

Don't fear Black-Clouds – Mechanismen künstlicher Intelligenz

Nina Eckertz, Øyvind Eide

A. Einleitung

Künstliche Intelligenz (KI) begegnet uns an zahlreichen Stellen im Alltag. Egal ob als Empfehlungsalgorithmus auf einer Social-Media-Plattform, in Form eines Saugroboters oder der Berechnung des schnellsten Weges zur Arbeit. Obwohl zumindest Teilaspekte der KI bereits vollständig in unserer alltäglichen Routine angekommen sind, ist nicht jeder positiv auf KI zu sprechen,¹ denn häufig ist es nicht direkt ersichtlich, wo, wie und welche Art von Daten verarbeitet werden oder wie genau die Methodik des KI-Systems funktioniert.

Dieses Phänomen der Nicht-Nachvollziehbarkeit über die Funktionalität des KI-Mechanismus kann mithilfe des Black-Box-Begriffs beschrieben werden. In der Systemtheorie bezeichnet „Black-Box“ ein nur durch das äußere Verhalten bekanntes System. So wird eine Black-Box allgemein als ein technischer Begriff ohne starke negative oder positive Konnotationen verstanden. Im Kontext des KI-Diskurses wird eine Black-Box jedoch oft als ein Begriff mit negativen Konnotationen empfunden. Die Kernfrage hinter der Black-Box im Rahmen von künstlicher Intelligenz lautet: Wie trifft eine künstliche Intelligenz eine Entscheidung? Um sich diesem Phänomen anzunähern, beschäftigt sich dieser Beitrag mit dem Black-Box-Begriff und der Nachvollziehbarkeit von KI in verschiedenen Paradigmen. Ziel dieses Beitrags ist es, einen Teil zur Aufklärung über KI beizutragen und zur Diskussion anzuregen. Dazu werden die Hintergründe des Black-Box-Begriffs sowohl mithilfe eines historischen Beispiels als auch den biologischen Vorbildern und deren informationstechnischer Umsetzung betrachtet. Hierzu werden zusätzlich aktuelle Modellierungen innerhalb der KI-Forschung geisteswissenschaftlich hinterfragt.

Hieraus ergibt sich, dass es viele unterschiedliche Faktoren gibt, die das Verständnis von KI-Systemen beeinflussen und deshalb der statische

1 TÜV-Verband e.V. (Hrsg.), Sicherheit und Künstliche Intelligenz, Erwartungen, Hoffnungen, Risiken. August 2021, S. 5.

Begriff der Black-Box unzureichend ist, um das Nicht-Nachvollziehbarkeit-Phänomen zu beschreiben. Diese unterschiedlichen Faktoren umfassen das persönliche Verständnis von KI-Systemen im alltäglichen Umgang, das gezielte *overselling* und *overbuying* von KI-Systemen und die technische Komplexität von KI-Systemen, angefangen bei der Modellierung von künstlicher Intelligenz bis hin zur praktischen Umsetzung.

Letztendlich kommen wir zu dem Schluss, dass jeder ein individuelles Verständnis von KI hat und sich KI anders erklärt. Aus diesem Grund möchten wir mit dem Black-Cloud-Begriff einen flexibleren Terminus für die Erfassung des Verständnisses von KI-Systemen vorschlagen. Die Idee hinter der Black-Cloud ist dabei, dass das Verständnis von KI sich nicht fest in einer Box messen lässt, sondern sich wie eine Wolke über die Zeit hinweg verändert. Dieser neue Begriff soll daher sowohl die individuelle Perspektive als auch zukünftige Entwicklungen abdecken und mehr Flexibilität bieten.

B. Black-Box-Begriff – Eine historisch basierte Einführung

Dieser Beitrag bezieht sich mit dem Begriff der „Künstlichen Intelligenz“ grundlegend auf eine Forschungsdisziplin innerhalb der Informatik, in welcher man die (menschliche) Intelligenz mithilfe verschiedener Technologien, z.B. komplexer Datenverarbeitung oder der Robotik, zu imitieren versucht.

KI bietet eine Vielzahl an Unterkategorien, wie bspw. *maschinelles Lernen* (engl. machine learning) sowie *computer vision* oder *natural language processing*, um die KI-Anwendung in den Kontext eines bestimmten Input-Mediums zu setzen. Diese Technologien werden dabei gerne im Kontext der Data Science betrachtet und benutzt. Je nachdem wie vielfältig die Fähigkeiten eines KI-Systems sind, kann zwischen einer schwachen und einer starken KI unterschieden werden.

Gemäß Russell und Norvig ist KI historisch betrachtet ein recht junges Feld innerhalb der Wissenschaft, denn die intensive Beschäftigung mit der KI startete erst kurz nach dem 2. Weltkrieg.² Heute, rund 70 Jahre später, ist es noch nicht gelungen, (menschliche) Intelligenz vollständig zu modellieren und technisch nachzubilden. Es ist also noch viel Platz für weitere „Edisons“ und „Einsteins“ innerhalb der KI-Forschung.³ Im Folgenden geben wir einen

2 S. Russell/P. Norvig, Artificial intelligence: a modern approach. Third, global edition, Boston: Pearson 2016, S. 16.

3 Russell/Norvig, Artificial intelligence (Fn. 2), S. 1.

Einblick in den aktuellen Diskurs der KI-Forschung und erklären das Phänomen Black-Box etwas detaillierter. Hierzu wird zunächst die Wahrnehmung von KI beleuchtet.

I. Wahrnehmung von KI – Black-Box-Hintergründe

Wie eingangs erwähnt, beziehen wir uns mit dem Begriff der künstlichen Intelligenz zunächst auf die Definition aus der Informatik. Im öffentlichen Diskurs ist KI jedoch weit mehr als eine reine Forschungsdisziplin. Anna Visvizi betont zum Beispiel, dass gerade im öffentlichen Diskurs KI sowohl als unendliche Chance als auch als Quelle für Risiko betrachtet werden würde.⁴ Man diskutiert hier oftmals nicht nur den aktuellen Forschungsstand und aktuelle Produkte, sondern auch die kurz- und langfristigen Folgen der KI-Forschung. Das liegt daran, dass in dieser Diskussion viele unterschiedliche Parteien teilnehmen, seien es Forschende, Politiker oder Wirtschaftsvertreter. Zusätzlich sind viele Bereiche betroffen, zum Beispiel die öffentliche Verwaltung, die Justiz oder das Militär. Ein wichtiger Aspekt rund um den KI-Diskurs ist daher nicht nur, in welchem thematischen Kontext KI diskutiert wird, sondern auch, wer an der Diskussion teilnimmt, wie viel (Vor-)Wissen man individuell mitbringt und wie voreingenommen, also *biased*, man ist.

Fortuna und Gorbaniuk schlagen daher beispielsweise in ihrer Studie zur Untersuchung, was Diskursteilnehmer mit dem „buzzword“ KI meinen, zwei Unterscheidungsmöglichkeiten der Sichtbarkeit von KI und von Nutzergruppen vor.⁵ Dies kann helfen, die Diskutanten erst mal grundsätzlich einzuordnen. Zum einen gibt es bei Fortuna und Gorbaniuk die unsichtbare KI, z.B. Verarbeitungsalgorithmen auf Social-Media-Plattformen, und die sichtbare KI, z.B. Sprachassistenten.⁶ Zum anderen bieten sie die Unterscheidung in IT-Experten und Laien an. Die Unterscheidung ergibt sich hierbei

4 A. Visvizi/M. Bodziany (Hrsg.), *Artificial Intelligence and Its Contexts: Security, Business and Governance. Advanced Sciences and Technologies for Security Applications*. Cham: Springer International Publishing, 2021, S. 13 f.; <https://link.springer.com/10.1007/978-3-030-88972-2> (besucht am 30.09.2022).

5 P. Fortuna/O. Gorbaniuk. What Is Behind the Buzzword for Experts and Laymen: Representation of „Artificial Intelligence“ in the IT-Professionals' and Non-Professionals' Minds, in: *Europe's Journal of Psychology* 18 (2022), 207 (209).

6 Fortuna/Gorbaniuk, *Buzzword* (Fn. 5), 207 (208 f.).

aus der beruflichen Qualifikation.⁷ Hierzu wurde auch manuell überprüft, wie viel Wissen die Studienteilnehmer über KI haben, mit dem Ergebnis, dass die IT-Experten deutlich mehr über KI wissen.⁸ Die Forscher kommen zu dem Schluss, dass es durchaus Unterschiede gibt, was mit KI gemeint ist. Zum einen ist mit KI bei Laien eher eine Imitation gemeint, zum anderen treffen die IT-Experten mehr Unterscheidungen von KI, z.B. nach Art von KI-Systemen und daher nach unterschiedlichen Anwendungsfällen.⁹ Das erweckt den Eindruck, dass es unterschiedliche Auffassungen des KI-Begriffs gibt.

Die Limitation dieser Studie liegt in ihrer Vereinfachung des bzw. der sichtbaren/unsichtbaren KI, aber auch in der Einordnung der Studienteilnehmer.¹⁰ Denn KI aus Perspektive der IT-Branche ist wie jedes Software-Projekt eine Teamleistung. Somit bietet die IT-Branche innerhalb der KI-Entwicklung nicht nur viele unterschiedliche Rollen, sondern auch vielfältige Aufgaben innerhalb dieser an. Eine feinkörnige Unterscheidung ist daher notwendig, um z.B. technische Rollen von semi- oder gar nicht-technischen Rollen abzugrenzen, denn diese bieten durchaus andere Perspektiven und anderes Vorwissen in Bezug auf den KI-Diskurs an. So macht es einen Unterschied, ob man sich mit einem Product Owner oder einem reinen Softwareentwickler austauscht, abhängig von deren Hintergründen.

Schaut man zunächst in die popkulturelle Darstellung und deren Narrative von künstlicher Intelligenz wie etwa der Androidin Ava in *Ex Machina*, dem *Terminator* oder Arisa in *Better Than Us*, zeigen viele Science Fiction-basierte Filme uns ein Bild auf, welches aus Sicht der heutigen Datenverarbeitung noch weit in der Zukunft oder eventuell gar nicht zu erreichen ist. So zeigen die genannten Exempel bspw. nur in sehr begrenzten Umfang den Schöpfungsprozess der KI auf und erklären nicht die technische Zusammensetzung ihrer KI.

Ein ähnlicher Ansatz ist in der Vermarktung von KI-Produkten zu beobachten. KI als Marketingbegriff und Buzzword ist bereits seit einigen Jahren in aller Munde, man kann hier durchaus von einem „Overselling“ sprechen. Ähnlich wie bei den Beispielen aus der Popkultur wird hier der KI-Begriff im Hype als „nie“ neue Technologie besprochen, wohingegen bspw. die Subkategorien und eigentlichen Innovationstreiber wie maschine-

7 Fortuna/Gorbaniuk, Buzzword (Fn. 5), 207 (209).

8 Fortuna/Gorbaniuk, Buzzword (Fn. 5), 207 (211).

9 Fortuna/Gorbaniuk, Buzzword (Fn. 5), 207 (216).

10 Fortuna/Gorbaniuk, Buzzword (Fn. 5), 207 (216).

lles Lernen nicht besprochen werden.¹¹ Das führt dazu, dass eine Lücke zwischen dem öffentlichen Bild von KI und den eigentlichen Fähigkeiten aktueller KI entsteht, wodurch Unternehmen bspw. ihren Einsatz von KI überdenken und ggf. davon absehen.¹²

Hierbei ist zu beachten, dass, aus wirtschaftlicher Sicht betrachtet, die zugrundeliegenden Technologien aufgrund des Betriebsgeheimnisses nicht genauer erklärt werden müssen. Auf rechtlicher Seite ist es gemäß Bibal et al. so, dass KI nur im öffentlichen Sektor eine Erklärung ihrer Entscheidungen bedarf, im privaten Sektor, egal ob Business-to-Business oder Business-to-Customer nur dann, wenn es ein spezifisches Gesetz dazu gibt.¹³ Das ist nicht nur im Sinne der Einordnung von „sichtbaren“ KI-Produkten (nach Definition von Fortuna/Gorbaniuk¹⁴) relevant, sondern gerade von unsichtbaren Produkten zum Personalisieren eines Produktes wie Empfehlungsalgorithmen auf sozialen Netzwerken. Insofern ist es nicht verwunderlich, wenn auf den ersten Blick intransparent ist, wie sich die KI hier zusammensetzt und wie sie ihre Entscheidungen trifft. Zusätzlich gibt es unterschiedliche Abstufungen für Transparenz und Erklärungsbedarf je nach Algorithmik-System, also z.B. ob es sich um ein künstliches neuronales Netz oder einen Entscheidungsbaum (engl. *decision tree*) handelt.¹⁵

II. Historisches Beispiel von KI-Overselling

Der Ansatz zur Übervermarktung von KI ist nicht neu. Eines der frühesten Experimente für Mustererkennung und regelbasierte Antwort-Systeme ist Eliza. Eliza wurde im Jahr 1966 von Joseph Weizenbaum entwickelt und gilt als erster richtiger Chatbot, als Wegbereiter für den Bereich *natural language understanding* und hatte damit massiven Einfluss auf den gesamten NLP-Bereich.

11 C. Weeks, Machine Learning: The Real Buzzword Of 2020. Forbes, 24. Mai 2020, <https://www.forbes.com/sites/forbestechcouncil/2020/03/24/machine-learning-the-real-buzzword-of-2020/?sh=lec990b63ea7> (besucht am 30.09.2022).

12 T. Fountaine/B. McCarthy/T. Saleh, Building the AI-Powered Organization. Harvard Business Magazine, Aug. 2019, <https://hbr.org/2019/07/building-the-ai-powered-organization> (besucht am 30.09.2022).

13 A. Bibal et al., Legal requirements on explainability in machine learning, in: Artificial Intelligence and Law 29.2 (Juni 2021), 149 (151); <https://link.springer.com/10.1007/s10506-020-09270-4> (besucht am 30.09.2022).

14 Fortuna/Gorbaniuk, Buzzword (Fn. 5), 207 (216).

15 Bibal et al., Explainability, (Fn. 13), 149 (159).

Die Grundfunktionalität von Eliza als virtuelle Therapeutin ist sehr ähnlich zu dem, was man noch heute in Chatbots findet. Gemäß Weizenbaum wird beim Eliza-Bot der Nutzerinput auf bestimmte Keywords untersucht, welche dann anhand von bestimmten Regeln („decomposition rules“) eine möglichst passende Antwort von Eliza hervorrufen.¹⁶ Hierzu hat Weizenbaum zahlreiche mögliche Konversationsverläufe modelliert und die Sprache der Antwort auf die eines echten Therapeuten angepasst (s. Abbildung 1). Für den Endnutzer bedeutet das, dass er seine Fragen und Antworten an Eliza selbst in einer bestimmten Art und Weise formulieren muss, damit Eliza versteht, was gemeint ist.

Ein solches regelbasiertes System bietet einige Herausforderungen. Weizenbaum nennt hierbei sowohl die Erkennung von Keywords und Kontext als auch die Weiterentwicklung des Programms an sich.¹⁷ Somit ist Eliza sehr inflexibel, da sie nur bestimmte Formulierungen erkennen und nicht selbstständig neue Situationen erfassen und verstehen kann. Insofern handelt es sich bei Eliza um eine schwache KI, wenn man sie denn überhaupt „intelligent“ nennen kann.

“[...]

It's true. I am unhappy.

DO YOU THINK COMING HERE WILL HELP YOU NOT TO BE UNHAPPY

I need some help, that much seems certain.

WHAT WOULD IT MEAN TO YOU IF YOU GOT SOME HELP

Perhaps I could learn to get along with my mother.

TELL ME MORE ABOUT YOUR FAMILY

My mother takes care of me.

WHO ELSE IN YOUR FAMILY TAKES CARE OF YOU

My father.

YOUR FATHER

You are like my father in some ways.

WHAT RESEMBLANCE DO YOU SEE

You are not very aggressive but I think you don't want me to notice that.

WHAT MAKES YOU THINK I AM NOT VERY AGGRESSIVE

You don't argue with me.

16 Joseph Weizenbaum, ELIZA – a computer program for the study of natural language communication between man and machine, in: Communications of the ACM 9.1 (1966), 36 (38), <https://dl.acm.org/doi/10.1145/365153.365168> (besucht am 29.09.2022).

17 Weizenbaum, ELIZA (Fn. 16).

WHY DO YOU THINK I DON'T ARGUE WITH YOU
You are afraid of me.
DOES IT PLEASE YOU TO BELIEVE I AM AFRAID OF YOU
My farther is afraid of everybody.
WHAT ELSE COMES TO MIND WHEN YOU THINK OF YOUR FATHER
Bullies.
DOES THAT HAVE ANYTHING TO DO WITH THE FACT THAT YOUR
BOYFRIEND MADE YOU COME HERE
[...]"

Abbildung 1: Ausschnitt einer Konversation mit Eliza. Regulärer Text: Nutzerinput. Text in Großbuchstaben: Antworten von Eliza.¹⁸

Trotzdem konnte Eliza einige Nutzer von sich überzeugen, insbesondere diejenigen, die dazu instruiert waren, mit Eliza wie mit einem echten Therapeuten zu interagieren.¹⁹ Die Diskussion um diese „Illusion“²⁰ einer Konversation mit einem echten Menschen ist bis heute noch relevant. Zum einen, weil es mit Modellen wie etwa dem Turing-Test bereits Ansätze gibt, welche aktiv hinterfragen, ob eine Maschine selbstständig denken kann.²¹ Zum anderen, weil es mit dem LaMDA-Modell von Google einen aktuellen Fall gibt, bei dem das Bewusstsein von KI diskutiert wurde.

Der Turing-Test wurde im Jahr 1950 vom britischen Logiker, Mathematiker, Kryptoanalytiker und Informatiker Alan Turing entwickelt.²² Im Turing-Test wird überprüft, ob eine Maschine Menschen davon überzeugen kann, dass sie selbst auch ein Mensch ist. Hierzu gibt es drei Testteilnehmer: den *interrogator* (menschlich), die Maschine und einen weiteren Menschen. Der *interrogator* stellt der Maschine und dem Menschen Fragen und bewertet abschließend, welcher Testteilnehmer der Mensch und welcher die Maschine ist. Hierzu muss das KI-System zahlreiche Fähigkeiten aufweisen, wie etwa das *machine learning*, um selbstständig neue Konzepte im Gespräch

18 Joseph Weizenbaum, ELIZA – a computer program for the study of natural language communication between man and machine, in: Communications of the ACM 9.1 (1966), 37 (39), <https://dl.acm.org/doi/10.1145/365153.365168> (besucht am 29.09.2022).

19 Weizenbaum, ELIZA (Fn. 16), 36 (42).

20 Weizenbaum, ELIZA (Fn. 16), 36 (42).

21 G. Oppy/D. Dowe, The Turing Test, Stanford Encyclopedia of Philosophy Archive, 4. Okt. 2021, <https://plato.stanford.edu/archives/win2021/entries/turing-test> (besucht am 30.09.2022).

22 A. Turing, Computing Machinery and Intelligence, in: Mind LIX.236 (Okt. 1950), S. 433 ff.

zu lernen, oder das *natural language processing*, um zu verstehen, was der Gesprächspartner kommuniziert.²³ Ursprünglich als *imitation game* gestartet, wird er bis heute dazu genutzt, um KI-Systeme zu testen und zu bewerten, zum Beispiel beim Loebner Prize.²⁴ Der Turing-Test ist bereits zur Zeit seiner Veröffentlichung mit sechs Einsprüchen (*objectives*) kritisch hinterfragt worden,²⁵ da der Test nicht einwandfrei beweisen kann, dass eine Maschine wirklich denken kann – oder aber dass ein Mensch nicht denkt. Es geht beim Turing-Test vor allem darum, den *interrogator* zu überzeugen. Hierzu stellt sich die Frage, ob dies mit heutigen KI-Systemen und den neuen Sprachmodellen besser funktioniert als mit den logikbasierten Systemen aus der Zeit Turings wie etwa Eliza. Ein gutes Beispiel ist hierfür der Diskurs rund um das LaMDA-Modell von Google. Hierzu schwappte erst kürzlich eine breite Diskussion um den „Eliza Effekt“ bei der Google KI LaMDA hoch,²⁶ bei welcher der Softwareentwickler Blake Lemoine in zahlreichen Blogartikeln berichtete, dass LaMDA ein echtes Bewusstsein besäße.²⁷

Bei LaMDA handelt es sich um ein sehr großes NLP-Modell, welches in Chatbots eingebaut werden kann, ähnlich wie zum Beispiel das BERT-Modell. Der Unterschied zwischen Modellen wie BERT und LaMDA ist, dass LaMDA vor allem auf Grundlage menschlicher Dialoge trainiert wurde.²⁸ Dadurch kann es in einem Chatbot verbaut, viel „menschlicher“ auf Input reagieren und viel eher einen echten Menschen imitieren als Eliza, denn LaMDA kann z.B. spezifische Antworten auf Fragen und auch in der Tonalität Nuancen erzeugen,²⁹ die LaMDAs Antworten überzeugend wirken lassen. Bei Eliza hingegen kommt die Illusion vor allem dadurch zustande, dass Weizenbaum aufgrund seiner persönlichen Erfahrung wusste, wie ein Therapeut

23 Russell/Norvig, Artificial intelligence (Fn. 2), S. 2.

24 D. Powers, The total Turing test and the Loebner prize, in: Proceedings of the Joint Conferences on New Methods in Language Processing and Computational Natural Language Learning – NeMLaP3/CoNLL '98, Sydney, Australia: Association for Computational Linguistics, 1998, 279 (279).

25 Oppy/Dove, Turing Test (Fn. 20).

26 J. Felton, The Eliza Effect: How A Chatbot Convinced People It Was Real Way Back In The 1960s. 22. Juni 2022, <https://www.iflscience.com/the-eliza-effect-how-a-chatbot-convinced-people-it-was-real-way-back-in-the-1960s-64155> (besucht am 20.09.2022).

27 C. Vallance, Google engineer says Lamda AI system may have its own feelings, in: BBC News (13. Juni 2022), <https://www.bbc.com/news/technology-61784011> (besucht am 21.09.2022).

28 E. Collins/Z. Ghahramani, LaMDA: our breakthrough conversation technology, <https://blog.google/technology/ai/lamda> (besucht am 21.09.2022).

29 Collins/Ghahramani, LaMDA (Fn. 27).

mit seinen Patienten interagiert – und wie erwähnt funktionierte dies vor allem nur dann gut, wenn die Nutzer dazu instruiert waren, sich wie ein Patient zu verhalten. LaMDA ist aufgrund seiner großen Trainingsdatensets viel flexibler.³⁰

Insofern liegt der wesentliche Unterschied zwischen diesen beiden Modellen bzw. Chatbots vor allem in ihrer technischen Entwicklung. Während bei Eliza viele der Regeln und Dialoge von Hand geschrieben und einprogrammiert wurden, haben große IT-Firmen wie Google ganz andere menschliche wie technische Möglichkeiten, um KI-Systeme zu entwickeln. Dadurch kann diese Illusion, also wieder eine Black-Box für den Endnutzer, entstehen, die die Dialoge für den Nutzer so überzeugend machen. Zudem kann man den Sachverhalt rund um LaMDA auch so betrachten, dass diese Berichte bewusst veröffentlicht wurden, um ggf. die Black-Box zu verstärken.

Dennoch bleibt LaMDA ein Modell und ist limitiert in seinen Fähigkeiten, in dem Sinne, wie die Entwickler es designen, genauso wie Eliza. Dabei scheint das Modell für den Nutzer mehr zu können als es eigentlich kann, es wird also erneut „oversellt“. Gleichzeitig muss man hier kritisch hinterfragen, ob nicht auch auf Seite der Rezipienten von LaMDA eine Art *overbuying* des KI-Systems gegeben hat. Bei Eliza gab es ein *overbuying*, d.h. das Gefühl von Menschlichkeit war stärker als Weizenbaum es gedacht hatte.³¹ Diese Beobachtung und der erwähnte Diskurs um LaMDA legen nahe, dass es dieses *overbuying* vielleicht auch bei LaMDA gegeben hat.

Aus diesem Grund sollte es nicht nur allgemein diesen Diskurs rund um Black-Boxen, Black-Clouds und die Illusion von KI geben, sondern es sollten auch sowohl Endnutzer, Software-Entwickler als auch Produktmanager aus verschiedenen Bereichen teilnehmen. Gleichzeitig braucht es, wie man am Beispiel vom Turing-Test gesehen hat, Metriken, um überprüfen und testen zu können, wie „bewusst“ KI-Systeme agieren können. Hierzu hat Google nach dem LaMDA-Diskurs angekündigt, an einer Reihe von neuen Tests zu arbeiten, die besser als der bisherige Turing-Test funktionieren sollen und sich dem Konzept der General Artificial Intelligence nähern sollen.³² Das bedeutet, dass hier nicht nur z.B. die Sprachfähigkeit eines KI-Systems getestet wird, sondern auch andere Fähigkeiten wie etwa das Schachspielen.

30 Vallance, Own Feelings (Fn. 26).

31 Weizenbaum, ELIZA (Fn. 16), 36 (42 f.).

32 B. Schwan, Künstliche Intelligenz und LaMDA: Warum Google einen neuen Turing-Test will, heise online, 21. Juni 2022, <https://www.heise.de/hintergrund/Kuenstliche-Intelligenz-Warum-Google-einen-neuen-Turing-Test-will-7145019.html> (besucht am 28.09.2022).

Insofern nähert man sich durch neue Innovation und Gegenentwürfe mit Testmodellen sowie deren Diskurs sehr stark an die Erforschung von KI und Wahrnehmung an. Gleichzeitig ist dies immer noch ein sehr individuelles Unterfangen, wodurch sich viele verschiedene Perspektiven in diesem Diskurs ergeben. In diesem Sinne sollten dabei nicht nur die Datensätze und Grundfunktionalitäten diskutiert werden, sondern auch die Wirkung eines KI-Systems auf den Nutzer. Wir möchten in dem folgenden Kapitel unseren Beitrag zu der Diskussion vertiefen, warum es nicht nur für den Endnutzer, sondern auch für manche IT-Experten KI-Systeme Black-Boxen sind. Im folgenden Kapitel wird dies anhand des Trends des künstlichen neuronalen Netzes und Agenten-Systems aufgezeigt, und deren biologische Vorbilder werden beleuchtet.

C. Technische Komplexität – Neuronale Netzwerke und Datenpipeline

Nachdem wir in den vorherigen Kapiteln einen Blick in die Hintergründe des Entstehens von Black-Boxen geworfen haben, beleuchten wir nun explizit die technische Seite von KI nach aktuellem Forschungsstand. KI kann nämlich nicht nur für Nutzer ohne KI-Hintergrund eine Black-Box und folgend eine Black-Cloud sein, sondern auch für IT-Experten. Dies liegt unserer Ansicht nach vor allem an der Modellierung hinter künstlicher Intelligenz und der (technischen) Komplexität der praktischen Umsetzung dieser Modelle. Dies wird in diesem Kapitel anhand des gegenwärtigen Trends des maschinellen Lernens beleuchtet.

I. Maschinelles Lernen – Grundbegriffe

Maschinelles Lernen (engl. *machine learning*, ML) bezeichnet die Fähigkeit eines KI-Systems, Muster in Daten zu erkennen und darauf basierende Vorhersagen zu treffen und Aktionen zu vollziehen. Grundsätzlich gibt es drei Paradigmen für das maschinelle Lernen:

1. das überwachte Lernen (engl. *supervised learning*),
2. das unüberwachte Lernen (engl. *unsupervised learning*) und
3. das verstärkende Lernen (engl. *reinforcement learning*).

Ein ML-System innerhalb eines der genannten Paradigmen ist dazu designt, ein bestimmtes Problem zu lösen. Dazu nutzt es weitere technische Kompo-

nenten wie z.B. *computer vision*, *natural language processing* oder die Robotik – je nachdem welche Form von Input vorliegt. Ein KI-System besteht also aus mehreren Komponenten, welche miteinander kombiniert und aufeinander angepasst werden müssen.

Typische Probleme und Sachverhalte, mit denen sich die heutigen KI-Systeme in den genannten Paradigmen auseinandersetzen, sind z.B. im überwachenden Lernen die Fähigkeit, eine bestimmte Tierart oder auch Menschen anhand von Trainingsdaten auf einem Bild zu erkennen (Klassifikation) oder die Kaufpreisentwicklung von einem Haus in einer bestimmten Gegend vorherzusagen (Regression). Darüber hinaus kann man mit dem verstärkenden Lernen einer KI etwa das Spielen von Schach beibringen.

Insofern sind die Fähigkeiten der derzeit gängigen ML-Systeme noch eingeschränkt bzw. „schwach“, da KI-Systeme nach heutigem Stand nur ein Problem oder eine eingeschränkte Anzahl an Problemen gleichzeitig lösen können. Eine vollständige Nachahmung einer menschlichen Intelligenz, also einer „starken“ KI, erscheint damit erst mal unwahrscheinlich.

Dennoch steckt hinter diesen vermeintlich „einfachen“ Systemen bereits einiges an Aufwand und interdisziplinärer Arbeit mit Ansätzen zu Fragen der Intentionalität und des Bewusstseins. Hierzu beleuchtet das folgende Kapitel die Umsetzung von KI als Modell von menschlicher und tierischer Intelligenz.

II. KI als Modell – Zwischen Intentionalität und Datenverarbeitung

Ein wichtiger Aspekt von aktuellen KI-Systemen ist die mathematische Erfassung und Modellierung von tierischer wie menschlicher Intelligenz. Die genannten ML-Paradigmen können mithilfe von künstlichen neuronalen Netzwerken (KNN, engl. *artificial neural network*) bearbeitet werden, einer direkten Modellierung der Informationsverarbeitung innerhalb des menschlichen wie tierischen Gehirns. Aufgrund der technologischen Hardware-Entwicklung, insb. von Grafikkarten, hat diese Form des maschinellen Lernens in den letzten 20 Jahren viel Aufmerksamkeit erhalten. Einige der dahinterliegenden Ideen, wie etwa die des künstlichen Neurons, werden jedoch schon seit den 1950er Jahren diskutiert.³³ KI wird hier betrachtet als ein Modell von biologischer Intelligenz (*modelling of* bzw. Modellierung von)

33 Russell/Norvig, Artificial intelligence (Fn. 2), S. 727 f.

und einer Art Versuch, Intelligenz mathematisch nachzubilden (*modelling for* bzw. Modellierung für).³⁴

Die KI-Forschung ist dabei nicht nur Modellierung, sondern mehr eine Studie und Reflexion unserer selbst, an der viele verschiedene Forscher aktiv teilnehmen, z.B. Biologen, Philosophen und Informatiker. Es ist also wichtig zu verstehen, was die Grundlagen der biologischen Existenz eines intelligenten Wesens sind, ob und wie man diese mathematisch erfassen und für einen Computer greifbar machen kann, sowie das KI-System und seine Nutzung kritisch im Sinne der Philosophie zu hinterfragen.

Um einer Maschine beizubringen, Muster zu erkennen und „selbstständig“ bzw. intentional zu lernen und zu handeln, hat man sich nicht nur intensiv mit der Weitergabe von Informationen bei Menschen und Tieren beschäftigt, sondern auch gezielt mit dem biologischen Lernprozess und der Interaktion zwischen einem Individuum und seiner Umwelt. Hieraus ergeben sich unterschiedliche Komplexitäten in den Anwendungsgebieten von KI-Systemen.

Während KI-Systeme vor allem im *supervised learning* „nur“ große Datenmengen verarbeiten und in ihrem Sichtfeld eingeschränkt sind – durch die Überwachung wird bereits das Ziel des Lernens vorgegeben – so bietet das Paradigma des *unsupervised learnings* und gerade das *reinforcement learning* viel Raum zum Diskurs rund um die Frage der Intentionalität.

Zur Frage der Intentionalität muss man zunächst über den Begriff „Intelligenz“ allgemein sprechen und definieren, was dieser eigentlich meint. Ähnlich wie bei den verschiedenen Paradigmen rund um das maschinelle Lernen gibt es auch beim Intelligenzbegriff verschiedene Unterkategorien, wie etwa soziale oder mathematische Intelligenz. Zudem stellt sich die Frage, ob sich die Intelligenz von Menschen und Tieren grundsätzlich unterscheidet oder ob dieselbe Art von Intelligenz vorliegt. Wie eingangs erwähnt, basieren aktuelle KI-Systeme vor allem auf der Modellierung menschlicher Intelligenz, daher ist hier durchaus Potenzial für weitere Forschung – und damit für weitere Black-Boxen und Black-Clouds. Bei dem Begriff der Intentionalität verhält es sich ähnlich. Hier kann man zwei Grundannahmen treffen:

1. die Intentionalität, welche einem Lebewesen oder Objekt zugeschrieben wird und
2. die Intentionalität, welche das Lebewesen oder Objekt eigentlich hat.

34 C. Geertz, *The interpretation of cultures: selected essays*. ACLS Humanities E-Book. New York: Basic Books, 1973, S. 93 f.

Viele Tierarten – so auch Menschen – besitzen Intentionalität. Maschinen hingegen wirken für viele Menschen so, als hätten sie Intentionalität (vgl. bspw. die LaMDA-Diskussion). Russell und Norvig stellen hierzu die Idee des „rationalen Agenten“ vor, der handelt, um das für sich bestmögliche Ergebnis zu erzielen.³⁵ Diese Idee kann eingeschränkt auch auf KI-Systeme übertragen werden, denn oftmals interagieren sie mit schwierigen Umgebungen.³⁶ Dabei werden sie etwas konkreter mit der Unterscheidung zwischen menschlichem und rationalem Handeln und Denken.³⁷

Auf der einen Seite liegt mit dem Zuschreiben von Rationalität bei Russell und Norvig bereits ein Zuschreiben von Intention vor, denn nach dem bestmöglichen Ergebnis zu handeln ist durchaus eine Intention. Dieser Gedanke ist vor allem im Paradigma des *reinforcement learning* interessant, da hier die Interaktion zwischen einem Individuum, einem Agenten und seiner Umgebung modelliert wird. Anhand von Belohnungen und Bestrafungen soll sich der Agent in seiner Umgebung zurechtfinden und ein Ziel erfüllen. Im Anwendungsfall in der Robotik, mit dem Roboter als Agenten, kann das Ziel sein, sich in einem Raum zu bewegen, ohne gegen ein Hindernis zu stoßen. Durch den Zusammenstoß kann der Roboter lernen, zukünftig Hindernisse besser zu erkennen und nicht mehr mit ihnen zu kollidieren. Hier ist die Intention also, zu lernen sich besser zu bewegen – ähnlich wie bei einem Kleinkind, was gerade das Laufen lernt.

Auf der anderen Seite kann man argumentieren, dass diese Rationalität eine Designentscheidung des Entwicklerteams einer KI ist. Der Grund dafür ist, dass auch diese KI als schwache KI nach wie vor in einem abgesteckten Szenario Anwendung findet. Wenn ein Entwicklerteam einer *reinforcement learning*-KI bspw. das Go spielen beibringt, so ist das vermeintlich rationale Ziel der KI das Go-Spiel gegen jeden Spieler zu gewinnen. Gleichzeitig ist dies aber auch das Ziel des Entwicklerteams, gemessen daran, dass die Sieg-Statistik von Systemen wie AlphaGo eine Metrik dafür ist, wie erfolgreich das AlphaGo-System ist und wie man es weiter verbessern kann.³⁸ Hier stellt sich nachfolgend die Frage, ob das KI-System die Entscheidung über den Spielzug aufgrund des statisch gesehen besten Wertes trifft oder ob es wirklich intentional und intelligent gehandelt hat. Maschinen können daher

35 Russell/Norvig, Artificial intelligence (Fn. 2), S. 4.

36 Russell/Norvig, Artificial intelligence (Fn. 2), S. 4 f.

37 Russell/Norvig, Artificial intelligence (Fn. 2), S. 2.

38 D. Silver et.al., Mastering the game of Go with deep neural networks and tree search, in: Nature 529.7587 (28.01.2016), 484 (484, 487); <https://www.nature.com/articles/nature16961> (besucht am 30.09.2022).

mit Menschen symbiotisch Intentionalität aufweisen, aber die Frage über die Maschinen-Intentionalität ist offen – genauso wie die der eigentlichen Intelligenz der Maschine.

Die Frage nach der Intention und Intelligenz von KI-Systemen ist also noch nicht komplett beantwortet. Hierzu wird sich voraussichtlich in der Forschung rund um die (menschliche) Intelligenz noch einiges tun. Es ist zu erwarten, dass sich die Forschung im Bereich der (menschlichen) Intelligenz noch weiter entwickeln wird. Dies ist ein weiterer wichtiger Grund für die Black Box und ein wichtiger Indikator für ihre weitere Entwicklung und Perspektiven. Im folgenden Kapitel wird das praktische Paradigma der KI beleuchtet und die konkrete technische Umsetzung von KI-Systemen analysiert.

III. KI-Projekte in der Praxis: Teamarbeit und technische Umsetzung

Ein weiterer Grund für die Komplexität von KI-Systemen und somit für die Black-Boxen und Black-Cloud ist die komplexe praktische Umsetzung von KI-Systemen.

KI-Systeme sind datenbasierte Projekte, welche durch die interdisziplinäre Zusammenarbeit von verschiedenen akademischen Feldern geprägt werden. Auch in der Praxis sind, ähnlich wie bei anderen IT-Projekten, meist mehrere Akteure mit unterschiedlichen Hintergründen beteiligt. Die genaue Größe und Zusammensetzung unterscheidet sich meist nach der Unternehmensgröße, es gibt kein perfektes Rezept zur Aufstellung eines KI-Teams. Stobierski schlägt beispielsweise eine Aufteilung in Datenwissenschaftler (eng. *data scientists*), Datenentwickler (engl. *data engineers*) und Datenanalysten (eng. *data analysts*) vor.³⁹ Diese Rollen haben unterschiedliche Aufgaben und Themenschwerpunkte. So kümmert sich der Datenentwickler vor allem um die Infrastruktur der Daten, z.B. um die Sicherung der Daten in Datenbanken, während der Datenwissenschaftler sich mehr mit dem Konzept und den Forschungsfragen zu den vorhandenen Daten auseinandersetzt. Darüber hinaus kann es noch weitere Akteure in dieser Art Projekt geben, wie etwa Projekt- und Produktmanager.

39 T. Stobierski, How to Structure Your Data Analytics Team. Harvard Business School Online. 9. März 2021, <https://online.hbs.edu/blog/post/analytics-team-structure> (besucht am 30.09.2022).

Aus diesen Positionen ergibt sich das folgende Schema zur praktischen Umsetzung datenbasierter Projekte:

1. Erfassen und Vorverarbeiten (engl. *preprocessing*) des Inputs,
2. Analysieren und Verarbeiten des Inputs mithilfe des KI-Systems,
3. Auswerten und Überprüfen des Ergebnisses des KI-Systems,
4. Verbesserung des KI-Systems und
5. Präsentation und Einbindung der Ergebnisse des KI-Systems z.B. in eine Präsentation oder in einen Chatbot.

Dieses Schema ist nicht als linearer Weg zu betrachten, sondern mehr als eine Art Zyklus, bei denen die Schritte ineinander fließen und auch wiederholt werden können, wenn sich z.B. Kundenanforderungen ändern oder man im Entwicklungsprozess neue interessante Erkenntnisse aus den vorliegenden Daten gewinnt. Daher ist für das Schema wichtig, dass die Intention und das Konzept hinter dieser Datenanalyse und praktischen Umsetzung sowie mögliche Forschungsfragen im praktischen Prozess genauso mitwachsen wie die eigentliche Analyse und Umsetzung.

Diese Arbeitsweise ist in der praktischen Umsetzung sehr agil und kann daher sehr gut in Kombination mit Frameworks zur Projektorganisation wie etwa Scrum oder Kanban umgesetzt werden. Gleichzeitig bedeutet dies aber auch mehr Komplexität innerhalb des Projekts durch den Mehraufwand an nötiger Kommunikation in Form von täglichen Meetings für Rücksprachen und offene Fragen.

IV. Technische Umsetzung von KI

Im Folgenden soll die technische Umsetzung eines KI-Systems grob anhand des zuvor erwähnten Schemas erläutert werden. Als konkretes Beispiel dient die Analyse von Social Media Daten.

1. Erfassung der Rohdaten

Grundsätzlich kann man sich nach einer ersten Definition des Konzepts oder einer Forschungsfrage die Frage stellen, ob es bereits einen passenden vorgefertigten Datensatz gibt, z.B. von der Plattform Kaggle oder der Google Data Search, oder ob man selbst manuell die Daten für das Projekt aggregieren muss. Im Fall der Analyse von Social-Media-Daten bietet sich das Crawling von sog. APIs an.

Einige soziale Netzwerke wie Twitter⁴⁰ oder Reddit⁴¹ haben eigene, öffentliche APIs, anhand denen man automatisiert auf bestimmte Daten der Plattformen zugreifen kann. Diese Plattformen bieten Metadaten zu einzelnen Beiträgen an, z.B. den Inhalt des Beitrags, den Autor oder den Veröffentlichungszeitpunkt. Darüber hinaus kann man Kommentare und deren Metadaten zu einem Post herunterladen. Diese Rohdaten können über einen bestimmten Zeitraum hinweg gemäß der Limitationen der API gesammelt und analysiert werden.

An dieser Stelle ist zu hinterfragen, ab welcher Größe der Datensatz für die Analyse relevant wird. Eine konkrete Antwort gibt es hierauf nicht, allerdings kann man schlussfolgern: Je größer und diverser ein Datensatz aufgestellt ist, desto besser. Im Sinne des Big Data-Gedankens leben die meisten Werkzeuge der Datenwissenschaft bzw. *data science* von Statistik – so auch das maschinelle Lernen. Daher macht es Sinn, so viele hochwertige Daten wie möglich zu sammeln. Dies ist auch notwendig, um genügend Daten für das Trainieren, Evaluieren und Testen eines KNNs zu haben, um die KNN-Performance korrekt beurteilen zu können.

2. Vorbereitung und Analyse der Daten

Hat man die notwendige Datenmenge zusammen, kann man bereits eine erste Analyse der Rohdaten vollziehen und die passende Analysemethode auswählen. Im Fall von Social Media-Daten kann dies z.B. Text sein. Die textuellen Social Media-Beiträge kann man mithilfe von bspw. Textklassifikation mit KNNs wie etwa BERT oder auch durch eine statische Analyse der Postinhalte mit *topic modelling* untersuchen. Hierfür müssen die Rohdatensätze etwas gesäubert und vorbereitet werden, z.B. durch die Lemmatisierung der Textinhalte aus den Social Media-Beiträgen.

Die Analyse von Daten mit künstlichen neuronalen Netzwerken ist dabei sehr komplex. Wie bei den Datensätzen hat man auch hier die Möglichkeit ein bereits vorgefertigtes Modell zu wählen, z.B. ImageNet für Bildklassifikation oder BERT für Textklassifikation, oder selbst ein KNN zu entwickeln.

40 Inc. Twitter. Twitter API Documentation | Docs | Twitter Developer Platform. 2022, <https://developer.twitter.com/en/docs/twitter-api> (besucht am 30.09.2022).

41 J. Wardle, API, 2022, <https://github.com/reddit-archive/reddit/wiki/API> (besucht am 30.09.2022).

V. Komplexität der praktischen Umsetzung von KNNs

Wie bereits in Kapitel unter C. II. dargelegt sind KNNs eine Modellierung der biologischen Informationsverarbeitung von Menschen und Tieren. In der Praxis kann dies über sog. *multilayer feed-forward neural networks* umgesetzt werden (s. Abbildung 2).

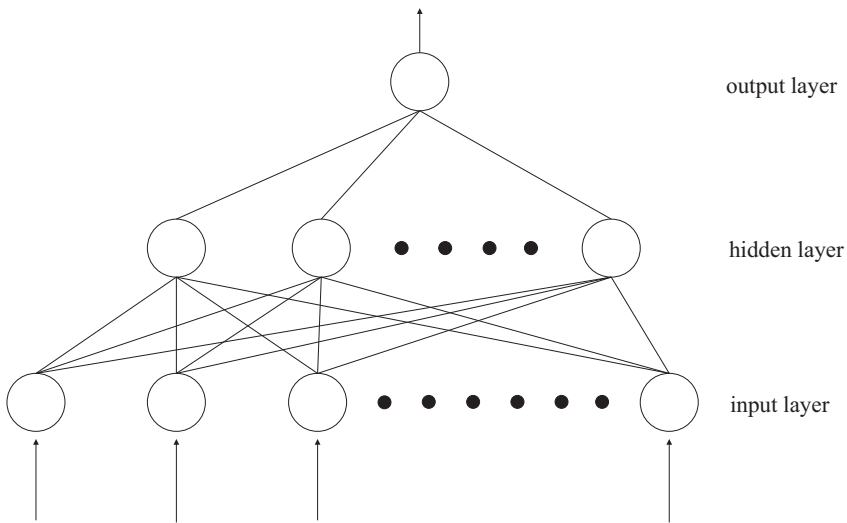


Abbildung 2: Schema eines multilayer feed-forward neural networks

Dieses Modell nimmt in den Input-Ebenen Input in Form von z.B. einem Text an und verarbeitet in versteckten Ebenen (engl. *hidden layers*) bestimmte Attribute dieses Inputs. In der *feature extraction* wird der Text-Input in der Input-Ebene in eine Form gebracht, mit welcher das KNN die Klassifikation durchführen kann. Hierzu wird mit Modellen wie etwa dem *bag-of-words* jeder Satz in einen Vektor umgewandelt, bei dem jede Stelle ein Wort impliziert. Durch den Abgleich und die Verrechnung der daraus generierten Features kann das KNN im Fall einer Textklassifikation dann vorhersagen, in welche Kategorie der Text gehört. Dies geschieht in den versteckten Ebenen. Die Output-Ebene führt die Ergebnisse dann in eine abschließende Vorhersage-Statistik zusammen.

Die künstlichen Neuronen des KNNs sind miteinander verbunden. Die Informationsverarbeitung läuft dabei linear von links nach rechts, durch die Verbindungen zwischen den Ebenen kann sich das KNN über das

Konzept der *backpropagation* im Trainingsvorgang selbst verbessern. Hier wird mithilfe einer Verlustfunktion bereits im Training das Zwischenergebnis der Klassifikation überprüft, und die Parameter des KNNs werden angepasst. Diese Parameter, z.B. die Gewichtung bestimmter Informationen, sind für die korrekte Klassifikation des Textes relevant. Ob eine Information anhand ihrer Gewichtung von einem künstlichen Neuron zum nächsten Neuron weitergegeben wird, ist davon abhängig, ob die Aktivierungsfunktion aktiviert werden kann. Das bedeutet, dass man initial als Entwickler die Gewichtungen und weitere Parameter wie die *learning rate* vorgibt, anhand derer das KNN sich sozusagen selbst verbessert und die Klassifikation berechnet. Die Berechnung der Anpassung der Gewichte des Netzes erfolgt selbstständig – man kann dies am Anfang als Entwickler einstellen und nach erfolgreichem Training nachjustieren. Darüber hinaus gibt es weitere Mechanismen wie etwa die Regularisierung (engl. *regularization*), die man einbauen kann, um das *overfitting* eines Modells, also das Auswendiglernen zu verhindern. Bei der Regularisierung werden z.B. einige Parameter des Modells bewusst weggelassen, um mehr Übersicht zu gewinnen. Wichtig ist zu betonen, dass all diese Bausteine je nach Problem unterschiedlich ausfallen können. So gibt es z.B. mehrere Arten von Aktivierungsfunktionen oder auch Verlustfunktionen. Sobald man die Komponenten richtig zusammengebaut und den Trainingsvorgang gestartet hat, kann man die Parameterweitergabe und Klassifikation nicht mehr beeinflussen bis man einen zweiten Trainingsvorgang startet. Insofern gibt es hier eine Black-Box by design, da man so das selbstständige Lernen des KNNs garantieren will.

Zusammenfassend ist die praktische Umsetzung von KI-Systemen ein komplexes und vielschichtiges Unterfangen, bei dem nicht umsonst viele verschiedene Akteure mit unterschiedlichen Aufgabengebieten in KI-Projekten teilnehmen. Neben der Black-Box by design bei KNNs als Analysemethode ist es ebenso wichtig, über die Erstellung der Datensätze und deren technische Umsetzung zu sprechen. Nur wenn ein umfassender und diverser Datensatz verfügbar ist, können qualitativ hochwertige Vorhersagen mit Black-Box-Systemen wie KNNs gemacht werden, wo ein Entwickler nicht in jeden Entscheidungsprozess des KNNs hineinschauen und nachjustieren kann. Erwähnt sei hier ebenfalls, dass es für diesen Diskurs ebenso wichtig ist, auch über die Präsentation und Visualisierung der Ergebnisse, z.B. in Form einer Konfusionsmatrix, zu sprechen, da Visualisierungen genauso Bias aufweisen und die Interpretation von KI beeinflussen können wie die eigentliche Analysemethode selbst.

D. Diskussion

Die vorangegangenen Kapitel haben gezeigt, wie vielfältig der Diskurs rund um KI-Systeme und ihre Funktionalität ist. Egal ob IT-Experte oder Laie, jeder bringt eine individuelle Perspektive in den Diskurs ein, abhängig von seinem Hintergrundwissen und (täglichen) Umgang mit KI-Systemen.

Um dieses individuelle Verständnis akkurat darzustellen, möchten wir nachfolgend die Black-Cloud-Formel vorstellen und einige Argumente dieses Aufsatzes kritisch im Ausblick hinterfragen.

I. Black-Cloud-Formel

Die vorliegende Formel soll das individuelle Verständnis einer Person von KI-Systemen mathematisch modellieren. Die Schwärze s jedes Punkts p in der Black-Cloud BC ist Resultat einer Funktion mit Vorkenntnissen (auch Kompetenz und Hintergrund) V und der technischen Komplexität K des KI-Systems als Parametern. D.h.

$$s_p = f(V, K)$$

und BC ist die Menge von allen p mit einem Wert höher als einem Grenzwert g .

$$p \in BC \Leftrightarrow s_p > g$$

Dadurch soll in diesem Modell die dynamische Entwicklung des Verständnisses dargestellt werden, durch welche die Black-Cloud flexibler ist und sich mit dem Untersuchungsobjekt verändert.

II. Ausblick

Nach einer ersten Beurteilung der Nutzer von KI-Systemen wäre es weiterhin interessant zu untersuchen, wie sich das Verständnis von KI-Systemen auf kurz- und langfristiger Basis verändert. Hierbei ist kritisch anzumerken, dass in einer Diskussion um KI-Systeme ebenfalls berücksichtigt werden sollte, ob sich die verschiedenen Nutzergruppen überhaupt für die Funktionalität von KI-Systemen interessieren oder ob sie diese Produkte ohne Hintergedanken nutzen wollen.

Außerdem ist zukünftig zu untersuchen, wie man die Funktionalität von KI-Systemen am besten erklärt. Wie im Kapitel *Technische Umsetzung von KI* aufgezeigt, besteht diese Art von Produkt aus mehreren Komponenten, an denen komplette Teams arbeiten. Es ist schwierig, dies alles innerhalb eines Beitrags zusammenzufassen – hier kann man noch viel mehr in die Tiefe gehen. Wir haben in dieser schematischen Erklärung beispielsweise viele mathematische Komponenten ausgelassen, die für manche Akteure des KI-Diskurses von größter Wichtigkeit sind.

E. Zusammenfassung

Zusammenfassend zeigt dieser Beitrag auf, wie wichtig und vielfältig der Diskurs rund um KI-Systeme ist. Die Diskussionen rund um das Eliza-, aber auch um das LamDA-Modell zeigen, wie viele verschiedene Perspektiven es gibt – und wie wichtig es ist, diese im Sinne des *overbuying* und *overselling* kritisch zu hinterfragen. Gleichzeitig gibt es bereits erste Untersuchungen, welche erfassen, wer wie viel über KI-Systeme weiß und auch wie man die Sichtbarkeit von KI-Systemen schematisieren kann.

Aus diesem Grund schlagen wir mit der Black-Cloud und ihrer Formel ein erstes Modell vor, um das individuelle Verständnis von KI-Systemen greifbar zu machen. Für die Black-Cloud ist dabei nicht nur das individuelle Vorwissen ein maßgeblicher Faktor, sondern auch die technische Komplexität. Die Diskussion rund um das *overselling* zeigt, dass aus juristischer Sicht in der Wirtschaft KI nicht erklärt werden muss. Dennoch ist KI, z.B. in Form von Machine Learning, eine komplexe Technologie, die ebenso für die Größe der eigenen Black-Cloud wichtig ist. Nicht nur, weil an der Entwicklung solcher Systeme viele verschiedene Akteure beteiligt sind und KI immer eine kollektive Teamleistung ist, sondern auch, weil mit der Modellierung biologischer Intelligenz manchen Systemen eine Art Intention zugeschrieben werden kann und so eine Black-Box by design entsteht. Gleichzeitig gibt es aber auch Argumente, die gegen eine Intentionalität von KI-Systemen sprechen. So können aktuelle KI-Systeme noch immer nur bestimmte Aufgaben erfüllen und sind inflexibel. Doch bereits hier stecken viele verschiedene Bausteine in solchen Systemen, wie das Beispiel von künstlichen neuronalen Netzwerken zeigt.

Es ist abzuwarten, wie sich diese Technologie bzw. dieses Forschungsfeld in den nächsten Jahren weiterentwickeln wird. Bis dahin bietet die Black-Cloud

eine Anregung zum Individualisieren des KI-Diskurses und zur Modellierung des Verständnisses von KI-Systemen.

