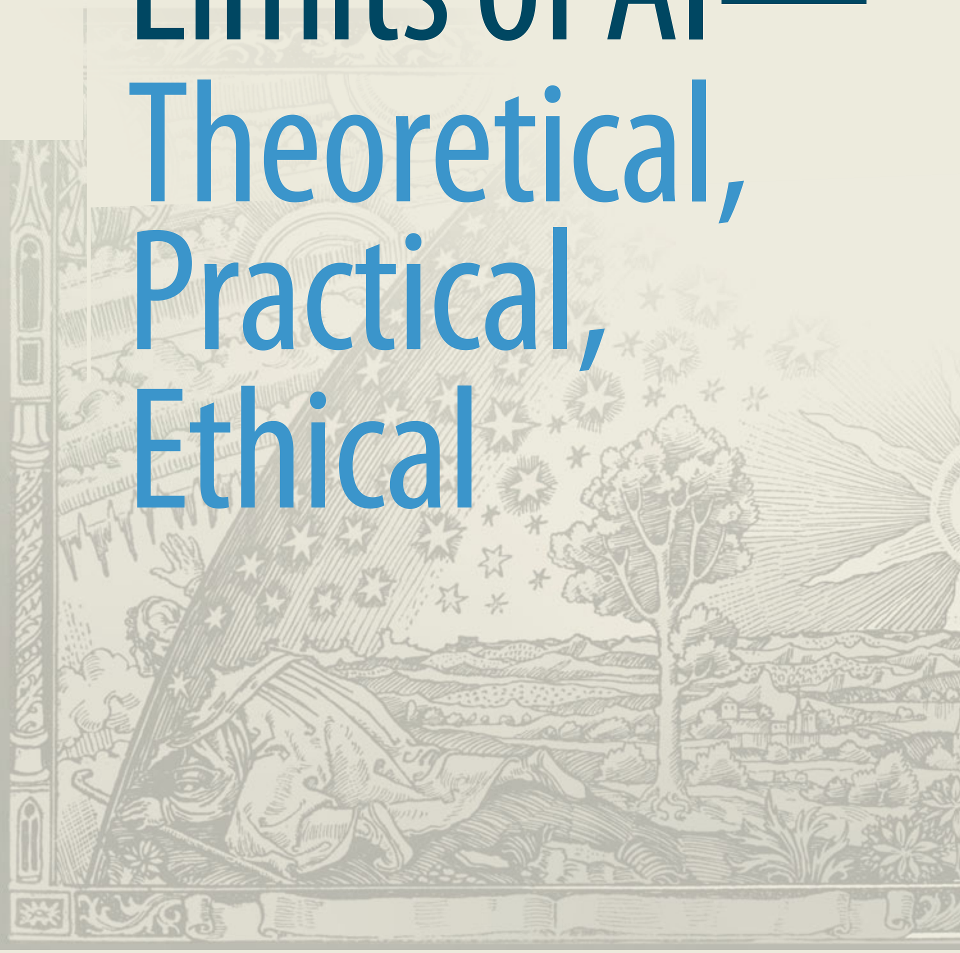
Klaus Mainzer
Reinhard Kahle

Limits of AI—Theoretical, Practical, Ethical



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.

Klaus Mainzer · Reinhard Kahle

Limits of AI—Theoretical, Practical, Ethical

Springer

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

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

ISSN 2194-0770 ISSN 2194-0789 (electronic)
Technik im Fokus
ISBN 978-3-662-68289-0 ISBN 978-3-662-68290-6 (eBook)
https://doi.org/10.1007/978-3-662-68290-6

© Springer-Verlag GmbH Germany, part of Springer Nature 2024

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer-Verlag GmbH, DE, part of Springer Nature.
The registered company address is: Heidelberger Platz 3, 14197 Berlin, Germany

Paper in this product is recyclable.

Preface (to the English translation)

The German version of this book appeared at the beginning of 2022 before the hype about the ChatCPT started. However, the limits of AI were already clearly shown there, also using the example of the ChatGPT and its precursors. Therefore, one cannot resist the remark that many of the commentators could have saved their dystopian or utopian remarks if they had read our book beforehand. Nevertheless, on the background of the analyses in this book, a summary about the ChatGPT has been added (Sect. 5.3). Although we utilized the translation software "DeepL," the authors bear responsibility for any errors. However, we retain the right to attribute any oversight of these errors to Artificial Intelligence for not alerting us. In many cases, the human authors corrected DeepL, because they have the better background knowledge and understand the universal language of mathematics.

Klaus Mainzer
Reinhard Kahle

Preface (to the first German edition)

In 1972, when the American philosopher Hubert Dreyfus published a bestseller entitled "What Computers Can't Do. The Limits of Artificial Intelligence", he was in fact only pointing out the limits of what we now call "symbolic AI". These were so-called expert systems, which combine the limited specialist knowledge of experts such as doctors and engineers in 'rule-based knowledge systems'. Dreyfus rightly pointed out the limits of this approach to intuitive knowing: The first hours of driving lessons can be taught in rules that are recorded in textbooks. But then intuitive learning begins and training is needed to become a really good driver. Anyone who has ever tried to perfect the stroke of a golf ball by following rules knows immediately what is meant.

After the paradigm of logic-based rule systems in the 1970s, the training of neural networks, the so-called connectionist paradigm, emerged. The connectionist paradigm overcame many of Dreyfus' limitations. The philosopher therefore somewhat meekly gave a later edition of his book the title "What Computers Still Can't Do".[1] Once again that one should be careful with apodictic demarcations. They can only apply to certain domains, systems, bodies of knowledge and preliminary stages of development.

[1] There is a certain ambiguity in the word "Still"; it could be understood as "not yet" or it come with the connotation "still, and never".

Even these boundaries, however, are only partially of interest. Still today, rule-based expert systems are highly elaborated and successfully applied in industry (e.g. logistics in the automotive industry) and medicine (e.g. control systems), without us perceiving them as spectacular "AI". The drawing of boundaries therefore does not mean that systems are outdated, but that we only know more precisely, what they can and cannot do.

Even more interesting are the limits that may arise from logic and mathematics. In logic and mathematics there exist problems which have not yet been solved or decided. Therefore, AI that depends on such problems will have only provisional limits. It is more interesting when we are dealing with problems that cannot be decided in principle. What is undecidable in principle? In this case, both natural and artificial intelligence reach their limits in principle. But the key question is: How does natural intelligence of mathematicians find solutions? An analysis of the mathematical background knowledge used by humans raises doubts as to whether AI would ever be able to do this. But it cannot be ruled out in principle.

Now one might think that these kinds of analyses are so abstract that they are irrelevant for the practical application of AI. Let some nerds in their ivory towers deal with it! In the meantime, the AI community will make a lot of money from "this side" AI and will shake up industry and society! But in fact the seemingly abstract mathematical questions we are referring, are directly connected with, for example, security issues in cryptography. This is not only when quantum computers are available! But their technical feasibility, together with the already implemented quantum communication, concerns the question of the mathematical limits of AI once again with additional explosiveness for practical applications. So let us enter the ivory towers of computer science, mathematics and philosophy, knowing very well that, only in this way, we will find the hidden dangers of technical civilisation as if under a magnifying glass.

Klaus Mainzer
Reinhard Kahle

Acknowledgements

Many of the examples given in the text benefited, in Reinhard Kahle's case, from discussions with Klaus Angerer (now Darmstadt), Philipp Hennig (Tübingen), Kristian Kersting (Darmstadt), Christoph Peylo (Renningen), Thomas Sattig (Tübingen) and Elektra Wagenrad (Berlin). Klaus Mainzer benefited from discussions in the steering group for a German Standardization Roadmap on Artificial Intelligence chaired by Wolfgang Wahlster, in the thematic network of the German National Academy of Science and Engineering (acatech) and as President of the European Academy of Sciences and Arts in discussions of the expert group on digitalisation and AI.

We are grateful to these colleagues, but we do not imply that they always share our view of the problems.

The second author was supported by the Udo Keller Foundation and the FCT – Fundação para a Ciência e a Tecnologia, I.P., under the scope of the projects UIDB/00297/2020 and UIDP/00297/2020 (Center for Mathematics and Applications)". The first author is chairman of the Board of Trustees of the Udo Keller Foundation.

Contents

1 The Concept of Artificial Intelligence

> 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].

Effective methods for problem-solving have been known since ancient mathematics. In geometry, the construction of a figure is split into elementary steps with compass and ruler. In arithmetic and algebra, methods of solving equations are split into elementary steps which, in principle, can be carried out by a machine. Thus, one speaks of algorithms, which are named after the Persian mathematician Al-Chwarizmi. Today, algorithms are executed by computer programs. The question is, to what extend steps cannot only be executed by a machine, but also found independently.

Artificial intelligence (AI) is therefore measured against human intelligence. According to the British logician and computer pioneer Alan M. Turing [2], a technical system is called "intelligent" if its answers and its way to solve problems cannot be distinguished from a human being. Originally, AI was oriented towards the rules and formulas of symbolic logic, which

© Springer-Verlag GmbH Germany, part of Springer Nature 2024

K. Mainzer and R. Kahle, *Limits of AI—Theoretical, Practical, Ethical*, Technik im Fokus, https://doi.org/10.1007/978-3-662-68290-6_1

were translated into suitable computer programs. One, therefore, also speaks of symbolic AI (Fig. 1.1). The underlying epistemological idea is that intelligence is primarily related to the ability of the human mind to derive logical conclusions.

One example was automatic reasoning, in which AI programmes simulated logical reasoning in logic calculi [3]. On this rule-based and symbolic basis, human planning, decision-making and problem solving of human experts should also be simulated in specialised fields of application. In corresponding expert systems or knowledge-based systems, the specific knowledge of an engineer or doctor, for example, is first translated into formal rules which should trigger a specific action automatically when a certain event occurs.

One product, which emerged from this approach is the programming language Prolog (French: programmation en logique), which still today enjoys a certain popularity, even though it is effectively used only in the theoretical sphere. It has not been able to establish itself in the industrial field for reasons that are certainly related to the limits of AI to be discussed here. In Prolog, (simple) rules can be formulated, for example, to store a network of flight connections.

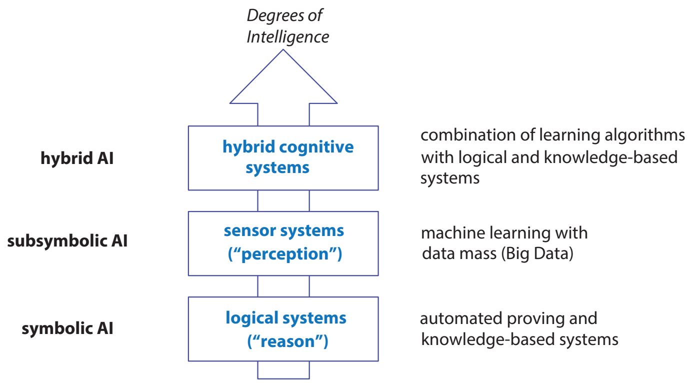


Fig. 1.1 From symbolic and sub-symbolic to hybrid AI

Example

```
reachable(X,Y) :- direct-flight(X,Y).
reachable(X,Y) :- direct-flight(X,Z), reachable(Z,Y).
directflight(NCY,DUB).
directflight(DUB,GWY).
directflight(DUB,ORK).
⋮
```

Prolog is a query language in which, for the given example, the question

```
?- reachable(NCY,GWY)
```

should return the answer `Yes`. ◀

To the extent that this type of knowledge representation was developed further, increasing problems of complexity arose—in two different meanings of "complexity": on the one hand, the general complexity of, for example, the grammar of a language is in general so complex that a simple translation into Prolog rules turns out to be impracticable. On the other hand, problems of computability complexity arise, for example, when querying all theoretically possible flight connections—mathematically the transitive closure of `reachable(X,Y)`—leads to calculation times that are no longer acceptable. Because of these problems, expert systems went out of fashion comparatively quickly.

However, it would be a misinterpretation to restrict research in artificial intelligence in the second half of the twentieth century to the field of expert systems. In particular SAT-solving has emerged from the considerations on automatic theorem proving. This SAT-solving has today far-reaching applications. In addition, motivated by the findings in neurological brain research, neural networks have also been developed as simplified computer simulations of the human nervous system, described with the aid of neurons. From the beginning, this approach was conceptually distinct from the rule-based systems, but hardly progressed beyond “toy applications“—not least due to the still comparatively limited memory and computing power of the available computers. These toy examples, although simulating neural networks in principle, did not yet allow any practical applications.

In a simplified form, the research fields of classical or old AI, as they emerged at the end of the twentieth century, can be summarised as follows:

Classic or Old AI

- *Expert systems*
  Prolog as a paradigmatic programming language.
- *SAT-Solving*
  problem-solving methods for propositional logic which solve complex problems that are just feasible in terms of complexity theory
- *Early neural networks*
  In the early phase of AI, only of very limited complexity.

The early neural networks were already a response to the fact that rule-based knowledge can never fully capture the intuitive skills of an expert. Knowledge is based on manifold experiences that are by no means symbolically represented in a textbook. An experienced driver realizes situations and reacts intuitively on the basis of a great deal of sensory data, without being aware of the logical processes in detail. In the same way an experienced doctor reacts in a critical situation as well as an experienced pilot in the cockpit of an aircraft. Intuition is by no means a mystical magic box. Rather, the recognition of data patterns and the estimation of expected probabilities can be trained and improved through experience.

In this context, logical rules, as in symbolic AI, are replaced by sensory data, in which statistical correlations and probabilities are determined. Learning from data is studied mathematically in statistical learning theory. Its algorithms form the basis of machine learning. From an epistemological point of view, these learning processes from sensory perceptual data take place unconsciously below conscious logical reasoning. This is why one also speak of subsymbolic AI (Fig. 1.1). Mathematically, the paradigm of logic is replaced by statistics and probability theory. The powerful computer technology of the past few years made it possible that machine learning with big data can now be implemented technically. Therefore, machine learning leads to new breakthroughs in the application of AI, e.g. in the development of drugs and vaccines.

Accordingly, at the beginning of the twenty-first century, a statistics-based or new AI has emerged, with the following characteristics.

Statistics-based or new AI

- *Machine learning*
  fed by
  - *Large amounts of data* ("Big Data")
    and often based on a high number of layers in neural networks, which enables
  - *deep learning*.

It should be noted, however, that the term "deep" is not to be understood in the sense of "profound", but only emphasises the aspect of a considerable extension of layers in the network which are comparable to the layers in a human brain. This new AI is thus a manifestation of subsymbolic AI and represents essentially a tool assisting human perception in a form that is optimised in many respects.

However, human intelligence can neither be reduced to the logic of the mind nor to the data of perception. Epistemologically, it depends on the connection between perception and understanding. In AI research, therefore, the future goal is to combine