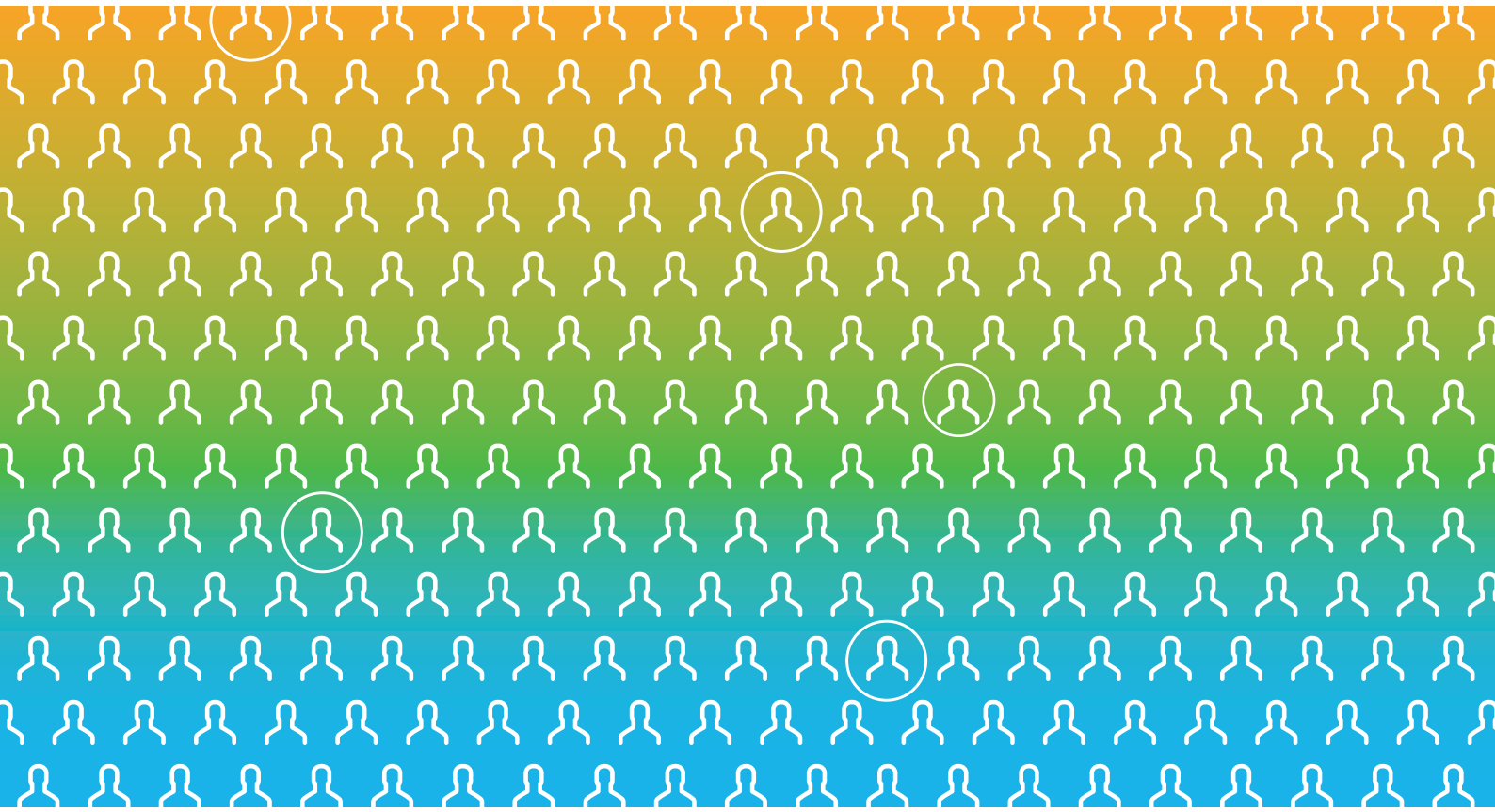
Automatisierte Entscheidungen und Künstliche Intelligenz im Personalmanagement

Ein Leitfaden zur Überprüfung essenzieller Eigenschaften KI-basierter Systeme für Betriebsräte und andere Personalvertretungen

Prof. Dr. Sebastian Stiller (TU Braunschweig)
Jule Jäger (TU Braunschweig)
Sebastian Gießler (AlgorithmWatch)
2. März 2020



Eine Publikation von

ALGORITHM WATCH

gefördert von der

Hans Böckler Stiftung

Inhalt

Die Automatisierung des Personalmanagements (Human Resources, Kurzform: HR) wird in Unternehmen weiter voranschreiten. Das bietet einerseits neue Chancen, andererseits birgt es auch Konfliktpotenzial im Betrieb. Denn diese Automatisierung wird die Organisation von Arbeit verändern und damit tiefgreifend die Unternehmensstruktur beeinflussen. Das macht die Einführung und Weiterentwicklung dieser Systeme zu einem wichtigen Thema für Betriebs- und Personalräte. Dieser Leitfaden soll Mitglieder dieser Interessenvertretungen in die Lage versetzen, den Prozess kompetent zu gestalten.

Aufbau des Leitfadens

Nach einer kurzen Einführung ins Thema wird eine Reihe von Fragen zu automatisierten Personalmanagementsystemen vorgestellt: Fragen, die Betriebsräte der Unternehmensleitung möglichst vor der Einführung neuer Systeme der Unternehmensleitung stellen sollten, die aber auch zu bereits eingeführten Systeme gestellt werden können. Zusammen mit den Fragen liefert dieser Leitfaden Erläuterungen, wie Antworten einzuordnen sind und was sie beinhalten sollten.

Einführung

Jeder kann sich in etwa vorstellen, was mit „Künstlicher Intelligenz“ (KI) gemeint ist: Computersysteme, die ungefähr so denken wie Menschen. Ein genaueres Verständnis ist nach allgemeiner Ansicht die Sache von Fachkundigen. Aus Expertensicht stellt sich KI deutlich anders dar: Systeme, die menschlichem Denken vergleichbar sind, sogenannte starke KI, gibt es nicht. Für „schwache“ KI gibt es dagegen keine verbindliche Definition. Der Begriff „KI“ wird so willkürlich und für so unterschiedliche Verfahren benutzt, dass man immer nachfragen sollte, wie das Verfahren funktioniert. Das gilt auch für andere Begriffe des automatisierten Personalmanagements wie „Talent Analytics“, „Workforce Analytics“, „People Analytics“ oder „Human Resources Analytics“. Die Antwort, das sei so ähnlich wie beim menschlichen Denken, ist eher irreführend als aufschlussreich.

Wie ein Verfahren funktioniert, ist nur zu einem kleinen, oft unwesentlichen Teil Sache von Expertinnen und Experten. Es geht immer um Verfahren, die auf der Grundlage von Daten zu Aussagen kommen. Weshalb und in welchem Sinne diese Aussagen wahr oder brauchbar sind, hängt von der Begründung für das Verfahren ab. Wesentliche Teile dieser Begründungen sind nicht mathematisch und können von Anwender·innen und Betroffenen mindestens so gut diskutiert werden wie von Entwicklerinnen und Entwicklern.

Hinzu kommt, dass viele Verfahren nicht einfach wahre oder falsche Aussagen treffen, sondern eher Hinweise und Empfehlungen geben. Um solche Hinweise verantwortlich zu verwenden, muss man wissen, wie sie entstanden sind. Ein solches Verständnis ist möglich. Und letztlich gilt: Die Begründung einer Aussage muss so gut sein, dass sie diejenigen überzeugt, die die Konsequenzen und die Verantwortung dafür tragen.

Im Rahmen der betrieblichen Mitbestimmung muss daher über die Funktionsweise von KI-Verfahren diskutiert werden können, die im Personalwesen eingesetzt werden. Dieser Leitfaden enthält Fragen, die dies ermöglichen, indem sie dabei helfen, zu den diskussionswürdigen Begründungszusammenhängen eines Softwaresystems vorzudringen.

Was unterscheidet moderne Personalmanagementsysteme mit dem Zusatz „Analytics“ (z. B. Human Resources Analytics, Kurzform: HRA) von klassischer Unternehmenssoftware, den sogenannten Human Resources Information Systems (HRIS)? Bisherige Systeme können Auskunft geben über direkt vorliegende Daten (z. B. Fortbildungtage von Angestellten, Abschlussnoten von Personen im Bewerbungsverfahren) oder einfache, klar definierte Aussagen aus den Daten ableiten (z. B. Mittelwert der Anzahl der Fortbildungstage aller Angestellten einer Abteilung, Durchschnittsnote der Personen im Bewerbungsverfahren). Neue Systeme treffen zusätzlich Aussagen, deren Zusammenhang mit den zugrunde liegenden

Daten nicht unmittelbar ersichtlich ist. Diese Aussagen reichen von Bewertungen und Empfehlungen (z. B. passgenaue Vorschläge für Fortbildungen, Vorauswahl von Bewerbungen) bis hin zu automatischen oder automatisch stark vorgeprägten Entscheidungen.

HRA bietet also vor allem eine Erweiterung um Komponenten, die Aussagen ermöglichen, deren Begründungen nicht mathematisch beweisbar sind, sondern von Argumentationen, Theorien und Erkenntnisse aus Psychologie, Betriebswirtschaftslehre, Personalwesen und Verhaltensökonomie geprägt werden. Es muss sowohl darüber gesprochen werden, ob die zugrunde gelegten Theorien akzeptiert werden, als auch darüber, ob die Umsetzung der Theorien in berechenbare Kriterien plausibel erscheint.

Natürlich sind solche HRA-Systeme nicht unfehlbar. Das liegt schon allein daran, dass es auf viele Fragen im Personalmanagement (Beurteilung von Bewerbungsunterlagen, Entscheidungen über Organisationsstruktur, Vergütungen) überhaupt keine eindeutigen und zweifelsfreien Antworten gibt. Dennoch kann es von Vorteil sein, HRA-Systeme mit Sachverstand einzusetzen. HRA-Systeme können oft sehr viel größere Zusammenhänge berücksichtigen als menschliche Entscheiderinnen und Entscheider. Menschliche Intuition, Erfahrung und Urteilsvermögen unterliegen vielen Schwächen, von denen HRA-Systeme frei sind. Wenn die Kriterien von HRA-Systemen transparent gemacht werden, kann das zu mehr Verlässlichkeit und Fairness im Betrieb führen.

Um ihre Entscheidung aus Daten abzuleiten, verwenden sie HRA-Systeme andererseits oft unpassende oder verengende Kriterien und können schlecht mit Einzelfällen umgehen. Darüber hinaus sind Systeme, die sogenanntes Maschinelles Lernen (ML) verwenden – d.h. aus Beispielentscheidungen der Vergangenheit ihre Regeln für die Zukunft ablesen –, davon abhängig, wie gut die Entscheidungen der Vergangenheit waren und wie gut sie zur Zukunft passen. Eine Grundannahme von ML-Methoden ist, dass die Zukunft im Wesentlichen genauso aussieht wie die Vergangenheit, aus der gelernt wird. Für Naturwissenschaften wie die Physik ist diese Annahme sehr vernünftig. Wo es um menschliche Entscheidungen geht, ist diese Annahme oft problematisch.

Die Schwächen von HRA-Systemen sind besonders dann kritisch, wenn ein System intransparent bleibt und nicht beurteilt, hinterfragt, kontrolliert, außer Kraft gesetzt und geändert werden kann. Umgekehrt kommen die Stärken von HRA-Systemen besonders gut zur Geltung, wenn ein System transparent ist und sinnvoll in das Unternehmen integriert wird. Deshalb beinhaltet dieser Leitfaden auch Fragen zur Integration des HRA-Systems in den Betrieb.

Vor- und Nachteile eines HRA-Systems zu beurteilen und abzuwägen, kann von Herstellern und Fachkundigen nicht allgemeingültig für alle Betriebe geleistet werden. Was für manche Betriebe passt, stört in anderen erheblich. Stattdessen stellt dieser Leitfaden Fragen zusammen, die auch Expertinnen und Experten stellen würden, um zu diskutieren, ob und wie ein HRA-System innerhalb eines bestimmten Betriebs eingesetzt werden soll.

Ein gut verstandenes und mit Umsicht in die betrieblichen Prozesse integriertes System kann für alle im Betrieb von Vorteil sein. Der Leitfaden soll helfen, die richtigen Fragen zu stellen, um dieses Verständnis zu gewinnen.

Unterstützung, Selbstständigkeit und Vorhersagen

Für die Beurteilung von HRA-Systemen ist es hilfreich, zwischen Systemen zu unterscheiden, die Entscheidungen unterstützen, und solchen, die autonom (selbstständig) entscheiden:

- Entscheidungsunterstützend ist ein Softwaresystem, wenn es einer Person am Bildschirm zwar Vorschläge liefert – z. B. wer befördert werden soll –, die Entscheidung aber auch anders ausfallen kann, weil sie vollends in der Hand dieser Person bleibt.

- Wird eine Software eingesetzt, die in gewissem Rahmen ohne menschliche Einbindung eigen-

ständig Entscheidungen fällen kann – etwa bei der Erstellung eines verbindlichen Schichtplans –, handelt es sich um ein autonomes System.

HRA-Systeme können deskriptiv (beschreibend), prädiktiv (vorhersagend) oder präskriptiv (vorschreibend) sein. Während die ersten beiden Varianten nur vorhandene Daten darstellen oder aus ihnen Vorhersagen ableiten, geben präskriptive Systeme Handlungsempfehlungen. Es treten auch Mischformen dieser drei Arten von Systemen auf.

Leitfragen

Der Leitfaden gliedert sich in vier Blöcke von Fragen. Im ersten Block soll möglichst konkret geklärt werden, welche Aussagen die Software trifft und welche Aufgaben sie im Betrieb hat. Der zweite Block legt den Fokus darauf, wie diese Aussagen begründet sind. Die Antworten hierauf sollten ebenso klar und plausibel sein, wie man es auch bei der Einführung nichtelektronischer Regelungen und Verfahren – etwa für Vergütung oder Arbeitsorganisation – erwarten würde. Im dritten Block sind Fragen zur Qualität von Algorithmen und Programmierung zusammengefasst, die von Fachkundigen beurteilt werden müssen. In den Antworten dort geht es daher, anders als im zweiten Block, nicht vordringlich um Verständnis und Plausibilität, sondern um verbindliche Zusagen und Prüfergebnisse von Expertinnen und Experten. Der vierte Block von Fragen dient der Klärung, ob vor dem Hintergrund der ersten drei Blöcke das System richtig eingesetzt und in die Prozesse des Betriebs integriert ist.

Q1 Welche Aussagen trifft und welche Entscheidungen berührt die Software?

Q1.1 Um welche Software und welche Komponenten davon geht es?

Softwaresysteme bestehen häufig aus mehreren Komponenten. Es können Komponenten und Funktionalitäten im Laufe des Einsatzes über Updates oder in einem Software-as-a-Service-Modell (SaaS) hinzukommen. Als Startpunkt für einen Dialog mit der Unternehmensleitung gilt es deshalb, eine Übersicht über das Softwaresystem zu gewinnen und zu klären, wie Komponenten behandelt werden, die in der Zukunft hinzukommen.

Wichtig ist darüber hinaus zu klären, wo die Daten des Unternehmens, vor allem die Beschäftigtendaten, gespeichert werden und wer zu welchem Zweck darauf Zugriff hat.

Q1.2 Welche Aussage trifft die Software und mit welchem Wahrheitsgehalt?

Die Aufgabe von Softwaresystemen ist es, Aussagen zu treffen, die von Nutzer·innen weiterverwendet werden oder sogar direkte Folgen nach sich ziehen. Dabei bietet ein Softwaresystem meist viele Funktionalitäten, die unterschiedliche Arten von Aussage treffen.

Aussagen sollten so konkret wie möglich benannt werden. Beispielsweise: Das Softwarepaket liefert Ergebnisse wie „Kandidat oder Kandidatin X hat Persönlichkeitsstruktur Y". Zu unkonkret ist dagegen folgende Antwort: „Dieses Softwareprodukt findet die idealen Mitarbeiterinnen und Mitarbeiter für das Unternehmen". Je konkreter die Aussage einer Software benannt wird, desto besser lässt sich später überprüfen, ob die Software hält, was versprochen wurde.

Nicht alle Aussagen sind einfach wahr oder falsch. Statistische Aussagen treffen z. B. mit einer bestimmten Wahrscheinlichkeit zu. Andere Aussagen sind eher Vermutungen oder Anregungen. In diesem Sinne hat jede Aussage einen Wahrheitswert. Der Wahrheitswert hängt davon ab, wie gut die Aussage begründet ist. Umgekehrt beeinflusst der Wahrheitswert, wie die Aussage verwendet werden kann. Es ist wichtig, den Wahrheitswert der Aussagen der Software deutlich zu machen. Ist er hoch, müssen die Verfahren sehr gut begründet sein. Ist er gering, kann man die Ergebnisse nur beschränkt einsetzen.

Q1.3 In welchem Bereich bereitet das Softwaresystem Entscheidungen vor oder entscheidet autonom (selbstständig)?

An der Antwort auf diese Frage sollte sich im Kern ablesen lassen, für welchen konkreten Zweck das Unternehmen das Softwaresystem anschaffen will. Sollen bestehende Verfahren (teil-)automatisiert oder gänzlich neue eingeführt werden? Welcher Klasse ist die Software zuzuordnen: Deskriptiv, prädiktiv, präskriptiv (s. Einführung)?

Q2 Wie kommt die Software zu ihren Aussagen?

Q2.1 Auf welche Daten hat die Software Zugriff?

Es ist aus mehreren Gründen wichtig zu klären, auf welche Daten (insbesondere Beschäftigtendaten) die Software Zugriff hat. Es betrifft Datenschutz- und Arbeitsrecht, welche Daten benutzt werden und welche Entscheidungen aufgrund welcher Daten gefällt werden.

Darüber hinaus sollte bei dieser Frage beurteilt werden, ob es schlüssig erscheint, die Aussagen der Software aus den verwendeten Daten abzuleiten. Auch Rückkoppelungseffekte sollten hier diskutiert werden: Wie beeinflusst es beispielsweise die Arbeitsweise der Angestellten, wenn sie wissen, dass ihr E-Mail-Verhalten ausgewertet wird? Sind die Aussagen des Systems dann noch sinnvoll oder werden sie durch mögliche Reaktionen der Beschäftigten entwertet?

Q2.2 Nach welchen Kriterien entscheidet die Software?

Ein Softwareprogramm erscheint auf den ersten Blick zu kompliziert, um von Laien verstanden zu werden. Aber bevor man anfängt zu programmieren, überlegt man sich Kriterien, mit denen man die gewünschte Aussage erzeugen kann. Solche Überlegungen enthalten Begründungen für die verwendeten Kriterien, die für alle Menschen verständlich sind. Die Diskussion dieser Kriterien erfordert keine Programmierkenntnisse. Sie kann von Anwendern und Betroffenen mindestens so gut geführt werden wie von denjenigen, die die Software erstellen.

Manche Softwaresysteme verwenden Kriterien, die direkt die Aussage der Software rechtfertigen. Das Kriterium eines Navigationssystems z. B. ist die Länge einer Route vom Start zum Ziel. Es sucht unter allen möglichen Routen diejenige aus, die entsprechend diesem Kriterium die kürzeste ist. Die Aussage des Navigationssystems: „Diese Route ist der kürzeste Weg“, ist also gut begründet. In moderner Personalsoftware werden häufig Aussagen getroffen, für die kein direktes Kriterium angeben werden kann. Ein Beispiel für eine solche Aussage wäre: „Mitarbeiterin XY ist für das Projekt besonders wichtig.“ Da „Wichtigkeit“ kein Wert ist, der eindeutig bestimmt werden kann, verwenden die Anbieter ein Ersatzkriterium. Beispielsweise berechnen sie, wie zentral Mitarbeiterin XY im Netzwerk der E-Mail-Kommunikation ist. Dieses Kriterium wird dann verwendet, um die Wichtigkeit von Mitarbeiterin XY zu bewerten. Damit die Software verantwortlich eingesetzt werden kann, müssen solche Ersatzkriterien diskutiert werden.

Für manche Aussagen fällt es uns Menschen schwer, exakte Kriterien anzugeben. Beispiele sind unser Musikgeschmack oder unsere Fähigkeit, befreundete Personen und Verwandte zu erkennen, ohne dass wir genau beschreiben können, woran wir das festmachen. Um mit Software solche Aussagen zu generieren, werden Methoden des Maschinellen Lernens verwendet. Ein ML-Verfahren erstellt automatisch anhand von Beispielen ein Kriterium, das häufig zu kompliziert ist, um von Menschen nachvollzogen zu werden. Werden solche Verfahren eingesetzt, sollten gesonderte Fragen gestellt werden, die hier folgen. Vorangestellt ist ein kurzer Einschub zum besseren Verständnis von ML.

Die Kriterien einer Software müssen genannt werden. In der Regel ist es möglich, die Kriterien hinreichend gut zu erklären, ohne so ins Detail gehen zu müssen, dass es wettbewerbsschädigend ist. Werden die Kriterien nicht genannt, kann man nicht nachvollziehen,

ob Aussagen begründet sind, und die Software letztlich nicht verantwortlich einsetzen.

Ein kurzer Einschub: Das Wichtigste über Maschinelles Lernen

Wir können viele Fragen beantworten, ohne ein klares Kriterium dafür angeben zu können. Zum Beispiel: Woran erkennt man Personen, die man gut kennt, auf Fotos wieder? Das Kriterium dafür muss sich irgendwie aus dem zusammensetzen, was man auf den Fotos sieht. Aber woraus sich das Kriterium genau zusammensetzt, können Menschen nicht auflisten. ML findet ein solches kompliziertes Kriterium, wenn man ihm genug Beispielbilder zum „Lernen" gibt und ungefähr sagt, worauf es darin achten soll.

Maschinelles Lernen besteht also immer aus zwei Phasen: In der ersten Phase wird ein Kriterium anhand von vielen Beispieldaten (man sagt „Trainingsdaten") gelernt. In der zweiten Phase wird das Kriterium eingesetzt, um neue Daten zu beurteilen. Die unabdingbare Grundlage für alle diese Verfahren ist, dass die Trainingsdaten und die Anwendungsdaten (zwar nicht dieselben aber) in etwa gleich sind. Sind in den Trainingsdaten nur Fotos enthalten, die eine Person von vorne zeigen, dann wird das Kriterium diese Person nicht auf Fotos erkennen, auf denen sie von hinten zu sehen ist.

Wie funktioniert Maschinelles Lernen? ML-Verfahren sind nicht direkt vergleichbar mit menschlichem Lernen. Häufig hört man, das ML-Verfahren nur von den Daten getrieben werden. Auch das ist nicht richtig. Jedes ML-Verfahren startet mit einer (sehr großen) Menge möglicher Kriterien, den Hypothesen. Man gibt also nicht das Kriterium selbst vor, aber die Art des Kriteriums. Ohne eine solche Vorgabe funktioniert es nicht. Mit Hilfe dieser Vorgabe steckt man Vorwissen – oder besser: eine Vormeinung – in das Verfahren.

Der sogenannte Trainingsalgorithmus sucht aus dieser Art von Kriterien ein Kriterium aus, das besonders gut zu den Trainingsdaten passt.

Das Wort „Lernen" erweckt den Eindruck, als führe ein ML-Verfahren früher oder später immer zur Wahrheit. Sehr viele Versuche, ein Problem mit ML-Verfahren zu lösen, scheitern jedoch. Darüber wird natürlich öffentlich nicht gerne gesprochen. Ob ein ML-Verfahren funktioniert oder nicht, lässt sich fast nur in der praktischen Erprobung beurteilen.

Die Qualität eines ML-Verfahrens wird mit einem weiteren Satz von Daten gemessen, den sogenannten Testdaten. Das ML-Verfahren kann ein Kriterium liefern, das bezogen auf seine Trainingsdaten immer richtige Aussagen trifft, angewendet auf die Testdaten aber versagt. Es ist wichtig, dass die Testdaten nicht schon im Training benutzt wurden. Ebenso wichtig ist, dass die Testdaten zu den Daten passen, auf die das gelernte Kriterium angewendet werden soll. Wenn man mithilfe eines ML-Verfahrens gut den Musikgeschmack von leitenden Angestellten trifft, kann das gleiche Verfahren bei Auszubildenden versagen.

Es gibt heute frei verfügbare Software, mit der man ML-Verfahren recht einfach selbst erstellen kann, ohne viel davon zu verstehen, wie und warum diese Verfahren funktionieren. Man baut ein Verfahren zusammen, trainiert es mit einem Trainingsdatensatz und probiert aus, ob es angewendet auf Testdaten gute Ergebnisse erbringt. Ein ML-Verfahren programmieren zu können, erfordert heute kein tiefes Verständnis davon, welche Schwierigkeiten ihre Anwendung haben kann.

Vor diesem Hintergrund sollten bei der Verwendung von ML-Verfahren folgende Fragen erörtert werden:

Q2.3 Welche Annahmen und wissenschaftlichen Theorien liegen dem verwendeten ML-Verfahren zugrunde und warum wurde dieses Verfahren gewählt?

Q2.4 Welche Trainingsdaten wurden für des ML-Verfahren verwendet?

Es muss sorgfältig erwogen werden, ob die Trainingsdaten und die Anwendungsfälle im Betrieb so gut

zusammenpassen, dass die Grundvoraussetzung für den Einsatz von ML-Verfahren erfüllt ist.

Die Trainingsdaten dienen dem ML-Verfahren als Beispiele, auf die es seine Entscheidungen gründet. Sind die Entscheidungen in den Beispielen tatsächlich mustergültig für den Betrieb? Wenn beispielsweise in der Vergangenheit vorwiegend Männer eingestellt wurden, wird ein ML-Verfahren, dass mit den Bewerbungsdaten erfolgreicher Personen trainiert wurde, möglicherweise Frauen diskriminieren.

Im Dialog mit der Unternehmensleitung sollte diese die Diskriminierungsfreiheit eines ML-Verfahrens nachweisen, statt umgekehrt lediglich anzunehmen, dass das System nicht diskriminiert, solange Diskriminierung nicht nachgewiesen wurde. Daher sollte folgende Frage gestellt werden:

Q2.5 Wie wurde das ML-Verfahren gegen Diskriminierung und andere ungewollte Einflüsse aus den Trainingsdaten gesichert?

Es reicht hier zum Beispiel nicht aus, das Geschlecht aus den Trainingsdaten für die Bewerbervorauswahl herauszuhalten. Das ML-Verfahren könnte mehr oder weniger eindeutig aus anderen Eigenschaften das Geschlecht ableiten und dadurch dennoch eine Geschlechterdiskriminierung als Muster aus den alten Trainingsdaten übernehmen.

Q2.6 Wie wurde das ML-Verfahren getestet?

In der Praxis wird die Qualität von ML-Verfahren nahezu immer durch Tests geprüft. Wie hoch sind die Wahrscheinlichkeiten für eine richtige und eine falsche Beurteilung, wenn das Verfahren auf Testdaten angewendet wird? Die Testdaten müssen vollkommen getrennt vom Training und der Entwicklung des ML-Verfahrens sein. Am besten sollten Tests unabhängig von den Anbietern der Systeme durchgeführt worden sein. Darüber hinaus müssen die Testdaten zu den Anwendungsfällen des Systems im Betrieb passen.

Q3 Wie ist die Qualität des Systems sichergestellt?

Q3.1 Setzt der Algorithmus die Kriterien exakt um?

Der Algorithmus im engeren Sinne ist ein (mathematisches) Verfahren, um das gewählte Kriterium möglichst schnell zu berechnen. Die Qualität von Systemen kann hier tatsächlich unterschiedlich sein. Es kann durchaus sein, dass beispielsweise der Algorithmus für ein Dienstplansystem nicht den Dienstplan erstellt, der nach den Vorgaben der beste wäre. Es ist schwierig, hier verlässliche Informationen zu erhalten. In der Regel hilft nur das Urteil von Sachverständigen oder ein Vergleich mit Konkurrenzprodukten.

Q3.2 Wie ist die Qualität der Implementierung (des Programmcodes) sichergestellt?

Hat man ein Kriterium und einen dazu passenden Algorithmus entworfen, muss dieser schließlich programmiert werden. Dabei entstehen Fehler, insbesondere da Softwaresysteme sehr große Mengen an Programmiercode umfassen. Die systematische Prüfung ist auch für Fachkundige äußerst schwierig. Im Allgemeinen gibt es drei Möglichkeiten:

1. Das Unternehmen, das die Software erstellt, kann seine Arbeitsprozesse nach Vorgaben organisieren, die dazu dienen, Fehler zu reduzieren und bereits während der Entwicklung zu finden.

2. Code kann nachträglich verifiziert werden. Dies ist jedoch nur sehr eingeschränkt möglich.

3. Der Hersteller behebt auftretende Fehler laufend. Auch hier ist ein direkter Einblick nicht möglich. Stattdessen zielt die Frage darauf, welche Maßnahmen zur Qualitätssicherung vom Hersteller unternommen wurden.

Q3.3 Wer hat die Software erstellt und welche Komponenten wurden von Dritten übernommen?

Softwarepakete sind oft umfangreich und bestehen aus einer Vielzahl von Bestandteilen. So ist es üblich, dass die Entwicklung von Komponenten an Dritte ausgelagert wird oder bestimmte Software-Dienste (z. B. Amazon Web Services (AWS), IBM Watson) von außen eingebunden werden. Die Antwort sollte Aufschluss darüber geben, ob die gesamte Software allein von der Anbieterfirma programmiert wurde und, wenn das nicht der Fall ist, welche Bestandteile welcher Drittanbieter eingebunden wurden.

Q4 Wie ist das System im Betrieb integriert?

Q4.1 Welche Fähigkeiten und Kenntnisse werden auf Seiten der Anwender·innen der Software benötigt?

Neben einer genauen Beschreibung der Qualifikation, die vorausgesetzt wird, könnte in der Antwort auch schon ein Weiterbildungs- und Trainingsbedarf beschrieben werden, der beim Arbeitgeber zu erwarten ist.

Q4.2 Wo liegt die Verantwortung für das Softwareprodukt im Unternehmen?

Die Antwort auf diese Frage sollte möglichst genau benennen, welche Abteilung/en und/oder Personen die Verantwortung für den Betrieb der Software tragen. Gibt es ein Verfahren, um Vorschläge, Beschwerden oder Bedenken zu melden? Bestehen Möglichkeiten, Verfahren zu ändern oder um Funktionen zu erweitern, die erst nach der Einführung gewünscht werden?

Q4.3 Wie und von wem wird entschieden, welche Funktionalität der Software genutzt wird?

Aus der Antwort sollte hervorgehen, wer im Unternehmen darüber entscheidet, welche Daten für den Einsatz der Software zur Verfügung gestellt und welche Auswertungen vorgenommen werden sollen. Dies kann die Verwendung von sensiblen personenbezogenen Daten (z. B. Inhalte von E-Mails oder die Anzahl von Krankheitstagen) betreffen. In der Antwort sollte auch enthalten sein, ob es einen klar festgelegten Entscheidungsprozess zu diesen Fragen gibt und wie dabei Mitbestimmungsrechte berücksichtigt werden.

Q4.4 Wer legt die Kennzahlen fest, anhand derer in der Software Ziele definiert werden?

People Analytics und verwandte Softwaresysteme benötigen einen Maßstab, um in ihren Analysen Bewertungen vorzunehmen, beispielsweise wann ein Ziel erreicht wurde oder wann etwas als „gut“, „passend“ oder „gelungen“ einzuordnen ist. Aus der Antwort auf die Frage sollte hervorgehen, ob das Unternehmen die entsprechenden Kennzahlen selbst festlegen und ändern kann oder ob alleine die Firma, die die Software herstellt, diese Kennzahlen ohne Einflussmöglichkeit von außen definiert.

Q4.5 Wie transparent ist der Entscheidungsweg?

Ist der Prozess, mit dem das System zu Entscheidungen kommt, transparent? Ist es möglich zu prüfen, ob die Entscheidung plausibel und nachvollziehbar ist? In der Antwort sollte beispielsweise beschrieben werden, ob und wie in der Benutzeroberfläche (User Interface) der Software intuitiv erfassbar und für den Menschen verständlich dargestellt wird, mit welcher Sicherheit die Software eine Aussage trifft.

Q4.6 Werden mögliche subtile Beeinflussungen durch die Gestaltung der Softwareoberfläche ausgeschlossen?

Art und Weise der Nutzerführung (UX, Usability) und die Darstellung von User-Interface-Elementen können bestimmtes Nutzungsverhalten fördern. Etwa wenn eine Schaltfläche zur Bestätigung groß und grün dargestellt wird, eine Schaltfläche zur Ablehnung dagegen klein und unscheinbar gestaltet ist. Stichworte sind hier „Nudging“ (auf Deutsch: „Stupsen“) und „Dark Patterns“ (auf Deutsch: „dunkle Muster“). Die Antwort sollte darlegen, dass das Softwareprodukt, vor allem dort, wo Entscheidungen bestätigt oder getroffen werden können, keine unterschwellige Beeinflussung durch solche Methoden ausübt.

Q4.7 Können automatische Entscheidungen korrigiert werden?

Die Antwort des Unternehmens sollte darlegen, ob bei Zweifeln an (teil-)automatisierten Entscheidungen eine Melde- und Eingriffsmöglichkeit besteht. Sieht das Softwareprodukt dies vor? Wie ist dieser Eingriff ausgestaltet? Ist er im Arbeitsalltag und in den Prozessabläufen realistisch anwendbar? Kann die Software manuell „überstimmt“ werden? Wer im Unternehmen hat das Recht dazu?

Q4.8 Wurde eine Risikoabschätzung vorgenommen?

Die Unternehmensleitung sollte bei der Einführung einer Software die Risiken in technischer Hinsicht, aber auch in Bezug auf datenschutzrechtliche Aspekte abschätzen lassen. Die Antwort auf die Frage sollte möglichst die Ergebnisse der Risikoabschätzung enthalten oder eine Begründung, warum keine erfolgt ist.

Automatisierte Entscheidungen und Künstliche Intelligenz im Personalmanagement

Ein Leitfaden zur Überprüfung essenzieller Eigenschaften KI-basierter Systeme für Betriebsräte und andere Personalvertretungen

Prof. Dr. Sebastian Stiller (TU Braunschweig)
Jule Jäger (TU Braunschweig)
Sebastian Gießler (AlgorithmWatch)
2. März 2020

Herausgeber:
AW AlgorithmWatch gGmbH
Linienstr. 13
10178 Berlin
Sitz der Gesellschaft: Bergstr. 22, 10115 Berlin

Kontakt: info@algorithmwatch.org

Korrektorat:
Karola Klatt

Layout:
Beate Autering
Tiger Stangl
www.beworx.de

Veröffentlicht im Rahmen des Forschungsprojekts
Automatisiertes Personalmanagement und Mitbestimmung

Webdossier: **algorithmwatch.org/auto-hr**

Gefördert durch die

Hans **Böckler** **Stiftung**