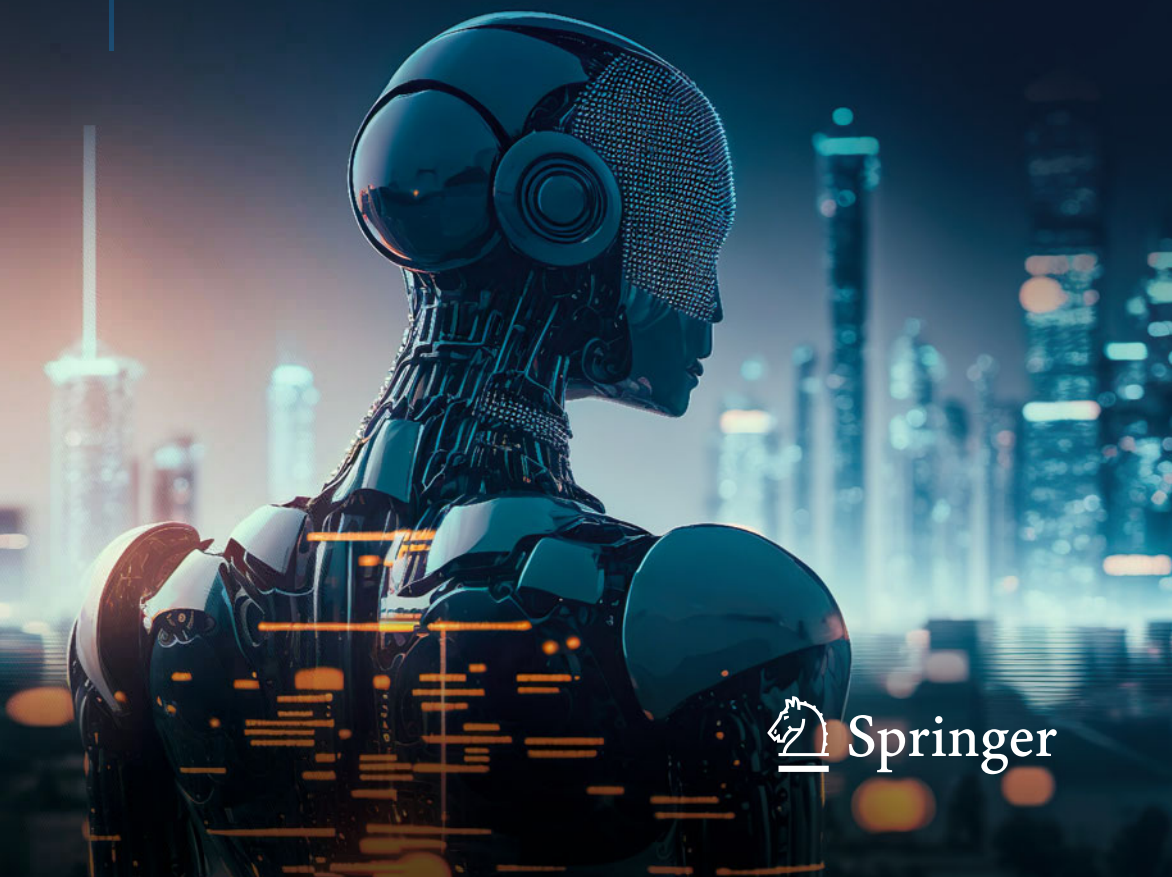
Patrick Krauss

Artificial Intelligence and Brain Research

Neural Networks, Deep Learning and the Future of Cognition



Springer

Artificial Intelligence and Brain Research

Patrick Krauss

Artificial Intelligence and Brain Research

Neural Networks, Deep Learning and the Future of Cognition

Springer

Patrick Krauss
University of Erlangen-Nuremberg
Erlangen, Germany

ISBN 978-3-662-68979-0 ISBN 978-3-662-68980-6 (eBook)
https://doi.org/10.1007/978-3-662-68980-6

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

For Sofie and Hannes

Preface

How does Artificial Intelligence work? How does the brain function? What are the similarities between natural and artificial intelligence, and what are the differences? Is the brain a computer? What are neural networks? What is Deep Learning? Should we attempt to recreate the brain to create real general Artificial Intelligence, and if so, how should we best proceed?

We are in an extremely exciting phase of cultural and technological development of humanity. Recently, Artificial Intelligence (AI) and Machine Learning have been making their way into more and more areas, such as medicine, science, education, finance, engineering, entertainment, and even art and music, and are becoming ubiquitous in twenty-first-century life. Particularly in the field of so-called Deep Learning, the progress is extraordinary in every respect, and deep artificial neural networks show impressive performance in a variety of applications such as processing, recognition, and generation of images or natural language. Especially in combination with a method called Reinforcement Learning, the networks are becoming increasingly powerful, for example when it comes to playing video games, or they even achieve superhuman abilities in complex board games like Go, when they are trained by playing millions of games against themselves.

Many of the algorithms, design principles, and concepts used in AI today, such as neural networks or the aforementioned reinforcement learning, have their origins in biology and psychology. Therefore, neuroscience lectures are becoming an integral part of courses such as computer science or artificial intelligence at more and more universities. But it is also worthwhile for brain researchers to engage with artificial intelligence, as it not only provides important tools for data evaluation, but also serves as a model for natural

intelligence and has the potential to revolutionize our understanding of the brain.

When considering the goals of AI and neuroscience, it becomes apparent that they are complementary to each other. The goal of AI is to achieve cognition and behavior at a human level, and the goal of neuroscience is to understand human cognition and behavior. One could therefore say that artificial intelligence and brain research are two sides of the same coin. The convergence of both research fields promises profound synergies, and it is already certain that the insights gained from this will shape our future in a sustainable way.

In recent years, I have given many lectures on these and related topics. From the subsequent discussions and numerous follow-up questions, I learned that the deep connection between AI and brain research is immediately apparent to most, but was not really conscious before. Although this is gradually beginning to change, most people associate AI exclusively with degree programs such as computer science or data science, and less so with cognitive science or computational neuroscience, even though these branches of science can contribute a lot to basic research in AI. Conversely, artificial intelligence has become indispensable in modern brain research. To understand how the human brain works, research teams are increasingly using models based on artificial intelligence methods, gaining not only neuroscientific insights, but also learning something about artificial intelligence.

There are already many excellent textbooks and non-fiction books in which the various disciplines are each presented in isolation. However, an integrated presentation of AI and brain research has not yet existed. With this book, I want to close this gap. Based on exciting and current research results, the basic ideas and concepts, open questions, and future developments at the intersection of AI and brain research are clearly presented. You will learn how the human brain is structured, what fundamental mechanisms perception, thinking, and action are based on, how AI works, and what is behind the spectacular achievements of AlphaGo, ChatGPT, and Co. Please note that I am not aiming for a comprehensive introduction to AI or brain research. You should only be equipped with what I consider to be the theoretical minimum, so that you can understand the challenges, unsolved problems, and ultimately the integration of both disciplines.

The book is divided into four parts, some of which build on each other, but can also be read independently of each other. So there are different ways to approach the content of this book. Of course, I would prefer if you read the book as a whole, preferably twice: once to get an overview, and a second

time to delve into the details. If you want to get an overview of how the brain works, then start with Part I. However, if you are more interested in getting an overview of the state of research in Artificial Intelligence, then I recommend you start with Part II. The open questions and challenges of both disciplines are presented in Part III. If you are already familiar with the basics and open questions of AI and brain research and are primarily interested in the integration of both research branches, then read Part IV.

I have tried to clarify complex issues through illustrative diagrams wherever possible. My children have actively supported me in creating these diagrams. English quotes have been translated by me, unless otherwise noted. Colleagues, friends, and relatives have greatly helped in correcting errors and improving the clarity and readability of the text. I would like to thank Konstantin Tziridis, Claus Metzner, Holger Schulze, Nathaniel Melling, Tobias Olschewski, Peter Krauß, and Katrin Krauß for this.

My special thanks go to Sarah Koch, Ramkumar Padmanaban, and Ken Kissinger from Springer Publishing, who have supported me in the realization of this book project.

My research work has been and continues to be supported by the German Research Foundation. I am grateful to those in charge. Without the inspiring working atmosphere at the Friedrich-Alexander University Erlangen-Nuremberg and the University Hospital Erlangen, many of my ideas and research projects would not have been possible. My special thanks go to Holger Schulze, Andreas Maier, and Thomas Herbst for their support, as well as Claus Metzner and Achim Schilling for the countless inspiring conversations. I sincerely thank my father for the many discussions on the various topics of this book. My greatest thanks go to my wife, who has always supported everything over the years and continues to do so. What I owe her, I cannot put into words. I dedicate this book to my children.

Großenseebach
in April 2023

Patrick Krauss

Contents

1

Introduction

An elephant is like a fan!

The fifth blind man

ChatGPT Passes the Turing Test

In the field of Artificial Intelligence (AI), there have been a number of spectacular breakthroughs in the last approximately 10–15 years—from *AlphaGo* to *DALL-E 2*to *ChatGPT* –, which until recently were completely unthinkable.

The most recent event in this series is certainly also the most spectacular: It is already clear that November 30, 2022 will go down in history. On this day, the company OpenAI made the Artificial Intelligence *ChatGPT* freely accessible to the public. This so-called Large Language Model can generate any type of text in seconds, answers questions on any topic, gives interviews and conducts conversations, remembering the course of which and thus responding adequately even in longer conversations. Since then, millions of people have been able to convince themselves daily of the amazing capabilities of this system. The responses and texts generated by *ChatGPT* are indistinguishable from those produced by humans. *ChatGPT* thus becomes the first in the history of Artificial Intelligence to pass the Turing Test, a procedure devised to determine whether a machine has the ability to think (Turing, 1950). An artificial system that passes the Turing Test has

P. Krauss, *Artificial Intelligence and Brain Research*,
https://doi.org/10.1007/978-3-662-68980-6_1

long been considered the Holy Grail of research in the field of Artificial Intelligence. Even though passing the Turing Test does not necessarily mean that *ChatGPT* actually thinks, you should still remember November 30, 2022 well. It not only represents the most important milestone in the history of Artificial Intelligence to date, but its significance is certainly comparable to the invention of the loom, the steam engine, the automobile, the telephone, the internet, and the smartphone, which often only turned out to be game-changers and decisive turning points in development in retrospect.

The Next Affront

In addition to the much-discussed consequences that *ChatGPT* and similar AI systems will have on almost all levels of our societal life, the astonishing achievements of these new systems also challenge our explanations of what fundamental concepts such as cognition, intelligence andconsciousness mean at all. In particular, the influence that this new type of AI will have on our understanding of the human brain is already immense and its impacts are not yet fully foreseeable.

Some are already talking about the next great affront to humanity. These are fundamental events or insights that have profoundly shaken man's self-understanding and his relationship to the world throughout history.

The Copernican affront, named after the astronomer Nicolaus Copernicus, refers to the discovery that the Earth is not the center of the universe, but revolves around the sun. This realization in the sixteenth century fundamentally changed the world view and led to a loss of self-centeredness and self-assurance. With the discovery of thousands of exoplanets, i.e., planets outside our solar system, in recent decades, this affront has even been intensified. This has shown that planetary systems are very common in our galaxy and that there may even be many planets that orbit in the habitable zone around their stars and thus represent possible places for life.

Another insight that affected man's self-understanding was the Darwinian affront. Charles Darwin's theory of evolution in the nineteenth century showed that man is not a species created by God, but has evolved like all other species through evolution. This discovery questioned man's self-understanding as a unique species separated from the rest of nature.

Another affront, which Sigmund Freud modestly named after the theory he developed, the psychoanalytic affront, refers to the discovery that human behavior and thinking are not always consciously and rationally controlled, but are also influenced by unconscious and irrational drives. This realization

shook man's confidence in his ability for self-control and rationality. The Libet experiments, which even question the existence of free will, further intensified the impact of this affront.

AI can be considered as a newly added fourth major affront to human self-understanding. Until now, our highly developed language was considered the decisive distinguishing feature between humans and other species. However, the development of large language models like *ChatGPT* has shown that machines are in principle capable of dealing with natural language in a similar way to humans. This fact challenges our concept of uniqueness and incomparability as a species again and forces us to at least partially rethink our definition of being human.

This "AI affront" affects not only our linguistic abilities, but our cognitive abilities in general. AI systems are already capable of solving complex problems, recognizing patterns, and achieving human-like or even superhuman performance in certain areas (Mnih et al., 2015; Silver et al., 2016, 2017a, b; Schrittwieser et al., 2020; Perolat et al., 2022). This forces us to reinterpret human intelligence and creativity, where we have to ask ourselves what role humans play in a world where machines can take over many of our previous tasks. It also forces us to think about the ethical, social, and philosophical questions that arise from the introduction of AI into our lives. For example, the question arises as to how we should deal with the responsibility for decisions made by AI systems, and what limits we should set on the use of AI to ensure that it serves the good of humanity (Anderson & Anderson, 2011; Goodall, 2014; Vinuesa et al., 2020).

Less than half a year after the publication of *ChatGPT*, its successor *GPT-4* was released in March 2023, which significantly surpasses the performance of its predecessor. This prompted some of the most influential thinkers in this field to call for a temporary pause in the further development of AI systems, which are even more powerful than GPT-4, in a widely noticed open letter[1] to prevent a potentially impending loss of control.

Artificial Intelligence and Brain Research

The remarkable achievements of *ChatGPT* and *GPT-4* also have direct implications for our understanding of the human brain and how it functions. They therefore not only challenge brain research, but even have the

[1] https://futureoflife.org/open-letter/pause-giant-ai-experiments/

potential to revolutionize it. Indeed, AI and brain research have always been closely intertwined in their history. The so-called cognitive revolution in the middle of the last century can also be seen as the birth of research in the field of AI, where it developed as an integral part of the newly emerged research agenda of cognitive sciences as an independent discipline. In fact, AI research was never just about developing systems to take over tedious work. From the beginning, it was also about developing and testing theories of natural intelligence. As we will see, some astonishing parallels between AI systems and brains have been uncovered recently. Therefore, AI plays an increasingly important role in brain research, not only as a pure tool for data analysis, but especially as a model for the function of the brain.

Conversely, neuroscience has also played a key role in the history of artificial intelligence and has repeatedly inspired the development of new AI methods. The transfer of design and processing principles from biology to computer science has the potential to provide new solutions for current challenges in the field of AI. Here too, brain research not only plays the role of providing the brain as a model for new AI systems. Rather, a variety of methods for deciphering the representation and calculation principles of natural intelligence have been developed in neuroscience, which can now in turn be used as a tool for understanding artificial intelligence and thus contribute to solving the so-called black box problem. An endeavor occasionally referred to as Neuroscience 2.0. It is becoming apparent that both disciplines will increasingly merge in the future (Marblestone et al., 2016; Kriegeskorte & Douglas, 2018; Rahwan et al., 2019; Zador et al., 2023).

Too Blind to See the Elephant

The realization that different disciplines must work together to understand something as complex as human-level cognition is of course not new and is vividly illustrated in the well-known metaphor of the six blind men and the elephant (Friedenberg et al., 2021):

> Once upon a time, there were six blind scientists who had never seen an elephant and wanted to research what an elephant is and what it looks like. Each examined a different part of the body and accordingly came to a different conclusion.
>
> The first blind approached the elephant and touched its side. “Ah, an elephant is like a wall,” she said.

The second blind touched the elephant's tusk and exclaimed: "No, an elephant is like a spear!"

The third blind touched the elephant's trunk and said: "You are both wrong! An elephant is like a snake!"

The fourth blind touched a leg of the elephant and said: "You are all wrong. An elephant is like a tree trunk."

The fifth blind touched the elephant's ear and said: "None of you know what you're talking about. An elephant is like a fan."

Finally, the sixth blind approached the elephant and touched its tail: "You are all wrong," he said. "An elephant is like a rope."

If the six scientists had combined their findings, they would have come much closer to the true nature of the elephant. In this story, the elephant represents the human mind, and the six blind people represent the various scientific disciplines that try to understand its functioning from different perspectives (Fig. 1.1). The punchline of the story is that while each individual's perspective is valuable, a comprehensive understanding of cognition

Fig. 1.1 The Blind Men and the Elephant. Each examines a different part of the body and accordingly comes to a different conclusion. The elephant represents the mind and brain, and the six blind represent different sciences. The perspective of each individual discipline is valuable, but a comprehensive understanding can only be achieved through collaboration and interdisciplinary exchange

can only be achieved when the different sciences work together and exchange ideas.

This is the founding idea of cognitive science, which began in the 1950s as an intellectual movement referred to as the cognitive revolution (Sperry, 1993; Miller, 2003). During this time, there were significant changes in the way psychologists and linguists worked and new disciplines such as computer science and neuroscience emerged. The cognitive revolution was driven by a number of factors, including the rapid development of personal computers and new imaging techniques for brain research. These technological advances allowed researchers to better understand how the brain works and how information is processed, stored, and retrieved. As a result of these developments, an interdisciplinary field emerged in the 1960s that brought together researchers from a wide range of disciplines. This field went by various names, including information processing psychology, cognition research, and indeed cognitive science.

The cognitive revolution marked a significant turning point in the history of psychology and related disciplines. It fundamentally changed the way researchers approach questions of human cognition and behavior, paving the way for numerous breakthroughs in areas such as artificial intelligence, cognitive psychology, and neuroscience.

Today, cognitive science is understood as an interdisciplinary scientific endeavor to explore the different aspects of cognition. These include language, perception, memory, attention, logical thinking, intelligence, behavior and emotions. The focus is primarily on the way natural or artificial systems represent, process, and transform information (Bermúdez, 2014; Friedenberg et al., 2021).

The key questions are: How does the human mind work? How does cognition work? How is cognition implemented in the brain? And how can cognition be implemented in machines?

Thus, cognitive science addresses some of the most difficult scientific problems, as the brain is incredibly difficult to observe, measure, and manipulate. Many scientists even consider the brain to be the most complex system in the known universe.

The disciplines involved in cognitive science today include linguistics, psychology, philosophy, computer science, artificial intelligence, neuroscience, biology, anthropology, and physics (Bermúdez, 2014). For a time, cognitive science fell somewhat out of fashion, particularly the idea of integrative collaboration between different disciplines was somewhat forgotten. Specifically, AI and neuroscience developed independently and thus also away from each other. Fortunately, the idea that AI and brain research are

complementary and can benefit greatly from each other is currently experiencing a real renaissance, with the term "cognitive science" apparently being interpreted differently in some communities today or considered too old-fashioned, which is why terms like *Cognitive Computational Neuroscience* (Kriegeskorte & Douglas, 2018) or *NeuroAI* (Zador et al., 2023) have been suggested instead.

The legacy of the cognitive revolution is evident in the many innovative and interdisciplinary approaches that continue to shape our understanding of the human mind and its functioning. Whether through state-of-the-art brain imaging techniques, sophisticated computer models, or new theoretical frameworks—researchers are constantly pushing the boundaries of what we know about the human brain and its complex processes.

Brain-Computer Analogy

Many researchers believe that computer models of the mind can help us understand how the brain processes information, and that they can lead to the development of more intelligent machines. This assumption is based on the brain-computer analogy (Von Neumann & Kurzweil, 2012). It is assumed that mental processes such as perception, memory, and logical thinking involve the manipulation of mental representations that correspond to the symbols and data structures used in computer programs (Fig. 1.2). Like a computer, the brain is capable of receiving, storing, processing, and outputting information.[2]

However, this analogy does not mean that the brain is actually a computer, but that it performs similar functions. By considering the brain as a computer, one can abstract from biological details and focus on the way it processes information to develop mathematical models for learning, memory, and other cognitive functions.

The brain-computer analogy is based on two central assumptions that underlie cognitive science. These are computationalism and functionalism.

[2] A fundamental difference is that a computer processes information with different components than those with which it stores the information. In the brain, both are done by the—sometimes same—neurons.

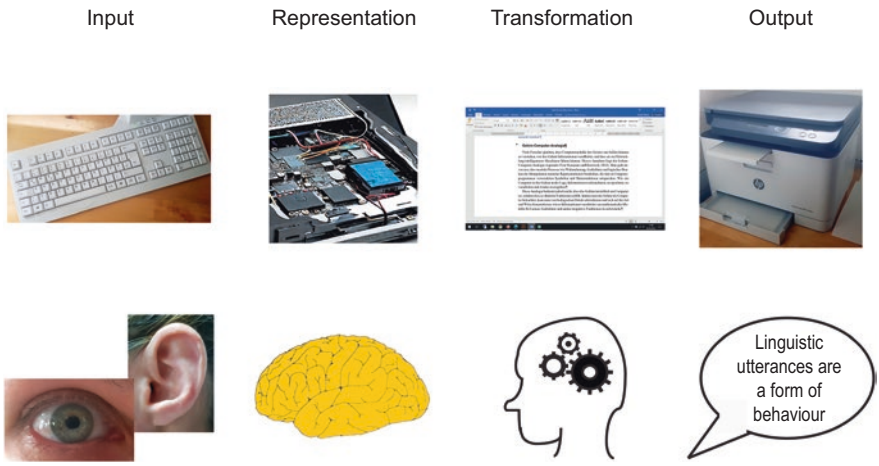


Fig. 1.2 Brain-Computer Analogy. Information processing includes the input, representation, transformation, and output of information. For a computer, the input may come from the keyboard, for a biological organism from the sensory organs. This input must then be represented: by storing it on a hard drive or in the computer's RAM, or in the brain as momentary neuronal activity in short-term memory or in long-term memory in the interconnection of neurons. Then a transformation or processing takes place, i.e., mental processes or algorithms must act on the stored information and change it to generate new information. For a computer, this could be text processing, for humans, for example, logical reasoning. Finally, the result of information processing is output. The output can be, for example, via a printer for a computer. In living beings, the output corresponds to observable behavior or, as a special case of behavior, to human linguistic utterances

Computationalism

In computationalism, it is assumed that cognition is synonymous with information processing, i.e., that mental processes can be understood as calculations and that the brain is essentially an information processing system (Dietrich, 1994; Shapiro, 1995; Piccinini, 2004, 2009). Like any such system, the brain must therefore represent information and then transform these represented information, i.e., there must be mental representations of information and there must be mental processes that can act on these representations and manipulate them. Computationalism has greatly influenced the way cognitive scientists and researchers in the field of artificial intelligence think about intelligence and cognition.

However, there is also criticism of this view, as evidenced by numerous ongoing debates in philosophy and cognitive science. Some critics argue, for example, that the computer model of the mind is too simple and cannot

fully capture the complexity and richness of human cognition. Others argue that it is unclear whether mental processes can really be understood as calculations or whether they fundamentally differ from the way processes occur in computers.

Functionalism

Is cognition only possible in a (human) brain? Functionalism clearly answers this question with a no. Accordingly, mental states and processes are defined exclusively by their functions or their relationship to behavior, not by their physical or biochemical properties (Shoemaker, 1981; Jackson & Pettit, 1988; Piccinini, 2004). What does this mean in concrete terms?

Please imagine a car in your mind's eye. And now remember the last situation in which you ate chocolate, and try to recall the taste as accurately as possible. Did you succeed? I assume you did. As I write these lines, I have brought to mind the same two mental states *"seeing a car"* and *"tasting chocolate"*. Obviously, each of us can activate the corresponding mental representations in our brains, even though you, I, and every other reader of these lines have completely different brains. All human brains are of course similar in their basic structure. But they are certainly not identical down to the smallest detail, especially not in the exact wiring of the neurons, if only because every person has had completely different, individual experiences in their life, which affect the wiring pattern of the brain. In computer science terminology, one would say that each person has a different, individual hardware. Yet we can all bring to mind the same mental state.

While in the previous example the systems were somehow very similar—they were always human brains—the following example may illustrate how much the different physical implementations of the same algorithm can differ from each other. Consider the addition of two numbers. The representation of these numbers, as well as the associated process or algorithm to add them, can be implemented in your brain when you "calculate in your head", or for example also in a laptop with spreadsheet program, a slide rule, a calculator or a calculator app on your smartphone. Each time, the same numbers are represented and added, while the information processing systems are completely different. This is the concept of *multiple realizability*.

Accordingly, the same mental state or process can in principle be realized by completely different natural or artificial systems. Put simply, this means that cognition and presumably also consciousness can in principle be implemented in any physical system capable of supporting the required

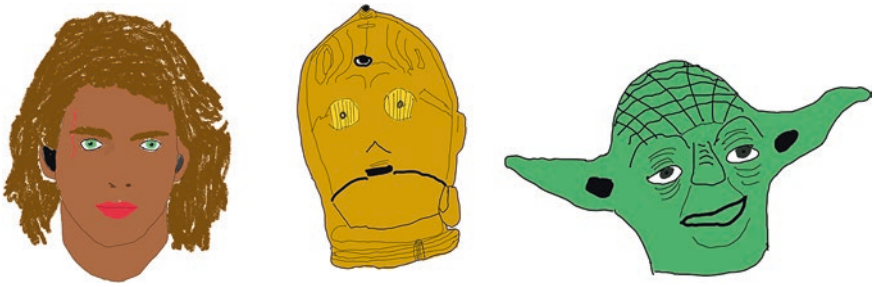


Fig. 1.3 Functionalism. Human-level cognition is not limited to a human brain, but could in principle also be implemented in any other system that supports the required calculations, such as correspondingly advanced robots or aliens

calculations. If many different human brains are already capable of this, why should this ability be limited to humans or biological systems? From the perspective of functionalism, it is therefore quite possible that the ability for human-like cognition can also be implemented in correspondingly highly developed machines or alien brains (Fig. 1.3).

Conclusion

In recent years, spectacular advances in artificial intelligence have turned our understanding of cognition, intelligence, and consciousness upside down and will have profound impacts on society and our understanding of the human brain. Cognitive science is the key to a deeper understanding of brain and mind, and computer models of the mind can help us understand how the brain processes information and contribute to the development of smarter machines. These models are based on the central assumptions of computationalism and functionalism, which emphasize the equivalence of cognition and information processing as well as the independence of cognitive processes from their physical implementation.

The advances in artificial intelligence have also led to the fields of neuroscience and computer science becoming increasingly intertwined. The transfer of construction and processing principles from biology to computer science promises new solutions for current challenges in artificial intelligence. Conversely, the close collaboration of these disciplines will become increasingly important in the future to understand complex systems like the human brain.

The recent advances in artificial intelligence and their applications have opened the door in neuroscience to new insights and technologies far beyond what was previously possible. We are only at the beginning of a new era of research and innovation, and it remains to be seen what fascinating discoveries and developments await us in the future.

References

Anderson, M., & Anderson, S. L. (Hrsg.). (2011). *Machine ethics*. Cambridge University Press.

Bermúdez, J. L. (2014). *Cognitive science: An introduction to the science of the mind*. Cambridge University Press.

Dietrich, E. (1994). *Computationalism. In thinking computers and virtual persons* (pp. 109–136). Academic.

Friedenberg, J., Silverman, G., & Spivey, M. J. (2021). *Cognitive science: An introduction to the study of mind*. Sage.

Goodall, N. J. (2014). Machine Ethics and Automated Vehicles. In G. Meyer & S. Beiker (Eds.), *Road vehicle automation. Lecture notes in mobility*. Springer. https://doi.org/10.1007/978-3-319-05990-7_9.

Jackson, F., & Pettit, P. (1988). Functionalism and broad content. *Mind,97*(387), 381–400.

Kriegeskorte, N., & Douglas, P. K. (2018). Cognitive computational neuroscience. *Nature neuroscience,21*(9), 1148–1160.

Marblestone, A. H., Wayne, G., & Kording, K. P. (2016). Toward an integration of deep learning and neuroscience. *Frontiers in computational neuroscience,10,* 94.

Miller, G. A. (2003). The cognitive revolution: A historical perspective. *Trends in cognitive sciences,7*(3), 141–144.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., … & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature,518*(7540), 529–533.

Perolat, J., De Vylder, B., Hennes, D., Tarassov, E., Strub, F., de Boer, V., … & Tuyls, K. (2022). Mastering the game of Stratego with model-free multiagent reinforcement learning. *Science,378*(6623), 990–996.

Piccinini, G. (2004). Functionalism, computationalism, and mental contents. *Canadian Journal of Philosophy,34*(3), 375–410.

Piccinini, G. (2009). Computationalism in the philosophy of mind. *Philosophy Compass,4*(3), 515–532.

Rahwan, I., Cebrian, M., Obradovich, N., et al. (2019). Machine behaviour. *Nature,568,* 477–486.

Schrittwieser, J., Antonoglou, I., Hubert, T., Simonyan, K., Sifre, L., Schmitt, S., … & Silver, D. (2020). Mastering Atari, Go, Chess and Shogi by planning with a learned model. *Nature,588*(7839), 604–609.

Shapiro, S. C. (1995). Computationalism. *Minds and Machines,5,* 517–524.

Shoemaker, S. (1981). Some varieties of functionalism. *Philosophical topics,12*(1), 93–119.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., … & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature,529*(7587), 484–489.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., … & Hassabis, D. (2017a). Mastering the game of Go without human knowledge. *Nature,550*(7676), 354–359.

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., … & Hassabis, D. (2017b). *Mastering Chess and Shogi by self-play with a general reinforcement learning algorithm*. arXiv preprint arXiv:1712.01815.

Sperry, R. W. (1993). The impact and promise of the cognitive revolution. *American Psychologist,48*(8), 878.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind,59*(236), 433–460.

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., … & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the sustainable development goals. *Nature Communications,11*(1), 233.

Von Neumann, J., & Kurzweil, R. (2012). *The computer and the brain*. Yale University Press.

Zador, A., Escola, S., Richards, B., Ölveczky, B., Bengio, Y., Boahen, K., … & Tsao, D. (2023). Catalyzing next-generation artificial intelligence through NeuroAI. *Nature Communications,14*(1), 1597.

Part I
Brain Research

In the first part of the book, the aim is to familiarize you with the most important aspects of the structure and function of the brain. In doing so, a detailed and systematic description of many molecular biological, physiological, and anatomical details is deliberately avoided. The presentation also makes no claim to completeness. Interested readers may deepen their knowledge with one of the many excellent textbooks available on psychology and neuroscience. Rather, these first chapters are intended to convey the basics necessary from the author's point of view, on the basis of which we want to show the numerous cross-connections to Artificial Intelligence in later chapters.

2

The Most Complex System in the Universe

There is always a bigger fish.

Qui-Gon Jinn

The Brain in Numbers

The human brain consists of approximately 86 billion nerve cells, known as neurons (Herculano-Houzel, 2009). These are the fundamental processing units responsible for the reception, processing, and transmission of information throughout the body. The neurons are connected via so-called synapses and form a gigantic neural network. On average, each neuron receives its input from about 10,000 other neurons and sends, also on average, its output to about 10,000 subsequent neurons (Kandel et al., 2000; Herculano-Houzel, 2009). The actual number of connections per neuron can vary significantly, over several orders of magnitude, which is why we also speak of a broad distribution of connections per neuron. Some neurons, such as those in the spinal cord, are only connected to a single other neuron, while others, for example in the cerebellum, can be connected to up to a million other neurons.

Based on the total number of neurons and the average number of connections per neuron, the total number of synaptic connections in the brain can be roughly estimated at one quadrillion. This is a number with 15 zeros and can also be written as 10^{15}. In recent years, we have become somewhat

P. Krauss, *Artificial Intelligence and Brain Research*,
https://doi.org/10.1007/978-3-662-68980-6_2

accustomed to amounts beyond a thousand billion, i.e., in the trillion range, in the context of politics and economics. The approximate number of synapses in the brain is a thousand times larger!

This may sound like a lot, but let's consider how many synapses would theoretically be possible in the brain. Each of the approx. 10^{11} neurons could in principle be connected to every other, with the information between any two neurons potentially running in two directions: either from neuron A to neuron B or vice versa. Additionally, each neuron can indeed be connected to itself. These special types of connections are called autapses. Purely combinatorially, this results in 10^{11} times 10^{11}, or 10^{22}, as the possible number of synapses. A comparison with the number of actually existing synapses shows that only about one in 10 million theoretically possible connections is actually realized. The network that the neurons form in the brain is therefore anything but dense *(dense),* but on the contrary extremely sparse *(sparse)* (Hagmann, 2008).

How Many Different Brains Can There Be?

In reference to the genome, which refers to the entirety of all genes of an organism, the connectome is the entirety of all connections in the nervous system of a living being (Sporns et al., 2005). To answer the question of how many different brains there can be, one must estimate how many different connectomes are combinatorially possible. At this point, it should be noted that not every theoretically possible connectome must result in a functioning viable nervous system. It turns out that it is quite complicated to calculate the exact number, which is why we want to settle here with an estimate for the lower limit of the actual number based on some simplifications. Let's assume for simplicity that each of the 10^{22} theoretically possible connections can either be present or not present. So we assume binary connections, with one representing an existing and zero a non-existing connection. As we will see later, reality is even more complicated. But even under this strong simplification, an absurdly high number of $2^{10^{22}}$ (read "two to the power of ten to the power of 22") results. This corresponds to a number with a trillion zeros. Of course, not every one of these connectomes leads to a powerful and viable nervous system, so the realizable number should only correspond to a tiny subset of all theoretically possible connectomes. On the other hand, the synaptic weights are not binary numbers, but can take any continuous value, which significantly increases the number of possibilities again.