

Entwicklung einer Methodik zur Prüfung der Wirksamkeit von künstlicher Intelligenz

Christian Lossos

Schwarz Dienstleistung KG
Stiftsbergstraße 1
74172 Neckarsulm
Christian.Lossos@gmail.com

Frank S. Morelli

Hochschule Pforzheim
Tiefenbronner Straße 65
75175 Pforzheim
Frank.Morelli@hs-pforzheim.de

Simon Geschwill

Schwarz Dienstleistung KG
Stiftsbergstraße 1
74172 Neckarsulm
Simon.Geschwill@mail.schwarz

ABSTRACT

Künstliche Intelligenz (KI) entwickelt sich durch optimierte Rechenleistungen, große Mengen an verfügbaren Daten und massiven Investitionen immer mehr zu einem unverzichtbaren Bestandteil des Alltags. Gleichzeitig führen die Komplexität der Technologie und fehlende Optionen zur eindeutigen Nachvollziehbarkeit von Entscheidungsprozessen zu vielfältigen, neuartigen Risiken. Vor dem Hintergrund möglicher Folgen durch maschinelle Fehlentscheidungen steht die Interne Revision in der Verantwortung, die Wirksamkeit von KI zu bewerten. Eine strikte Anwendung generischer Prüfungshandlungen erweist sich nach dem aktuellen Stand der Technik als nicht durchführbar. Für die Wirksamkeitsprüfung wird deshalb ein indirektes Bewertungsverfahren im spezifischen Anwendungsfall vorgeschlagen.

SCHLÜSSELWÖRTER

Künstliche Intelligenz (KI), Literaturanalyse, Risikomanagement, Interne Revision, Prüfungsmethodik

MOTIVATION

Im Sommer des Jahres 1955 wurden durch die Dartmouth-Konferenz im US-Bundesstaat New Hampshire die Grundlagen für das Fachgebiet der künstlichen Intelligenz (KI) erarbeitet. Knapp 60 Jahre später gewann erstmals ein Computerprogramm auf Basis von KI den Wettkampf im strategischen Brettspiel Go gegen den weltbesten Spieler - ein Meilenstein für das Gebiet um KI (Silver, Schrittwieser, Simonyan, Antonoglou, & Huang, 2016).

Heute sind digitale Assistenten wie Siri und Alexa in vielen Haushalten präsent und selbst Kinder nutzen diese wie selbstverständlich. Gleichzeitig wird KI in vielen Bereichen und Branchen eingesetzt und erzielt mittlerweile bessere Ergebnisse als Menschen, beispielsweise in der Erkennung von Hautkrebs (Haenssle, Fink, & Schneiderbauer, 2018).

Die Auswirkungen von KI auf das gesellschaftliche und wirtschaftliche Leben sind enorm und können bisher nur schwer quantifiziert werden. Analysten der Unternehmensberatung McKinsey & Company bewerten die Auswirkungen von KI auf die Weltwirtschaft höher als die durch die Erfindung der Dampfmaschine im 19. Jahrhundert und dem Aufkommen von Informations- und Kommunikationstechnologien im 21. Jahrhundert.

Bis zum Jahr 2030 ergibt sich eine jährliche Steigerung des durchschnittlichen Bruttoinlandsprodukts um 1,2 Prozentpunkte sowie ein globaler Wertschöpfungsbeitrag in Höhe von 13 Milliarden US-Dollar durch das Aufkommen von KI (Bughin, Seong, & Manyika, 2018, S. 1-3).

Problemstellung

KI wird zunehmend in vielen Wirtschaftsbereichen eingesetzt und ermöglicht es Unternehmen ihre Handlungsmöglichkeiten zu erweitern und zu optimieren. Damit kann die Technologie einen entscheidenden Beitrag leisten, die jeweilige Marktposition zu sichern bzw. zu verbessern. Diese Chance geht jedoch mit verschiedenen Risiken einher, welche auf der „Blackbox“ als Grundprinzip von KI beruhen. Konventionelle Computerprogramme verarbeiten Daten durch explizite Anweisungen bzw. Befehle, die von Softwareentwicklern zur Lösung einer Problemstellung definiert wurden. Im Gegensatz dazu basiert KI auf selbstständigen Lernprozessen, wodurch eine Autonomie des Systems entsteht, die zu völlig neuartigen Problemlösungsansätzen führt. Durch die Komplexität der verwendeten Algorithmen ist es für Menschen nur mit erheblichem Aufwand möglich, die Entscheidungswege der KI nachzuvollziehen.

Als Resultat der fehlenden Nachvollziehbarkeit kam es bereits in der Vergangenheit zu Fehlfunktionen von KI in der Praxis. Ein selbstlernender Chatbot des Technologiekonzerns Microsoft adaptierte beispielsweise das Feedback einiger Twitter-Nutzer und veröffentlichte nach kurzer Zeit rassistische und sexistische Mitteilungen (Lee, 2016). Ähnliche Erfahrungen sammelte der Onlineversandhändler Amazon mit einer Software zur Unterstützung des Auswahlverfahrens für Bewerber auf Basis von KI, welche ebenfalls sexistische Tendenzen entwickelte und Männer besser bewertete als Frauen (Dastin, 2018).

Die genannten Beispiele zeigen, dass KI für Unternehmen sowohl neue Chancen als auch Risiken eröffnet. Dabei muss beachtet werden, dass hierbei, außer der negativen Berichterstattung in den Medien, bislang keine nennenswerten Folgen oder Schäden für die jeweils verantwortlichen Unternehmen entstanden sind.

Beim Blick in die Zukunft und möglichen Beteiligungen von KI an folgenschweren Entscheidungen stellt sich die Frage, wie man die Wirksamkeit von KI in der Praxis gewährleisten kann. Dabei bezieht sich die Wirksamkeit von KI im Rahmen des vorliegenden Artikels auf die prozessuale Effizienz sowie die Funktionsweise im Hinblick auf interne und externe Vorgaben.

Für die Interne Revision als „unabhängige und objektive Prüfungs- und Beratungs[einheit]“ (Deutsches Institut für Interne Revision e.V., 2015, S. 5) besteht die Notwendigkeit, angemessen auf den Einsatz von KI im Unternehmen zu reagieren. Dabei steht die Entwicklung innovativer Methoden zur Prüfung der Wirksamkeit von KI im Fokus, um weiterhin objektiv unternehmerische Risiken für Entscheidungsträger bewerten zu können (Clark, 2018, S. 1).

Stand der Forschung

In Theorie und Praxis erfolgt momentan eine intensive Bearbeitung des KI-Forschungsgebiets, woraus eine hohe Anzahl an wissenschaftlichen Veröffentlichungen resultiert. Im Rahmen der vorliegenden Arbeit kann zwischen Veröffentlichungen zur KI selbst sowie Veröffentlichungen zur Erklärbarkeit von KI (englisch: Explainable Artificial Intelligence, XAI) unterschieden werden. Die Analyse von Veröffentlichungen zu KI zeigt zwei Trends:

- Zum einen werden Verfahren und Methoden von KI verbessert, um resultierende Ergebnisse zu optimieren - beispielsweise Systeme zur Bilderkennung (Yi, Li, Feng, & Shi, 2018) und Systeme zur Durchführung von tiefem Lernen (Doshi, Hung, Liang, & Chiu, 2016). Zum anderen erfolgt eine Untersuchung der praktischen Anwendung von KI. Dies betrifft viele verschiedene Branchen und Use Cases, beispielsweise in den Bereichen Medizin (Kumar & Rao, 2018), Internet der Dinge (Abraham, et al., 2018) und autonomes Fahren (Kaur & Sobti, 2018).
- Ein weiterer Schwerpunkt der Forschungen betrifft die Auswirkungen von KI auf das wirtschaftliche und gesellschaftliche Leben. Darunter zählen beispielsweise ethische (Kumar, Kharkwal, Kohli, & Choudhary, 2016), rechtliche (Doshi-Velez & Korts, 2017) und soziale Aspekte (Krügel, Schütze, & Stoklas, 2018). Diese münden oftmals in der allgemeinen Fragestellung, inwiefern KI reguliert werden sollte (Etzioni, 2018).

Das Forschungsgebiet um XAI befasst sich mit der fehlenden Transparenz als Herausforderung im Sinne der „Blackbox“-Theorie. Während herkömmliche Software Schlussfolgerungen explizit auf Basis vorab implementierter Regeln trifft, können Systeme auf Basis von KI eigenständige Entscheidungen treffen. In diesem Kontext stellt die Fähigkeit zum selbstständigen Lernen von KI eine zusätzliche Herausforderung dar.

Um die Nachvollziehbarkeit von Entscheidungen autonomer Systeme auf Basis von KI zu gewährleisten, forschen Wissenschaftler an der Entwicklung entsprechender Methoden und Verfahren (Adadi &

Berrada, 2018; Dosilovic, Brcic, & Hlupic, 2018). Viele Fortschritte und Ergebnisse von XAI basieren jedoch auf äußerst komplexen, mathematischen Verfahren und sind gegenwärtig noch nicht für die Anwendung in der Praxis konzipiert. Weitergehende Analysen und Betrachtungen sind im vorliegenden Artikel daher nicht integriert.

Zielsetzung

Der vorliegende Artikel verfolgt zwei Zielsetzungen, um die Interne Revision auf den aufkommenden Einsatz von KI vorzubereiten:

- Zunächst wird durch eine Literaturanalyse der aktuelle Forschungsstand (State-of-the-Art) für den praktischen Einsatz von KI dokumentiert um mögliche, zukünftige Prüfungsfelder zu identifizieren.
- Ferner erfolgt die Entwicklung einer Methodik zur Prüfung der Wirksamkeit von KI im praxisorientierten Einsatz. Diese unterstützt die Interne Revision in der objektiven und unabhängigen Bewertung von Auswirkungen durch KI auf Unternehmensprozesse.

THEORETISCHE GRUNDLAGEN VON KI

Als Teilgebiet der Informatik stellt KI im Allgemeinen das Bestreben dar, menschliche Intelligenz durch Computer nachzubilden und in praktischen Anwendungsfällen einzusetzen. Zur Realisierung von KI werden Methoden und Verfahren der Disziplinen des maschinellen Lernens (englisch: Machine Learning) sowie des tiefen Lernens (englisch: Deep Learning) verwendet. Diese Differenzierung wird nachfolgend erläutert, um eine einheitliche, theoretische Grundlage für die weiteren Kapitel zu gewährleisten.

Künstliche Intelligenz

Obwohl der Trend zu KI erst in den vergangenen Jahren an Popularität zugenommen hat, wurden erste philosophische Überlegungen zur Abbildung von menschlichen Handelns in mechanische Terme bereits im 17. Jahrhundert durch Thomas Hobbes dokumentiert (Hobbes, 1651). Im weiteren Verlauf entstand in den 1960er-Jahren das Forschungsgebiet um intelligente Systeme, welches insbesondere durch Alan Turing und dessen Konzeption des Turing-Tests geprägt wurde (Turing, 1950).

Der Begriff „Künstliche Intelligenz“ wurde im Jahr 1955 durch McCarthy et al. im Rahmen der Dartmouth-Konferenz geprägt. Während dieser Veranstaltung entstand nicht nur die Begrifflichkeit KI, sondern auch das dazugehörige Forschungsgebiet. Zudem entstand durch die Konferenz eine erste KI-Definition:

“The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it.” (McCarthy, Minsky, Rochester, & Shannon, 1955, S. 1)

Seit der Dartmouth-Konferenz wurde das Gebiet um KI stetig weiterentwickelt, weshalb die damalige Definition

durch McCarthy et al. sich mittlerweile als nicht mehr zeitgemäß erweist. Zudem ist die Einigung auf eine allgemeingültige Definition von KI auf Grund der interdisziplinären Verwendung der Technologie nicht möglich. Beispielsweise wird KI in den Ingenieur-, Medizin- und Wirtschaftswissenschaften verwendet, wodurch sich je Fachrichtung andere Prioritäten bzw. Interpretationen und Ansichten ergeben. Dies erschwert die Entwicklung einer eindeutigen Definition (Ertel, 2016, S. 1). Weiterhin fällt bei der separaten Analyse der Bestandteile des Terms „künstliche Intelligenz“ auf, dass selbst für den Begriff Intelligenz keine allgemeingültige Definition besteht (Gentsch, 2018, S. 17).

Auf Basis des aktuellen Stands der Forschung wird folgende Arbeitsdefinition für KI angewendet: Künstliche Intelligenz (KI) ist ein Teilgebiet der Informatik mit der Zielsetzung, durch IT-Systeme menschliche Fertigkeiten und Fähigkeiten ab- bzw. nachzubilden, um dadurch eigenständig Aufgaben zu bearbeiten. Die technische Umsetzung erfolgt durch Algorithmen, welche sich selbstständig und anhand von Erfolgen und Fehlern sowie durch Lernprozesse kontinuierlich optimieren. Mögliche Einsatzgebiete umfassen beispielsweise das autonome Fahren, Systeme zur Spracherkennung und Sprachausgabe sowie die Abstraktion von Wissen.

Maschinelles Lernen

Zur Umsetzung von KI hat sich maschinelles Lernen etabliert. Wie der Name bereits erkennen lässt, liegt hierbei der Fokus auf selbstständigen Lernprozessen im Gegensatz zu expliziten Vorgaben durch Entwickler. Im Detail lassen sich zwei Aspekte im Kontext von maschinellem Lernen hervorheben, die Mitchell sowie Samuel wie folgt charakterisieren:

“A field of study that gives computers the ability to learn without being explicit programmed” - Arthur Samuel (Simon, 2012, S. 89)

“Machine Learning is the study of computer algorithms that improve automatically through experience.” (Mitchell, 1997)

Zunächst beschreibt Samuel den Aspekt, dass Algorithmen die Fähigkeit zum Lernen besitzen, ohne dass menschliche Entwickler explizite (Programmier-) Anweisungen hierfür implementieren. Zusätzlich wird durch Mitchell der Aspekt hervorgehoben, dass sich Algorithmen automatisch durch Erfahrungen verbessern. Beide Aspekte zeigen den Kerngedanken von maschinellem Lernen, dass Algorithmen eigenständig und ohne das Zutun von menschlichen Entwicklern die Fähigkeit haben zu lernen und sich auf Basis durchgeführter Tätigkeiten zu verbessern. Im Gegensatz zur herkömmlichen Softwareentwicklung geschieht dies bei maschinellem Lernen durch eine Anpassung der Algorithmen durch Feedback, welches abhängig vom jeweils erzielten Ergebnis ist. Hierzu wird durch den Algorithmus zunächst eine Hypothese erstellt, welche anschließend mit dem jeweiligen Ergebnis abgeglichen wird. Entsprechend der jeweiligen Abweichungen zwischen der Hypothese und dem Ergebnis erfolgt eine

iterative Anpassung des Algorithmus durch das System (Kirste & Schürholz, 2019, S. 24).

Eine grundsätzliche Einteilungsmöglichkeit von maschinellem Lernen besteht in der Differenzierung der Lernverfahren in überwachtes Lernen (englisch: supervised learning), unüberwachtes Lernen (englisch: unsupervised learning) sowie verstärktes Lernen (englisch: reinforced learning). Zentrale Unterschiede werden in Abbildung 1 dargestellt.

Parameter Lernverfahren	Initiale Informationen	Zielsetzung	Vorgehensweise
Überwachtes Lernen	Eingangsdaten und Ergebnisse auf Basis von Eingangsdaten	Vorhersage von Ergebnissen für unbekannte Eingangsdaten	Analyse von Zusammenhängen zwischen Ein- und Ausgabedaten
Unüberwachtes Lernen	Eingangsdaten	Clustering / Komprimierung von Eingangsdaten	Identifikation von Zusammenhängen innerhalb der Eingangsdaten
Verstärktes Lernen	Eingangsdaten und Belohnungsfunktion	Maximierung der Belohnungsfunktion und dadurch Problemlösung.	Trial & Error

Abbildung 1: Lernverfahren des maschinellen Lernens

Tiefes Lernen

Die momentanen Fortschritte von KI basieren vor allem auf Erkenntnissen im Gebiet des tiefen Lernens. Hierbei versteht man die künstliche Nachbildung von biologischen Nervennetzen durch Computersysteme zur Verarbeitung von Informationen. Die Grundlagen für diese Entwicklung bilden Forschungsergebnisse des Neurologen Warren S. McCulloch und des Logikers Walter Pitts aus dem Jahr 1943, welche sich mit der Funktionsweise des Gehirns und den ersten Neuronenmodellen beschäftigten (McCulloch & Pitts, 1943).

Grundlegend besteht ein Gehirn aus Millionen von Zellen (Neuronen), welche durch die Weitergabe elektrischer Impulse miteinander kommunizieren. Dabei sammelt ein Neuron elektrische Impulse verschiedener vorgelagerter Neuronen. Sobald ein definierter Schwellenwert elektrischer Impulse erreicht wird, gibt das Neuron einen elektrischen Impuls an nachgelagerte Neuronen weiter. Dieser Prozess verläuft über viele Ebenen von Neuronen, bis letztendlich eine Ausgabe erfolgt, beispielsweise durch einen Reflex. Durch diese Verkettung verschiedener Neuronen ähnelt das Gehirn einem komplexen neuronalen Netzwerk.

Die künstliche Nachbildung von neuronalen Netzen wurde im Jahr 1952 durch Alan Lloyd Hodgkin und Andrew Fielding Huxley geprägt. Durch sie entstand das Hodgkin-Huxley-Modell zur technischen Simulation einzelner biologischer Neuronen (Hodgkin & Huxley, 1952). Basierend auf den damaligen Erkenntnissen werden im heutigen Zeitalter durch die Informatik künstliche neuronale Netzwerke mit Millionen von Neuronen realisiert.

Grundlegend bestehen künstliche neuronale Netzwerke aus der Eingangs-, verschiedenen Verarbeitung- sowie der Ausgangsebene. Dabei erhalten die Neuronen der verschiedenen Schichten jeweils Informationen durch vorgelagerte Neuronen. Das betroffene Neuron verarbeitet diese und gibt wiederum Informationen an nachgelagerte Neuronen weiter. Die Verarbeitung von Informationen durch ein individuelles Neuron ist insbesondere abhängig von der Gewichtung der Verbindung mit dem vorgelagerten Neuron sowie dessen individuellen Schwellwerts (Kirste & Schürholz, 2019, S. 31).

Weitergehend lassen sich drei Typen von neuronalen Netzwerken unterscheiden:

- Verkoppelte neurale Netzwerke (englisch: Feedforward Neural Networks) zeichnen sich durch einen stringenten Informationsfluss von der Eingabe-, über die Verarbeitungs- zur Ausgabeebene aus.
- Im Gegensatz dazu können die Informationen bei wiederkehrenden neuronalen Netzwerken (englisch: Recurrent Neural Networks) über Rücksprünge bzw. Rückkopplungen zwischen den verschiedenen Ebenen im Wechsel fließen.
- Eine besondere Stellung nehmen faltende neuronale Netzwerke (englisch: Convolutional Neural Networks) ein. Diese werden insbesondere zur Verarbeitung von Bild-, Video und Audiodaten eingesetzt und basieren auf Faltungsfunktionen bzw. Konvolutionen.

Die erläuterten, technischen Grundlagen von KI stellen an dieser Stelle eine hinreichende Basis für die weiteren Kapitel gemäß der initialen Zielsetzung des vorliegenden Artikels dar. Weitere Informationen lassen sich beispielsweise den Ausführungen von Russel et al. (2010) und Goodfellow et al. (2017) entnehmen.

LITERATURANALYSE

Die Vielfältigkeit von KI stellt die Interne Revision vor die Herausforderung, deren Auswirkungen auf Prozesse in diversen Unternehmensbereichen zu bewerten. Um bereits im Vorfeld einen Ausblick auf zukünftige, mögliche Prüfungsfelder bereitzustellen, wird im Folgenden eine extensive und systematische Literaturanalyse des State-of-the-Art durchgeführt. Hierbei besteht die Zielsetzung in der Identifizierung von praktischen Anwendungsfällen von KI, welche momentan durch die Wissenschaft untersucht werden. Auf weitere Aspekte, wie beispielweise Ansätze zur Optimierung oder Entwicklungen von Verfahren auf Basis von KI, verzichten die Autoren an dieser Stelle bewusst.

Literaturanalysen haben nach Cooper das Ziel „[...] to describe, summarize, evaluate, clarify, and/or integrate the content of the primary reports.“ (Cooper, 1988, S. 107). Dazu werden wissenschaftliche Veröffentlichungen zum jeweiligen Themengebiet recherchiert und anschließend analysiert. Webster und Watson präzisieren das Vorgehen bei Literaturrecherchen, indem wissenschaftliche

Datenbanken nach entsprechender Literatur abzufragen sind (Webster & Watson, 2002, S. xvi). Die Durchführung von Literaturanalysen, vor allem die Phase der Literaturrecherche, ist im Forschungsbereich der Wirtschaftsinformatik äußerst aufwändig, da ständig neue Forschungsergebnisse veröffentlicht werden und damit ein Überblick nur schwer möglich erscheint (vom Brocke, et al., 2009, S. 2).

Um im Kontext von KI einen ganzheitlichen Überblick über den aktuellen Forschungsstand zu gewährleisten, wird für die nachfolgende Literaturanalyse ein systematischer Ansatz gewählt. Dieser orientiert sich am mehrphasigen Framework zur Durchführung von Literaturanalysen nach vom Brocke et al. welches, im Gegensatz zu vergleichbaren Vorgehensmodellen zur Literaturanalyse (wie z.B. nach Fettke (2006)), einen starken Fokus auf die Dokumentation legt: “[...] documenting the literature search process is a crucial part in any review article“ (vom Brocke, et al., 2009, S. 10). Zur Durchführung einer Literaturanalyse empfehlen vom Brocke et al. ein mehrphasiges Framework (siehe Abbildung 2).

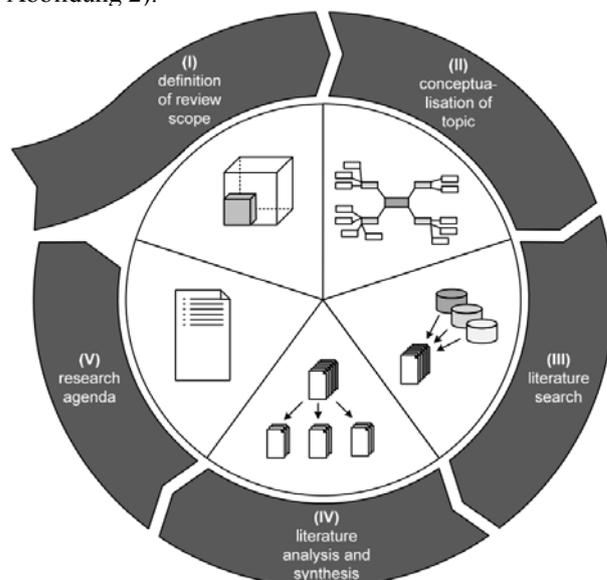


Abbildung 2: Verwendetes Framework für Literaturanalysen (vom Brocke, et al., 2009, S. 7)

Phase I: Definition des Umfangs

Durch Phase I wird das Ziel der jeweiligen Literaturanalyse definiert. Wie bereits beschrieben, sollen praktische Anwendungsfälle von KI identifiziert werden, um der Internen Revision einen Ausblick über mögliche zukünftige Prüfungsfelder zu ermöglichen.

Phase II: Konzeptualisierung des Themas

In Phase II erfolgt eine Einführung in die zu analysierende Thematik sowie in verwandte Themengebiete, um ein grundlegendes theoretisches Verständnis zu entwickeln. Hierbei wird auf die bereits beschriebenen Grundlagen von KI in diesem Artikel verwiesen.

Phase III: Literaturrecherche

Im Rahmen der vorliegenden Literaturanalyse verwenden die Verfasser als Suchbegriff den Terminus „Artificial Intelligence“. Um die Qualität der zu untersuchenden Literaturquellen zu gewährleisten, findet die Untersuchung ausschließlich in wissenschaftlich anerkannten Datenbanken statt. Eine wesentliche Basis für die zugehörige Fundierung stellen Peer-Review-Verfahren dar.

Aus diesem Grund sind die Datenbanken des Institute of Electrical and Electronics Engineers (IEEE Xplore Digital Library) sowie der Association for Computing Machinery (ACM Digital Library) für die Literaturrecherche ausgewählt worden.

Phase IV: Literaturanalyse und -synthese

Konzeptorientierte Vorgehensweisen spiegeln die Vorgehensweise vieler Literaturanalysen wider (Webster & Watson, 2002, S. xvi). Auch in der vorliegenden Untersuchung wird diese Methode eingesetzt, um die verschiedenen praktischen Anwendungsfälle von KI zu differenzieren. Webster und Watson empfehlen hierzu die Anwendung einer zweidimensionalen Konzeptmatrix gemäß Abbildung 3.

Articles	Concepts				
	A	B	C	D	...
1		*	*		*
2	*	*			
...			*	*	

Abbildung 3: Konzeptmatrix zur Literaturanalyse (Webster & Watson, 2002, S. xvii)

Im Verlauf der Literaturanalyse und -synthese erfolgt vertikal die Auflistung der analysierten Literaturquellen sowie horizontal eine Auflistung der identifizierten Konzepte. Zudem wird parallel eine Analyse der jeweils verwendeten Datengrundlage durchgeführt. Hierbei erfolgt die Zuordnung jeder einzelnen analysierten Literaturquelle zu einer der folgenden Kategorien:

- Bild- / Videodaten
- Audiodaten
- Alphanumerische Daten
- Numerische Daten
- Keine Spezifizierung

Durch die Anwendung dieser Methode soll die Erfassung von relevanten Anwendungsfällen von KI gewährleistet werden. Zudem lassen sich auf diese Weise Zusammenhänge zwischen den individuellen Literaturquellen identifizieren.

Phase V: Resultierende Erkenntnisse

Im Rahmen der Literaturanalyse erfolgte eine Abfrage der genannten Datenbanken zum Stichtag des 03.12.2018. Hierbei wurden insgesamt 128 Literaturquellen identifiziert und analysiert, welche den definierten Anforderungen des vorliegenden Artikels entsprechen.

Innerhalb der Wissenschaft wird das Themengebiet um die praktische Anwendung von KI momentan intensiv erforscht. Dies zeigt bereits die Anzahl an verfügbaren Quellen im Jahr 2018, auf die sich die vorliegende Literaturrecherche beschränkt.

Verwendete Datengrundlagen

KI kann und wird auf verschiedene Datengrundlagen angewendet, was die vorliegende Literaturanalyse bestätigt. Abbildung 4 zeigt die Verteilung der jeweiligen Datengrundlagen auf die Literaturquellen in absteigender Reihenfolge entsprechend ihrer Häufigkeit.

Datengrundlage	Absoluter Anteil (n=128)	Relativer Anteil (in Prozent)
Numerische Daten	56	44
Bild- / Videodaten	42	33
Keine Spezifizierung	23	18
Audiodaten	4	3
Alphanumerische Daten	3	2

Abbildung 4: Verteilung der Datengrundlagen auf die analysierten Literaturquellen

Verarbeitungen numerischer Daten werden bereits seit vielen Jahren, beispielsweise im Kontext des Business Intelligence, durchgeführt, wodurch sich deren hoher Anteil erklärt. Die dabei verwendeten und eingesetzten Methoden lassen sich durch den ergänzenden Einsatz von KI deutlich optimieren. Die Existenz entsprechender Methoden, Verfahren und Datenbestände erklärt, warum numerische Daten die am häufigsten verwendete Datenbasis im Rahmen der Literaturanalyse repräsentieren.

Ebenfalls eine hohe Verbreitung hat die Verarbeitung von Bild- und Videodaten. Sie gilt als eine der zentralen Stärken von KI. Lange Zeit handelte es sich um eine Domäne, in welchem ausschließlich Menschen eingesetzt werden konnten. Durch das verstärkte Aufkommen effizienter Algorithmen auf Basis von KI und der Verfügbarkeit hoher Rechenleistungen ist es jedoch mittlerweile möglich, entsprechende Tätigkeiten automatisiert durchzuführen.

Während viele der analysierten Literaturquellen auf spezifischen Datengrundlagen basieren, enthalten einige Arbeiten keine zugehörigen Hinweise. Als Grund für diese Feststellung lassen sich Arbeiten aufführen, welche die Anwendung von KI innerhalb eines spezifischen Forschungsgebiets thematisieren - beispielsweise im Rahmen einer Literaturanalyse. Zudem werden durch die Wissenschaft auch Arbeiten zur Anwendung von KI ohne technischen Hintergrund verfasst. Diese befassen sich beispielsweise mit ethischen, sozialen oder ökonomischen Auswirkungen von KI.

Konzeptionierung gemäß der Anwendungsfälle

Im Rahmen der Literaturanalyse wurden insgesamt 128 Literaturquellen bzw. Anwendungsfälle von KI analysiert. Die Anwendungsfälle wurden durch das

beschriebene, iterative Vorgehen zu 20 entwickelten Kategorien zugeordnet.

- Autonomes Fahren / Fahrerassistenzsysteme
- Bildung
- Chatbots
- Energie
- Entscheidungsunterstützung
- Finanzen
- Handschriftanalyse
- Internet der Dinge
- IT-Sicherheit
- Biologie
- Logistik
- Medizin
- Mobilität
- Öffentliche Verwaltung
- Personenanalyse
- Prozessoptimierung
- Qualitätssicherung
- Smart City
- Sprachanalyse
- Steuerungssysteme

PRÜFUNG VON KI

Fortschritte durch Wissenschaft und Praxis bilden die Grundlagen für einen zukünftig verstärkten Einsatz von KI in Unternehmen. Diese Entwicklung bietet zwar vielfältige Chancen, führt im Kontext des Risikomanagements jedoch auch zu Herausforderungen. Insbesondere die Interne Revision muss auf diesen Trend reagieren, um weiterhin die objektive Prüfung von Unternehmensprozessen zu gewährleisten.

Das Institute of Internal Auditors (IIA), ein internationaler Zusammenschluss von Revisoren, definiert die Rolle der Internen Revision im Kontext des zunehmenden Einsatzes von KI in Unternehmen wie folgt:

„[...] internal audit can help an organization evaluate, understand, and communicate the degree to which artificial intelligence will have an effect (negative or positive) on the organization's ability to create value in the short, medium, or long term.“ (Institute of Internal Auditors, 2017, S. 5)

Risiken beim Einsatz von KI

Sowohl technische als auch ökonomische Risiken resultieren aus dem Einsatz von KI in Unternehmen:

Technische Risiken

- Logische Fehler und Programmierfehler
Wie jede Software unterliegt auch KI dem Risiko von logischen Fehlern und Implementierungsfehlern. Dadurch kann die Effektivität von Algorithmen beeinflusst werden, was ggf. zu verringerter Qualität von Ergebnissen und damit zu massiven Auswirkungen auf den Anwendungskontext führt (Institute of Internal Auditors, 2017, S. 4).
- Qualität der verwendeten Daten

Sollte die verwendete Datengrundlage unidentifizierte, verzerrte Muster (algorithmischer Bias) beinhalten, kann diese Voreingenommenheit durch die KI verstärkt aufgenommen werden und dadurch das Datenmodell beeinflussen. Dies führt ggf. dazu, dass die Ergebnisse von KI diskriminierende Faktoren beinhalten und dadurch Personengruppen benachteiligen (Institute of Internal Auditors, 2017, S. 4).

- Cybersicherheit
Digitalen Angreifern erhalten durch den Einsatz von KI einen zusätzlichen Angriffsvektor. Zudem kann sich durch die Art der jeweils verarbeiteten Informationen (Personenbezogene- / sensible Unternehmensdaten) eine zusätzliche Motivation für Angriffe ergeben (Institute of Internal Auditors, 2017, S. 4).
- Regulatorische Anforderungen
Nationale und internationale Gesetzgebungen sind (noch) nicht vollumfänglich auf den Einsatz von KI angepasst. Dies führt zu teilweise unklaren Rechtslagen und bei beteiligten Stakeholdern zu rechtlichen Risiken.
Zeitgleich bestehen für Teilbereiche von KI strenge Regularien, beispielsweise im Kontext der Massenverarbeitung personenbezogener Daten durch die Europäische Datenschutz-Grundverordnung (DSGVO). Verfehlungen der Anforderungen in diesem sensiblen Bereich werden durch hohe Bußgelder sanktioniert.

Ökonomische Risiken

- Wirksamkeit im Anwendungsfall
Der Einsatz der von KI muss einen definierten, wertschöpfenden Beitrag leisten. Andernfalls werden vorhandene Potenziale des individuellen Anwendungsfalls nur unzureichend ausgeschöpft.
- Reputations- und finanzielle Schäden
Viele Bereiche, beispielsweise das Gesundheits- und Finanzwesen, erheben, verarbeiten und speichern persönliche Daten mittels KI und leiten daraus folgenschwere Entscheidungen ab.
„Voreingenommene, fehleranfällige, gehackte oder unethische Systeme bergen daher erhebliche Reputationsrisiken“ (Boillet, 2018, S. 2) sowie die Gefahr enormer finanzieller Schäden für das jeweils verantwortliche Unternehmen.
- Ablehnung von KI
Obwohl durch den Einsatz von KI vielfältige Chancen generiert werden, kann deren Einsatz von verschiedenen Personengruppen (z.B. Kunden oder Mitarbeitern) abgelehnt werden (Institute of Internal Auditors, 2017, S. 4). Als zugehörige Ursachen lassen sich beispielsweise die Angst vor dem Verlust von Arbeitsplätzen oder Vorbehalte gegenüber der Technologie auf Grund der Verarbeitung von persönlichen Informationen identifizieren.
- Umsetzung von Rentabilitätsanforderungen
Investitionen in KI (Hard- und Software, Gehälter etc.) gewährleisten nicht die Erfüllung definierter

Ziele (Institute of Internal Auditors, 2017, S. 5) per se. Daher kann es bei der Implementierung von KI vorkommen, dass vorab geplante Kennzahlenwerte nicht erfüllt werden.

Prüfmethodik für KI

Zum gegenwärtigen Zeitpunkt stellen Prüfungshandlungen im Kontext von KI in vielen Unternehmen zwar noch eine Seltenheit dar, werden jedoch in Zukunft vermehrt erfolgen. Zur Durchführung und Unterstützung der Internen Revision ist deshalb die Entwicklung zugehöriger neuartiger Methoden und Verfahren essenziell. Im Folgenden wird daher eine Methodik zur Prüfung der Wirksamkeit von KI beschrieben. Diese basiert auf insgesamt sieben Prüfungsmodulen und soll bei zugehörigen Prüfungshandlungen unterstützen.

Neben den eingesetzten Algorithmen sowie den dadurch erzeugten Datenmodellen umfassen Prüfungshandlungen im Kontext von KI weitere Bereiche. Beispielsweise beeinflussen eingesetzte IT-Infrastrukturen, Aspekte des Datenschutzes, Cybersicherheit, Beziehungen zu externen Partnern, Vorgehen bei Veränderungen, der Umgang mit Partnern sowie betriebswirtschaftliche Kennzahlen die Wirksamkeit von KI erheblich. Hinzu kommt der Faktor der beschriebenen Blackbox und der dadurch hohe Aufwand, Entscheidungswege von KI nachzuvollziehen.

Nachfolgend werden die Prüfmodule aufgeführt sowie deren Grundlagen individuell beschrieben.

Entwicklungsvorgehen

Ergebnisse von KI sind stark abhängig von der Qualität und Quantität der verwendeten Daten, was der Grundsatz „Garbage In, Garbage Out“ veranschaulicht. Somit bildet das Prüfmodul „Entwicklungsvorgehen“ einen elementaren Bestandteil der Prüfmethodik.

Dem Prüfmodul liegt das Prozessmodell Cross-Industry Standard Process for Data Mining (CRISP-DM) zu Grunde, welches in den 1990er-Jahren ursprünglich für Data Mining-Projekte entwickelt wurde. Mittlerweile wird CRISP-DM jedoch auch für weitere Projekte und Anwendungsfälle im Kontext von Data Science angewendet. Data Science zielt in diesem Zusammenhang als interdisziplinärer Ansatz darauf ab, Erkenntnisse aus Daten unterschiedlichster Bereiche z.B. in Form von „Big Data“ zu gewinnen. Zugehörige Verfahren und maschinelle Lernalgorithmen sollen Datenmodelle generieren, die deskriptive, prädiktive oder präskriptive Analysen ermöglichen. Unter anderem wird in der Literatur auch bei der Prüfung von KI das CRISP-DM-Prozessmodell vorgeschlagen und eingesetzt (Clark, 2018).

CRISP-DM besteht aus sechs Phasen zwischen denen, bei Bedarf, Rücksprünge erfolgen können. Grundlegend orientiert sich dieser Ansatz an der Vorgehensweise im Rahmen von Softwareprojekten: Beginnend bei betriebswirtschaftlichen Anforderungserhebungen des jeweiligen Anwendungsfalls, über Analysen und Anpassungen der jeweils vorliegenden Datenbestände

bis zur abschließenden Bereitstellung der entwickelten Lösung deckt CRISP-DM viele Aspekte ab.

Initial beinhaltet CRISP-DM keine Phase, welche zukünftige Entwicklungen / Anpassungen des entwickelten Modells explizit behandelt. Aus diesem Grund wird dieser Ansatz um die Phase „Future“ erweitert. Diese behandelt die Fragestellung, ob bereits in der Vergangenheit und Gegenwart Überlegungen und Entscheidungen zu möglichen Anpassungen des Modells in der Zukunft getroffen wurden - sowohl bedingt durch die Regulatorik als auch durch wechselnde, betriebswirtschaftliche Anforderungen. Die Thematik ist von besonderer Relevanz, da in vielen Bereichen ein hoher Kostendruck herrscht und mittlerweile auch ein Fokus auf der Nachhaltigkeit von IT-Investitionen liegt. Nachfolgend werden die Bestandteile (Kategorien) der einzelnen CRISP-DM-Phasen im Kontext der zu entwickelnden Prüfmethodik aufgeführt.

Business Understanding

- Anforderungsmanagement
- Regulatorik
- Risikomanagement
- Erfahrungswerte
- Saisonale Aspekte

Data Understanding

- Regulatorik
- Datenquelle
- Datenbeschaffenheit

Data Preparation

- Extraktion
- Transformation
- Laden

Modeling

- Verwendete Technologien
- Fortschrittskontrolle

Evaluation

- Anforderungsmanagement
- „Overfitting“ / „Undertraining“-Kontrolle (Ausgleich zwischen Generalisierbarkeit und Präzision von Modellen)

Deployment

- Reproduzierbarkeit
- Leistungskontrolle
- Innovationsmanagement

Future

- Regulatorik
- Flexibilität
- Wissensmanagement

IT-Infrastruktur

Als technische Grundlage von KI bildet die jeweils verwendete IT-Infrastruktur einen wesentlichen Bestandteil von Prüfungshandlungen. Dementsprechend widmet sich das Prüfmodul IT-Infrastruktur den eingesetzten Hard- und Softwarekomponenten insofern, als man hierbei deren Konfiguration bewertet. Der Fokus liegt im Einzelnen auf den folgenden Bereichen:

- Lizenzen
- Schichtenarchitektur
- Datensicherung
- Protokollierung
- Uhrzeiten
- Systemanforderungen
- Versionskontrolle

Die genannten Prüfansätze basieren auf der Norm ISO/IEC 27001:2013, die beispielsweise durch das Bundesamt für Sicherheit in der Informationstechnik (BSI) im Rahmen der BSI-Standards empfohlen wird.

IT-Sicherheit

Durch Digitalisierungsprojekte in sämtlichen Branchen hat sich IT-Sicherheit mittlerweile als fester Bestandteil in Unternehmen etabliert. KI, als Ergänzung zu bestehenden IT-Systemen, erzeugt in diesem Kontext einen weiteren Angriffsvektor.

Sämtliche Aktivitäten im zugehörigen Kontext unterliegen dem Ziel „elektronisch gespeicherte [...] Informationen und deren Verarbeitung“ (Bedner & Ackermann, 2010, S. 323) zu schützen. Dazu werden die grundlegenden Schutzziele nachfolgend erläutert:

- *Integrität*
Dieser Sachverhalt beinhaltet die Gewährleistung der „Vollständigkeit und Korrektheit der Daten (Datenintegrität) sowie die korrekte Funktionsweise des Systems (Systemintegrität)“ (Bedner & Ackermann, 2010, S. 326).
- *Vertraulichkeit*
Enthaltene Informationen eines IT-Systems sind ausschließlich berechtigten Personen zugänglich (Bedner & Ackermann, 2010, S. 323).
- *Verfügbarkeit*
„Zum einen muss die Datenverarbeitung inhaltlich korrekt sein und zum anderen müssen alle Informationen und Daten zeitgerecht zur Verfügung stehen und ordnungsgemäß verarbeitet werden“ (Bedner & Ackermann, 2010, S. 326).

Im Rahmen der IT-Sicherheit wird für jedes System eine Schutzbedarfsfeststellung gemäß den Standards des BSI durchgeführt. Dabei lassen sich den individuellen Schutzziele quantitative Schutzbedarfskategorien zuordnen (normal; hoch; sehr hoch). Entsprechend der definierten Schutzbedarfskategorie je Schutzziel werden weitergehende Maßnahmen zur Umsetzung der Anforderungen durchgeführt.

Bei Anwendung der Prüfmethodik erfolgt zunächst die Prüfung, ob eine Schutzbedarfsfeststellung durchgeführt wurde. Weitergehend wird geprüft, ob Maßnahmen zur Gewährleistung der Schutzziele gemäß des individuell definierten Schutzbedarfs umgesetzt worden sind.

Datenschutz

Spätestens seit Inkrafttreten der DSGVO ist das Thema Datenschutz in das Bewusstsein der Bevölkerung und von Unternehmen gerückt. Hierbei liegt der Fokus auf dem Schutz personenbezogener Daten, welche gemäß den gesetzlichen Vorgaben wie folgt definiert sind.

“Personenbezogene Daten [sind] alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person [...] beziehen [...]“ Art. 4 Abs. 1 DSGVO

Werden personenbezogene Daten gemäß Art. 4 Abs. 1 DSGVO verarbeitet, sind gemäß Art. 1 Abs. 1 DSGVO Maßnahmen zum Schutz der personenbezogenen Daten umzusetzen. Basierend auf den gesetzlichen Grundlagen der DSGVO können entsprechende Prüfungshandlungen durchgeführt werden.

Veränderungsmanagement

Entsprechend dem Motto „Never change a running system“ bergen Änderungen an Hard- oder Softwarekomponenten oftmals diverse Risiken. Dementsprechend müssen Änderungen mit Hilfe von Standardprozessen durchgeführt werden, womit sich das Auftreten sowie die Intensität von möglichen Risiken minimieren lässt.

Im Rahmen der Prüfmethodik werden folgende Kategorien definiert:

- Prozessvorgehen
- Verantwortlichkeiten
- Dokumentation
- Regulatorik
- Risikoanalyse
- Nachvollziehbarkeit
- Exklusivität

Die Prüfungshandlungen in den genannten Kategorien basieren auf der bereits genannten Norm ISO/IEC 27001:2013 sowie auf COBIT 5.

Partnermanagement

Externe Partner stellen durch deren Wissen und Fähigkeiten, insbesondere im Technologieumfeld, einen wichtigen Mehrwert zur erfolgreichen Durchführung von Entwicklungsprojekten dar. Dieser Aspekt wird gewürdigt, indem man vertragliche Vereinbarungen mit externen Partnern bewertet. Die vorliegende Prüfmethodik beinhaltet folgende Kategorien:

- Vertraglicher Rahmen
- Rechte und Pflichten der Beteiligten
- Laufzeit des Vertrags
- Leistungskontrolle des externen Partners

Als Grundlage zur Konzeption der Prüfungshandlungen werden sowohl der Prüfungsstandard Nr. 4 des Deutschen Instituts für Interne Revision e.V. als auch der Global Technology Audit Guide 12 des Institute of Internal Auditors verwendet. Beide Veröffentlichungen erläutern die Prüfung von IT-Projekten, welche auch die Prüfung von beteiligten, externen Partnern beinhaltet. Die jeweiligen Empfehlungen zu Prüfansätzen sind daher den Veröffentlichungen entnommen und an den vorliegenden Anwendungsfall angepasst worden.

Erfolgsmessung

Durch die Erhebung und Analyse von Kennzahlen wird die Wirtschaftlichkeit von KI in individuellen Anwendungsfällen bewertet. Hierzu ist das Prüfmodul

Erfolgsmessung in die Bereiche der „generischen Aspekte“ sowie der „spezifischen Aspekte“ gegliedert. Die generischen Aspekte nehmen Bezug auf Kennzahlen, welche bereits durch die Entwickler bzw. das Entwicklungsteam definiert wurden. Der Bereich „spezifische Aspekte“ bezieht sich auf Kennzahlen, die durch die verantwortlichen Prüfer generiert werden. Als zentraler Ausgangspunkt fungiert dementsprechend die Frage, welche weiteren Kennzahlen für den jeweiligen Anwendungsfall definiert werden sollen. Nachfolgend werden die Kategorien, gegliedert nach den genannten Bereichen, aufgeführt:

Generische Aspekte

- Entwicklung
- Kontrolle
- Kommunikation
- Gegenmaßnahmen
- Mehrwert
- Disruption

Spezifische Aspekte

- Offene Fragen

FAZIT

Im Rahmen der Literaturanalyse sind die identifizierten und analysierten Anwendungsfälle insgesamt 20 Kategorien zugeordnet worden. Auf Grund des breiten Themenspektrums der Kategorien lässt sich die Aussage treffen, dass man KI gegenwärtig in vielen Bereichen einsetzt und sich deshalb keine klare Abgrenzung vollziehen lässt, welche konkreten Anwendungsfälle von KI momentan untersucht werden. Vielmehr ergeben sich durch die genannten Kategorien verschiedene Prüfungsfelder, auf welche die Interne Revision im Rahmen zukünftiger Prüfungshandlungen einen besonderen Fokus legen sollte.

Im Hinblick auf die KI-Prüfung muss konstatiert werden, dass sich eine umfassende Prüfung der Wirksamkeit von KI mit dem momentanen Stand der Technik nur teilweise durchführen lässt. Diese Problematik basiert auf dem Sachverhalt, dass KI auf maschinellem und tiefem Lernen mit einem hohen Anteil von Autonomie der eingesetzten Verfahren basiert, welche die Nachvollziehbarkeit von Entscheidungsprozessen bzw. von getroffenen Entscheidungen für Anwender erheblich erschwert.

Zudem muss bei Prüfungshandlungen im Bereich von KI jeder Einzelfall individuell betrachtet werden, da eine strikte Anwendung generischer Prüfungshandlungen nicht möglich ist. Eine Wirksamkeitsprüfung von KI in Anwendungsfällen sollte daher über ein indirektes Prüfverfahren erfolgen. Dabei werden relevante Bereiche von KI wie z.B. das Entwicklungsvorgehen, die eingesetzte IT-Infrastruktur, Parameter der IT-Sicherheit etc. geprüft. Auf Basis der erhaltenen Erkenntnisse kann eine Einschätzung getroffen werden, inwiefern die Wirksamkeit der KI im jeweils geprüften Anwendungsfall vorhanden ist.

ZUKÜNFTIGE FORSCHUNG

Angesichts der durchgeführten Literaturanalyse lässt sich aussagen, dass die Breite und Tiefe des momentan bestehenden Themenspektrums an praktischen Anwendungsfällen eine solide Grundlage für die verstärkte Kooperation zwischen Wissenschaft und Wirtschaftsunternehmen bietet. Gegenwärtig setzen viele Unternehmen auf Grund mangelnder Ressourcen und fehlender Innovationskraft keine KI ein. Dadurch gehen Chancen verloren, die Effizienz von Geschäftsprozessen zu optimieren und somit Unternehmen die Möglichkeit zu eröffnen, Wettbewerbsvorteile zu sichern bzw. auszubauen.

Insbesondere wissenschaftliche Einrichtungen mit einem starken Anwendungsbezug können einen Beitrag leisten, durch bestehendes Wissen und Methodenkenntnissen die Wirtschaft zu stärken. Gleichzeitig besteht die Chance existierende Verfahren und Methoden im Bereich von KI am praktischen Beispiel zu erproben und zu optimieren. Neben der praxisorientierten Anwendung von KI sollte durch die Wissenschaft auch die Gewährleistung der Wirksamkeit von KI intensiv bearbeitet werden. Hierbei sind zwei Themenbereiche differenziert zu betrachten:

- Im Kontext der Prüfung von KI müssen neue Verfahren und Methoden entwickelt werden, um auf den technologischen Fortschritt adäquat zu reagieren.
- Parallel stellt der Themenbereich um XAI einen relevanten Ansatz dar, die Wirksamkeit von KI zu gewährleisten.

Die vorliegende Prüfmethodik für KI bietet eine Grundlage für formale, technische und prozessuale Prüfungshandlungen. Aufgrund der bisher weitgehend theoretisch geprägten Fortschritte mit wenigen anwendungsbezogenen Erkenntnissen für die Unternehmenswelt, wurde eine Bearbeitung der XAI-Thematik im Rahmen des vorliegenden Artikels hingegen nicht verfolgt. Die Autoren sind der Ansicht, dass die Wissenschaft diesem Themenbereich einen größeren Fokus geben sollte, um bereits bestehende Erkenntnisse und Methoden für die Praxis zu weiterzuentwickeln. Von besonderem zukünftigem Interesse könnte dabei die Weiterentwicklung von Systemen sein, die autonom durchgeführte Entscheidungsprozesse einfacher nachvollziehbar und transparenter visualisieren.

LITERATUR

- Abraham, A. M., Kulkarni, N., Clement, N., Bhat, L., Misale, N., Hussain, T., . . . Ramamurthy, B. (2018). Augmenting IoT-based Systems with Intelligence. *2018 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, (pp. 1-6). Bangalore.
- Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access: Volume 6*, 52138 - 52160.

- Bedner, M., & Ackermann, T. (2010). Schutzziele der IT-Sicherheit: 34. *Datenschutz und Datensicherheit*, 323-328.
- Boillet, J. (2018). *AI: A risk and a way to manage risk*. Paris: Ernst & Young.
- Bughin, J., Seong, J., & Manyika, J. (2018). *Notes from the AI frontier: Modeling the impact of AI on the world economy*. New York City: McKinsey Global Institute.
- Clark, A. (2018). The Machine Learning Audit - CRISP-DM Framework. *ISACA Journal*(1).
- Cooper, H. M. (1988). *Knowledge in Society*. Heidelberg: Springer.
- Dastin, J. (2018). *Amazon scraps secret AI recruiting tool that showed bias against women*. Canary Wharf: Reuters.
- Deutsches Institut für Interne Revision e.V. (2015). *Online-Revisionshandbuch für die Interne Revision in Kreditinstituten*. Frankfurt am Main.
- Doshi, R., Hung, K. W., Liang, L., & Chiu, K. H. (2016). Deep Learning Neural Networks Optimization using Hardware Cost Penalty. *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*, (pp. 1954-1957). Montreal.
- Doshi-Velez, F., & Kortz, M. (2017). Accountability of AI Under the Law: The Role of Explanation. *Harvard Public Law Working Paper No. 18-07*.
- Dosilovic, F. K., Brcic, M., & Hlupic, N. (2018). Explainable artificial intelligence: A survey. *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, (pp. 210-215). Opatija.
- Ertel, W. (2016). *Grundkurs Künstliche Intelligenz*. Wiesbaden: Springer Vieweg.
- Etzioni, O. (2018). Point: Should AI technology be regulated? Yes, and Here's How. *Communications of the ACM: Volume 61, Issue 12*, 30-32.
- Fettke, P. (2006). State-of-the-Art des State-of-the-Art. *Wirtschaftsinformatik 48*, 257-266.
- Gentsch, P. (2018). *Künstliche Intelligenz für Sales, Marketing und Service*. Wiesbaden: Springer Gabler.
- Goodfellow, I., Bengio, B., & Courville, A. (2017). *Deep Learning*. Cambridge: The MIT Press.
- Haenssle, H. A., Fink, C., & Schneiderbauer, R. (2018). *Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists*. Oxford: Annals of Oncology.
- Hobbes, T. (1651). *Leviathan: Or the Matter, Form, and Power of a Commonwealth, Ecclesiastical and Civil*. London.
- Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology: Volume 117, Issue 4*, 500-544.
- Institute of Internal Auditors. (2017). *Artificial Intelligence - Considerations for the Profession of Internal Auditing: Special Edition*. Lake Mary, Florida.
- Kaur, P., & Sobti, R. (2018). Scenario-based simulation of intelligent driving functions using neural networks. *2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, (pp. 1-6). Bangalore.
- Kirste, M., & Schürholz, M. (2019). Einleitung: Entwicklungswege zur KI. In V. Wittpahl, *Künstliche Intelligenz* (pp. 21-35). Berlin, Heidelberg: Springer Vieweg.
- Krügel, T., Schütze, B., & Stoklas, J. (2018). Legal, ethical and social impact on the use of computational intelligence based systems for land border crossings. *2018 International Joint Conference on Neural Networks (IJCNN)*, (pp. 1-8). Rio de Janeiro.
- Kumar, K., & Rao, A. C. (2018). Breast Cancer Classification of Image using Convolutional Neural Network. *2018 4th International Conference on Recent Advances in Information Technology (RAIT)*, (pp. 1-6). Dhanbad.
- Kumar, N., Kharkwal, N., Kohli, R., & Choudhary, S. (2016). Ethical Aspects and Future of Artificial Intelligence. *2016 International Conference on Innovation and Challenges in Cyber Security (ICICCS-INBUSH)*, (pp. 111-114). Noida.
- Lee, D. (2016). Tay: Microsoft issues apology over racist chatbot fiasco. *BBC News*.
- McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955). *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*. Hanover.
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics: Volume 5, Issue 4*, 115-133.
- Mitchell, T. M. (1997). *Machine Learning*. New York City: McGraw-Hill Education.
- Russell, S. J., & Norvig, P. (2010). *Artificial Intelligence: A Modern Approach*. Upper Saddle River: Prentice Hall.
- Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., & Huang, A. (2016). *Mastering the Game of Go without Human Knowledge*. London: DeepMind.
- Simon, P. (2012). *Too Big to Ignore: The Business Case for Big Data*. Hoboken: John Wiley & Sons, Inc.
- Turing, A. M. (1950). *Computing Machinery and Intelligence*. Oxford: Oxford University Press.
- vom Brocke, J., Simons, A., Niehaves, B., Reimer, K., Plattfaut, R., & Cleven, A. (2009). Reconstructing the giant: On the importance of

rigour in documenting the literature search process. *European Conference on Information Systems*.

Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, xiii - xxiii.

Yi, N., Li, C., Feng, X., & Shi, M. (2018). Research and Improvement of Convolutional Neural Network. *Proceedings - 17th IEEE/ACIS International Conference on Computer and Information Science, ICIS 2018*, (pp. 637-640). Singapur.

KONTAKT

Christian Lossos
Schwarz Dienstleistung KG
Stiftsbergstraße 1
74172 Neckarsulm
Christian.Lossos@gmail.com

Frank S. Morelli
Hochschule Pforzheim
Tiefenbronner Straße 65
75175 Pforzheim
Frank.Morelli@hs-pforzheim.de

Simon Geschwill
Schwarz Dienstleistung KG
Stiftsbergstraße 1
74172 Neckarsulm
Simon.Geschwill@mail.schwarz