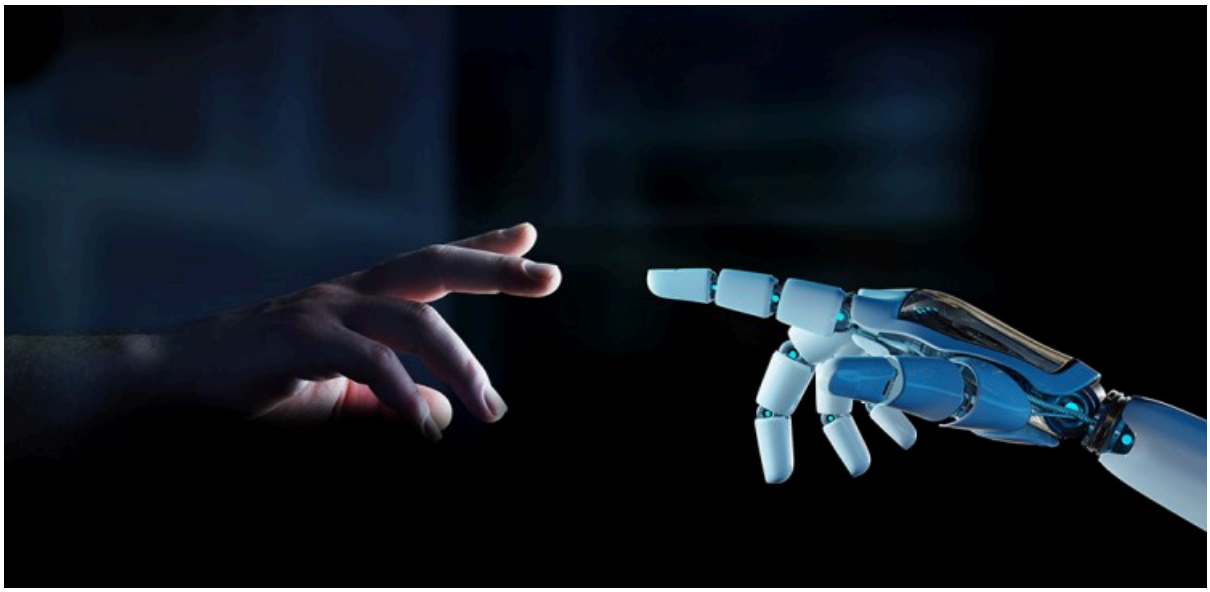
Autor: Marc-Aurel Luca Lewald (PPE)
Course of Studies: SIRP / ITPP - Special Issue: Transformative Research Project
Institution: Karlshochschule International University
Professor(s): Prof. Dr. Wendelin Küpers
17. Juni 2023

What Are The Key Principles And Limitations Of Artificial Intelligence?



Siemens (2023)

1. Introduction

In an era marked by rapid technological advancements and digital transformation, artificial intelligence (AI) emerges as a cornerstone of innovation, reshaping industries and redefining the way we interact with the world. The exponential growth of AI has undoubtedly already fueled the integration of intelligent systems into various domains, from healthcare and finance to education and transportation, propelling humanity into a new age of possibilities (Microsoft, 2018), and representing one of the biggest shifts in human history (Roser, 2023). Among the most well-known and famous companies behind the development of AI are e.g. Amazon, IMB, Microsoft, C3AI, Google, Palantir, Salesforce, Nvidia, and many many more. These companies are key players and play an important role in the development of AI.

At it's core, AI refers to the development of computer systems capable of performing tasks that typically require human intelligence, such as e.g. problem-solving, pattern recognition, learning, or even decision-making. In his paper: "What is artificial Intelligence" from 2007, John McCarthy defines AI as:

> *"the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable." (McCarthy, 2007, Basic Questions)*

This is a widely used definition that has also been broadly accepted. However, there are also voices that claim that there is no uniform and clear definition of and about AI, especially with regard to legal requirements (Schuett, 2019), but also when it comes to the actual application areas of AI (Schuett, 2021).

Nevertheless, the general significance of AI in today's world cannot be overstated. It has revolutionized the way we process vast amounts of data, unlocking unprecedented insights and facilitating informed decision-making across a multitude of sectors, changing the world and society as we know it (Makridakis, 2017). AI-driven applications have the potential to improve efficiency, enhance user experiences, and tackle complex challenges that were once considered insurmountable. Experts estimate the increase in labour productivity related to AI to be between 11 and 37 % by the year 2035 (European Parliament, 2022).

As AI becomes increasingly pervasive, it is essential to understand its underpinnings and inherent limitations to ensure that its development and deployment remain responsible, ethical, and equitable (European Parliament, 2020). All the more so in view of a study by the WEF, at which 90% of participants (AI experts) consider it plausible that that the first human-level AI's could be developed within the next 100 years (Roser, 2023).

In the light of AI's growing prominence, this paper seeks to address the research question, "**What are the key principles and limitations of artificial intelligence?**“ It will do so by delving into the foundations of AI, providing a comprehensive understanding of the core concepts of AI, differentiating between artificial intelligence, machine learning, deep learning and neural networks while elucidating their respective characteristics and relationships.

Furthermore, it will explore AI learning processes including the challenges and limitations associated with AI learning processes, examining how these constraints may influence its capabilities and effectiveness. This includes the investigation and examination of the role of AI in reinforcing or mitigating societal biases. As AI systems increasingly make decisions that impact human lives, understanding the sources and consequences of biases within these algorithms becomes more and more important (Srinivasan & Chander, 2021).

Lastly, it will pick up on ethical considerations surrounding AI, exploring the implications of its use in various sectors and the potential consequences of its limitations on current and future applications. It quickly becomes clear that there are countless ethical questions that need to be addressed in the field of artificial intelligence. By providing a holistic perspective on the key principles and limitations of artificial intelligence, this paper aims to foster a nuanced understanding of AI's potential and challenges, providing an overview about the development of AI and empowering readers to engage in informed discussions and contribute to the responsible development and deployment of AI technologies.

2. Background

2.1. Brief history of AI development

The evolution of artificial intelligence has a rich and storied history, with the seeds of its development planted in ancient myths and philosophical inquiries that contemplated the nature of intelligence and the possibility of creating artificial beings with human-like cognitive abilities. A prime illustration of this concept is Hephaistos, the Greek deity of blacksmithing and craftsmanship, who employs a self-operating bellows for carrying out basic, monotonous tasks (Truitt, 2021).

However, the birth of AI as a recognized field of study can be traced back to the mid-20th century, when pioneers such as Alan Turing, John McCarthy, Marvin Minsky, Geoffrey Hinton and others began to lay the groundwork for the development of intelligent machines, making them among the most influential people in the field of AI.

- The already quoted **John McCarthy** is one of the first proponents of the development of AI and the inventor of the programming language LISP. He coined the term Artificial Intelligence in 1955 for the first time at the Dartmouth Conference. (Britannica, 2022)

- **Marvin Minsky** coined the term Artificial Intelligence together with John McCarthy and was the founder of the Artificial Intelligence Laboratory at the Massachusetts Institute of Technology (MIT), probably one of the most notorious laboratories in the field. (Dennis, 2023)

- **Alan Turing** developed the so-called Turing test in the course of a paper, a research method to determine whether a computer is capable of thinking like a human being or not. Even though the

test dates back to the 1950s, the Turing test remains state of the art to this day. (Turing, 1950; SEP, 2023)

- Last but not least, **Geoffrey Hinton**, who has often been called the Godfather of AI and is largely responsible for the development of machine & deep learning over the past decades. However, he recently turned against the strong advance of AI and is now warning about the dangers that the new systems bring with them. In the wake of this, he left his job at Google and is now working to educate people about the risks and dangers of AI (Metz, 2023).

So now one of the men responsible for the current development in AI is fighting against it. In his view, AI poses a much greater threat to humanity and the planet than climate change (Barkhausen, 2023).

However, these early efforts led to groundbreaking advancements in various subfields, including e.g. symbolic AI, which relied on explicit knowledge representation and logical reasoning, as well as connectionism, an approach that sought to model cognitive processes using artificial neural networks inspired by the structure and function of biological neurons (Goel, 2021).

As AI research progressed over the years and decades, the field witnessed periods of rapid advancement and excitement (AI Summers), followed by episodes of stagnation and disillusionment (AI Winters) (Toosi et al., 2021). Through this cyclical process of growth and retreat, AI research has gradually matured, giving rise to the modern era of machine learning and deep learning as we know it today. With the current hype around AI and the current development, it is clear that we are in times of an AI summer. It will be rather interesting to see whether there will still be another winter in the future with the current exponential development.

2.2. Distinction between Artificial Intelligence, Machine Learning, Deep Learning and Neural Networks

In the context of artificial intelligence, the terms machine learning, deep learning or neural networks are often used. But what exactly are these concepts? And how are these concepts related to the general notion of artificial intelligence and to each other? Simply put, all forms are different sub-groups and sub-fields of AI that have emerged over the years from the general field of AI (figure 1).

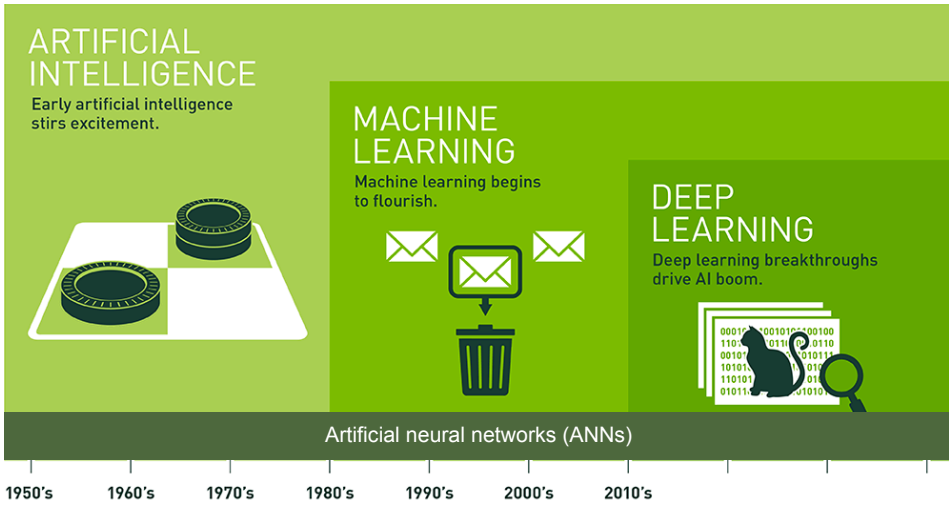


Figure 1. (Samson, 2017)

Let's start with Neural Networks. Artificial neural networks (ANNs) are designed to emulate the human brain by employing a series of algorithms, consisting of different layers. They are the starting point for the development of today's technologies in the field of AI.

In simple terms, machine learning is then a neural network with more than 3 layers. With that, machine learning provides a more powerful and sophisticated approach to modeling complex relationships within data (Kavlakoglu, 2020). Thus, machine learning is already a further development of neural networks and deep learning is then a further stage of development.

As the name implies, deep learning is then about the depth of layers. Deep Learning is "based on the development of artificial neural networks that mimic the way that the human brain works … Deep learning technology has demonstrated its strength to recognize images, to communicate, and to translate from one language to another" (Aggarwal et al., 2022).

AI itself, then, is a field that has been around for quite a long time and has also involved many different phases, but has stretched as an all-encompassing term from the beginning of history to the present day, encompassing Machine Learning, Deep Learning, and Neural Networks. To develop a comprehensive understanding of the fundamentals of the field and the rapidly evolving landscape of AI research and applications, it is imperative to understand the interwoven relationships between these different areas.

3. Key principles of AI Learning Processes

Now that the basic structure and understanding of AI has been clearly laid out and broken down into its various parts, the next step is to take a closer look at the basic principles on which AI systems are based. This comprises a set of techniques and methods that, when combined, give AI systems the ability to possess cognitive capabilities.

One of the core principles within the sphere of AI are "**problem-solving and search techniques**", which involve the identification of optimal or near-optimal solutions to intricate problems by methodically traversing the solution space (Viharos & Kemény, 2007). Simply put, it's like trying to find the quickest route from home to a destination in a big city. AI uses these principles to methodically sift through all possible options, just like checking all roads.

Another pivotal principle is "**knowledge representation and reasoning**". It concentrates on the encoding, organization, and manipulation of information in a way that promotes intelligent reasoning and decision-making (Lakemeyer & Nebel, 1994). It is comparable with a librarian organizing books in a library. Just as the librarian sorts books so one can find the book needed, AI organizes information in a way that it can use to make smart choices or come up with new ideas.

Learning methods and adaptability are another cornerstone of AI. They in general allow systems to accumulate and refine knowledge based on experience and thus improve their performance over time. It's like a toddler that learns to walk. They try, they fall, and they try again, each time improving a little bit. Similarly, AI systems learn from their mistakes and successes, getting better and smarter as they gain more experience, just like the toddler eventually learns to walk and then run. From an evolutionary point of view, models or algorithms must also adapt to new environments and challenges out of a sense of urgency to "survive" (eLife, 2021).

In addition, **natural language perception and understanding** also serve as key principles within AI. These tasks revolve around the question: "how computers can be used to understand and manipulate text or speech in natural language to do useful things" (Chowdhury, 2003, Introduction). They enable systems to process and interpret sensory information or human speech. With that AI learns to understand human language, be it written or spoken, so it can interact with us more naturally. It's like teaching a computer to read a book, understand a conversation, or even reply to your text messages just like a human would. In this way, AI systems can interact with and understand the world around them.

Last but not least, also **robotics and embodied intelligence** form a crucial facet of artificial intelligence. It's in general about the combination of key principles such as perception, reasoning, learning, and decision-making within physical agents that actively interact with their environment. Although one has to note, that "there is not a consensus as to what precisely makes an intelligence "embodied" so far" (Roy et al., 2021, The Challenges of Embodied Intelligence).

This dimension of AI transcends purely computational tasks and encompasses the design, control, and coordination of robotic agents capable of navigating and manipulating the physical world. This multifaceted discipline requires the integration of sensors and actuators for perception and action, as well as the development of algorithms and control strategies that enable robots to reason, learn, and adapt to their surroundings. (Starzyk, 2008)

By bridging the gap between the digital and physical worlds, robotics and embodied intelligence play a central role in advancing AI research and applications, and expanding the possibilities of AI-driven innovation in areas such as manufacturing, logistics, healthcare, agriculture and personal assistance. This makes the knowledge of this interplay and its possibilities highly relevant when addressing the challenges or limitations in the AI field. (Howard et al., 2018)

The inclusion of robotics and embodied intelligence not only broadens the range of possible applications, but also deepens the understanding of the interplay between AI and the physical world, paving the way for increasingly sophisticated and versatile AI systems across different industries and use cases. Taken together, these fundamental principles form the backbone of AI research and development and shape the capabilities and potential of AI systems. However, along with all these capabilities and potentials, there also arise numerous barriers and obstacles.

4. Challenges and limitations in AI learning processes

Now that the basics of understanding and functioning of systems with artificial intelligence have been shown, it's about to continue with the question of where exactly the current challenges or limitations of these basic principles lie and how they can be explained.

One of the primary challenges in AI learning processes is striking a balance between **overfitting and generalization**. When a model is excessively trained on a particular dataset, it becomes overly focused and struggles to perform well with unfamiliar data. It begins generating inaccurate predictions when presented with new information, making the model ineffective despite its accuracy on the original training data. This phenomenon is referred to as overfitting as it „is no longer generalizable to the overall population" (Mutasa et al., 2020, Overfitting).

Conversely, generalization pertains to the capability of the model to effectively function with novel, unobserved data by grasping the inherent trends and associations present in the data it was trained on (Google, 2022). Techniques such as regularization, network-reduction, data-expansion and early stopping have been developed to mitigate overfitting and enhance generalization, ensuring that AI models remain robust and useful across diverse problem domains (Ying, 2019).

Another problem is the **quality and quantity** of the available data. Data serves as the lifeblood of AI learning processes, and both its quality and quantity pose significant challenges in developing effective AI models. High-quality data is crucial for training AI systems, as inaccurate, incomplete, or noisy data can lead to suboptimal model performance and hinder the learning process. Ensuring that data is cleaned, curated, and accurately labeled is paramount to the success of AI models. Researchers identified major challenges in terms of completeness, accuracy of features, and accuracy of targets in the respective areas. (Budach et. al, 2022)

Additionally, obtaining a sufficient quantity of data is often a challenge, as many AI learning techniques, particularly deep learning, require large amounts of data to achieve satisfactory performance. (McKendrick, 2021)

As AI models become increasingly complex and sophisticated, their **interpretability and explainability** also become more challenging. Interpretability and explainability is in most cases understood as “the ability to explain or to present in understandable terms to a human” (Linardatos et al., 2020, Fundamental Concepts and Background). The complex nature of advanced AI models, especially those based on deep learning, often results in them being labeled as "black boxes." This is due to the challenges in comprehending their internal mechanisms and the logic behind their decisions. Such opacity can affect confidence in AI systems and slow their integration into important sectors such as healthcare, finance, and law, where the ability to explain the mechanisms behind is absolutely critical. (Linardatos et al., 2020)

The **computational complexity and resource requirements** of AI learning processes present another significant challenge in the development and deployment of AI models. Many state-of-the-art AI models, particularly deep learning architectures, require vast computational resources for training and inference. (Thompson et al., 2020)

This high demand for computational power can lead to increased costs, energy consumption, and environmental impact (Magubane, 2023), potentially limiting the accessibility and scalability of AI technologies. If the development of AI were not stopped, or not developed effectively in terms of resource consumption, it would lead to ever-increasing resource consumption and thus an ever-increasing problem. Of course, many industries and scientists have already recognized this problem and are working on ways to avoid the resource and energy consumption of artificial intelligence. (Wu et al., 2022)

5. AI and societal biases

Another major problem one has to consider and which will be considered separately and specifically in the course of this paper, is the bias of artificial intelligence. Since it is a human-built machine with human inputs and human data, it is quite clear that our biases, opinions, etc. are incorporated into the machines' algorithms.

As artificial intelligence systems become more pervasive in our daily lives, their potential impact on society, both positive and negative, must be carefully considered (Ntoutsi et al., 2020). Already today, several cases have come to light where biases have significantly influenced decisions and represent the extent of biases in artificial intelligence for our society and in the everyday life.

- In 2018, Amazon stopped using an AI recruiting tool because it showed gender bias towards male candidates for technical jobs. The AI, trained on a decade's worth of resumes mostly from men, learned to favor male candidates. This even went so far, that the tool was downgrading resumes containing the word „women's". (Reuters, 2018)

- In 2019, a study revealed that a healthcare algorithm, used by US hospitals and affecting over 200 million people, was inadvertently biased against black patients. The algorithm predicted patients' extra medical care needs based on their healthcare cost history, not directly on race. Yet, as black patients often incurred lower costs than white patients with the same conditions, the algorithm disproportionately favored white patients. (Vartan, 2019)

- But one of the biggest and most known biases found in AI-Systems so far was in COMPAS. That’s an algorithm utilized across U.S. state courts to estimate the risk of criminals reoffending, that has been scrutinized for racial bias. Despite Equivant, the software's creator, denying these claims, a ProPublica investigation suggested otherwise. The analysis concluded that the software is as reliable as random, untrained internet users, indicating significant flaws. Furthermore, the results showed a disparity in risk prediction, overestimating the likelihood of black defendants reoffending while underestimating the same for white defendants. In fact, black defendants were nearly twice as likely to be incorrectly categorized as high risk compared to their white counterparts. (Larson et al., 2016)

These and many more examples show that it’s important to examine the origins of biases in AI systems, explore the consequences of biased AI implementations, discuss strategies for mitigating and reinforcing biases in AI algorithms, and consider the limitations in addressing and detecting such biases. By understanding and confronting these challenges, one can work towards developing more equitable and responsible AI systems that better serve the diverse needs of society and be aware of the constant „danger“ of bias a great part of the world we live in.

5.1. Understanding biases in AI systems

Biases in AI systems can arise from various sources, including the data used for training, the design of the algorithms themselves, or the social context in which these systems are deployed. Often, biased data is the primary culprit, as AI systems learn from and model the patterns and relationships present in the training data. If the data contains underlying biases, whether due to historical injustices, sampling errors, or other factors, these biases may be inadvertently learned and perpetuated by the AI system. (Srinivasan & Chander, 2021)

Additionally, biases may also emerge from the design choices and assumptions made by developers during the creation of AI algorithms, reflecting their conscious or unconscious beliefs and values. In this respect, the tool ChatGPT, which was released in 2023, had major problems. The language model was said to be politically left-wing, and some right-wing content was simply not reflected (Baum & Villasenor, 2023). As this is by far the fastest growing platform in the world, caution is definitely warranted, and it shows the importance of discussing biases in AI.

Last but not least and as already shown by various examples, biased AI systems can have far-reaching and detrimental consequences, particularly when employed in such high-stakes decision-making contexts like hiring, lending, healthcare, and criminal justice. This can lead to the reinforcement existing inequalities, perpetuate stereotypes, and negatively impact the lives of individuals from marginalized or underrepresented groups. (Lloyd, 2018)

5.2. Strategies for mitigating and reinforcing biases in AI algorithms

Due to the significant impact of biases in algorithms researchers and practitioners have developed a range of strategies and techniques to address and mitigate biases in AI systems. These include data preprocessing methods, which aim to balance the representation of different groups within the training data; algorithmic fairness techniques, which incorporate fairness constraints or objectives directly into the learning process; and post-processing methods, which adjust the outputs of AI systems to reduce disparities between groups. (Mehta, 2022)

However, this is also a much more far-reaching problem than just from a technical perspective. Therefore a interdisciplinary approach is needed that also incorporates perspectives from fields such as sociology, ethics, or human rights that might help to identify and address potential biases and their implications more effectively. These fields have been studying the issue of biases for some time, and can help incorporate these findings into the development of AI.

IBM, one of the first companies to address AI and its ethical implications, advocates e.g. for the following 5 points (Dortch & Hobson, 2021) :

- Strengthen AI literacy throughout society
- Require assessments and testing for high-risk AI systems
- Require AI transparency through disclosure
- Require mechanisms for consumer insight and feedback
- Establish universal use limitations of AI and adopt responsible licensing practices.

It can quickly be seen that these points go far beyond technical boundaries and include the whole of society. While these points would definitely help combat human prejudice against AI, it remains to be seen to what extent prejudice within an AI's decision-making power can also be eradicated?

So despite huge progress in identifying and reducing biases in AI systems, numerous obstacles persist. To start, pinpointing and measuring biases is a nuanced and context-specific endeavor, given that varying scenarios call for different interpretations of fairness and bias. Furthermore, there can be a tension between fairness and other performance indicators, complicating the task of simultaneously optimizing AI systems for all targeted goals. So combating biases in AI systems calls for an overarching emphasis on transparency, accountability, and ethical factors throughout the AI development cycle. This comprehensive approach may call for shifts in organizational ethos and procedures, as exemplified by IBM. (Lee et al., 2019)

6. Ethical considerations of AI

Consequently, the rapid advancements and widespread applications of artificial intelligence have given rise to numerous ethical questions and concerns. As AI systems increasingly influence our lives, it is crucial to explore and address these ethical questions.

AI systems, particularly those that rely on large-scale data collection and analysis, can pose significant risks to **privacy and personal autonomy**. The widespread use of AI-enabled surveillance technologies, such as facial recognition e.g., has raised concerns about the erosion of privacy and the potential for abuse by governments or corporations. To address such concerns, policymakers, technologists, and civil society must work together to establish appropriate legal frameworks, technological safeguards, and public oversight mechanisms to balance the benefits of AI with the need to protect individual privacy and civil liberties. (Kerry, 2020)

As previously discussed, biased AI systems can perpetuate and exacerbate existing **societal inequalities**, leading to unfair treatment and discrimination. Therefore ethical considerations in AI decision-making involve developing transparent, fair, and accountable AI systems that respect the rights and dignity of all individuals, regardless of their background or social status. This requires a commitment of politics, economics and society in understanding and addressing the sources of bias in AI systems, incorporating diverse perspectives in the design and evaluation of AI technologies, and engaging in ongoing monitoring and auditing of AI systems to ensure fairness and equity.

Furthermore, the development and deployment of increasingly powerful and autonomous AI systems in the society raises concerns about the general safety and the potential for unintended consequences of AI. It involves a host of potential risks, many of which might result from unforeseen implications of their programming or misuse (Amodei et al., 2016).

Therefore, ethical considerations in AI applications also include the development of transparent and interpretable AI systems that allow for meaningful human oversight and control, as well as the establishment of legal and regulatory frameworks that ensure compliance with ethical norms and protect the rights of affected individuals. (Dastani & Yazdanpanah, 2023; Ashrafian, 2015)

As has also already been established, AI systems can sometimes make mistakes, leading to **undesirable or harmful outcomes**. As such, it would be crucial to establish clear lines of accountability and responsibility in AI applications. This involves determining who should be held responsible when AI systems fail, whether it be developers, users, or even the AI systems themselves.

This raises one of the most pressing ethical questions in the field of AI: The question whether robots need or should have rights? What may sound like a joke to some at first is slowly but surely becoming a hot topic of discussion in the field of artificial intelligence. Do robots need rights, if so, what should they look like and on the basis of which moral principles should they be enacted? How would this change responsibility, or can systems take responsibility at all? (Gunkel, 2018; Gunkel, 2018)

Another big problem for many people is the potential of AI technologies to **disrupt labor markets** and reshape the global economy, leading to concerns about job displacement, wage inequality, and social unrest. While AI-driven automation may create new job opportunities and increase productivity, it may

also render certain jobs obsolete, disproportionately impacting low-skilled workers and exacerbating existing economic disparities (Abrardi et al., 2019). Admittedly, it must be said that all technical revolutions in the history of mankind have contributed to a widening of the gap between rich and poor.

Accordingly ethical considerations in AI require policymakers, businesses, and educational institutions to anticipate and address these potential impacts, developing policies and programs that promote workforce adaptation, social safety nets, and equitable economic growth in the age of AI (Ernst, Merola & Samaan, 2019).

This shows the variety of topics and areas that scientists, companies and society are dealing with in relation to artificial intelligences and their ethical implications and questions.

7. Current and future applications of AI

As AI technologies continue to evolve and advance, they hold the potential to revolutionize the way we live, work, and interact with one another. It's therefore helpful to be aware of the current and future applications of AI, including an overview of AI applications across industries, the limitations shaping the AI landscape, emerging AI technologies and their potential impact. AI technologies have already been adopted across a wide range of industries, if not nearly everyone, transforming traditional processes and enabling new capabilities.

In **healthcare**, AI systems are utilized for tasks such as diagnostics, drug discovery, and personalized medicine. An illustration of this is Recursion Pharmaceuticals, a firm known for pioneering transformative treatments by integrating automation and machine learning. They are one of the leading AI-powered, technology-focused biotech firms globally, with the capacity to assess thousands of chemical reactions daily in the quest for novel medicines. (Allen & Nilsson, 2022)

The **automotive industry** employs AI in the development of self-driving vehicles, while AI-powered virtual assistants and chatbots have become commonplace in customer service and communication applications. The most well-known example of this is the globally renowned company Tesla, led by Elon Musk (Ajitha & Nagra, 2021). In the automotive industry, however, it's not just about smart cars, but also the achievement of smart factories that promise more efficient workflows. (Breunig et al., 2017)

AI also finds its applications in sectors such as the **defense industry**. In the case of Palantir, in form of an AI analysis platform, that is used to process battlefield data as quickly as possible and draw appropriate conclusions from it (Palantir, 2023). It is certain that the current Ukraine conflict was significantly influenced by the use of this software (Mashur, 2023). Thus, AI systems are being deployed in even the most critical and dangerous industries.

These examples, among many others, showcase the versatility and potential of AI technologies to reshape the way we approach problems and deliver innovative solutions across diverse sectors. Nevertheless, it must be clearly stated that the progress of artificial intelligence also raises very significant questions.

7.1. Emerging AI technologies and their potential impact

As AI research progresses, new technologies and techniques continue to emerge every single day, offering novel ways to address current limitations and expand the scope of AI applications. Some promising areas of research include transfer learning, which enables AI systems to leverage knowledge from one domain to apply it to another (Shin'ya et al., 2022); neuromorphic computing, which aims to develop AI hardware that mimics the architecture and efficiency of the human brain (Fraunhofer, 2023); and artificial general intelligence (AGI), also known as Strong-AI, which seeks to develop AI systems capable of performing any intellectual task that a human can do (IBM, 2023).

These and other emerging AI technologies hold the potential to dramatically impact various aspects of the artificial intelligence landscape which is still heavily impacted from the last achievements. This shows how fast-moving the topic of artificial intelligence is. As soon as a study or similar is published, its content is actually almost outdated again. This makes research in the field of AI all the more difficult but also all the more important in order not to lose track of current progress.

7.2. The role of regulation and policy in shaping the future of AI

This rapid development and deployment of AI technologies necessitates the creation and implementation of appropriate regulatory frameworks and policies to ensure that AI systems are developed and used responsibly and ethically. Governments, industry stakeholders, and civil society must work in joint forces to develop regulations that address issues such as data protection, algorithmic fairness, transparency, and accountability.

Policies must strike a balance between fostering innovation and ensuring that AI technologies do not exacerbate existing societal disparities or create new risks. By proactively engaging with the ethical, legal, and social implications of AI, policymakers and stakeholders can help shape a future in which AI technologies are harnessed for the greater good of all. There is a growing focus on AI ethics, with 2023 expected to be a year of acceleration in AI ethics legislation (Gorden, 2022). This is in response to the increasing development, deployment, and interaction with AI technologies by companies.

As of 2023, there are several regulations and frameworks in place for artificial intelligence. On January 26, 2023, the National Institute of Standards and Technology (NIST), an agency of the US Department of Commerce, released its Artificial Intelligence Risk Management Framework 1.0. This is a voluntary, non-sector-specific, use-case-agnostic guide for technology companies that are designing, developing, deploying, or using AI systems to help manage the many risks of AI (NIST, 2023). The EU is also currently working on a suitable framework for the use, classification and handling of AI, the so-called AI-ACT. (European Parliament, 2023)

In summary, the current regulations and frameworks for AI are designed to manage the risks associated with AI, ensure compliance, spur innovation, and uphold ethical standards in AI development and use. However, as quickly becomes clear, there is still a great need for regulations and frameworks around AI. We are only at the beginning of the development of these extraordinary systems, but also of their framework conditions.

8. Conclusion

Artificial intelligence has demonstrated tremendous potential to transform various aspects of our lives, offering solutions to complex problems and opening up new avenues for innovation. However, realizing the full potential of AI also requires a deep understanding of its underlying principles, limitations, and ethical implications.

Throughout this discussion, the key principles of AI were presented. These principles form the foundation of AI systems, enabling them to exhibit human-like cognitive abilities and facilitate their application across a wide range of tasks and domains. Alongside these principles, also the challenges and limitations inherent in AI learning processes, were featured.

As AI continues to permeate various industries, understanding its limitations becomes crucial for the responsible development and deployment of AI systems. By acknowledging and addressing these limitations, researchers, developers, and policymakers can work towards creating AI systems that are not only more robust and effective but also adhere to ethical norms and societal values. This involves tackling challenges related to biases, privacy, accountability, and safety, among others, ensuring that AI technologies are designed and implemented with the interests of all stakeholders in mind.

The future of AI research and applications is marked by both exciting prospects and formidable challenges. As AI technologies continue to advance, new opportunities emerge for their application across various industries, with the potential to revolutionize namely every sector. At the same time, AI researchers and practitioners must grapple with the ethical and societal implications of their work, striving to minimize the risks and unintended consequences associated with AI deployment.

In conclusion, the key to harnessing the transformative power of AI lies in our collective understanding of its principles, limitations, and ethical considerations. By embracing a holistic and responsible approach to AI development and deployment, we can ensure that AI technologies contribute positively to our society, economy, and environment, paving the way for a future marked by progress, prosperity, and inclusivity.

9. Bibliography

Abrardi, L., Cambini, C. & Rondi, L. (2019). The economics of artificial intelligence: A survey. *Robert Schuman Centre for Advanced Studies Research Paper No. RSCAS*, *58*.

Allen, A. & Nilsson, L. (2021). The Drug Factory: Industrializing How New Drugs Are Found. *SLAS DISCOVERY: Advancing the Science of Drug Discovery, 26(9), 1076-1078.*

Aggarwal, K., Mijwil, M., Garg, S., Al-Mistarehi, A. W. et al. (2022). Has the Future Started? The Current Growth of Artificial Intelligence, Machine Learning, and Deep Learning. *Iraqi Journal for Computer Science and Mathematics. 3. 115-123.*

Ajitha, P.V. & Nagra, A. (2021). An Overview of Artificial Intelligence in Automobile Industry – A Case Study on Tesla Cars. *Solid State Technology. Volume: 64 Issue: 2.*

Amodei, D., Olah, C., Steinhardt, J., Christiano, P. et al. (2016). Concrete problems in AI safety. *Working paper paper led by Google Brain researchers.* Last Accessed 17.06.2023: https://arxiv.org/pdf/1606.06565.pdf%20http://arxiv.org/abs/1606.06565.pdfv

Ashrafian, H. (2015). Artificial intelligence and robot responsibilities: Innovating beyond rights. *Science and engineering ethics, 21, 317-326.*

Baum, J. & Villasenor, J. (2023). The politics of AI: ChatGPT and political bias. *Brookings*. [Online] Last Accessed 17.06.2023: https://www.brookings.edu/blog/techtank/2023/05/08/the-politics-of-ai-chatgpt-and-political-bias/

Barkhausen, B. (2023). KI-Pionier hält Künstliche Intelligenz für gefährlicher als den Klimawandel. *Gründerszene - Business*. [Online] Last Accessed 17.06.2023: https://www.businessinsider.de/gruenderszene/business/ki-pionier-ki-gefaehrlicher-als-klimawandel/

Breunig, M., Kässer, M., Klein, H. & Stein, J. P. (2017). Building smarter cars with smarter factories: How AI will change the auto business. *McKinsey Digital, McKinsey & Company*. [Online] Last Accessed 17.06.2023: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/building-smarter-cars

Britannica (2022). John McCarthy. *Encyclopedia Britannica*. [Online] Last Accessed 17.06.2023: https://www.britannica.com/biography/John-McCarthy

Budach, L., Feuerpfeil, M., Ihde, N., Nathansen, A., et al. (2022). The Effects of Data Quality on Machine Learning Performance. *arXiv preprint arXiv:2207.14529*.

Chowdhury, G. (2003). Natural language processing. *Annual Review of Information Science and Technology, 37. pp. 51-89.*

Dastani, M. & Yazdanpanah, V. (2023). Responsibility of AI Systems. *AI & Soc **38**, 843–852.*

Dennis, M. A. (2023). Marvin Minsky. *Encyclopedia Britannica*. [Online] Last Accessed 17.06.2023: https://www.britannica.com/biography/Marvin-Lee-Minsky

Dortch, A. & Hobson, S. (2021). Mitigating Bias in Artificial Intelligence. *IBM Policy*. [Online] Last Accessed 17.06.2023: https://www.ibm.com/policy/mitigating-ai-bias/

eLife (2021). Can AI learn to adapt?. *eLife Magazin*. [Online] Last Accessed 17.06.2023: https://elifesciences.org/digests/66273/can-ai-learn-to-adapt

Ernst, E., Merola, R. & Samaan, D. (2019). Economics of Artificial Intelligence: Implications for the Future of Work. *IZA Journal of Labor Policy, 9 (1)*.

European Parliament (2020). The ethics of artificial intelligence: Issues and initiatives. *EPRS | European Parliamentary Research Service - STUDY Panel for the Future of Science and Technology - Scientific Foresight Unit (STOA) PE 634.452.*

European Parliament (2022). Artificial intelligence: threats and opportunities. *European Parliament News.* [Online] Last Accessed 17.06.2023: https://www.europarl.europa.eu/news/en/headlines/eu-affairs/20230427STO83302/europe-day-2023-celebrating-european-unity

European Parliament (2023). AI Act: a step closer to the first rules on Artificial Intelligence**.** *European Parliament News.* [Online] Last Accessed 17.06.2023: https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence

Fraunhofer (2023). Neuromorphic Computing. *Fraunhofer IPMS*. [Online] Last Accessed 17.06.2023: https://www.ipms.fraunhofer.de/de/Strategic-Research-Areas/Neuromorphic-Computing.html

Goel, A. K. (2021). Looking back, looking ahead: Symbolic versus connectionist AI. *AI Magazine 42: 83–85.*

Gorden, C. (2022). 2023 Will Be The Year Of AI Ethics Legislation Acceleration. *Forbes Magazine*. [Online] Last Accessed 17.06.2023: https://www.forbes.com/sites/cindygordon/2022/12/28/2023-will-be-the-year-of-ai-ethics-legislation-acceleration/

Google (2022). Generalization. *Google Machine Learning Grundlagenkurse.* [Online] Last Accessed 17.06.2023: https://developers.google.com/machine-learning/crash-course/generalization/video-lecture?hl=de

Gunkel, D. J. (2018). The other question: can and should robots have rights?. *Ethics and Information Technology*, *20*, *87-99.*

Gunkel, D. J. (2018). Robot rights. *The MIT Press.* ISBN: 9780262038621

Howard, D., Eiben, A.E., Kennedy, D.F., Mouret, J. et al. (2019). Evolving embodied intelligence from materials to machines. *Nat Mach Intell **1**, 12–19.*

IBM (2023). What is strong AI?. *IBM*. [Online] Last Accessed 17.06.2023: https://www.ibm.com/topics/strong-ai

Kavlakoglu, E. (2020). AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What's the Difference? *IBM Blog*. [Online] Last Accessed 17.06.2023: https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks

Kerry, C. F. (2020). Protecting privacy in an AI-driven world. *Brookings*. [Online] Last Accessed 17.06.2023: https://www.brookings.edu/research/protecting-privacy-in-an-ai-driven-world/

Larson, J., Mattu, S., Kirchner, L. & Angwin, J. (2016). How We Analyzed the COMPAS Recidivism Algorithm. *ProPublica*. [Online] Last Accessed 17.06.2023: https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

Lakemeyer, G. & Nebel, B. (1992). Foundations of Knowledge Representation and Reasoning. *Part of the* **Lecture Notes in Computer Science** *book series (LNAI,volume 810)*

Lee, N. T., Resnick, P. & Barton, G. (2019). Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms. *Brookings*. [Online] Last Accessed 17.06.2023: https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/

Lloyd, K. (2018). Bias amplification in artificial intelligence systems. [Online] Last Accessed 17.06.2023: https://arxiv.org/pdf/1809.07842.pdf

Linardatos, P., Papastefanopoulos, V., Kotsiantis S. (2021). Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy. 23(1):18.*

McCarthy, J. (2007). What is artificial Intelligence?. *Computer Science Department Stanford University.* [Online] Last Accessed 17.06.2023: https://www.diochnos.com/about/McCarthyWhatisAI.pdf

McKendrick, J. (2021). The Data Paradox: Artificial Intelligence Needs Data; Data Needs AI. *Forbes Magazine.* [Online] Last Accessed 17.06.2023: https://www.forbes.com/sites/joemckendrick/2021/06/27/the-data-paradox-artificial-intelligence-needs-data-data-needs-ai/

Magubane, N. (2023). The hidden costs of AI: Impending energy and resource strain. *Penn Today, University of Pennsylvania.* [Online] Last Accessed 17.06.2023: https://penntoday.upenn.edu/news/hidden-costs-ai-impending-energy-and-resource-strain

Makridakis, S. (2017). The Forthcoming Artificial Intelligence (AI) Revolution: Its Impact on Society and Firms. *Futures, 90. pp. 46-60.*

Masuhr, N. (2023). Die Invasion der Ukraine nach einem Jahr - Ein militärischer Rück- und Ausblick. *RusslandAnalysen, 432. pp. 5-10. ,*

Mehta, S. (2022). A guide to different bias mitigation techniques in machine learning. *AIM Analytics India Magazine.* [Online] Last Accessed 17.06.2023: https://analyticsindiamag.com/a-guide-to-different-bias-mitigation-techniques-in-machine-learning/

Metz, C. (2023). The Godfather of A.I.' Leaves Google and Warns of Danger Ahead. *New York Times.* [Online] Last Accessed 17.06.2023: https://www.nytimes.com/2023/05/01/technology/ai-google-chatbot-engineer-quits-hinton.html

Microsoft (2018). The global impact of AI across industries. *Microsoft Transform.* [Online] Last Accessed 17.06.2023: https://news.microsoft.com/transform/the-global-impact-of-ai-across-industries/

Mutasa, S., Sun, S., Ha, R. (2020). Understanding artificial intelligence based radiology studies: What is overfitting? *Clin Imaging, 65. pp. 96-99.*

NIST (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). *National Institute of Standards and Technology.* [Online] Last Accessed 17.06.2023: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf

Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V. et al. (2020). Bias in data-driven artificial intelligence systems—An introductory survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(3), e1356.*

Palantir (2023). Palantir AIP. *Palantir.* [Online] Last Accessed 17.06.2023: https://www.palantir.com/platforms/aip/

Reuters (2018). Amazon ditched AI recruiting tool that favored men for technical jobs. *Guardian Magazin.* [Online] Last Accessed 17.06.2023: https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine

Roser, M. (2023). Here's how experts see AI developing over the coming years. *WEF World Economic Forum.* [Online] Last Accessed 17.06.2023: https://www.weforum.org/agenda/2023/02/experts-ai-developing-over-the-coming-years

Roy, N., Posner, I., Barfoot, T., Beaudoin, P. et al. (2021). From Machine Learning to Robotics: Challenges and Opportunities for Embodied Intelligence. *ArXiv, abs/2110.15245.*

Samson, O. (2017). Deep learning weekly piece: the differences between AI, ML, and DL. *Medium.* [Online] Last Accessed 17.06.2023: https://towardsdatascience.com/deep-learning-weekly-piece-the-differences-between-ai-ml-and-dl-b6a203b70698

Schuett, J. (2019). A Legal Definition of AI. *Goethe University Frankfurt - SSRN Electronic Journal.* DOI:10.2139/ssrn.3453632

Schuett, J. (2021). Defining the Scope of AI Regulations. *Forthcoming in Law, Innovation and Technology, Legal Priorities Project Working Paper Series No. 9.*

SEP (2023). Alan Turing. *Stanford Encyclopedia of Philosophy.* [Online] Last Accessed 17.06.2023: https://plato.stanford.edu/Entries/turing/

Shin'ya Y., Sekitoshi K., Atsutoshi K., Daiki C. et al. (2022). Transfer Learning with Pre-trained Conditional Generative Models. *Kyoto University.* arXiv:2204.12833v2

Siemens (2023). Intelligent produktion med Artificial Intelligence i industrien. *Siemens*. [Online] Last Accessed 17.06.2023: https://www.siemens.com/dk/da/produkter/industri/fokusomraader/artificial-intelligence.html

Srinivasan, R., & Chander, A. (2021). Biases in AI systems. *Communications of the ACM, 64, 44 - 49.*

Starzyk, Janusz. (2008). Motivation in Embodied Intelligence. *In book: Frontiers in Robotics, Automation and Control.*

Thompson, N., Greenewald, K., Lee, K., & Manso, G. F. (2020). The Computational Limits of Deep Learning. *IDE Discussion Paper No. 1048.*

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind, Volume LIX, Issue 236, October 1950, pp. 433–460.*

Truitt, E. R. (2021). Surveillance, Companionship, and Entertainment: The Ancient History of Intelligent Machines. *The MIT Press Reader.* [Online] Last Accessed 17.06.2023: https://thereader.mitpress.mit.edu/the-ancient-history-of-intelligent-machines/

Toosi, A., Bottino, A. G., Saboury, B., Siegel, E. et al. (2021). A Brief History of AI: How to Prevent Another Winter (A Critical Review). *PET clinics, 16 (4), 449–469.*

Vartan, S. (2019). Racial Bias Found in a Major Health Care Risk Algorithm. *Scientific American*. [Online] Last Accessed 17.06.2023: https://www.scientificamerican.com/article/racial-bias-found-in-a-major-health-care-risk-algorithm/

Viharos, Z. & Kemény, Z. (2007). AI techniques in modelling, assignment, problem solving and optimization. *Engineering Applications of Artificial Intelligence. 20. 691-698*.

Wu, C. J., Raghavendra, R., Gupta, U., Acun, B. et al. (2022). Sustainable ai: Environmental implications, challenges and opportunities. *Proceedings of Machine Learning and Systems, 4, 795-813.*

Ying, X. (2019). An overview of overfitting and its solutions. *In Journal of physics: Conference series (Vol. 1168, p. 022022). IOP Publishing.*