PLATTFORM INDUSTRIE 4.0

Mit Überraschungen umgehen

Künstliche Intelligenz und Akzeptanz

Ein Wegweiser

März 2021

Einleitung

Die zunehmende Digitalisierung vereint mit „Internet of Things“ und Industrie 4.0 die digitale und physikalische Welt. Auf diese Weise werden Kosten durch Effektivitäts- und Effizienzgewinne reduziert, die Innovationsfähigkeit von Produkten, Geschäftsmodellen und Dienstleistungen gesteigert und die Nachhaltigkeit im Umgang mit Ressourcen, Umwelt und der Gesellschaft gefördert.

Der Einsatz von Künstlicher Intelligenz (KI) im industriellen Umfeld ist ein weiterer Schritt und Treiber, Geschäftsmodelle neu zu denken. Industrielle technische Systeme können – so der Gedanke – durch KI funktional erweitert, sowie autonomer und resilienter gegen äußere Einflüsse gestaltet werden.

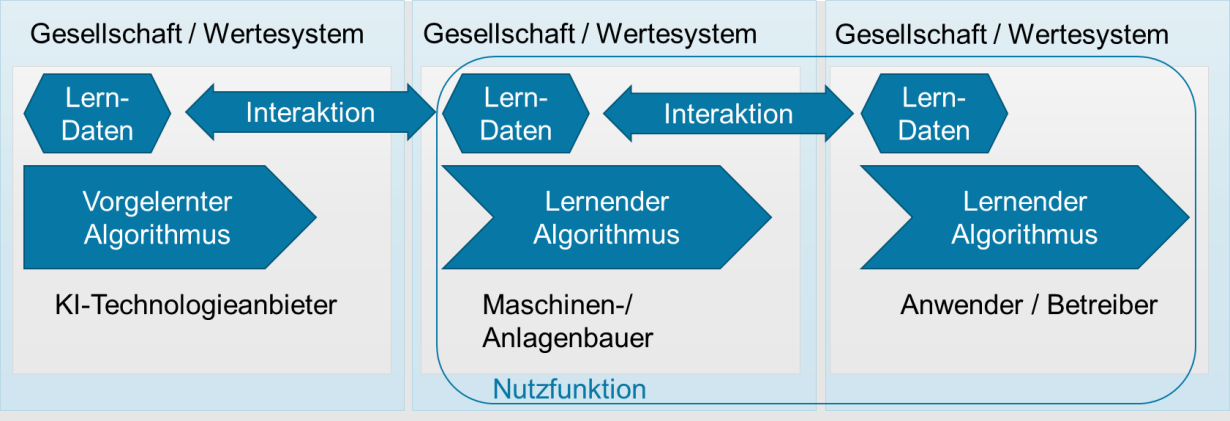
Im Umfeld autonomer industrieller Prozesse und Produktionsanlagen wird vielfach die Methode des maschinellen Lernens eingesetzt. Die mithilfe der KI autonom agierenden Systeme liefern ein neues Werteversprechen für den Anlagenbetreiber hinsichtlich der Effizienz, bedeuten aber auch veränderte Produktionssituationen über den kompletten Lebenszyklus der angeschafften Anlagen hinweg. Insbesondere wenn Anlagen global in unterschiedlichen Rechtsräumen und Wertesystemen eingesetzt werden, ergeben sich für den Maschinenbetreiber und seine Mitarbeiter spezifische Herausforderungen und Risiken. Neben dem ökonomischen Nutzen ist aus Anwendersicht auch zunehmend die Vertrauenswürdigkeit einer KI-Lösung von Bedeutung. Die Akzeptanz einer KI-Lösung ist dabei in großem Maße vom Vertrauen des Anwenders abhängig. Der Maschinen- und Anlagenbauer muss die Herausforderungen und den Bedarf an Vertrauen verstehen und beides in seiner Anlagenauslegung und den verbundenen Dienstleistungen adressieren.

Um diese Facetten von KI zu beleuchten, beschäftigt sich die Arbeitsgruppe „Technologie- und Anwendungsszenarien“ der Plattform Industrie 4.0 – nach den Veröffentlichungen „Künstliche Intelligenz in der Industrie 4.0“ und „KI in der Industrie 4.0: Orientierung, Anwendungsbeispiele, Handlungsempfehlungen“ – mit dem Aspekt der Akzeptanz. Es ist klar, dass die bisherigen Technologie-Akzeptanzmodelle aufgrund des nicht-deterministischen Verhaltens des KI-Einsatzes hier nicht anwendbar sind, da die Softskills der KI berücksichtigt werden müssen.

Ziel der Arbeitsgruppe ist die Bereitstellung einer Methode, um in diesem Kontext mithilfe einer einfach verständlichen Metrik die Akzeptanz einer KI-basierten Lösung in ihren unterschiedlichen Varianten und Einsatzszenarien durch den Anlagenbetreiber zu verstehen sowie den Maschinen- und Anlagenbauern einen entsprechenden Orientierungsrahmen durch Handlungsempfehlungen zu geben. Dies geschieht in Ergänzung zu Veröffentlichungen und Standards, wie z.B. die „Ethics Guidelines for a trustworthy AI“ oder „AI Roadmap - A human-centric approach to AI in aviation“ der EU.

Technische Parameter für KI im Spannungsfeld von Werteversprechen und Akzeptanz

Künstliche Intelligenzen verhalten sich wie „Black Boxes“, deren Funktionsweisen sich im Nachhinein vom Menschen nicht mehr entschlüsseln lassen. Dies wird insbesondere dann problematisch, falls überraschende, nicht erwartete Vorfälle oder Ergebnisse auftreten. Damit werden unter anderem Fragen aufgeworfen, wie Künstliche Intelligenzen kontrolliert werden können, wie nachverfolgbar deren Anwendungen sind und welche Systemgrenzen gesetzt werden sollen. Antworten auf diese Fragestellungen werden von den Erwartungshaltungen und Wertesystemen der Gesellschaften bestimmt.
Die nachfolgenden Abschnitte liefern erste Hinweise und eine Systematik, die es den Herstellern und Anbietern von KI-Lösungen bzw. technischen Systemen mit integrierter KI ermöglichen wird, die Auswirkungen von sich verändernden Maschinen einzuschätzen und die Akzeptanz durch Hinweise für das Angebot, den Betrieb und die betroffenen Mitarbeiter zu erhöhen.
Für den Lernprozess beim maschinellen Lernen haben die Entstehung und der Austausch von Daten sowie die zur Steuerung genutzten KI-Modelle einen direkten Einfluss auf die Fähigkeiten, Qualität, Verlässlichkeit und Betriebssicherheit in einem industriellen Prozessablauf. Damit verbunden ist das Werteversprechen des Herstellers für sein KI-basiertes Produkt bzw. dessen direkten Nutzen für einen Kunden, die bestehenden technischen Produktmerkmale um variable – durch KI veränderte oder neu entstehende Merkmale – zu erweitern. Die zunehmende globale Vernetzung ermöglicht es, Daten aus unterschiedlichen geografischen Bereichen mit zum Teil konträren Regularien, Wertvorstellungen oder auch kundenspezifischen Eigenschaften einzusetzen. Der Erfolg eines Maschinenherstellers hängt am Ende von der Akzeptanz des Nutzers ab, der neben der technischen Funktionsbeschreibung der Maschine oder Anlage auch die Beschreibung der durch KI gestalteten technischen Möglichkeiten erkennen und bewerten muss. Dabei spielen einerseits die Systemgrenzen beim Lernen und beim Einsatz der KI, sowie andererseits die Sichtbarkeit der eigenen Daten innerhalb und außerhalb dieser Systemgrenzen eine wesentliche Rolle. Akzeptanz wird darüber hinaus durch Transparenz bei Datenquellen, -strömen und Algorithmen geschaffen sowie durch vertrauensbildende Maßnahmen im Entwicklungs- und Nutzungsprozess der mit KI gestalteten Maschine oder Anlage bzw. auch der Erstellung und Nutzung der KI-Algorithmen etabliert.
Um dies aktiv zu gestalten, ist es notwendig, die Herkunft von KI-Technologie und Daten zu berücksichtigen. Dies gilt in der Phase des Vorlernens (der Konditionierung) des Systems beim KI-Technologieanwender, während des Lernprozesses im Rahmen der Entwicklung beim Maschinen- und Anlagenbauer, sowie beim weiteren Lernen in der Betriebsphase beim Anwender/Betreiber (Siehe nachfolgendes Bild).



Technische Parameter und Systemgrenzen

Diese, als technische Daten für Maschinen und Anlagen auszuweisenden Merkmale, sind insbesondere für die eigentliche Nutzfunktion zu betrachteten. Wird beispielweise über KI-Methoden eine Regelung (Nutzfunktion) optimiert, so ist zu berücksichtigen, dass die Lerneffekte durch Nutzdaten verschiedenster Maschinen bei unterschiedlichen Anwendern/Betreibern in z.B. unterschiedlichen Gesellschaften und Wertesystemen erstellt werden. Bereits bei der Konstruktion der Maschine/Anlage mit integrierter KI ist zu beachten, dass die Datenquellen des Anwenders/Betreibers aus unterschiedlichen Wertesystemen kommen können, die einheitlich, adaptiv oder auch konträre Rahmenbedingungen enthalten können. Für den Maschinen-/Anlagenbauer sind je nach Einsatzfall somit unterschiedliche Verwendungsarten notwendig und evtl. sogar rechtlich erforderlich, wenn z.B. personenbezogene Daten Verwendung finden.

Sowohl beim Lernen als auch beim Einsatz der KI spielen die Verwendbarkeit und Sichtbarkeit eigener Daten innerhalb und außerhalb von Systemgrenzen eine wesentliche Rolle. Dies ist die Voraussetzung für eine technisch und vertrieblich notwendige Akzeptanz, die durch die Transparenz bei Datenquellen, -strömen und Algorithmen geschaffen wird. Dadurch ergeben sich vertrauensbildende Aspekte im Entwicklungs- und Nutzungsprozess der mit KI gestalteten Maschine oder Anlage und auch bei der Erstellung und Nutzung der KI-Algorithmen.

„Überraschtheit“ als Dreh- und Angelpunkt

Akzeptanz beim Anwender und Nutzer setzt auf dem Vertrauen auf, dass sich aus der Übereinstimmung von Erwartung und Wirkung und in einem gewissen Maß durch die Nachvollziehbarkeit von technischen Systemen entwickelt. Die Erwartung beschreibt hierbei das durch den Nutzer erwartete Verhalten einer Maschine. Die Wirkung sagt etwas über das tatsächliche Verhalten aus. Dieses ist auf der einen Seite die wahrnehmbare funktionale Maschineninteraktion eines sich durch KI verändernden Maschinenverhaltens, wie auch die durch das

Wertesystem des Nutzers erwartbare Reaktion. Diese ist je nach Wertesystem unterschiedlich und basiert auf Erfahrungen im Umgang mit Maschinen und Anlagen, betrieblichen Vorgaben und rechtlichen Rahmenbedingungen.
Stimmen Erwartung und Wirkung nicht überein, ist das in den meisten Fällen überraschend. Dieser Aspekt der Überraschung sollte aus Sicht der Arbeitsgruppe als wesentlicher Parameter in die Betrachtung der Akzeptanz einbezogen werden. Dabei kann die wahrnehmbare Überraschung im Zusammenhang mit der Akzeptanz zu unterschiedlichen Reaktionen führen. Wir unterscheiden Überraschung in vier Spielarten und bezeichnen diese im Weiteren auch als „Überraschtheit“:

0. Ich bin nicht überrascht
1. Ich bin überrascht, aber akzeptiere (da ich vertraue)
2. Ich bin überrascht, und verlange eine Erklärung und/oder Änderung, die dazu führen sollte, dass ich akzeptiere
3. Ich bin überrascht und lehne ab

Um diese nicht-deterministische Sichtweise zu erfassen, spielen für die Wahrnehmung von Überraschungen vier ‚Filter‘ eine wichtige Rolle:

- Ein erster Filter wird durch Heuristiken der Informationsverarbeitung gebildet. Heuristik der Informationsverarbeitung soll hier die Fähigkeit des Nutzers widerspiegeln, der mit unvollständigen Informationen oder Wissen, wenig Zeit oder Antrieb zu einer annehmbaren Entscheidung kommen kann. Jeder Mensch baut durch die Wahrnehmung von und den Umgang mit bestimmen Dingen Erfahrungswissen auf, mit dessen Hilfe dann analoge Situationen wahrgenommen und bewältigt werden. Dies stellt einen ersten Filter dar, denn Überraschungen treten auf, wenn sich eine Situation der Erfahrung entzieht oder ihr widerspricht und mehr noch, wenn die Heuristiken der Informationsverarbeitung scheitern.

- Der zweite Filter wird durch kognitiv-affektive Faktoren gebildet. Wenn das situative Erleben den individuellen oder kollektiven Überzeugungssystemen zuwiderläuft, dann formt das ein „Überrascht-Sein“. Da hierbei individueller bzw. kollektiver Glaube, in Form von Zuneigung oder Stigmata, ins Spiel kommen, ist die Relevanz dieser Form der Überraschung tendenziell sehr hoch.

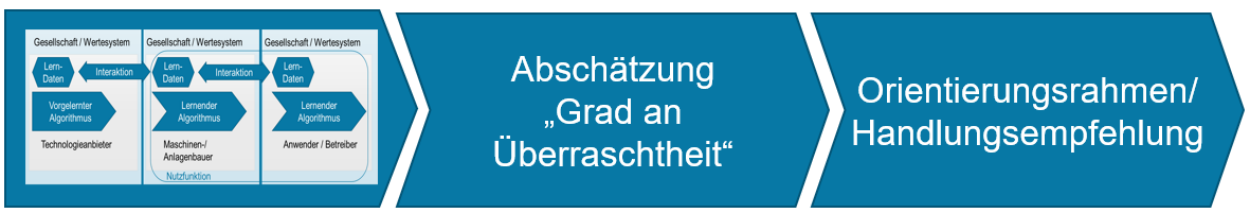
- Die anderen beiden Filter beziehen sich auf den politisch-institutionellen Rahmen und auf den kulturellen Hintergrund. Darin spiegeln sich die Werte und Regeln einerseits, andererseits die Grundüberzeugungen eines Kollektivs als Identität wider.

Diese vier Filter können werteraumspezifisch für die Beschreibung der Überraschtheit als Faktorisierung der Überraschung herangezogen werden. Unterschiedliche Werteräume können mit solchen kontextsensitiven Filtern detaillierter betrachtet werden.

Hilfestellung zur Erwartungsbetrachtung und Akzeptanz

Unter der Prämisse „Wie kann eine Nicht-Akzeptanz abgewendet werden?“ sind die beiden Fragestellungen, „Wie kann nachhaltiges Vertrauen geschaffen werden?“ und „Wie kann das Verständnis für die Überraschtheit gesteigert werden?“, wesentliche Kernelemente. Bei linear konstruierten Maschinen ist die Wirkungsweise hinsichtlich ihres Verhaltens und der Bandbreite ihrer Interaktion vielfach eindeutig zu beschreiben. Kommt KI zum Einsatz kann sich das Verhalten ändern. Je selbstständiger (abhängig vom Autonomie Level) und umfangreicher die KI arbeiten kann, hat dieses auch Auswirkungen in der Mensch-Technik-Interaktion, die in engen Grenzen (z.B. ein Regler wird in seiner Performance optimiert) von nicht durch den Menschen direkt wahrgenommenen bis hin zu überraschenden Interaktionen (z.B. ein Roboter reicht dem Menschen ein Bauteil in einer anderen für den Menschen neuen, aber optimierten Haltung an) führen kann.
Als Eingangsgrößen für die weiteren Handlungsempfehlungen sind Datenquellen, der lernende Algorithmus, die erwartete Verhaltensveränderung auf den Nutzer und der vorliegende Rechtsraum bzw. das bestehende Wertesystem zu betrachten. Im Ergebnis kann mittels der technischen Betrachtung und der Wahrnehmung über die vier Filter, die Überraschung aufgrund einer möglichen Verhaltensänderung als Abschätzung der Überraschtheit abgeleitet werden. Hieraus werden sich Handlungsempfehlungen ergeben, die dem Hersteller Hinweise für die weitere Vorgehensweise geben.

Die Handlungsempfehlungen reichen von „Keine Maßnahmen erforderlich“ bis hin zu einer umfangreicheren Analyse der Situation und der zu ergreifenden Maßnahmen. Wird z.B. von einer möglichen Ablehnung ausgegangen, ist Feedback vom Anwender einzuholen, ob die Veränderung auch vom Nutzer so bewertet wird. Die Erwartungskonformität aus der Betrachtung der Wertesysteme ist explizit zu ermitteln und in Folge durch z.B. eine verbesserte Beschreibung oder ein angepasstes Training eine mögliche Verhaltensänderungen herzustellen.
Werden nutzerbezogene Daten verwendet, ist eine weitere rechtliche Betrachtung notwendig, die hier nicht weiter ausformuliert wird. Grundsätzlich gilt immer, dass die Betrachtung der funktionalen Sicherheitstechnik weiterhin für die Sicherheit von Menschen, Maschinen und der Umwelt benötigt wird.



Ablauf für die Erstellung des Orientierungsrahmens

Zusammenfassung

Zusammenfassend hat die Arbeitsgruppe die Grundzüge und Mechanismen für einen konstruktiven Umgang mit Überraschungen aufbereitet. Ausgehend von der Semantik der „Autonomie Level“ und des Verhaltens von KI wurde der Begriff Überraschtheit als Indikator für die Akzeptanz und als Hinweisgeber für die Empfehlung für Maschinen- und Anlagenbauer eingeführt. Damit eröffnet sich ein europäischer Weg des KI-Einsatzes, der besonderen Wert für die Technikgestaltung hat und wichtig für den Export von KI-basierten Maschinen und Anlagen ist.

Autoren

Diese Publikation ist ein Ergebnis der Arbeitsgruppe „Technologie- und Anwendungsszenarien" der Plattform Industrie 4.0

Dieses Papier entstand unter der Mitarbeit von Johannes Kalhoff (Phoenix Contact GmbH & Co. KG), Johannes Diemer (Diemer Consulting 4.0 UG), Alexander Fay (Institut für Automatisierungstechnik), Stefan Böschen (RWTH Aachen University), Stefan Elmer (Festo SE & Co. KG), Christian Görg (TRUMPF Werkzeugmaschinen GmbH + Co. KG), Christoph Legat (Hekuma GmbH), Olga Mordvinova (incontext.technology GmbH), Andreas Nettsträter (Fraunhofer IML), Thomas Stiedl (Robert Bosch GmbH), Marco Ulrich (ABB AG Forschungszentrum Deutschland), Thomas Gries (Institut für Textiltechnik der RWTH Aachen University) und Gerd Bachmann (VDI Technologiezentrum GmbH).

Bildnachweis: Adobe Stock / Gorodenkoff; Plattform Industrie 4.0

Kontakt: Geschäftsstelle Plattform Industrie 4.0, Bülowstraße 78, 10783 Berlin

geschäftsstelle@plattform-i40.de