Auswirkungen künstlicher Intelligenz auf Normung und Arbeitsschutz

André STEIMERS, Thomas BÖMER

Institut für Arbeitsschutz
der Deutschen Gesetzlichen Unfallversicherung (IFA)
Alte Heerstraße 111, D-53757 Sankt Augustin

Kurzfassung: Der schnell wachsende Markt der künstlichen Intelligenz (KI) wird durch Erfolge auf den Gebieten der Bilderkennung, Spracherkennung sowie des automatisierten Fahrens in den kommenden Jahren zunehmend an Bedeutung für den Arbeitsschutz gewinnen. Trotz der rasch voranschreitenden positiven Entwicklungen und neuen Möglichkeiten für den Arbeitsschutz, bringt diese Entwicklung jedoch auch unbekannte Risiken mit sich. Bewährte Methoden zur Risikoreduzierung in der Softwareentwicklung können diesen neuen Herausforderungen nur bedingt begegnen. In diesem Beitrag werden einige der im Zusammenhang mit dieser Technologie neu auftretenden Risiken analysiert und neue Konzepte zur Risikoreduzierung vorgestellt, um ein Bewusstsein für die neuen Herausforderungen, die Systeme der künstlichen Intelligenz mit sich bringen, zu schaffen.

Schlüsselwörter: Künstliche Intelligenz, Machine Learning, Arbeitsschutz, Normung

1. Einleitung

Methoden der künstlichen Intelligenz (KI) werden bereits heute erfolgreich zur Lösung von Problemen in einer Vielzahl von Anwendungsgebieten eingesetzt. Insbesondere durch Methoden des Maschinellen Lernens und speziell des Deep Learnings lassen sich zunehmend große Fortschritte auf den Gebieten der Bilderkennung, Spracherkennung, Wissensextraktion sowie des automatisierten Fahrens aber auch der Analyse großer Datenmengen erkennen. Erfolge auf diesen Gebieten werden in den kommenden Jahren zunehmend an Bedeutung für das industrielle Umfeld, aber insbesondere auch für den Arbeitsschutz gewinnen.

Laut einer Studie durch Accenture Plc, Irland könnte sich 2025 die Wirtschaftskraft dieses Sektors (Automatisierte Roboter und Fahrzeuge, Datenanalyse, etc.) zwischen 6,5 und 12 Trillionen Euro pro Jahr bewegen (Purdy & Daugherty 2016). Eine weitere Studie, basierend auf einem makroökonomischen Modell für 12 Länder und 16 Wirtschaftszweige mit KI als zusätzlichem Produktionsfaktor, geht für Deutschland ab 2035 von einer zusätzlichen Bruttowertschöpfung von 1,6 % pro Jahr aus (Purdy & Daugherty 2017).

Bereits jetzt gibt es einige Anwendungen, die zeigen welchen Nutzen Methoden der künstlichen Intelligenz für den Arbeitsschutz haben können. Als Beispiel sei hier ein Projekt der Firma Microsoft Corp., USA genannt, das zum Ziel hat eine umfassende Überwachung von Baustellen und Lagerbereichen zu bieten. So können mit Hilfe einer speziellen Software nicht nur eine Überwachung der Benutzung persönli-

cher Schutzausrüstung wie Helme, Warnwesten und Handschuhe erfolgen, sondern auch Gefährdungen wie heiße Maschinen oder austretende gefährliche Flüssigkeiten erkannt und vor ihnen gewarnt werden.

Trotz der rasch voranschreitenden positiven Entwicklungen und neuen Möglichkeiten für den Arbeitsschutz, bringt diese Entwicklung jedoch auch bisher nicht bekannte Risiken mit sich. So ist bereits heute eine stetig zunehmende Anzahl von Unfällen, in denen Systeme der künstlichen Intelligenz beteiligt sind, zu verzeichnen. Zwei besonders schwerwiegende Unfälle, seien hier als Beispiele näher beschrieben:

- 03/2018 USA: Tödlicher Unfall automatisiertes Uber Taxi
  Während einer Testfahrt eines automatisiertes Uber Taxis im US Bundesstaat Arizona, erfasst das Taxi mit etwa 61 km/h eine die Fahrbahn überquerende Fußgängerin. Laut eines Berichtes des National Transportation Safety Boards (NTSB 2018-1) wurde etwa sechs Sekunden vor dem Unfall sowohl vom verbauten Radar- sowie LIDAR-Sensor ein Objekt im Weg des Fahrzeugs detektiert. Der Algorithmus scheiterte jedoch zuerst an der genauen Klassifizierung, welche erst 1,3 Sekunden vor dem Zusammenstoß vorgenommen werden konnte. Die Einleitung einer Notbremsung schlug anschließend jedoch fehl, weil das entsprechende vom Fahrzeughersteller verbaute System für Notbremsungen im automatischen Fahrbetrieb deaktiviert war. Die Testfahrerin leitete die Bremsung erst etwa eine Sekunde nach dem Zusammenstoß ein, da sie zuvor durch das Schauen einer TV Sendung abgelenkt war (Somerville & Shepardson 2018).
- 03/2018 USA: Tödlicher Unfall Tesla Modell X
  Im US Bundesstaat Kalifornien fährt ein im „Autopilot-Modus“ fahrender Tesla Modell X mit etwa 115 km/h in eine Fahrbahnbegrenzung, welche zwei Spuren voneinander trennt und kollidiert im Anschluss mit zwei weiteren Fahrzeugen. Laut eines Berichtes des National Transportation Safety Boards (NTSB 2018-2) ereignet sich der Unfall beim Abbiegevorgang, um einen mehrspurigen Highway über eine Ausfahrt zu verlassen. Dabei missinterpretierte die teilautomatische Steuerung die Fahrbahnbegrenzungen dahingehend, dass es die linke Begrenzungslinie der bisherigen Spur sowie die rechte Begrenzungslinie der Abbiegespur als neue Begrenzung heranzog und den von ihnen aufgespannten Zwischenbereich als neue Fahrspur interpretierte. Kurz vor dem Aufprall auf die sich dort befindenden Barrieren, welche als Fahrbahnbegrenzung dienen, beschleunigte die Steuerung das Fahrzeug nochmals, anstatt den Bremsvorgang einzuleiten.

2. Neue Risiken der Künstlichen Intelligenz

Die im letzten Abschnitt vorgestellten Unfälle verdeutlichen, dass im Bereich der künstlichen Intelligenz im Zusammenhang sicherheitsgerichteter Systeme Handlungsbedarf besteht. Es hat sich gezeigt, dass die modernen Methoden der KI, bisher unbekannte Risiken hervorbringen, die zum Versagen des Systems beitragen können. Insbesondere Algorithmen des maschinellen Lernens und dem untergeordnet auch des Deep Learnings erfordern eine genaue Analyse ihrer individuellen Eigenheiten bei der Erstellung einer Risikoanalyse. So lassen sich generell folgende Ausfallursachen für diese Modelle ausmachen:

- Unzureichende oder fehlerhafte Definition der Systemspezifikation
  In der Design- und Spezifikationsphase eines Projekts, werden die Anforderungen an das zu erstellende System analysiert, aus denen sich schließlich die

genauen Spezifikationen des Systems ableiten. Werden diese Anforderungen nicht vollständig betrachtet oder basieren auf fehlerhaften Annahmen, setzt sich dies unmittelbar auf die umzusetzende Spezifikation fort. Fehler, die in dieser Phase gemacht werden, können daher nur sehr schwierig oder gar nicht mehr behoben werden.

In Systemen welche Verfahren der KI einsetzen, beziehen sich diese Fehler hauptsächlich auf die genaue Formulierung des vom Algorithmus zu lösenden Problems. Werden hier die Eigenheiten verschiedener Algorithmen nicht mitbetrachtet, führt dies zur Auswahl eines Algorithmus, der nur unzureichend auf die Aufgabe des Systems hin angepasst werden kann.

- Fehlerhafte Implementierung des Algorithmus

  Viele Verfahren der KI basieren noch heute auf nicht lernenden Methoden. Wie in der Entwicklung einer jeden anderen Software gilt auch hier, dass sich Fehler in der Implementierung des Algorithmus unmittelbar auf das Ergebnis auswirken. Lernende Algorithmen haben dahingegen die Eigenschaft, dass sie sich in ihrer Trainingsphase selbstständig anpassen. Dieser Prozess wird jedoch ebenfalls von der Implementierung des Lernalgorithmus bestimmt. Die Ergebnisse des anschließend erzeugten bzw. trainierten Algorithmus hängen somit mittelbar von der Qualität der richtigen Implementierung des Lernalgorithmus ab.

  Insbesondere ist zu beachten, dass Methoden der KI oft hochdimensionale Probleme darstellen, was oftmals den Einsatz von GPUs (Graphics Processing Unit) erfordert. Deren Programmierung beinhaltet jedoch auch neue Fehlermöglichkeiten. Exemplarisch genannt seien hier nur Race Conditions, Deadlocks und Messeffekte.

- Ungeeignete Auswahl der Features

  Werden zur Entscheidungsfindung durch den trainierten Algorithmus falsche Features herangezogen, kann dies zu einem unerwünschten Systemverhalten führen. So könnte ein automatisiert fahrendes Fahrzeug, welches nur auf deutschen Landstraßen trainiert wurde, beispielsweise die Grasnarbe am Straßenrand als wichtiges Merkmal für die Fahrbahnbegrenzung heranziehen und somit in Gegenden mit kaum vorhandener Vegetation versagen.

  Am im vorherigen Abschnitt beschriebenen Beispiel des tödlichen Unfalls eines Tesla Modell X wurde zudem deutlich, dass hier schon kleine Fehlinterpretationen fatale Auswirkungen haben können.

  Dieses Risiko steht im engen Zusammenhang mit dem Begriff Transparenz, was gerade bei Methoden des Deep Learnings eine hohe Bedeutung aufweist. Diese Methoden werden vor allem für komplexe Aufgaben eingesetzt, die von einem Menschen nur unter immensen Aufwand oder gar nicht algorithmisch umgesetzt werden können. Ein Beispiel hierfür ist die Klassifikation von Objekten, wie einem Auto oder Menschen in einem Bild. Diese Komplexität bringt den Nachteil mit sich, dass der Weg der Entscheidungsfindung des bspw. angelernten Neuronalen Netzes nicht mehr nachvollziehbar bzw. intransparent ist, wodurch ungeeignete Features nicht mehr einfach identifiziert und entfernt werden können.

- Bias oder Rauschen im Trainingsdatensatz

  Methoden des maschinellen Lernens werden dazu eingesetzt, auf Basis vorhandener Daten Erkenntnisse zu gewinnen und diese so zu verallgemeinern, dass sie für die Analyse von neuen, bisher unbekannten Daten verwendet werden können. Die Basis dieser Erkenntnisse sind somit die zum Trainieren des Algorithmus verwendeten Daten. Beinhalten diese Daten Bias, so wird auch dieser als Teil des zu generalisierenden Modells angesehen, wodurch letztlich unerwünschte Ergebnisse produziert werden.

  Weiterhin hat sich gezeigt, dass auch Rauschen einen äußerst negativen Einfluss auf das Modell haben kann. Dies geht sogar so weit, dass durch künstlich erzeugtes aber vom Menschen nicht zu erkennendes Rauschen das, zu den zum Training verwendeten Daten hinzugefügt wird, vollkommen neue Ergebnisse provoziert werden können. So ist es beispielsweise möglich einen zuvor erkannten Menschen aus einem Bild zu löschen und stattdessen eine freie Straße zu erzeugen (Cisse 2017).
- Falsche Parametrisierung

  Neben die zum Anlernen des Algorithmus verwendeten Daten ist die Parametrisierung des Trainingsalgorithmus von entscheidender Bedeutung für die Qualität des trainierten Algorithmus und somit die Qualität der Entscheidungsfindung. Beispiele für wichtige Parameter sind hier Anzahl der Schichten eines neuronalen Netzwerks oder der Grad des Polynoms eines regressionsbasierten Algorithmus.

  Werden diese Parameter zu niedrig angesetzt, können sie die Komplexität der Aufgabe nicht zureichend abbilden, werden sie jedoch zu hoch angesetzt besteht die Gefahr der Überanpassung (engl. Overfitting). Als Überanpassung wird eine zu genaue Anpassung an die zum Training verwendeten Daten bezeichnet. Daraus ergibt sich der Nachteil, dass das trainierte Modell zwar sehr gut die vorhandenen Daten repräsentiert, jedoch nur noch unzureichend die eigentlich gesuchte zugrundeliegende Struktur der Daten beschreibt.
- Hardwarebezogene Fehler

  Treten während der Trainingsphase oder im Betrieb eines KI-Systems hardwarebezogene Fehler auf, so können diese die korrekte Ausführung des Algorithmus negativ beeinträchtigen oder sogar vollständig unterbinden.

  Allgemein lassen sich hardwarebezogene Fehler in drei Gruppen einteilen. Die erste Gruppe bezieht sich dabei auf klassische Hardwarefehler, die auf defekten Bauteilen basiert. Die zweite Fehlergruppe stellen Soft-Errors dar. Als Soft Errors werden unerwünschte temporäre Zustandsänderungen von Speicherzellen oder Logikbauteilen bezeichnet, die zumeist durch energiereiche Strahlung verursacht werden. Die letzte Gruppe bilden fehlerhafte Treiber und, insofern genutzt, fehlerhafte Module der Applikationssoftware der verwendeten Hardware.
- Menschliche Faktoren

  Der letzte hier betrachtete Risikofaktor in einem System, bei dem es zu einer Interaktion zwischen einem Menschen und einer Maschine kommen kann, stellt der Mensch dar. Hier lassen sich drei Risikoursachen identifizieren.

  Die erste Risikoursache beschreibt den falschen Gebrauch eines KI-Systems. Wird dieses für Aufgaben eingesetzt für das es nicht konzipiert wurde, kann dies zu ernsthaften Konsequenzen führen. Beispiele hierfür lassen sich heute im Bereich der Nutzung teilautomatisierter Steuerungen finden, bei denen der Nutzer die vorgeschriebene ständige Überwachung unterlässt.

  Ein weiteres Risiko stellt der Missbrauch der Technologie dar. So könnte ein

Assistenzsystem zur Wissensextraktion den Menschen durch eine im Sinne des Herstellers liegende Auswahl von Ergebnissen beeinflussen.

Allerdings ist es auch möglich, dass sich aus dem Nichtgebrauch dieser Technologie Risiken ergeben. So ist ein System, das für eine spezielle Aufgabe konstruiert wurde und über eine entsprechende Sensorik verfügt, meist wesentlich sensitiver als ein Mensch. Da sich in einem solchen Fall die Entscheidungsfindung des Algorithmus auf eine bessere Datenbasis stützt, wäre der automatische Betrieb hier dem Manuellen vorzuziehen. Ein Beispiel hierfür sind erste Assistenzsysteme für Flurförderfahrzeuge.

3. Internationale Standardisierung im Bereich der künstlichen Intelligenz

Den im vorherigen Kapitel genannten neuen Herausforderungen können die bewährten Methoden zur Risikoreduzierung in der Softwareentwicklung nur bedingt begegnen und auch der internationale Standardisierungsstand bietet hier aktuell noch keine Abhilfe.

Aus diesen Gründen beschlossen die internationalen Normungsorganisationen ISO und IEC im Jahr 2017 die Gründung eines gemeinsamen Standardisierungsgremiums, das sich umfassend mit dieser Thematik auseinandersetzt. Dieses Gremium ist als Subcommittee 42 (SC 42) dem Joint Technical Committee 1 (JTC 1) „Information Technology“ zugeordnet und ist zurzeit wie folgt untergliedert:

- ISO/IEC JTC 1/SC 42/JWG 1 Joint Working Group SC 42 - SC 40:
  Governance implications of AI
  Die Joint Working Group 1 zwischen SC 42 „Artificial Intelligence“ und SC 40 „IT Service Management and IT Governance“ beschäftigt sich mit der Ausarbeitung eines Frameworks, das zum Verständnis der Auswirkungen des Einsatzes verschiedener KI-Technologien beitragen soll sowie mit Verfahren die es den verantwortlichen Stellen ermöglichen sollen die Einführung solcher Technologien zu überwachen und zu evaluieren.
- ISO/IEC JTC 1/SC 42/SG 1
  In dieser Studiengruppe werden die besonderen Eigenschaften und Charakteristika verschiedener AI-Methoden untersucht und die Architekturen ihrer Algorithmen in der Tiefe analysiert. Weitere Aufgaben der SG 1 sind die Beobachtung des aktuellen Forschungsumfeldes sowie die Erarbeitung von Vorschlägen für neue Normungsvorhaben auf diesem Gebiet.
- ISO/IEC JTC 1/SC 42/WG 1 Foundational standards
  Diese Arbeitsgruppe beschäftigt sich neben der Aufstellung eines grundlegenden Vokabulars, mit der Definition wichtiger Begriffe sowie der Entwicklung von Taxonomien, wie beispielsweise der Aufstellung eines Lebenszyklus für KI-Systeme.
- ISO/IEC JTC 1/SC 42/WG 2 Big Data
  Die laufenden Projekte der WG 2 beschäftigen sich mit der Erstellung eines gemeinsamen Vokabulars für Big Data, der Schaffung eines allgemeinen Überblicks, aber auch der Entwicklung einer Referenzarchitektur für Big Data.
- ISO/IEC JTC 1/SC 42/WG 3 Trustworthiness
  Die Arbeitsgruppe Trustworthiness beschäftigt sich mit der Frage, wie sich dieser Begriff für ein KI-System definieren lässt sowie den Verfahren, welche zu einem vertrauenswürdigen/ zuverlässigen KI-System führen können.

Hierfür werden unterschiedliche Arten von Bias identifiziert und analysiert, aber auch Security- und Datenschutzaspekte betrachtet, um davon ausgehend Eigenschaften zu identifizieren, über die ein solches System verfügen muss und Verfahren zu finden, mit deren Hilfe diese Eigenschaften implementiert werden können.

- ISO/IEC JTC 1/SC 42/WG 4 Use cases and applications

  In der WG 4 werden neue potentielle Anwendungsfelder identifiziert und Empfehlungen zur Best Practice beim Einsatz der künstlichen Intelligenz in diesen Anwendungsfeldern erstellt.

4. Diskussion

Der rapide wachsende Markt der künstlichen Intelligenz wird zurzeit durch die starke strategische Förderung durch nationale aber auch internationale Institutionen sowie die Privatwirtschaft vorangetrieben. Es ist zu beobachten, dass Methoden der künstlichen Intelligenz auf immer mehr Gebieten Anwendung finden. Den sich daraus ergebenen positiven Entwicklungen stehen jedoch auch große gesellschaftliche und technologische Risiken gegenüber. Diesen Risiken zu begegnen stellt eine große Herausforderung dar. Die internationale Normung hat sich inzwischen diesem Gebiet angenommen und arbeitet intensiv an neuen Projekten. Durch die langen Projektlaufzeiten internationaler Normen, sind solche jedoch erst in einigen Jahren zu erwarten. Es ist jedoch geplant eine Serie von „Technical Reports“ herauszubringen, welche bereits gute Unterstützung dabei bieten können ein KI-System nach anerkannten Regeln zu erstellen.

5. Literatur

Cisse, M., Adi, Y., Neverova, N.,Keshet, J. (2017) Houdini: Fooling Deep Structured Visual and Speech Recognition Models with Adversarial Examples. In NIPS 2017, 6980-6990.

National Transportation Safety Board (2018-1) Preliminary Report Highway HWY18MH010. www.ntsb.gov, Aufruf: 26.11.2018.

National Transportation Safety Board (2018-2) Preliminary Report Highway HWY18FH011. www.ntsb.gov, Aufruf: 26.11.2018.

Purdy M., Daugherty P. (2016) Why AI is the future of growth. Accenture Plc.

Purdy M., Daugherty P. (2017) How AI boosts industry profits and innovation. Accenture Plc.

Somerville H., Shepardson D. (2018) Uber car's 'safety' driver streamed TV show before fatal crash: police, www.reuters.com/article/us-uber-selfdriving-crash-idUSKBN1JI0LB, Aufruf: 26.11.18.

Gesellschaft für
Arbeitswissenschaft e.V.

Arbeit interdisziplinär
analysieren – bewerten – gestalten

65. Kongress der
Gesellschaft für Arbeitswissenschaft

Professur Arbeitswissenschaft
Institut für Technische Logistik und Arbeitssysteme
Technische Universität Dresden

Institut für Arbeit und Gesundheit
Deutsche Gesetzliche Unfallversicherung

27. Februar – 1. März 2019

GfA-Press

Bericht zum 65. Arbeitswissenschaftlichen Kongress vom 27. Februar – 1. März 2019

Professur Arbeitswissenschaft, Institut für Technische Logistik und Arbeitssysteme, Technische Universität Dresden;
Institut für Arbeit und Gesundheit, Deutsche Gesetzliche Unfallversicherung, Dresden

Herausgegeben von der Gesellschaft für Arbeitswissenschaft e.V.
Dortmund: GfA-Press, 2019
ISBN 978-3-936804-25-6

NE: Gesellschaft für Arbeitswissenschaft: Jahresdokumentation

Als Manuskript zusammengestellt. Diese Jahresdokumentation ist nur in der Geschäftsstelle erhältlich.

Schriftleitung: Matthias Jäger
im Auftrag der Gesellschaft für Arbeitswissenschaft e.V.

Screen design und Umsetzung
© 2019 fröse multimedia, Frank Fröse
office@internetkundenservice.de · www.internetkundenservice.de