidenz, dass prädiktive Analysen in der Praxis angewendet werden und funktionieren. Dass regelmäßiges Feedback grundsätzlich einen signifikanten Einfluss auf die Arbeitsleistung hat, konnte durch wissenschaftliche Studien immer wieder gezeigt werden. Beispielsweise wirkt sich Feedback positiv auf die Leistung von Bankangestellten aus, gemessen über das Wachstum der Spareinlagen (Bartel 2004) oder die wahrgenommene Servicequalität (Liao et al. 2009). Nur liegen ähnliche Ergebnisse für den Einsatz von digitalen Ziel- und Feedbacksystemen (noch) nicht vor.

Während Befürworter in der erhöhten Frequenz und Transparenz sowie der vermehrten Einbindung von Kollegen als Beurteilenden einen Schritt in Richtung Netzwerkorganisation, intrinsische Motivation durch Visualisierung des Big Pictures und dem Sinn von Arbeit sehen, verstehen Kritiker digitale Tools zum Performance Management vor allem als **Instrumente zur Überwachung und Leistungssteigerung** durch Aufbau von Druck und Stress. Ziel dieser Kritik sind Anwendungen, die E-Mails, Tastenanschläge, Mausbewegungen, Webseitenaufrufe, Interaktions- und Kommunikationsströme und weitere Verhaltensweisen (siehe dazu Abschn. 4.4.2) messen und als Leistungsindikatoren auswerten. Ein kontrovers diskutiertes Beispiel ist das Leistungsbewertungs- und Feedbacksystem von Zalando (siehe Beispiel-Box).

Beispiel: Zonar – das Leistungsbewertungs- und Feedbacksystem von Zalando

Seit 2016 ist beim Onlineversandhändler Zalando ein Software-System namens Zonar im Einsatz, das die Leistungsbewertung von ca. 5000 Mitarbeitenden durch gegenseitiges Feedback mittels einer App ermöglicht (Staab und Geschke 2019, S. 11; Zalando 2019a). In Zonar fließen die Daten aus den halbjährlichen Mitarbeiterbeurteilungen und Echtzeitbewertungen ein. So können die Mitarbeitenden mit der App strukturiertes und unstrukturiertes Feedback unmittelbar nach Meetings, Workshops, Projektabschlüssen, Kampagnen, etc. geben. Das heißt, nach einem solchen Ereignis können die Mitarbeitenden Bewertungen von ihren Führungskräften, Kollegen oder auch internen Kunden anfordern oder

solche unaufgefordert erhalten (Ewenstein et al. 2016, S. 67). Die Bewertung besteht aus einem quantifizierenden Rating, das durch eine qualitative Rezension begründet wird. Ebenso hat der/die Feedbackgebende die Möglichkeit, Notizen, Sprachnachrichten oder Dokumente an das Feedback anzuhängen. Zudem gehen in Zonar Daten aus der Produktion in den Algorithmus ein, z. B. die Anzahl an abgeschlossenen Produktbearbeitungen im Online-Shop (Ewenstein et al. 2016, S. 67 f.; Staab und Geschke 2019, S. 15).

Aktuell nominieren Mitarbeitende zweimal jährlich bis zu acht Kollegen, die sie beurteilen, wobei Führungskräfte die Auswahl mitbestimmen. Der Zonar-Algorithmus teilt die Mitarbeitenden dann aufgrund der erhaltenen Bewertungen in drei Gruppen ein: Low, Good und Top Performer, je nachdem ob sie zu den 10 % mit den schlechtesten bzw. besten Bewertungen gehören oder dazwischen liegen (Ewenstein et al. 2016, S. 67 f.). An diese halbjährliche Gesamtbewertung sind Konsequenzen geknüpft. So sollen sich alle Mitarbeiter mit ihren persönlichen Entwicklungszügen, Stärken und Schwächen auseinander setzen, und Top Performer qualifizieren sich als Potenzialkandidaten für eine Karriere im Unternehmen (Staab und Geschke 2019, S. 19; Zalando 2019a). Der Entwicklungsgedanke scheint dem Beurteilungsimpetus untergeordnet: Im Kern geht es bei Zonar um hochfrequente horizontale Ratings, die automatisiert auf Basis eines Algorithmus zu individuellen Scores und Gruppenbildungen verarbeitet werden (Staab und Geschke 2019, S. 8, 55).

Ein großer Kritikpunkt ist die Intransparenz des Algorithmus, sodass (auch den Mitarbeitenden) nicht klar ist, auf Basis welcher Kriterien und Regeln die individuelle Eingruppierung geschieht (Staab und Geschke 2019, S. 44). Der Algorithmus scheint so konzipiert zu sein, dass Schwächen im quantitativen Rating höchstens so günstig ausfallen können, wie die am schlechtesten bewertete Stärke, d. h. sobald nur drei von fünf Sternen im Rating der Stärken vergeben wurden, stehen nur noch die drei ungünstigeren Auswahlmöglichkeiten im Rating der Schwächen zur Verfügung (Staab und Geschke 2019, S. 31). Außerdem werden Werte im mittleren Skalenbereich als Good Performer (=100 %) definiert und Werte darüber sind sehr außergewöhnlich und so gut wie nicht erreichbar (Staab und Geschke 2019, S. 30). Solche systemseitig vorgegebenen Verzerrungen beinträchtigen die Validität eines Bewertungssystems – und führen zum Vorwurf, dass Zonar ein Instrument der Kostensenkung durch systematische Lohnrepression sei, weil es darauf angelegt wurde, möglichst geringe Gehälter zu zahlen (Staab und Geschke 2019, S. 57). Ungeachtet der Unwissenheit über die genaue Funktionsweise des Algorithmus, haben sich zumindest manche Mitarbeiter Verhaltensweisen überlegt, die sie besserstellen sollen. So geben sie an, grundsätzlich etwas bessere Bewertungen zu vergeben und Ra-

tings von Kollegen einzuwerben, von denen sie eine gute Bewertung erwarten (Staab und Geschke 2019, S. 31)

Zalando (2019b) hat einigen Kritikpunkten widersprochen (insbesondere jenen zur Illegalität und fehlenden Einbindung des Betriebsrats) und weist auf die mangelnde Grundgesamtheit der Studie von Staab und Geschke hin: Diese basiert auf Firmendokumenten und zehn Interviews, was angesichts der 5000 an Zonar teilnehmenden Mitarbeitenden tatsächlich wenig erscheint (0,2 %). Allerdings können auch aus wenigen qualitativen Daten wichtige Themen identifiziert werden (Reichertz 2016, S. 130 ff.). Zudem findet sich keine Klarstellung zur Kritik, wie genau der Algorithmus funktioniert.

4.4.1.3 Was man noch machen kann

Verfolgt man den offiziellen Ansatz von Zalando, so sind die Bewertungen der **Ausgangspunkt für Lern- und Entwicklungsmöglichkeiten**, weshalb das Performance Management hier einen Übergangspunkt zur Personalentwicklung hat. Je nach Stärke können Mitarbeiter zum Coach oder Mentor in diesem Bereich werden – und für die identifizierten Schwächen werden ihnen eLearnings, Trainings, Coaching, etc. angeboten. Meist wird nur auf die Schwächen rekurriert (siehe Abschn. 4.5).

Zur Identifikation von Stärken und Schwächen lassen sich **externe Quellen automatisiert einbinden**. In Business-Netzwerken, wie z. B. LinkedIn, können sich die Teilnehmer gegenseitig Fähigkeiten und Qualifikationen bestätigen. Entweder lassen sich diese Daten crawlern oder sie können über integrierte Schnittstellen von Performance-Management-Softwarelösungen (z. B. BambooHR, WebHR) für die Validierung von Kompetenzen herangezogen werden. Dabei ist die Aktualität der gegenseitigen Bewertungen und Referenzen wahrscheinlich größer als ihre Validität, weil von vielen Gefälligkeitsvoten auszugehen ist und Kompetenzen ja nur bestätigt und nicht aberkannt oder negativ eingeschätzt werden können (Kern und Haep 2016, S. 14). Interessant sind in diesem Kontext die Übernahmen von LinkedIn, GitHub, Lynda und Glint durch Microsoft. Damit hat Microsoft sowohl Zugriff auf Daten über Qualifikationen, Kompetenzen, Stärken und Schwächen von Mitarbeitern aus ihren Online-Profilen (LinkedIn, GitHub), ihren Feedbacks (Glint) und ihrem tagtäglichen Arbeitsverhalten (Office-Produkte) als auch die Möglichkeit, ihnen Weiterbildungssangebote über die Lernplattform Lynda anzubieten.

Eine weitere Form der Nutzung der Feedbacks sind **Auswertungen hinsichtlich möglicher Kündigungen oder Missstände rund um den Arbeitsplatz**. Bewertungen in Form von strukturierten Daten (Punkte, Sternchen oder andere ordinal skalierte Daten) können über deskriptive Analysen oder (multiple) Regressionen ausgewertet werden (siehe Abschn. 3.4 und 3.6). Freitext-Feed-

backs können mit Text-Mining-Verfahren analysiert werden (z. B. Sentiment-Analysen), um kritische Themen zu identifizieren, was z. B. B.near und IBM anbieten (Schweyer 2018, S. 9; siehe Abschn. 4.2.1.3). Die Auswertung von Ziel- und Feedbacksystemen kann auch bei Veränderungsprojekten hilfreich sein. Solche **Change Analytics** können nicht nur Feedbackprozesse durch Instant bzw. Echtzeit-Feedbacks beschleunigen, sondern es lassen sich Miss- oder Widerstände direkt erkennen und Lernangebote sowie Change-Botschaften personalisieren (Stenius 2017, S. 60 ff.). Für Führungskräfte können solche Daten über die individuelle und teambasierte (Un-)Zufriedenheit in einem „Employee Experience Dashboard“ aufbereitet werden. So lassen sich Determinanten und Effekte der Arbeitszufriedenheit, z. B. Führungsverhalten, schnell visuell erfassen (siehe Beispiel-Box).

Beispiel: Visualisierung von Treibern mit dem Impact-Grid von functionHR (<https://www.functionhr.de>)

Das Start-up functionHR bietet mit seinem „HR Keyboard“ eine Reihe kleinerer Software-Lösungen („Module“) entlang der gesamten HR Wertkette an. Die einzelnen Module ermöglichen es Anwendern, Daten zunächst zu erheben und zu analysieren, um z. B. einen Überblick über die aktuelle Personalstruktur zu erhalten oder die wichtigsten Treiber bzw. Einflussfaktoren für die Arbeitsleistung und -zufriedenheit zu identifizieren. Die dafür benötigten Daten werden über Mitarbeiterumfragen speziell erhoben, sodass gezielt geprüft werden kann, ob theoretisch mögliche Treiber auch tatsächlich vorliegen. Zudem werden die Ergebnisse grafisch aufbereitet, sodass sie einfacher interpretiert werden können. Mit Maßnahmen, die auf die wichtigsten Faktoren einwirken, lassen sich dann Veränderungen prognostizieren. Sollte sich beispielsweise auf Basis einer multiplen Regressionsanalyse der Daten herausstellen, dass „gute Führung“ ein wichtiger Treiber für Arbeitszufriedenheit ist, lässt sich dieser Zusammenhang – inklusive der Merkmale, was „gute Führung“ ausmacht und Auswirkungen einer Veränderung – in einem „Impact-Grid“ visualisieren (siehe Abb. 4.5).

Neben klassischen multivariaten Regressionsanalysen kommen auch komplexere Verfahren wie Ridge-Regressionen (eine Variante von Regressionsanalysen) und KNN zum Einsatz. Da über die Mitarbeiterumfragen nicht nur numerische Feedbacks eingeholt werden, sondern in Freitextfeldern auch Kommentare gegeben werden, kommen bei der Auswertung auch NLP-Verfahren zum Einsatz. Beispielsweise können Themen, die in den Kommentaren immer wieder genannt werden, über Topic Models (z. B. Latent Dirichlet Allocation; siehe Abschn. 4.2.2.2) erfasst werden. In der Praxis hat sich gezeigt,

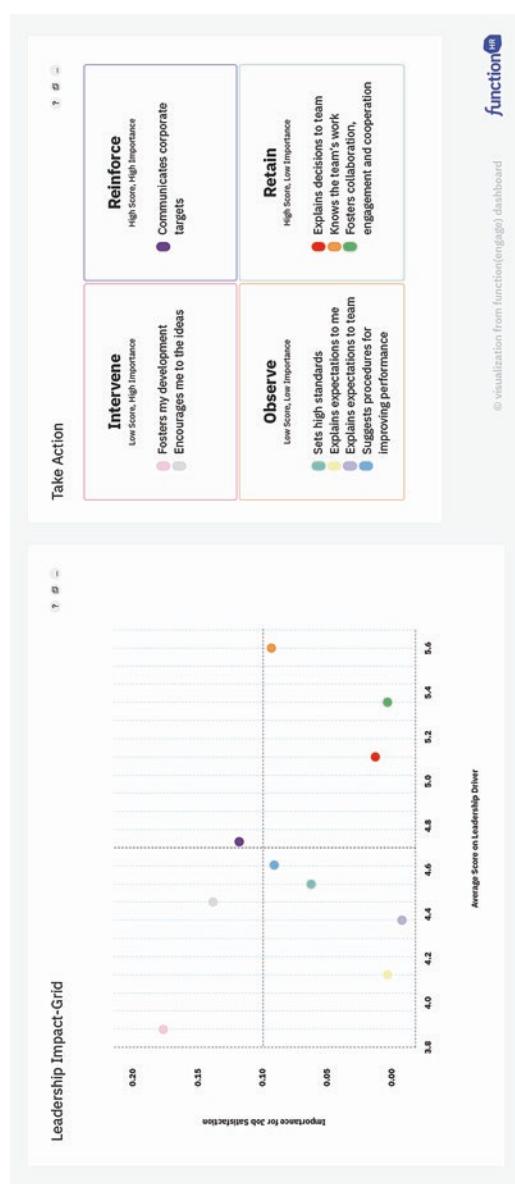


Abb. 4.5 Treiberanalyse („Impact-Grid“) von functionHR

dass schon mit klassischen Methoden gute Ergebnisse bei einer Reihe von Fragestellungen erreicht werden können, z. B., warum Mitarbeitende (un-)zufrieden sind, das Unternehmen verlassen, die Führungskraft schlecht bewerten, von den Vergütungskomponenten nicht motiviert werden oder kein Wissen im Team teilen. Auf Basis solcher Analysen konnten beispielsweise bei FlixBus unterschiedliche Maßnahmen eingeführt werden, die innerhalb von sechs Monaten den Anteil loyaler Mitarbeiter um 50 % erhöhten.

Je nachdem, welche Themen in den Feedbacks genannt bzw. abgefragt werden, können Organisationen auch untersuchen, **welche Vergütungsbestandteile die größte Wirkung entfalten** (Schweyer 2018, S. 8 f.). Gerade für große Unternehmen, in denen die Gesamtvergütung oft aus sehr vielen Elementen und Zusatzleistungen besteht (z. B. Grundgehalt, Bonus, Firmenwagen, Firmenkindergarten, betriebliche Altersvorsorge), ist es wichtig zu wissen, welche davon den größten Nutzen für welche Mitarbeiter(gruppen) und das Unternehmen stiften. Zur Berechnung können bereits multivariate Regressionsmodelle ausreichen. Falls kein linearer Zusammenhang zwischen Vergütungsbestandteilen und Motivation oder Leistung angenommen werden soll (was durchaus sinnvoll ist), bieten sich nicht-lineare Klassifizierungsalgorithmen wie Entscheidungsbäume oder KNN an, die Muster zwischen der Motivation oder Leistung und den damit in Verbindung stehenden Vergütungsbestandteilen identifizieren können (siehe Abschn. 3.4).

Der Vergütung liegt in großen Unternehmen oftmals eine **Stellenbewertung (Grading)** zugrunde, mit der die Anforderungen einer Stelle anhand definierter Kriterien bewertet werden, z. B. welche Breite und Tiefe an Fach- und Führungskompetenzen für erfolgreiches Arbeiten auf der Stelle gefordert sind oder wie groß der Einfluss auf das Geschäftsergebnis ist. Im Unterschied zur Leistungsbeurteilung wird bisher nicht der Output, sondern der erwartete Input gemessen, um Stellen intern in eine Rangordnung zu bringen und die Vergütung anforderungs- und marktadäquat festzulegen (Skenes und Kleiner 2003). Am Markt gibt es dafür einige digitale Tools von etablierten Anbietern wie Mercer, Willis Towers Watson oder der Hay Group, aber auch Start-ups wie gradar und Content. Beispielsweise hat Content ein Modell entwickelt, das alle Aufgaben sowie die damit einhergehenden Kapazitäten und Kosten in einer relationalen Datenbank abbildet. Der Vorteil im Vergleich zu den Bewertungsverfahren von Mercer und Co. ist, dass **Änderungen von Stellenzuschnitten einfacher abgebildet** werden können, weil die kleinsten Einheiten einer Stelle – die erbrachten Aufgaben und Services – identifiziert und im System hinterlegt sind. Diese Flexibilität ist in einer dynamischen Arbeitswelt wichtig. Noch einen Schritt weiter geht die **Verbindung zwischen Input (Grading) und Output (Leistungsbeurteilung)**. Im einfachsten Fall gehen die Informationen über

die Anforderungen einer Stelle in sogenannte **Kalibrierungsrunden** ein und werden dort zur Festlegung von Zielerreichung und Bonus genutzt, wobei die Leistung durch Menschen beurteilt wird. Bei Zalando ist dies beispielsweise so: Im Zonar-System sind die Stellenstruktur und die Entwicklungspfade hinterlegt, um die Erwartungen und Anforderungen an einen Job, Beförderungsbedingungen und Entwicklungsstufen zu dokumentieren, und die Kalibrierung erfolgt im „People Review Committee“ (Staab und Geschke 2019, S. 17, 20). Eine weitergehende Idee ist, die Aufgaben nicht nur mit Kosten und Kapazitäten, sondern auch der erbrachten Leistung in einem IT-System zu verknüpfen. Letztlich wird dies mit Zonar durch die Peer-to-Peer-Leistungsbeurteilungen teilweise realisiert – mit all den Problemen und der Kritik, die ein solcher Ansatz mit sich bringt.

Um Mitarbeiter zusätzlich für die Erreichung ihrer Ziele zu motivieren, setzen manche Anbieter auf die **Integration von Gamification-Elementen** (Levels, Punkte- bzw. Sternebewertungen, Bestenlisten, Badges), also Mechaniken, die in Spielen angewendet werden, um die (intrinsische) Motivation der Spieler hoch zu halten. Solche Funktionen bieten beispielsweise Betterworks und engagedly: Mitarbeitende können Punkte sammeln, die dann für Gutscheinkarten oder andere Belohnungen eingelöst werden können, oder sie können einen Level aufsteigen, wenn sie Ziele erreicht haben. Die Sichtbarkeit des Erreichten innerhalb der Organisation ist ein oft genutzter Hebel, um immaterielle Anreize wie Anerkennung und Statusgewinn anzubieten. Letztlich sollen Mitarbeitende durch Gamification intrinsisch motiviert und spielerisch zur Selbstoptimierung animiert werden (Reeves und Read 2009, S. 73, 224). Der Hype um den Einsatz von Gamification am Arbeitsplatz ist mittlerweile allerdings einer kritisch-realistischeren Betrachtung gewichen (Vesa und Harviainen 2019).

4.4.1.4 Was (noch) nicht geht

Permanentes (Rück-)Melden kann als Totalüberwachung wahrgenommen werden. Diese Wahrnehmung können Mitarbeitende bei allen Performance-Management-Methoden haben, egal ob es um Zielsysteme, Instant-, Peer2Peer- oder 360°-Feedbacks geht. Durch die aktuellen technologischen Möglichkeiten bekommen diese Methoden allerdings eine neue Qualität: Die Daten werden mehr, vielfältiger, kommen zeitnäher und aus unterschiedlicheren Quellen als dies ohne digitale Tools möglich wäre. Das allein dadurch erzeugte Unbehagen steigt weiter, falls die Auswertungsmechanik und Entscheidungslogiken nicht nachvollziehbar sind. Die Kritik an solchen IT-basierten Personalmanagementsystemen ist nicht neu. Ortmann (1984) hat ausführlich erläutert, inwiefern die automatisierte Verarbeitung von Informationen über das Personal dem gleich kommt, was als Bentham's **Panopticon** bekannt ist: Die Architektur eines Gefängnisses, bei dem in der

Mitte ein Überwachungsturm steht und ringförmig um diesen Turm die Zellen angeordnet sind. Aus dem Turm heraus sind somit alle Insassen einsehbar, während der Turm uneinsehbar bleibt (bei entsprechendem Lichteinfall bzw. speziellen Turmfenstern). Die Insassen können nicht sehen und damit auch nicht wissen, ob sie überwacht werden – aber sie sollten davon ausgehen. Ob sie es wirklich tun, ob sich alle der disziplinierenden Gewalt fügen, die von einer prinzipiell alles sehenden und selbst uneinsehbaren Technik ausgehen kann, **hängt jedoch von einigen Faktoren ab** (siehe Ortmann 1984, S. 142 ff.):

1. Technologie bietet immer nur Möglichkeiten, die noch **von Menschen genutzt** werden müssen. Es ist unrealistisch, dass tatsächlich alle alles sehen und alles sagen, weil es für alle Nutzer gute Gründe gibt, Beobachtungen und Beurteilungen nicht unzensiert weiterzugeben (Neuberger 2000, S. 19). Wie eine Technologie genutzt wird, liegt zudem an den etablierten Normen und Werten der jeweiligen Unternehmenskultur (z. B. Fairness, Offenheit, Aggressivität, etc.). Insofern bekommt jede Organisation mit ihrem Performance-Management-System das Maß an Überwachung, Fairness und Entwicklungsorientierung, das sie verdient.
2. Von einer Totalüberwachung ist nicht auszugehen. Einerseits, weil die technologische Architektur eines elektronischen Panopticons nicht an die Perfektion von Bentham's Idee(!) heranreicht, und andererseits Unternehmen offene Sozialsysteme sind, sodass es **Rückzugsorte, Lücken oder gar die Möglichkeit zur Manipulation** des Systems durch die Überwachten gibt. Da Algorithmen meist nur teilweise und nicht völlig intransparent sind, entstehen genau dadurch Lücken, die von den Überwachten ausgenutzt werden können (siehe die Einwerbung wohlgesonnener Feedbacks als Umgangsweise mit Zonar, die wortwörtlich als Umgehen der intendierten Logik gesehen werden kann).
3. Die Intransparenz führt nicht nur zu Angst und einem wahrgenommenen Kontrollverlust über die Beurteilung. **Intransparenz legitimiert auch den Widerstand** gegen ein solches System, weil auf Datenschutzgesetze und weithin akzeptierte Werte wie Transparenz, Fairness, Beteiligung bei Entscheidungen, etc. zurückgegriffen werden kann (siehe die öffentliche Empörung, die Zonar ausgelöst hat, bis hin zur Überprüfung durch die Berliner Datenschutzbehörde; Zeit-Online 2019).

In **manchen Berufen und manchen Unternehmen scheint die Totalüberwachung bereits umgesetzt** zu sein. Das Eingangsbeispiel über die Lagermitarbeiter bei Amazon ist ein Beispiel dafür (siehe Kap. 2). Auch außerhalb der Lagerhallen setzt **Amazon** auf Leistungskontrolle, die eine subtile Form der Machtausübung ist, da keine Führungskraft mehr persönliche Anweisungen gibt oder gar herum-

schreit und droht – das übernimmt eine App, verpackt in motivierende Mitteilungen. Seufert (2019, S. 77), ein Journalist der ZEIT, berichtet von seiner Tätigkeit als Paketzusteller für Amazon und der antreibenden Wirkung von digitalen Zielvorgaben, Lob und Aufmunterungen einerseits sowie physisch erlebten Ängsten andererseits:

„Ich bekomme motivierende Mails, Wochenzusammenfassungen, etwa: 32 erfolgreiche Zustellungen, davon 31 pünktlich, 100 Prozent erfolgreiche Zustellungen für den gesamten Zeitraum, vielen Dank, dass Sie mit Amazon Flex Lächeln zustellen! Als ich länger keine Schicht bekomme, schickt Amazon mir Tipps: Aktualisieren Sie häufig, seien Sie flexibel, seien Sie schnell! Jede dieser Motivations-mails triggert aber auch Ängste. Was, wenn der Wert sinkt? Wie oft kann man zu spät kommen, bevor man gesperrt wird? Diese Unsicherheit treibt mich an“.

Andere Logistikdienstleister wie **UPS** haben ähnliche Systeme (Ajunwa et al. 2017, S. 743 f.). Zudem sind die Überwachungsmöglichkeiten nicht auf produktionsnahe Jobs beschränkt: Aus **Call-Centern** ist bekannt, dass dort die Tastaturschläge und die Dauer der Bearbeitung eines Falles oder auch nur bestimmter Felder einer Softwaremaske gemessen und zur Bewertung verwendet werden. Und Büroangestellte hinterlassen mit ihren Schlüsselkarten bzw. Mitarbeiterausweisen Daten über ihre Anwesenheit und physischen Bewegungsmuster, während ihre Browserverläufe Auskunft über ihre digitalen Bewegungen geben (Ajunwa et al. 2017, S. 742 f.).

Doch selbst bei diesen eindringlichen Beispielen sind **Überwachung und Leistungskontrolle nicht grenzenlos**. Mitarbeitende finden immer wieder Möglichkeiten mit den Regeln des Systems gegen das System zu spielen. So ist beispielsweise bekannt, dass Paketboten viele Pakete bei Nachbarn, in Abholshops oder anderswo abstellen, um Zeit zu sparen. Bei UPS führte dies z. B. dazu, dass zwar mehr Pakete ausgeliefert wurden, darunter aber viele Fehlzustellungen waren (Gärtner und Cläßen 2016, S. 39). Solche Umgangsmöglichkeiten (im doppelten Wortsinne) gibt es immer, sobald es Regeln gibt – nicht, weil diese Lücken lassen (das kommt natürlich auch vor), sondern, weil sie *etwas* regeln und die Akteure mit den Regeln gegen die Intention der Organisation vorgehen können. Einen Schritt weiter gehen Sabotageakte, wie sie Seufert (2019, S. 77) berichtet: im Namen von Amazon pampig zu den Kunden sein, ein Päckchen fallen lassen – unabsichtlich natürlich – oder den Paketinhalt einfach mitzunehmen und dafür gefeuert zu werden.

Die Zusicherung anonymer Beurteilungen – alle können alles sagen, ohne erkannt zu werden – kann unerwünschte Folgen nach sich ziehen, die aus sozialen Medien bekannt sind: **öffentliche Diffamierungen aus der Anonymität heraus** fallen leichter als Kritik, die von Angesicht zu Angesicht geäußert wird. In Bezug

auf Zalandos Zonar-System sprechen Staab und Geschke (2019, S. 34) von einer „Verantwortungsdiffusion“, die durch die Beschaffenheit der Technologie erzeugt wird. Ob diese Folgewirkungen tatsächlich entstehen, hängt wesentlich von den ersten beiden oben genannten Faktoren ab.

Unabhängig von der technischen Möglichkeit zur Totalüberwachung, gibt es **Datenschutzgesetze** und insbesondere in Deutschland die **betriebliche Mitbestimmung**. Dass eine Datensammlung, die rund um die Uhr und ohne konkrete Zwecke erfolgt, zumindest nicht mit deutschen Gesetzen vereinbar ist, hat das Bundesarbeitsgericht im April 2017 entschieden (NZA 2017). Selbst bei anonymen Stichpunkt-erhebungen und Einhaltung hoher Datenschutzstandards können Betriebs- und Personalräte der Einführung einer technischen Einrichtung zur Verhaltens- und Leistungsüberwachung – und als solche sind Ziel- und Feedbacksysteme zu werten – widersprechen. Dies ist der Fall, wenn **die freie Entfaltung der Persönlichkeit** der Mitarbeitenden beeinträchtigt wird oder andere besondere Gefahren von der Überwachung ausgehen (§ 75 Abs. 2 und § 87 Abs. 1 Nr. 6 BetrVG). Eine umfangreiche Überwachung kann durch eine schriftliche, freiwillige Einwilligung der Mitarbeitenden in die konkrete Datenerhebung legitimiert sein. Aufgrund des bestehenden Machtungleichgewichts zugunsten des Arbeitgebers wird eine freie Entscheidung des Arbeitnehmers als nahezu unmöglich betrachtet. Zudem sind der Einwilligung solche Formen der Datenverarbeitung entzogen, die die Arbeitnehmenden zum bloßen Objekt machen, wie dies insbesondere beim Erstellen von Persönlichkeitsprofilen der Fall ist (Kraus 2018, S. 705). In diesem Zusammenhang ist eine entscheidende Frage, ob durch das Performance-Management-System eine betriebliche Ungleichheit etabliert und legitimiert wird, die die freie Entfaltung der Persönlichkeit beeinträchtigt. Kritiker argumentieren, dass Beurteilungen über Scoring-Werte die Belegschaft anhand von Zahlen in Gruppen einteilen (z. B. Top-Performer und Low-Performer) und aufgrund dieser Festlegung gruppenspezifische Privilegien oder Sanktionen zugewiesen werden (Staab und Geschke 2019, S. 556). Ob dem so ist, muss im Einzelfall geprüft werden. Unabhängig davon ist die herrschende Meinung, dass **Zielvereinbarungen nicht für die Einteilung in High und Low-Performer ausreichen**, da es sich lediglich um einen abstrakten Berechnungsfaktor handelt, aus dem nicht auf die geschuldeten Normalleistung rückgeschlossen werden kann. Für eine vergleichende Einteilung bedarf es einer Gruppe an Mitarbeitenden, deren Tätigkeit hinsichtlich Art, Menge, Umfang und Qualität vergleichbar ist. Und für eine Kündigung muss zusätzlich eine länger anhaltende, signifikante Leistungsdifferenz vorliegen, d. h. ein Leistungsunterschied von ca. einem Drittel (Kraus 2018, S. 702).

Ein datenbezogenes Hindernis für den Einsatz von ML bzw. ADM im Performance Management sind die **geringen Fallzahlen, was die Mustererkennung erschwert**. Es liegen meist zu wenig vergleichbare Ziele und Ratings vor, da die

Ziele von Mitarbeitenden auf unterschiedlichen Stellen sich normalerweise auch unterscheiden. Wahrscheinlich dominieren in der Praxis auch deshalb deskriptive Analysen. Wie oben bereits erwähnt, gibt es keine großzahlig empirische Evidenz, dass Predictive Analytics zur Anwendung kommen. Wenn mehrere hundert oder gar mehrere tausend Ereignisse vorliegen müssen, um ML-Algorithmen anwenden zu können (Ratner 2017, S. 7 ff.), dann dürfen und können aus individuellen Zielvereinbarungen keine Rückschlüsse gezogen werden. Wer mit Hilfe von Instant Feedbacks, Pulse Checks oder anderen anlassbezogenen Befragungen statistische Treiberanalysen vornehmen möchte, muss auf die statistische Vergleichbarkeit der Daten achten und darf diese nicht durch wechselnde Themen oder Skalenformate verhindern (Werther und Woschée 2018, S. 234). Allerdings liegt das Problem der **geringen Fallzahl nicht bei jeder Fragestellung** vor. Es schwächt sich bereits ab, wenn es um die Frage der Wirkung von Entgeltbestandteilen geht, da hier erwartungsgemäß größere Stichproben vorliegen. Noch kleiner ist es, wenn Feedbacks ausgewertet werden sollen, weil hier einerseits viele Texte generiert werden und es andererseits auch lexikonbasierte Text-Mining-Verfahren gibt, die nicht mit Trainingsdatensätzen angelernt werden müssen und daher auch keine großen Fallzahlen benötigen (siehe Abschn. 4.2.1.2).

In der Praxis ist immer wieder eine gewisse **Befragungsmüdigkeit und Überforderung bei häufigen Zielanpassungen** zu beobachten. Insbesondere die OKR-Methode geht mit einem hohen Abstimmungsaufwand einher, wenn Führungskräfte monatlich Ziele und Erreichungsgrade mit allen Mitarbeitenden klären müssen und die Ziele auch noch zwischen Teams oder Abteilungen harmonisiert werden sollen. Auch wer wöchentlich oder noch häufiger aufgefordert ist Feedback zu geben, kann dies als Stressfaktor empfinden. Zudem wird die Erwartung geweckt, dass sich aufgrund des Feedbacks etwas ändert – und da sich in den seltens-ten Fällen wöchentlich Strukturen und Arbeitsabläufe ändern, wird diese Erwartung wöchentlich enttäuscht. **Ärger, Frust, niedrige Teilnahmekquoten und Widerstand** gegen das Feedback- bzw. Bewertungssystem sind die häufige Folge (Werther und Woschée 2018, S. 235). Diese Reaktionen verstärken sich noch, wenn das System aus Sicht der Mitarbeitenden nicht mit den offiziellen Zweck-setzungen (Agilität, Wertschätzung, Entwicklung) übereinstimmt, wie das Zalando-Beispiel zeigt (Abschn. 4.4.1.2).

Leistungsbewertung und Leistungssteuerung sind nur zwei der **vielen Funktionen von Performance Management**, auch wenn diese meist nicht offen ausgesprochenen werden: Es geht neben der Beseitigung von Fehlern, Leistungs- und Identifikationsbarrieren auch um Steigerung der Motivation der Organisationsmitglieder und der Kundenzufriedenheit, Verbesserung der Arbeitszufriedenheit, Koordination der Aktivitäten, Umsetzung der angestrebten Unternehmenskultur durch Diszipli-

nierung und Sozialisation, Stärkung der Eigen- und Mitverantwortung, Erhöhung der Transparenz, Gelegenheit zum ‚Dampfablassen‘ (das Anprangern von Defiziten erspart deren Beseitigung), Schaffung einer Grundlage für individualisierte Personalentwicklung, u. v. m. (Neuberger 2000, S. 15). Mit dieser Aufzählung wird deutlich: Performance Management ist ein **interessenbeladenes Konfliktfeld**, das **digitale Tools allein nicht befrieden können**.

4.4.2 Performance Management mit Workplace Analytics

4.4.2.1 Worum es geht

Der Begriff Workplace Analytics wird meist synonym zu People oder HR Analytics verwendet (siehe Marler und Boudreau 2017). Im Folgenden sollen allerdings nicht Personendaten (z. B. Persönlichkeitseigenschaften, Kompetenzen, Verweildauer im Job), sondern **Verhaltensdaten** im Vordergrund stehen, vor allem abgeholtene Meetings, E-Mail- und Face-to-Face-Kommunikation und Bewegungsabläufe. Diese Daten sagen etwas über den **Prozess des Zusammenarbeiten** am Arbeitsplatz aus – und durch Korrelation mit Erfolgsmaßen wie abgeschlossenen Verkäufen, erledigten Beschwerden oder verbesserter Teamarbeit lassen sich Erkenntnisse für das Performance Management ziehen. Weil die gesammelten Daten Rückschlüsse über die Interaktionen und Beziehungen am Arbeitsplatz ermöglichen, sprechen manche auch von „**relational analytics**“ (Leonardi und Contractor 2018; Bernstein et al. 2019). Gemeinsam ist den unterschiedlichen Ansätzen, dass die erhobenen Daten auf Interaktionen sowie deren Auswirkungen im Arbeitsalltag zielen und nicht nur auf statische Persönlichkeitseigenschaften oder die Leistung einer einzelnen Person (wie es etwa bei der Messung von Tastaturanschlägen, Webseitenbesuchen oder der Bewertung der Kompetenzen eines Kollegen der Fall ist). Vorausgeschickt sei allerdings, dass **nicht immer der Fokus auf den Beziehungen** zwischen Menschen, also dem sozialen Netzwerk, liegt. Stattdessen werden die Daten doch wieder auf Einzelpersonen zurückgeführt und die Mitarbeitenden in High und Low Performer unterschieden.

4.4.2.2 Wie es geht

Für Workplace Analytics sind aktuell zwei Quellen zentral für die **Datenerhebung: Sociometric Badges und Office365**. Ersteres sind kleine Geräte, die Mitarbeiter z. B. um den Hals tragen können und die mittels GPS, Bluetooth, Infrarotschnittstelle und Sprachaufnahmefunktion aufzeichnen, wer wann wo mit wem wie lange interagiert hat (Fischbach et al. 2010; Olguín-Olguín und Pentland 2010; Pentland 2012). Solche Wearables werden sowohl für die Analyse manuell domi-

nierter Tätigkeiten als auch für Büroarbeit eingesetzt. Die Office-Produkte von Microsoft (Word, Excel, PowerPoint, Outlook, etc.) sind weltweit bekannt. Weniger bekannt ist, dass Microsoft Analyse-Tools anbietet, mit deren Hilfe E-Mails, Kalendereinträge und andere Office365-Daten ausgewertet werden können, um die Zusammenarbeit zu analysieren (Fuller 2014, 2015; Höller und Wedde 2018). Im Folgenden wird zunächst Office365 als Datenquelle diskutiert.

Microsoft Office365

sammelt Daten aus der Nutzung der einzelnen Office-Produkte (Microsoft 2019a, b; Höller und Wedde 2018, S. 11; Leonardi und Contractor 2018, S. 78 f.):

- An welchen Kontakt wurde mit welcher Betreffzeile und in welchem Zeitfenster eine E-Mail geschrieben (Datenerfassung durch Outlook)?
- Wer war mit wem wie lange in Meetings bzw. Videokonferenzen (Outlook, Skype)?
- Wie lange wurde von wem an einem Dokument gearbeitet (PowerPoint, Word, Excel)?
- Mit wem hat man sich vernetzt und gechattet (Teams)?

Aus den Daten werden Vorschläge zur Optimierung des Arbeitsalltags generiert, wobei die zugrunde liegende **Office-Graph-Technologie** nach und nach Verbindungen zwischen Personen, Dokumenten und Interaktionen knüpft und dabei lernt (siehe Abb. 4.6). Eine Funktion ist aus den Streams auf den Startseiten von sozialen Medien bekannt: Auf der Office365-Benutzeroberfläche namens Delve werden dem Nutzer die relevantesten Dokumente angezeigt, wobei die Relevanz von der Graph-Technologie berechnet wird, und zwar anhand der Aktivitäten des Nutzers und seiner Kollegen, der Interaktionen zwischen den Mitarbeitenden, der Bearbeitungen, die Kollegen gemeinsam an Dokumenten vornehmen, der Intensität des E-Mail-Austausches, usw. (Höller und Wedde 2018, S. 27). **Ob ein Dokument angezeigt wird oder nicht, hängt also von den Interaktionen mit dem Dokument und im sozialen Netzwerk ab.**

Weitere **Auswertungsfunktionen** stellen die Microsoft-Dienste **MyAnalytics** und **Workplace Analytics** bereit. MyAnalytics wertet die Intensität des E-Mail-Verkehrs zwischen einer einzelnen Nutzerin und ihren Kollegen aus und zählt zudem, an welchen Sitzungen sie gemeinsam teilgenommen haben. Damit lässt sich bestimmen, mit wem man oft und eng zusammenarbeitet – gemessen über die Zeit, die gemeinsam in Meetings verbracht wurde (Höller und Wedde 2018, S. 28). Zudem erinnert MyAnalytics die Benutzer z. B. daran, sich Zeit zu blocken, bevor man noch eine Meeting-Anfrage akzeptiert und dann im Meeting-Hamsterrad

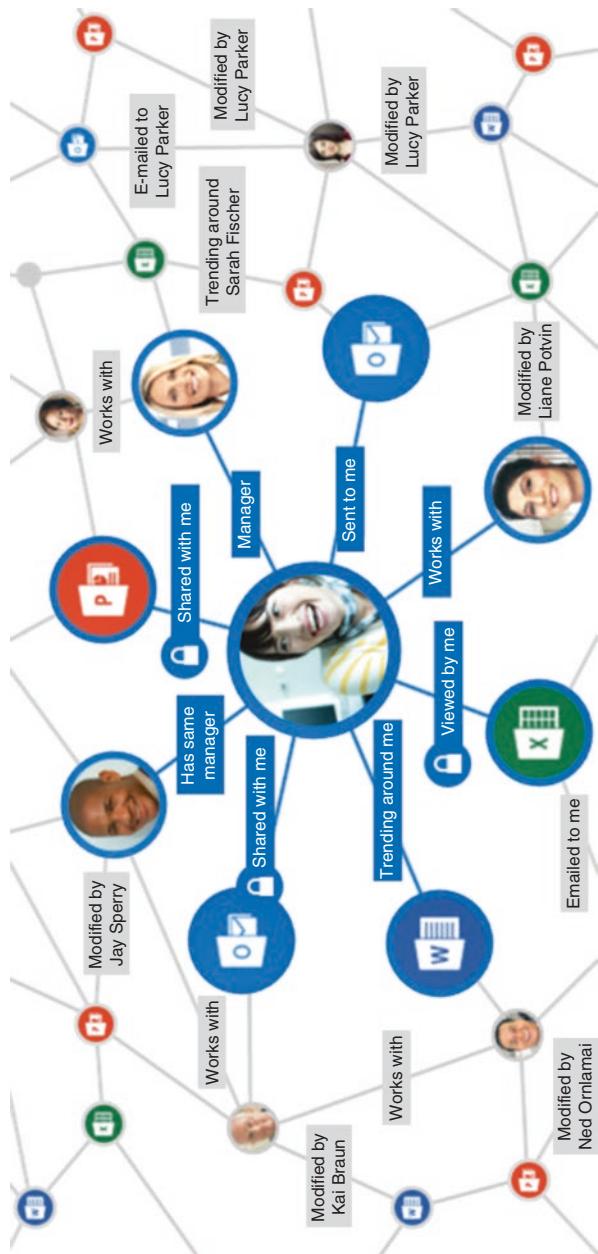


Abb. 4.6 Microsoft Office365 Graph (Microsoft 2019d)

untergeht, oder hebt die Auswirkungen der nach Feierabend geschriebenen E-Mails auf andere hervor (Bernstein et al. 2019, S. 34). Während MyAnalytics Daten über und für Einzelpersonen auswertet, ist Workplace Analytics eine eigene Produktlinie für Unternehmen. Workplace Analytics geht über die Analyse von E-Mail und Meetingdaten hinaus, weil diese Informationen mit Erfolgsmaßen korreliert werden, um herauszufinden, **welche Verhaltensweisen zu hoher Leistung führen** (Fuller 2014, 2015; Nielsen 2016).

Als Beispiel dient eine Studie über mehrere tausend Mitarbeitende im Vertrieb eines Softwareunternehmens (im Folgenden Gärtner 2018b, S. 151 f.). Über einen Zeitraum von sechs Monaten hinweg wurden deren E-Mails und vor allem das Meeting-Verhalten ausgewertet: Wer hatte laut Outlook-Meeting-Einladung mit wem eine Besprechung und wie lange dauerte diese? Aus den Daten lassen sich zum Teil **innovative Kennzahlen** zusammensetzen, z. B. wie viel Zeit eine Vertriebsmitarbeiterin bei (potenziellen) Kunden bzw. mit ihrer Vorgesetzten verbracht hat und wie groß ihr soziales Netzwerk in der eigenen Organisation ist. Es lässt sich auch abschätzen, wie viel Zeit am Stück Mitarbeitende mit konzentrierter Arbeit verbringen konnten, indem die Zeitblöcke ohne Unterbrechung durch Meetings gezählt werden – solche Phasen der Stillarbeit sind insbesondere bei wissensintensiven und kreativen Jobs wichtig (Bernstein et al. 2019, S. 30; Newport 2016). Im Fokus der Studie standen die Zeitspanne, die eine Person mit Interessenten, Kunden oder Kollegen in Meetings verbrachte, die Frequenz der Besuche und die Netzwerkgröße. Die Leistung wurde über die Vertragsabschlussrate gemessen. Durch **lineare Korrelation** der Kommunikations- und Netzwerksdaten mit den Vertragsabschlussraten ergaben sich folgende Ergebnisse für die Top-Performer:

- Top-Performer verbrachten 18 %–33 % mehr Zeit mit Kunden (und haben 40 % weniger Kunden als der Durchschnitt bei Low-Performern)
- Top-Performer hatten ein 30–40 % größeres internes Netzwerk (entspricht 10–20 Personen mehr)
- Top-Performer arbeiteten 2–4 h pro Woche länger
- Top-Performer verbrachten mehr Zeit mit ihren Vorgesetzten bzw. dem Senior Management

Bei den weniger erfolgreichen Vertriebsmitarbeitern ergab sich ein umgekehrtes Bild: sie verbrachten weniger Zeit mit Kunden und Senior Managern und sie hatten ein kleineres Netzwerk. Laut Fuller (2014, S. 2; s. a. 2015, S. 1) sind es genau diese drei Faktoren – Zeit mit Kunden, Größe des internen Netzwerks und Zeit mit dem Management –, die am höchsten mit Vertriebserfolg (Abschlüssen)

korrelieren und zwar unabhängig von der Region (z. B. Asien vs. USA) oder dem Kunden (z. B. Großunternehmen vs. Mittelständler).

Die **Implikationen für das Performance Management** liegen auf der Hand: Jene Vertriebler, die durchschnittlich weniger Abschlüsse vorweisen können, sollten mehr Zeit mit (weniger) Kunden sowie ihren Vorgesetzten verbringen und zugleich auf den Aufbau und die Pflege ihres internen Netzwerks achten. Soweit ist dies kaum verwunderlich und deckt sich weitgehend mit anderen Studien (siehe Khusainova et al. 2018). Die **Stärke von Workplace Analytics ist, konkrete Werte für das Leistungsoptimum** ausrechnen zu können, natürlich nur für die Daten aus dem gegebenen Kontext. In diesem Fall ist es z. B. besser, wenn Vertriebler ihre Zeit für Kunden (z. B. 15 Stunden pro Woche) nicht auf 15 Accounts aufteilen, sondern wenn sie stattdessen jeweils 3 Stunden bei insgesamt 5 Kunden verbringen würden (Fuller 2014, 2015). Darüber hinaus könnten Personaler schon beim **Onboarding auf soziale Vernetzung achten**, indem neue Mitarbeiter aktiv mit Kollegen aus unterschiedlichen Abteilungen in Kontakt kommen.

Eine Auswertungsmöglichkeit, die in der Studie nur ansatzweise genutzt wird, sind **Rückschlüsse über soziale Beziehungen** aus den Daten über das interne Netzwerk. Es wird mit dem Zentralitätsgrad nur eine Maßzahl eingeführt, die den Einfluss einer Person in ihrem Netzwerk abbilden soll (Fuller 2015). Die theoretische Grundlage für diese Interpretation liefert die soziale Netzwerktheorie bzw. die **organisationale Netzwerkanalyse**. Eine solche Analyse repräsentiert Akteure in Organisationen – eine einzelne Person oder eine Gruppe – als Knoten eines Netzes (bzw. Graphens) und die Beziehungen zwischen ihnen als Kanten. Spezielle Software (z. B. InFlow, Kenelyze, NodeXL, UCINet) hilft bei der grafischen Aufbereitung der Charakteristika von Netzwerken (als Graph). Zudem gibt es seit über 50 Jahren robuste Forschungserkenntnisse zu einer Vielzahl an netzwerkbasierten Fragestellungen (Granovetter 1973; Burt 2004; Jansen 2003):

- Wieso machen bestimmte Personen Karriere und andere nicht?
- Wer ist mächtig oder ein informeller Entscheidungsträger?
- Wer beeinflusst wen bei einer Entscheidung?
- In welchen Beziehungsstrukturen entstehen neue Ideen?
- Wer ist eher ein Wissensmakler, wer ein in sich gekehrter Experte?
- Wer trägt zur Vernetzung zwischen Teams, Abteilungen und Organisationen bei?
- Wessen E-Mails oder Kommunikationsbeiträge im Intranet erzeugen keine Reaktion, kümmern also offenbar niemanden?
- Welche Kollegen werden oft um Rat gefragt (per E-Mail oder per Besprechungseinladung)?

Antworten auf diese Fragen geben mathematische Auswertungen zur Netzwerkstruktur und den Beziehungen. Zu den wichtigsten Kennzahlen zur Beschreibung von organisationalen Netzwerken gehören **Zentralitätsmaße und Prestigemaße** zur Messung der Bedeutung und des Einflusses von Akteuren bzw. ihrem Ansehen (Jansen 2003, S. 127 ff.). Beispielsweise berechnet man mit der **Betweenness-Centrality** eines Knotens, wie viel kürzeste Verbindungen zwischen zwei Knoten durch diesen Knoten führen, und erhält damit ein Maß für ihre Bedeutung. Akteure (Knoten) mit hoher Betweenness-Centrality haben sehr viele direkte Kontakte oder liegen oft „auf dem Weg“ oder „zwischen“ zwei anderen Akteuren. Solche Akteure dienen als Mittler, weil über sie Informationen und Wissen auf kurzen Wegen weitergegeben werden können. Sie haben auch die Möglichkeit, Kontrolle auszuüben, weil andere Akteure nur unter hohem Aufwand an ihnen vorbeikommen. Im Extremfall verbindet ein solcher Akteur zwei Subgruppen, die ohne ihn nicht interagieren könnten. So ein Akteur hat starke Kontroll- und Profitmöglichkeiten. Weitere Zentralitätskennzahlen beschreiben, wie nah oder fern sich Akteure sind (**Closeness Centrality**) und mit wie vielen Akteuren ein Akteur in Kontakt steht (**Degree Centrality**). Letzteres steht im Fokus der oben erläuterten Microsoft-Studie.

Ähnlich wie für Zentralität gibt es auch für die Berechnung des Prestiges mehrere Kennzahlen (Jansen 2003, S. 152). Mit dem **Degree-Prestige** wird erfasst, wie oft ein Akteur von anderen für eine Kontaktaufnahme gewählt wird, wie oft also ein Pfeil im Graphen auf ihn zeigt. Die Kennzahl **Rang-Prestige** berücksichtigt zudem, wie wichtig der Akteur ist, der die Wahl ausspricht – sie gibt der Verbindung bzw. Wahl mehr oder weniger Gewicht, weshalb das Prestige eines Akteurs (langfristig) steigt, wenn er von anderen prestigeträchtigen Akteuren gewählt wird (auf dieser Idee basiert auch Googles ursprünglicher Page-Rank-Algorithmus). **Proximity-Prestige** wiederum berücksichtigt die Pfadlänge und gilt als Maß der relativen Erreichbarkeit von Knoten, also Akteuren, Gruppen oder allgemein: Objekten.

Mit diesen Kennzahlen können bereits einige **Aussagen zur Netzwerkstruktur** getroffen **und Optimierungen** abgeleitet werden: Wo sollte es zentrale Positionen (Schlüsselpersonen, Hubs) geben, wo sollten die Kommunikationsströme auf mehrere Akteure verteilt sein (Analyse der Degree Centrality)? Welche Akteure können anderen mit ihrem Wissen und ihrem Einfluss zu besserer Leistung oder einem Karriereschritt verhelfen (Abschätzung über die Betweenness-Centrality; siehe Granovetter 1973)? Wo gibt es strukturelle Löcher, sodass Akteure vom Informationsfluss abgeschnitten sind, und welche Akteure könnten am einfachsten Brücken zu den Abgeschnittenen bauen (Berechnungen über die Closeness und Degree Centrality; siehe Burt 2004)?

Im Unterschied zu früher können heute **soziale Netzwerke digital und als Nebenprodukt der täglichen Arbeit** erfasst werden – z. B. über Office365. Insbe-

sondere im Bereich der Analyse und Optimierung von Teamarbeit ergeben sich dadurch neue Möglichkeiten (Kozlowski et al. 2015; Kozlowski und Chao 2018; Wang und Katsamakas 2019). Beispielsweise haben Wang und Katsamakas (2019) in ihrer Studie von 2621 Softwareentwicklern die Auswirkung einer Reihe von Netzwerkmaßen (z. B. Degree, Betweenness und Closeness Centrality) auf die **Performance von Softwareentwicklern und gesamter Projekte** untersucht. Als Erfolgsmaß wurde unter anderem gezählt, wie oft Entwickler etwas zu den gemeinsamen Programmierungsbibliotheken beitragen. Dabei kann aus den Daten abgeschätzt werden, wer z. B. eine Projektleitungsrolle hat, nämlich Akteure, die an vielen Projekten beteiligt sind (hohe Degree Centrality), aber pro Projekt nur wenig in die Programmierungsbibliotheken einstellen. Anders verhält es sich mit Akteuren, die nur an einem oder wenigen Projekten beteiligt sind und wenig zu den Bibliotheken beitragen: Hier handelt es sich voraussichtlich um wenig produktive Entwickler (Wang und Katsamakas 2019, S. 32). Über die deskriptive Netzwerkanalyse hinaus haben Wang und Katsamakas (2019, S. 40 ff.) weitere Algorithmen aus dem Bereich der Inferenzstatistik bzw. des Machine Learnings angewendet. Z. B. wurden Zeitreihenanalysen verwendet, um die Interaktionen und Performance im Zeitverlauf zu analysieren. Außerdem noch Cluster-Verfahren, um Sub-Gruppen von Softwareentwicklern zu identifizieren. Ein Ergebnis war, dass 73 Entwickler keiner Gruppe angehörten – aus Personalperspektive kann dies Handlungsbedarf bedeuten, wenn sich der Verdacht bestätigt, dass diese vereinzelten Akteure tatsächlich vom Informations- und Wissensfluss abgeschnitten sind. Die Studie ist eine der ersten, die eine **soziale Netzwerkanalyse auf Basis digital erfasster Daten mit Algorithmen aus dem Machine Learning-Bereich verknüpft**, um Vorhersagen über die Einzel- und Teamleistung zu treffen.

Sociometric Badges

sind, wie bereits erwähnt, kleine Geräte, die mittels **verschiedener Sensoren** (GPS, Bluetooth, Infrarotschnittstelle) und **Sprachaufnahmefunktion** aufzeichnen, wer wann wo mit wem wie lange interagiert hat (Fischbach et al. 2010; Pentland 2012). Entwickelt wurden sie ursprünglich am Massachusetts Institute for Technology, aber auch andere Organisationen haben mittlerweile ähnliche Geräte konstruiert, z. B. Hitachi (Yano et al. 2015). Mit Sociometric Badges sollen **Verhaltensdaten am Arbeitsplatz** in verschiedenen Arbeitssituationen anonymisiert erfasst werden, angefangen bei körperlichen Bewegungen (Sitzen, Stehen, Gehen, Rennen), über Sprachmerkmale (leises Sprechen, lautes Schreien, Nachfragen) – ohne Auswertung des tatsächlichen Dialoginhaltes – bis hin zur zwischenmenschlichen Interaktion (gemessen über die Nähe zweier oder mehrerer Badges über eine gewisse Zeit). Mit den erhobenen Daten kann z. B. herausgefunden werden, ob zwei Personen sich während einer gemeinsamen Arbeitssitzung tatsächlich gut verstehen.

den werden, an welchen Orten die Mitarbeitenden besonders oft und intensiv kommunizieren (meist ist das in der Kaffeeküche oder am Wasserspender der Fall), womit Arbeitsumgebungen so gestaltet werden können, dass der informelle Austausch gefördert wird (Fischbach et al. 2010; Pentland 2012).

In der zentralen Studie aus dem Performance-Management-Bereich wurden die Mitarbeitenden im Call Center einer Bank in Prag mit den Badges ausgestattet (Pentland 2012). Die Ausgangsthese ergab sich aus einer Reihe von Vorstudien, u. a. in der Marketingabteilung einer deutschen Bank: 35 % der Unterschiede in der Leistung eines Teams sind allein auf die Anzahl der persönlichen Gespräche zwischen den Teammitgliedern zurückzuführen und die Arbeitsleistung verringert sich bei mehr als ein paar Dutzend Dialogen pro Arbeitsstunde (Pentland 2012, S. 67). Entscheidend ist es also, die **optimale Anzahl an kommunikativen Interaktionen zu bestimmen**, anstatt bloß auf die Relevanz von Kommunikation für die (Team-)Leistung hinzuweisen. Dazu wurden drei Kennzahlen berechnet:

1. Energie der Kommunikation, gemessen an der Art und Anzahl der Interaktionen
2. Engagement, das die Verteilung der Energie unter den Teammitgliedern widerspiegelt
3. Exploration, was die Anzahl an kommunikativen Interaktionen, die Mitglieder außerhalb ihres Teams betreiben, misst

Als Maß für die Leistung bzw. die Produktivität diente die durchschnittliche Zeit, die ein Call-Center Agent braucht, um ein Kundenanliegen zu bearbeiten (Average Handling Time, AHT). Die besten **Prädiktoren für die Produktivität waren die Energie und das Engagement** eines Teams außerhalb formaler Meetings. Zusammengenommen erklärten diese beiden Faktoren ein Drittel der Unterschiede in der Produktivität zwischen den Gruppen. Außerdem suchen leistungsstärkere Teams mehr Kontakt außerhalb ihres Teams (Pentland 2012, S. 61 f.). Auf Basis der Analysen empfahlen die Forscher dem Leiter des Call Centers, den Kaffeepausenplan der Mitarbeiter so zu überarbeiten, dass alle im Team gleichzeitig eine Pause einlegen können, weil dies den Mitarbeitenden mehr Zeit geben würde, sich mit ihren Teamkollegen über Berufliches und Privates auszutauschen. Natürlich mit dem Ziel, dadurch die Leistung zu verbessern, weil sich durch den informellen Austausch soziale Beziehungen etablieren und Tipps zur Lösung von Kundenproblemen einfacher die Runde machen. Für eine solche **Empfehlung braucht es belastbare Daten**, denn in einem Call Center bedeuten gleichzeitige Pausen, dass ein ganzes Team ausfällt und keine Kundenanfragen bearbeitet. Laut Pentland (2012, S. 69) zahlte es sich für den Leiter des Call Centers aus, auf die Daten zu hören: die AHT fiel um 20 % bei den leistungsschwächeren Teams (sie wurden also besser)

und im gesamten Call Center ging sie um 8 % zurück. Auch **individuelle Interventionen** wurden auf Basis der persönlichen Badges von Mitarbeitenden vorgenommen. Wenn etwa einzelne Mitarbeitende einen niedrigen Energie- und Engagement-Wert zeigten, wurden die Daten dahingehend untersucht, ob die Person vielleicht zu leise spricht, sich vor der Gruppe versteckt oder von anderen das Wort abgeschnitten bekommt (was durch Überschneidungen der Sprachdaten zweier Badges zu einem Zeitpunkt festgestellt werden kann) (Pentland 2012, S. 67).

Eine andere Studie basiert auf Daten der tragbaren Sensorgeräte von Hitachi, die ähnliche technische Funktionalitäten bieten wie jene vom MIT. Allerdings ging es nicht um Kommunikationsverhalten, sondern um die Verbindung von Verhaltensdaten mit einem psychologischen Konstrukt: Glück (happiness). Die dahinter liegende Annahme ist, dass **Glück ein wesentlicher Treiber für die Produktivität in wissensintensiven und kreativen Arbeitsumgebungen** ist (Yano et al. 2015; im Folgenden Gärtner 2018b, S. 153 f.). Zunächst wurden 468 Mitarbeiter aus 10 unterschiedlichen Abteilungen mit Sensorgeräten ausgestattet, sodass Daten über ca. 5000 Personentage vorlagen. Parallel wurde ein Fragebogen verteilt, der mit 20 Fragen (u. a. zu Konzentration, Freude, Appetit, Angst, Einsamkeit) bemaß, wie glücklich die Teilnehmer waren. Danach wurde der Zusammenhang zwischen diesem Glücksmass und der physischen Aktivität (z. B. Gehen, Kopfnicken, Tippen) bestimmt. Das Ergebnis: Die Variation der Daten über Glück lässt sich auch in den Daten über die Variation des Aktivitätslevels (aktiv vs. inaktiv) wiederfinden, und zwar als Wahrscheinlichkeit in Inaktivität zu verfallen, die umgekehrt proportional zur Zeitdauer der Aktivität ist (sogenannte „1/T-rule“, wobei T die Dauer der Aktivität ist). Relevant ist deshalb die Fluktuation bzw. Änderung der Aktivitätslevel (sogenannte „1/T-fluctuation“). Die Forscher konnten zeigen, dass die 1/T-Fluktuation stark positiv mit Glück korrelierte und schlussfolgern deshalb, dass **Glück sich quantitativ über Veränderungen in der physischen Aktivität messen lässt** und diese wiederum mittels der tragbaren Sensoren objektiv bestimmt werden kann. Ein Umweg über Selbsteinschätzungen oder andere Faktoren ist nicht mehr nötig.

Anschließend wurden 215 Personen eines Outbound-Call-Centers für 29 Tage mit Sensorgeräten ausgestattet (insgesamt 6 Mrd. Datenpunkte). Nach der Datensammlung wurde der Zusammenhang zwischen der 1/T-Fluktuation der physischen Aktivität und den erfolgreichen Abschlussraten analysiert. Das Ergebnis: an Tagen mit einer überdurchschnittlichen 1/T-Fluktuation waren die Verkaufsabschlüsse um 34 % höher als an Tagen mit einer unterdurchschnittlichen 1/T-Fluktuation. Für die Autoren belegt diese Analyse, was sie als Dreifaltigkeit bezeichnen: **physische Aktivität = Glück = Produktivität**. Eine weitere Implikation ist, dass **produktive Arbeitsumgebungen nicht vollkommen entspannt oder stressfrei sind**, sondern mit einem mittleren Aktivitätsniveau einhergehen.

Das kann als Bestätigung der Flow-Theorie von Csikszentmihalyi interpretiert werden. Ähnliches wird auch in einer aktuellen Studie bestätigt: Zu wenig kommunikative Interaktion ist ebenso produktivitätsmindernd wie zu viel (Bernstein et al. 2019). Entscheidend ist, bemessen zu können, wo das Optimum liegt – und dafür liefern Sociometric Badges genügend Daten.

Ein weiteres Beispiel für den Einsatz von Sensoren sind Wearables in produktionsnahen Bereichen, wie sie von Kinexon, Sigfox, Intranav oder MotionMiners angeboten werden (siehe Beispiel-Kasten). Im Unterschied zu den bisher erläuterten Ansätzen und Arbeitskontexten geht es hier um die Analyse von Intra-Logistik-Prozessen. Der Einsatz der Technologie soll einerseits **Durchlaufzeiten optimieren** und andererseits die **Ergonomie von Arbeitsabläufen verbessern**, um die Gesundheit der Mitarbeitenden zu schützen. Noch expliziter auf das letzte Ziel ausgerichtet sind Anbieter von Motion Capturing Anzügen (z. B. XSense), die winzkelgenaue Ergonomiestudien ermöglichen.

Beispiel: Prozessanalyse und -optimierung mit MotionMiners (<https://www.motionminers.com>)

MotionMiners, ein Spin-off des Fraunhofer-Instituts für Materialfluss und Logistik, offeriert eine Technologie zur automatisierten Prozessanalyse. Hardwareseitig werden mobile Sensoren und Kleinstfunksender (Beacons) verwendet, um die Prozesse aufzunehmen, während sie ablaufen (siehe Abb. 4.7). Softwareseitig werden aus den Daten automatisch einzelne Arbeitsschritte durch ein Deep-Learning-Netzwerk identifiziert und einer weiteren Analyse zugänglich



Abb. 4.7 MotionMiners: Hardware-Ausrüstung für die Prozessmessung

gemacht. Dazu werden die Bewegungen, Körperhaltungen und auch die Arbeitssituation detektiert und den Tätigkeiten im Arbeitsprozess zugeordnet. Diese Fähigkeit, Körperbewegungen aus Prozessdaten abzuleiten, unterscheidet die MotionMiners-Technologie von anderen Anbietern, die sich auf die reine Ortung von Behältern, Transportvehikeln, Maschinen und Werkzeugen fokussieren. Mit der Kombination aus Hard- und Software können Prozessanalysen durchgeführt werden, ohne dass eine Prozessingenieurin mit Klemmbrett und Stoppuhr neben den Mitarbeitenden stehen muss. Im Vergleich zu manuellen Prozessaufnahmen kann der Aufwand erheblich reduziert und den Mitarbeitenden eine höhere Anonymität zugesichert werden, als dies bei direkter Beobachtung oder Videoaufzeichnung der Fall ist.

Beispielsweise wurde der Einlagerungsprozess bei Hugo Boss analysiert und optimiert. Im Fokus standen drei Einlagerwagen und der anfallende Handhabungsaufwand. Dazu wurden Bewegungsdaten einer Mitarbeitergruppe über einen Zeitraum von vier Wochen aufgezeichnet und sowohl hinsichtlich der Prozessdurchlaufzeit (Wegzeiten, Verräumzeiten, Such- und Sortierzeiten, Wartezeiten) als auch bezüglich ergonomischer Aspekte (z. B. wie oft muss der Wagen im Laufe eines Einlagervorgangs durch den Mitarbeiter bewegt werden) ausgewertet. Mit diesen Kennzahlen wurde eine Nutzwertanalyse der Einlagerwagen erstellt und ein präferierter Wagentyp identifiziert. Für den Einlagerprozess selbst konnte der unproduktive Anteil an Such- und Sortierzeiten bemessen werden, was wiederum in Empfehlungen für eine Prozessverbesserung mündete.

Der Einsatz von digitalen Tools verlangt insbesondere im Bereich von Performance Management ein hohes Maß an **Datenschutzvorkehrungen** (z. B. anonyme Datenerhebung und -auswertung, Manipulationssicherheit) und **Transparenz** (welche Daten werden wie erhoben und ausgewertet). **Vollkommene Transparenz ist wenig hilfreich**. Einerseits genügt es z. B. nur zu analysieren, was erfolgreiche Teams ausmacht, ohne transparent machen zu müssen, welche Teams die leistungsstärksten sind. Andererseits meint Transparenz auch nicht, dass die Leistungsdaten allen Organisationsmitgliedern zur Verfügung gestellt werden, sondern, dass jeder Einzelne die über ihn erhobenen Daten einsehen kann. Leonardi und Contractor (2018) empfehlen z. B. den Mitarbeitenden die gesammelten Daten einmal pro Jahr zur Verfügung zu stellen und hierfür mit möglichst einfachen, visuellen Mitteln zu arbeiten. Fraglich ist, inwieweit Mitarbeitende die Implikationen unterschiedlicher Zentralitäts- und Reputationsmaße nachvollziehen können. Und wenn alle Kennzahlen erklärt werden ist zu fragen, ob **nicht trotz, sondern wegen dieser vollkommenen Transparenz die Analysen nicht mehr begriffen werden können**, weil es zu kompliziert geworden ist.

Überraschend mag in diesem Zusammenhang sein, dass z. B. das Tragen von Sociometric Badges in den beschriebenen Studien **nicht als absolute Überwachung aufgefasst**, sondern von Mitarbeitenden entweder positiv aufgenommen wird oder sie nach einer Eingewöhnungszeit vergessen haben, dass sie Badges tragen (Fischbach et al. 2010; Pentland 2012; Yano et al. 2015). Inwiefern Transparenz die Wahrnehmung des Überwachtwerdens und der Fremdkontrolle auslöst, scheint allerdings auch an der jeweiligen National- und Unternehmenskultur zu liegen.

4.4.2.3 Was man noch machen kann

Die Einsatzmöglichkeiten von Netzwerkanalysen sind breit gefächert, weil **Beziehungen sowohl unser wirtschaftliches als auch gesellschaftliches und privates Leben beeinflussen**. Die Fragestellungen zielen letztlich immer auf die Wirkung von Einflussnahme (Macht) und Prestige auf die Ausgestaltung von Kooperations- und Kommunikationsbeziehungen: Wer beeinflusst wen – sei es nun eine Kollegin, Kundin, Chefin oder ein anderes Team oder den Betriebsrat? Sofern Daten über Microsoft Office365 gesammelt werden, sind sie natürlich auf die Arbeit mit eben diesen Softwareprodukten beschränkt. Anbieter wie TrustSphere oder KeenCorp beziehen die Daten für ihre Netzwerkanalysen aus vielen **weiteren Anwendungen wie Slack, Salesforce, SugarCRM, IBM Connections und Notes**. Dadurch können nicht nur intra- und interorganisationale Kommunikationsströme ausgewertet werden, sondern auch die Interaktion mit Kunden, was Leistungsmessungen anhand von kundenbezogenen Kennzahlen ermöglicht.

Über Aspekte des Performance Managements hinaus gibt es Anwendungsfälle aus dem Bereich **Diversity & Inclusion**. Z. B. wurde bei einem Finanzdienstleistungskonzern untersucht, inwiefern interne Netzwerke förderlich für die Karriere von weiblichen Mitarbeitern sind (TrustSphere 2019). Ähnliche Fragestellungen sind für andere Mitarbeitergruppen denkbar: Werden Mitarbeitende mit anderer Hautfarbe, Religion bzw. ethnischer Abstammung in die organisationalen Netzwerke eingebunden oder eher ausgeschlossen? Welchen Einfluss können sie geltend machen und welches Prestige haben sie? Die Ergebnisse, die durch solche oder ähnliche „Diversity Analytics“-Projekte generiert werden, können Augen öffnen und Wandelprozesse anstoßen (Arbeitskreis Controlling heterogener Personalstrukturen 2018, S. 86 ff.).

Sociometric Badges werden relativ oft beim **Design von Büroarchitekturen** eingesetzt. Produkte und Dienstleistungen in diesem Bereich bietet Humanyze an, eine Firma, die von Absolventen des MIT Media Labs gegründet wurde. Unternehmen wie die Bank of America, Deloitte und The Boston Consulting Group haben mit Hilfe der Humanyze-Badges Antworten auf Fragen wie die folgenden gefunden (Waber et al. 2014; Bernstein und Turban 2018; Bernstein und Waber 2019):

- Wie beeinflusst ein Großraumbüro die Face-to-Face Kommunikation?

- Wie viele Kaffeecke bzw. Wasserspender sind pro Etage optimal und wie groß sollten diese Räume sein, sodass sich nicht zu viele oder zu wenige Mitarbeiter tummeln, sondern genau so viele, dass diese sich möglichst gut informell austauschen?
- Wo sollten Abteilungen angesiedelt werden, damit sich möglichst viele ungezwungene „Zusammenstöße“ ergeben?

Auch wenn die Antworten nicht einfach zu verallgemeinern sind, weil die jeweiligen Arbeitskontakte berücksichtigt werden müssen, so scheint es dennoch ein recht robustes Ergebnis zu sein, dass Großraumbüros die Anzahl an Face-to-Face-Kommunikation deutlich senken (bis zu 70 %). Ein Grund ist, dass Mitarbeiter sich ob der permanenten Fragen (und Blicke) ihrer Kollegen gestresst fühlen und deshalb auf digitale Kommunikationswege (E-Mail, Slack, WhatsApp etc.) ausweichen, weil sie bei den digitalen Kanälen selber entscheiden können, wann sie antworten (Bernstein und Turban 2018, S. 6).

Zwar werten die Anbieter von Sociometric Badges die aufgezeichneten Sprachinhalte noch nicht aus, aber der Schritt in diese Richtung ist zu erwarten. Aktuell erfassen die Geräte „nur“ Metadaten wie Lautstärke, Sprechgeschwindigkeit, Abwechslung von Gesprächsanteilen und die Deckung von Sprachstilen (Language Style Matching; siehe Abschn. 4.5.2.3). Beispielsweise analysiert Ambit, ein Spin-off des Stanford Research Institute, das Stimmprofil von Mitarbeitenden und verfolgt in Echtzeit nicht nur das Hin und Her im Gespräch (anhand der Länge und Abwechslung von Gesprächsanteilen), sondern auch, wann ein Mitarbeiter ängstlich, wütend, freudig, traurig oder zaghaft klingt (Bernstein et al. 2019, S. 34). Solche **Kommunikationsdaten können als Leistungsfaktoren im Performance Management** dienen, wenn ihr Zusammenhang mit Kreativität (Anzahl entwickelter Ideen), Entscheidungsgeschwindigkeit (Dauer der Gespräche bis Entscheidung gefällt wird), Teamkohäsion oder Mitarbeitermotivation (wenn Mitarbeitende im wahrsten Sinne des Wortes „gehört werden“ und „alle gleichermaßen zu Wort kommen“) berechnet wird und dieser sich als statistisch signifikant erweist. Die Daten können den Teams auch **in Echtzeit über Grafiken zurückgespielt** werden, sodass diese ihr Kommunikationsverhalten visualisiert sehen – und steuernd einwirken können (Quwaider und Biswas 2010; Zhang et al. 2018).

Selbst bei **Führungsverhalten** und der **Qualität sozialer Interaktionen** kann die Analyse von Verhaltensweisen zu Erkenntnissen führen, mit deren Hilfe Trainings handlungsorientierter gestaltet werden können (siehe Beispiel-Kasten). Neuere Studien widmen sich emergenten **Phänomenen auf Teamebene**, z. B. dem Teamgeist bzw. Teamkohäsion, gemeinsamen mentalen Modellen oder Teamemotionen (Kozlowski und Chao 2018; Zhang et al. 2018). Solche Phänomene werden zwar immer

wieder von Praktikern beschrieben, sie sind aber schwierig zu messen und unklar in ihrer Entstehung sowie ihrem Verlauf. Immerhin ist recht klar, dass wertschätzende Kommunikation und Vertrauen hierbei eine wichtige Rolle spielen. Dabei erwächst Vertrauen im Zeitverlauf aus solch wertschätzenden Interaktionen, wirkt dann wieder zurück auf die Zusammenarbeit und lässt qualitativ hochwertigere Teambeziehungen entstehen (Busch und von der Oelsnitz 2016). Vertrauen bildet zugleich eine wichtige Voraussetzung, um das entstehen zu lassen, was als psychologische Sicherheit bezeichnet wird (Edmondson 1999). Ist sie etabliert, können Teams gut lernen und innovativ sein, weil sich die Teammitglieder trauen, ihre Meinung offen zu äußern und jeder Ideen, Fragen oder Vorschläge einbringen, aber auch Fehler eingestehen kann, ohne deswegen negative Konsequenzen befürchten zu müssen.

Beispiel: Die Vermessung des Führungsstils

Welche konkreten Verhaltensweisen kennzeichnen einen partizipativen bzw. transformationalen Führungsstil? Damit Führungskräfte diese Führungsstile erlernen können, muss man wissen, welches verbale und non-verbale Verhalten zu den erwünschten Ergebnissen führt. Genau das wurde in einer Studie gemessen, und zwar mittels kleiner tragbarer Geräte, die mit Bewegungssensoren, Bluetooth- und Infrarotschnittstellen sowie Mikrofonen ausgestattet sind, um Interaktionen zwischen Personen aufzuzeichnen (Meyer et al. 2016).

Ein partizipativer Führungsstil ist vor allem dadurch gekennzeichnet, dass Führungskräfte alle Teammitglieder nach ihrem Input fragen. Beispielsweise können Fragen zu bestimmten oder weiteren Informationen, zur persönlichen Meinung oder zu neuen Vorschlägen gestellt werden. Das stimuliert Reflektion und trägt dazu bei, dass sich ein einheitliches Verständnis des Gesamtbildes etabliert. Empathie lässt sich vor allem dadurch fördern, dass die Gestik, Mimik, Haltung und Bewegungen der anderen nachgeahmt werden (Meyer et al. 2016, S. 778). Nicht nur non-verbales Verhalten kann gespiegelt werden. Die Nachahmung der Sprache der anderen (Geschwindigkeit, Tonalität, Begriffe, etc.) führt dazu, dass sich die Beteiligten eher mögen und leichter in sozialen Austausch kommen und Vertrauen aufbauen (Meyer et al. 2016, S. 779).

Für die Messung wurden 165 Personen verkabelt. In Dreierteams musste eine Rekrutierungsaufgabe gelöst werden, wobei relevante Informationen über die Bewerber ungleich in der Gruppe verteilt waren. Bei 28 Teams wurde ein partizipativer Führungsstil angewendet, bei den restlichen 27 Teams ein direkter. Die Forscher bestimmten die Ausprägung des Frageverhaltens anhand von Videoaufnahmen. Das nachahmende Verhalten wurde über die Geräte gemessen: insgesamt 18 Sensoren zeichneten auf, ob die Führungskräfte innerhalb von 60 Sekunden ähnliche Verhaltensweisen wie ein anderes Teammitglied zeigten (z. B. Kopfnicken, Arme verschränken, Gestikulieren).

Die Ergebnisse: partizipativ agierende Führungskräfte nutzen tatsächlich mehr Fragen, was sich in besseren Bewertungen der Führung und qualitativ besseren Teamentscheidungen niederschlägt. Ebenfalls hat die Nachahmung des Verhaltens anderer einen positiven Effekt auf die Bewertung der Führungskraft im Entscheidungsprozess – sie führt aber nicht zu besseren Entscheidungen (Meyer et al. 2016, S. 784).

4.4.2.4 Was (noch) nicht geht

Wie bereits erwähnt, ist in Deutschland die Implementierung von technischen Überwachungseinrichtungen – und dazu zählen Microsoft Office365 und Socio-metric Badges – **mitbestimmungspflichtig**. Außerdem stehen der Vorratsdatensammlung und der zweckfreien Auswertung **datenschutzrechtliche Grundsätze und Verbote** entgegen (siehe Abschn. 4.4.1.4). Mindestens ist deshalb der Abschluss von Betriebsvereinbarungen erforderlich und es sollten nur zeitlich begrenzte Erhebungen mit definiertem Zweck und begründetem Interesse des Arbeitgebers verfolgt werden. Zudem kann leistungsschwachen Mitarbeitenden nicht einfach gekündigt werden, weil Daten aus einer umfangreichen Kontrolle des Leistungsverhaltens im Kündigungsschutzprozess nicht verwertet werden können: Der Verwertbarkeit stehen Beweisverwertungsverbote entgegen, die aus dem Recht auf informationelle Selbstbestimmung gem. Art. 2 Abs. 1 i. V. m. Art. 1 Abs. 1 GG der betroffenen Arbeitnehmer resultieren (Kraus 2018, S. 705).

Sind die Erhebung und Auswertung gesetzlich und betrieblich erlaubt, gilt es die **Daten in Bezug auf ihre inhaltliche Aussagekraft zu validieren**. Beispielsweise muss ein Meeting-Eintrag in Outlook nicht heißen, dass die betreffende Person tatsächlich bei der Besprechung war. Selbst wenn sie das Meeting wahrgenommen hat, muss dies nicht bedeuten, dass tatsächlich eine Beziehung zu anderen aufgebaut wurde. Und falls ein engerer Kontakt stattfand, muss dessen affektive Qualität bemessen werden, denn häufige Interaktion kann ebenso eine freundschaftliche Beziehung bedeuten wie einen immer wieder aufkeimenden Streit. Auch die Daten der Sociometric Badges sind kritisch zu hinterfragen, einerseits in Bezug auf die Frage, was sie inhaltlich eigentlich aussagen, und andererseits auf die Verlässlichkeit der generierten Daten. Ersteres meint, dass aus den reinen Metadaten ohne inhaltliche Auswertung des Gesagten nur bedingt Rückschlüsse auf die soziale Qualität der Beziehung gezogen werden können (z. B. lautes Sprechen an sich muss nicht heißen, dass jemand kritisch angegangen oder angeschrien wird, sondern kann in der Persönlichkeit des Sprechers, der Umgebungslautstärke oder Entfernung vom Gesprächspartner liegen). Weiteres bezieht sich unter anderem auf Abweichungen, die auch bei baugleichen Geräten festzustellen sind, sodass

beispielsweise Mikrofone unterschiedlich sensibel reagieren (Chaffin et al. 2017, S. 5).

Die Ergebnisse, die über Microsoft Workplace Analytics geliefert werden, kommen **ohne Rückgriff auf (sozial-)psychologische Konstrukte** wie Persönlichkeitseigenschaften, Leistungsmotivation oder intrinsische Motivation, etc. aus. Solche Konstrukte werden meist über Selbsteinschätzungen erhoben, weshalb mit Verzerrungen zu rechnen ist (z. B. aufgrund sozial erwünschter Antworten). Diese Stärke ist gleichzeitig auch eine Schwäche, weil **mit dem reinen Fokus auf Verhaltensdaten die sozialpsychologische Komplexität von Arbeit unterschätzt wird** (Khusainova et al. 2018, S. 12 f.). Zwar lassen sich dadurch relativ konkrete Handlungsempfehlungen geben, aber diese können kontraproduktiv sein, weil sozialpsychologische Faktoren vernachlässigt werden. Nimmt man z. B. eine Empfehlung wie „Erhöhe die Anzahl und Dauer von Meetings mit Kunden“ und denkt dann an einen pomadigen, unangenehmen oder unzuverlässigen Vertriebler, so wird einsichtig, dass es wahrscheinlich zu negativen Kundenreaktionen kommt, wenn sie nun von diesem Vertriebler öfter besucht werden (Fuller 2014). Außerdem wurden nur Korrelationen und nicht Kausalitäten in einem Längsschnittdesign analysiert, was Fuller (2015) auch selbstkritisch anmerkt.

Eine weitere Limitation der beschriebenen Studien ist, dass **Erfolg nur eindimensional gemessen** wurde (anhand von Abschlussraten). Was aber ist, wenn die Anzahl nicht vergleichbar ist, etwa weil unterschiedlich komplexe Produkte verkauft werden? Was ist mit weiteren, (nicht-)monetären Aspekten von Erfolg wie beispielsweise Margen, Wiederkaufsraten, Kundenzufriedenheit oder Motivation? Die Wechselwirkungen, die in Unternehmen herrschen, können durch eine Kennzahl nur unzureichend abgebildet werden. Eher ist bei solch eindimensionalem Management von **interessengeleiteten Handeln** auszugehen – sowohl auf Seiten der Führungskräfte als auch auf Seiten der Mitarbeitenden. Getreu dem Motto: ‚Tell me how you will measure me, and I tell you how I will behave‘, womit die Kennzahlen nur noch abbilden, wie gut sich jemand an sie angepasst hat (auch bekannt als Goodhart's Laws bzw. Campbell's Law). Diese Logik verstärkt sich durch das Gefühl des Überwachtwerdens, sodass nur noch die Verhaltensweisen gezeigt werden, die beobachtet werden, während kreative, risikoreiche oder unbeobachtbare ausbleiben (Gärtner 2018b, S. 157).

4.5 Personalentwicklung

Personalentwicklung ändert sich durch den Einsatz von digitalen Tools in mehreren Dimensionen, was sich in unterschiedlichen Übersichten bzw. „Landkarten“ widerspiegelt, z. B. der „Global Learning Landscape“ von HolonIQ oder der „Glo-

bal EdTech Landsacape“ von Navitas. Weltweit gibt es im Umfeld von Lernen und (Weiter-)Bildung tausende digitale Tools, wobei in den letzten Jahren vor allem die Aspekte **Analytics und Chatbots für adaptives, personalisiertes Lernen** sowie **Virtual und Augmented Reality** in den Fokus der Betrachtung gerückt sind (Jenewein 2018, S. 266; Hilbig et al. 2019, S. 5). Diese Themen sollen auch in diesem Kapitel im Fokus stehen.

Es ändern sich jedoch nicht nur Technologien, sondern auch die Arbeitsbedingungen und damit die geforderten Kompetenzprofile, was wiederum eine **Anpassung des strategischen und operativen Wissens- und Kompetenzmanagements** nach sich zieht (North und Maier 2018). Dazu gehören strategische Fragen, welches Wissen und welche Kompetenzen zukünftig aufgrund der Unternehmensstrategie gebraucht werden und welche Geschäftsmodelle aufgrund veränderter Kompetenzen möglich sind. Diese Fragen werden im Folgenden ebenso ausgeblendet wie Aspekte der vor- bzw. innerbetrieblichen Sozialisation (z. B. Onboarding, Kulturentwicklung) und des Karrieremanagements (z. B. Karrierepfade wie Führungs-, Fach- und Projektkarrieren). Im Fokus stehen die Themen **betriebliche Weiterbildung und Kompetenzentwicklung** (Kauffeld 2016, S. 8) bzw. das, was North und Maier (2018, S. 675) als operatives Wissensmanagement bezeichnen: das richtige Wissen und die richtigen Kompetenzen zur richtigen Zeit am richtigen Ort verfügbar zu machen.

Je schneller Wissen und Kompetenzen veralten, desto relevanter wird Weiterbildung bzw. Kompetenzentwicklung als zentraler Baustein der Personalentwicklung. Unabhängig davon, ob dies auf eine Organisation zutrifft oder nicht, bleibt die Frage nach dem **Nutzen von Personalentwicklung ein Dauerbrenner**. Und da in profitorientierten Unternehmen tendenziell in Kategorien von Kosten und Gewinnen bzw. Nutzen gedacht und gehandelt wird, muss die Personalentwicklung ihren Mehrwert hinsichtlich dieser Kategorien aufzeigen – ungeachtet der Tatsache, dass auch andere Ergebnisse eine Rolle spielen können (z. B. Selbstverwirklichung bzw. persönliche Weiterentwicklung). Die Zielsetzung ist, dass die Kompetenzentwicklung adaptiv (**personalisiert und kontextspezifisch**) sowie **selbstorganisiert** ablaufen soll, das heißt, die einzelnen Mitarbeiter und Teams sind aufgefordert ihr Wissen und ihre Kompetenzen aktuell und passend für die konkrete Aufgabe zu halten. Hierfür kommen eine Reihe von (Learning-)Analytics-Lösungen in Betracht, die vor allem die Protokolldateien von Lernplattformen als Datengrundlage nutzen (siehe Abschn. 4.5.1).

Während Unternehmen schon seit Jahren auf eLearning bzw. Blended-Learning-Konzepte setzen, ist der Einsatz von tragbaren Geräten (**Wearables**) noch kaum verbreitet (siehe Liebert und Talg 2018; Meier et al. 2018; Wilson 2013). Der Vorteil ist, dass mit den Sensoren der Wearables genau die großen, vielfältigen und

personalisierten Datensätze generiert werden können, die für Learning Analytics gebraucht werden. Ähnliche Möglichkeiten bietet der Einsatz von **Virtual und Augmented Reality**, wobei hier der Vorteil einerseits darin liegt, dass das Lernsetting risikoarm und kostengünstig ist, und andererseits Fragen nach dem Datenschutz weniger virulent werden als beim permanenten Tragen von Wearables (siehe Abschn. 4.5.2). Auch wenn sich die in den beiden Abschnitten genannten Technologien unterscheiden, verfolgt man mit ihrem Einsatz im Allgemeinen dennoch die gleichen Ziele: **Individualisierung, Kontextualisierung und Selbstorganisation der Kompetenzentwicklung**.

Im Einzelnen ändern sich durch den Einsatz digitaler Tools mehrere Aspekte der Weiterbildung bzw. Kompetenzentwicklung (siehe Meier et al. 2018):

- Die **Lerninhalte** lassen sich durch „Learning Analytics“ (synonym: „Educational Data Mining“) besser individualisieren und an den Lernfortschritt bzw. hinsichtlich der Lernziele anpassen.
- Die Weiterentwicklungs**methoden** und **Lernorte** werden diverser: Die Palette reicht vom informellen und haptischen Lernen am Arbeitsplatz mit bzw. von anderen Lernenden, über virtuelle Weiterbildung am Computer bzw. per Smartphone mit Hilfe von Video-Bibliotheken, kuratierten Blogposts, E-Coachings, spielerischen Tests oder Simulationen bis hin zum Einsatz von Sensoren, Augmented und Virtual Reality. Durch diese Medien und Technologien wird die vormals textlastige Inhaltsvermittlung nun bildhaft, vertont, interaktiv und haptisch. Digitale Technologien ermöglichen somit multimodale und interaktives Lernen, was beides als **qualitätssteigernd** gelten kann (Niegemann und Heidig 2019, S. 3). In dem Maße, wie digitale Tools die selbstorganisierte oder (teil-) automatisierte Erstellung von Lerninhalten unterstützen, geht es allerdings weniger um Qualität und mehr um **Kostenreduktion** (Meier et al. 2018, S. 22, 25).
- Beim **Controlling** der Personalentwicklungsmaßnahmen kann auf (Nahezu-) Echtzeitdaten aus der Trainingssituation zugegriffen werden. Zu den klassischerweise erhobenen Daten über Teilnehmer, Trainer und Trainings(ausstattung) (z. B. jeweilige Anzahl, Kosten, Zeiten, Kompetenzen) kommen nun noch Daten über den eigentlichen Trainingsverlauf und den Lernfortschritt.
- Aus den eben genannten Veränderungen ergeben sich neue Anforderungen an die **Kompetenzen** von Personalentwicklern. Insbesondere müssen sie die Möglichkeiten und Grenzen der Tools des digitalisierten Lernens einschätzen können. Denn egal, ob diese als Fluch oder Segen betrachtet werden, es ist offensichtlich, dass **digitale Technologien Medien des Wissens und Handelns** sind: Wir wissen mehr, wenn wir Zugriff auf weltweite Datenbanken haben, und wir agieren anders, weil wir diese Möglichkeiten haben.

- Mit den Aufgaben verändern sich auch **Rollen** bzw. Erwartungen an Personalentwickler, Führungskräfte, Mitarbeiter und Anbieter von Bildungsservices. Weil die digitalen Tools neue Formen des bedarfsgtriebenen, selbstgesteuerten und sozialen Lernens ermöglichen, rücken die Lernenden sowohl stärker in den Fokus als auch in Treiber-Rolle, während HR zum Kurator von Lerninhalten wird (Jenewein 2018, S. 273; Liebert und Talg 2018, S. 203 f.): Die Lernenden bestimmen Form, Inhalt, Zeit und Mitlernende eher selbst und die entsprechende Nachfrage wird über firmeninterne oder -externe Cloudangebote befriedigt (**vom Push- zum Pull-Learning**). HR ist dann in der Verantwortung, kontinuierlich relevante, hochwertige und aktuelle Lerninhalte zu suchen und zur Verfügung zu stellen – und mutiert zum Spotify des Lernens, weil Inhalte wie eine personalisierte Playlist dargeboten werden. Gleichzeitig werden die Lernenden aber noch mehr als früher für ihre Weiterentwicklung und „employability“ verantwortlich gemacht. Ein nicht-technisches Argument dafür ist, dass bei hoher Spezialisierung und Dynamik in der Wirtschaft zunehmend nur die Führungskräfte und Mitarbeiter selbst wissen, was sie brauchen. Ein anderes Argument ist, dass die Weiterbildungsbudgets den Investitionsbedarf nicht decken und „Employee Self-Service“ billiger für das Unternehmen ist.

Im Folgenden wird der Fokus auf dem **Weg zum personalisierten und adaptiven Lernen** liegen, da sich hier die Weiterbildungsaspekte Inhalt, Methodik, Rolle und Controlling schneiden. Zudem werde ich Begriffe wie Weiterbildungsmaßnahme, Training und Seminar als Synonyme behandeln. Gleiches gilt für Lernen, Wissens- und Kompetenzaufbau, was zudem nur in Bezug auf die individuelle und nicht die kollektive Ebene betrachtet wird (für differenziertere Ausführungen siehe Kauffeld 2016; Duscheck und Gärtner 2018).

4.5.1 Learning Analytics & Learning Management Systeme: Personalisiertes, adaptives und selbstorganisiertes Lernen

4.5.1.1 Worum es geht

Seit Jahrzehnten steht die Frage im Raum, wie der **gesamte Nutzen von Weiterbildungen** gemessen werden kann. Zwischen hoch wissenschaftlichem Vorgehen (quasi-experimentelle Designs mit mehreren Messzeitpunkten und Kontrollgruppen) und einfachen „Happiness Sheets“, die nur das Befinden der Teilnehmer abfragen, hat sich eine ganze Evaluationsindustrie gebildet. Auf der Suche nach

ebenso validen wie handhabbaren Herangehensweisen hat sich das Vier-Ebenen-Modell von Kirkpatrick (1967) als gangbarer Mittelweg etabliert. Die vier Ebenen, auf denen die Evaluation eines Trainings stattfinden soll, sind: Reaktion, Lernen, Verhalten und Resultate. Die erste Ebene (**Zufriedenheitserfolg**) erfasst die Wirkung des Trainings auf die Teilnehmenden direkt im Anschluss an die Weiterbildungsmaßnahme. Hinsichtlich der zweiten Ebene (**Lernerfolg**) wird gefragt, inwieweit die Teilnehmenden etwas durch die Weiterbildung gelernt haben, das sie vorher noch nicht wussten. Die Fragen, ob die Mitarbeitenden ihr Verhalten tatsächlich ändern (Ebene 3: **Transfererfolg**) und dies auch zu erwünschten Ergebnissen bei der Arbeit führt (Ebene 4: **Unternehmenserfolg**), können nur durch zeitversetzte Erhebungen beantwortet werden – und oft auch nur, indem die Perspektive Dritter (Kollegen, Führungskräfte, Kunden) eingeholt wird (Kauffeld 2016, S. 114 f.).

Learning Analytics fokussiert primär den Lernerfolg und teilweise den Transfererfolg. Erreicht wird dies, indem Daten über Lernprozesse gemessen, gesammelt und ausgewertet werden, um diese Informationen den Lernenden und Lehrenden zur Verbesserung der Lerninhalte und -abläufe (am besten in Echtzeit) zur Verfügung zu stellen (Ifenthaler und Drachsler 2018, S. 3). Insbesondere soll Learning Analytics den Lernerfolg erhöhen, indem durch Datenanalysen ein personalisiertes und adaptives Lernen unterstützt wird. Der Vorteil von Learning Analytics ist, dass detailgenaue Beobachtungen von Lernprozessen bei sehr großen Teilnehmerzahlen vorgenommen und Muster in unterschiedlichen Dimensionen identifiziert werden. Diese Muster beziehen sich meist auf erfolgversprechende Verhaltensweisen beim Lernen, z. B. die Anzahl an Interaktionen mit bestimmten Lernobjekten oder anderen Lernenden, die Dauer der Auseinandersetzung mit Lerninhalten, (nicht-)erfolgreiche Lernpfade und -inhalte. **Learning Analytics ist ein Wegweiser, nicht der Weg.** Mit den Zahlen sind noch keine Lerninhalte erstellt und es hat auch noch keine wie auch immer geartete Weiterbildungsmaßnahme stattgefunden. Um tatsächlich personalisiertes und adaptives Lernen zu ermöglichen, müssen noch **Lerninhalte, -methoden, -technologien, -orte und -rollen gestaltet** werden. Als Richtlinien für diese Gestaltung gelten viele Praktiker und Wissenschaftler aktuell die Prinzipien der Zentrierung auf den/die Lernenden und deren Selbstregulation (Kauffeld 2016, S. 7, 156 f.; Ifenthaler und Drachsler 2018, S. 10; Meier et al. 2018, S. 19). Das bedeutet, dass die Lernenden entweder selbst über die Inhalte/ Ziele, Methoden, Technologien und Orte des Lernens entscheiden oder zumindest weitgehend mitentscheiden dürfen.

Sowohl die adaptive Personalisierung als auch die Selbstorganisation lässt sich am einfachsten ermöglichen, wenn das Lernen virtuell stattfindet. Hierfür gibt es unzählige **Learning Management Systeme bzw. Lernplattformen** (z. B. Adobe Cap-

tivate, Canvas, Cornerstone Learning, SAP SuccessFactors/SAP Learning Hub, Workday bzw. quofox, coursera, degreed, IMC, Lynda, Udemy, LinkedIn Learning). Bei diesen Systemen werden den Lernenden (mehr oder weniger) personalisiert unterschiedlichste Themen und Medien angeboten (Videos, Podcasts, Artikel, Chats, TED-Talks, Spiele, etc.) und die (Selbst-)Organisation des Lernens softwarebasiert abgewickelt. Da es schon länger nicht mehr um die Verwaltung von Weiterbildungsmaßnahmen, sondern die Personalisierung des Lernerlebnisses geht, bezeichnen sich manche Anbieter auch als **Learning Experience Platform**. Daneben gibt es themenspezifische Tools, z. B. die Sprachlernsysteme Carnegie Speech und Duolingo oder Nachhilfesysteme wie Carnegie Cognitive Tutor for Mathematics. Schon seit Jahrzehnten gibt es Testverfahren, bei denen die Auswahl an Fragen computergestützt funktioniert (Computer Adapted Testing, CAT). In der umfänglichsten Ausbaustufe gehen sie weit über die reine Trainingsadministration (An- und Abmeldungen, Teilnehmerkommunikation, Zertifikate, etc.) hinaus und bieten Bewertungs- bzw. Prüfungsoptionen (Quiz-Fragen, Online-Tests, etc.), bereiten unterschiedliche Lerninhalte und -formate auf (Texte, Videos, Lernkarten, etc.) und ermöglichen den Austausch zwischen den Lernenden (Chats, virtuelle Gruppenräume, etc.). Trotz aller Funktionalitäten: Diese Systeme unterstützen nur die Wissensvermittlung, nicht den Aufbau von Kompetenzen. Dass digitale Technologien allein die hohen Ziele der Selbstorganisation bzw. -regulation und des erhöhten Lernerfolgs nur teilweise erfüllen können, wird unten noch diskutiert (siehe Abschn. 4.5.1.4).

4.5.1.2 Wie es geht

Im engeren Sinn geht es bei Learning Analytics um deskriptive und diagnostische Analysen sowie prädiktive ML-Verfahren mit dem Ziel den Lernerfolg zu erhöhen. Deskriptive Auswertungen, z. B. über die Anzahl absolviert er Lerneinheiten, erfolgreich bestandener Tests (oder Anteil fehlerhafter Antworten), werden den Lernenden zurückgespielt, meist über ein **Learning Analytics Dashboard**. Der Lernprozess wird damit vor allem über Kennzahlen abgebildet. Ziel ist, dass die Lernenden einen höheren Grad an Selbstbestimmung erleben als es bei Trainings ohne dieses Feedback-Element der Fall wäre. Dazu müssen sie auf die rückgemeldeten Informationen **aufmerksam werden, sie reflektieren und interpretieren**, um dann **entsprechend handeln** zu können. Diese Teilprozesse sollen Dashboards unterstützen, indem sie unterschiedliche Elemente und Ergebnisse darstellen (Klerkx et al. 2017, S. 144 f.):

- Artefakte, die die Lernenden produziert haben (Dokumente, Posts, Software, etc.)
- Soziale Interaktionen (Blog-Einträge und Kommentare, Twitter-Nachrichten, etc.)
- Nutzung von Objekten (Webseiten- und Videoaufrufe, Handbücher, Präsentationen, etc.)

- Zeitdauer, die mit den Lerninhalten und anderen Objekten verbracht wurde (inkl. wann ein Online-Kurs verlassen und nicht weitergeführt wurde)
- Testergebnisse (aufgeschlüsselt nach Wissensfeldern bzw. -lücken)
- Wahrscheinlichkeit des Bestehens (kann auch nur den Trainern angezeigt werden)

Die so vorgenommene **Quantifizierung und Visualisierung des Lernfortschritts bilden das zentrale Feedback-Instrument** bei Learning Analytics und sollen den Lernenden helfen, Lernschwierigkeiten aufzuspüren und ihr bisheriges Verhalten anzupassen – oder die im Learning Analytics Dashboard dargestellten Informationen kritisch in Frage zu stellen (Ifenthaler und Drachsler 2018, S. 10, 13). Im Einzelfall kann das ein Warnhinweis sein, der sich auf das generelle Verhalten bezieht (Engagement, Fleiß, Präsenz) oder auf Details im Lernprozess (falsche Arbeitstechniken bzw. -routinen, Kenntnislücken).

Die **Daten für Learning Analytics** kommen aus unterschiedlichen elektronischen Systemen, die meist privat(wirtschaftlich) betrieben werden, entweder als **Learning Management Systeme** (LMS) oder als **Lernplattformen**. (siehe Beispiel-Kasten). Dabei werden verschiedene Daten wie die Profilinformation des Nutzers, seine Aktivität auf der Plattform (z. B.: Was gibt die Lernende in der Suche ein? Welchen Links folgt sie? Wo verweilt sie besonders lange? Bei welchen Inhalten werden vertiefende Lektionen aufgerufen?), die Aktivität vergleichbarer Nutzer und die Attribute der Lernobjekte selbst genutzt (Liebert und Talg 2018). In einer Meta-Studie wurden mehrheitlich **positive Lerneffekte von adaptiven Lernsystemen** konstatiert (Escueta et al. 2017). Neben diesen geschützten Datenbeständen gibt es auch eine zunehmende Anzahl von öffentlichen Datensätzen (siehe das Learning Analytics Community Exchange-Projekt: www.LACEproject.eu oder www.educationaldatamining.org; die dort verfügbaren Daten stammen allerdings aus dem schulischen und akademischen Ausbildungskontext und nicht aus der arbeitsplatzbezogenen Weiterbildung). Eine weitere Datenquelle können Audience Response Systeme sein, mit Hilfe derer man Rückmeldungen während einer Weiterbildung maßnahmen einholen kann (z. B. in Form von Umfragen oder votings). Es empfiehlt sich trotz des Rückgriffs auf externe Angebote den Zugriff auf die Prozessdaten zu wahren, um als Personalabteilung selbst personalisierte Bildungsdienstleistungen anbieten zu können (Meier 2019, S. 41).

Beispiel: Lernplattform von quofox (<https://quofox.com/de>)

Quofox ist 2015 gegründet worden und betreibt eine der größten Lernplattformen für den deutschsprachigen Raum sowie ein Lernmanagementsystem. Angeboten werden unterschiedlichste Lernformate: Online- und Offline-Kurse,

Learning Nuggets, Web Based Trainings, Lerngruppen (Social Learning), Coachings & Mentoring, eBooks, Blended Learning und eine App für Smartphones, um auf die Lernressourcen auch von unterwegs aus zuzugreifen. Über unterschiedliche (ML-)Algorithmen werden individuelle Weiterbildungsangebote und angepasste sowie anpassbare Lernpfade angeboten. Im persönlichen Dashboard werden Dokumente und Lernaktivitäten dokumentiert: Es gibt eine Merkliste, eine Anzeige von Nachrichten und nächsten Kursen, einen Überblick über den Lernfortschritt und natürlich eine Kurssuche sowie Bestellhistorie. Insbesondere der Lernfortschritt und die nächsten Kurse (auf die es meist Rabbatte gibt) sollen zu neuem Lernen motivieren. Elemente wie der Lernfortschritt können schon als Gamification betrachtet werden. Zusätzliche Features wie Sternebewertungen, Badges und Bestenlisten werden kursspezifisch angeboten und sollen spielerisch zum Lernen anregen. Ein weiterer Schritt in Richtung Gamification und Personalisierung ist die Präsentation von Lerneinheiten als Missionen, die von den Lernenden bewältigt werden müssen.

Liegen genügend Daten über den Verlauf und die Ergebnisse vergangener Trainings vor, dann können **Vorhersagen** getroffen werden, z. B. über die Wahrscheinlichkeit des Bestehens oder welche weiteren Lerninhalte und -medien in welcher Reihenfolge interessant sein könnten. Die eingesetzten statistischen Verfahren kommen sowohl aus dem ML-Umfeld (siehe Abschn. 3.4) als auch aus Gebieten der deskriptiven Statistik und empirischen Sozialforschung:

- Clusteranalysen, um **Typen von Lernenden** zu identifizieren oder **Themen-Cluster**, die für (andere) Lernende interessant sein könnten und deshalb empfohlen werden, oder an denen Lernende immer wieder scheitern (solche **Empfehlungsalgorithmen** kennt man von Produktbewerbungen: Kunden, die X kauften, interessierten sich auch für Y).
- Regressionen, um **Erfolgsfaktoren** zu identifizieren, die mit hohen Testergebnissen korrelieren. Damit können Vorhersagen für die Erreichung der Lernziele modelliert oder kritische Schwellenwerte ermittelt werden, um entsprechend Warnhinweise zu geben.
- Klassifikationsalgorithmen, um die **Wahrscheinlichkeit des Bestehens** (bzw. Nicht-Bestehens) anhand verschiedener Faktoren (Lerntypus, Inhalte, Methoden, Zeit, etc.) zu prognostizieren.
- Textanalysen, um **Feedbacks der Lernenden zu Lerneinheiten** auszuwerten, z. B. können mit Sentiment-Analysen positiv bzw. negativ bewertete Aspekte identifiziert werden.

- Expertensysteme und Wissensbasen, um **Lernmaterialien zu strukturieren** und entsprechenden Kompetenzniveaus zuzuordnen.
- Soziale Netzwerkanalysen, um die **Interaktionen von Lernenden** untereinander oder zwischen Gruppen zu analysieren und zu optimieren.

Die Vorhersage von Lernleistungen ist ein immer wiederkehrendes Thema und wird oft über **Klassifikationsalgorithmen** wie Entscheidungsbäume, SVM oder KNN modelliert (z. B. Acharya und Sinha 2014; Kumar et al. 2018). Z. B. zeigt die Studie von Acharya und Sinha (2014), dass SVM studentische Leistungen am besten vorhersagt, jedoch empfehlen sie den Einsatz von Entscheidungsbäumen, weil diese nicht signifikant schlechtere Ergebnisse liefern und gleichzeitig verständlicher sind. Ein weiterer großer Nachteil ist, dass keine Analysen entlang von Interaktionsverläufen vorgenommen werden können. Hierfür eignen sich rekurrente künstliche neuronale Netze (siehe Beispiel-Kasten).

Beispiel: Lernempfehlung bei Online-Kursen

In der Studie von Tang et al. (2017) geht es darum, Lernenden Empfehlungen zu geben, was sie als nächstes tun sollten, um ihre Erfolgswahrscheinlichkeit zu erhöhen. Als Input konnte auf die Interaktionen der Lernenden mit den über 3500 Objekten (Videos, Präsentationen, Tests, etc.) eines Massive Open Online Courses (MOOC) zurückgegriffen werden. Der Vorteil dieser Datenbasis ist nicht nur die schiere Menge (17 Millionen Ereignisse/Datenpunkte), sondern vor allem, dass Interaktionsverläufe nachgezeichnet werden konnten: Welche Ressourcen schauen sich Lernende zuerst an und wie ist die Reihenfolge dessen, was danach kommt? Da es mehrere Methoden gibt, um Längsschnittdaten zu analysieren, wurden zwei Verfahren eingesetzt – n-Gramm-Techniken und Recurrent Neural Networks (LSTM) –, um das Modell zu finden, das die besten Rückschlüsse auf die Leistung zulässt. Beide Verfahren sagen das Lernverhalten ca. drei Mal besser vorher als der Lehr- und Veranstaltungsplan. Die Studie fokussiert sich allerdings auf die Vorhersage der nächsten Interaktion. Ob der Lehrplan tatsächlich ein schlechter Wegweiser durch den Lernprozess ist, kann so nicht direkt gesagt werden – immerhin jedoch, dass Lernende sich nicht entlang des Lehrplans durch das MOOC hangeln (Tang et al. 2017, S. 230).

Mit Learning Analytics auf Basis von ML-Algorithmen können Muster in großen Datenmengen analysiert werden, um entsprechende Empfehlungen für erfolgreiche Lernpfade zu geben. Was fehlt, ist eine **Verknüpfung der maschinellen mit der pädagogischen Perspektive** (Kravčík et al. 2018, S. 149). Zu den pädagogischen Grundregeln gehört, Lernziele zu definieren, die Vorkenntnisse zu bestimmen und

geeignete Methoden der Wissensvermittlung einzusetzen, die helfen, das Lernziel zu erreichen und dabei den Lernprozess attraktiv und effektiv gestalten. Neben dem übergeordneten Ziel, den Lernerfolg zu erreichen, sind deshalb einige Nebenziele zu berücksichtigen: Frustration bzw. Langeweile durch zu hohe bzw. zu niedrige Anforderungen sollten vermieden werden, Anpassungen der Lernziele und -inhalte aufgrund geänderter Präferenzen oder Vorkenntnisse sollten möglich sein und insgesamt muss ein sicheres System entworfen werden, das den Lernenden nicht nur den Schutz ihrer Daten zugesteht, sondern ihnen auch die psychologische Sicherheit, dass ein Versagen nicht zu ernsthaften Konsequenzen führt, vermittelt (Kravčík et al. 2018, S. 140). Eine Lösung, die all das zumindest ansatzweise leistet, ist APPsist (siehe Beispiel-Kasten).

Beispiel: Adaptive und personalisierte Wissensvermittlung mit APPsist

Ziel der App für mobile Endgeräte ist es, Wissen sowohl formell als auch informell zu vermitteln, und zwar angepasst auf das jeweilige Expertiseniveau und die fachlichen Aufgaben von Mitarbeitern in der Produktion (Ullrich et al. 2018, S. 108). Die App schlägt individuell ausgewählte Inhalte und nächste Schritte vor und unterstützt dadurch die Benutzer sowohl bei bisher nicht oder wenig beherrschten Tätigkeiten als auch beim generellen Wissenserwerb. Dafür wurden mit Praxispartnern – unter anderem Festo und MBB – eine Reihe von intelligenten Diensten entwickelt, die mit Hilfe von KI/ML-Verfahren Inhalte flexibel an die Lernenden anpassen.

Im Kern basiert APPsist auf drei Komponenten: einem Wissensmodell, einem Lernermodell und einem Bildungsmodell (im Folgenden Ullrich et al. 2018, S. 109 ff.). Das Wissensmodell strukturiert Lernobjekte in Form von annotiertem Lehr-Lern-Material. Das Lernermodell bildet die Basis für die Personalisierung des Lernens. Dafür enthält es eine Abschätzung des Wissenstands, der Eigenschaften der Lernenden und ihren Interaktionen mit Lernobjekten (z. B. wie oft ein Dokument angesehen wurde, wie oft ein Produktionsgegenstand verwendet und wie oft ein Assistenzschritt gestartet, abgebrochen und beendet wurde). Das Wissensmodell und das Lernermodell werden über das Bildungsmodul so verknüpft, dass die Lernobjekte zu den Eigenschaften der Lernenden passen. Darüber hinaus ist didaktisches Wissen in Form von Wenn-dann-Regeln in einem Expertensystem abgebildet, sodass ein Mindestmaß an didaktischen Grundregeln eingehalten wird. Zusätzlich werden weitere relevante Benutzerinformationen gespeichert, z. B. die Arbeitsplatzgruppe des Mitarbeiters, Freigabe-Regeln (welche Tätigkeiten der Mitarbeiter durchführen darf), ob er/sie sich aktuell in der Haupttätigkeit (beispielsweise der Montage) oder Nebentätigkeit befindet und was die Entwicklungsziele sind, die im Zielvereinbarungsgespräch zwischen Mitarbeiter und Führungskraft festgelegt wurden.

Auslöser für die Verarbeitung und Bereitstellung von Informationen sind Ereignisse, die von drei Seiten aus ausgelöst werden können (Ullrich et al. 2018, S. 108): von Mitarbeitenden (z. B. durch Benutzereingaben), dem Kontext (z. B. durch Änderungen des Maschinenstatus) oder den Diensten selbst (z. B. basierend auf eingetragenen Regeln). Konkret sollen Mitarbeiter sowohl in der aktuellen Arbeitssituation als auch perspektivisch unterstützt werden. Erstes kann beispielsweise bedeuten, dass eine Maschine meldet, dass ein benötigter Werkstoff nicht mehr in ausreichendem Maße vorhanden ist. Im APPsist-System wird der Mitarbeiter nicht nur darauf hingewiesen, sondern bekommt passende Maßnahmen angezeigt. Zweiteres bezieht die im Mitarbeitergespräch vereinbarten Entwicklungsziele mit ein, sodass im APPsist-System Inhalte und Maßnahmen vorgeschlagen werden, die sich auf die Entwicklungsperspektive des Mitarbeiters beziehen. Beispielsweise können ungelernte Mitarbeiter durch Lerninhalte und schrittweise Empfehlungen dazu befähigt werden, den Wechsel eines Werkstoffes in einer teilautomatisierten Montagelinie selbstständig durchzuführen (Ullrich et al. 2018, S. 110 f.).

Manchmal ist die Herausforderung nicht, dass keine Daten vorliegen, sondern, dass es **zu viele Daten** gibt, die als vorhersagekräftige Faktoren gelten können: bearbeitete Dokumente, Anzahl an Interaktionen, Testergebnisse, Einzelaufgaben eines Tests, Zeit, die mit einem Dokument oder Video oder Test, usw. verbracht wurde und (zu) vieles mehr. Um die Diagnostik und Vorhersage zu verbessern, kann dann auf die neuere **Forschung zu Learning Analytics** und auf die jahrzehntelange **Forschung zu Personalentwicklung im Allgemeinen** zurückgegriffen werden. Studien aus dem Learning-Analytics-Umfeld konnten beispielsweise zeigen, dass ein höherer Lernerfolg mit folgenden Faktoren einhergeht: die Lernenden zeigen eine höhere Beteiligung (MacFadyen und Dawson 2010) und liefern ihre Aufgaben rechtzeitig ab (Vermeulen 2014), sie interagieren öfters mit anderen Lernenden in Chats oder Teamforen (Mangaroska und Giannakos 2018) und diese Interaktion beruht auf Gegenseitigkeit bzw. Reziprozität (Haya et al. 2015). Diese Faktoren gilt es dann zu messen und in ein Learning Analytics-Modell zu integrieren. Die „klassische“ Erforschung von Personalentwicklungsmaßnahmen wiederum hat zu einer fast schon unübersichtlichen Vielzahl an wissenschaftlich validierten Tests und Checklisten für alle Phasen eines Trainingsprozesses geführt: angefangen bei der Anforderungsanalyse und der Lernzielsetzung über die Festlegung des Bewertungsprozesses und die Erfolgsfaktoren für den Transfer bis hin zur Auswahl von Methoden (siehe dazu Neuberger 1994; Kauffeld 2016).

4.5.1.3 Was man noch machen kann

Eng verknüpft mit datenbasierten Auswertungen ist **Gamification** also der Einsatz von Spieldesign-Elementen in einem nicht spielerischen Kontext (Freire et al.

2016). Typische Elemente sind Fortschrittsbalken, Sternebewertungen, Badges, Bestenlisten und User-Challenges, die vor allem die intrinsische Motivation triggern sollen (siehe den Beispiel-Kasten zu quofox oben). Daneben vergeben manche Unternehmen, wie z. B. Porsche, die Deutsche Telekom oder Daimler, für die Lernenden auch Punkte: Für den Abschluss eines Lernmoduls verdienen sich die Teilnehmer Credit Points. Je nach Ausgestaltung können diese Punkte in so etwas Materielles wie Einkaufsgutscheine transferiert werden, oder man braucht sie, um eine höhere Stufe und weitere Lernressourcen zu erreichen. Unbekannt ist, ob und wie weit eine Verknüpfung der verdienten Punkte mit anderen Daten des Lernverhaltens erfolgt (Dauer für ein Modul, Anzahl an Wiederholungen beim Ansehen einer Lernressource wie einem Dokument oder Video, Antworten auf Fragen anderer im Chat, etc.). In vielen Studien zeigte sich ein **positiver Einfluss von Spielen auf die Fantasie, Neugier, Motivation, Auseinandersetzung mit dem Thema sowie den Lernerfolg** (Unger et al. 2015, S. 176 f.; Kettler und Kauffeld 2019, S. 251). Digital repräsentierte Spiel-Design-Elemente motivieren nicht nur, sie generieren umfangreiche Interaktionsdaten und helfen so, Lernaktivitäten mess- und sichtbar zu machen (Seufert et al. 2017, S. 15). Allerdings ist zu beachten, dass der Lernerfolg multikausal bedingt ist und beispielweise Faktoren wie Vorwissen, Einstellung zu und Erfahrung mit (Lern-)Spielen, Interesse am Thema oder Aufmerksamkeit im Spiel den Lernprozess und -erfolg beeinflussen (Kettler und Kauffeld 2019, S. 251). In der Praxis ist der Hype um Gamification im Lernbereich schon wieder abgeflaut (Simões 2015, S. 54 f.; siehe auch Vesa & Harviainen 2019).

Einige Organisationen wie SAP und Siemens setzen intelligente **Tutoring-Systeme** bzw. **Chatbots als Lerncoaches** ein (Jenewein 2018, S. 268; Liebert und Talg 2018, S. 204; siehe auch Kerry et al. 2008; Winkler und Soellner 2018). Ähnlich wie bei den Dialogsystemen im Recruiting gibt es einfachere, regelbasierte Chatbots, die z. B. Frequently Asked Questions rund um Trainingsbudgets, An- und Abmeldeprozesse, Zertifikate, etc. beantworten. Beispielsweise wurde für den SAP Learning Hub „Ed the Bot“ als **virtueller Assistent** eingeführt, um die Moderatoren bei der automatischen Beantwortung wiederkehrender Fragen zu unterstützen – personalisierte Rückmeldungen und die Empfehlung passender Lerninhalte sind weitere Funktionen (Jenewein 2018, S. 268; Pesch 2018, S. 24). Mit Learning Analytics hat dies insofern zu tun, als dass deskriptive oder diagnostische Analysen dabei helfen relevante FAQs zu definieren und die eingesetzten Lern-Chatbots die gleichen ML-Verfahren nutzen wie die Recruiting-Chatbots (siehe Abschn. 4.2.2.2). Auch automatisierte Push-Nachrichten während des Lernprozesses sind relativ weit verbreitet, um **situative Lernhilfen** zu geben (siehe Beispiel-Kasten zu „WalkMe“). Natürlich können auch Rückmeldungen an die Lernenden automatisiert ausgespielt werden, etwa um sie zu motivieren, noch die

letzten Minuten eines Videos anzuschauen, einen Test zu machen oder sich einer Lerngruppe anzuschließen (Liebert und Talg 2018; Winkler und Soellner 2018). Ebenso können Alert-Meldungen an Trainer versendet werden, wenn z. B. überproportional viele Lernende an bestimmten Aufgaben scheitern, zu gut abschneiden oder einen Kurs abbrechen. Der IBM Watson Career Coach und CareerFoundry schlagen zusätzlich interessante Stellenangebote auf Basis von gespeicherten Kompetenzprofilen und Lebenslaufdaten vor, um die Nutzer durch ihren Karrierepfad zu begleiten. Ebenfalls auf Basis von IBM Watson läuft CARL, ein Chatbot, der Mitarbeitenden von Siemens Rede und Antwort steht – soweit sie ihm antrainiert wurden (siehe Beispiel-Kasten zu „CARL“). **Generative Chatbots sind noch selten** (Pesch 2018, S. 24). Solche Bots, die Antworten immer wieder neu generieren oder gar Trainingsinhalte vermitteln (und nicht nur vorab definierte Fragen-Answer-Paare ausspielen), gibt es vor allem im (schulischen) Bereich des Erlernens neuer Sprachen und Mathematik (z. B. Botsify; Winkler und Soellner 2018).

Beispiel: WalkMe – ein Assistenzsystem zur Nutzung digitaler Tools (<https://www.walkme.com/germany>)

WalkMe wurde 2011 gegründet und bezeichnet sich als einer der ersten Anbieter einer Digitalen Adaptionsplattform (DAP). Mit Hilfe der DAP lernen Nutzer schneller und einfacher, wie sie (neue) Unternehmenssoftware, mobile Anwendungen oder Websites bedienen sollen, das heißt welche Funktionalitäten ein Softwareprodukt hat. Je nachdem, welche Maske oder welches Feld eine Nutzerin sieht, bekommt sie hilfreiche Informationen zur Bedienung und Schritt-für-Schritt-Anweisungen angezeigt (sogenannte Walk-Thrus). Das Versprechen ist, dass Nutzer selbst komplexe Prozesse ohne vorherige Erfahrung ausführen können, wenn sie den schrittweisen Anweisungen der Pop-ups folgen.

Die anfallenden Daten werden gesammelt und mit Hilfe von ML-Verfahren analysiert, um die bereitgestellten Informationen inhaltlich, zeitlich und personenbezogen zu optimieren. Ohne dass die Nutzerin die Hilfsfunktion aktivieren muss, werden ihr personalisierte Handlungsanleitungen angezeigt, die auf Ort und Zeit der Nutzung, Verhaltensprognosen und Software-Engagement basieren. Laut Patentanmeldung kommen mehrere Algorithmen zum Einsatz, wobei Support Vector Machines und Decision Trees explizit genannt werden (WalkMe 2017, S. 19 f., 24). Als Folge der Nutzung von WalkMe sollen neue IT-Systeme schneller und reibungsloser eingeführt werden können, weil weniger Schulungszeit benötigt wird und die neuen Softwareprodukte von den Mitarbeitenden eher akzeptiert werden. Außerdem können repetitive Arbeitsschritte automatisiert werden und es gibt die Möglichkeit Chatbots zu integrieren (WalkMe 2019).

Beispiel: Cognitive Advisor CARL bei Siemens

Seit 2017 setzt Siemens den kognitiven HR-Assistenten CARL ein, der als Einstiegsplatz für Mitarbeiterfragen an HR dient (Rinser-Willuhn und Greiner im Druck, S. 2). CARL steht einerseits für Cognitive Advisor for interactive user Relationship and content Learning und ist andererseits der Name eines Sohnes von Firmengründer Werner von Siemens (Matzer und Litzel 2017). CARL ist ein textbasierter Chatbot und basiert auf der IBM Watson-Technology, mit der sowohl Sprachverarbeitung als auch das Anlegen, Durchsuchen und Pflegen von Inhalten möglich ist (Rinser-Willuhn und Greiner im Druck, S. 8). Im Kern liegt ein supervised-learning-Modell zugrunde, das heißtt, dass Experten sowohl mögliche Fragen als auch Antworten auf diese Fragen vorgeben und CARL nur Ähnlichkeiten zwischen den Fragestellungen erkennen muss, um entsprechende Antworten auszugeben. Das initiale Anlernen wurde durch das zentrale Team in Deutschland erledigt, darüber hinaus erfolgt die Trainingsarbeit von HR-Mitarbeitenden der jeweiligen Siemens Landesgesellschaften (Rinser-Willuhn und Greiner im Druck, S. 6). Durch die Suchfunktion und das Erlernen passender Antworten nimmt CARL den Nutzern den Aufwand ab, sich durch alle Informationstypen (Rundschreiben, Formulare, Fachinhalte, Prozesse, zugehörige IT-Systeme) zu einem Themengebiet durchzusuchen (Rinser-Willuhn und Greiner im Druck, S. 7).

Das Team aus Siemens- und IBM-Mitarbeitenden startete mit CARL als zusätzlicher Funktion zur klassischen schlagwortbasierten HR-Suchseite im Intranet. Das heißtt, dass CARL auf Fragen z. B. zu Urlaubsansprüchen oder dem Kinderbetreuungszuschuss die Dokumente ausspielt, die dazu hinterlegt sind, die Nutzer sich aber auch durch die Themen klicken oder direkt nach ihnen suchen können (Rinser-Willuhn und Greiner im Druck, S. 2 f.). Anfangs standen nur fünf solcher HR-relevanten Themen bereit, was ein Grund sein mag, warum zu Beginn „nur“ 70 Prozent der 1600 Pilot-Anwender mit CARL zufrieden waren (Matzer und Litzel 2017). Aktuell kann CARL auf über 200 Personalthemen in fünf Sprachen antworten und steht für rund 280.000 Siemens-Mitarbeiter in mehr als 20 Ländern zur Verfügung (IBM 2019). Dass noch nicht jedes Land CARL in die Landing Page eingebunden hat, liegt vor allem an den jeweiligen Landessprachen (z. B. müssten in der Schweiz vier Sprachen berücksichtigt werden). In absehbarer Zukunft soll CARL nicht nur Antworten auf Fragen rund um Personalentwicklungsthemen geben können, sondern für alle HR-bezogenen Inhalte auskunftsähig sein (Krankmeldung, Urlaubs- und Reiseplanung, etc.) – und zwar für alle Siemens-Mitarbeiter weltweit in ihrer jeweiligen Landessprache (IBM 2019).

Eine wichtige Lektion, die die Verantwortlichen lernen mussten, war, dass Gespräche zwischen Menschen und dem Chatbot nicht auf das bloße Bereitstellen von Links oder Antworten auf Frequently Asked Questions reduziert werden kann. Stattdessen müssen Gesprächsverläufe konzipiert, die Anzahl und Länge von Antworten sowie deren Tonalität (inkl. informelle Konversationsanteile, sogenannter „Chit Chats“) auf die landeskulturellen Spezifika angepasst werden. Letztlich kann ein Chatbot nur so gut sein wie der eingepflegte Inhalt – und wie gut dieser auf das Nutzerverhalten ausgerichtet ist (Rinser-Willuhn und Greiner im Druck, S. 11).

Bislang ist der CARL noch nicht mit HR-Backend-Systemen verknüpft, so dass er lediglich Auskünfte erteilen bzw. Informationen zurückspielen, aber keine automatische Änderung von Daten in einem (HR-)System vornehmen kann (IBM 2019). Noch muss CARL solche Aufträge an das zuständige HR-Team weitergeben.

Ein Chatbot-Spezialfall ist Hubert (hubert.ai): Der **Chatbot Hubert gestaltet Umfragen interaktiver**, indem er sie wie einen Dialog und nicht wie eine Abfrage aussehen lässt. Der Mehrwert für Personalentwicklung liegt vor allem in der Evaluation aber auch im Feedback an die Lernenden. Einerseits kann man über oberflächliche Happiness-Fragen hinausgehen und dynamisch Nachfragen zu den Gründen für ein gutes oder schlechtes Weiterbildungserlebnis stellen (und so tiefere und handlungsleitende Erkenntnisse gewinnen). Andererseits kann Hubert, während eine Weiterbildungsmaßnahme noch läuft, Chats mit Studierenden beginnen und inhaltlich auswerten, aber auch mit natürlich anmutenden Nachfragen das Engagement stimulieren. Die wissenschaftliche Validierung des Mehrwerts von Chatbots steht noch am Anfang.

Ebenfalls auf Sprachanalysen bzw. Natural Language Processing basieren Anwendungsfälle, bei denen KI zur **automatischen Übersetzung von eLearning-Inhalten**, der automatischen **Transkription** von Videos (inklusive Erzeugung von Untertiteln) und **Verschlagwortung** von Trainingsinhalten eingesetzt wird (Liebert und Talg 2018, S. 200). Für einfache Trainingsinhalte, bei denen es nur um die Übersetzung in eine andere Sprache geht, sind solche Systeme schon im Einsatz, z. B. in der Siemens Learning World (Kunz et al. 2019, S. 215 f.). Wer YouTube-Videos zum Lernen benutzt, kann ebenfalls auf automatische Transkriptionen und zum Teil auch Übersetzungen zugreifen.

Sollte keine Lernplattform bzw. ein Learning Management System vorhanden sein, so kann **RPA als Brückentechnologie** helfen und standardisierte Aktivitäten automatisieren, z. B. bei der Planung von Trainingsterminen und -ressourcen oder der Sentiment-Analyse von Freitexten aus Feedback-Systemen (Liebert und Talg 2018, S. 204).

4.5.1.4 Was (noch) nicht geht

Selbstorganisation bzw. selbstgesteuertes Lernen **beinhaltet mehr als nur das Lernen mit digital aufbereiteten Inhalten**. Selbstgesteuertes Lernen beginnt damit, dass der Lernbedarf erkannt und Lernziele formuliert werden (Kauffeld 2016, S. 63). Schon allein deshalb ist Lernen ein aktiver und kognitiv aufwändiger Prozess, der nicht aus einer reinen Konsumhaltung heraus geschieht. Zudem braucht es interaktive Methoden wie Gruppendiskussionen, Info-Messen/Gruppenpuzzles, Projektarbeiten, etc., weil sich manche Fähigkeiten nur schwer anhand von Texten und Bildern einüben lassen (z. B. Selbstmanagement, interkulturelle Kompetenzen, Handgriffe in der Produktion) (Kauffeld 2016, S. 63 f.). Zwar kann man auch mittels digitaler Technologien in virtuellen Räumen oder an geteilten Dokumenten zusammenarbeiten, aber die Interaktionsqualität und auch die -dynamik ist eine andere als bei Gruppen, die sich persönlich treffen. Denn dadurch entstehen emergente Phänomene wie Vertrauen, Teamkohäsion bzw. ein Teamgeist leichter, während sie sich im digitalen Raum deutlich schwerer herausbilden (Antoni und Syrek 2017, S. 250; siehe auch Busch und von der Oelsnitz 2016). Außerdem ist das Erkennen von individuellen Lernbedarfen nur der erste Schritt, denn es muss ja auch ein entsprechendes Angebot geben – und da die **Erstellung von individuellen Lernmaterialien aufwändig ist**, bricht die Individualisierung bzw. das adaptive Lernen oft an dieser Stelle ab.

Bei virtuellen Trainings gibt es **kaum oder gar keine menschlichen Rückmeldungen**, es sei denn, es werden zeitgleich Online-Lerngruppen, E-Coachings oder Lerntandems gebildet (siehe z. B. CareerFoundry). Feedback bleibt auch bei der digitalen Personalentwicklung wichtig und lässt sich teilweise sehr zeit- und verhaltensnah geben, was ein Vorteil der Digitalisierung ist. Automatisierte Dialogsysteme bzw. **Chatbots, die diese Aufgabe übernehmen können, stehen noch am Anfang der Entwicklung** (Rüdian und Pinkwart 2019). Außerdem zeigt sich der Nutzen einer Weiterbildungsmaßnahme erst am Arbeitsplatz und damit nach der Maßnahme. Auch für diesen Transfererfolg ist Feedback wichtig, hier kommt es allerdings auf die Rückmeldungen der Führungskräfte, Kollegen, Kunden oder Lieferanten an – und die erfährt man direkt und nicht digital vermittelt.

Durch Learning Analytics kann auch das **Gegenteil von Selbstbestimmung und Selbstorganisationskompetenzen** erreicht werden. Wenn den Lernenden immer empfohlen wird, was sie als nächstes tun sollten, ist dies bestenfalls eine Selbstbestimmung innerhalb eines vorgegebenen Rahmens, der den Gestaltungsspielraum eingrenzt. Kritisch betrachtet ist es eine fremdgesteuerte Selbstorganisation, die (Weiter-)Entwicklung nur entlang der Vorgaben zulässt. Der VR-Pionier Laron Lanier (2017, S. 59) hat dieses Risiko als „**Skinner-Box**“ beschrieben, mit Rekurs auf die berühmten Rattenexperimente von B.F. Skinner: Die Lernenden werden auf bestimmte Verhaltensweisen konditioniert, ohne zu merken, dass sie nur auf dargebotene Reize reagieren. Selbst, wenn es nicht zu diesen negativen

Wirkungen kommt, lässt sich fragen, inwiefern die Aus- und Weiterbildung von Kompetenzen wie Kreativität, kritischem Denken, sozialer und interkultureller Kompetenz sowie gemeinsamer Zusammenarbeit durch Learning Analytics und digitale Tools unterstützt wird. Je wichtiger diese Kompetenzen für ein Unternehmen sind, desto eher müsste darüber nachgedacht werden, ob solche Lernsettings tatsächlich hilfreich sind oder nicht der Weiterentwicklung erwünschter Kompetenzen im Weg stehen. Jede Technologie unterstützt bestimmte kognitive und sensomotorische Fähigkeiten, aber vernachlässigt gleichzeitig andere (Greenfield 2009, S. 71).

Bisher werden bei Learning Analytics **vor allem einfach zu erhebende Kennzahlen** berücksichtigt. Einfach meint hier sowohl den Kontext als auch die Art der Daten. Die meisten Studien und Beispiele stammen aus dem universitären oder schulischen Umfeld, ein Kontext, in dem Daten leichter erhoben werden können. Ausnahmen bilden große Unternehmen wie SAP und Siemens, die erste Projekte im Learning-Analytics-Umfeld umgesetzt haben (Jenewein 2018; Liebert und Talg 2018). Zudem ist die Art der Daten meist auf quantitative Aspekte beschränkt, allen voran die Anzahl an Interaktionen mit Lernressourcen oder anderen Lernenden, Zeitspannen (wie lange eine Lernressource angeschaut wurde, wie lange und in welcher Frequenz die Lernende eingeloggt war oder wann sie aus einem Kurs austrat), Testergebnisse und Feedbacks im Sinne von Happiness Sheets und Bewertungen der Lernressourcen. Um den Lernerfolg bei Weiterbildungen in der unternehmerischen Praxis besser erklären zu können, **müssten auch unstrukturierte Daten aus dem Arbeitsumfeld** analysiert werden, z. B. indem man Texte in Chats (Kommentare, Fragen, Beschwerden, etc.) oder Videos bzw. Videokonferenzen in Teamräumen auswertet. Das ist allerdings nicht nur deutlich aufwändiger und anspruchsvoller, was die technischen Fähigkeiten angeht, es ist auch rechtlich und ethisch bedenklich bzw. verboten (siehe Kap. 6).

Unterschiedliche **Datenformate lassen sich oft schwer integrieren**, weil die Daten in unterschiedlichen Systemen gespeichert sind (firmeninternes LMS, externe Lernplattform, internes Enterprise Social Network, Personalverwaltungssystem). Die technischen Herausforderungen lassen sich auch durch Schnittstellen (API) nur bedingt lösen, weil die einzelnen Systeme zum Großteil proprietär sind, was es schwerer macht, die Lerndaten mit anderen Personal- und Verhaltensdaten zu verbinden und Handlungsempfehlungen auszuspielen (Freire et al. 2016, S. 19).

Bei der Evaluation von Personalentwicklungsmaßnahmen spielt **Mikropolitik** eine große Rolle, weil es um Zeit, Geld und persönliche wie geschäftliche Interessen geht (Kauffeld 2016, S. 128). Selbst mit ausführlichem Zahlenwerk lassen sich machtpolitisch motivierte Entscheidungen in Bezug auf Trainingsinhalte, -orte, -methode und Trainer nicht aushebeln. Es ist sogar das Gegenteil denkbar: Ge-

wiefe Mikropolitiker beeinflussen, welche Kennzahlen wie gemessen werden, um die Wahrscheinlichkeit zu erhöhen, dass die Zahlen und damit die Entscheidung zugunsten der von ihnen bevorzugten Option ausfällt.

Eine weitere Herausforderung bei der personalisierten und adaptiven Wissensvermittlung ist, dass bei webbasierten Lernformaten **nicht alle Daten über die Lernende, ihre Historie und ihre aktuelle Situation** vorliegen. Dies liegt einerseits am Datenschutz und ethischen Bedenken: Wie viele Daten dürfen gesammelt werden, um personalisierte Angebote auszuspielen (siehe Kap. 6)? Andererseits liegt es auch am schieren Umfang der relevanten Daten. Dazu gehören nämlich auch der emotionale Zustand, die physische Belastung und unterschiedliche Formen der Ablenkung (z. B. eingehende E-Mails oder störende Gespräche am Lernort). Zum Teil lassen sich diese Daten über tragbare Geräte mit physiologischen und Kontextsensoren erheben. Davon handelt der nächste Abschnitt.

4.5.2 Wearables, Augmented und Virtual Reality für personalisiertes und adaptives Lernen

4.5.2.1 Worum es geht

Neben Learning Analytics und Lernplattformen zeichnen sich zwei weitere Trends beim Einsatz digitaler Tools in der Personalentwicklung ab. Erstens der Einsatz von **Wearables**, also tragbaren Geräten, die mit Sensoren ausgestattet sind, um (Lern-)Aktivitäten und körperliche Reaktionen wie Stress oder physische Belastung aufzuzeichnen (Bürgy 2018, S. 145, 154; Wilson 2013). Mit Hilfe dieser Sensoren sollen Daten über den Lernenden und seine Situation besser und in Echtzeit erfasst werden können. Dazu gehören beispielsweise Daten über die Präferenzen und emotionale Lage des Lernenden, die sich beide relativ schnell ändern können und deshalb in Echtzeit erhoben werden sollten. Zweitens **Augmented bzw. Virtual Reality**, bei denen zwar ebenfalls tragbare Geräte (vor allem Brillen und Handschuhe) benutzt werden, allerdings weniger als Messinstrumente, sondern eher im Sinne eines Mediums, das neue Interaktionen ermöglicht, weshalb sie nicht nur mit Sensoren, sondern auch Akten ausgestattet sind.

Mit Wearables ergeben sich neue Potenziale für Weiterbildung und Lernen, die sich in vier Kategorien unterteilen lassen (siehe im Folgenden Bürgy 2018, S. 151 ff.):

1. **Aktives Lernen** umfasst den Lernprozess der Nutzer und diesen können Wearables durch Handlichkeit und Freihändigkeit unterstützen. Ein Beispiel wäre das Sprachenlernen mit drahtlosen Kopfhörern und integriertem Mikrofon

- (Hearables), womit umgebungs- und zeitunabhängiges Lernen möglich wird. Unterstützt werden kann dies durch Datenbrillen, die das Sprachlernprogramm visuell umsetzen (z. B. visualisierte Vokabeln, Rechtschreibung etc.).
2. Beim **Lernen über sich selbst** erhalten Nutzer Rückmeldungen über ihre Leistung sowie ihr Lernverhalten („Quantified Self“). Möglich wird dies über Sensoren, die eine Reihe von Daten messen (z. B. Blutdruck, Schlafrythmus, Bewegungsmuster, körperliche Belastungen etc.).
 3. Das **Lernen von Zusammenhängen** ermöglicht Lernenden, gesammelte Daten und Erfahrungen in Relation zu stellen (z. B. mithilfe von ML-Verfahren): Wer oder was befindet sich wo zu einem bestimmten Zeitpunkt? Im Gegensatz zu den ersten beiden Arten bewegt sich diese nicht auf der Ebene einer einzelnen Person, sondern auf einer kollektiven Ebene, um generelle Aussagen über den Einzelfall hinaus zu bestimmen.
 4. Die gewonnenen Daten können für das **Lernen über die Nutzung** von Wearables verwendet werden. So ersetzen Smartphones mit Internetanbindung heute beispielsweise den Griff zur Straßenkarte.

Die **Vorteile** des Einsatzes tragbarer Systeme liegen zunächst darin, dass die Aus- und Weiterbildung noch stärker personalisiert und kontextspezifisch durchgeführt werden kann. Mit der Kombination aus persönlichen, sensorbasierten Daten und Log-Daten aus der Lernumgebung können lernrelevante Faktoren wie Stress, Ablenkung, physische Belastung und Emotionen objektiver messbar gemacht und in Bezug zum Lernprozess und -erfolg gesetzt werden. Die Interaktion mit den virtuellen Objekten macht das Erlernte erlebbar und erhöht dadurch den Erinnerungswert (Niegemann und Heidig 2019, S. 10 ff.). Kosteneinsparungen sind ebenfalls ein Grund für die Nutzung von Wearables bzw. VR/AR im Weiterbildungsbereich. Man denke nur an den Wegfall von Reisekosten zum Trainingsort oder die Einsparungen, die Flugsimulatoren im Vergleich zu realen Flügen bei der Pilotenausbildung ermöglichen. Auch die Automobilindustrie ist an solchen Technologien interessiert, da hier ein hohes Volumen an relativ teuren Produkten gefertigt wird und VR/AR zu einem effizienteren Training der Fertigungsmitarbeiter führen kann (Dörner et al. 2019a, S. 373 ff.). Weitere Vorteile sind die zeitliche und räumliche Unabhängigkeit der Lernprozesse, geringe Fehlerfolgekosten (weil in einem sicheren virtuellen Lernumfeld gearbeitet wird, womit potenziell schwerwiegende körperliche, aber auch finanzielle Folgen ausbleiben) und die Möglichkeit zur Etablierung hoher Schulungsstandards im In- und Ausland (Metzger et al. 2018, S. 916). Diese Vorteile sind der Grund, warum Unternehmen wie BMW, Audi, Deutsche Bahn, Siemens, Hagedorn, Avaloq und viele weitere VR/AR im Weiterbildungsbereich einsetzen (Hafen 2018).

Die Systeme werfen aber auch **Fragen** auf und sind kritisierbar (für Details siehe Abschn. 4.5.2.4): Inwiefern liegt eine naive Technologiegläubigkeit vor, die sich entweder nur auf Technologie verlässt oder ihr zumindest eine höhere Relevanz für den Lernerfolg zuschreibt als inhaltlichen, didaktischen und sozialpsychologischen Aspekten? Mit welchen Methoden kann man die lernrelevanten Faktoren messen bzw. wie sollte man die Daten interpretieren, sodass der Rückschluss auf die interessierenden Faktoren mit geringer Fehlerwahrscheinlichkeit gelingt? Wie geht man mit dem Datenschutz und ethischen Problemen um, die eine solche Überwachung (am Körper der Lernenden und in ihrem Umfeld) aufwirft?

4.5.2.2 Wie es geht

Wearables gibt es in den unterschiedlichsten Ausprägungen, wobei sie sich vor allem darin unterscheiden, mit welchen Sensoren sie bestückt sind und wie (einfach) sie getragen werden. Gängige Sensoren (die zum Großteil auch in Smartphones und Tablets verbaut sind) sind: **Kameras, GPS, Mikrofone, Akzeleratoren, Gyroskope, elektrodermale Aktivitätssensoren**. Weniger häufig finden sich Herzfrequenzmesser, Elektroenzephalographen (EEG) oder Luftverschmutzungssensoren (Bauer et al. 2018, S. 32; Yun et al. 2016, S. 30). Die Sensoren können in unterschiedlicher Form getragen werden: im Smartphone, als eigens entwickelte Geräte wie Handschuhe, Arm- oder Stirnbänder und Umhänge-Badges oder als unaufdringliche Systeme, die in die Kleidung oder gar den Körper integriert werden. Immer geht es darum, die Daten über den Lernenden und den Lernfortschritt zu erfassen und für die Unterstützung des Lernprozesses zu verwenden (Bauer et al. 2018, S. 30).

Die Daten, die mit diesen Technologien erhoben werden, lassen **Rückschlüsse auf sensomotorische, kognitive und emotionale Aspekte** beim Lernen zu. Es können Lernzeiten und der Verlauf verschiedener Lernaktivitäten über die Zeit gemessen, aber auch die Konzentrationsfähigkeit oder emotionale Lage bestimmt werden. Beispielsweise ist seit Langem bekannt, dass Emotionen oder Stress mit einer Veränderung des elektrischen Leitungswiderstandes der Haut einhergehen – und der lässt sich mit Hilfe der elektrodermalen Aktivität messen (siehe Boucsein 2012, S. 369–391). Emotionen und Emotionsregulation wiederum haben einen großen Einfluss auf den Lernprozess und die Lernleistung (Yun et al. 2016, S. 32). Der emotionale Zustand kann auch über bild- und sprachanalytische Verfahren abgeschätzt werden, bei denen dann die in Abschn. 3.4 diskutierten ML-Verfahren zum Klassifizieren der Bild- und Sprachdaten in Emotionskategorien eingesetzt werden. Allerdings sind die Erkennungsraten eher schlecht: Fang et al. (2018) berichten von ca. 60 % richtig klassifizierten Emotionen. Ein anderes Beispiel ist das Erlernen neuer Arbeitsabläufe in produzierenden Unternehmen, etwa, wenn die

Bedienung neuer Maschinen eingeübt werden soll. Mit Hilfe der Sensoren kann gemessen werden, in welcher Frequenz und Geschwindigkeit eine Tätigkeit ausgeführt wird und ob die Abfolge korrekt ist (Strohmeier et al. 2016, S. 842).

Ähnliche Anwendungsszenarien sind bei der **Kundeninteraktion** im Einzelhandel bzw. im Vertrieb und generell bei **Präsentationen in der Öffentlichkeit denkbar**. Hier kommen Geräte zum Einsatz, die sowohl die Stimme als auch die Körperbewegungen der Nutzer messen, um ihnen dann in Echtzeit Feedback zu ihrem Verhalten zu geben. Ein solches Tool haben Schneider et al. (2016) entwickelt und für Verkaufspräsentationen unter Zeitdruck („Elevator Pitches“) getestet. Das Feedback an die Präsentierenden führte zu einer Steigerung der Motivation, des Selbstvertrauens und der Selbstwahrnehmung sowie der Gesamtleistung (Schneider et al. 2016, S. 324 f.).

Anwendungen aus den Bereichen **AR/VR setzen ebenfalls Wearables, Sensoren und Analytics ein** – und gehen doch einen etwas anderen Weg bei der Personalisierung des Lernprozesses. Mit Hilfe von Helmdisplays (Head Mounted Displays, HMD) werden Bilder auf einem Bildschirm in der Nähe der Augen angezeigt. Bekannte Beispiele sind AR- oder VR-Brillen. Die Anbieter solcher Brillen (z. B. Oculus Rift, HTC Vive Pro Eye oder Pimax 8K) liefern oft die passenden Eingabegeräte wie Datenhandschuhe, 3D-Mäuse oder Flysticks mit dazu, es gibt aber natürlich auch Spezialanbieter (z. B. den Datenhandschuh Manus VR Glove). Solche Systeme erlauben die handbasierte Steuerung der Aktionen eines Nutzers in der virtuellen Welt, während andere Technologien, z. B. eine virtuelle Kamera, auch eine blickgerichtete oder berührungslose Steuerung mittels Gestenerkennung ermöglichen – dann haben die Nutzer beide Hände frei (Dörner et al. 2019a, S. 236 f.). Um die Interaktion mit der virtuellen Umgebung zu ermöglichen, bedarf es daher einer ganzen Reihe an Geräten und Sensoren (z. B. Kameras, Displays, Beschleunigungssensoren, Gyroskope, Magnetometer) sowie leistungsfähige Computergrafikprogramme (Dörner et al. 2019a, S. 225 ff.). Damit die virtuelle (Lern-)Welt ein überzeugendes Erlebnis bietet und die Lernenden tatsächlich in die Thematik eintauchen, müssen die Displays hochauflösend sein (die aktuell höchste Pixelzahl bietet die Pimax 8K, nämlich 3840 mal 2160 Pixel), möglichst viele Bilder pro Sekunde senden (die Oculus Quest liefert 72 Bilder pro Sekunde), ein weites Sehfeld bieten (bei der Pimax-Brille beträgt es 200 Grad, während die Oculus Quest nur 100 Grad ermöglicht) und die Nutzer beim Tragen der Geräte möglichst wenig einschränken. Größere Beeinträchtigungen erfahren die Nutzer durch das Gewicht und die Handhabung der Geräte sowie die meist noch notwendige Verkabelung. Erst 2019 kam mit der Oculus Quest die erste VR-Brille auf den Markt, die eine kabellose Bewegung im Raum ermöglicht. Der Preis für die Mobilität ist die begrenzte Rechenleistung der kabellosen Brillen (Albrecht 2020, S. 33). Allerdings sind die **technischen Voraussetzungen nur ein Baustein – ebenso wichtig sind**

die Lerninhalte und die Didaktik. Für die inhaltlich-didaktische Gestaltung virtueller Aus- und Weiterbildungsprogramme geben Niegemann und Niegemann (2018) einige Hinweise.

Die Vorteile des Einsatzes von **AR und VR in der Aus- und Weiterbildung** sind zunächst, dass die Trainings risikolos und kostengünstiger ablaufen, aber auch, dass das Eintauchen in die Lernwelt den Lerneffekt erhöht (Broll 2019, S. 346; Dörner et al. 2019a). Die ersten beiden Vorteile sind einfacher zu realisieren als die Steigerung des Lerneffekts, die von vielen weiteren Faktoren abhängt (siehe Abschn. 4.5.2.4). Wenig verwunderlich ist es deshalb, dass die Technologien vor allem dort eingesetzt werden, wo Schulungsmaßnahmen teuer sind und Fehler während des Trainings ein hohes Risiko für Leib und Leben bedeuten, z. B. in der Flugzeugindustrie. Eine der ersten kommerziellen Anwendungen von AR im Trainingsumfeld war, dass Techniker bei Boeing damit auf das Verlegen von Kabelbäumen in Flugzeugen geschult wurden (Broll 2019, S. 345). Ebenfalls lange bekannt sind Flugsimulatoren, in denen Piloten trainieren können – neu ist jedoch der Einsatz von VR in der Pilotenausbildung. Im Dezember 2019 hat Airbus einen Heimsimulator für Piloten, den Airbus VR Flight Trainer, angekündigt, mit dem (angehende) Piloten auch von zu Hause aus in einem virtuellen Cockpit trainieren können (Albrecht 2020, S. 33). Analog dazu gibt es unterschiedliche Simulatoren im medizinischen Bereich. So ermöglicht z. B. „Fundamental Surgery“ (2019) das Üben chirurgischer Eingriffe mithilfe von computer-generierten Simulationen und haptischem Feedback. Auch bei Zahnmedizinern sind VR-Simulatoren im Einsatz und es konnte gezeigt werden, dass ein Simulator mit haptischer Technologie die manuelle Geschicklichkeit förderte, und zwar sowohl bei erfahrenen Zahnmedizinern als auch bei Studierenden (Ben Gal et al. 2010, S. 503). Aber auch über riskante und teure Berufsfelder hinaus gibt es Anwendungsbeispiele. Der Einzelhandelskonzern Walmart trainiert seine Beschäftigten in einem virtuellen Laden darauf, wie sie z. B. mit einem Kundenansturm umgehen sollen (Albrecht 2020, S. 33). Ein Beispiel für die **Kombination von AR und VR bei der Aus- und Weiterbildung** von Servicetechnikern ist das Projekt GLASSROOM (siehe Beispiel-Kasten).

Beispiel: GLASSROOM-Projekt

Beim GLASSROOM-Projekt geht es um die Aus- und Weiterbildung von Technikern im Außeneinsatz, vorrangig im Bereich des Maschinen- und Anlagenbaus. Ziel ist, ein bedarfsoorientiertes Bildungskonzept durch Nutzung der Potenziale der virtuellen und erweiterten Realität zu schaffen (Metzger et al. 2018, S. 903 f.). Beispielsweise soll der Wechsel von Motoren als Lernszenario über VR/AR-Technologien abgebildet werden. Auf der einen Seite soll ein Kompetenzaufbau durch VR ermöglicht werden. Fokus sind hier Schulungsszenarien in der beruflichen

Bildung. Der Kompetenzaufbau auf Seiten der erweiterten Realität vervollständigt die berufliche Weiterbildung durch Nutzung von AR-Brillen (z. B. Google Glass). Dabei soll der Übergang von der Simulation zur Arbeit an realen Geräten z. B. durch sukzessiv abnehmende Anleitungen begleitet werden.

Virtuelle Realität (VR-Brillen, v. a. Oculus Rift DK2) unterstützt bei der Aus- und Weiterbildung der Techniker an mobilen VR-Stationen in Schulungszentren. Smart Glasses erweitern die Realität und unterstützen Techniker am Einsatzort bei Arbeitsprozessen durch Einblendung relevanter Informationen, wie beispielsweise schrittweisen Anleitungen (Bues et al. 2018, S. 114 ff.). Um dies zu ermöglichen, hat das AR-System drei Teilmodule: Das Recording-Tool ermöglicht eine direkte Aufnahme von Prozessen mit den Smart Glasses durch die Fachexpertin. Es können Anleitungen durch Spracherkennung, Fotos oder Videos aufgenommen werden. Das Support-Tool ist die eigentliche Assistenz und sorgt für die Darstellung von Schritt-für-Schritt-Anleitungen, Auftragsübersichten und zusätzlichen Informationen auf dem Display der Smart Glasses. Die Interaktion erfolgt über Spracherkennung oder Touch Interface. Das Management-Modul dient zur Bearbeitung und Verwaltung der Assistenzprozesse am Desktop (Schwantzer 2018, S. 94 ff.).

Die Kombination von VR und Learning Analytics liegt nahe, weil im virtuellen Raum nicht nur relativ einfach Daten über das Lernverhalten gesammelt werden können, sondern auch die Darbietung von Lerninhalten individuell und dynamisch an die Situation des Lernenden angepasst werden kann. Dabei umfasst die adaptive Darbietung sowohl den Inhalt als auch die Benutzerführung, um beispielsweise die Navigation durch eine Lernplattform zu optimieren (Bauer et al. 2018). Neben Anwendungen im universitären Umfeld gibt es eine Reihe von ähnlich gelagerten Projekten in der Wirtschaft, sowohl von Forschungseinrichtungen (z. B. TU Ilmenau, Universität Stuttgart, Institut für Sozialwissenschaftliche Forschung München) als auch von Firmen (z. B. die Virtuelle Werkstatt bei MAHA von UReality; <https://www.immersivelearning.news>). Viele dieser Initiativen sind im industriellen Bereich, insbesondere bei der risikoreichen Wartung und Reparatur von Maschinen oder Gebäuden angesiedelt. So auch das Projekt, das seit 2017 von der Universität Stuttgart im Verbund mit Trumpf und weiteren Firmen durchgeführt wird (siehe Beispiel-Kasten).

Beispiel: Das Projekt VASE (<https://www.ife.uni-stuttgart.de/bpt/forschung/vase/>)

Das Projekt VASE (Virtual and Analytics Service im Maschinen- und Anlagenbau) ist ein Verbund von Maschinenbauunternehmen, Technologieunternehmen für 3D-Lern- und Arbeitswelten, IT-Servicedienstleistern, Betriebswirtschaftlern

und Technikdidaktikern. Die beteiligten Firmen wie Trumpf, Festo oder Stihl sind als Hersteller von teuren (Werkzeug-)Maschinen daran interessiert, Trainings unabhängig von Ort, Zeit und der Frage, ob Anlagen in materieller Form vorhanden sind, abhalten zu können, um sowohl Risiken und Kosten zu reduzieren als auch eine erhöhte Flexibilität bei der Schulung der Nutzer zu ermöglichen (Zinn 2018, S. 8). Dazu wird die Maschine (z. B. der 3-Drucker TruPrint 3000 von Trumpf) als digitaler Zwilling in der virtuellen Lernumgebung repräsentiert, welche über verschiedene Benutzerschnittstellen, z. B. ein Head-Mounted-Display, zugänglich gemacht werden (Duffke 2019; Zinn 2018, S. 6). Sowohl eine freie als auch eine geführte Interaktion mit der Maschine ist möglich, um Szenarien wie die Bedienung, Wartung oder Reparatur einzuführen. Die einzelnen Szenarien können im Single- oder Multi-User-Modus bearbeitet werden (Duffke 2019).

Im Projekt wird ein weiterer Aspekt von virtuellen Lernumgebungen genutzt: Die Daten, welche sowohl den Lernprozess als auch Maschinenlaufzeitdaten umfassen, werden mit Hilfe von ML-Verfahren analysiert. Das macht eine Reihe von virtuellen Services möglich, z. B. automatisierte und individuelle Hilfestellungen, die bei Fehlern während der Interaktion mit dem digitalen Maschinen-Zwilling eingespielt werden (Zinn 2018, S. 6). Dabei können die Lernenden mit dem virtuellen Avatar Marcus kommunizieren. Das immersive Lernen steigert die Motivation der Lernenden und die Anpassung der Szenarien an das Wissensniveau der Lernenden trägt zur Zufriedenheit und dem Lernerfolg bei (Duffke 2019).

Für den Trainingserfolg ist der Grad an Immersion wichtig, weil nur so die Lernenden nicht ständig durch unrealistische Reaktionen der virtuellen Objekte abgelenkt werden. So kann es passieren, dass ein in VR bewegtes Werkzeug, das über ein Tracking-System an die Handbewegungen des Nutzers gekoppelt ist, nicht direkt mit der Hand mitgeführt, sondern mit etwas Abstand nachgezogen wird (Buhr et al. 2019, S. 268). Für ein Höchstmaß an Immersion müssen Objekte in einer so hohen Auflösung bzw. die Interaktionen mit ihnen mit einer so kleinen Zeitverzögerung (Latenz) dargestellt werden, dass die **Auflösung oberhalb und die Latenz unterhalb der menschlichen Wahrnehmungsschwelle** liegen. Um das zu gewährleisten, braucht es entsprechend leistungsstarke Soft- und Hardware. Beispielsweise ermöglicht der Datenhandschuh Manus Glove das Greifen, Loslassen und Werfen von virtuellen Gegenständen und es lässt sich der Grad an Kraft- und Druckanwendung durch ein haptisches Feedback erkennen. Damit sorgt der Datenhandschuh für eine feinkörnige, fingergenaue und natürliche Interaktion (Manus 2019).

Allerdings sollte man dem Impuls widerstehen, möglichst viele Wearables und Sensoren zu verwenden, um dem Nutzer möglichst viele Stimuli zurückzumelden.

Zu viele Geräte reduzieren normalerweise den Grad an Immersion und es ist oft unklar, was mit den erhobenen Daten ausgewertet werden soll (Dörner und Steinicke 2019a, S. 30). Deshalb sollte man immer die jeweilige Aufgabe im Blick haben und für diese analysieren, **welche Wearables und Sensoren wirklich notwendig sind.**

4.5.2.3 Was man noch machen kann

Eine relativ einfache Möglichkeit, Wearables mit Personalentwicklung zu verbinden, sind **Lern-Apps**. Diese haben zwei zentrale Vorteile: Erstens besitzen die meisten Lernenden nicht nur ein Handy, sondern tragen es ständig bei sich, womit situative und kurze Wissensvermittlung möglich wird (Pesch 2018, S. 22). Zweitens können **Gamification**-Elemente einfach integriert und so der Lerntransfer unterstützt werden (Limpächer und Maier 2016, S. 56; Jenewein 2018). Denn hier gibt es in der Praxis oft die größten Hürden, weil Stress, gewohnte Routinen und fehlende Unterstützung sowie Wertschätzung dazu führen, dass keine Verhaltensänderung eintritt und somit ca. 80 % der Seminare am Transfer scheitern (Kauffeld 2016, S. 4). Ein Beispiel ist die Leadership Evolution App, die alle Teilnehmer eines Führungskräfteseminars dazu motivieren soll, sich dauerhaft und nachhaltig mit dem Gelernten auseinanderzusetzen (siehe Limpächer und Maier 2016). Auch in **virtuellen Welten** lässt sich Gamification relativ einfach einsetzen und steigert den Lernerfolg (Oberdörfer und Latoschik 2019). Ein Beispiel ist „The VR Dive“, ein virtuelles Führungskräftereining von Resourceful Humans, in dem Teams (4 Spieler, 2 Beobachter) eine U-Boot-Mission bewältigen müssen. Im Debrief werden die Verhaltensweisen dann reflektiert.

AR/VR bzw. Wearables insgesamt lassen sich gut in **Lernfabriken integrieren**. Lernfabriken sind physische Orte, in denen Produktionsprozesse und -bedingungen nachgestellt werden, um ein möglichst realitätsnahes Lernumfeld sowie problem- und handlungsorientiertes Lernen zu ermöglichen (Abele et al. 2015, S. 1). Um Kosten und Risiken zu minimieren und gleichzeitig eine Vielzahl von Szenarien bei hoher Motivation trainieren zu können, kommen virtuelle Objekte, Räume und Daten über Maschinen, Produktions- und Lernprozesse vermehrt zum Einsatz (Abele et al. 2015, S. 5; Juraschek et al. 2018, S. 153; siehe auch die „International Association of Learning Factories“). Je immersiver und umfangreicher das Lernerlebnis ist, desto eher kann VR sogar die physische Lernfabrik ersetzen.

Eine Weiterentwicklung von Tutoring-Systemen bzw. Chatbots (siehe Abschn. 4.5.1.3) sind persönliche **Learning Avatare**, die den Lernenden als Lerncoaches in der virtuellen Realität gegenüberstehen (Vladova et al. 2019). Zukünftig soll es möglich sein, mit den Avataren in einen Dialog zu treten bzw. über Sprachkommandos mit virtuellen Objekten zu interagieren (Dörner et al. 2019a, S. 224). Die Vision ähnelt der Künstlichen Intelligenz „Jarvis“ in den Iron-Man-Verfilmungen: Man kann den Avatar quasi alles fragen und erhält immer die richti-

gen Antworten (in den ursprünglichen Comics war Edwin Jarvis der Butler der Familie Stark). Wann es die digitalen Butler aus Hollywood in die reale Wirtschaftswelt schaffen, steht noch in den Sternen.

Da in modernen Organisationen **Sprache** eine zentrale Rolle spielt, ist die Frage, wie Menschen (besser) kommunizieren (können), schon lange ein Thema der Personalentwicklung. Einerseits im Hinblick auf Kommunikationsstile als Inhalt bzw. Ziel von Weiterbildungsmaßnahmen und andererseits in Bezug auf Sprache als Medium in Trainings (siehe Lung 1996). Der Einsatz von Wearables ermöglicht nun eine situationsbezogene **Sammlung von Sprach- und Dialogdaten** im Zeitverlauf und die Analysemethoden aus dem Bereich der Text-Mining-Verfahren eröffnen neue Formen der Auswertung dieser großen Datenmengen. Beispielsweise versuchen unterschiedliche Ansätze unter dem Label „**Language Style Matching**“ herauszufinden, wie man messen kann, ob die Kommunikationsstile von Menschen zu gelungener Zusammenarbeit – ob nun im Team oder zwischen Vorgesetzter und ihren Mitarbeitern – beitragen oder nicht. Dafür liegt der Analysefokus nicht auf den Wortinhalten, sondern auf sogenannten Funktionswörtern (Präpositionen, Pronomen, Artikel), weil die Art und Weise, wie wir diese verwenden, einen Einfluss auf unsere Beziehungen zu anderen hat (Müller-Frommeyer et al. 2019, S. 1343). Empathisches Führungsverhalten zeigt sich beispielsweise darin, dass Vorgesetzte Mitarbeiter häufiger verbal nachahmen (z. B. in der Sprechgeschwindigkeit und Syntax der Mitarbeiter), Personalpronomen der zweiten Person (du, Sie, ihr) seltener verwenden und den Äußerungen ihrer Mitarbeiter häufiger zustimmen (Meinecke und Kauffeld 2019, S. 493 f.). Die Messung solcher sprachlichen Muster kann Führungskräften sehr konkrete Hinweise zu Verhaltensänderungen geben (siehe den Beispiel-Kasten zur Vermessung des Führungsverhaltens, Abschn. 4.4.2.3). Auf Team-Ebene kommt eine Studie von Pentland (2012, S. 66) zu dem Ergebnis, dass erfolgreiche Call-Center-Teams dadurch gekennzeichnet sind, dass alle Teammitglieder ähnlich oft sprechen und zuhören, die kommunikativen Interaktionen eher kurz und nett sind, die Mitarbeiter sich auch über inoffizielle Kanäle untereinander und nicht nur mit der Führungskraft austauschen. Während sich die wissenschaftliche Debatte allerdings noch um die Frage dreht, wie man die Sprachstile und kommunikativen Interaktionen methodisch valide, objektiv und ethisch vertretbar messen kann (siehe Müller-Frommeyer et al. 2019), haben Anbieter wie Precire schon begonnen das Feld der Weiterbildung von und über Sprache zu beackern. Neben Trainings zur Analyse der verbalen Kommunikation und Verbesserung ihrer Wirkung bietet das Aachener Start-up auch einen E-Mail-Assistenten für die schriftliche Kommunikation an.

4.5.2.4 Was (noch) nicht geht

Noch sind Lerncoaches Zukunftsmusik und die Auswertungs- und Optimierungsmöglichkeiten von Sprache nur im Ansatz verstanden. In der Gegenwart finden AR/

VR-Trainings vor allem in den Gebieten statt, in denen **sensomotorische Fähigkeiten** gefragt sind, z. B. in der Produktion, Logistik, Industrie, Medizin, Feuerwehr und Armee (siehe Dörner et al. 2019a, S. 357 f.). Für wissensintensive Berufe, die komplexe Problemstellungen nicht nur bearbeiten, sondern als solche zunächst identifizieren und näher beschreiben müssen, eignen sie sich weniger.

Ähnlich wie bei Learning Analytics ist die **Technologie allein noch nicht hinreichend für den Lernerfolg**. Wer Wearables einsetzt, sollte sich bewusst machen, dass die Lernwirksamkeit von mehreren Faktoren beeinflusst wird (Niegemann und Heidig 2019, S. 11):

- von der Qualität der Information, die Lernende durch Interaktion mit der Lernumgebung gewinnen können,
- von der Qualität der kognitiven Operationen, die durch Interaktion mit der Lernumgebung initiiert werden,
- von Art und Ausmaß der Belastung des Arbeitsgedächtnisses der Lernenden während des Lernprozesses,
- vom aktivierten Vorwissen der jeweiligen Lernenden,
- von den in der jeweiligen Lernsituation realisierten selbstregulatorischen Fähigkeiten des jeweiligen Lernenden sowie
- von Persönlichkeitsmerkmalen, motivationalen und emotionalen Zuständen der Lernenden während des Lernprozesses.

Zwei Punkte muss man dieser ohnehin schon langen Liste noch hinzufügen: Der **Lernerfolg hängt auch vom didaktischen Konzept und der sozialen Beziehung** zu bzw. Interaktion mit den Lehrenden und Mitlernern ab (Hattie 2012). Das heißt, dass auch virtuelle Lernumgebungen und -szenarien didaktisch durchdacht werden müssen und Lernen nicht ohne Andere stattfindet. AR/VR bzw. Wearables mögen zwar adaptives Lernen ermöglichen, aber sie tun dies nur mit **Fokus auf einen einzelnen Lernenden** (auch wenn es Ausnahmen gibt, wie z. B. die VR Escape Games von EXIT-VR, sind diese auf ein paar wenige Teilnehmer limitiert). Das ist zum Großteil der Technologie geschuldet, da es aufwändig ist, wenn sich mehrere Lernende im selben virtuellen Raum bewegen sollen: Jede Teilnehmerin benötigt eine eigene Workstation mit Hochleistungskomponenten, eine Lizenz der VR-Software und zudem sind bei Systemen ohne Backpacks die Headsets per Kabel mit den Workstations verbunden, sodass man sich nicht frei bewegen kann. Eine andere Option, kabellos und im Team in eine virtuelle Welt einzutauchen, bieten CAVEs (Cave Automatic Virtual Environment): Bei dieser Technologie werden 3D-Bilder auf eine oder mehrere Wände eines Raums projiziert, die Bewegungen der Nutzer im Inneren der „Höhle“ werden registriert und

die virtuelle Welt wird entsprechend verändert (Hafen 2018, S. 28). Dass damit die Wahrnehmung der Lernsituation eine andere ist, liegt auf der Hand. Offen ist, wie sich das auf das Lernerleben und den Lernerfolg auswirkt (siehe Dörner et al. 2019a, S. 259).

Weitgehend unerforscht sind **moralische, ethische und rechtliche Herausforderungen**, die durch die Nutzung von VR und Wearables entstehen. Klar ist, dass die Nutzer sich nicht nur als Beobachter, sondern als Teil der virtuellen Welt verstehen, womit sich die Erlebnisse potenziell stärker auf Psyche und Selbstmodell, Gefühle und Verhaltensweisen auswirken (Dörner et al. 2019a, S. 256). Bereits durch die Wahl seines Avatars bestimmen die Lernenden Größe, Hautfarbe, Geschlecht und Körperfdimensionen, was **Vorurteile und stereotype Reaktionen** in der virtuellen Welt – und auch der realen – nach sich ziehen kann oder eine Form der Flucht aus der Realität ist (Dörner et al. 2019a, S. 256, 258). Man muss nicht direkt an Cybermobbing, unsittliche Übergriffe oder Gewalt in virtuellen Welten denken, um die Grenzen der Technologie bzw. die negativen Seiten, dessen, was sie ermöglicht, zu erkennen. Selbst scheinbar banale Wirkungen wie **Cybersickness**, d. h. physisches Unwohlsein, das sich vor allem aus inkonsistenten Sinneswahrnehmungen ergibt (Dörner und Steinicke 2019a, S. 68), sind Lernhindernisse. Weitgehend unerforscht sind die positiven wie negativen **Langzeitfolgen**, wenn Menschen immer mehr Zeit in virtuellen Welten verbringen, dort Kontakte pflegen und Erfahrungen machen, die sie in der realen Welt aus Angst, Scham oder man gelnden Möglichkeiten nicht machen (Dörner et al. 2019a, S. 259).

Bei der Kombination von AR/VR bzw. Wearables mit Learning Analytics kann es dazu kommen, dass sich die Lernenden **permanent überwacht** fühlen, was grundsätzlich die (Lern-)Leistung beeinträchtigen kann. Die Hauptgründe dafür sind, dass sich der Fokus der Aufmerksamkeit von der Tätigkeit auf die Darstellung der Tätigkeit nach Außen verschiebt und auf Kontrolle unter anderem mit Reaktanz geantwortet wird (siehe Abschn. 2.1). Für den Bereich des Lernens wurde bereits angemerkt, dass Selbstorganisation, Kreativität und Kollaboration mit Anderen sowie (risikofreies) Feedback durch Andere zentral sind (Abschn. 4.5.1.4). Mit anderen Worten: Vertrauen darauf, dass die gesammelten Daten nicht gegen einen verwendet werden, sowie eine psychologische Sicherheit in der (Lern-)Gruppe ermöglichen (Team-)Lernen (Edmondson 1999). Genau diesen beiden Faktoren läuft eine permanente Überwachung zuwider, zumal, wenn die Gefahr einer missbräuchlichen Nutzung der Daten nicht ausgeschlossen werden kann. Deshalb braucht es auch ein **mehrschichtiges Sicherheitssystem**: auf Ebene der Lernplattform bzw. der Lernumgebung, auf Ebene des Computer-Backends für die Analysen und auf Ebene der verwendeten Geräte (Yun et al. 2016, S. 35).

4.6 Personalbindung und -freisetzung

Personalbindung und -freisetzung sind sowohl in Zeiten knapper Fachkräfte als auch wirtschaftlicher Umbrüche **hoch relevante Themen für das Personalmanagement** (Kanning 2017, S. 2 f.). Bei der **Personalbindung** geht es darum, das Arbeitgebersversprechen im Alltag einzulösen und die Mitarbeitenden für die Arbeit im Unternehmen zu begeistern, sodass sie sich dem Arbeitgeber in hohem Maße verbunden fühlen. Ist das geschafft, so die dahinterstehende Logik, werden sich die Mitarbeitenden eher für die Ziele des Unternehmens **engagieren** und anstrengende Veränderungen akzeptieren sowie dem Unternehmen **treu bleiben**, selbst wenn sich attraktive Alternativen ergeben (Kanning 2017, S. 2 f.).

Es kann allerdings wirtschaftlich geboten und sozial angemessen sein, sich **aktiv von Mitarbeitenden zu trennen** oder auf **freiwillige Fluktuation** zu setzen. Dies gilt insbesondere in einem wirtschaftlichen Umbruch, wenn also neue Qualifikationen und Ideen gebraucht werden. Zudem kann damit gerechnet werden, dass mit dem neuen Wissen auch neue Energie („Schwung“) in die Organisation kommt, da die Neuen meist hoch motiviert sind. Allerdings setzt sich zurzeit der Gedanke durch, dass Personalbindungsmaßnahmen und internes Recruiting mehr Vor- als Nachteile bringen (Cappelli 2019, S. 51).

In ihrer gesamten thematischen Breite können beide Themenbereiche hier nicht behandelt werden. Stattdessen wird der Fokus auf den digitalen Tools zur **Vorhersage der Kündigungswahrscheinlichkeit** und der **Nachfolgeplanung** liegen.

4.6.1 Abschätzung der Kündigungswahrscheinlichkeit

4.6.1.1 Worum es geht

Die Vorhersage, wer bzw. welche Personengruppen am ehesten kündigen werden, ist ein immer wieder genanntes Anwendungsszenario von Predictive Analytics im Personalmanagement (z. B. Lee et al. 2017, S. 203; Tambe et al. 2019, S. 19, 26). Der Grund für die hohe Relevanz ist einfach: Aus Arbeitgebersicht will man entweder wissen, ob **leistungsstarke Mitarbeitende abwanderungsgefährdet** sind oder ob man **leistungsschwächere mit mehr oder weniger (monetärem) Aufwand loswerden** kann. Aus Sicht von Personalvermittlern (Head Hunter bzw. Executive Search oder Active Sourcern) ist die Kündigungswahrscheinlichkeit ein Indikator, mit welchem (monetären) Aufwand man eine Person zum Wechsel bewegen kann. Das Problem dabei ist nicht, dass es keine Forschungsergebnisse zu den Faktoren, die Kündigungen vorhersagen können, gibt. Das Problem ist, dass **zu viele mögliche Prädiktoren** identifiziert wurden: Alter, Entgelt, allgemeine Ar-

beitszufriedenheit, Zufriedenheit mit den Arbeitsbedingungen, Burnout, Anerkennung, Wachstumspotenzial, wahrgenommene Fairness am Arbeitsplatz, Dienstalter im Job, Teamgröße und -leistung, Geschlecht, ethnische Zugehörigkeit, Bildung, Familienstand, etc. (Holtom et al. 2008; Lee et al. 2017). Gleichzeitig haben die meisten Studien nur eine mäßige Vorhersagegüte, weil sie nur ca. 25 % der Kündigungsvarianz erklären (Lee et al. 2017, S. 205). Wichtig ist deshalb, dass eine Organisation für ihre **spezifischen Gegebenheiten** herausfindet, was zu freiwilligen Kündigungen führt, und dabei am besten Daten zu mehreren Faktoren über einen längeren Zeitraum analysiert. In den letzten Jahrzehnten wurden Verfahren entwickelt, die genau das versprechen.

4.6.1.2 Wie es geht

Da die Vorhersage über eine Klassenzugehörigkeit – kündigen oder nicht kündigen – getroffen werden soll, eignen sich prinzipiell **alle Klassifikationsalgorithmen** (Abschn. 3.4), insbesondere logistische Regression, Random Forest, Gradient Boosting Machines, Support Vector Machines, Naïve Bayes und k-Nearest Neighbors (Zhao et al. 2019; de Oliveira et al. 2019). Betrachtet man das Kündigungsrisiko nicht als dichotome, sondern stetige Variable, können **auch Regressionen** vorgenommen werden. Entscheidend ist deshalb eher die Frage, wie viele Merkmale (Features) in das Modell eingehen sollen, welcher Art diese sind (Personen- vs. Verhaltensdaten, nominale vs. ordinal vs. metrische, externe vs. interne) und welche Annahmen über die Verteilung der Variablen getroffen werden. Ein Beispiel für die Verwendung von internen Daten bietet die Berechnung der Kündigungs wahrscheinlichkeit bei Hewlett Packard (siehe Beispiel-Kasten).

Beispiel: Flight Risk Score bei Hewlett Packard

Hewlett Packard sah sich in einigen Bereichen mit hoher Fluktuation konfrontiert, insbesondere in der Abteilung, die die globale Vertriebsvergütung konzipiert und verantwortet, belief sich die Flukturationsrate auf rund 20 % pro Jahr (Siegel 2013, S. 48). Aufgrund der komplexen Tätigkeit in dieser Abteilung sind die Kosten für die Neubesetzung sehr hoch, weil man nicht nur geeignete Spezialisten finden muss, sondern diese auch noch in die komplizierten Abläufe und die vielschichtige Organisationsstruktur von Hewlett Packard eingearbeitet werden müssen.

In einem Pilotprojekt mit ca. 300 Mitarbeitenden sollte ein „Flight Risk Score“ entwickelt werden, also ein Wert, der die Kündigungs wahrscheinlichkeit eines Mitarbeitenden abbildet (Siegel 2013, S. 37). Zunächst wurden Vergangenheitsdaten zu Gehältern, Gehaltserhöhungen, Stelleneinstufungen und Jobwechsel über einen Zweijahreszeitraum gesammelt, um damit einen

ML-Algorithmus zu trainieren. Unter den 40 % der HP-Mitarbeitenden mit dem höchsten Flight Risk waren 75 % derjenigen, die tatsächlich kündigten. Die Vorhersagekraft des Modells war damit fast doppelt so hoch wie bei einer Zufallsauswahl, wenn man also raten würde (Siegel 2013, S. 49). Aus den Daten und dem Modell wurde deutlich, dass Beförderungen zwar das Risiko, dass ein Mitarbeiter kündigt, senkt – aber nur, wenn zeitgleich eine deutliche Gehaltserhöhung mit einhergeht. War dies nicht der Fall, führten häufige Beförderungen zu einer hohen Kündigungswahrscheinlichkeit. Positiv wirkte sich hingegen Job-Rotation aus, was wohl an den hochspezialisierten Tätigkeiten lag (Siegel 2013, 49 f.).

Ein Wert für die Kündigungswahrscheinlichkeit kann natürlich einiges an Sprengkraft beinhalten, zumindest ist er aber immer fehlerbehaftet und muss deshalb mit Vorsicht verwendet werden. Aus diesem Grund wurden die Reports mit den Flight Risk Scores nur an drei Manager aus dem globalen Vergütungsteam verteilt und diese wurden in der Interpretation des Wertes, samt seiner Grenzen, geschult. Über die Schulung hinaus wurde auf eine leicht interpretierbare Darstellung geachtet. Bei HP konnten alles in allem positive Wirkungen erzielt werden, insbesondere sank die Fluktuation auf 15 %, zumindest in manchen Regionen, und es ergab sich ein (rechnerisches) Einsparpotenzial von \$ 300 Millionen aufgrund von Einsparungen, wenn Mitarbeitende mit hohem Kündigungsrisiko zum Bleiben bewegt werden können. Sowohl der behutsame Umgang mit den Flight Risk Scores als auch die Einsparpotenziale waren wichtige Argumente dafür, dass der Score nach der Pilotphase auf die gesamte Belegschaft ausgedehnt und quartalsweise aktualisiert wurde (Siegel 2013, S. 50 f.).

Es gibt mehrere Verfahren, die eine solche Abschätzung leisten können. Welches sich am besten eignet, ist vor allem von den zur Verfügung stehenden Daten (kleiner vs. großer Datensatz, ausbalanciert vs. unausgeglichen, strukturierte vs. unstrukturierte Daten) abhängig. Aber auch Fragen der Machbarkeit (inklusive der Ausführungszeit für ein ML-Modell) und Nachvollziehbarkeit spielen eine Rolle. Deshalb gibt es keinen one-best-way. Allerdings hat sich immer wieder gezeigt, dass **Entscheidungsbäume und logistische Regressionen** einfach nachvollziehbar sind und zu guten Ergebnissen führen (Frierson und Si 2018, S. 255 f.; Zhao et al. 2019, S. 752 ff.).

Die grundsätzliche Logik von Entscheidungsbäumen wurde bereits erläutert (Abschn. 3.4 und Abb. 3.4). An dieser Stelle wird deshalb ein Beispiel für einen **Entscheidungsbaum auf Basis eines von IBM zur Verfügung gestellten Datensatzes** gezeigt. Der Datensatz, der auf Kaggle (2019) zu finden ist, enthält 1470 Datenpunkte (davon 237, die gekündigt haben) und 34 Merkmale. Für die Modellierung des Entscheidungsbaumes wurden alle Merkmale aus dem Datensatz entfernt, die entweder

für alle Datenzeilen den gleichen Wert enthielten (z. B. EmployeeCount) oder Ausprägungen, die zur eindeutigen Identifikation benutzt werden (z. B. die Employee-Number), sodass 28 Merkmale übrigblieben. Für die Berechnung wurde das Programm RapidMiner 9.5 benutzt und die maximale Tiefe des Baumes auf 5 limitiert. Mit großem Abstand ist das monatliche Einkommen der Faktor, der am besten zwischen kündigen („leaves“) und nicht kündigen („stays“) diskriminiert (siehe Abb. 4.8). Die erste Abzweigung im Modell („MonthlyIncome = 1151“) führt zwar nur zu drei Datensätzen, zeigt aber, dass bei einem Gehalt von weniger als \$ 1151 eine Kündigung sehr wahrscheinlich ist. Eine ähnlich geringe Anzahl an Datensätzen, nämlich acht, liegen im Teilbaum „< 19,5“ Jahre. Bei solchen Werten kann die Logik des Baumes hinterfragt und optimiert werden (auch wenn die Vorhersagegüte des Modells akzeptabel bis gut ist: accuracy = 82,9 %, precision = 83,1 %, recall = 99,7 %).

Bei einer einfachen **logistischen Regression** muss die abhängige Variable binär bzw. dichotom sein, also in eine von zwei Kategorien fallen: kündigen oder nicht kündigen (bei der multinominalen logistischen Regression kann die Zielvariable auch in mehrere Klassen fallen). Ein großer Vorteil der logistischen Regression ist, dass die unabhängigen Variablen, die die Kündigung vorhersagen sollen, beliebig skaliert sein können (nominale, ordinal, metrische). Für die Abschätzung der Kündigungswahrscheinlichkeit könnten deshalb sowohl die Merkmale Gehalt (metrisch), Zufriedenheit mit der Vorgesetzten (ordinal) oder Art des Jobs (nominal) benutzt werden siehe (Abschn. 3.4 und Abb. 3.3). Der Vorteil einer solchen Analyse ist, dass die **Merkmale nach ihrem Einfluss auf die Kündigungswahrscheinlichkeit gewichtet** werden (durch Angabe der β -Koeffizienten) und so eine detaillierte Aussage darüber möglich ist, welche Faktoren mehr oder weniger relevant für eine Kündigung sind. Erinnert sei aber nochmals daran, dass die **Gewichtungskoeffizienten nicht linear interpretiert** werden dürfen (anders als bei der

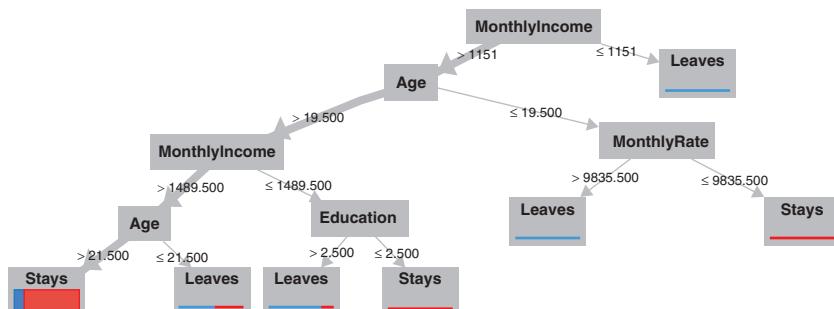


Abb. 4.8 Entscheidungsbaum zur Kündigungsvorhersage (eigene Erstellung mit RapidMiner 9.5)

linearen Regression, wo die Änderung von x um eine Einheit zu einer Änderung von y um den Faktor β führt). Meist werden deshalb Effektkoeffizienten angegeben, bei denen Werte kleiner als 1 einen negativen Einfluss und Werte größer als 1 einen positiven Einfluss auf die Odds, d. h. das Verhältnis von $P(Y = 1)$ zur Gegenwahrscheinlichkeit $1 - P(Y = 1)$, haben (Backhaus et al. 2016, S. 308).

Die meisten Softwareanbieter haben die Berechnung der Kündigungs wahrscheinlichkeit als Funktion in ihre Workforce-Analytics-Module integriert (z. B. Claro, Visier, Workday, SAP SuccessFactors, Talentwunder). Damit lassen sich interne Daten für die Berechnung des Kündigungsrisikos leicht verwerten. Bei SAP SuccessFactors gehen beispielsweise vor **allem personenbezogene und firmeninterne Merkmale** in das „Employee Flight Risk“ ein (Shih 2017a):

- Demografie (Alter, Geschlecht, Behinderung, ethnische Zugehörigkeit usw.),
- Vergütung (Gehalt, Aktienoptionen usw.),
- Berufserfahrung (Besoldungsgruppe, Dauer der Organisationszugehörigkeit, Zeit auf der Position usw.),
- Entwicklung (Schlüsselposition, Leistungsbewertung, Potenzialbewertung usw.),
- Beschäftigung (Stellenkategorie, Mitarbeiterklasse, Beschäftigungsniveau, Note usw.),
- Nachfolge (kritischer Arbeitsplatz, Nachfolgebewertung, Nachfolgebereitschaft usw.)

Die Auswertung erfolgt über Algorithmen, die in der SAP Automated Predictive Library verfügbar sind, insbesondere eine Ridge-Regression und weitere Funktionen zur automatischen Extraktion von Vergangenheitsdaten (über einen Dreimonthszeitraum) und zur Modelloptimierung (Shih 2017b). **Die Ridge-Regression ist eine Variante von Regressionsanalysen**, die, insbesondere in Kombination mit einer strukturellen Risikominimierung (Structural Risk Minimization), Über- und Unteranpassung vermeidet, wenn viele Merkmale und/oder wenige Trainingsinstanzen verwendet werden (Provost und Fawcett 2017, S. 173). Mit ihr kann man auf Basis der Merkmale (Vergütung, Berufserfahrung, etc.) z. B. berechnen, wie viele Tage es noch dauert, bis ein Mitarbeitender kündigt. Verwendet man diese Information, um ein paar Wochen vorher auf die Person zuzugehen, könnte sich eine Kündigung vermeiden lassen. Da sich statistische Verfahren aber nie auf den konkreten Einzelfall beziehen können, bleibt ein Rest an Unsicherheit, ob die auf Basis der gewichteten Treiber vorgeschlagenen Maßnahmen wirken. Allerdings: **Ohne HR Analytics** muss die HR-Abteilung das Kündigungsrisiko von Mitarbeitenden entweder erraten oder sich auf die **Einschätzung der jeweiligen Führungskraft verlassen**. Welche Variante die akkurate, schnellere und billigere Prognose liefert, muss im Einzelfall geprüft werden. Dazu sollten die Perfor-

mance-Parameter des ML-Modells mit dem Anteil vergangener Fehleinschätzungen einer Führungskraft verglichen werden (über Ersteres liegen Daten immer vor, über Zweiteres meist nur vereinzelt). Laut SAP-Angaben kann mit ihrem ML-Modell bereits beim Training mit nur 30 % der Daten eine Verbesserung von 167 % gegenüber der reinen Zufallseinschätzung erreicht werden (Shih 2017b). Zudem sollte eine Kosten-Nutzen-Betrachtung stattfinden: Wie viel (monetärer) Aufwand muss betrieben werden, um abwanderungsbereite Personen zum Bleiben zu bewegen? Nimmt man z. B. an, dass Gehaltserhöhungen der wichtigste Treiber für die Minimierung des Kündigungsrisikos sind, dann kann für ein gegebenes Budget berechnet werden, wie viel man für welche Personen ausgeben sollte. In einer vereinfachten Variante werden nur die Gesamtkosten, die durch eine Kündigung verursacht werden, abgeschätzt (siehe das Modell von Predict42 im Beispiel-Kasten).

Beispiel: Kalkulation von Kündigungsrisiko und -kosten mit Predict42

(<https://www.predict42.com/>)

Die Gründer des Start-ups Predict42 übertragen ihre Expertise aus dem Online-Marketing-Umfeld auf Themen der Personalarbeit. Neben der Berechnung der Kündigungsreichweite („Employees at Risk“) kalkulieren sie in ihren Modellen auch die Kosten, die normalerweise mit einer Kündigung einhergehen („Cost of Attrition“). Werden diese Daten noch in Verbindung zur Mitarbeiterleistung gesetzt, die man z. B. über die durchschnittliche Zielerreichung („Target Achievement“) messen kann, so lässt sich erkennen, wie viele leistungsstarke Mitarbeiter kündigungsbereit sind, was ein Unternehmen natürlich bedauern würde (siehe Abb. 4.9).

Für die Abschätzung des Kündigungsrisikos bei einem großen Beratungsunternehmen wurden mehrere Modelltypen mit unterschiedlichen ML-Softwarepaketen gegeneinander getestet: Random Forest, Logistic Regression und Gradient Boosting. Letzteres hatte die besten Ergebnisse hinsichtlich der klassischen Gütekriterien (siehe Abschn. 3.4). Da eine bloße Abschätzung der Kündigungsreichweite noch zu wenig Information für eine Entscheidung bietet, wurden in einem zweiten Schritt die Kosten einer Kündigung berechnet. Mit dieser Information haben Führungskräfte und HR-Professionals eine bessere Datengrundlage für die Entscheidung, welche Maßnahmen getroffen werden. Um die Kosten zu kalkulieren wurde auf Durchschnittswerte, wie sie z. B. von der Society of Human Resource Management angegeben werden, zurückgegriffen. Demnach verursacht die Wiederbesetzung einer Führungskraft Kosten in Höhe von 250 % der Jahresgesamtvergütung der Position, während es bei Einsteiger-Jobs „nur“ 50 % der jährlichen Bezüge sind.

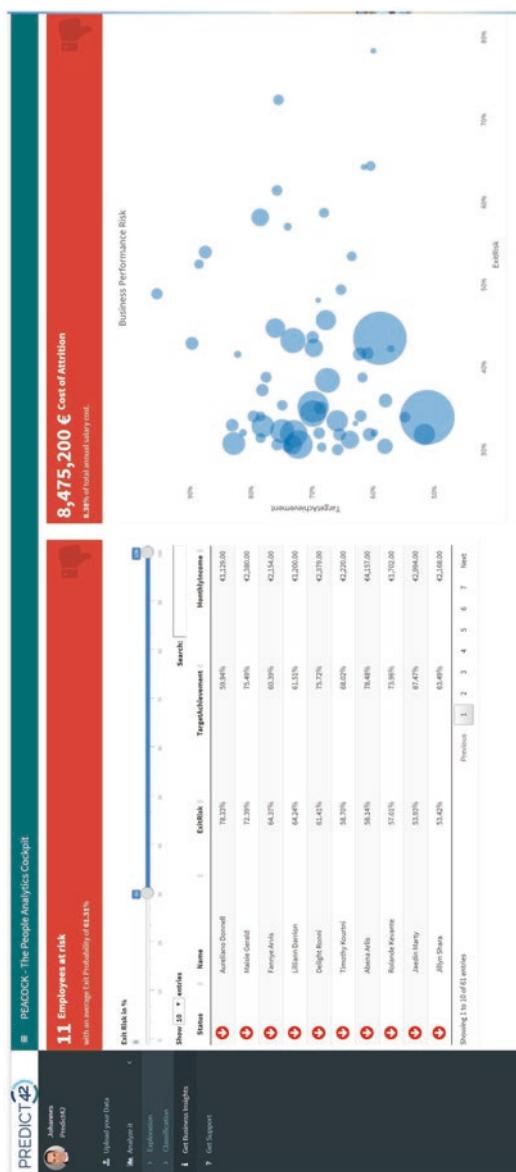


Abb. 4.9 Abschätzung des Kündigungsrisikos und der -kosten (Predict42)

Ein Ansatz, der auch **externe Daten** berücksichtigt, ist die Berechnung des **Kündigungsrisikos auf Basis von Social Media-Verhaltensdaten**. Die Idee ist, Daten zu sammeln, die auf die Suche nach einem neuen Job hindeuten – und da die Jobsuche ein emotionales Ereignis ist, geht sie mit gesteigerter Aktivität in den sozialen Medien einher. Als Quelle dienen die öffentlich zugänglichen Daten aus sozialen Medien, z. B., bei welchen Themen der User Hashtags auf Twitter setzt, wann und wie oft der Lebenslauf auf LinkedIn oder Xing aktualisiert wurde (z. B. Änderung des Job-Seeker-Status, Hinzufügen neuer Qualifikationen oder Bestätigungen/Empfehlungen von anderen Mitgliedern), welchen Inhalten die Nutzerin folgt oder sie liked (z. B. andere Jobs, Firmen oder Personalvermittler) oder wie häufig die Nutzerin mit Karriereseiten interagiert. Ein Anbieter, der solche Daten verwertet, ist Joberate (die seit 2018 als Claro.hr firmieren). Sie berechnen mit geheim gehaltenen ML-Algorithmen einen sogenannten J-Score, der in Echtzeit die Social Media-Aktivitäten auswertet (Joberate 2016). Die Daten über diese Aktivitäten werden von sogenannten Social Data Aggregatoren wie GNIP und Hootsuite gekauft (diese Applikationen ermöglichen es Nutzern, ihre Social Media-Aktivitäten über eine Oberfläche zu managen, also z. B. den Kundendienst über Twitter, die Marketing-Kampagne auf Instagram, die Stellenbewerbung auf LinkedIn oder die Kommunikation mit Facebook-Fans zu verwalten – die dabei entstehenden Daten verkaufen sie weiter an Dritte). Der J-Score reicht von 4 bis max. 70, wobei 4 keine Aktivitäten in Richtung Jobwechsel und damit kein Kündigungsrisiko bedeutet, während ein Wert von 70 auf aktive Jobsuche hinweist. Aus Gründen der Übersichtlichkeit werden drei Kategorien gebildet: Citizens (J-Score von 4–36), Tourists (37–59) und Migrants (60–70) (Joberate 2016, S. 2). Ein deutscher Anbieter, der einen ähnlichen Score aus Daten von über 75 **Business- und Social-Media-Netzwerken** berechnet, ist Talentwunder: Als Tool für Active Sourcing zeigt es zu jedem Kandidaten dessen Wechselbereitschaft an (Abschn. 4.2.2.2). Berechnet wird der Wert aus Daten wie vergangenen Arbeitsplatzwechseln, Profil-updates, der Entfernung zwischen Wohn- und Arbeitsort sowie dem Job Seeker-Status (Dittes 2018, S. 219). Genau wie beim J-Score ist auch bei Talentwunder nicht öffentlich bekannt, welcher ML-Algorithmus verwendet wird.

Ein ganz anderer Ansatz zur Abschätzung des Kündigungsrisikos ist die **Auswertung von E-Mails**, um darauf basierend die **sozialen Netzwerke** und die **Stimmungslage** innerhalb eines Unternehmens zu identifizieren. Entsprechende Studien wurden in den letzten Jahren von einer Forschungsgruppe, die ihren Ursprung am MIT hat, vorgelegt (Gloor et al. 2017; de Oliveira et al. 2019; siehe auch Höller und Wedde 2018). Im Mittelpunkt steht die Frage, was das E-Mail-Kommunikationsverhalten über das Kündigungsrisiko von Managern aussagt.

Zum Einsatz kamen neben der deskriptiven Analyse von Netzwerkmetriken (siehe Abschn. 4.4.2.2) noch ein Naïve Bayes-Algorithmus zum Klassifizieren der Sentsments sowie eine logistische Regression zur Überprüfung der Faktoren, die die Kündigung am besten vorhersagen (Gloor et al. 2017, S. 346 f.). Statt der logistischen Regression wurde in der zweiten Studie ein KNN verwendet, das zeitlich verteilte Daten berücksichtigen kann. Diese sogenannten Gated Recurrent Units (GRU) sind eine Spielart von LSTM-Netzwerken und stammen damit aus der Klasse der Recurrent Neural Networks (de Oliveira et al. 2019, S. 49).

In der ersten Studie untersuchten die Forscher die E-Mails von 866 Managern eines großen Dienstleistungsunternehmens über einen 18-monatigen Zeitraum hinweg, währenddessen 111 Manager das Unternehmen verließen (Gloor et al. 2017). Aus den Kommunikationsmustern wurde auf das soziale Netzwerk der Manager geschlossen und anhand von Netzwerkmerkmalen sowie inhaltlichen Indikatoren (z. B. Emotionalität und Komplexität der verwendeten Sprache) wurde ein Modell entwickelt, um die Kündigungswahrscheinlichkeit vorherzusagen. Ein besonderes Augenmerk wurde auf die 4–5 Monate, bevor die Manager gingen, gelegt. In deren Kommunikationsnetzwerk und -verhalten zeigten sich signifikante Unterschiede: Das Netzwerk war durch eine geringere Nähe (siehe Abschn. 4.4.2.2) und Responsivität gekennzeichnet, das heißt, ihre Verbindungen zu anderen Personen/Knoten im Netzwerk sind durchschnittlich länger (gehen also über andere Knoten), sie erhalten langsamer Rückmeldungen auf ihre E-Mails und müssen öfter nachfragen, bevor sie eine Antwort erhalten. Außerdem zeigte eine Sentiment-Analyse, dass ihre E-Mails von einem häufigeren Wechsel zwischen positiven und negativen Stimmungslagen gekennzeichnet sind (Gloor et al. 2017, S. 347). Zudem konnte ein Wechsel im Kommunikationsverhalten festgestellt werden: Fünf Monate bevor die jeweiligen Manager kündigten mussten sie andere öfter anschreiben, um eine Antwort zu erhalten. Über den Grund können die Forscher nur spekulieren, aber es könnte sein, dass diese Manager schon innerlich gekündigt hatten, dies in ihrem Verhalten (unbewusst) zeigten und deshalb von anderen eher gemieden wurden (Gloor et al. 2017, S. 348).

Auch in der zweiten Studie wurden Netzwerkanalysen auf Basis der E-Mail-Kommunikation durchgeführt. Sie unterscheidet sich aber in zweierlei Hinsicht von der ersten Studie: Es wurden die E-Mails von 3952 Managern über einen 12-Monatszeitraum analysiert und durch den Einsatz des GRU-Netzes kam ein Klassifizierungsalgorithmus zum Einsatz, der Daten im Zeitverlauf erfassen kann. Das bedeutet, dass **Änderungen in der Beziehung zwischen Netzwerkmetriken und der Kündigungswahrscheinlichkeit über die Zeit hinweg ausgewertet** werden können (die Daten der 12 Monate wurden in Zeitfenster à 15 Tage unterteilt). Das ist ein großer Vorteil gegenüber den anderen bereits beschrie-

benen Verfahren und kommt der Realität viel näher, weil sich die Kündigungsbelegschaft und ihre Gründe im Zeitverlauf ändern können (de Oliveira et al. 2019, S. 44, 56). Das Modell erreichte eine akzeptable bis gute Vorhersagegüte, gemessen an den klassischen Werten wie *accuracy* (richtig zugeordnete Daten: 80,0 %), *precision* (richtig positiv zugeordnete Daten dividiert durch die Summe aus falsch und richtig Positiven: 74,5 %) und *recall* (richtig Positive dividiert durch die Summe aus richtig Positiven und falsch Negativen: 84,4 %) (de Oliveira et al. 2019, S. 43, 54). Verbesserungen bestehen einerseits hinsichtlich des Algorithmus, der nur aus einer Schicht und einmal aus 8 und einmal aus 16 Neuronen bestand, was noch kein Deep Learning im ursprünglichen Sinn erlaubt. Außerdem war der Datensatz nicht ausgewogen, weil vergleichsweise wenige Manager tatsächlich kündigten (78 der 3952). Weitere detaillierte Fehleranalysen stehen noch aus. Dennoch präsentiert die Studie einen interessanten Ansatz, der vor allem die Dynamik der Faktoren für freiwillige Kündigungen berücksichtigt und eine netzwerkbasierte Analyse für die Monate vor der Kündigung liefert. Personalverantwortliche können damit ein **Frühwarnsystem für die Mitarbeiterfluktuation** aufbauen (siehe aber Abschn. 4.6.1.4). Wenn HR-Professionals in ihrer Organisation beobachten, dass eine Managerin plötzlich ein kleineres Netzwerk hat (niedrige Degree Centrality), ihre Nähe zu anderen drastisch sinkt (niedrige Closeness Centrality), ihre E-Mails kürzer werden und sie E-Mails langsamer beantwortet, dann könnten sie daraus ableiten, dass die Managerin in 5–7 Monaten wahrscheinlich kündigen wird. Ob ihr dann Optionen angeboten werden und welche das sein sollen, kann der Algorithmus jedoch nicht vorhersagen.

4.6.1.3 Was man noch machen kann

Bei der Personalbindung bestehen viele Parallelen und Verbindungen zur Personalrekrutierung. Zunächst sind die Vorzeichen bei bestehenden Beschäftigungsverhältnissen nur umgedreht: Während Recruiter Mitarbeitende (ab-)werben wollen, versucht man sie beim sogenannten Retention Management zu halten. Deshalb können die oben beschriebenen Vorgehensweisen bzw. ML-Algorithmen auch beim **Active Sourcing** eingesetzt werden. Außerdem sollten Unternehmen zur Vermeidung von Fluktuation und den damit entstehenden Aufwänden bereits bei der Rekrutierung ansetzen: Je passgenauer die neuen Mitarbeitenden angezogen und ausgewählt werden, desto geringer ist die Gefahr der Enttäuschung (bzw. desto höher ist die Zufriedenheit) und desto höher die Bindung (Allen et al. 2010, S. 55 f.; Kanning 2017, S. 207 f.). Das wiederum unterstreicht die Relevanz von **Matching-Verfahren** im Recruiting (Abschn. 4.2.2).

Mit einer datenbasierten Auswertung der Kündigungstreiber lassen sich **tradiertere Wahrnehmungsmuster hinterfragen**. Sowohl Studien von Wissenschaft-

lern als auch Praktikern haben gezeigt, dass Personaler nicht wirklich wissen, was die Mitarbeitenden wertschätzen. Insbesondere unterschätzen HR-Professionals die Relevanz von monetären Aspekten und überschätzen weiche Faktoren: Während Personaler denken, dass Arbeitgeberwechsel vor allem durch Faktoren wie Arbeitsklima, verantwortungsvolle Aufgaben und Selbstverwirklichung motiviert sind (Staufenbiel und Kienbaum 2016), steht für Mitarbeitende eine bessere Bezahlung (Gehalt und Boni) auf Platz 1 (EY 2016a). Zu ähnlichen Ergebnissen kommen wissenschaftliche Studien (de Vos und Meganc 2009; Grund 2009). Beispielsweise untersuchte Grund (2009) fast 5000 freiwillige Kündigungen (also tatsächliche Arbeitsplatzwechsel und nicht nur abgefragte Präferenzen wie in vielen anderen Studien) und kam zu dem Ergebnis, dass das Gehalt der wichtigste Faktor für die Kündigung ist. Mit den oben beschriebenen ML-Algorithmen können **für das eigene Unternehmen die relevanten Wirkungsbeziehungen identifiziert werden**, zumindest auf Basis von Vergangenheitsdaten.

4.6.1.4 Was (noch) nicht geht

Der Veränderungswunsch ergibt sich aus einer wahrgenommenen Arbeitsunzufriedenheit und **der Realisierbarkeit des Wunsches aus den sich bietenden Jobalternativen** – das haben March und Simon (1958, S. 112 ff.) schon vor über 60 Jahren festgestellt. Insofern greifen alle Ansätze, die nur interne Daten verwenden, zu kurz, weil sie keine Informationen darüber berücksichtigen, welche Stellen zurzeit ausgeschrieben und welche Fähigkeiten besonders gefragt sind. In diesen Kontext gehört auch die Feststellung, dass Mitarbeitende **nicht nur von etwas weg-, sondern auch zu etwas hin gehen**. Das bedeutet, dass HR-Professionals zwar das Kündigungsrisiko richtig eingeschätzt haben mögen und auf die entsprechende Mitarbeiterin zugegangen sind, diese aber nicht zum Bleiben zu bewegen ist, weil die Alternative in zu vielen Aspekten verlockender erscheint. An dieser Stelle können Algorithmen logischerweise nicht weiterhelfen.

Ihre unterstützende Wirkung können algorithmische Modelle aber schon früher im Retention-Prozess verlieren, nämlich dann, wenn herausgefunden werden soll, was der Mitarbeiterin überhaupt angeboten werden soll. Zwar gewichten die meisten Verfahren unterschiedliche Faktoren und geben damit Anhaltspunkte für Bleibeangebote, aber **im Einzelfall können ganz andere, individuelle Beweggründe** vorliegen, die sich nicht mit den statistisch abgeschätzten decken (siehe Siegel 2013, S. 50). Erschwerend kommt hinzu, dass die Motive und damit die Kündigungswahrscheinlichkeit keine statischen Größen sind, sondern sich **im Zeitverlauf ändern**. Hierfür braucht es dann Verfahren, die Daten entlang der Zeitachse berücksichtigen können, wie z. B. das beschriebene GRU-Netzwerk. Diese Ansätze sind aber noch nicht der Standard. Selbst wenn die Dynamik der

Kündigungsbereitschaft berücksichtigt wird, bleiben **andere Optionen als zu kündigen** typischerweise ausgeblendet. Das heißt, es wird vom Vorliegen bestimmter Merkmale auf die Handlung „kündigen“ geschlossen, ohne andere Verhaltensweisen im Modell mit abzubilden (z. B. könnten die Mitarbeitenden ihre Beiträge bzw. ihren Arbeitsaufwand senken oder versuchen ihre Situation zu verbessern, anstatt zu kündigen de Oliveira et al. 2019, S. 56).

Bei der Datensammlung und -auswertung ist darauf zu achten, **welche Daten berücksichtigt werden**. Mit Blick auf das oben beschriebene Beispiel von SAP wird beispielsweise deutlich, dass zwar eine Reihe von personenbezogenen Daten in das Modell eingehen, aber keine Daten über **(sozial-)psychologische Konstrukte wie Arbeitszufriedenheit, Commitment oder soziale Einbindung**, obwohl diese zu den klassischen Treibern von Kündigungen zählen (Holtom et al. 2008, S. 234, 237; Lee et al. 2017, S. 202). Die Vermutung liegt nahe, dass man solche Daten eigens hätte erheben müssen, während man die anderen einfach aus HR-Systemen ziehen kann. Allerdings sollte nicht das Vorhandensein von Daten das Analysevorgehen bestimmen. Selbst wenn (sozial-)psychologische oder emotionale Faktoren erhoben werden, ist fraglich, wie valide dies ist. Beispielsweise muss das Konstrukt Arbeitszufriedenheit in Zahlen bzw. abgestufte Kategorien (z. B. 1–5) umgewandelt werden, um in die Berechnungen einfließen zu können. Was für den einen Mitarbeiter aber eine 3 bedeutet, wird höchstwahrscheinlich nicht damit übereinstimmen, was für eine andere Mitarbeiterin eine 3 bedeutet. Dieses Problem existiert allerdings bei allen quantitativen Erhebungsmethoden, unabhängig von der Datenmenge und dem statistischen Verfahren.

In diesem Kontext ist auch das **Aufwand-Nutzen-Verhältnis** zu diskutieren. Zwar liefern ML-Algorithmen auch bei kleinen Datensätzen akzeptable Ergebnisse und die Größe an sich hat nur einen kleinen Einfluss (Zhao et al. 2019, S. 747, 749), aber manche Informationen können auch **einfacher und präziser durch Gespräche** gewonnen werden (entweder durch regelmäßige Mitarbeitergespräche oder Austrittsinterviews). Gerade (sozial-)psychologische Faktoren können Menschen mit normal ausgeprägten Empathiewerten schneller und einfacher erfassen als Maschinen (siehe Gärtner 2017). Zwar lassen sich Text-Mining-Verfahren zur Analyse von E-Mails, Intranet-Posts oder Exit-Interviews nutzen, aber noch sind die Ergebnisse zu allgemein, weil entweder nur Sentiments oder allgemeine Topics identifiziert werden (siehe Abschn. 4.2.1.4).

Gegen die Berechnung des Kündigungsrisikos können **datenschutzrechtliche und ethische Gründe** sprechen (siehe Kap. 6). Insbesondere die Auswertung von Kommunikationsdaten ist kritisch zu sehen. Selbst dann, wenn sie rückwirkend, anonymisiert und mit einem konkreten Zweck verbunden geschieht, weil der so trainierte Algorithmus ja doch auf die aktuellen Mitarbeitenden angewandt werden

muss und damit **personenbezogene Daten produziert**. Zudem sollten für das Training des ML-Algorithmus datenschutzkritische Merkmale herausgenommen werden (Alter, Geschlecht, Ethnie), was wiederum zu realitätsferneren Vorhersagen führen kann (insbesondere Alter und Geschlecht haben sich immer wieder als relevante Einflussfaktoren für freiwillige Kündigungen herausgestellt; Holtom et al. 2008, S. 237, 249).

4.6.2 Nachfolgeplanung

4.6.2.1 Worum es geht

Die Nachfolgeplanung ist ein Element der strategischen Personalplanung und -entwicklung, bei dem die Fluktuationsvorhersage als Information genutzt wird. Sofern Organisationen eine systematische Nachfolgeplanung betreiben, fokussiert sich diese meist auf Top-Management-Positionen oder vereinzelte Schlüsselpersonen und beruht auf **Nominierungsprozessen**: Entweder erfolgt die Nominierung von potenziellen Nachfolgekandidaten durch hierarchisch höher gestellte Positionen, die HR-Abteilung oder durch Selbstnominierung. Durch den **Einsatz digitaler Tools** lässt sich einerseits die Datenbasis für die Planung verbreitern und andererseits können die Analyseergebnisse mit Maßnahmen der anderen Personalteifunktionen verzahnt werden. Deshalb ist die Nachfolgeplanung als Einzelmaßnahme kaum sinnvoll, sondern immer **eingebettet in die Planung, Entwicklung, Vergütung und Bindung** von Talenten – weshalb man sie oft im Zusammenhang mit dem Talent Management bringt (siehe Hoeffler et al. 2015, S. 201).

4.6.2.2 Wie es geht

Nachfolgeplanungsfunktionen sind meist in den Applikationen der bekannten Softwarehäuser integriert. Oft bieten sie neben der qualifikatorischen Eignung und Potenzialeinschätzung (durch die Führungskraft) noch die Möglichkeit die Kandidaten in einem Pool zu verwalten und anhand eines (Echtzeit-)Organigramms vakante Stellen und mögliche Karrierepfade aufzuzeigen. Weitere Informationen beziehen sich auf die Anzahl unbesetzter Schlüsselpositionen sowie die Bereitschaft des jeweiligen Mitarbeitenden eine Nachfolge anzutreten. Meist wird über ein Dashboard eine vergleichende Übersicht aller potenziellen Nachfolgekandidaten angeboten, die **Informationen über Anzahl der Kandidaten, Qualifikation, Potenzial, Nachfolge- und Abwanderungswahrscheinlichkeit sowie die allgemeine Fluktuationsrate** von Talenten enthält (Hoeffler et al. 2015, S. 201 f.; siehe Beispiel-Kasten). Sollten Lücken zum geplanten Bedarf offenbar werden, so können Maßnahmen aus den anderen Personalteifunktionen angestoßen werden: Melden eines Rekrutierungsbedarfs, Anbieten von Weiterbildungsmaßnahmen oder

Anreizen zur Erhöhung (Senkung) der Bleibewahrscheinlichkeit (Kündigungs- wahrscheinlichkeit). Während für die meisten nachfolgerelevanten Informationen deskriptive Analysen ausreichen, bedarf es für die Berechnung der Kündigungs- wahrscheinlichkeit und Empfehlung von Maßnahmen prädiktiver Analysen (siehe Abschn. 4.4.2, 4.5.1 und 4.6.1).

Beispiel: Succession-Risk-Management-App von Siemens (https://www.youtube.com/watch?v=hpy6I_dL6kE&feature=youtu.be)

Siemens hat eine App für Tablets und Smartphones entwickelt, die das Nachfolgemanagement erleichtern soll. Die sogenannte Succession-Risk-Management-App zielt auf einen kleineren Nutzerkreis ab: Vorstandsmitglieder, Führungskräfte auf der 1. und 2. Ebene, HR Business Partner, HR Pipeline & Placement. In der App wird für alle Geschäftsfelder des Unternehmens aufgezeigt, welche Stellen demnächst neu besetzt werden müssen. Das Top-Management kann zu jeder Zeit den aktuellen Status der Talente-Pipeline für kritische Funktionen nachvollziehen und bearbeiten.

Klickt man auf eine bestimmte Person/Position, werden die potenziellen Nachfolger aus dem aktuellen Pool angezeigt. Neue Kandidaten lassen sich einfach zur bestehenden Nachfolgeplanung hinzufügen und der Status der erfassten Kandidaten kann entlang von drei Stufen angepasst werden: Ready Now, Ready Future, Candidate Ideas. Jede vorgenommene Änderung wird dabei in einem Change-Log angezeigt und ist für alle anderen User sichtbar. Dadurch entsteht eine erhöhte Transparenz in einem Bereich, der früher traditionell weniger transparent war.

Über die Ausführungen in den genannten Abschnitten hinaus, möchte ich auf die Anwendung einer Klasse von ML-Algorithmen verweisen, die noch kaum besprochen wurde: **Clustering-Verfahren**. Mit solchen nicht überwachten ML-Algorithmen können automatisch Gruppen von Mitarbeitenden segmentiert werden und zwar auf Basis der Ähnlichkeit von Anforderungs- und Fähigkeitsprofilen. Als Inputdaten für die Nachfolgeplanung dienen einerseits Stellenbeschreibungen oder Anforderungsprofile und andererseits die Fähigkeiten der Kandidaten. Durch Clustering-Algorithmen wie DBSCAN werden nun die Anforderungs- und Fähigkeitsprofile hinsichtlich ihrer Ähnlichkeit gruppiert. Dadurch entstehen zweielementige Kombinationen von Anforderungs- und Fähigkeitsprofilen und alle Kombinationen innerhalb einer identifizierten Gruppe stellen Kandidat-Position-Matches dar und sind damit Handlungsalternativen für das Nachfolgemanagement (Piazza 2010, S. 79, 181 f.). Ein Vorteil eines solchen Clustering-Verfahrens ist, dass eine Reihe potenziell passender Matches präsentiert wird und nicht nur eine Kandidatin für eine Stelle. Dies kommt einer Nachfolgeplanung, die in **Szenarien** denkt, entgegen.

Ein Nachteil ist: Um diese algorithmenbasierte Identifikation durchführen zu können, bedarf es nicht nur der Informationen über Anforderungen und Fähigkeiten. Diese müssen auch noch in **strukturgleichen Datenobjekten** vorliegen, d. h. die beiden Profile müssen durch die gleichen Merkmale beschrieben werden (Piazza 2010, S. 182). Gleiches gilt, wenn weitere Informationen (z. B. Präferenzen oder Wechselbereitschaften der Kandidaten) in eines der Profile aufgenommen werden sollen.

4.6.2.3 Was man noch machen kann

Da sich wie bereits erwähnt weitere Personalinstrumente an die Nachfolgeplanung anschließen, sind auch prinzipiell alle der bisher genannten Tools anwendbar. Ein Beispiel ist Textkernel (Abschn. 4.2.2.2): Zwar geht es bei der angebotenen Dienstleistung vornehmlich um Recruiting, jedoch können die Text-Mining-Verfahren auch für das automatische Auslesen und Matchen von Anforderungs- und Stellenprofilen im Rahmen des Nachfolgemanagements sowie beim **Offboarding und Exit Management** benutzt werden. Für Letzteres werden die Fähigkeitsprofile einfach nicht mit internen, sondern externen Stellenvakanzen abgeglichen.

4.6.2.4 Was (noch) nicht geht

Cluster-Verfahren sind darauf ausgerichtet, innerhalb von Gruppen möglichst ähnliche Daten zu sammeln und zwischen den Gruppen einen möglichst großen Abstand zu erzeugen, das heißt, die Gruppen sind sich unähnlich. Beim Clustern von Anforderungs- und Fähigkeitsprofilen handelt es sich um in sich gemischte Gruppen, weil sie zwei verschiedene Arten von Objekten enthalten. Dann ist es allerdings denkbar, dass ein einzelnes Fähigkeitsprofil über mehr Ähnlichkeit zu einem Anforderungsprofil einer anderen Gruppe als zu jenen der gleichen Gruppe verfügt und somit die Anwendung von Cluster-Algorithmen zu einer **suboptimalen Nachfolgeplanung** führt (Piazza 2010, S. 184).

Die Bildung von Clustern wird auch beim **Text-Mining** verwendet. Beispielsweise hat Starbucks die über 5000 Kommentare aktueller und ehemaliger Mitarbeitender auf Glassdoor.com analysieren lassen, um die Gründe für die hohe Personalfliktuation herauszufinden (Sakellariadis 2015). Ziel war es, die Narrative zu identifizieren, die in den Bewertungen auftauchen. Die Texte wurden also auf (un-)ähnliche Erzählungen hin gruppiert, die von „Starbucks ist der Stern“ über „Glorifiziertes Fastfood“ bis „Teilzeit-Pariah“ reichen. Wichtig sind hier jedoch weniger die Cluster an Narrativen, sondern vielmehr die einzelnen Topics (z. B. kostenlose Getränke für die Belegschaft, intransparente Beförderungspraxis durch das Mittelmanagement, schlechte Karriereoptionen; Sakellariadis 2015), weil nur sie Handlungsorientierung geben.

Literatur

- Abel, J., Klohs, K., Lehmann, H., & Lantow, B. (2017). Sentiment-analysis for German employer reviews. In W. Abramowicz (Hrsg.), *International conference on business information systems* (Bd. 303, S. 37–48). Cham: Springer.
- Abele, E., Metternich, J., Tisch, M., Chryssolouris, G., et al. (2015). Learning factories for research, education, and training. *Procedia CIRP*, 32, 1–6.
- Acharya, A., & Sinha, D. (2014). Early prediction of students performance using machine learning techniques. *International Journal of Computer Applications*, 107(1), 37–43.
- Ajunwa, I., Crawford, K., & Schultz, J. (2017). Limitless worker surveillance. *California Law Review*, 105, 735–776.
- Albrecht, H. (2020). Besser als in echt. *Die Zeit*, Nr. 2, S. 33.
- Allen, D. G., Bryant, P. C., & Vardaman, J. M. (2010). Retaining talent: Replacing misconceptions with evidence-based strategies. *The Academy of Management Perspectives*, 24(2), 48–64.
- Antoni, C. H., & Syrek, C. (2017). Digitalisierung der Arbeit: Konsequenzen für Führung und Zusammenarbeit. *Gruppe. Interaktion. Organisation. Zeitschrift für Angewandte Organisationspsychologie*, 48(4), 247–258.
- Arbeitskreis Controlling heterogener Personalstrukturen (2018). Diversity Analytics: Stand und Perspektiven. In: Krause S. & Pellenz B. (Hrsg.) *Betriebswirtschaftliche Implikationen der digitalen Transformation. ZfbF-Sonderheft* (pp. 81–101). Wiesbaden: Springer-Gabler.
- Arnaud, S., & Chandon, J. L. (2013). Will monitoring systems kill intrinsic motivation? An empirical study. *Revue de gestion des ressources humaines*, 4(90), 35–53.
- Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. (2016). *Multivariate Analysemethoden – Eine anwendungsorientierte Einführung* (14., überarb. u. ak. Aufl.). Berlin/Heidelberg: Springer.
- Bartel, A. P. (2004). Human resource management and organizational performance: Evidence from retail banking. *Industrial and Labor Relations Review*, 57(2), 181–203.
- Bauer, M., Bräuer, C., Schuldt, J., Niemann, M., & Krömker, H. (2018). Application of wearable technology for the acquisition of learning motivation in an adaptive e-Learning platform. In T. Z. Ahram (Hrsg.), *Advances in human factors in wearable technologies and game design, Proceedings of the AHFE 2018 international conference on human factors in wearable technologies and human factors in game design and virtual environments* (Bd. 795, S. 29–40). Cham: Springer.
- Ben Gal, G., Weiss, E. I., Gafni, N., & Ziv, A. (2010). Preliminary assessment of faculty and student perception of haptic virtual reality simulator for training dental manual dexterity. *Journal of Dental Education*, 75, 496–504.
- Berendes, K., & Werner, C. (2015). Human Resource Intelligence und Analytics für die Strategische Planung. In S. Strohmeier & F. Piazza (Hrsg.), *Human Resource Intelligence und Analytics* (S. 251–273). Wiesbaden: Springer.
- Bernstein, E., & Waber, B. (2019). The truth about open office. *Harvard Business Review*, 97(6), 82–91.
- Bernstein, E., Shore, J., & Lazer, D. (2019). Improving the rhythm of your collaboration. *MIT Sloan Management Review*, 61(1), 29–36.

- Bernstein, E. S., & Turban, S. (2018). The impact of the ‚open‘ workspace on human collaboration. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1753), 20170239.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Bleidorn, W., & Hopwood, C. J. (2019). Using machine learning to advance personality assessment and theory. *Personality and Social Psychology Review*, 23(2), 190–203.
- Blum, B., & Kainer, F. (2019). Rechtliche Aspekte beim Einsatz von KI in HR: Wenn Algorithmen entscheiden. *Personal Quarterly*, 71(3), 22–27.
- Bondarouk, T., Parry, E., & Furtmueller, E. (2017). Electronic HRM: Four decades of research on adoption and consequences. *The International Journal of Human Resource Management*, 28(1), 98–131.
- Booth, R. (2019). Unilever saves on recruiters by using AI to assess job interviews. <https://www.theguardian.com/technology/2019/oct/25/unilever-saves-on-recruiters-by-using-ai-to-assess-job-interviews>. Zugegriffen am 01.11.2019.
- Boucsein, W. (2012). *Electrodermal activity* (2. Aufl.). Heidelberg/New York: Springer.
- Broll, W. (2019). Augmentierte Realität. In R. Dörner, W. Broll, B. Jung & P. Grimm (Hrsg.), *Virtual und Augmented Reality (VR/AR): Grundlagen und Methoden der Virtuellen und Augmentierten Realität* (2. Aufl., S. 315–356). Berlin/Heidelberg: Springer.
- Bues, M., Schultze, T., & Wingert, B. (2018). Konzeption und Implementierung einer VR-Lernumgebung für technische Dienstleistungen. In O. Thomas, D. Metzger & H. Niegemann (Hrsg.), *Digitalisierung in der Aus- und Weiterbildung* (S. 113–123). Berlin/Heidelberg: Springer Gabler.
- Buettner, R. (2017). Prädiktive Algorithmen zur Persönlichkeitsprognose auf Basis von Social-Media-Daten. *Personal Quarterly*, 3, 22–27.
- Buhr, M., Pfeiffer, T., Reiners, D., Cruz-Neira, C., & Jung, B. (2019). Echtzeitaspekte von VR-Systemen. In R. Dörner, W. Broll, B. Jung & P. Grimm (Hrsg.), *Virtual und Augmented Reality (VR/AR): Grundlagen und Methoden der Virtuellen und Augmentierten Realität* (2. Aufl., S. 266–314). Berlin/Heidelberg: Springer.
- Bürgy, C. (2018). Mobile Learning mit Wearables. In C. de Witt & C. Gloerfeld (Hrsg.), *Handbuch Mobile Learning* (S. 141–159). Wiesbaden: Springer.
- Burrell, J. (2016). How the machine ‚thinks‘: Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 1–12.
- Burt, R. S. (2004). Structural holes and good ideas. *American Journal of Sociology*, 110(2), 349–399.
- Busch, M. W., & von der Oelsnitz, D. (2016). Emergente Teamphänomene – Warum sich Erfolge eines Teams nicht einfach kopieren lassen. *Gruppe. Interaktion. Organisation. Zeitschrift für Angewandte Organisationspsychologie*, 47(4), 345–355.
- Campion, M. C., Campion, M. A., Campion, E. D., & Reider, M. H. (2016). Initial investigation into computer scoring of candidate essays for personnel selection. *Journal of Applied Psychology*, 101(7), 958.
- Cappelli, P. (2019). Your approach to hiring is all wrong. *Harvard Business Review*, 97(3), 48–58.
- Cappelli, P., & Tavis, A. (2016). The performance management revolution. *Harvard Business Review*, 94(10), 58–67.
- Cassens-Röhrlig, G. (2019). Die Vorstufen zur künstlichen Intelligenz. *Personalmagazin*, 06, 76–77.

- Castelvecchi, D. (2016). Can we open the black box of AI? *Nature*, 538, 20–23.
- Chaffin, D., Heidl, R., Hollenbeck, J. R., Howe, M., Yu, A., Voorhees, C., & Calantone, R. (2017). The promise and perils of wearable sensors in organizational research. *Organizational Research Methods*, 20(1), 3–31.
- Chalfin, A., Danieli, O., Hillis, A., Jelveh, Z., Luca, M., Ludwig, J., & Mullainathan, S. (2016). Productivity and selection of human capital with machine learning. *American Economic Review*, 106(5), 124–127.
- Colomo-Palacios, R., González-Carrasco, I., López-Cuadrado, J. L., Trigo, A., & Varajao, J. E. (2014). I-Competere: Using applied intelligence in search of competency gaps in software project managers. *Information Systems Frontiers*, 16(4), 607–625.
- Cowgill, B. (2017). Automating judgement and decision-making: Theory and evidence from résumé screening. In *Columbia University, 2015 empirical management conference*.
- Dabirian, A., Paschen, J., & Kietzmann, J. (2019). Employer branding: Understanding employer attractiveness of IT companies. *IT Professional*, 21(1), 82–89.
- Deloitte. (2017). *Rewriting the rules for the digital age: 2017 Deloitte Human Capital Trends*. Deloitte University Press.
- Diensthuber, A., & Gromer, L. V. (2018). Machine Learning meets Pflege(dokumentation) – Erfahrungen aus dem Pilotprojekt BHB Linz. <https://www.pflegekongress.at/html/publicpages/154417460560149.pdf>. Zugegriffen am 27.11.2019.
- Diercks, J. (2020). Online-Assessment. In T. Verhoeven (Hrsg.), *Digitalisierung im Recruiting* (S. 79–100). Wiesbaden: Springer Gabler.
- Diercks, J., & Kupka, K. (2014). Recruitainment – Bedeutung, Einflussfaktoren und Begriffsbestimmung. In dies. (Hrsg.), *Recruitainment: Spielerische Ansätze in Personalmarketing und -auswahl* (S. 1–18). Wiesbaden: Springer Gabler.
- Dittes, A. (2018). Die digitale Revolution – Neue Möglichkeiten im Recruiting. In Fortmann & B. Kolocek (Hrsg.), *Arbeitswelt der Zukunft* (S. 209–220). Berlin/Wiesbaden: Springer.
- Doerr, J. (2018). *OKR: Objectives & Key Results: Wie Sie Ziele, auf die es wirklich ankommt, entwickeln, messen und umsetzen*. München: Vahlen.
- Dörner, R., & Steinicke, F. (2019a). Wahrnehmungsaspekt von VR. In R. Dörner, W. Broll, B. Jung & P. Grimm (Hrsg.), *Virtual und Augmented Reality (VR/AR): Grundlagen und Methoden der Virtuellen und Augmentierten Realität* (2. Aufl., S. 43–78). Berlin/Heidelberg: Springer.
- Dörner, R., Kuhlen, T. W., Matthys, G., et al. (2019a). Fallbeispiele für VR/AR. In R. Dörner, W. Broll, B. Jung & P. Grimm (Hrsg.), *Virtual und Augmented Reality (VR/AR): Grundlagen und Methoden der Virtuellen und Augmentierten Realität* (2. Aufl., S. 357–392). Berlin/Heidelberg: Springer.
- Dörner, R., Broll, W., Jung, B., Grimm, P., & Göbel, M. (2019a). Einführung in Virtual und Augmented Reality. In R. Dörner, W. Broll, B. Jung & P. Grimm (Hrsg.), *Virtual und Augmented Reality (VR/AR): Grundlagen und Methoden der Virtuellen und Augmentierten Realität* (2. Aufl., S. 1–42). Berlin/Heidelberg: Springer.
- Dörner, R., Geiger, C., Oppermann, C., Paelke, V., & Beckhaus, S. (2019a). Interaktionen in Virtuellen Welten. In R. Dörner, W. Broll, B. Jung & P. Grimm (Hrsg.), *Virtual und Augmented Reality (VR/AR): Grundlagen und Methoden der Virtuellen und Augmentierten Realität* (2. Aufl., S. 219–265). Berlin/Heidelberg: Springer.
- Drucker, P. F. (2005). *Was ist Management? Das Beste aus 50 Jahren* (4. Aufl.). Berlin: Econ.
- Dudler, L. (2020). Wenn Bots übernehmen – Chatbots im Recruiting. In T. Verhoeven (Hrsg.), *Digitalisierung im Recruiting* (S. 101–111). Wiesbaden: Springer Gabler.

- Duffke, G. (2019). Verbundprojekt VASE. <https://www.linkedin.com/pulse/verbundprojekt-vase-virtual-analytics-service-im-und-gerd-duffke/>. Zugegriffen am 23.11.2019.
- Duschek, S., & Gärtner, C. (2018). Integratives Kompetenzmanagement. In S. Duschek & F. Schramm (Hrsg.), *Kompetenzmanagement in Clustern und Organisationen* (S. 79–93). München/Mering: Hampp.
- Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, 44(2), 350–383.
- Ehl, C., & Schumann, F. (2019). Von MBO zu OKR. *OrganisationsEntwicklung*, 04, 54–55.
- Escueta, M., Quan, V., Nickow, A. J., & Oreopoulos, P. (2017). *Education technology: An evidence-based review* (National Bureau of Economic Research Working paper no. 23744, S. 1–102).
- Ewenstein, B., Hancock, B., & Komm, A. (2016). Ahead of the curve: The future of performance management. *McKinsey Quarterly*, 2, 64–73.
- EY (2016a). Ernst & Young: Jobstudie 2016. [https://www.ey.com/Publication/vwLUAssetsPI/EY_Jobstudie_2016/\\$FILE/EY-Jobstudie-2016-Karriere-und-Wechselbereitschaft.pdf](https://www.ey.com/Publication/vwLUAssetsPI/EY_Jobstudie_2016/$FILE/EY-Jobstudie-2016-Karriere-und-Wechselbereitschaft.pdf). Zugegriffen am 01.07.2017.
- Fang, C., Liu, Z., Li, Q., & Liu, S. (2018). Sensor data-driven emotion perception in physical learning spaces: A review and prospect. In *Proceedings of the seventh international conference of educational innovation through technology (EITT 2018)*, Auckland, New Zealand (S. 1–5). IEEE.
- Feloni, R. (2017). Consumer-goods giant Unilever has been hiring employees using brain games and artificial intelligence – And it's a huge success. <https://www.businessinsider.de/unilever-artificial-intelligence-hiring-process-2017-6>. Zugegriffen am 01.11.2019.
- Fischbach, K., Gloor, P. A., Lassenius, C., Olguin, D. O., Pentland, A. S., Putzke, J., & Schoder, D. (2010). Analyzing the flow of knowledge with sociometric badges. *Procedia Social and Behavioral Sciences*, 2(4), 6389–6397.
- Florkowski, G. W. (2019). HR technologies and HR-staff technostress: An unavoidable or combatable effect? *Employee Relations*, 41(5), 1120–1144.
- Freire, M., Serrano-Laguna, Á., Iglesias, B. M., Martínez-Ortiz, I., Moreno-Ger, P., & Fernández-Manjón, B. (2016). Game learning analytics: Learning analytics for serious games. In M. Spector, B. Lockee & M. Childress (Hrsg.), *Learning, design, and technology* (S. 1–29). Cham: Springer.
- Frierson, J., & Si, D. (2018). Who's next: Evaluating attrition with Machine Learning algorithms and survival analysis. In F. Y. L. Chin et al. (Hrsg.), *International conference on big data* (S. 251–259). Cham: Springer.
- Fuller, R. (20. August 2014). 3 behaviors that drive successful salespeople. *Harvard Business Review (Web Article)*, S. 1–4.
- Fuller, R. (8. July 2015). What makes great sales people? *Harvard Business Review (Web Article)*, S. 1–5.
- Furkel, D. (2018). Eine neue Technologiewelle. *Personalmagazin*, 12, 32–36.
- Gärtner, C. (2013). Cognition, knowing and learning in the flesh: Six views on embodied knowing in organization studies. *Scandinavian Journal of Management*, 29(4), 338–352.
- Gärtner, C. (2016). Modernisierter Klassiker: Mitarbeitergespräche zeitgemäß führen. *change! 01*, 14–16.
- Gärtner, C. (2017). Mensch oder Maschine: Wer trifft die besseren Personalentscheidungen? *HR Performance*, 2, 14–17.

- Gärtner, C. (2018a). Human Resource Management bei und von XING. In C. Gärtner & C. Heinrich (Hrsg.), *Fallstudien zur Digitalen Transformation: Case Studies für die Lehre und praktische Anwendung* (S. 129–147). Berlin/Wiesbaden: Springer.
- Gärtner, C. (2018b). Workplace Analytics und Performance Management im Vertrieb. In F. M. Binninger, T. Weise & A. Mues (Hrsg.), *Moderne Personalpolitik in Handel und Vertrieb* (S. 143–160). München: UTZ.
- Gärtner, C., & Claßen, M. (2016). Die gläsernen Mitarbeiter. *Personalmagazin*, 08, 38–39.
- Geister, S., & Rastetter, D. (2009). Aktueller Stand zum Thema Online-Tests. In H. Steiner (Hrsg.), *Online-Assessment* (S. 3–16). Berlin/Heidelberg: Springer.
- Gelbard, R., Ramon-Gonen, R., Carmeli, A., Bittmann, R. M., & Talyansky, R. (2018). Sentiment analysis in organizational work: Towards an ontology of people analytics. *Expert Systems*, 35(5), e12289.
- Gloor, P. A., Colladon, A. F., Grippa, F., & Giacomelli, G. (2017). Forecasting managerial turnover through e-mail based social network analysis. *Computers in Human Behavior*, 71, 343–352.
- Goler, L., Gale, J., & Grant, A. (2016). Let's not kill performance evaluations yet. *Harvard Business Review*, 94(11), 90–94.
- Gombolay, M. C., Gutierrez, R. A., Clarke, S. G., Sturla, G. F., & Shah, J. A. (2015). Decision-making authority, team efficiency and human worker satisfaction in mixed human-robot teams. *Autonomous Robots*, 39(3), 293–312.
- Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6), 1360–1380.
- Greenfield, P. M. (2009). Technology and informal education: What is taught, what is learned. *Science*, 323(5910), 69–71.
- Grund, C. (2009). Jobpräferenzen und Arbeitsplatzwechsel. *Zeitschrift für Personalforschung*, 23(1), 66–72.
- Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys*, 51(5), 93:1–93:42.
- Habermas, J. (2004). Freiheit und Determinismus. *Deutsche Zeitschrift für Philosophie*, 52(6), 871–890.
- Hafen, T. (2018). Künstliche Welten steigern die Effizienz. *com! professional*, 1, 26–31.
- Han, J., Zhang, Z., & Schuller, B. (2019). Adversarial training in affective computing and sentiment analysis: Recent advances and perspectives. *IEEE Computational Intelligence Magazine*, 14(2), 68–81.
- Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. New York: Routledge.
- Haya, P. A., Daems, O., Malzahn, N., Castellanos, J., & Hoppe, H. U. (2015). Analysing content and patterns of interaction for improving the learning design of networked learning environments. *British Journal of Educational Technology*, 46(2), 300–316.
- Hilbig, R., Renz, A., & Schildhauer, T. (2019). Data analytics – The future of innovative teaching and learning. In *Proceedings of ISPIM innovation conference – Celebrating innovation: 500 years since DaVinci (Florence, 16.–19. June 2019)* (S. 1–16).
- Hill, J., Ford, W. R., & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human-human online conversations and human-chatbot conversations. *Computers in Human Behavior*, 49, 245–250.

- Hoeffler, D., Kunowsky, O., Müller, S., Niendorf, T., & Pfalzgraf, J. (2015). Personalwirtschaftliches Reporting mit SAP Netweaver und Business Objects. In S. Strohmeier & F. Piazza (Hrsg.), *Human Resource Intelligence und Analytics. Grundlagen, Anbieter, Erfahrungen und Trends* (S. 187–220). Wiesbaden: Springer Gabler.
- Hoffman, M., Kahn, L. B., & Li, D. (2017). Discretion in hiring. *The Quarterly Journal of Economics*, 133(2), 765–800.
- Höller, H.P., & Wedde, P. (2018). Die Vermessung der Belegschaft: Mining the Enterprise Social Graph. *Mitbestimmungspraxis*, Nr. 10, S. 1–38.
- Holtom, B. C., Mitchell, T. R., Lee, T. W., & Eberly, M. B. (2008). Turnover and retention research: A glance at the past, a closer review of the present, and a venture into the future. *The Academy of Management Annals*, 2(1), 231–274.
- Huff, J., & Götz, T. (2019). Evidenz statt Bauchgefühl? Möglichkeiten und rechtliche Grenzen von Big Data im HR-Bereich. *Neue Zeitschrift für Arbeitsrecht (NZA)* 36(24) (Beilage 2/2019), 73–78.
- IBM. (2019). Siemens AG: A one-of-a-kind AI-based chatbot innovates employee support. <https://www.ibm.com/case-studies/siemens-ag-watson-ai>. Zugegriffen am 01.10.2019.
- Ifenthaler, D., & Drachsler, H. (2018). Learning analytics. In H. Niegemann & A. Weinberger (Hrsg.), *Lernen mit Bildungstechnologien* (S. 1–20). Berlin: Springer.
- Ivens, S. (2018). *Unternehmensreputation im digitalen Zeitalter: Wie Mitarbeiter die Online-Reputation bei Kunden, Kollegen und Bewerbern beeinflussen*. Wiesbaden: Springer Gabler.
- Jansen, D. (2003). Einführung in die Netzwerkanalyse: Grundlagen, Methoden, Forschungsbeispiele. 2., erw. Aufl., Wiesbaden: Springer.
- Jenewein, T. (2018). Ansätze zum Lernen im Digitalen Zeitalter. Darstellung am Beispiel SAP. In T. Petry & W. Jäger (Hrsg.), *Digital HR. Smarte und agile Systeme, Prozesse und Strukturen im Personalmanagement* (S. 259–274). Freiburg: Haufe Lexware.
- Joberate. (2016). *White paper: Understanding J-Score™ predictive analytic @Joberate*. https://de.slideshare.net/Joberate/white-paper-understanding-jscore-predictive-analytic-joberate?from_action=save. Zugegriffen am 18.12.2019.
- Juraschek, M., Büth, L., Posselt, G., & Herrmann, C. (2018). Mixed reality in learning factories. *Procedia Manufacturing*, 23, 153–158.
- Kaggle. (2019). IBM HR analytics employee attrition & performance. <https://www.kaggle.com/pavansubhashini/ibm-hr-analytics-attrition-dataset>. Zugegriffen am 08.06.2019.
- Kanning, U. P. (2015). *Personalauswahl zwischen Anspruch und Wirklichkeit: eine wirtschaftspsychologische Analyse*. Berlin/Heidelberg: Springer.
- Kanning, U. P. (2017). *Personalmarketing, Employer Branding und Mitarbeiterbindung: Forschungsbefunde und Praxistipps aus der Personalpsychologie*. Berlin/Heidelberg: Springer.
- Kauffeld, S. (2016). *Nachhaltige Personalentwicklung und Weiterbildung* (2. Aufl.). Berlin/Heidelberg: Springer.
- Kern, D., & Haep, M. (2016). Wie Konzerne ihr Performance Management modernisieren können (Interview von G. Birkner). *Comp & Ben Magazin*, 2, 12–14.
- Kerry, A., Ellis, R., & Bull, S. (2008). Conversational agents in E-Learning. In T. Allen, R. Ellis & M. Petridis (Hrsg.), *Applications and innovations in intelligent systems* (S. 169–182). London: Springer.
- Kettler, C., & Kauffeld, S. (2019). Game-based Learning. In S. Kauffeld & J. Othmer (Hrsg.), *Handbuch innovative Lehre* (S. 249–253). Wiesbaden: Springer.

- Khusainova, R., De Jong, A., Lee, N., Marshall, G. W., & Rudd, J. M. (2018). (Re) defining salesperson motivation: Current status, main challenges, and research directions. *Journal of Personal Selling & Sales Management*, 38(1), 2–29.
- Kirkpatrick, D. L. (1967). Evaluation of training. In R. L. Craig (Hrsg.), *Training and development handbook: A guide to human resources development* (S. 87–112). New York: McGraw-Hill.
- Klahre, A.-M. (2017). 3 Ways Johnson & Johnson is taking talent acquisition to the next level. <https://www.jnj.com/innovation/3-ways-johnson-and-johnson-is-taking-talent-acquisition-to-the-next-level>. Zugegriffen am 08.08.2019.
- Kleinberg, J., & Raghu, M. (2018). Team performance with test scores. *ACM Transactions on Economics and Computation (TEAC)*, 6(3–4), 17–45.
- Kleindienst, B. (2017). *Performance Measurement und Management: Gestaltung und Einführung von Kennzahlen- und Steuerungssystemen*. Wiesbaden: Springer Gabler.
- Klerkx, J., Verbert, K., & Duval, E. (2017). Learning analytics dashboards. In C. Lang, G. Siemens, A. Wise & D. Gasevic (Hrsg.), *Handbook of learning analytics* (S. 143–150). SOLAR, Society for Learning Analytics and Research.
- Kozlowski, S. W., & Chao, G. T. (2018). Unpacking team process dynamics and emergent phenomena: Challenges, conceptual advances, and innovative methods. *American Psychologist*, 73(4), 576–592.
- Kozlowski, S. W., Chao, G. T., Chang, C. H., & Fernandez, R. (2015). Team dynamics: Using „big data“ to advance the science of team effectiveness. In S. Tomidandel, E. B. King & J. M. Cortina (Hrsg.), *Big data at work: The data science revolution and organizational psychology* (S. 273–309). New York: Routledge.
- Kraus, C. (2018). Digitalisierung der Arbeitswelt – das Ende der Low Performer? *Der Betrieb*, 12, 701–705.
- Kravčík, M., Ullrich, C., & Igel, C. (2018). Das Internet der Dinge als Lern- und Ausbildungshilfe in der digitalen Ära. In K. Schwuchow & J. Gutmann (Hrsg.), *HR-Trends 2019: Strategie, Digitalisierung, Diversität, Demografie* (S. 137–151). Freiburg: Haufe.
- Kuksenok, K., & Martyniv, A. (2019). Evaluation and improvement of chatbot text classification data quality using plausible negative examples. *arXiv preprint arXiv:1906.01910*.
- Kumar, A. D., Selvam, R. P., & Kumar, K. S. (2018). Review on prediction algorithms in educational data mining. *International Journal of Pure and Applied Mathematics*, 118(8), 531–537.
- Kunz, C., Leubner, T., Liebert, K., Reuter, M., Siebel, J., & Kinschel, M. (2019). Industrie 4.0 – Ein digitales Transfermodell für Aus- und Weiterbildung. In H. Lee & S. Pfeiffer (Hrsg.), *Industrie 4.0: Risiken und Chancen für die Berufsbildung* (2. Aufl., S. 205–216). Bielefeld: wbv.
- Küpers, W., & Weibler, J. (2005). *Emotionen in Organisationen*. Stuttgart: Kohlhammer.
- Kutschka, M., & Herbers, J. (2016). An insight to aviation: Rostering ground personnel in practice. In M. Lübbecke, A. M. Koster, P. Letmathe, R. Madlener, B. Peis & G. Walther (Hrsg.), *Operations research proceedings 2014* (S. 349–355). Cham: Springer.
- Lambrecht, A., & Tucker, C. (2019). Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads. *Management Science*, 65(7), 2966–2981.
- Lanier, J. (2017). *Dawn of the new everything: A journey through virtual reality*. London: Bodley Head.

- Laumer, S., Weitzel, T., & Luzar, K. (2019). Robo-Recruiting: Status quo und Herausforderungen für die KI in der Personalgewinnung. *Personal Quarterly*, 71(3), 10–15.
- Ledford, G. E., Benson, G., & Lawler, E. E. (2016). Aligning research and the current practice of performance management. *Industrial and Organizational Psychology*, 9(2), 253–259.
- Lee, T. W., Hom, P. W., Eberly, M. B., & Mitchell, T. R. (2017). On the next decade of research in voluntary employee turnover. *Academy of Management Perspectives*, 31(3), 201–221.
- Leonardi, P., & Contractor, N. (2018). Better people analytics. *Harvard Business Review*, 96(6), 70–81.
- Levy, P., Tseng, S., Rosen, C., & Lueke, S. (2017). Performance management: A marriage between practice and science – Just say „I do“. In R. M. Buckley, A. R. Wheeler & J. R. B. Halbesleben (Hrsg.), *Research in personnel and human resources management* (Bd. 35, S. 155–213). Bingley: Emerald.
- Liao, H., Toya, K., Lepak, D., & Hong, Y. (2009). Do they see eye to eye? Management and employee perspectives of high-performance work systems and influence processes on service quality. *Journal of Applied Psychology*, 94(2), 371–391.
- Libuda, I., & Fleischmann, F. (2018). Strategische Personalplanung und People Analytics. In K. Schwuchow & J. Gutmann (Hrsg.), *HR-Trends 2019: Strategie, Digitalisierung, Diversität, Demografie* (S. 233–243). Freiburg: Haufe Lexware.
- Liebert, K., & Talg, A. (2018). Künstliche Intelligenz und das Lernen der Zukunft. In K. Schwuchow & J. Gutmann (Hrsg.), *HR-Trends 2019: Strategie, Digitalisierung, Diversität, Demografie* (S. 197–208). Freiburg: Haufe Lexware.
- Limpächer, S., & Maier, L. (2016). Der Transfer des Gelernten. *Personalführung*, 6, 56–60.
- Liu, B. (2015). *Sentiment analysis: Mining opinions, sentiments, and emotions*. New York: Cambridge University Press.
- Lochner, K., & Preuß, A. (2018). Digitales Recruiting. *Gruppe. Interaktion. Organisation. Zeitschrift für Angewandte Organisationspsychologie*, 49(3), 193–202.
- Lung, H. (1996). *Sprache und Didaktik im Seminar*. München/Basel: Ernst Reinhardt.
- Lunsford, D. L., & Phillips, P. P. (2018). Tools used by organizations to support human capital analytics. *Performance Improvement*, 57(3), 6–15.
- MacFadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an „early warning system“ for educators: A proof of concept. *Computers & Education*, 54(2), 588–599.
- Mahmud, J. (2015). IBM Watson Personality Insights: The science behind the service. <https://developer.ibm.com/watson/blog/2015/03/23/ibm-watson-personality-insights-science-behind-service/>. Zugegriffen am 29.07.2019.
- Mangaroska, K., & Giannakos, M. (2018). Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning. *IEEE Transactions on Learning Technologies*, 12(4), 516–534.
- Manus VR (2019). The Manus VR Gloves. <https://manus-vr.com/gloves/>. Zugegriffen am 31.10.2019.
- March, J. G., & Simon, H. A. (1958). *Organizations*. New York: Wiley.
- Marler, J. H., & Boudreau, J. W. (2017). An evidence-based review of HR analytics. *The International Journal of Human Resource Management*, 28(1), 3–26.
- Martini, M. (2019). *Blackbox Algorithmus – Grundfragen einer Regulierung Künstlicher Intelligenz*. Berlin/Heidelberg: Springer.

- Matzer, M., & Litzel, N. (2017). Digitalisierung schlau umgesetzt. <https://www.bigdata-insider.de/digitalisierung-schlau-umgesetzt-a-654757/>. Zugegriffen am 01.10.2019.
- McChesney, C., Covey, S., & Huling, J. (2012). *The 4 disciplines of execution: Achieving your wildly important goals*. New York: Free Press/Simon & Schuster.
- Meier, C. (2019). Vom Anbieter zum Ermöglicher. *Personalmagazin*, 4, 39–42.
- Meier, C., Bäcker, D., & Seibold, D. (2018). *Digitale Transformation und L&D: Ergebnisse einer Standortbestimmung und Handlungserfordernisse*. scil Arbeitsbericht Nr. 29.
- Meinecke, A. L., & Kauffeld, S. (2019). Engaging the hearts and minds of followers: Leader empathy and Language style matching during appraisal interviews. *Journal of Business and Psychology*, 34(4), 485–501.
- Metzger, D., Jannaber, S., Berkemeier, L., & Thomas, O. (2018). Glassroom: Aus- und Weiterbildung mit Smart Glasses und Virtual Reality-Brillen im technischen Kundendienst. In C. de Witt & C. Gloerfeld (Hrsg.), *Handbuch Mobile Learning* (S. 901–919). Wiesbaden: Springer.
- Meyer, B., Burtscher, M. J., Jonas, K., Feese, S., Arnrich, B., Tröster, G., & Schermuly, C. C. (2016). What good leaders actually do: Micro-level leadership behaviour, leader evaluations, and team decision quality. *European Journal of Work and Organizational Psychology*, 25(6), 773–789.
- Microsoft. (2019a). Bayer – Innovation project. <http://download.microsoft.com/download/8/1/5/8158F66A-448F-4C5E-B75B-C15C4F551933/Bayer-BoFramework-CognitiveServices.pdf>. Zugegriffen am 07.10.2019a
- Microsoft. (2019b). Workplace analytics. <https://docs.microsoft.com/de-de/workplace-analytics/index-orig>. Zugegriffen am 03.12.2019.
- Microsoft. (2019c). Datenschutzhandbuch für myAnalytics-Administratoren. <https://docs.microsoft.com/de-de/workplace-analytics/myanalytics/overview/privacy-guide>. Zugegriffen am 10.12.2019.
- Microsoft. (2019d). How does Office Delve know what's relevant to me? <https://support.office.com/en-us/article/How-does-Office-Delve-know-what-s-relevant-to-me-048d502e-80a7-4f77-ac5c-f9d81733c385>. Zugegriffen am 10.12.2019.
- Mou, Y., & Xu, K. (2017). The media inequality: Comparing the initial human-human and human-AI social interactions. *Computers in Human Behavior*, 72, 432–440.
- Mühlbauer, D., Huff, J., & Süß, J. (2018). People Analytics und Arbeit 4.0. In S. Werther & L. Bruckner (Hrsg.), *Arbeit 4.0 aktiv gestalten* (S. 107–132). Berlin/Wiesbaden: Springer.
- Müller-Frommeyer, L. C., Frommeyer, N. A., & Kauffeld, S. (2019). Introducing rLSM: An integrated metric assessing temporal reciprocity in language style matching. *Behavior Research Methods*, 51(3), 1343–1359.
- Neculoiu, P., Versteegh, M., & Rotaru, M. (2016). Learning text similarity with siamese recurrent networks. In The Association for Computational Linguistics (Hrsg.), *Proceedings of the 1st workshop on representation learning for NLP* (S. 148–157) Berlin.
- Neuberger, O. (1994). *Personalentwicklung* (2. Aufl.). Stuttgart: Enke.
- Neuberger, O. (2000). *Das 360°-Feedback: Alle fragen? Alles sehen? Alles sagen?* München/Mering: Hampp.
- Newport, C. (2016). *Deep work: Rules for focused success in a distracted world*. London: Hachette.
- Niegemann, H., & Niegemann, L. (2018). Design digitaler Aus- und Weiterbildungsszenarien. In O. Thomas, D. Metzger & H. Niegemann (Hrsg.), *Digitalisierung in der Aus- und*

- Weiterbildung: *Virtual und Augmented Reality für Industrie 4.0* (S. 75–91). Berlin/Heidelberg: Springer Gabler.
- Niegemann, H. M., & Heidig, S. (2019). Interaktivität und Adaptivität in multimedialen Lernumgebungen. In H. Niegemann & A. Weinberger (Hrsg.), *Lernen mit Bildungstechnologien* (S. 1–25). Berlin/Heidelberg: Springer.
- Nielsen, C. (16. Februar 2016). What work email can reveal about performance and potential. *Harvard Business Review (Web Article)*, S. 1–4.
- Nienhüser, W. (2018). Bilder der Austauschbeziehung zwischen Kapital und Arbeit in der Personalwirtschaftslehre – Grundzüge einer Ideengeschichte. In W. Matiaske & W. Weber (Hrsg.), *Ideengeschichte der BWL* (S. 443–486). Wiesbaden: Springer Gabler.
- Norten, K. (2014). Unconscious bias at work. <https://library.gv.com/unconscious-bias-at-work-22e698e9b2d#.ceqzmagrj>. Zugegriffen am 18.09.2017.
- North, K., & Maier, R. (2018). Wissen 4.0 – Wissensmanagement im digitalen Wandel. *HMD Praxis der Wirtschaftsinformatik*, 55(4), 665–681.
- NZA (2017). Mitbestimmung bei technischer Überwachungseinrichtung – Belastungsstatistik. *Neue Zeitschrift für Arbeitsrecht (NZA)*, 34(18/2017), 1205–1212.
- Oberdörfer, S., & Latoschik, M. E. (2019). Knowledge encoding in game mechanics: Transfer-oriented knowledge learning in desktop-3D and VR. *International Journal of Computer Games Technology*, 2019, 1–17.
- Olguín-Olguín, D., & Pentland, A. (2010). Sensor-based organisational design and engineering. *International Journal of Organisational Design and Engineering*, 1(1–2), 69–97.
- de Oliveira, J. M., Zylka, M. P., Gloor, P. A., & Joshi, T. (2019). Mirror, Mirror on the wall, who is leaving of them all: Predictions for employee turnover with gated recurrent neural networks. In Y. Song, F. Grippa, P. Gloor, & J. Leitão (Hrsg.), *Collaborative innovation networks* (S. 43–59). Cham: Springer.
- Ortmann, G. (1984). *Der zwingende Blick. Personalinformationssysteme – Architektur der Disziplin*. Frankfurt/New York: Campus.
- Ott, A. (2017). A year with our recruiting chatbot. <https://medium.com/@frauott/a-year-with-our-recruiting-chatbot-e77f52264a>. Zugegriffen am 17.07.2019.
- Oviatt, S., Schuller, B., Cohen, P. R., Sonntag, D., Potamianos, G., & Krüger, A. (Hrsg.). (2019). *The handbook of multimodal-multisensor interfaces: Language processing, software, commercialization, and emerging directions – Volume 3*. New York: ACM/Morgan & Claypool.
- Pandey, S., & Pandey, S. K. (2019). Applying natural language processing capabilities in computerized textual analysis to measure organizational culture. *Organizational Research Methods*, 22(3), 765–797.
- Pentland, A. (2012). The new science of building great teams. *Harvard Business Review*, 90(4), 60–69.
- Pesch, U. (2018). Neue Wege zum Wissen. *Personalwirtschaft (Sonderheft)*, 07, 22–25.
- Piazza, F. (2010). *Data Mining im Personalmanagement*. Wiesbaden: Springer Gabler.
- Poria, S., Peng, H., Hussain, A., Howard, N., & Cambria, E. (2017). Ensemble application of convolutional neural networks and multiple kernel learning for multimodal sentiment analysis. *Neurocomputing*, 261, 217–230.
- Poria, S., Majumder, N., Hazarika, D., Cambria, E., Gelbukh, A., & Hussain, A. (2018). Multimodal sentiment analysis: Addressing key issues and setting up the baselines. *IEEE Intelligent Systems*, 33(6), 17–25.

- Portugal, I., Alencar, P., & Cowan, D. (2018). The use of machine learning algorithms in recommender systems: A systematic review. *Expert Systems with Applications*, 97, 205–227.
- Prabha, I. M., & Srikanth, G. U. (2019). Survey of sentiment analysis using deep learning techniques. In *1st international conference on innovations in information and communication technology* (S. 1–9) Chennai.
- Provost, F., & Fawcett, T. (2017). *Data Science für Unternehmen: Data Mining und datenanalytisches Denken praktisch anwenden*. Frechen: mitp.
- Quast, C., Villeumier, F., Kirchner, J., & Fedossov, A. (2018). *Programmatic Job Advertising – Mit Big Data zum planbaren Bewerber-Rücklauf* (Whitepaper). Wollmilchsau.
- Queb. (2019). KI und Sprachanalyse im Recruiting: L'Oréal setzt auf SEEDLINK. <https://www.queb.org/blog/ki-und-sprachanalyse-im-recruiting-loreal-setzt-auf-seedlink/>. Zugriffen am 23.08.2019.
- Quwaider, M., & Biswas, S. (2010). DTN routing in body sensor networks with dynamic postural partitioning. *Ad Hoc Networks*, 8(8), 824–841.
- Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2019). Mitigating bias in algorithmic employment screening: Evaluating claims and practices. *arXiv preprint arXiv:1906.09208*.
- Rambachan, A., & Roth, J. (2019). Bias in, bias out? Evaluating the folk wisdom. *arXiv preprint arXiv:1909.08518*.
- Ratner, B. (2017). *Statistical and machine-learning data mining: Techniques for better predictive modeling and analysis of big data* (3. Aufl.). Boca: Chapman & Hall/CRC.
- Reeves, B., & Read, L. J. (2009). *Total engagement: Using games and virtual worlds to change the way people work and businesses compete*. Boston: Harvard Business Press.
- Reichertz, J. (2016). *Qualitative und interpretative Sozialforschung: eine Einladung*. Wiesbaden: Springer VS.
- Remus, R., Quasthoff, U., & Heyer, G. (2010). SentiWS. A publicly available German-language resource for sentiment analysis. In N. Calzolari et al. (Hrsg.), *Proceedings of the 7th international language resources and evaluation* (S. 1168–1171). Valletta: European Language Resources Association.
- Rüdian, S., & Pinkwart, N. (2019). Towards an automatic Q&A generation for online courses – A pipeline based approach. In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren & R. Luckin (Hrsg.), *Artificial intelligence in education* (S. 237–242). Cham: Springer.
- Sakellariadis, S. (2015). Making sure the cup stays full at starbucks: Leveraging narratives from Glassdoor.com to improve recruitment and retention. http://www.huffingtonpost.com/sophie-sakellariadis/making-sure-the-cup-stays_b_7935760.html. Zugriffen am 01.07.2019.
- Sánchez-Monedero, J., & Dencik, L. (2019). *The datafication of the workplace. Working paper*. Cardiff University.
- Sanchez-Monedero, J., Dencik, L., & Edwards, L. (2019). What does it mean to solve the problem of discrimination in hiring? Social, technical and legal perspectives from the UK on automated hiring systems. *arXiv preprint arXiv:1910.06144*.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124(2), 262–274.

- Schmidt, F. L., Oh, I. S., & Shaffer, J. A. (2016). *The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 100 years of research findings* (Fox School of Business research paper).
- Schmidt-Atzert, L., Künecke, J., & Zimmermann, J. (2019). TBS-DTK Rezension: „PRE-CIRE JobFit“. *Report Psychologie*, 44(7/8), 19–21.
- Schneider, J., Börner, D., Van Rosmalen, P., & Specht, M. (2016). Can you help me with my pitch? Studying a tool for real-time automated feedback. *IEEE Transactions on Learning Technologies*, 9(4), 318–327.
- Scholz, T. M. (2017). *Big data in organizations and the role of human resource management: A complex systems theory-based conceptualization*. Frankfurt a. M.: Peter Lang.
- Scholz, C., & Scholz, T. (2019). *Grundzüge des Personalmanagements* (3. Aufl). München: Vahlen.
- Schwantzer, S. (2018). Konzeption und Implementierung eines Smart-Glasses-basierten Informationssystems für technische Dienstleistungen. In O. Thomas, D. Metzger & H. Niegemann (Hrsg.), *Digitalisierung in der Aus- und Weiterbildung* (S. 94–112). Berlin/Heidelberg: Springer Gabler.
- Schweyer, A. (2018). *Predictive analytics and artificial intelligence in people management*. Incentive Research Foundation.
- Seufert, J. (2019). Ich liefere. *Die Zeit*, Nr. 46 (07.11.2019), S. 76–77.
- Seufert, S., Preisig, L., Krapf, J., & Meier, C. (2017). *Von Gamification zum systematischen Motivationsdesign mit kollaborativen und spielerischen Gestaltungselementen: Konzeption und Anwendungsbeispiele*. scil Arbeitsbericht Nr. 27.
- Shi, Y., Zhu, L., Li, W., Guo, K., & Zheng, Y. (2019). Survey on classic and latest textual sentiment analysis articles and techniques. *International Journal of Information Technology & Decision Making*, 18(4), 1243–1287.
- Shih, J. (2017a). How predictive analytics helps improve human capital management. <https://blogs.sap.com/2017/03/19/how-predictive-analytics-helps-improve-human-capital-management/>. Zugegriffen am 18.12.2019.
- Shih, J. (2017b). <https://blogs.sap.com/2017/07/20/applying-predictive-analytics-to-manage-employee-turnover/>. Zugegriffen am 18.12.2019.
- Siegel, E. (2013). *Predictive analytics: The power to predict who will click, buy, lie, or die*. Hoboken: Wiley.
- Simões, J. (2015). *Using gamification to improve participation in social learning*. Doctoral Thesis. University of Vigo.
- Skenes, C., & Kleiner, B. H. (2003). The HAY System of compensation. *Management Research News*, 26(2/3/4), 109–115.
- Speer, A. B. (2018). Quantifying with words: An investigation of the validity of narrative-derived performance scores. *Personnel Psychology*, 71(3), 299–333.
- Spitzer, D., Tschürtz, S., & Burel, S. (o. J.). Deutsche Stellenausschreibungen unterscheiden zwischen Mann und Frau. Whitepaper 100 Worte GmbH. https://www.100worte.de/wp-content/uploads/2019/02/Artikel_Genderbias.pdf. Zugegriffen am 08.08.2019.
- Squires, P., Kaufman, H. G., Togelius, J., & Jaramillo, C. M. (2017). A comparative sequence analysis of career paths among knowledge workers in a multinational bank. In *IEEE International Conference on Big Data (Big Data)* (pp. 3604–3612). IEEE.
- Staab, P., & Geschke, S. C. (2019). *Ratings als arbeitspolitisches Konfliktfeld: Das Beispiel Zalando* (Study der Hans-Böckler-Stiftung, No. 429). Düsseldorf: Hans-Böckler-Stiftung.

- Staufenbiel & Kienbaum. (2016). RecruitingTrends 2017: Was HR-Verantwortliche wissen müssen. https://www.staufenbiel.de/fileadmin/fm-dam/PDF/Studien/RecruitingTrends_2017.pdf. Zugegriffen am 01.07.2017.
- Stenius, H. (2017). *Change analytics: Exploring the data-driven management of organizational change*. Helsinki: Aalto University.
- Stone, P., & Chatagny, J. (2019). Analytics als Wandelmotor nutzen. Ein Erfahrungsbericht zur Neugestaltung interner Dienste bei der Swiss Re. *OrganisationsEntwicklung*, Nr. 3/2019, S. 6–12.
- Strohmeier, S. (2015). Analysen der Human Resource Intelligence und Analytics. In S. Strohmeier & F. Piazza (Hrsg.), *Human Resource Intelligence und Analytics* (S. 3–47). Wiesbaden: Springer.
- Strohmeier, S. (2019). Einsatzpotenziale und -grenzen von künstlicher Intelligenz in der Personalarbeit (Interview mit T. Biemann). *Personal Quarterly*, 3, 6–9.
- Strohmeier, S., Majstorovic, D., Piazza, F., & Theres, C. (2016). Smart HRM – das „Internet der Dinge“ im Personalmanagement. *HMD Praxis der Wirtschaftsinformatik*, 53(6), 838–850.
- 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.
- Tallgauer, M., Festing, M., & Fleischmann, F. (2020). Big Data im Recruiting. In T. Verhoeven (Hrsg.), *Digitalisierung im Recruiting* (S. 25–39). Wiesbaden: Springer Gabler.
- Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial intelligence in human resources management: Challenges and a path forward. *California Management Review*, 61(4), 15–42.
- Tang, S., Peterson, J., & Pardos, Z. (2017). Predictive modelling of student behaviour using granular large-scale action data. In C. Lang, G. Siemens, A. Wise & D. Gasevic (Hrsg.), *Handbook of learning analytics* (S. 223–233). SOLAR, Society for Learning Analytics and Research.
- TrustSphere. (2019). How a large financial services organization used Organizational Network Analytics to understand female promotability. <https://www.trustsphere.com/wp-content/uploads/2019/04/Diversity-Inclusion-Case-Study-201903-1-Letter.pdf>. Zugegriffen am 12.12.2019.
- Ullrich, C., Hauser-Ditz, A., Kreggenfeld, N., Prinz, C., & Igel, C. (2018). Assistenz und Wissensvermittlung am Beispiel von Montage- und Instandhaltungstätigkeiten. In S. Wischmann & E. A. Hartmann (Hrsg.), *Zukunft der Arbeit: Eine praxisnahe Betrachtung* (S. 107–122). Berlin/Heidelberg: Springer.
- Unger, T., Goossens, J., & Becker, L. (2015). Digitale Serious Games. In U. Blötz (Hrsg.), *Planspiele und Serious Games in der beruflichen Bildung: Auswahl, Konzepte, Lernarrangements, Erfahrungen-Aktueller Katalog für Planspiele und Serious Games* (S. 157–179). Bielefeld: Bertelsmann.
- Unilever. (2018). Frequently asked questions: Unilever Graduate Programme/Unilever Future Leadership Programme. https://www.unilever.co.za/Images/uflp-digital-selection-faq_tcm1262-506251_en.pdf. Zugegriffen am 01.11.2019.
- Van Laar, E., Van Deursen, A. J., Van Dijk, J. A., & De Haan, J. (2017). The relation between 21st-century skills and digital skills: A systematic literature review. *Computers in Human Behavior*, 72, 577–588.

- Verhoeven, T. (2020). Künstliche Intelligenz im Recruiting. In ders. (Hrsg.), *Digitalisierung im Recruiting* (S. 113–128). Wiesbaden: Springer Gabler.
- Vermeulen, A. E. (2014). *Improving student engagement through visualization of course activities*. Doctoral dissertation.
- Vesa, M., & Harviainen, J. T. (2019). Gamification: Concepts, consequences, and critiques. *Journal of Management Inquiry*, 28(2), 128–130.
- Vetter, C. (2018). Strategie und Arbeit 4.0. In S. Werther & L. Bruckner (Hrsg.), *Arbeit 4.0 aktiv gestalten* (S. 132–146). Berlin: Springer.
- Vetter, C., & Zacharides, K. (2016). Datengetriebene Personalarbeit. *Personalmagazin*, 3, 56–58.
- Vincent, J. (2018). Amazon reportedly scraps internal AI recruiting tool that was biased against women. <https://www.theverge.com/2018/10/10/17958784/ai-recruiting-tool-bias-amazon-report>. Zugegriffen am 08.08.2019.
- Vladova, G., Haase, J., Rüdian, L. S., & Pinkwart, N. (2019). Educational chatbot with learning avatar for personalization. In *Twenty-fifth Americas conference on information systems, Cancun, 2019*.
- de Vos, A., & Meganck, A. (2009). What HR managers do versus what employees value: Exploring both parties' view on retention management from a psychological contract perspective. *Personnel Review*, 38(1), 45–60.
- Waber, B., Magnolfi, J., & Lindsay, G. (2014). Workspaces that move people. *Harvard Business Review*, 92(10), 68–77.
- Wagner, M. (2009). *Modellbasierte Arbeitskräfteplanung für stochastische Instandhaltungsereignisse in der zivilen Luftfahrt*. Dissertation TU Berlin.
- WalkMe (2017). *Autonomous behavior reasoning analysis*. World Intellectual Property Organization, Publication no. WO 2017/212466 A1.
- WalkMe. (2019). *Chat-based application interface for automation*. United States Patent Application, Publication no. US 2019/0215283 A1.
- Wang, N., & Katsamakas, E. (2019). A network data science approach to people analytics. *Information Resources Management Journal*, 32(2), 28–51.
- Warszta, T., Oldenburger, J., Severin, A., Thiessen, L., & Lange, M.-C. (2020). Präferenzen von eignungsdiagnostischen Instrumenten in Selbst- und Fremdselektion. *Personal Quarterly*, 1, 34–41.
- Wen, Q., Gloor, P. A., Fronzetti Colladon, A., Tickoo, P., & Joshi, T. (2019). Finding top performers through email patterns analysis. *Journal of Information Science*, 1–20.
- Werther, S. (2018). Mit künstlicher Intelligenz zu smarteren Prozessen. *Personalwirtschaft Sonderheft*, 7, 28–30.
- Werther, S., & Woschée, R. (2018). Die Zukunft von Feedback in Unternehmen – zwischen mobilen Apps und Echtzeit-Dashboards? In I. Jöns & W. Bungard (Hrsg.), *Feedbackinstrumente im Unternehmen* (S. 229–242). Wiesbaden: Springer Gabler.
- Wilson, H. J. (2013). Wearables in the workplace. *Harvard Business Review*, 91(11), 27–28.
- Winkler, R., & Soellner, M. (2018). Unleashing the potential of chatbots in education: A state-of-the-art analysis. *Academy of Management Proceedings*, 2018(1), 15903.
- Wittenhagen, J. (2019). KI ist (noch) kein Problemlöser. *Lebensmittelzeitung*, 33, 39.
- Yano, K., Akitomi, T., Ara, K., Watanabe, J., Tsuji, S., Sato, N., Hayakawa, M., & Moriwaki, N. (2015). Measuring happiness using wearable technology: Technology for boosting productivity in knowledge work and service businesses. *Hitachi Review*, 64(8), 517–524.

- Yun, H., Domanska, M., Fortenbacher, A., Ghomi, M., & Pinkwart, N. (2016). Sensor data for learning support: Achievements, open questions and opportunities. In R. Zender (Hrsg.), *Proceedings of DeLF1 workshops 2016 co-located with 14th e-learning conference of the German Computer Society* (S. 28–39). Potsdam.
- Zalando. (2019a). Über Zonar. <https://corporate.zalando.com/de/newsroom/de/news-storys/ueber-zonar>. Zugegriffen am 03.12.2019.
- Zalando. (2019b). <https://corporate.zalando.com/de/newsroom/de/news-storys/unser-statement-zur-studie-der-hans-boeckler-stiftung>. Zugegriffen am 03.12.2019.
- Zeit-Online. (2019). Datenschutzbehörde prüft Mitarbeitersoftware von Zalando. <https://www.zeit.de/arbeit/2019-11/zonar-zalando-mitarbeiter-scoring-software>. Zugegriffen am 03.12.2019.
- Zhang, X., Zhou, Y., Ma, Y., Chen, B. C., Zhang, L., & Agarwal, D. (2016). Glmix: Generalized linear mixed models for large-scale response prediction. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (S. 363–372). San Francisco: ACM.
- Zhang, Y., Olenick, J., Chang, C. H., Kozlowski, S. W., & Hung, H. (2018). TeamSense: Assessing personal affect and group cohesion in small teams through dyadic interaction and behavior analysis with wearable sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(4), 39.1–22.
- Zhao, Y., Hryniwicki, M. K., Cheng, F., Fu, B., & Zhu, X. (2019). Employee turnover prediction with machine learning: A reliable approach. In K. Arai et al. (Hrsg.), *Proceedings of SAI intelligent systems conference* (S. 737–758). Cham: Springer.
- Zimmermann, T., & Romeiser, A. (2019). Flexibilisierung der Schichtarbeit: Potsdamer Modell schafft Gestaltungsspielräume für Chemieunternehmen. *CHEManager*, 6, 8.
- Zinn, B. (2018). *Ergebnisbericht zum Verbundvorhaben „Virtual and Analytics Service im Maschinen- und Anlagenbau – VASE“*. https://www.ife.uni-stuttgart.de/dokumente/bpt/bpt_forschung/VASE-Ergebnisbericht-2018.pdf. Zugegriffen am 23.11.2019.



Unterstützung und Ersetzung durch RPA

5

Zusammenfassung

In diesem Kapitel liegt der Fokus auf der Automatisierung von HR-Prozessen, insbesondere unter Verwendung von Robotic Process Automation (RPA). Zunächst werden die generellen Anwendungskriterien und Einsatzszenarien von RPA erläutert, bevor konkrete Use Cases aus vier HR-Bereichen beschrieben werden: Recruiting, On- und Offboarding, Stammdatenmanagement und Reporting, Personalbetreuung, Gehaltsabrechnung bzw. Reisekostenabrechnungsprozess. Bei der Diskussion der Beispiele werden kritische Aspekte aufgezeigt und auf die zentrale Rolle von Change Management eingegangen.

Softwareroboter können bei regelbasierten und wiederkehrenden Aufgaben eingesetzt werden, vor allem um Kosten zu sparen, eine fehlerfreie Aufgabenabwicklung zu sichern und Mitarbeitende von nervenden Routinetätigkeiten zu entlasten. Durch Einbindung von Algorithmen für die Bild-, Text- und Spracherkennung sowie die Anbindung an Produktivsysteme wird der Roboter „intelligenter“. Im Folgenden werden diese Nutzenversprechen und Anwendungskriterien systematisiert und dann auf etablierte Anwendungsfälle von RPA eingegangen (der Einsatz intelligenter Roboter, vulgo Chatbots, wurde bereits in den Abschnitten von Kap. 4 erläutert).

5.1 Nutzenversprechen und Anwendungskriterien

Wie oben bereits kurz beschrieben (Abschn. 3.7), können durch RPA sowohl die Produktivität erhöht als auch die Kosten verringert werden, wenn ein **Prozess strukturiert sowie regelbasiert** ist und **sehr häufig** durchgeführt wird. Neben diesen Hauptvorteilen nennen Berater und Wissenschaftler weitere (siehe Abb. 5.1): Da Softwareroboter Prozesse so ausführen, wie es ihnen vorgegeben wurde, und alle Datenänderungen dokumentiert werden, gibt es keine Probleme bei der Einhaltung gesetzlicher oder firmeninterner Regeln (Compliance). Zudem ist anders als bei traditionellen Automatisierungstechnologien wie Business Process Management Systems keine aufwändige Umprogrammierung der Softwareapplikation oder Backend-Integration notwendig, weshalb der Einsatz von RPA innerhalb von wenigen Wochen geschehen und der Return on Investment entsprechend frühzeitig generiert werden kann. Über betriebswirtschaftliche und technische Argumente hinaus wird gerne angeführt, dass durch die Automatisierung von repetitiven Prozessen die Mitarbeiter von „stupider“ Arbeit befreit würden.

Im Falle von **RPA** tragen zur Erreichung dieser Ziele vor allem folgende Aspekte der Softwarebots bei (Czarnecki et al. 2019, S. 804 f.; Lacity und Willcocks 2016b, S. 3; Scheer 2017):

- niedrigere Kosten (geringere direkte bzw. indirekte Personalkosten wie Gehälter bzw. Ausfallzeiten durch Arbeitspausen, Krankheitstage, gesunkene Raum-, Energie- und Softwarelizenzkosten)
- 24/7 Verfügbarkeit und direkte Reaktion



Abb. 5.1 Nutzenversprechen von RPA (in Anlehnung an Capgemini Consulting 2016, S. 11; Lacity und Willcocks 2016a, b)

- gleichbleibende, fehlerlose Qualität (auf Basis dessen, was programmiert wurde)
- lückenlose Dokumentation ist möglich (Compliance-Anforderungen können erfüllt werden)

Im Vergleich zu traditionellen Ansätzen der Prozessautomatisierung ist bei RPA von einem geringeren Umsetzungsaufwand bei vergleichbaren Nutzenpotenzialen auszugehen. Die sich daraus ergebende hohe Rentabilität ist durch unterschiedliche Projektbeispiele dokumentiert, wobei die Angaben zu **Kosteneinsparungen von 35 bis 80 %** reichen (Capgemini Consulting 2016; Lacity und Willcocks 2016a, b; Scheer 2017; Gartner 2018) und auch die Spanne für den **Return on Investment (ROI) mit 200 bis 800 %** sehr unterschiedlich ist (Lacity et al. 2015; Lacity und Willcocks 2016b). Ein Grund dafür ist: Die bei traditionellen Ansätzen der Prozessautomatisierung notwendigen Abstimmungen zwischen technischen Rahmenbedingungen und fachlichen Anforderungen entfällt. Allerdings ist zu beachten, dass ein Re-Design von Prozessen kein eigentliches Ziel ist und die Prozess-optimierung vor der Automatisierung stattfinden muss, da das RPA-System auf Basis der Ist-Prozesse angelernt wird (Langmann und Turi 2020, S. 23).

Der erste Stolperstein bei RPA-Projekten ist die **Auswahl eines geeigneten Prozesses**, denn schließlich kann es nicht darum gehen, etwas zu automatisieren, nur weil es technisch möglich ist (**technische Dimension**). Bei Kosten-Nutzen-Überlegungen (**wirtschaftliche Dimension**) sollte berücksichtigt werden, dass die RPA-Einführung nicht immer so reibungslos verläuft wie vom Anbieter versprochen. Z. B. kann schon eine nur leicht veränderte Benutzeroberfläche oder ein Update des RPA-Herstellers dazu führen, dass der Bot nicht mehr das richtige Feld für die Dateneingabe findet. Ebenso sollte der Qualifizierungsaufwand für die Fachabteilung monetär berücksichtigt werden, wenn sie das, was IT-Abteilungen regelmäßig tun, nun selbst machen sollen (Anforderungen in Software umsetzen, Testen, Bugs fixen, etc.). Außerdem können nachgelagerte Kosten entstehen, wenn sich in den Fachbereichen eine Schatten-IT etabliert, die dann doch irgendwann wieder zentralisiert oder abteilungsübergreifend harmonisiert werden soll. Wie hoch dieses **Lehgeld** ist, lässt sich schwer bestimmen, auch weil davon meist nur am Rande berichtet wird (siehe Schuster 2020). Über die technische und wirtschaftliche Dimension hinaus ist auch zu berücksichtigen, welches Verhältnis von menschlicher Arbeit und Maschinenarbeit angestrebt werden soll (**ethisch-moralische Dimension**) und was für die gegebenen Ressourcen einer Organisation machbar ist (**pragmatische Dimension**). Letzteres beinhaltet die Frage, wie mit Widerständen seitens der Mitarbeitenden umgegangen wird, die oft den Wegfall ihres Arbeitsplatzes befürchten und in jedem Falle neue Fähigkeiten erlernen müssen. Je nach Ausmaß der Automatisierung fallen nicht nur Rollen weg, sondern es entstehen auch neue, z. B. Prozessdesigner, Tester und RPA-Manager. Diese vier Dimensionen werden in

Tab. 5.1 Anwendungskriterien für RPA

Kriterien für RPA	Ausprägung der Anforderung
Abwicklungsvolumen	mittel bis hoch
regelbasierte Verarbeitung	hoch
Komplexität des Prozesses (Ausnahmen, Veränderungen des Input-Output-Verhältnisses)	niedrig bis mittel (Varianten eines Prozesses können abgebildet werden, veränderliche Inputs/Outputs kaum)
Reifegrad des Prozesses	hoch
Grad der Interoperabilität (unterschiedliche involvierte Systeme)	egal (je mehr, desto eher zeigen sich die Vorteile von RPA)
Grad des Compliance Risikos	egal (je höher, desto eher zeigen sich die Vorteile von RPA)
Umsetzbarkeit und Akzeptanz	ausreichende finanzielle und personelle Ressourcen, wenig Widerstand seitens der Mitarbeitenden, Sicherung der Ergebnisqualität, Verfügbarkeit, Korrektheit und Kundenzufriedenheit

Anwendungskriterien für RPA in Anlehnung an Lacity und Willcocks (2016a, b); Scheer (2017); Czarnecki und Auth (2018); Czarnecki et al. (2019)

Listen mit Auswahlkriterien für RPA nicht immer vollständig berücksichtigt, weil zwar auf die **Relevanz von Change Management** hingewiesen wird, der Fokus aber auf den technologischen und wirtschaftlichen Dimensionen liegt (Lacity und Willcocks 2016a, S. 47; Hindle et al. 2019, S. 27 f.; für ein Beispiel siehe Abschn. 5.5).

Von den immer wieder genannten **Selektionskriterien** (siehe Tab. 5.1) zur Auswahl geeigneter Prozesse für RPA sind die ersten die wichtigsten, denn ist der Prozess zu komplex oder nicht ausnahmslos regelbasiert, steigt der Aufwand und der Nutzen von RPA sinkt. Die Höhe des Abwicklungsvolumens wirkt hier als Multiplikator – sowohl in Richtung steigender Nutzen als auch in Richtung abnehmender, wenn der Prozess zu selten durchgeführt wird, weil dann der Aufwand für die Programmierung des Bots vergleichsweise hoch ist.

Langmann und Turi (2019, S. 21) haben ein **Scoring-Modell** für die RPA-Prozessauswahl entwickelt, das neben den oben genannten Kriterien weitere Aspekte beinhaltet, denen eine Gewichtung zugewiesen wird (z. B. erhalten die Minimalkriterien ein dreifaches Gewicht). Für alle Kriterien werden die Prozesse anhand einer Skala von 1–5 bewertet. Der „Automatisierungswert“ ergibt sich als gewichtete Summe über alle Kriterien und Bewertungen. Je höher der Wert, desto eher ist der Prozess für RPA geeignet (beispielhaft für einen Prozess siehe

Tab. 5.2). Da sich die in der Literatur veröffentlichten Kriterienlisten zumindest im Detail unterscheiden, muss **jede Organisation für sich definieren**, wie beispielsweise „Komplexität“ konkret gemessen werden soll. Ebenfalls ist über die Gewichtungen zu diskutieren, da die vorgeschlagenen **Werte auf praxisorientierten Daumenregeln** und nicht auf Gesetzmäßigkeiten beruhen. Zudem ist in der Praxis zu beobachten, dass Organisationen **Grenzwerte festlegen**, die erreicht werden müssen, damit RPA-Projekte angegangen werden. Bei Telefónica o2 gibt es beispielsweise die Regel, dass mindestens drei Vollzeitarbeitsplätze eingespart werden müssen, damit ein RPA-Projekt gestartet werden kann (Lacity und Willcocks 2016b, S. 31). Unter Berücksichtigung der vier, bereits oben genannten

Tab. 5.2 Scoring-Modell zur Auswahl von Prozessen für RPA

Prozesse	Prozess 1	Prozess 2	Prozess 3
Gewichtete Scoring-Summe (Automatisierungswert)	90
<i>Minimalkriterien (3-faches-Gewicht)</i>	57		
Grad der Regelbasiertheit	2		
Frequenz	3		
Grad der Standardisierung	4		
Standardisierte Datentypen	5		
Grad der Wiederholung	5		
<i>Zusatzkriterien (2-faches-Gewicht)</i>	28		
Umfang des Prozessvolumens	2		
Komplexitätsgrad	2		
Anzahl Ausnahmen	2		
Anzahl verwendeter Applikationen	3		
Anzahl Entscheidungspunkte	2		
Anzahl manueller Eingriffe	1		
Anzahl beteiligter Benutzer / User	2		
<i>Sonderkriterien (1-faches-Gewicht)</i>	5		
Anzahl verwendeter Sprachen	1		
Ausprägung des Sicherheitsrisikos	4		
Scoring-Modell für RPA-Prozessauswahl (in Anlehnung an Langmann & Turi 2019, S. 21)			

Scoring-Modell für RPA-Prozessauswahl (in Anlehnung an Langmann und Turi 2019, S. 21)

Dimensionen ist zudem zu überlegen, ob nicht eine ganz andere **Alternative zu RPA** zu bevorzugen wäre. Z. B. könnte der Funktionsumfang der Anwendungssoftware bzw. des Business-Process-Management-Systems angepasst werden, was zwar Programmieraufwand bedeutet, aber nicht nur einfach ein Workaround ist (manchmal „heilt“ RPA schlechte Prozesse bzw. IT-Systeme nur an der Oberfläche, nämlich der Repräsentationsschicht).

Prüft man HR-Prozesse auf diese Kriterien hin, so wird deutlich, dass es auf Sachbearbeitungsebene eine Reihe von regelgebundenen Prozessen gibt, die oft ablaufen und aufgrund unterschiedlicher Anwendungssoftware ein Hin und Her zwischen verschiedenen Masken und Dokumenten erfordern. Häufig genannt werden (Teil-)Prozesse in folgenden Bereichen (siehe Capgemini Consulting 2016, S. 27; Anagnoste 2018, S. 64 f.; Haliva 2018; UiPath 2019a, b):

- Recruiting, On- und Offboarding (siehe Abschn. 5.2),
- Personalplanung (siehe Abschn. 5.3),
- Personalbetreuung (siehe Abschn. 5.4)
- Vergütungsmanagement (siehe Abschn. 5.5)

Nachdem geeignete Prozesse identifiziert sind, muss eine **detaillierte Prozessdokumentation** bis zur Click-Ebene (Keystroke Level) erstellt werden, weil der Roboter genau wie ein Mensch auch alle Klicks und Tastaturanschläge ausführen muss. In diesem Detailgrad sind Prozessbeschreibungen meist noch nicht vorhanden, können aber vom Fachbereich relativ einfach erstellt werden. Neben Screenshots oder Video-Bildschirmaufnahmen bieten sich Swim-Lane-Diagramme an, weil mit ihnen die verantwortlichen Rollen bzw. organisatorischen Stelleneinheiten sowie beteiligten Softwareanwendungen abgebildet werden können. Für eine umfassende Beurteilung, ob sich die Automatisierung wirklich lohnt, sollten neben der reinen Dokumentation des Prozessablaufs noch folgende Informationen eingeholt werden: **Kosten** der einzelnen Prozessschritte, **Kapazitäten** und **Engpässe** im Prozess, notwendige **Prozessvarianten** und **Ausnahmen** sowie organisationale **Regeln**, die es zu beachten gilt (Langmann und Turi 2020, S. 23). Damit sind die notwendigen Aspekte der technischen und betriebswirtschaftlichen Dimension eines RPA-Projektes beschrieben. Zu erinnern ist allerdings noch an die ethisch-moralischen und pragmatischen Aspekte, die Haltung und ein **professionelles Change Management** erfordern. Letzteres sollte sowohl individuelle Faktoren (insbesondere die Fähigkeiten der Anwender und ihre Motivation) als auch strukturelle Elemente (insbesondere die Einbindung des Projekts in die organisationalen Strukturen und Abläufe) berücksichtigen (siehe Lacity und Willcocks 2016b).

Da für jeden der beschriebenen Anwendungsfälle ein eigener Softwareroboter lizenziert werden muss, wächst die Zahl der eingesetzten Bots schnell auf eine zwei- oder gar dreistellige Anzahl. Bei Ernst & Young in London sind z. B. 23 Roboter allein im HR-Bereich im Einsatz (Zielinski 2018). Dann kann ein **Robot-Controller die einzelnen Bearbeitungsfälle den entsprechenden Robotern zu-teilen**, indem er eingehende E-Mails nach Anhaltspunkten für den jeweiligen Fall durchsucht (z. B. Erfassung von An- und Abwesenheiten bzw. Änderung von erfassten Zeiten, Reports erstellen, etc.) und dann die E-Mail den zuständigen Robotern zuweist (Scheer 2017, S. 31).

5.2 Recruiting, On- und Offboarding

Im Recruiting gibt es eine Vielzahl von administrativen Aufgaben, die bei großen Unternehmen **sehr oft und immer wieder gleich ablaufen**: Bewerbungsunterlagen auf Vollständigkeit prüfen, Hintergrundchecks durchführen (z. B. ob die Berufszulassung eingetragen und noch gültig ist), Qualifikationen mit relevanten Stellenanforderungen vergleichen, Einladungen zu Interviews oder Ablehnungen (nach vordefinierten Regeln) verschicken, Personalstammdaten von Neueingestellten oder Anträge einpflegen (z. B. für die Anmeldung zur betrieblichen Altersvorsorge), Genehmigungsprozesse starten (z. B. Zugangskarten, Zugriffsrechte auf Laufwerke und IT-Systeme beantragen, Visitenkarten bestellen) oder neuen Mitarbeitern wichtige Informationen zum ersten Arbeitstag schicken (Anreise, Parken, Fahrradstellplätze, Kontakt zu Schlüsselpersonen, etc.). Gerade beim Onboarding einer neuen Mitarbeiterin sind oft **unterschiedliche Abteilungen mit jeweils unterschiedlicher Anwendungssoftware** beteiligt. Beispielsweise nimmt die HR-Abteilung die Daten einer Bewerberin aus der Software für Online-Assessments oder dem Application-Tracking-System, pflegt sie als neue Mitarbeiterin in ein HR-Stammdatenverwaltungssystem ein und übermittelt die Daten dann der IT und dem Facility Management, die daraufhin virtuelle und physische Zugänge verwalten. Beim Offboarding, also wenn Mitarbeiter kündigen, gekündigt werden oder ihr Arbeitsvertrag endet, laufen diese Prozesse nochmals mit anderen Vorzeichen ab (dass Offboarding auch Wissenssicherung und Nachfolgemanagement beinhaltet, bleibt hier ausgeblendet).

Beispiel: Onboarding bei UiPath (<https://www.uipath.com/blog/hr-automation-rpa>)

Einer der größten Anbieter von RPA-Lösungen, UiPath, hat den eigenen Onboarding-Prozess mittels RPA vereinfacht (UiPath 2019a, b). Neue Mitarbeiter scannen ihre Dokumente und schicken sie elektronisch zur Bearbeitung an die HR-Abteilung von UiPath. Da die verwendete RPA-Lösung das Modul zur optischen Zeichenerkennung (OCR) beinhaltet, kann der Softwareroboter erkennen, was auf dem Computerbildschirm angezeigt wird. Der Roboter ist so konfiguriert, dass er die Informationen, die HR benötigt, aus den gescannten Dokumenten herausliest und sie dann in die erforderliche HR-Datenbank einträgt. Nach dem automatischen Auslesen und Eintragen muss HR noch die Richtigkeit der vom Roboter in das System eingegebenen Informationen überprüfen. Zumindest ein Teil des Onboarding-Prozesses erfolgt aber automatisiert, was insgesamt dazu geführt hat, dass der Prozess nur noch ein Sechstel der Zeit benötigt.

Natürlich lassen sich nicht immer alle Teilschritte automatisieren. **Medienbrüche** (von der digitalen in die physische Welt und zurück) sind beispielsweise von einem Softwareroboter nur zu handhaben, wenn es sich um das Einscannen oder Ausdrucken von Dokumenten handelt – einen Brief oder ein Formular (z. B. einen ausgedruckten Lebenslauf) in den Scanner legen, kann der Softwareroboter natürlich nicht. Ebenfalls nicht geeignet sind Vorgänge, bei denen es **viele Ausnahmen und Sonderfälle** geben kann oder die **nicht regelgebunden** sind. Beispielsweise können Qualifikationen nur mit Stellenanforderungen abgeglichen werden, wenn die eingehenden Bewerbungen immer an der gleichen Stelle Informationen zur geforderten Qualifikation auflisten. Das ist letztlich nur über digitale Bewerbungsformulare zu gewährleisten. Selbst wenn die Automatisierung möglich ist, kann es unwirtschaftlich sein, einen Roboter einzusetzen, wenn z. B. pro Jahr nur ein paar Dutzend Bewerbungen eingehen. Auch kann die automatisierte Bearbeitung als unangemessen und nicht legitim wahrgenommen werden, z. B. bei automatischen Absagen oder Hintergrundchecks in Polizeiregistern oder sozialen Netzwerken. Deshalb kommt es in der Praxis oft zu einer **Arbeitsteilung zwischen Roboter und Mensch**: Während der Roboter die unbedenklichen Standardfälle bearbeitet, prüft der Mensch die komplexeren Fälle oder die, bei denen der Roboter sich nicht sicher ist (die Mindestwahrscheinlichkeit an Sicherheit kann von Prozessschritt zu Prozessschritt und zwischen Unternehmen variieren).

5.3 Personalplanung: Stammdatenmanagement und Reporting

Wie bereits beim Onboarding-Anwendungsszenario deutlich wurde, eignet sich RPA für die **Personalstammdatenverwaltung**, d. h. für alle Eintragungen und Änderungen von Grunddaten wie Name, Adresse, Familienstand, Kontoverbindung, Krankenversicherung, etc. von aktuellen und früheren Mitarbeitern, Bewerbern und Freelancern. Die Eignung ergibt sich daraus, dass einerseits meist strukturierte Datenformate vorliegen und andererseits ein ebenso fehlerloses wie konsistentes Vorgehen über mehrere Datenbanken notwendig ist. In Organisationen gibt es häufig unterschiedliche Softwareapplikationen, weil entweder auf die beste Lösung für ein Teilproblem statt auf Standardsysteme gesetzt wurde (Best-of-Breed-Ansatz) oder bei Zukäufen oder der Expansion ins Ausland nicht 100 %-ig sichergestellt werden konnte, dass vorhandene Systeme abgeschaltet werden. **Je heterogener die Softwareapplikationen sind, desto unterschiedlicher sind die Datenformate**, in denen die Organisationseinheiten ihre Personalinformationen erfassen, was auch durch Schnittstellen nicht immer sauber gelöst werden kann (Deloitte 2017, S. 88, 91). Trotz einiger Standardisierungs- und Automatisierungsanstrennungen werden in HR-Abteilungen immer noch viele Daten und Dateien manuell in Softwareapplikationen geladen, geändert und weitergegeben. Da dies aufwändig und fehleranfällig ist, bietet sich der Einsatz von RPA an. Allerdings ist zu klären, ob die Anzahl an Vorgängen die Einführung von Softwarerobotern wirtschaftlich rechtfertigt.

Softwareroboter eignen sich für die **Erstellung von rein quantitativen Reports**, weil hier strukturierte Daten regelbasiert verarbeitet werden. Ein Beispiel sind regelmäßige Auswertungen über den Personalbestand, wenn diese nach unterschiedlichen Mitarbeitergruppen, Divisionen und Regionen sortiert werden sowie Ein- und Austritte enthalten sollen. In Organisationen, die hierfür manuell Reports aus unterschiedlichen Applikationen (z. B. SAP, Excel) ziehen, um dann die Daten übersichtlich zusammenzuführen, ermöglicht RPA eine deutliche Reduktion des Aufwands und des Fehlerrisikos. Schuster (2020) berichtet z. B. von einem Fall, bei dem die Erstellung eines Reports für die Personalplanung von 8 Stunden pro Woche auf 6 Min pro Woche reduziert werden konnte – innerhalb von zweieinhalb Tagen und ohne die Fehler, die vorher beim Kopieren von Daten und Erstellen von Pivot-Tabellen regelmäßig gemacht wurden. Die Kennzahlen, die in den Bericht eingehen (u. a. Headcount im Konzern sowie in allen Tochtergesellschaften, Ein- und Austritte, Fehlzeiten, Krankenstand) waren über neun Exceldateien mit insgesamt 17 Blättern und ca. 1000 Zeilen verteilt, weshalb Fehler bei der manuellen

Bearbeitung wahrscheinlich sind. Bei der Erstellung des Reports agiert der Roboter wie ein Mitarbeiter: Er loggt sich in die Quellapplikation ein, führt die gewünschten Abfragen durch, kopiert die abgefragten Daten in eine zentrale Excel-Datei und versendet den Report an die relevanten Stellen in der Organisation, hier an das Management Board. Neben der Zeitersparnis und gestiegenen Qualität des Reports konnte in diesem Fall die Mitarbeitermotivation gesteigert werden, weil eine unliebsame Tätigkeit wegfiel. Einen ähnlichen Use Cases beschreibt UiPath (2019b).

5.4 Personalbetreuung

Eine häufige Aufgabe in HR-Abteilungen ist es, Dokumente über Mitarbeitende zu erstellen oder weiterzuleiten, z. B. **Krankenscheine, An- und Abwesenheiten bzw. Urlaube, Überstundenanträge und -genehmigungen, Arbeitszeugnisse, Dokumente für Kindergarten- oder Hortbetreuung, Abrechnung von Dienstreisen, Administration von Trainings** (Information, Genehmigung, Einschreibung und Abrechnung) (Capgemini Consulting 2016, S. 27 f.; Anagnoste 2018, S. 64 f.; Haliva 2018). Die automatisierte Bearbeitung kann sich insbesondere für Personaldienstleister lohnen, also jene Firmen, die HR-Prozesse von ihren Kunden übernehmen. Nimmt man beispielsweise für die Bearbeitung eines Krankenscheins im Durchschnitt vier Minuten pro Vorgang an und rechnet mit 2500 Krankenscheinen pro Monat, so ergeben sich durch die Automatisierung von 90 % aller Fälle bereits Einsparungen in Höhe von 900 Minuten oder 15 Stunden pro Monat, womit sich die RPA-Einführung nach 6 Monaten amortisiert (UiPath 2019b). Auch bei der Erfassung von Arbeitszeiten wird RPA von großen Organisationen eingesetzt, insbesondere zur Validierung der Daten. Die Softwareroboter vergleichen dann z. B. die eingereichten An- und Abwesenheitsnachweise mit den Login-Zeiten im Firmennetzwerk und senden eine Nachricht bei fehlenden Informationen oder Inkonsistenzen (Haliva 2018).

Manchmal stößt ein Ereignis unterschiedliche, aber **miteinander verbundene HR-Prozesse** an. Z. B. wird aller Voraussicht nach die Versetzung einer Mitarbeiterin von Spanien nach Deutschland eine neue Bedarfsanforderung im spanischen Recruiting-System nach sich ziehen und eine Gehaltsanpassung dieser Mitarbeiterin in der Gehaltsabrechnung bzw. der Payroll-Software. Zusätzlich könnte es sein, dass eine Relocation-Firma, mit der das Unternehmen einen Rahmenvertrag hat, beauftragt wird, für diese Mitarbeiterin eine neue Wohnung zu suchen, ihren Umzug zu organisieren und einen Sprachkurs zu buchen. Manche HR-Software-Suiten können die Teilprozesse über die entsprechenden Software-Module verknüpfen.

Wenn aber unterschiedliche Softwareprogramme verwendet werden, geht das nicht oder nur unter hohem Programmieraufwand für die Schnittstellen. Hier kann sich RPA lohnen.

Moderne HR-Applikationen stellen viele **Self-Service-Funktionen** rund um das Antrags- und Auskunftsmanagement zur Verfügung, weshalb die oben genannten Fälle eher von den Mitarbeitenden als von der HR-Abteilung erledigt werden. Dann braucht es **keine Investitionen in RPA**. Wer zudem Kosten für die HR-Software sparen möchte, kann die kostenlose Lösung von Factorial testen – und sollte dabei aber den alten Merksatz bedenken „Wenn es nichts kostet, hat man mit seinen Daten dafür bezahlt“. Factorial ermöglicht unter anderem die digitale Verwaltung von Unternehmens- und Mitarbeiterdokumenten, die Arbeitszeiterfassung (inklusive Urlaubskalender) und bietet ein Mitarbeiterportal.

Personalbetreuung ist ein Feld, in dem **Chatbots einen qualitativen Mehrwert** bieten können. Da es in diesem Abschnitt um RPA gehen und einige Chatbot-Anwendungen bereits weiter oben vorgestellt wurden, werde ich hier nur ein Beispiel beschreiben. Bei KPMG soll der Chatbot Aava die HR Service Line entlasten und die Antwort- sowie Servicezeiten verbessern. Obwohl der Chatbot nur Fragen zum Entwicklungsgespräch beantworten kann, und das auch nur mit vorformulierten Antworten, konnten durch die Implementierung die Telefonanfragen um bis zu 55 % gesenkt werden (KPMG 2019).

5.5 Gehaltsabrechnung, Reisekostenabrechnungs-Prozess

Der Einsatz von RPA im Vergütungsumfeld fokussiert sich auf Aspekte der Gehalts- und Reisekostenabrechnung, ist aber auch hier nicht unumstritten. Während manche ein hohes Automatisierungspotenzial sehen (Capgemini Consulting 2016, S. 27; EY 2016a, b, S. 2), betonen andere Studien die Komplexität von Vergütungsthemen, die von Ausnahmen bzw. situativen Anpassungen getrieben ist (z. B. Saison- und Mehrarbeit, Boni- und Akkordlohnberechnung, lokale Gesetzgebung) (Cloudpay 2018, S. 7, 14). Einigkeit besteht darin, dass vor allem vorbereitende und prüfende Tätigkeiten automatisiert werden können. Dazu gehören die **Sammlung und Validierung der benötigten Daten** für die Gehaltsabrechnung, die **Meldung bei Behörden, Sozialversicherungen und Krankenkassen** und das Prüfen der berechneten Brutto-Netto-Geldbeträge. Hier spielt RPA seine Stärken aus, weil Daten zwischen unterschiedlichen Applikationen transferiert werden müssen und die korrekte Verarbeitung von hoher Relevanz ist.

Die Prüfung auf korrekte Überweisung erweist sich nicht nur dann als aufwändig, wenn Boni verrechnet werden sollen, sondern vor allem, wenn Reisekosten rückvergütet und die Beträge mit den eingereichten **Reisekostenabrechnungen** gegengecheckt werden müssen. Sind die Abrechnungen und entsprechenden Belege wie Hotelrechnungen oder Taxiquittungen eingescannt, muss der Softwareroboter die Informationen darauf erkennen. Dazu benötigt er mindestens eine Texterkennungsfunktion (Optical Character Recognition), welche heutzutage über ein KNN/DL implementiert ist (Abschn. 3.5). Sind die unterschiedlichen Belegarten, Zahlungsbeträge und Zahlungsformen erkannt, berechnet der Softwareroboter die zu erstattenden Beträge gemäß der Reisekostenrichtlinie. Dann prüft er die Passung zwischen Belegbeträgen und Reisekostenabrechnung und fordert gegebenenfalls fehlende Belege oder Informationen an. Sollte alles korrekt sein, veranlasst er den Überweisungsvorgang (Scheer 2017 S. 34). Zwar können HR-Software-Suiten diese Prozessschritte auch automatisieren oder unterstützen, aber weil in vielen Organisationen unterschiedliche Applikationen am Prozess beteiligt sind, kann RPA sinnvoll sein. Beispielsweise erfolgen bei der Deutschen Bahn die Anträge zur Abrechnung einer Dienstreise bzw. der damit anfallenden Reisekosten dank einer RPA-Lösung zu 92,6 % automatisch (Swiss Post Solutions für die Deutsche Bahn 2019).

Dass der Softwareroboter tatsächlich den kompletten Gehalts- oder Reisekostenabrechnungsprozess übernimmt, ist selten der Fall. In der Praxis lässt sich eher beobachten, dass **Teilaktivitäten automatisiert** werden, z. B. bei OpusCapita, einem finnischen Dienstleister für Auftragsabwicklungs- und Bezahlprozesse. Hier wurde nur die Qualitätskontrolle der Daten automatisiert, die von Kunden an OpusCapita zur Weiterverarbeitung geliefert werden. Der Softwareroboter namens Roger nimmt einen Abgleich zwischen geänderten Mitarbeiterdatensätzen im Personalstammdatensystem beim Kunden und den Daten im Gehaltsabrechnungssystem von OpusCapita vor, und sendet den Payroll-Mitarbeitern danach einen Statusbericht per E-Mail (siehe Abb. 5.2). Vor der RPA-Lösung wurden manuell Screenshots mit Notizen versehen und verschickt, was einen täglichen Aufwand von 30 bis 60 Minuten bedeutete, der nun entfällt (Hallikainen et al. 2018, S. 44).

Selbst bei diesen wenigen Aktivitäten gab es bei OpusCapita **drei Probleme**. Erstens wurden ein paar wenige Ausnahmefälle nicht berücksichtigt, z. B. die Gehaltsabrechnung für Mitarbeitende, die nur für einen Ferienjob eingestellt wurden. Zweitens traten Fehlermeldungen auf, weil der Roboter zu schnell arbeitete und nicht auf die Antworten der verknüpften Anwendungen wartete. Drittens war die Verbindung zwischen der SAP-HR-Anwendung des Kunden und dem Gehaltsabrechnungssystem bei OpusCapita nicht immer funktionsfähig, sodass der Roboter nicht alle Daten empfangen konnte (Hallikainen et al. 2018, S. 46). Zwar konnten

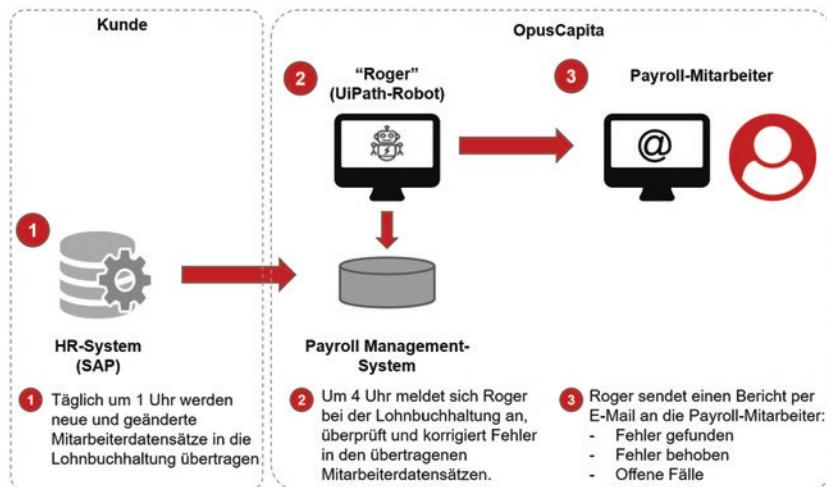


Abb. 5.2 Automatisierter Qualitätskontrollprozess für Payroll-Daten (eigene Erstellung in Anlehnung an Hallikainen et al. 2018, S. 45)

all diese technischen Probleme behoben werden, aber es zeigt dennoch, dass RPA-Lösungen neue Probleme mit sich bringen. Zumal die personalbezogenen Herausforderungen damit noch gar nicht thematisiert sind. Zu der **Befürchtung, dass Menschen durch Roboter ersetzt werden**, kommt der **Schulungsaufwand** für die RPA-Software hinzu, die dann auch noch zu **Stress und Frust statt Mitarbeiterzufriedenheit** führen kann, wenn sie nicht richtig funktioniert.

Diese „weichen Faktoren“ kommen bei RPA-Einführungen nicht selten zu kurz, obwohl **Change Management erfolgskritisch** ist (Hindle et al. 2019, S. 27). **Vor der Einführung** gilt es, Befürchtungen ernsthaft zu begegnen, Mythen über RPA zu widerlegen und die Mitarbeitenden in der Bedienung der Software zu schulen. Da RPA typischerweise eher von der Fachabteilung als von der IT-Abteilung verantwortet wird, müssen die Fachexperten aber nicht nur die Software bedienen können, sie müssen auch kompetent im Umgang mit Datenschutz- und Compliance-Themen sein (Lacity und Willcocks 2016a, S. 47 f.). **Während der Einführung** muss damit gerechnet werden, dass die anfängliche Euphorie und Motivation aufgrund von (Software-)Problemen bei der Automatisierung und Rückschlägen während des Projekts zurückgehen. In dieser Phase kann es dazu kommen, dass Erwartungen enttäuscht werden, weil die Prozesse oder die Softwareumgebungen doch komplexer sind als gedacht und organisationale Regeln oder Compliance-Aspekte (insbesondere Datensicherheit und -zugriff) einer Automatisierung entgegenstehen

(Langmann und Turi 2020, S. 53, 40 f.). Spätestens **nach der Einführung** müssen neue Rollen und Karrierepfade implementiert werden. Manche der neuen Rollen sind schon für die Einführung relevant, z. B. Prozessdesigner und Tester. Andere Rollen sind erst wirtschaftlich sinnvoll, wenn ein Unternehmen viele Softwareroboter im Einsatz hat, z. B. RPA-Manager in einer zentralen Abteilung, die alle RPA-Projekte steuert (Lacity und Willcocks 2016a, S. 47 f.; Hindle et al. 2019, S. 29 f.). Auch für die Mitarbeitenden, bei denen Aktivitäten durch RPA ersetzt wurden, müssen, je nach Ausmaß der Automatisierung, neue Einsatzmöglichkeiten gefunden werden. In diesem Kontext ist es interessant, dass Mitarbeitende Entlassungen eher akzeptieren, wenn der eigene Arbeitsplatz durch eine Maschine oder einen Roboter ersetzt wird. Dies hat eine Studie mit mehr als 2000 Teilnehmenden ergeben (Granulo et al. 2019). Ein Grund dafür ist das eigene Selbstwertgefühl: Da wir uns mit anderen Menschen, nicht aber Maschinen vergleichen, fühlen wir uns nicht so gekränkt, wenn wir durch einen (Software-)Roboter ersetzt werden. Das gilt jedoch nicht, wenn der Arbeitsplatz eines Kollegen auf dem Spiel steht. Die Implikation: Möglicherweise ist der Widerstand gegen Arbeitsplatzverluste, die durch Robotereinsatz verursacht werden, geringer als wenn sie durch Outsourcing an andere Arbeitnehmende außerhalb der Firma verloren gehen (Granulo et al. 2019, S. 1068).

Literatur

- Anagnosete, S. (2018). Robotic automation process – The operating system for the digital enterprise. *Proceedings of the International Conference on Business Excellence*, 12(1), 54–69.
- Capgemini Consulting. (2016). Robotic process automation – Robots conquer business processes in back offices. <https://www.capgemini.com/consulting-de/wp-content/uploads/sites/32/2017/08/robotic-process-automation-study.pdf>. Zugegriffen am 30.08.2017.
- Cloudpay. (2018). *RPA in global payroll: How robotic process automation is improving global payroll* (Whitepaper).
- Czarnecki, C., & Auth, G. (2018). Prozessdigitalisierung durch Robotic Process Automation. In T. Barton, C. Müller & C. Seel (Hrsg.), *Digitalisierung in Unternehmen* (S. 113–131). Wiesbaden: Springer.
- Czarnecki, C., Bensberg, F., & Auth, G. (2019). Die Rolle von Softwarerobotern für die zukünftige Arbeitswelt. *HMD Praxis der Wirtschaftsinformatik*, 56(4), 795–808.
- Deloitte. (2017). *Rewriting the rules for the digital age: 2017 Deloitte Human Capital Trends*. Deloitte University Press.
- EY. (2016a). Ernst & Young: Jobstudie 2016. [https://www.ey.com/Publication/vwLUAsset-SPI/EY_Jobstudie_2016/\\$FILE/EY-Jobstudie-2016-Karriere-und-Wechselbereitschaft.pdf](https://www.ey.com/Publication/vwLUAsset-SPI/EY_Jobstudie_2016/$FILE/EY-Jobstudie-2016-Karriere-und-Wechselbereitschaft.pdf). Zugegriffen am 01.07.2017.
- EY. (2016b). *Robotic process automation for HR & payroll* (White paper).

- Gartner. (2018). Manage robotic process automation. <https://www.gartner.com/en/finance/trends/robotic-process-automation>. Zugegriffen am 26.08.2019.
- Granulo, A., Fuchs, C., & Puntoni, S. (2019). Psychological reactions to human versus robotic job replacement. *Nature Human Behaviour*, 3(10), 1062–1069.
- Haliva, F. (2018). Top 10 HR tasks that are perfect for robotic process automation. <https://blog.kryonsystems.com/rpa/top-ten-hr-tasks-that-are-perfect-for-rpa>. Zugegriffen am 12.10.2019.
- Hallikainen, P., Bekkhus, R., & Pan, S. L. (2018). How OpusCapita used internal RPA capabilities to offer services to clients. *MIS Quarterly Executive*, 17(1), 41–52.
- Hindle, J., Willcocks, L. P., & Lacity, M. (2019). Keys to RPA success. https://www.blue-prism.com/uploads/resources/white-papers/KCP_Summary-Executive_Research_Report_Final.pdf. Zugegriffen am 12.12.2019.
- KPMG. (2019). *Chatbot Aava*. Internes Firmendokument.
- Lacity, M., Willcocks, L. P., & Craig, A. (2015). *Robotic process automation at Telefónica O2* (The outsourcing unit working research paper series). London School of Economics and Political Science, Paper 15/02.
- Lacity, M., & Willcocks, L. P. (2016a). A new approach for automating services. *MIT Sloan Management Review*, 58, 40–49.
- Lacity, M., & Willcocks, L. P. (2016b). Robotic process automation at Telefónica O2. *MIS Quarterly Executive*, 15, 21–35.
- Langmann, C., & Turi, D. (2020). *Robotic Process Automation (RPA)-Digitalisierung und Automatisierung von Prozessen*. Wiesbaden: Springer Gabler.
- Scheer, A.-W. (2017). *Performancesteigerung durch Automatisierung von Geschäftsprozessen* (Whitepaper). Saarbrücken: Scheer Holding.
- Schuster, A. (2020). Robotic Process Automation im Human Resource Management. In C. Gärtner (Hrsg.), *Smart human resource management* (S. X–Y). Wiesbaden: Springer Gabler.
- UiPath. (2019a). RPA for HR – Creating a more human workplace. https://dfe.org.pl/wp-content/uploads/2019/02/RPA-for-HR_Creating-a-More-Human-Workplace.pdf. Zugegriffen am 03.12.2019.
- UiPath. (2019b). Modern companies automate HR processes with RPA. <https://www.uipath.com/blog/hr-automation-rpa>. Zugegriffen am 13.12.2019.
- Zielinski, D. (2018). Robotic process automation gains a foothold in HR. <https://www.shrm.org/resourcesandtools/hr-topics/technology/pages/robotic-process-automationhr.aspx>. Zugegriffen am 13.12.2019.



Ethische und gesetzliche Richtlinien für Smart HR

6

Zusammenfassung

Mit dem Einsatz von digitalen Tools in der Personalarbeit stellen sich insbesondere aufgrund der automatisierten Datenverarbeitung und Entscheidungsfindung neue Fragen, die sowohl rechtliche als auch ethisch-moralische Aspekte betreffen. Zunächst wird deutlich gemacht, welche unterschiedlichen allgemeinen und personalspezifischen Fragestellungen zurzeit diskutiert werden. Anschließend werden einschlägige Gesetzesvorgaben und deren Implikationen erläutert. Aus dem Bereich der Ethik werden mehrere Richtlinien aus dem globalen wie nationalen Kontext in ihren Grundzügen beschrieben. Insbesondere werden die Richtlinien des Ethikbeirats HR Tech vorgestellt und auf offene Fragen aufmerksam gemacht.

Ohne Daten sind Algorithmen so nutzlos wie ein Kochrezept ohne Zutaten. Im Personalumfeld sind die **Daten oft sensibel, weil sie an Personen geknüpft sind** und nicht (Produktions-)Maschinen: Daten aus der Personalakte (z. B. Personalstammdaten, Abmahnungen, Zeugnisse, amtliches Führungszeugnis), Beurteilungs- und Gehaltsdaten oder Daten aus der Kontrolle des Arbeitsplatzes (z. B. Videoüberwachung). Weil die digitalen Tools mit ihren Algorithmen solche Daten verarbeiten, werden ethische und rechtliche Maßstäbe relevant, was zu vielen Diskussionen innerhalb der Personalszene und darüber hinaus geführt hat (siehe

z. B. Ajunwa et al. 2017; Blum und Kainer 2019; Bomhard 2018; Gärtner und Claßen 2016). Debattiert werden sowohl allgemeine als auch sehr spezifische Fragen, wobei deutlich wird, dass die Antworten umso differenzierter ausfallen sollten, je spezifischer die Fragen auf eine Personengruppe oder ein Thema abstellen (siehe Tab. 6.1):

Aus dieser kleinen Aufzählung wird deutlich: Die Themen, die es aus rechtlicher und ethischer Perspektive zu diskutieren gilt, sind breit gefächert und berühren nicht nur die genannten HR-Prozesse, sondern auch weitere. Es geht mehrheitlich um **Datenschutz** und **Privatsphäre**, **Transparenz** über und **Verantwortung** für automatisierte Entscheidungen, **Gerechtigkeit** und **Fairness** der automatisierten Entscheidungen sowie **gefälschte oder bösartige Inhalte** (Jobin et al. 2019, S. 391 f.). Und je nach Fall und Thema werden die Betroffenen und auch die Gesetzeshüter unterschiedlich sensibel reagieren und andere Anforderungen an die Anwendbarkeit von ADM-Systemen bzw. digitalen Tools stellen – es macht eben einen Unterschied, ob einer Mitarbeiterin ein Lernvideo aus ihrem Fachgebiet vorgeschlagen wird oder ob ihr gekündigt werden soll. Da die Themen nicht nur vielfältig, sondern viele davon gerade im HR-Umfeld kritisch sind, können (der europäische) Datenschutz und ethische Vorkehrungen zum Wettbewerbsfaktor für Anbieter digitaler Tools werden: Unternehmen, die nicht nur eine gesetzestreue, sondern auch „gute“ Lösung für kritische Anwendungsfälle von KI bzw. ADM-Systemen gefunden haben, können diese besser verkaufen.

Im Folgenden können nicht alle, aber doch die zentralen Aspekte der deutschen gesetzlichen Vorgaben (Abschn. 6.1) und der ethischen Diskussion (Abschn. 6.2) erörtert werden. Eine detaillierte Aufarbeitung müsste konkrete Fälle vor dem Hintergrund eines bestimmten Rechtssystems betrachten. Denn es liegt auf der Hand, dass die Europäische Datenschutzgrundverordnung nicht deckungsgleich mit US-amerikanischem Recht und schon gar nicht mit den dahinterliegenden Werten und Weltanschauungen ist.

6.1 Zentrale gesetzliche Vorgaben

Genau wie jede andere Form der Personalarbeit, muss „Smart HRM“, das sich auf Daten und digitale Tools stützt, innerhalb des rechtlichen Rahmens stattfinden. In Deutschland sind die wichtigsten rechtlichen Normen in unterschiedlichen Gesetzen verankert: dem **Grundgesetz** (GG), dem **Allgemeinen Gleichbehandlungsgesetz** (AGG), dem **Betriebsverfassungsgesetz** (BetrVG) und dem **neuen Bundesdatenschutzgesetz** (BDSG-neu), das eine Konkretisierung und Ergänzung der **Europäischen Datenschutzgrundverordnung** (DSGVO) darstellt. Darüber hinaus können fallweise weitere Gesetzesbücher einschlägig sein (siehe Kramer 2019).

Tab. 6.1 Zentrale ethische und rechtliche Fragestellungen beim Einsatz von KI/ADM-Systemen

Allgemeine ethisch-rechtliche Fragen	Personalspezifische Fragestellung (Auswahl)
Darf es den gläsernen Menschen geben?	Dürfen Unternehmen Informationen über Bewerber in den sozialen Medien sammeln und für die Personalauswahlscheidung verwenden? Wie genau dürfen Arbeitgeber überwachen, ob die Mitarbeitenden die vertraglich geschuldete Arbeitsleistung erbringen?
Wie transparent sollen KI/ADM-Systeme sein, wenn sie Empfehlungen abgeben oder Entscheidungen treffen? Wie können Black Boxes vermieden werden (siehe DL)? Wie soll man Nutzern bzw. Betroffenen die Algorithmen erklären, wenn sie transparent gemacht wurden?	Wie transparent muss ein Algorithmus sein, der eine Bewerberin zur Einstellung vorschlägt? Wie transparent muss einer sein, der Mitarbeitern Vorschläge zur gegenseitigen Vernetzung (oder Trainings, interessante Inhalte, Dokumente/Links für eine spezifische Fragestellung) macht? Wie stellt man sicher, dass Bewerber und Mitarbeitende die zugrunde liegenden Algorithmen bzw. Modelle verstanden haben?
Sollen KI/ADM-Systeme sich als solche zu erkennen geben?	Sollen sich (Recruiting-)Chatbots als solche zu erkennen geben?
Wie soll man das Spannungsfeld aus Personalisierung und Anonymisierung handhaben, um einerseits nutzerorientierte Angebote zu ermöglichen und andererseits Datenschutz zu garantieren?	Wie können in der Personalentwicklung personalisierte Inhalte oder Maßnahmen zur Verfügung gestellt werden, wenn die Nutzer anonym bleiben sollen? Wie kann ein Recruiting-Chatbot den Verlauf einer Bewerbung einem Kandidaten zuordnen, wenn die Daten geschützt sein sollen?
Inwieweit wollen wir es zulassen, dass das menschliche autonome Entscheiden durch Maschinen ersetzt wird, auf die Gefahr hin, dass wir zum Objekt einer uns nicht bewussten Steuerung werden?	Soll man Menschen durch Maschinen beurteilen oder auswählen lassen? Sollen Wearables Daten über die physische Arbeitsbelastung (z. B. Schadstoffbelastungen, Lärm, Belastung des Muskel-Skelett-Systems) sammeln, für Analysen bereitstellen und Empfehlungen geben, auch wenn uns nicht bewusst ist, warum diese Vorschläge gemacht werden? Sollen Mitarbeitende Vorschläge für Lerninhalte, Vernetzungsmöglichkeiten, Karriereoptionen, etc. erhalten, trotz des Risikos, dass diese selektiv sind, zu einer Filterblase führen können und dadurch auch eine unbewusste Steuerung entfalten?

(Fortsetzung)

Tab. 6.1 (Fortsetzung)

Allgemeine ethisch-rechtliche Fragen	Personalspezifische Fragestellung (Auswahl)
Wie sollte Verantwortung zwischen den Entwicklern und Anwendern von ADM-Systemen verteilt werden?	Wer verantwortet eine getroffene Einstellungs- oder Beförderungsentscheidung und wie wird sie gerechtfertigt?
Wenn KI/ADM-Systeme weniger Fehler und akkurate Vorhersagen machen als Menschen, macht man sich dann schuldig, wenn man sie nicht einsetzt?	Wenn KI/ADM-Systeme zumindest bei manchen Aufgaben oder personalbezogenen Entscheidungen überlegen sind, machen sich HR-Professionals dann nicht schuldig, wenn sie sie nicht einsetzen?
Darf man KI/ADM-Systeme einsetzen, wenn dies sozial unerwünschte Konsequenzen haben kann?	Darf man KI/ADM-Systeme einsetzen, wenn dies zur Folge hat, dass Menschen arbeitslos werden und/oder zu viel(!) Freizeit haben?
Was ist legal (innerhalb des rechtlichen Rahmens) und was gerade noch legitim (von einer interessierten Öffentlichkeit oder einer Gesellschaft akzeptiert)?	Dürfen und sollen automatische Tests zur sprachbasierten Persönlichkeitsdiagnostik eingesetzt werden? Darf ein Arbeitgeber Hintergrundchecks über Social-Media-Plattformen (mit Erlaubnis) durchführen, auch wenn dadurch Informationen erlangt werden können, die nach geltendem Recht unzulässige Fragen des Arbeitgebers wären (z. B. die Frage nach dem Bestehen einer Schwangerschaft)?
Wie soll die Gefahr von Fake-Inhalten, die entweder durch KI-Systeme erzeugt wurden oder diese manipulieren sollen, bewertet und gebannt werden?	Wie soll die Gefahr, dass Bewerber Unterlagen oder Tests fälschen, bewertet und gebannt werden? Wie sollen HR-IT-Systeme vor manipulativen Eingriffen (z. B. Hacks) geschützt werden?

Ethisch-rechtliche Fragestellungen beim Einsatz von KI/ADM-Systemen (eigene Erstellung)

Beispielsweise fällt die Direktansprache, wie sie beim Active Sourcing der Fall ist, auch unter das Gesetz gegen den unlauteren Wettbewerb (UWG), weshalb der Abwerbende bei dem Kandidaten dessen Einwilligung zum Abwerbeversuch einholen muss (Diercks 2019a, S. 101). Meist geht es bei der Diskussion um ADM und digitale Tools in der Personalarbeit jedoch um Themen wie eine mögliche **Diskriminierung** durch Algorithmen, **Transparenz** der Datenverarbeitung, die Einschränkung der **informationellen Selbstbestimmung** oder Behinderung der **freien Entfaltung der Persönlichkeit**. Für diese Themen, die im Folgenden näher erläutert werden sollen, sind Regelungsgrundsätze vor allem im GG (insbesondere § 3), dem AGG (insbesondere § 1, 3, 5, 8–10), sowie dem BDSG-neu in Verbindung mit der DSGVO zu finden (insbesondere §§ 4, 22–39 BDSG-neu sowie Art. 4, 9, 12–14, 22, 88 DSGVO).

6.1.1 DSGVO und BDSG: Transparenz und Informationspflicht, Zweckbindung und Datenminimierung

Weil bei allen digitalen Tools der Personalarbeit Daten verarbeitet werden, ist die Definition **personenbezogener Daten** relevant. Gemäß Art. 4 Nr. 1 DSGVO sind dies „alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person beziehen“. Das umfasst Daten wie Name, E-Mail-Adresse, Alter, IP-Adresse, Cookies oder Standortdaten, weil sie eine Person identifizierbar machen (Diercks 2019b, S. 186 f.). Dabei ist der Personenbezug auch gegeben, wenn er zwar nicht in den erhobenen Daten selbst, aber durch Hinzuziehen weiterer Informationen möglich wird – womit letztlich alle Daten zu personenbezogenen werden können (Huff und Götz 2019, S. 74). Damit ist die DSGVO das zentrale Normenwerk für Analytics und Automatisierung in der Personalarbeit.

Ein wesentliches Ziel der DSGVO ist **Transparenz über Datenverarbeitung**. So müssen Arbeitgeber gemäß Art. 30 DSGVO ein Verzeichnis aller Verarbeitungstätigkeiten anlegen, also auflisten, zu welchem Zweck welche Daten für welchen Empfängerkreis erhoben, gespeichert, verändert, übermittelt oder gelöscht werden, und welche technisch-organisatorischen Maßnahmen sowie Fristen für diese Verarbeitungsformen vorgesehen sind (Diercks 2019b, S. 187). **HR muss zu diesen Informationen einen leichten Zugang ermöglichen**, zumindest für die Betroffenen, die sowohl die Mitarbeiter des Unternehmens als auch Bewerber und Besucher der (Karriere-)Webseite sein können (Diercks 2019b, S. 189). Dies ist lediglich der administrative Aspekt der Transparenz- und Informationspflicht. Mehr Sprengkraft entfaltet die Frage, inwieweit der **Transparenzgrundsatz dem Einsatz selbstlernender Algorithmen entgegensteht**: Bilden ADM-Systeme eine Black Box, weil die Berechnungen, die zu einem Output führen, nicht mehr nachvollziehbar sind, ist die Verarbeitung nicht rechtmäßig (Blum und Kainer 2019, S. 26 f.). Für KNN/DL-Systeme ist das eher ein Stolperstein als für andere ML-Verfahren, deren Rechenwege ohne größeren Aufwand transparent gemacht werden können (siehe Abschn. 3.4 und 3.5).

Neben dem Transparenzgrundsatz und der Informationspflicht betont die DSGVO (insbesondere Art. 5) die **Datenminimierung**, die zudem mit dem **Zweckbindungsgrundsatz** verknüpft ist: Personenbezug soll überall dort vermieden werden, wo er zur Zweckerreichung nicht notwendig ist (Huff und Götz 2019, S. 74). Die Konsequenzen dieser Grundsätze wurden bereits im Zusammenhang mit Performance-Management-Systemen erläutert (Abschn. 4.4.1.4): Eine dauerhafte Datensammlung ohne konkreten Zweck ist gesetzeswidrig (z. B. die Aufzeichnung von Tastatureingaben und Mausbewegungen oder der Ein- und

Auslogdaten am Computer). Das ist kritisch für alle ML-Verfahren, die auf große Datenmengen angewiesen sind, und noch kritischer, wenn vor dem Einsatz von Algorithmen nicht klar ist, wofür man die Daten und die möglicherweise noch zu findenden Muster in den Daten gebrauchen könnte. Außerdem läuft es der Neigung mancher Data-Mining-Experten zuwider, auch einmal mit Daten „zu spielen“, die in ganz anderen Kontexten (und damit zu anderen Zwecken) erhoben wurden oder einfach da sind (z. B. Posts in sozialen Medien analysieren und auf Korrelationen mit Leistungsdaten hin untersuchen). Insofern passen die Forderungen der Datensparsamkeit und Zweckbindung nicht zu den Besonderheiten von Big-Data-Analysen. Da aber bei einem professionellen Analytics-Vorgehen die Definition des zu lösenden Problems der erste Schritt ist (siehe Abb. 3.8), handelt es sich grundsätzlich **nicht um zielloses Sammeln, sondern um gezielte Analysen** (Huff und Götz 2019, S. 75).

Da die Datensammlung nicht nur minimiert werden soll, sondern ein grundsätzlicher Datenschutz den Kerngedanken der DSGVO bildet, verlangt sie eine **Erlaubnisgrundlage** für den Einsatz datenbasierter Tools in der Personalarbeit. Eine Erlaubnisoption ergibt sich bei mitbestimmten Betrieben, da dort mit dem Betriebsrat eine **Betriebsvereinbarung** abgeschlossen werden kann. Eine weitere Möglichkeit ist es, von den Betroffenen eine **Einwilligung** einzuholen – diese Option kennen wir auch als Endverbraucher, wenn wir Häkchen in Pop-up-Fenster setzen und mit einer Antwort auf daraufhin eingehende E-Mails unsere Einwilligung zur Datennutzung im Rahmen eines Double-Opt-In-Verfahrens geben. Die Einwilligung ist nur dann wirksam, wenn sie freiwillig erfolgt, wobei die Freiwilligkeit im Rechtssinne dann gegeben ist, wenn dem Betroffenen auch bei Verweigerung keine Nachteile entstehen. Das führt im Fall eines (sich anbahnenden) Beschäftigungsverhältnisses zu der Frage, wie das Abhängigkeitsverhältnis zwischen Arbeitnehmer und Arbeitgeber zu beurteilen ist. Typischerweise wird von einem Machtungleichgewicht zugunsten des Arbeitgebers ausgegangen, womit eine freie Entscheidung des Arbeitnehmers als nahezu unmöglich betrachtet wird (Kraus 2018, S. 705). Z. B. wird immer wieder angeführt, dass von einer Freiwilligkeit bei Bewerbern oder Mitarbeitenden nicht ausgegangen werden kann, weil wohl nur die wenigsten diese Einwilligung verweigern würden, aus Angst, bereits dadurch Nachteile zu erleiden (Blum und Kainer 2019, S. 26). Dagegen verweist Diercks (2019b, S. 98) darauf, dass nach wie vor **personenbezogene Daten von Beschäftigten für Zwecke des Beschäftigungsverhältnisses verarbeitet werden dürfen**, z. B. wenn die Daten für die Einstellung erforderlich sind. Deshalb braucht es für die Verarbeitung von Bewerberdaten grundsätzlich keine Einwilligung, die Bewerber müssen nur über die Datenverarbeitung informiert werden. Je nach Anwendungsfall wird aber zu diskutieren sein, wie die Macht- bzw. Abhängigkeitsverhältnisse

nisse zwischen Arbeitgebern, Mitarbeitenden und Bewerbern sowie Kandidaten ausfallen.

Klar ist, dass **nicht in allen Fällen eine Einwilligung eingeholt werden muss**. Dies gilt für Bewerbungen und Datenanalysen, an denen der Arbeitgeber ein berechtigtes Interesse hat, ohne dass schutzwürdige Belange der Beschäftigten gefährdet wären. Darunter fallen beispielsweise **Analysen, die Faktoren für die Mitarbeiterzufriedenheit** herausfinden sollen, um dann z. B. Führungskräften Indikatoren zu liefern, welche Dimension wichtig für eine hohe Arbeitszufriedenheit sind. Die dafür benötigten Daten können – zweckgebunden – durch Befragungen oder aus elektronischen Personalakten gewonnen werden. Da die Identität der Beschäftigten hierfür uninteressant ist, können die Mitarbeiterdaten anonymisiert werden, womit Datenschutzbedürfnisse berücksichtigt sind (Huff und Götz 2019, S. 77). Ähnliches gilt für die **Berechnung der Wechselwahrscheinlichkeit**, die ebenfalls auf einem aggregierten Niveau mit Mindestgruppengrößen stattfinden kann. Obwohl personenbezogene Daten verarbeitet werden, können einzelne Mitarbeiter in der Ergebnisausgabe nicht identifiziert werden – für die Analyse sind nur die allgemeinen Treiber des Fluktuationsrisikos von Interesse. Komplizierter wird es, wenn Daten aus der individuellen Leistungsbeurteilung mit in die Analysen eingehen. Problematisch ist das, wenn leistungsstarke von leistungsschwachen Mitarbeitenden unterschieden werden, weil es dann nicht mehr nur um Bindungsmaßnahmen geht, sondern die Daten auch für Sanktionierung der weniger Produktiven verwendet werden oder diese zumindest Nachteile befürchten könnten (Huff und Götz 2019, S. 78). Als Lösung schlagen Huff – Mitgründer von functionHR, einem auf People Analytics spezialisiertem Start-up – und Götz vor, einen externen Auftragsverarbeiter mit der Analyse zu betrauen und als Ergebnis nur die Leistungsstarken mit erhöhtem Fluktuationsrisiko anzuzeigen (Huff und Götz 2019, S. 78). Inwiefern dies alle nicht gelisteten Mitarbeiter vor nachteiligen Maßnahmen (z. B. Kündigungen, Pflichtweiterbildungen) durch den Arbeitgeber schützt, bleibt allerdings offen.

Weitere Beispiele für weiterhin erlaubte digitale Tools in der Personalarbeit sind **Video-Interviews, Online-Assessments, Kulturmatcher und Active Sourcing, solange einige Voraussetzungen erfüllt sind**. Für Active Sourcing müssen drei Voraussetzungen gegeben sein: (1) ein berechtigtes Interesse in Form eines Personalgewinnungsinteresses auf Seiten des Suchenden; (2) kein entgegenstehendes, überwiegendes Interesse auf Seiten der Kandidatin; (3) durch die Kandidatin selbst offensichtlich öffentlich gemachte Daten (Diercks 2019a, S. 100). Davon kann insbesondere in engen Arbeitsmärkten und wenn die Kandidaten in den einschlägigen Business-Netzwerken ihren Jobsucher-Status auf aktiv gesetzt habe ausgegangen werden. Bei den Personalauswahlverfahren ist für die Frage der Zulässigkeit § 26

Abs. 1 BDSG-neu einschlägig: Es muss ein legitimer Zweck verfolgt werden, der nicht mit einem gleich geeigneten, aber mildernden Mittel erreicht werden kann, und die Datenverarbeitung muss verhältnismäßig im engeren Sinne sein (Diercks 2019a, S. 103). Beim Einsatz der jeweiligen Tools darf es sich nicht um eine „nicht erforderliche“ Maßnahme (88 DGSVO i.V.m. § 26 BDSG-neu) handeln, denn diese wäre unzulässig. Für die Erforderlichkeit müssen fachliche Tatsachen berücksichtigt werden, hier insbesondere die Erkenntnisse der Eignungsdiagnostik zur Vorhersagegütigkeit der jeweiligen Verfahren aber auch Daten über die Arbeitsmarktsituation. Berücksichtigt man zusätzlich die Arbeitgeber- und Bewerberinteressen, wird man zu dem Ergebnis kommen, dass diese digitalen Tools zulässig sind (Diercks 2019a, b). Aber: Werden Videointerviews benutzt, um über die Auswertung von Mimik und Sprache Rückschlüsse auf die Persönlichkeit zu ziehen, so wird der **Nachweis der Erforderlichkeit deutlich schwieriger**, weil die Konstruktvalidität zu kritisieren ist und die prognostische Güte (Vorhersage von Berufserfolg durch Persönlichkeit) gering ist (siehe Diercks 2017 und Abschn. 4.2.2.4).

Ein weiterer Regelungsgedanke der DSGVO ist, dass **Menschen und nicht Maschinen die Hoheit bei kritischen Entscheidungen behalten** (Art. 22). Um solche Entscheidungen handelt es sich beispielsweise bei Einstellungen, Beförderungen oder Entlassungen, weil die Kandidaten durch die Entscheidung potenziell einen Nachteil erleiden und sich ihr kaum entziehen können. Das Ergebnis der Datenanalyse darf die Personalentscheidung nicht vollends vorwegnehmen, oder anders formuliert: Eine Automatisierung rechtlich erheblicher Personalentscheidungen verbietet sich grundsätzlich nach Art. 22 DSGVO. Wird eine solche Personalentscheidung durch ein automatisiertes System allerdings vorbereitet, indem z. B. Punktwerte im Rahmen eines „Profiling“ berechnet werden (siehe DSGVO Erwägungsgrund 71), so sind diese vom Verbot des Art. 22 DSGVO nicht erfasst, sondern es ist eine **Interessenabwägung** nach § 26 Abs. 1 BDSG-neu vorzunehmen (Blum und Kainer 2019, S. 26; Martini 2019, S. 173). Ein grundlegendes Interesse des Arbeitgebers ist z. B., herauszufinden, wie sich die Arbeitsleistung entwickelt, denn diese ist ein wesentlicher Bestandteil des Arbeitsvertrags und der Beziehung zwischen einem Unternehmen und seinen Beschäftigten.

Es liegt in der Natur der Sache, dass Rechtsnormen als **allgemeine Regeln** formuliert sind, aber **auf die spezifische Situation eines einzelnen Falls angewendet werden müssen**, womit immer wieder Diskussionen um die Auslegung von Gesetzen entstehen (Ortmann 2003, S. 189 ff.). Beispielsweise waren die Datenschutzbehörden in NRW und Berlin der Auffassung, dass Video-Interviews in der Personalauswahl nicht zulässig wären, weil sie nicht objektiv geboten seien, um Stellen zu besetzen. Berücksichtigt man jedoch nicht nur die Richtlinien der DS-

GVO und Normen des BDSG, sondern auch Erkenntnisse der Eignungsdiagnostik und die allgemeine Arbeitsmarktsituation, so ändert sich die Einschätzung – und Video-Interviews gelten als zulässig (Diercks 2017, S. 751, 756). Dass die tagtägliche Praxis zudem eine andere ist, darf zwar nicht dazu führen, vom Sein auf das Sollen zu schließen, es zeigt aber, dass Bewerber und Personaler entweder kein Unrechtsempfinden beim Einsatz von Video-Interviews haben oder dies zumindest nicht zur Ablehnung der Tools führt.

6.1.2 AGG: Diskriminierung

Während die DSGVO und mithin das BDSG-neu vor allem die Themen Transparenz bzw. Informationspflicht, Datenminimierung und -schutz regeln, ist das **AGG auf Diskriminierungsfragen** ausgerichtet. Zunächst ist festgeschrieben, was geschützte Merkmale (Diskriminierungsmerkmale) sind, nach denen Personen nicht in ungerechtfertigter Weise benachteiligt werden dürfen: (ethnische) Herkunft, Geschlecht, Religion und Weltanschauung, Behinderung, Alter, sexuelle Identität bzw. Orientierung (§ 1 AGG). Über Art. 3 GG und Art. 9 DSGVO sind weitere Merkmale wie die politische Meinung, Sprache, Gewerkschaftszugehörigkeit, genetische oder biometrische Merkmale hinzuzunehmen. **Ungerechtfertigt ist eine Ungleichbehandlung dann, wenn ein sachlicher Grund fehlt.** Das bedeutet, dass eine Diskriminierung auch akzeptabel sein kann, wenn ein gesellschaftlich anerkannter sachlicher Grund dafür besteht, beispielsweise wenn es um die Vermeidung von Gefahren geht (z. B. beim Ausschluss mancher Personengruppen von der Benutzung bestimmter Geräte oder Fahrzeuge), die Intimsphäre anderer Personen geschützt werden soll (z. B. beim Ausschluss bestimmter Personen von der Nutzung sanitärer Anlagen) oder wenn bei einer Personengruppe kein Interesse an der Durchsetzung der Gleichbehandlung besteht (z. B. bei Preisnachlässen aus sozialen Gründen für Studierende oder Rentner oder günstigeren Preisen für Männer bei Tanzkursen, bei denen ein Frauenüberhang besteht) (Orwat 2019, S. 113).

Beim Einsatz von ADM-Systemen kann es neben der unmittelbaren Diskriminierung aufgrund der Korrelation mit einem Diskriminierungsmerkmal auch zu einer ungerechtfertigten Ungleichbehandlung kommen, wenn scheinbar neutrale Ersatzinformationen („Proxies“) herangezogen werden (um etwa Zufriedenheits- oder Produktivitätsdaten zu vergleichen). Eine solche **statistische Diskriminierung** liegt z. B. vor, wenn in der Analyse das Merkmal „Teilzeitbeschäftigung“ verwendet wird, weil mit damit eine Korrelation zum geschützten Merkmal Geschlecht besteht, da Frauen öfter in Teilzeit arbeiten (Orwat 2019, S. 28). In der Praxis werden häufig solche Ersatzmerkmale oder Näherungsgrößen verwendet, und zwar

aus **rationalen Gründen, die dem Wirtschaftlichkeitsprinzip folgen:** Die gewünschten Daten liegen oft entweder nicht vor oder können nur unter erheblichem Aufwand erhoben und analysiert werden. Zudem können unterschiedliche Angebots- und Preisgestaltungen im Sinne der Anbieter und Nutzer sein oder folgen der Differenz aus Angebot und Nachfrage. Insbesondere im Recruiting wird diese Idee der Zielgruppensegmentierung übernommen, wenn beispielsweise Stellenanzeigen über Social Media (mehr oder weniger) personalisiert ausgespielt werden (siehe Abschn. 4.2.2.2 und 4.2.2.3). Dass es dabei zu Ungleichheiten kommt, zeigt die oben bereits zitierte Studie von Lambrecht und Tucker (2019): Die Preise für das Ausspielen einer Anzeige sind für Frauen höher als für Männer. Ob dies – nach deutschem Recht – eine sachlich gerechtfertigte Diskriminierung ist, bleibt unklar, auch weil es schwierig ist die globalen Plattformbetreiber mit einem landesspezifischen Gesetz zu belangen. Doch auch innerhalb der Landesgrenzen ist die Frage nach der (un-)gerechtfertigten statistischen Diskriminierung selten einfach zu beantworten. Beispielsweise wurde beim Anwendungsfall der Abschätzung der Kündigungswahrscheinlichkeit gezeigt, dass Unternehmen die Entfernung zwischen Arbeits- und Wohnort als Merkmal heranziehen (siehe Abschn. 3.4 und 4.6.1.2). Dann gehen allerdings indirekt Informationen über den sozioökonomischen Status und unter Umständen sogar die ethnische Herkunft ein, weil die Wohnadresse auch Aussagen über Kaufkraft und Ethnie erlaubt (im Münchener Umland ist die Kaufkraft höher als im Stadtteil Hasenbergl und es leben dort weniger Menschen mit Migrationshintergrund). Dass die Information über den Wohnort verwendet wird, röhrt allerdings nicht aus einer intrinsischen Abneigung gegen eine bestimmte Personengruppe, sondern erfolgt mit einer anderen Zielsetzung und aus einem rationalen Kalkül heraus, nämlich um angesichts eines Informationsdefizits möglichst einfach und effizient die Entfernung zum Arbeitsplatz zu berechnen. Aus dem Spannungsfeld von Gesetz, Ökonomie und Gesellschaft ergibt sich also die **Frage, was rechtlich (noch) legal und was legitim ist.** Das Gesetz liefert dafür formalisierte Maßstäbe, nach denen im Einzelfall zu beurteilen ist, ob eine Form der statistischen Differenzierung als ungerechtfertigt gilt oder nicht (Orwat 2019, S. 30). Bei der Analyse großer Datenmengen erfolgt die Diskriminierung also nicht, um Personengruppen zu benachteiligen, sondern um verschiedene Eigenschaften oder Merkmale und Muster zwischen diesen zu identifizieren (Hagendorff 2019, S. 55). Eine unrechtmäßige (illegale) oder ungerechtfertigte (nicht legitimate) Diskriminierung ist das nicht per se. Spätestens an dieser Stelle sollte deutlich geworden sein, dass rechtliche und ethische Aspekte bei der Nutzung von Daten und digitalen Technologien ineinander übergehen.

6.2 Ethikrichtlinien

Werden Fragen nach dem Sollen und Dürfen gestellt, berührt man damit die Reflexion auf das, was allgemein als gut und schlecht bzw. richtig gilt. Man ist also im Bereich der Ethik und denkt über moralisches Handeln (und dessen Bedingungen) nach: **Welchen Werten, Normen, Prinzipien und Tugenden sollte unser Handeln folgen?** Werden sie eingehalten bzw. verletzt, reagieren wir mit Akzeptanz bzw. Ablehnung – in vielen Fällen herrscht auch Toleranz oder Ignoranz. Dies nicht zuletzt, weil es im Umfeld von Ethik und Moral mehr Fragen als allgemeingültige Antworten gibt. Bei der Antwortfindung sollen Ethikrichtlinien helfen. Von diesen wurden weltweit in den letzten Jahren eine ganze Reihe seitens staatlicher Einrichtungen, Forschern und Praktikern entwickelt – Jobin et al. (2019) untersuchen beispielsweise „nur“ 84 genauer. Diese Richtlinien sollen zwar nicht als Ersatz für rechtliche Regelungen verstanden werden, sehr wohl jedoch als Ergänzung und Vorarbeit für künftige Rechtsnormen. Dass es solcher Erweiterungen bedarf, liegt unter anderem daran, dass es manchen nicht genügt, dass Daten freiwillig gegeben und anonymisiert ausgewertet werden, weil sie nicht davon ausgehen, dass alle Personen die Tragweite der Datenfreigabe abschätzen können. Was erlaubt ist und was erlaubt sein sollte, sind deshalb trotz aller Berührungspunkte zwei unterschiedliche Fragen. Für eine abschließende Beurteilung, inwieweit die Ethikrichtlinien („soft law“) in Gesetze („hard law“) münden, ist es noch zu früh, jedoch müsste eine globale Harmonisierung erfolgen, die sehr schwierig ist (Jobin et al. 2019, S. 396 f.).

Im Kontext von KI werden viele Werte, Prinzipien etc. als wichtig erachtet, allen voran **Transparenz, Gerechtigkeit und Fairness, Vermeidung von Schäden bzw. Wohlwollen, Verantwortung, Privatheit, Freiheit und Autonomie, Vertrauen, Nachhaltigkeit, Würde und Solidarität** (Jobin et al. 2019, S. 395 f.). Die Werte variieren von Land zu Land. Aus Sicht der Datenethikkommission der Bundesregierung (2019, S. 43 ff.) sind folgende Handlungsmaßstäbe unverzichtbar: Würde des Menschen, Selbstbestimmung, Privatheit, Sicherheit, Nachhaltigkeit, Gerechtigkeit und Solidarität. Auch Effektivität wird als wichtiger Wert, insbesondere für Unternehmen, genannt, jedoch soll er nicht als höchster Wert betrachtet werden und darf die Entfaltung des Menschen nicht substanzell einschränken. Dies wäre z. B. der Fall, wenn Arbeitnehmer durch technische Systeme dazu gezwungen würden, die gesamten Arbeitsabläufe in den Dienst maximaler Effektivität zu stellen und dabei ihre Privatsphäre oder persönliche Integrität verletzt würde (Datenethikkommission 2019, S. 40).

Die Vielzahl an genannten Werten verrät, dass es sich nicht um einen kohärenten Kanon handelt, sondern dass die **vielen Werte und Richtlinien** zueinander in Verbindung gebracht werden müssen und dass dabei **Konflikte** entstehen. Z. B. lässt sich Diskriminierung kaum ohne Rekurs auf Fairness diskutieren, womit nicht nur Fragen der inhaltlichen und prozeduralen Fairness berührt sind, sondern auch, ob man Fairness als universellen Zweck setzen darf, der alle oder zumindest viele Mittel heiligt. Z. B. eine bestimmte Person(engruppe) zu benachteiligen (z. B. Aktionäre), um für viele alle einen Vorteil herauszuschlagen (z. B. für die Belegschaft), oder einen persönlichen Vorteil zu erreichen, ohne anderen zu schaden, zumindest nicht zum jetzigen Zeitpunkt (Neuberger 2006, S. 345 ff.). Dass unterschiedliche Gruppen unterschiedliche Interessen verfolgen, gilt auch im digitalen Zeitalter und diese Interessenvielfalt lässt sich kaum numerisch abbilden oder gar auf einen Nenner bringen. Verkompliziert wird die Gemengelage durch den Zeitbezug, da nie ganz klar sein kann, was spätere Generationen wollen und wie sie unser jetziges Handeln beurteilen werden. Eine andere Rolle spielt Zeit noch in Bezug auf das Davor und Danach von Entscheidungen, weil sich durch Entscheidungen Situationen ändern und ein Zustand herbeigeführt werden kann, der aus der einen Perspektive als wünschenswert gilt und aus einer anderen nicht. Ein Beispiel ist die oben bereits thematisierte 4/5-Regel (siehe Abschn. 4.2.2.2): Gemäß der „Uniform Guidelines on Employment Selection Procedures“ darf die Selektionsrate von Gruppen mit bestimmten Merkmalen (z. B. Hautfarbe, Geschlecht, Alter) nicht kleiner als 4/5 der Gruppe mit der höchsten Selektionsrate sein. Die Quote soll sicherstellen, dass gesellschaftlich benachteiligte Personengruppen nicht im Personalauswahlprozess diskriminiert werden („Procedure having adverse impact constitutes discrimination unless justified“). Damit steht die Vorgabe jedoch im Konflikt zu einer Regel der Amerikanischen Psychologenvereinigung (APA), die besagt, dass es keine feste Quotenregelung *nach* einem Auswahlverfahren geben sollte (Raghavan et al. 2019, S. 5). Was nach einer Auswahlentscheidung gelten soll hängt in diesem Fall davon ab, ob man der APA angehört oder Verfechter der 4/5-Regel ist. Nicht nur solche Situationen bedürfen einer situativen Güterabwägung, sondern letztlich gilt dies für alle Entscheidungssituationen, auch wenn uns im Alltag diese Probleme selten bewusst werden, was aber nur daran liegt, dass wir gelernt haben, wie *man das hier so macht* (siehe Ortmann 2003, S. 132 ff.).

Dass auch **Datenanalysten viele Entscheidungen treffen, die große Auswirkungen haben** können, ist in den letzten Jahren deutlich geworden (Diesner 2015). Letztlich müssen entlang aller neun Schritte des Vorgehens bei People-Analytics-Projekten (siehe Abb. 3.8) Entscheidungen getroffen werden, angefangen

bei der Frage, welches Problem überhaupt analysiert und optimiert werden soll, über die Suche und Bereinigung von Datensätzen sowie die Selektion von ML-Algorithmen bis hin zur Visualisierung und den Umsetzungsempfehlungen. Die Beteiligten treffen also viele Entscheidungen, die nicht wertfrei sein können, weil sie in ein soziales Gefüge eingebettet sind. Beispielsweise müssen sie im Recruiting-Bereich eine Definition der „Relevanz“ unterschiedlicher sozialer Medien und Kanäle entwickeln, wobei die Operationalisierung dieses Konstrukts mehr oder weniger gut gelingen kann (Welche Social Media-Plattformen werden überhaupt berücksichtigt? Gehen neben Messgrößen wie Klickzahlen auch Kriterien wie die Qualität der angezogenen Kandidaten ein? usw.). Selbst wenn es eine allgemein akzeptierte Definition und einen Datensatz für alle gäbe, könnten unterschiedliche Datenanalysten immer zu unterschiedlichen Ergebnissen kommen, wenn sie nicht die gleichen Variablen, Hypothesen und Algorithmen für ihre Analyse wählen. Eindrucksvoll zeigt diesen Effekt eine Studie von Silberzahn und Uhlmann (2015): 29 Forscherteams erhielten alle den gleichen Datensatz über Fußballspieler, aber nur 20 Teams fanden eine statistisch signifikante Korrelation zwischen Hautfarbe und roten Karten. Das lag allerdings nicht daran, dass die anderen unwissenschaftlich oder schlecht arbeiteten, sondern, dass sie mit ihren – gut begründeten – Analysemodellen schlicht andere Ergebnisse erhielten.

Weil viele kleine Entscheidungen große Auswirkungen haben können, ist es grundsätzlich ratsam, die verwendeten ML-Modelle offenzulegen, um die gewählten Annahmen und Algorithmen überprüfen zu können. **Transparenz** und **Nachvollziehbarkeit** bilden deshalb so etwas wie die Minimalforderungen vieler Ethikrichtlinien (siehe Zook et al. 2017; Jobin et al. 2019; Ethikbeirat HR Tech 2019 sowie die Informationsbox zur „Hambacher Erklärung zur Künstlichen Intelligenz“ der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder). Darüber hinaus gibt es weitergehende Forderungen, was die Kontrollierbarkeit angeht. **Wenn personenbezogene Daten verarbeitet werden oder die Entfaltung der Persönlichkeit betroffen** ist, dann werden spezielle **Einrichtungen zur Überwachung und Kontrolle** der verwendeten Daten und Algorithmen gefordert. Diese können technischer und/oder organisationaler Natur sein (siehe Abschn. 6.3). Das ist beispielsweise der Fall, wenn digitale Tools Entscheidungen treffen, die einzelne Menschen (z. B. Bewerberinnen, Mitarbeitende) oder die Gesellschaft als Ganzes betreffen und ihnen (dauerhaften) Schaden zufügen können (z. B. Verwehrung der Einstellung oder Beförderung).

Hambacher Erklärung zur Künstlichen Intelligenz (https://www.datenschutzkonferenz-online.de/media/en/20190405_hambacher_erklaerung.pdf)

1. KI darf Menschen nicht zum Objekt machen

Die Garantie der Würde des Menschen (Art. 1 Abs. 1 GG, Art. 1 GRCh) gebietet, dass insbesondere im Fall staatlichen Handelns mittels KI der Einzelne nicht zum Objekt gemacht wird. Vollständig automatisierte Entscheidungen oder Profiling durch KI-Systeme sind nur eingeschränkt zulässig. Entscheidungen mit rechtlicher Wirkung oder ähnlicher erheblicher Beeinträchtigung dürfen gemäß Art. 22 DS-GVO nicht allein der Maschine überlassen werden. Wenn der Anwendungsbereich des Art. 22 DS-GVO nicht eröffnet ist, greifen die allgemeinen Grundlagen des Art. 5 DS-GVO, die insbesondere mit den Grundsätzen der Rechtmäßigkeit, Zurechenbarkeit und Fairness die Rechte des Einzelnen schützen. Betroffene haben auch beim Einsatz von KI-Systemen den Anspruch auf das Eingreifen einer Person (Intervenierbarkeit), auf die Darlegung ihres Standpunktes und die Anfechtung einer Entscheidung.

2. KI darf nur für verfassungsrechtlich legitimierte Zwecke eingesetzt werden und das Zweckbindungsgebot nicht aufheben

Auch für KI-Systeme gilt, dass sie nur zu verfassungsrechtlich legitimierten Zwecken eingesetzt werden dürfen. Zu beachten ist auch der Grundsatz der Zweckbindung (Art. 5 Abs. 1 lit. b DS-GVO). Zweckänderungen sind mit Art. 6 Abs. 4 DS-GVO klare Grenzen gesetzt. Auch bei KI-Systemen müssen erweiterte Verarbeitungszwecke mit dem ursprünglichen Erhebungszweck vereinbar sein. Das gilt auch für die Nutzung personenbezogener Daten zu Trainingszwecken von KI-Systemen.

3. KI muss transparent, nachvollziehbar und erklärbar sein

Personenbezogene Daten müssen in einer für die betroffene Person nachvollziehbaren Weise verarbeitet werden (Art. 5 Abs. 1 lit. a DS-GVO). Dies erfordert insbesondere eine transparente Verarbeitung, bei der die Informationen über den Prozess der Verarbeitung und ggf. auch über die verwendeten Trainingsdaten leicht zugänglich und verständlich sind (Art. 12 DS-GVO). Entscheidungen, die auf Grundlage des Einsatzes von KI-Systemen erfolgen, müssen nachvollziehbar und erklärbar sein. Es genügt nicht die Erklärbarkeit im Hinblick auf das Ergebnis, darüber hinaus muss die Nachvollziehbarkeit im Hinblick auf die Prozesse und das Zustandekommen von Entscheidungen gewährleistet sein. Nach der DS-GVO ist dafür auch über die involvierte Logik ausreichend aufzuklären. Diese Transparenz-Anforderungen sind fortwährend zu erfüllen, wenn KI-Systeme zur Verarbeitung von personenbezogenen Daten eingesetzt werden. Es gilt die Rechenschaftspflicht des Verantwortlichen (Art. 5 Abs. 2 DS-GVO).

4. KI muss Diskriminierungen vermeiden

Lernende Systeme sind in hohem Maße abhängig von den eingegebenen Daten. Durch unzureichende Datengrundlagen und Konzeptionen kann es zu Ergebnissen kommen, die sich als Diskriminierungen auswirken. Diskriminierende Verarbeitungen stellen eine Verletzung der Rechte und Freiheiten der betroffenen Personen dar. Sie verstößen u. a. gegen bestimmte Anforderungen der Datenschutz-Grundverordnung, etwa den Grundsatz der Verarbeitung nach Treu und Glauben, die Bindung der Verarbeitung an legitime Zwecke oder die Angemessenheit der Verarbeitung.

Diese Diskriminierungsneigungen sind nicht immer von vornherein erkennbar. Vor dem Einsatz von KI-Systemen müssen deshalb die Risiken für die Rechte und Freiheiten von Personen mit dem Ziel bewertet werden, auch verdeckte Diskriminierungen durch Gegenmaßnahmen zuverlässig auszuschließen. Auch während der Anwendung von KI-Systemen muss eine entsprechende Risikoüberwachung erfolgen.

5. Für KI gilt der Grundsatz der Datenminimierung

Für KI-Systeme werden typischerweise große Bestände von Trainingsdaten genutzt. Für personenbezogene Daten gilt dabei auch in KI-Systemen der Grundsatz der Datenminimierung (Art. 5 Abs. 1 lit. c DS-GVO). Die Verarbeitung personenbezogener Daten muss daher stets auf das notwendige Maß beschränkt sein. Die Prüfung der Erforderlichkeit kann ergeben, dass die Verarbeitung vollständig anonymer Daten zur Erreichung des legitimen Zwecks ausreicht.

6. KI braucht Verantwortlichkeit

Die Beteiligten beim Einsatz eines KI-Systems müssen die Verantwortlichkeit ermitteln und klar kommunizieren und jeweils die notwendigen Maßnahmen treffen, um die rechtmäßige Verarbeitung, die Betroffenenrechte, die Sicherheit der Verarbeitung und die Beherrschbarkeit des KI-Systems zu gewährleisten. Der Verantwortliche muss sicherstellen, dass die Grundsätze nach Art. 5 DS-GVO eingehalten werden. Er muss seine Pflichten im Hinblick auf die Betroffenenrechte aus Art. 12 ff DS-GVO erfüllen. Der Verantwortliche muss die Sicherheit der Verarbeitung gemäß Art. 32 DS-GVO gewährleisten und somit auch Manipulationen durch Dritte, die sich auf die Ergebnisse der Systeme auswirken, verhindern. Beim Einsatz eines KI-Systems, in dem personenbezogene Daten verarbeitet werden, wird in der Regel eine Datenschutz-Folgenabschätzung gemäß Art. 35 DS-GVO erforderlich sein.

7. KI benötigt technische und organisatorische Standards

Um eine datenschutzgerechte Verarbeitung sicherzustellen, sind für Konzeption und Einsatz von KI-Systemen technische und organisatorische Maßnahmen gem. Art. 24 und 25 DS-GVO zu treffen, wie z. B. Pseudonymisierung. Diese erfolgt nicht allein dadurch, dass der Einzelne in einer großen Menge personenbezogener Daten scheinbar verschwindet. Für den datenschutzkonformen Einsatz von KI-Systemen gibt es gegenwärtig noch keine speziellen Standards oder detaillierte Anforderungen an technische und organisatorische Maßnahmen. Die Erkenntnisse in diesem Bereich zu mehren und Best-Practice-Beispiele zu entwickeln ist eine wichtige Aufgabe von Wirtschaft und Wissenschaft. Die Datenschutzaufsichtsbehörden werden diesen Prozess aktiv begleiten.

Ähnliche Forderungen bzw. Richtlinien hat der Ethikbeirat HR Tech (2019) formuliert:

1. Vor der Einführung einer KI-Lösung muss die Zielsetzung für die Nutzung geklärt werden. In diesem Prozess sollen alle relevanten Interessensgruppen identifiziert und eingebunden werden.
2. Wer KI-Lösungen anbietet oder nutzt, muss darauf achten, dass diese empirisch evaluiert sind und über eine theoretische Grundlage verfügen.

3. Wer KI-Lösungen einsetzt, muss sicherstellen, dass die Handlungsträgerschaft der Menschen bei wichtigen Personalentscheidungen nicht eingeschränkt wird.
4. Wer KI-Lösungen in seiner Organisation nutzt, muss diese in ihrer Logik verstehen und erklären können.
5. Organisationen, die KI-Lösungen nutzen, sind für die Ergebnisse ihrer Nutzung verantwortlich.
6. Wer personenbezogene Daten für KI-Lösungen nutzt, muss im Vorfeld definieren, für welche Zwecke diese verwendet werden und sicherstellen, dass diese Daten nur zweckdienlich erhoben, gespeichert und genutzt werden.
7. Vor bzw. beim Einsatz einer KI-Lösung müssen die davon betroffenen Menschen über ihren Einsatz, ihren Zweck, ihre Logik und die erhobenen und verwendeten Datenarten informiert werden.
8. Für die Nutzung in KI-Lösungen dürfen keine Daten erhoben und verwendet werden, welche der willentlichen Steuerung der Betroffenen grundsätzlich entzogen sind (Achten der Subjektqualität).
9. Wer KI-Lösungen entwickelt oder nutzt, muss sicherstellen, dass die zugrunde liegenden Daten über eine hohe Qualität verfügen und systembedingte Diskriminierungen ausgeschlossen werden.
10. Wer KI-Lösungen nach den vorliegenden Richtlinien einführt, soll transparent sicherstellen, dass die Richtlinien auch bei der betrieblichen Umsetzung und der Weiterentwicklung beachtet werden.

Zwar sind die Richtlinien auf den ersten Blick sinnvoll, jedoch gibt es (mindestens) **sechs noch zu lösende Problemfelder**:

1. Wie lässt sich die Forderung nach **vollständiger Transparenz umsetzen**? Drei Hindernisse müssten umgangen werden: Schutz von Firmengeheimnissen, das Black-Boxing-Problem und das Transparenzparadox. Erstens haben sowohl Start-ups als auch Softwarekonzerne im Rahmen einer Marktwirtschaft ein berechtigtes Interesse am Schutz ihrer entwickelten Verfahren. Selbst wenn der unwahrscheinliche Fall eintritt, dass durch Selbstverpflichtungen der Industrie das erhoffte Maß an Transparenz und Kontrolle gewährleistet wird, bleibt zweitens noch das Black-Boxing-Problem (siehe Abschn. 4.2.2.4). Zwar gibt es einige Tests und Methoden, mit deren Hilfe die Entscheidungslogiken von DL/ML-Modellen nachvollzogen werden können, jedoch gibt es noch nicht einmal eine allgemein akzeptierte Definition darüber, was eine Erklärung bzw. Offenlegung der Black Box genau beinhalten muss (Guidotti et al. 2018, S. 36). Drittens erfährt die Forderung nach vollständiger Transparenz eine weitere Verkomplizierung dadurch, dass in manchen Fällen weniger Transparenz als möglich oder sogar mehr Intransparenz geboten ist. Dies ist dann der Fall, wenn Verständlichkeit optimiert oder Manipulationsrisiken minimiert werden sollen.

- Wenn jeder Algorithmus, jedes Merkmal, jede Bedingung und jede Ausnahme beschrieben wird, wissen wir, dass es unwahrscheinlich ist, dass die Beschreibung verstanden oder gar gelesen wird (Nissenbaum 2011, S. 36). Das ist Transparenzparadox gemein: Mehr Transparenz über das ML-Modell führt nicht unbedingt zu mehr Klarheit, schon gar nicht bei den Nutzern, die den Wald vor lauter Bäumen nicht mehr sehen (wollen). Hinzu kommt, dass Manipulationen durch Hacker aber auch Nutzer erleichtert werden, wenn alle Spezifikationen offen liegen. So wurde z. B. immer wieder gezeigt, dass Klassifikationsalgorithmen, insbesondere KNN/DL, durch speziell konstruierte Beispiele („adversarial examples“), auf die das Modell sensibel reagiert, manipuliert werden können (Szegedy et al. 2013). Zwar sind solche Attacken auch möglich, wenn der ML-Algorithmus für die Angreifer eine Black Box ist (Brendel et al. 2017), aber dennoch werden Manipulationsversuche mit steigender Transparenz einfacher.
2. Wenn die Transparenz nicht im gewünschten Umfang gelingt, wie können dann auf Anwenderseite die **notwendigen Kompetenzen über ML** aufgebaut werden, sodass diese verstehen, wie die Algorithmen und die Technologie hinter den Angeboten von Start-ups und Softwarekonzernen funktionieren? Selbst, wenn man eine weitgehende Transparenz auf Anbieterseite annimmt, scheinen HR-Professionals aktuell weder das Wissen noch die technischen Mittel zu haben, um die richtigen Fragen zu stellen und die Antworten zu verstehen. Folgerichtig lautet eine zentrale Forderung des Ethikbeirats HR Tech (2019, S. 20), dass der notwendige Sachverstand aufgebaut werden muss. Das gilt für die ML-Modelle genauso wie für die Methoden zum Öffnen der Black Box. Beides ist aufwändig und anspruchsvoll, weshalb sicher nicht alle Anwender die entsprechende Expertise sowie Hard- und Software erwerben werden (können).
 3. Es wird nicht klar, **welcher ethische Ansatz den Richtlinien zugrunde** liegt. Das ist problematisch, weil es mehrere Theorien darüber gibt, welches Handeln als moralisch gilt – und diese nicht widerspruchsfrei sind. Um nur zwei herauszufiltern: Aus utilitaristischer Perspektive sollte eine Person ihre Handlungen rational abwägen und dann die Entscheidung treffen, bei der für die größte Zahl der größtmögliche Nutzen resultiert. Das kann der Tugendethik zuwiderlaufen, weil es dieser nicht um die Konsequenzen der Handlung geht, sondern um die Gesinnung, Werthaltung und die Pflichten, aus denen heraus eine Handlung zu erfolgen hat. Beispielsweise könnte argumentiert werden, dass eine sprachbasierte Personaldiagnostik vielleicht dem Einzelnen schaden könnte, aber Erleichterungen für eine Vielzahl an anderen Beteiligten im Recruitingprozess (z. B. Personaler, Fach- und Führungskräfte) bringt und so den größtmöglichen Gesamtnutzen generiert. Dagegen könnten Tugend- oder auch Vernunftethiker argumentieren, dass es nicht sein darf, dass Menschen aufgrund von Verfahren ausgewählt werden, deren Funktionslogik sie nicht verstehen und die sie nicht willentlich beeinflussen können.

4. Da auch menschliche Entscheidungen nicht diskriminierungsfrei sind: **Welche Alternative – maschinelles oder menschliches Versagen – ist akzeptabler** oder moralischer oder weniger diskriminierend? Diese Frage führt unweigerlich zur Diskussion, auf Basis welcher Ethiktheorie (z. B. Utilitarismus, Pflicht-/Tugendethik, etc.) diese Einschätzung begründet wird und warum nicht eine andere ethische Grundposition „besser“ ist – womit wiederum geklärt werden müsste, aus welcher Position und Argumentation heraus diese Bewertung vorgenommen wird, usw., ad infinitum. Mehr noch: Wer die Tools anderer als unethisch brandmarkt, kann dem Vorwurf ausgesetzt sein, dies nur aus Eigeninteresse zu tun, um seine eigenen Angebote besser verkaufen zu können. Dann neutralisieren oder konterkarieren die guten (ökonomischen) Folgen die reine (moralische) Gesinnung (Neuberger 2006, S. 334).
5. Wie kann die **Achtung der Subjektqualität prüfbar umgesetzt werden** und **warum ist dies eine notwendige Forderung?** Mit der Forderung des Ethikbeirat HR Tech (2019, S. 21), dass keine Daten erhoben oder verwendet werden dürfen, die der willentlichen Steuerung der Betroffenen entzogen sind, begibt man sich auf ein philosophisch hoch komplexes Gebiet. Der „Wille“ ist ein umstrittenes Konstrukt und die damit verbundenen Fragen nach dem Zusammenhang von Geist, Bewusstsein und Gehirn bzw. Materie sind seit Jahrhunderten ungelöst (siehe Libet 2004, S. 281 f.; Gärtner 2007, S. 149 ff.; Russell und Norvig 2013, S. 27 ff.). Unabhängig davon scheinen die Autoren bei dieser Forderung implizite Bewertungsmaßstäbe (siehe Tugendethik) anzulegen, was akzeptabel ist und was nicht (z. B. wird ein Lügendetektor-Test als nicht akzeptabel beschrieben). Wie verhält es sich aber mit Sensoren, mit Hilfe derer die Temperatur oder Lautstärke an einem Arbeitsplatz gemessen wird, um daraufhin die Belastungen zu analysieren und zu reduzieren? Die Datensammlung läge ebenfalls außerhalb der willentlichen Steuerung des Einzelnen – und wäre wohl dennoch für die Allermeisten akzeptabel. Auch müsste die gleiche Frage an nicht-digitale Personalarbeit gestellt werden: Wie viel willentliche Kontrolle haben Bewerber über das Zustandekommen der Bewertungen in einem klassischen Assessment Center oder einem multimodalen Interview? An diesem Beispiel zeigt sich, dass es nicht nur um die Steuerung der eigenen Verhaltensweisen, sondern auch um die Interpretationen der Handlungen anderer geht – und die sind meist der „willentlichen Steuerung der Betroffenen entzogen“, weshalb die Forderung kaum umsetzbar ist.
6. Wie lassen sich die **Nachteile (und Utopien) von voller Transparenz und absolutem Datenschutz vermeiden, ohne ihre Vorteile aufzugeben?** Wie bereits erläutert ist volle Transparenz weder realistisch noch in allen Fällen erwünscht. Gleicher gilt für totalen Datenschutz bzw. absolute Privatheit: Ein (normales)

Leben ohne Preisgabe von personenbezogenen Daten erscheint weder erstrebenswert noch möglich, weil wir dann keine Navigationssoftware benutzen, dem Arbeitgeber unsere Sozialversicherungsdaten nicht aushändigen oder unseren Ärzten keine Informationen über (Vor-)Erkrankungen geben dürfen. Privatsphäre ist kein binärer Wert (privat vs. öffentlich oder voll vs. nicht vorhanden), sondern sie ist situationsspezifisch. Ethikrichtlinien müssten also auf eine situative Privatheit bzw. die kontextgebundene Aufgabe des Datenschutzes abstellen. Diese Idee ist im Konzept der Kontextintegrität ausgearbeitet („**contextual integrity**“; Nissenbaum 2011), findet sich aber nicht in den oben zitierten Richtlinien.

6.3 Technische und organisatorische Vorkehrungen zur Lösung ethisch-rechtlicher Probleme

Diskriminierung durch Algorithmen ist seit einigen Jahren ein prominentes Thema unter Datenanalysten und KI-Forschern. Es haben sich mehrere Gruppen gebildet, die sich zum Ziel gesetzt haben, gegen mögliche algorithmische Diskriminierungen vorzugehen und für eine höhere Transparenz bei ML-Modellen zu sorgen. Zu finden sind diese Ansätze unter den Begriffen „Discrimination-aware Data Mining“ (DADM), „Fairness, Accountability and Transparency in Machine Learning“ (FATML) und „Explainable Artificial Intelligence“ (XAI). Vorgeschlagen wurden sowohl **technische als auch organisatorische Vorkehrungen entlang des Analyseprozesses**, also angefangen bei der Datensuche, der -aufbereitung und -auswertung bis hin zur Präsentation der Ergebnisse. Zum Teil werden die Maßnahmen auch von gesetzlichen Richtlinien gefordert (z. B. Art. 32 DSGVO). Von den vielen Vorschlägen und Vorgaben habe ich die aus meiner Sicht wichtigsten ausgewählt.

Technische Vorkehrungen

- **Anonymisierung** und Verschlüsselung von (Input- und Output-) Daten, um sie ohne Personenbezug zu verarbeiten (anonymisierte Daten können einer Person nicht mehr zugeordnet werden). Z. B. kann ein Hash-Wert aus der Personalnummer der Beschäftigten erzeugt werden, der zwar eine eindeutige Identifikation des Datensatzes erlaubt, aber von dem aus nicht mehr auf die Personalnummer zurückgeschlossen werden kann (Huff und Götz 2019, S. 75). Unter dem Begriff „**Differential Privacy**“ werden eine Reihe von Verfahren diskutiert, die das Problem lösen sollen, nichts über ein Individuum zu erfahren, während gleichzeitig nützliche Informationen über eine Gruppe generiert werden (Dwork und Roth 2014).

- **Pseudonymisierung** und Verschlüsselung von (Input- und Output-)Daten. Pseudonymisierung liegt beispielsweise vor, wenn die Personalnummer eines Mitarbeitenden verwendet wird, weil grundsätzlich mit weiteren Daten auf die betreffende Person zurückgeschlossen werden kann (man denke nur an ein Team, in dem alle zu unterschiedlichen Zeitpunkten in die Firma eingetreten sind, weshalb das Eintrittsdatum eine eindeutige Zuordnung der Personalnummer zu einer Person zulässt). Für die Pseudonymisierung gibt es mehrere Lösungen, unter anderem die Nichtangabe (die zu schützenden Daten werden weggelassen), Maskierung (die zu schützenden Daten werden mit einem konstanten oder sich ändernden Wert ersetzt, z. B. Austausch der letzten zwei Ziffern einer Postleitzahl mit zwei Nullen), und Shuffling (die Daten unterschiedlicher Datensätze werden miteinander vermischt, z. B. erhält Person „B“ die Adresse von Person „A“) (Isele et al. 2018, S. 21 f.).
- **Minimalprofile** durch Mitarbeitende erstellen lassen. Dazu legen Nutzer eigene Profile für eine HR-Lösung an, die nur rudimentäre Daten enthalten, wobei die Mitarbeitenden selber entscheiden, welche Daten sie einstellen.
- **Daten vor Verwendung prüfen.** Gute Datensätze spiegeln Offenheit und Vielfalt wider, während Datenmengen, die zu klein oder homogen sind oder zu viele seltene bzw. zufällige Ereignisse beinhalten, nicht verwendet werden sollten (Tambe et al. 2019, S. 18). Im Prinzip geht es darum, ML-Algorithmen in der Lernphase nur mit solchen Daten zu trainieren, die keine Verzerrungen haben und in keinem Zusammenhang mit ungerechtfertigter Diskriminierung stehen (Hagendorff 2019, S. 60). Dabei beginnt eine Verzerrung bereits bei nicht-ausbalancierten Datensätzen, wenn z. B. bei Sentiment-Analysen keine Balance zwischen negativen und positiven Texten vorliegt (siehe Abschn. 4.2.1.4). Sie geht aber weiter bis zu strukturellen bzw. gesellschaftlichen Verzerrungen, die sich in Datensätzen widerspiegeln können (siehe Abschn. 4.2.2.4). In Bezug auf die Zufälligkeit von Ereignissen können unsupervised ML-Algorithmen wie Cluster- oder Ausreißer-Analysen die menschliche Expertise komplementieren. In einem ersten Schritt werden die Datensätze auf unerwartete und irreführende Muster hin abgesucht, die bei der späteren Verwendung weiterer Algorithmen zu Diskriminierungen führen könnten. Sind solche Muster, Trends oder Abweichungen erkannt, können die entsprechenden Merkmale eliminiert werden oder zumindest die späteren Ergebnisse besser interpretiert und bewertet werden (Hagendorff 2019, S. 61). So könnte sich beispielsweise eine Erkrankung, die sich auf die Leistung auswirkt, als zufälliges Ereignis herausstellen oder als chronische Erkrankung. Und wenn ein Unternehmen nur einmal im Jahr eine Senior-Management-Position besetzt, so findet dieses Ereignis zu selten statt und man sollte die Daten nicht verwenden. Manche Software-Tools prüfen Datensätze automatisch auf solche Probleme und machen den Anwender gegebenenfalls darauf aufmerksam.

Organisatorische Vorkehrungen

- **Anonymisierung** von (Input- und Output-) Daten, um sie ohne Personenbezug zu verarbeiten. Z. B. können Mindestgrößen für die Stichproben sowie die Ergebnisausgabe festgelegt werden, sodass einzelne Mitarbeitende trotz des Personenbezugs nur unter Umständen identifiziert werden können, was die Intensität des Eingriffs in die Persönlichkeitsrechte stark abmildert (Huff und Götz 2019, S. 75). Eine andere Variante ist es, personenbezogene oder anderweitig sensible Merkmale erst gar nicht zu erheben („fairness through unawareness“). Auch die Verteilung der Zugriffsrechte auf die Daten an unterschiedliche Stellen, z. B. externe Dienstleister, kann einen höheren Grad an Anonymisierung ermöglichen (Hagendorff 2019, S. 62). Dennoch bleibt bei diesen Varianten ein Restrisiko der Deanonymisierung anonymisierter Daten, wenn und weil manchmal wenige (nicht-sensible) Merkmale genügen, um Personen doch zu identifizieren. So genügen laut einer Studie 15 demografische Merkmale, um 99,98 % aller Amerikaner trotz anonymisierter Datensätze zu identifizieren (Rocher et al. 2019). Als Daumenregel gilt: je sensibler die Daten sind, desto aufwendiger sollten die Maßnahmen zur Anonymisierung und Datensicherung sein, was sowohl die physische Speicherung der Daten als auch Zugriffsregeln (für die Administratoren und Nutzer im Unternehmen sowie die Software-Anbieter) beinhaltet. Diese Regelungen sollten in einer Betriebsvereinbarung festgehalten werden.
- Durchführung **regelmäßiger Sicherheitsaudits** zur Prüfung der Datensätze auf Schutz vor **unbefugten** Zugriffen sowie **regelmäßige Prüfung und Anpassung** der Modelle auf veränderte Situationen. Dazu gehört auch, dass die verwendeten Daten, Hypothesen und Modelle hinterfragt werden und die Interpretationen der vom ML-Modell gelieferten Ergebnisse evaluiert werden, um Fehlinterpretation zu minimieren bzw. daraus entstehende Risiken zu bewerten (siehe Silberzahn und Uhlmann 2015; Zook et al. 2017). Solche Überprüfungen sind aufwändig und es bedarf sowohl datenanalytischer und technologischer als auch fachlicher Expertise, um z. B. die Analyse der Kündigungswahrscheinlichkeit angemessen zu interpretieren und entsprechende Maßnahmen abzuleiten. Aufgrund der interdisziplinären Anforderungen wird es kaum reichen, nur eine Stelle einzurichten bzw. eine Person mit der Prüfung zu beauftragen. Adäquater wären funktionsübergreifende Projektteams, die zudem regelmäßig geschult werden müssen, damit sie die Möglichkeiten und Grenzen verschiedener und innovativer ML-Modelle einschätzen können. Schließlich eignen sich Algorithmen nur für bestimmte Fragestellungen (siehe Tab. 3.1 und 3.2).

- **Zertifizierungen und Prüfsiegel** können Laienanwendern den Aufwand abnehmen, sich ein eigenes Verständnis über die Funktionsweise von ML-Modellen bzw. KI-Anwendungen zu erarbeiten.
- **Visualisierung** der Logik der Algorithmen und der Ergebnisse. Es soll visuell nachvollzogen werden können, welche Daten in den Entscheidungsprozess eingeflossen sind und auf Basis welcher Einflussfaktoren eine Entscheidung gefällt wurde. Das beinhaltet auch eine Darstellung der Faktoren, die nicht oder nur in geringem Maße berücksichtigt wurden. Die Ergebnisse sollten so aufbereitet sein, dass sie auch ohne tiefes mathematisch-statistisches Wissen verstanden werden können. Dabei entsteht ein Spannungsverhältnis zwischen korrekter und nutzerfreundlicher Darstellung, das bei jeder Analyse neu austariert werden muss (siehe das Transparenzparadox in Abschn. 6.2). Zu beachten ist einerseits, welches Erkenntnissinteresse verfolgt wird, und andererseits, welche Vorkenntnisse die Adressaten haben (siehe Sinar 2018, S. 5 ff.). Dass zudem keine Manipulation der Interpretation durch die Darstellungsart verfolgt werden sollte, versteht sich von selbst. Gleichzeitig wird die Wahrnehmung und Meinungsbildung immer gelenkt, z. B. durch die Wahl von Farben, Größen(-verhältnissen) oder der Reduktion komplexer Themen auf einen Wert oder eine Signalfarbe. Die Vor- und Nachteile sowie Auswirkungen von Visualisierungen sind breit erforscht (z. B. Meyer et al. 2013; Gärtner und Huber 2015).
- **Sensibilisierung für kulturelle Einflüsse und Entwicklung der Organisationskultur:** Einerseits spiegeln Daten die kulturellen Gegebenheiten einer Organisation wider, andererseits ist die Art wie sie analysiert werden eingebettet in einen Kulturraum, sodass in den Algorithmen auch die Werte, Normen und Routines ihrer Macher eingeschrieben sind (Crawford 2013). Dass z. B. Unternehmen aus der Softwareindustrie dazu neigen eine „Bro Culture“ auszubilden, weil dort männliche und technikgläubige Weiße in der überwiegenden Mehrheit sind und untereinander ein überwiegend kumpelhaftes, aber anderen gegenüber aggressives Verhalten an den Tag legen, wurde mehrfach beschrieben (Zuboff 2015, S. 78 ff.; Gärtner 2018, S. 145). Eine Empfehlung ist, auf solche Auswüchse zu achten und ihnen durch Diversität und Dynamik in Bezug auf die sozio-demografische Zusammensetzung der Belegschaft entgegenzuwirken. Das schafft allerdings noch keine Abhilfe bei einem weiteren Problem: Wie kommen die Werte und Normen in ein Rechenmodell? Für ADM oder People Analytics müssten Normen und wertbeladene Begriffe (z. B. Fairness, Anti-Diskriminierung) numerisch formalisiert werden, soweit sie in die Datensammlung, -analyse und -präsentation eingehen. An diesem Problem scheitern und scheiden sich die (großen) Geister der Philosophie (man denke nur an den frühen und den späten Wittgenstein).

Die Empfehlungen und Forderungen sind unterschiedlich konkret und einzelne Maßnahmen allein werden nicht alle Herausforderungen lösen können. Manche empfehlen deshalb die getroffenen technischen und organisatorischen Vorkehrungen sowie etwaige (Ethik-)Richtlinien in einem Rahmenwerk im Sinne einer „Data & Analytics Governance“ festzuhalten. Ein solches Dokument bietet die Grundlage für den Umgang mit Daten in einer Organisation und sendet entsprechende Signale an die Stakeholder (Krotova und Fritsch 2019). Trotz aller Vorkehrungen werden sich alle Beteiligten und Betroffenen aber letztendlich damit arrangieren müssen, dass **es keine perfekte (diskriminierungs- und konfliktfreie) Personalarbeit gibt**.

Literatur

- Ajunwa, I., Crawford, K., & Schultz, J. (2017). Limitless worker surveillance. *California Law Review*, 105, 735–776.
- Blum, B., & Kainer, F. (2019). Rechtliche Aspekte beim Einsatz von KI in HR: Wenn Algorithmen entscheiden. *Personal Quarterly*, 71(3), 22–27.
- Bomhard, D. (2018). *Automatisierung und Entkollektivierung betrieblicher Arbeitsorganisation: Herausforderungen einer digitalen Arbeitswelt*. Wiesbaden: Springer.
- Brendel, W., Rauber, J., & Bethge, M. (2017). Decision-based adversarial attacks: Reliable attacks against black-box machine learning models. *arXiv preprint arXiv:1712.04248*.
- Crawford, K. (2013). The hidden biases in big data. <http://blogs.hbr.org/2013/04/the-hidden-biases-in-big-data>. Zugegriffen am 08.01.2020.
- Datenethikkommission der Bundesregierung. (2019). *Gutachten der Datenethikkommission der Bundesregierung*. https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/it-digitalpolitik/gutachten-datenethikkommission.pdf?__blob=publication-File&v=4. Zugegriffen am 05.01.2020.
- Diercks, N. (2017). Video-Interviews in Personalauswahlverfahren. *Datenschutz und Datensicherheit*, 41(12), 750–756.
- Diercks, N. (2019a). Datenschutz im Beschäftigtenverhältnis – Teil 1. *Zeitschrift für das Forderungsmanagement*, 3, 97–104.
- Diercks, N. (2019b). Datenschutz im Beschäftigtenverhältnis – Teil 2. *Zeitschrift für das Forderungsmanagement*, 5, 186–194.
- Diesner, J. (2015). Small decisions with big impact on data analytics. *Big Data & Society*, 2(2), 1–6.
- Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4), 211–407.
- Ethikbeirat HR Tech. (2019). Richtlinien für den verantwortungsvollen Einsatz von Künstlicher Intelligenz und weiteren digitalen Technologien in der Personalarbeit. https://www.ethikbeirat-hrtech.de/wp-content/uploads/2019/06/Ethikbeirat_und_Richtlinien_Konsultationsfassung_final.pdf. Zugegriffen am 01.07.2019.

- Gärtner, C. (2007). *Innovationsmanagement als soziale Praxis: grundlagentheoretische Vorarbeiten zu einer Organisationstheorie des Neuen*. München/Mering: Rainer Hampp.
- Gärtner, C. (2018). Human Resource Management bei und von XING. In C. Gärtner & C. Heinrich (Hrsg.), *Fallstudien zur Digitalen Transformation: Case Studies für die Lehre und praktische Anwendung* (S. 129–147). Berlin/Wiesbaden: Springer.
- Gärtner, C., & Claßen, M. (2016). Die gläsernen Mitarbeiter. *Personalmagazin*, 8, 38–39.
- Gärtner, C., & Huber, C. (2015). Drawing the line: Outlining the concepts of mindful organizing and sketching out the role of visual templates. In J. Reb & P. Atkins (Hrsg.), *Mindfulness in organizations* (S. 305–331). Cambridge: Cambridge University Press.
- Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys*, 51(5), 93:1–93:42.
- Hagendorff, T. (2019). Maschinelles Lernen und Diskriminierung: Probleme und Lösungsansätze. *Österreichische Zeitschrift für Soziologie*, 44(1), 53–66.
- Huff, J., & Götz, T. (2019). Evidenz statt Bauchgefühl? Möglichkeiten und rechtliche Grenzen von Big Data im HR-Bereich. *Neue Zeitschrift für Arbeitsrecht (NZA)* 36(24) (Beilage 2/2019), 73–78.
- Isele, C., Kaufmann, P., Mempel, L., Schütze, B., & Spyra, G. (2018). Sicherheit personenbezogener Daten: Umgang mit Art. 32 DS-GVO. https://www.gesundheitsdatenschutz.org/download/Art.32_SicherheitVerarbeitung.pdf. Zugegriffen am 08.01.2020.
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399.
- Kramer, S. (Hrsg.). (2019). *IT-Arbeitsrecht. Digitalisierte Unternehmen: Herausforderungen und Lösungen* (2. Aufl.). München: C.H. Beck.
- Kraus, C. (2018). Digitalisierung der Arbeitswelt – das Ende der Low Performer? *Der Betrieb*, 12, 701–705.
- Krotova, A., & Fritsch, M. (2019). *Berücksichtigung von Daten in digitalen Reifegradmodellen*. Köln: Institut der deutschen Wirtschaft.
- Lambrecht, A., & Tucker, C. (2019). Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads. *Management Science*, 65(7), 2966–2981.
- Libet, B. (2004). Haben wir einen freien Willen? In C. Geyer (Hrsg.), *Hirnforschung und Willensfreiheit. Zur Deutung der neuesten Experimente* (S. 268–289). Frankfurt a. M.: Suhrkamp.
- Martini, M. (2019). *Blackbox Algorithmus – Grundfragen einer Regulierung Künstlicher Intelligenz*. Berlin/Heidelberg: Springer.
- Meyer, R. E., Höllerer, M. A., Jancsary, D., & Van Leeuwen, T. (2013). The visual dimension in organizing, organization, and organization research: Core ideas, current developments, and promising avenues. *Academy of Management Annals*, 7(1), 489–555.
- Neuberger, O. (2006). *Mikropolitik und Moral in Organisationen*. Stuttgart: Lucius & Lucius.
- Nissenbaum, H. (2011). A contextual approach to privacy online. *Daedalus*, 140(4), 32–48.
- Ortmann, G. (2003). *Regel und Ausnahme. Paradoxien sozialer Ordnung*. Frankfurt a. M.: Suhrkamp.
- Orwat, C. (2019). *Diskriminierungsrisiken durch Verwendung von Algorithmen*. Berlin: Antidiskriminierungsstelle des Bundes.

- Raghavan, M., Barcas, S., Kleinberg, J., & Levy, K. (2019). Mitigating bias in algorithmic employment screening: Evaluating claims and practices. *arXiv preprint arXiv:1906.09208*.
- Rocher, L., Hendrickx, J. M., & De Montjoye, Y. A. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nature Communications*, 10(1), 1–9.
- Russell, S., & Norvig, P. (2013). *Künstliche Intelligenz – Ein moderner Ansatz* (3., ak. Aufl.). München: Pearson.
- Silberzahn, R., & Uhlmann, E. L. (2015). Crowdsourced research: Many hands make light work. *Nature*, 526(7572), 189–191.
- Sinar, E. F. (2018). Data visualization: Get visual to drive HR's impact and influence. Society for Human Resource Management (SHRM) & Society for Industrial Organizational Psychology (SIOP). https://www.shrm.org/hr-today/trends-and-forecasting/special-reports-and-expert-views/Documents/2018%2003_SHRM-SIOP%20White%20Paper_Data%20Visualization.pdf. Zugegriffen am 18.08.2019.
- Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., & Fergus, R. (2013). Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*.
- Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial intelligence in human resources management: Challenges and a path forward. *California Management Review*, 61(4), 15–42.
- Zook, M., Barcas, S., Boyd, D., Crawford, K., et al. (2017). Ten simple rules for responsible big data research. *PLoS Computational Biology*, 13(3), e1005399.
- Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30(1), 75–89.

Smart HRM: Wohin geht die Reise?

7

Zusammenfassung

Als Fazit ergibt sich aus den detaillierten Analysen der anderen Kapitel, dass Globalurteile über die Sinnhaftigkeit von digitalen Tools in der Personalarbeit sinnlos sind. An einer genaueren Beschäftigung sollten deshalb alle HR-Professionals ein Interesse haben, um weder die Möglichkeiten noch die Gefahren der Ergänzung und Ersetzung von menschlicher Arbeitskraft durch Maschinen falsch einzuschätzen.

Globalurteile über die Sinnhaftigkeit von digitalen Tools in der Personalarbeit sind sinnlos. Statt allgemeiner Aussagen beginnt die Smart-HRM-Reise mit Detailanalysen, die konkrete Antworten auf vier Fragen liefern müssen:

1. Welcher Automatisierungsgrad ist effektiv und effizient, das heißt, welche Aufgaben aus dem vielfältigen Portfolio der Personalarbeit können Maschinen besser und/oder billiger als Menschen erledigen (sowohl für die Kunden von HR als auch innerhalb der Personalabteilung)?
2. Wie sind die Chancen und Risiken des Einsatzes digitaler Tools in der Personalarbeit zu bewerten?
3. Welche rechtlichen und ethischen Grenzen gilt es zu beachten?
4. Wie groß ist der Aufwand für die Datensammlung und -analyse und welcher Mehrwert ist zu erwarten?

Entlang dieser Fragen gilt es zu bedenken, dass viele menschliche Entscheidungen durch Vorurteile, Stereotypen oder sonstige Verzerrungen geprägt sind (Kahneman 2012, S. 109 ff.). Demgegenüber sind algorithmenbasierte Entscheidungen neutraler, rationaler und zuverlässiger, was zu effizienteren und effektiveren Ergebnissen führt – sofern sie nicht auf verzerrten Daten beruhen (weil die Daten nicht repräsentativ sind bzw. selektiv ausgewählt wurden, sich auf seltene Ereignisse beziehen oder gefälscht wurden), unpassende Modelle verwenden (weil die Anwendungsbedingungen und Grenzen der Algorithmen nicht beachtet wurden) oder rechtliche bzw. ethische Richtlinien verletzen. Die Vorkehrungen, die zur Minimierung dieser Gefahren getroffen werden können, sind bekannt und erprobt. Sie sind aber nicht perfekt und müssen permanent an den technischen und wissensbezogenen Fortschritt angepasst werden. Die Smart-HRM-Reise ist also nie zu Ende.

Damit sich HR-Professionals überhaupt auf die Reise hin zu smarter Personalarbeit machen können, **muss noch einiges an Wissen und technischen Mitteln aufgebaut** werden (Angrave et al. 2016). Bei vielen Firmen steckt das Thema People Analytics noch immer in den Kinderschuhen (Marler und Boudreau 2017; Van den Heuvel und Bondarouk 2017). Kenntnisse in Statistik und Data Mining sind zwar notwendig, aber nicht hinreichend, um den datengetriebenen strategischen Mehrwert zu liefern. Hinzukommen muss ein profundes Verständnis betriebswirtschaftlicher und sozialpsychologischer Zusammenhänge – allgemein und speziell auf das Unternehmen bezogen, weil Befunde über Ursache-Wirkungs-Muster im Managementbereich mehrdeutig sind und variieren können. Auch das Wissen darüber, was in einer Organisation als akzeptabel gilt (oder zumindest ein Gespür dafür), gehört dazu. Nicht alles, was technisch möglich und ökonomisch erwünscht ist, sollte und darf auch umgesetzt werden. Hier können Personaler ihr Wissen und ihre Erfahrung in Stellung bringen (Gärtner und Claßen 2016). Dazu gehören auch Erfahrungen im Umgang mit dem Betriebsrat und datenschutzrechtlichen Fragen. Da diese allerdings alles andere als einfach zu lösen sind, beginnen Firmen ihre Reise oft mit einem kleinen Pilotprojekt und präferiert im Ausland, um die Komplexität überschaubar zu halten – und um dem Betriebsrat zumindest am Anfang der Reise aus dem Weg zu gehen (z. B. in den USA, UK, Schweiz, Indien oder China).

Aktuell scheinen die Verantwortlichen dem möglichen Nutzen digitaler Tools für die Personalabteilung entweder sehr skeptisch gegenüberzustehen oder ihn euphorisch herbeizusehnen (BPM 2019; DGFP 2019). Ähnliches gilt für die Frage, inwieweit KI/ML und andere Automatisierungstechnologien wie RPA zu einer **Ersetzung von Arbeit(skräften)** führen. Die Meinungswelle schwapppt zwischen „Maschinen können Menschen nicht ersetzen“ und „KI und Automatisierung ersetzen Menschen und vernichten Arbeitsplätze“ hin und her. Um hierauf eine fundierte Antwort zu geben, müssen die jeweils spezifischen Aufgaben entlang einer Reihe von Prüfkriterien analysiert werden:

- Ausmaß an Feinmotorik und Fühlen
- Anteil an Kreativität bzw. Erfindungsreichtum
- Bedarf an Sinnstiftung während bzw. für die Aufgabe
- Häufigkeit, etwas mit Macht auch gegen Widerstände Anderer durchzusetzen
- Grad an Mitgefühl bzw. Empathie, um soziale Beziehungen aufzubauen
- Anteil an situationsspezifischen Entscheidungen bzw. flexibler Interaktion

Je stärker diese Faktoren ausgeprägt sind, desto eher ist die menschliche Arbeitskraft im Vorteil (anders formuliert: desto schwieriger lassen sich die Aufgaben durch Maschinen ersetzen). Es ist z. B. nicht davon auszugehen, dass Menschen bei der Aufgabe, Teamarbeit zu gestalten, bald vollständig durch Maschinen ersetzt werden, weil dies eine sozialpsychologisch komplexe und kontextsensitive Aufgabe ist – und davon gibt es bei mitarbeiterbezogenen Fragestellungen in Unternehmen sehr viele. Wovon auszugehen ist, dass Maschinen Menschen unterstützen und innovative Abläufe bzw. Arbeitsergebnisse ermöglichen, während gleichzeitig einzelne Prozessschritte und standardisierte Abläufe automatisiert werden.

Statt Entscheidungen nur datengetrieben vorzunehmen und zu rechtfertigen, werden individuelle Kunstfertigkeit und Team-Spirit noch lange Zeit relevant bleiben – insbesondere bei komplexen Aufgaben. Diese zeichnen sich dadurch aus, dass mehrere, miteinander verbundene Aspekte zu beachten sind, die sich im Zeitverlauf ändern bzw. situativ angepasst werden müssen und die widersprüchlich sein können. Die vier Kriterien Kreativität, Sinnstiftung, Macht und Mitgefühl können als Umgangsformen mit Widersprüchen oder Dilemmata gelesen werden, von denen es in Organisationen nach wie vor sehr viele gibt (siehe Gärtner 2019). Die **Lösung komplexer, konfliktbeladener sozialer Situationen ist ein Merkmal von Intelligenz, das Maschinen nicht verwirklichen können**, solange sie keinen (vergänglichen) Körper, (Überlebens-)Willen und gesunden Menschenverstand sowie die Fähigkeit zur kommunikativen, kreativen und empathischen Auseinandersetzung mit Menschen haben. Dabei sollen und dürfen die menschlichen (und organisationalen bzw. institutionellen) Fähigkeiten zur Lösung solcher Situationen nicht überschätzt werden: Wir machen dabei viele Fehler und kommen bei manchen Problemen nur sehr langsam voran. Aber die Fehler entlang des Weges lösen häufig Lernprozesse aus, an deren Ende eine gute Antwort auf komplexe Probleme steht – wobei mit „gut“ die Integration ökonomischer, technischer, rechtlicher sowie ethisch-moralischer Aspekte gemeint ist, auch wenn diese konfliktär zueinanderstehen. Hingegen können ML-Algorithmen immer nur auf das zurückgreifen, was bisher schon praktiziert wurde, weil es nur darüber Daten gibt – selbst wenn andere Praktiken besser wären, könnte sie der Algorithmus nicht entdecken oder vorschlagen, weil es zu ihnen eben keine Daten gibt (Strohmeier 2019, S. 9).

Jene HR-Professionals, die die Reise bereits angetreten haben (siehe Petry und Jäger 2018; Gärtner 2020), wenden digitale Tools vor allem im **Recruiting-Bereich**

reich oder zur **Auswertung strukturierter Daten** an. Vier weitere vielversprechende Anwendungsgebiete liegen in der Auswertung von Textdaten, sozialen Beziehungen (Organisationale Netzwerkanalyse), der Nutzung von Chatbots und dem Einsatz von RPA. Aber selbst wenn Personalverantwortliche nur einen Bruchteil der in Kap. 4 beschriebenen Tools einsetzen, so stellt sich automatisch die Frage, wie die unterschiedlichen Daten, Systeme und Analysen integriert und ohne großen händischen Aufwand angewendet werden können. Es fehlt noch eine „plug & play“-Lösung, bei der die Applikationen problemlos miteinander interagieren können.

Die Reise ist also keineswegs einfach. So lange aber die Antworten auf die eingangs (Kap. 1) gestellten Fragen nicht ganz klar sind – und das ist der Fall –, können uns digitale Tools und Verfahren aus dem Bereich der Künstlichen Intelligenz lehren, **die richtigen Fragen über smartes Handeln zu stellen**. Spätestens, wenn wir die Tools anwenden, sind wir gezwungen, gründlich darüber nachzudenken, was wirklich wichtig und richtig ist. Ich hoffe, diesen Erkenntnisprozess angestoßen und angeleitet zu haben.

Literatur

- Angrave, D., Charlwood, A., Kirkpatrick, I., Lawrence, M., & Stuart, M. (2016). HR and analytics: why HR is set to fail the big data challenge. *Human Resource Management Journal*, 26(1), 1–11.
- BPM. (2019). Zwischen Euphorie und Skepsis: KI in der Personalarbeit. https://www.bpm.de/sites/default/files/bpm-service-24-ki_in_der_personalarbeit.pdf. Zugegriffen am 09.03.2019.
- DGFP. (2019). Künstliche Intelligenz in HR. https://www.dgfp.de/fileadmin/user_upload/DGFP_e.V/Medien/Publikationen/2019/Befragung-KI-in-HR_September-2019.pdf. Zugriffen am 20.10.2019.
- Gärtner, C. (Hrsg.). (2020). *Smart human resource management*. Wiesbaden: Springer Gabler.
- Gärtner, C. (2019). Ambivalenzen und Dilemmata in digital transformierten Arbeitswelten. *supervision: Mensch – Arbeit – Organisation*, 4, 10–17.
- Gärtner, C., & Claßen, M. (2016). Die gläsernen Mitarbeiter. *Personalmagazin*, 8, 38–39.
- Kahneman, D. (2012). *Thinking, fast and slow*. London: Penguin Books.
- Marler, J. H., & Boudreau, J. W. (2017). An evidence-based review of HR analytics. *The International Journal of Human Resource Management*, 28(1), 3–26.
- Petry, T. & Jäger, W. (Hrsg.). (2018). *Digital HR: Smarte und agile Systeme, Prozesse und Strukturen im Personalmanagement*. Freiburg: Haufe.
- Strohmeier, S. (2019). Einsatzpotenziale und -grenzen von künstlicher Intelligenz in der Personalarbeit (Interview mit T. Biemann). *Personal Quarterly*, 3, 6–9.
- Van den Heuvel, S., & Bondarouk, T. (2017). The rise (and fall?) of HR analytics: A study into the future application, value, structure, and system support. *Journal of Organizational Effectiveness: People and Performance*, 4(2), 157–178.