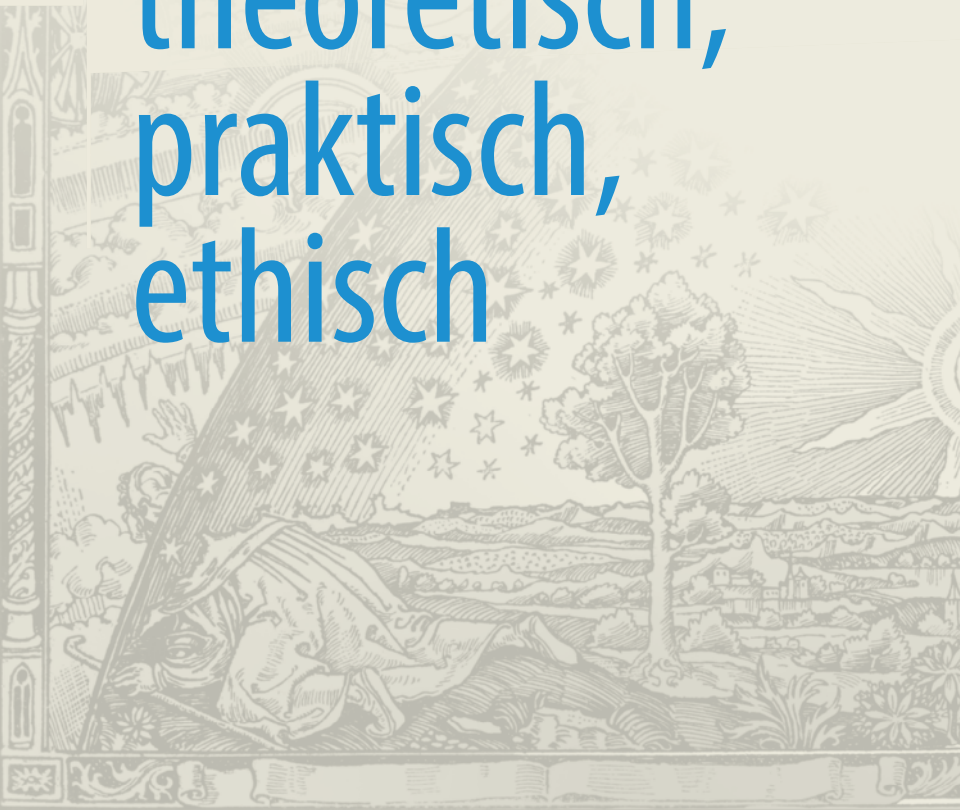


Klaus Mainzer
Reinhard Kahle

Grenzen der KI theoretisch, praktisch, ethisch



SACHBUCH



Springer

Technik im Fokus

Die Buchreihe Technik im Fokus bringt kompakte, gut verständliche Einführungen in ein aktuelles Technik-Thema.

Jedes Buch konzentriert sich auf die wesentlichen Grundlagen, die Anwendungen der Technologien anhand ausgewählter Beispiele und die absehbaren Trends.

Es bietet klare Übersichten, Daten und Fakten sowie gezielte Literaturhinweise für die weitergehende Lektüre.

Weitere Bände in der Reihe <https://link.springer.com/bookseries/8887>

Klaus Mainzer · Reinhard Kahle

Grenzen der KI – theoretisch, praktisch, ethisch

 Springer

Klaus Mainzer
TUM Senior Excellence Faculty
Technische Universität München
München, Deutschland

Reinhard Kahle
Carl Friedrich von Weizsäcker-
Zentrum, Universität Tübingen
Tübingen, Deutschland

ISSN 2194-0770

ISSN 2194-0789 (electronic)

Technik im Fokus

ISBN 978-3-662-65010-3

ISBN 978-3-662-65011-0 (eBook)

<https://doi.org/10.1007/978-3-662-65011-0>

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert an Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2022

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von allgemein beschreibenden Bezeichnungen, Marken, Unternehmensnamen etc. in diesem Werk bedeutet nicht, dass diese frei durch jedermann benutzt werden dürfen. Die Berechtigung zur Benutzung unterliegt, auch ohne gesonderten Hinweis hierzu, den Regeln des Markenrechts. Die Rechte des jeweiligen Zeicheninhabers sind zu beachten.

Der Verlag, die Autoren und die Herausgeber gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag, noch die Autoren oder die Herausgeber übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen. Der Verlag bleibt im Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutionsadressen neutral.

Planung/Lektorat: Michael Kottusch

Springer ist ein Imprint der eingetragenen Gesellschaft Springer-Verlag GmbH, DE und ist ein Teil von Springer Nature.

Die Anschrift der Gesellschaft ist: Heidelberger Platz 3, 14197 Berlin, Germany

Vorwort

Als der amerikanische Philosoph HUBERT DREYFUS 1972 einen Bestseller unter dem Titel „*Was Computer nicht können. Die Grenzen künstlicher Intelligenz*“ veröffentlichte, zeigte er tatsächlich nur die Grenzen dessen auf, was wir heute „symbolische KI“ nennen. Es handelte sich um sogenannte Expertensysteme, die beschränktes Spezialwissen von Experten wie z. B. Ärzte und Ingenieure in „regelbasierten Wissenssystemen“ abbildeten. Dreyfus zeigte mit Recht die Grenzen dieses Ansatzes für intuitives Können auf: Die ersten Stunden des Fahrunterrichts lassen sich in Regeln vermitteln, die in Textbüchern festgehalten werden können. Dann beginnt aber intuitives Lernen und Trainieren, um ein wirklich guter Autofahrer zu werden. Wer je versucht hat, den Abschlag eines Golfballs durch Befolgen von Regeln zu perfektionieren, weiß sofort, was gemeint ist.

Nach dem Paradigma von Logik-basierten Regelsystemen trat daher in den 1970er Jahren das Trainieren von neuronalen Netzen, das sogenannte konnektionistische Paradigma, auf, das viele der Grenzziehungen von Dreyfus überwand. Der Philosoph gab daher einer späteren Auflage seines Buchs etwas kleinlaut den Titel „*Was Computer (noch) nicht können?*“.¹ Es zeigte

¹ Der englische Titel *What Computer Still Can't Do* läßt sich aber auch als *Was Computer noch immer nicht können* übersetzen – so in der Übersetzung des Vorworts des Buches, das in der Deutschen Zeitschrift für Philosophie veröffentlicht wurde.

sich wieder einmal, daß man mit apodiktischen Grenzziehungen vorsichtig sein sollte. Sie können sich eben nur auf bestimmte Domänen, Systeme, Wissensbestände und vorläufige Entwicklungsstände beziehen.

Selbst diese Grenzen haben aber die Praxis nur teilweise interessiert. Regelbasierte Expertensysteme sind nach wie vor bis heute hoch perfektioniert und erfolgreich in Industrie (z. B. Logistik der Automobilindustrie) und Medizin (z. B. Kontrollsysteme) im Einsatz, ohne daß wir sie als spektakuläre „KI“ empfinden. Grenzziehung bedeutet also nicht, daß Systeme überholt sind, sondern daß wir nur genauer wissen, was sie können und was sie nicht können.

Um so interessanter sind Grenzen, die sich möglicherweise aus Logik und Mathematik ergeben. In Logik und Mathematik gibt es eben Probleme, die „noch“ nicht gelöst oder entschieden sind. KI, die davon abhängt, hätte also nur vorläufige Grenzen. Spannend wird es, wenn es sich um prinzipiell nicht entscheidbare Probleme handelt. Was ist aber prinzipiell nicht entscheidbar? In diesem Fall stoßen sowohl natürliche als auch künstliche Intelligenz an prinzipielle Grenzen. Die Schlüsselfrage lautet aber: Wie findet die natürliche Intelligenz von Mathematikern Lösungen? Eine Analyse des von Menschen verwendeten mathematischen Hintergrundwissens läßt Zweifel aufkommen, ob KI dazu jemals in der Lage wäre. Prinzipiell ausschließen läßt es sich aber nicht.

Nun könnte man meinen, diese Art von Analysen sind so abstrakt, daß sie für die praktische Anwendung der KI keine Rolle spielen. Sollen sich doch einige Nerds in ihren Elfenbeintürmen damit auseinandersetzen! Die KI-Community wird in der Zwischenzeit viel Geld mit der „diesseitigen“ KI verdienen und Industrie und Gesellschaft aufmischen! Tatsächlich sind aber die scheinbar abstrakten mathematischen Fragen, die hier gemeint sind, mit z. B. Sicherheitsfragen der Kryptologie unmittelbar verbunden. Das gilt nicht erst dann, wenn Quantencomputer vorliegen! Aber ihre technische Möglichkeit zusammen mit der bereits realisierten Quantenkommunikation verleiht der Frage nach mathematischen Grenzen der KI noch einmal zusätzliche praktische Brisanz. Begeben wir uns also in die Elfenbeintürme

der Informatik, Mathematik und Philosophie, wohl wissend, daß wir erst damit die verborgenen Brüche der technischen Zivilisation wie unter einem Brennglas erkennen können.

Klaus Mainzer
Reinhard Kahle

Danksagung

Viele der im Text gegebenen Beispiele profitierten im Fall von Reinhard Kahle von Diskussionen mit Klaus Angerer (jetzt Darmstadt), Philipp Hennig (Tübingen), Kristian Kersting (Darmstadt), Christoph Peylo (Renningen), Thomas Sattig (Tübingen) und Elektra Wagenrad (Berlin). Klaus Mainzer profitierte von Diskussionen in der Steuerungsgruppe für eine KI-Normungsroadmap unter Leitung von Wolfgang Wahlster, im Themennetzwerk der Deutschen Akademie der Technikwissenschaften (acatech) und als Präsident der Europäischen Akademie der Wissenschaften und Künste von Diskussionen der dortigen Expertengruppe für Digitalisierung und KI. Mit dem Dank an diese Kolleginnen und Kollegen wollen wir aber nicht unterstellen, daß sie auch immer unsere Sicht der Problematik teilen würden.

Inhaltsverzeichnis

1	Zum Begriff der Künstlichen Intelligenz	1
	Literatur.	7
2	Praktische Grenzen	9
2.1	Das Schicksal der Expertensysteme.	9
2.2	Kausalität versus Statistik	18
2.3	Vom Bayesschen Lernen zu neuronalen Netzen	29
2.4	Datenmenge und Datenqualität	47
2.5	Die Rückkehr des „Rahmen-Problems“.	52
	Literatur.	54
3	Theoretische Grenzen	57
3.1	Läßt sich „Rechnen“ statistisch erlernen?	57
3.2	Kontinuierliche versus diskrete Fragestellungen	61
3.3	Welche Rolle spielt Zufall in der KI?	69
3.4	Welche Rolle spielt Chaos in der KI?	73
3.5	Gibt es eine Berechenbarkeits- und Komplexitätstheorie für die KI?	76
	Literatur.	82
4	Konzeptionelle Grenzen	85
4.1	Die Frage „Warum?“	85
4.2	Kann sich die KI „erinnern?“	87
4.3	Läßt sich Programmieren automatisieren?	89
4.4	Läßt sich Beweisen automatisieren?	92
4.5	„What you give is what you get?“	98

4.6	Hintergrundtheorie	102
4.7	Ethische und gesellschaftliche Grenzen der KI	105
	Literatur	117
5	Ausblicke auf Hybride KI	121
5.1	Möglichkeiten und Grenzen neuromorpher KI . . .	121
5.2	Möglichkeiten und Grenzen der Quanten-KI. . . .	131
5.3	Quo vadis KI?	138
	5.3.1 Eine optimistische Vision.	138
	5.3.2 Eine skeptische Sicht	140
	Literatur	146
	Namensverzeichnis	149
	Sachverzeichnis	151



Zum Begriff der Künstlichen Intelligenz

1

The term AI contains an explicit reference to the notion of intelligence. However, since intelligence (both in machines and in humans) is a vague concept, although it has been studied at length by psychologists, biologists, and neuroscientists, AI researchers use mostly the notion of rationality, which refers to the ability to choose the best action to take in order to achieve a certain goal, given certain criteria to be optimized and the available resources.

European Commission's High-Level Expert Group on Artificial Intelligence [1].

Effektive Problemlösungsverfahren sind seit der antiken Mathematik bekannt. In der Geometrie wird die Konstruktion einer Figur in elementare Konstruktionsschritte mit Zirkel und Lineal zerlegt. In Arithmetik und Algebra werden Lösungsverfahren für Gleichungen in elementare Rechenschritte zerlegt, die im Prinzip von einer Maschine ausgeführt werden können. Man spricht dann auch von Algorithmen, die nach dem persischen Mathematiker ALCHWARIZMI benannt werden. Heute werden Algorithmen durch Computerprogramme ausgeführt. Die Frage ist, bis zu welchem

Grad Lösungsschritte nicht nur von einer Maschine ausgeführt, sondern auch selbständig entwickelt werden können.

Künstliche Intelligenz (KI) wird daher an der menschlichen Intelligenz gemessen. Nach dem britischen Logiker und Computerpionier ALAN M. TURING [2] wird ein technisches System „intelligent“ genannt, wenn es in seinen Antworten und seinem Lösen von Problemen nicht von einem Menschen unterschieden werden kann. Ursprünglich orientierte sich KI an Regeln und Formeln der symbolischen Logik, die in passende Computerprogramme übersetzt wurden. Man spricht deshalb auch von symbolischer KI (Abb. 1.1). Dahinter steht die erkenntnistheoretische Vorstellung, daß Intelligenz vor allem mit der Fähigkeit logischen Schließens im menschlichen Verstand verbunden ist.

Ein Beispiel war das automatische Beweisen, bei dem KI-Programme das logische Schließen in Logikkalkülen simulierten [3]. Auf dieser regelbasierten und symbolischen Grundlage sollte auch menschliches Planen, Entscheiden und Problemlösen simuliert werden, wie es menschliche Experten in spezialisierten Anwendungsgebieten tagtäglich realisieren. In entsprechenden Expertensystemen bzw. wissensbasierten Systemen wird zunächst das spezifische Fachwissen eines z. B. Ingenieurs oder Arztes in formale Regeln übersetzt, die bei Eintreten eines bestimmten Ereignisses eine bestimmte Handlung automatisch auslösen sollen.

Ein Erbe dieses Ansatzes ist die Programmiersprache Prolog (vom Französischen: *programmation en logique*), die sich auch heute noch einer gewissen Beliebtheit erfreut, auch wenn sie effektiv nur im theoretischen Bereich benutzt wird und sich im industriellen Anwendungsbereich nicht durchsetzen konnte – aus Gründen, die mit den hier zu diskutierenden Grenzen der KI durchaus in Zusammenhang stehen. In Prolog können (einfache) Regeln formuliert werden, zum Beispiel um ein Netz von Flugverbindungen zu speichern.

Beispiel

```
erreichbar(X, Y) :- direktflug(X, Y).  
erreichbar(X, Y) :- direktflug(X, Z),  
erreichbar(Z, Y).
```

```
direktflug(LIS, FRA).  
direktflug(FRA, MUC).  
direktflug(FRA, STR).  
⋮
```

Prolog ist eine *Abfragesprache* bei der für das gegebene Beispiel die Frage

```
?- erreichbar(LIS, STR).
```

die Antwort Yes . liefern sollte. ◀

In dem Maße, in dem diese Art von Wissensrepräsentation weiter ausgebaut werden sollte, traten zunehmend Komplexitätsprobleme auf – in zwei verschiedenen Bedeutungen von „Komplexität“: Zum einen ist die *allgemeine Komplexität* zum Beispiel der Grammatik einer Sprache, häufig so groß, daß sich eine einfache Übersetzung in Prolog-Regeln als praktisch undurchführbar herausstellte. Zum anderen treten Probleme der *Berechenbarkeitskomplexität* auf, die zum Beispiel bei der Abfrage aller theoretisch möglichen Flugverbindungen – mathematisch der transitive Abschluß von $\text{erreichbar}(X, Y)$ – zu nicht mehr hinnehmbaren Rechenzeiten führen. Aufgrund dieser Probleme sind die Expertensysteme vergleichsweise schnell wieder aus der Mode gekommen.

Allerdings wäre es eine grobe Verkürzung, die Forschung in der Künstlichen Intelligenz in der zweiten Hälfte des 20. Jahrhunderts auf den Bereich der Expertensysteme zu beschränken. Aus den Überlegungen zum automatischen Beweisen hat sich insbesondere das *SAT-solving* (deutsch nur etwas unbeholfen als „Erfüllbarkeitslösen“ wiederzugeben) als eine Disziplin entwickelt, die heute weitreichende Anwendungen hat. Daneben wurden – moti-

viert durch die Erkenntnisse der neurologischen Hirnforschung – auch *neuronale Netze* als vereinfachende Computersimulationen des mit Hilfe von Neuronen beschriebenen menschlichen Nervensystems im Gehirn eingeführt. Dieser Ansatz war von Beginn an konzeptionell von den regelbasierten Systemen unterschieden, kam aber – nicht zuletzt durch die noch vergleichsweise begrenzte Speicher- und Rechenleistung der zur Verfügung stehenden Computer – kaum über „Spielzeuganwendungen“ hinaus. Er erlaubte zwar das Prinzip neuronaler Netze zu simulieren und zu studieren, ermöglichte aber noch kaum praktische Anwendungen.

In vereinfachender Form lassen sich damit die Forschungsgebiete der *klassischen* oder *alten KI*, wie sie sich zum Ende des 20. Jahrhunderts darbot, wie folgt zusammenfassen:

Klassische oder alte KI

- *Expertensysteme*
Prolog als paradigmatische Programmiersprache.
- *SAT-Solving*
Aussagenlogisches Problemlösungsverfahren, das komplexitätstheoretisch gerade noch handhabbare Fragestellungen behandeln kann.
- *Frühe neuronale Netze*
In der frühen Phase der KI nur von sehr beschränkter Komplexität.

Die frühen neuronalen Netze waren schon eine Reaktion darauf, daß regelbasiertes Wissen nie das intuitive Können eines Experten vollständig erfassen kann. Können beruht auf vielfältigen Erfahrungen, die nur schwer symbolisch in einem Lehrbuch repräsentiert werden können. Ein erfahrener Autofahrer erfaßt Situationen und reagiert intuitiv auf der Grundlage vieler sensorischer Daten, ohne sich der logischen Abläufe im Einzelnen bewußt zu sein. Ähnlich reagiert ein erfahrener Arzt in einer kritischen Situation oder ein

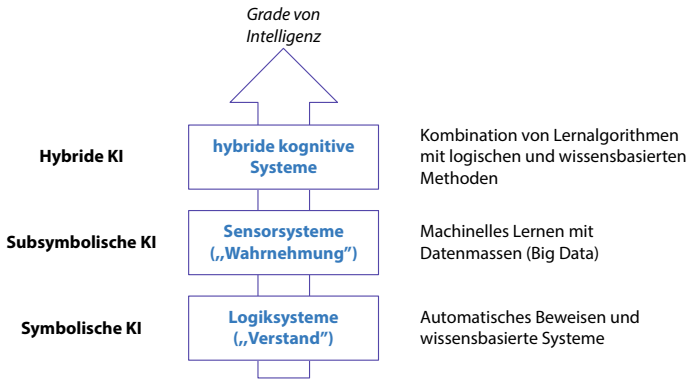


Abb. 1.1 Von der symbolischen und subsymbolischen zur hybriden KI

erfahrener Pilot im Cockpit eines Flugzeuges. Intuition ist keineswegs ein mystischer Zauberkasten. Vielmehr kann Erkennen von Datenmustern und Einschätzung von Erwartungswahrscheinlichkeiten durch Erfahrung trainiert und verbessert werden.

An die Stelle logischer Regeln, wie sie in der symbolischen KI verwendet werden, treten nun sensorielle Daten, in denen statistische Korrelationen und Wahrscheinlichkeiten bestimmt werden. Lernen aus Daten wird mathematisch in der statistischen Lerntheorie untersucht. Ihre Algorithmen liegen dem maschinellen Lernen zugrunde. Erkenntnistheoretisch laufen diese Lernprozesse aus sensoriiellen Daten der Wahrnehmung unbewußt unterhalb des bewußten logischen Schließens ab. Man spricht deshalb auch von subsymbolischer KI (Abb. 1.1). Mathematisch wird dabei das Paradigma der Logik durch Statistik und Wahrscheinlichkeitstheorie ersetzt. Die starke Rechnertechnologie der vergangenen Jahre hat es ermöglicht, daß das maschinelle Lernen mit großen Datenmassen nun auch technisch realisiert werden kann und zu neuen Durchbrüchen der KI-Anwendung z.B. bei der Entwicklung von Medikamenten und Impfstoffen führte.

Entsprechend hat sich im beginnenden 21. Jahrhundert eine *statistikbasierte* oder *neue KI* herausgebildet, deren Charakteristik sich wie folgt beschreiben läßt.