The Turkish Online Journal of Design, Art and Communication
E-ISSN: 2146-5193
DOI: 10.7456/tojdac

Research Article / Araştırma Makalesi

CONCEPTUAL REVIEW OF ARTIFICIAL INTELLIGENCE; DIFFERENCES BETWEEN HUMAN AND MACHINE LEARNING

YAPAY ZEKANIN KAVRAMSAL İNCELENMESİ; İNSAN VE MAKİNE ÖĞRENİMİ ARASINDAKİ FARKLAR

Büşra Fadim SARIKAYA[1]

ORCID: B.F.S. 0000-0002-9492-7493

Corresponding author/Sorumlu yazar:
[1] Büşra Fadim Sarıkaya
Türk-Alman University, Türkiye
E-mail/E-posta: busra.sarikaya@tau.edu.tr

Received/Geliş tarihi: 03.04.2024

Benzerlik Oranı/Similarity Ratio: %3

Revision Requested/Revizyon talebi: 07.05.2024

Last revision received/Son revizyon teslimi: 20.05.2024

Accepted/Kabul tarihi: 25.05.2024

Etik Kurul İzni/ Ethics Committee Permission: Bu çalışma için etik kurul izni gerekmemektedir.

Citation/Atıf: Sarıkaya, B. F. (2024). Conceptual Review Of Artificial Intelligence; Differences Between Human And Machine Learning. The Turkish Online Journal of Design Art and Communication, 14 (3), 648-659. https://doi.org/10.7456/tojdac.1464262.

Abstract
Machine learning and artificial intelligence produce algorithms that appear to make "intelligent" decisions akin to humans but operate differently from human thought processes. Understanding the background of these recommendations is crucial for humans to make decisions based on machine suggestions. Since humans are oriented towards understanding human intelligence, it is not yet fully understood whether they can truly comprehend the "thinking" generated by machine learning or merely project human-like cognitive processes onto machines. In everyday life, systems designed to assist human tasks and decisions based on smart algorithms are increasingly encountered. These algorithms predominantly rely on machine learning technologies, enabling the discovery of previously unknown correlations by analyzing large amounts of data. The analysis of thousands of X-ray images of both healthy and diseased individuals by machines can be demonstrated. Making this system operational, it is necessary to determine the patterns distinguishing "healthy" images from those labeled as "diseased" and to find an algorithm defining the latter. Algorithms trained in this manner are utilized in various applications, such as in preselection for job applications or in communication via voice assistants. After the conceptual explanation of artificial intelligence in the first part of the study, the concepts of weak and strong artificial intelligence will be examined. Subsequently, after describing the subcategories of artificial intelligence, distinctions between human learning and machine learning will be addressed. The potential risks and opportunities created by machine learning will be discussed in the conclusion section.

Anahtar Kelimeler: Yapa Zeka, Makine Öğrenimi, İletişim, Ekilesim, Sanal Ortam.

Öz
Makine öğrenimi ve yapay zeka, insanlarınkine benzer *"akıllı"* kararlar verebiliyor gibi görünen, ancak insan düşüncesinden farklı işleyen algoritmalar üretmektedir. İnsanların makine önerilerine dayalı kararlar verebilmesi için bu önerilerin arka planını anlayabilmesi önem arz etmektedir. İnsanlar, insan zekasını anlamaya yönelik olduklarından, makine öğrenimi tarafından yaratılan *"düşünmeyi"* gerçekten anlayıp anlayamayacakları ya da yalnızca insan benzeri bilişsel süreçleri makinelere yansıtıp yansıtmadıkları henüz tam anlaşılmamıştır. Ayrıca yapay zekanın medya temsilleri, yapay zekayı gerçekte sahip olduğundan daha yüksek yetenekler ve insan benzerliği varmış gibi lanse etmektedir. Günlük hayatta, akıllı algoritmalar temelinde insan görevlerini ve kararlarını kolaylaştırmak için tasarlanmış yardım sistemleriyle giderek daha fazla karşılaşılmaktadır. Bu algoritmalar ağırlıklı olarak, büyük miktarda veriyi analiz ederek önceden bilinmeyen korelasyonları keşfetmeyi mümkün kılan makine öğrenimi teknolojilerine dayanmaktadır. Örnek olarak, hasta ve sağlıklı insanlara ait binlerce röntgen görüntüsünün makine tarafından analiz edilmesi gösterilebilmektedir. Bu sistemi çalışır hale getirmek için, *"sağlıklı"* olarak not düşülen görüntülerin hangi kalıplarla *"hasta"* olarak not düşülenlerden ayırt edilebileceğini belirlemek ve ikincisini tanımlayan bir algoritma bulmak gerekmektedir. Bu şekilde oluşturulan *"eğitilmiş"* algoritmalar, yalnızca tıbbi teşhisler için değil, aynı zamanda bir iş ilanı için başvuranların ön seçiminde veya iletişimde de sesli asistanlar yardımıyla olmak üzere çeşitli uygulama alanlarında kullanılmaktadır. Çalışmanın ilk bölümünde yapay zekâ kavramsal olarak açıklandıktan sonra, zayıf ve güçlü yapay zekâ kavramları irdelenecektir. Daha sonra, yapay zekânın alt kategorileri açıklandıktan sonra, insan öğrenmesi ve makine öğrenmesi arasındaki ayrımlar ele alınacaktır. Makine öğrenmesi ve derin öğrenme kavramlarının incelenmesinin ardından, sonuç bölümünde makine öğrenmesinin yarattığı potansiyel riskler ve fırsatlar tartışılacaktır.

Keywords: Artificial Intelligence, Machine Learning, Communication, Interaction, Virtuality.

INTRODUCTION

The genesis of AI traces back to the seminal works of pioneers such as Alan Turing and John McCarthy, who laid the theoretical groundwork for the development of intelligent systems. At its core, AI encompasses the endeavor to imbue machines with the capacity to perceive, reason, and act autonomously, thus mirroring the cognitive faculties of human beings. This pursuit has given rise to a spectrum of AI manifestations, ranging from rudimentary rule-based systems to sophisticated neural networks capable of complex pattern recognition (Turing, 1950; McCarthy, 1955). Alan Turing's seminal paper *"Computing Machinery and Intelligence"* published in 1950, proposed the famous Turing Test as a criterion for determining a machine's intelligence (Turing, 1950). John McCarthy, in his 1955 paper *"Proposal for the Dartmouth Summer Research Project on Artificial Intelligence"* coined the term *"artificial intelligence"* and outlined the goals and scope of the field, thereby catalyzing its development (McCarthy, 1955). At its essence, AI constitutes the pursuit of endowing machines with the ability to perceive, reason, and act autonomously, thereby emulating the cognitive capacities inherent to human beings (Russell & Norvig, 2016). This endeavor has precipitated a diverse spectrum of AI applications, ranging from rudimentary rule-based systems to sophisticated neural networks capable of intricate pattern recognition (LeCun et al., 2015; Goodfellow et al., 2016). At its essence, AI constitutes the pursuit of endowing machines with the ability to perceive, reason, and act autonomously, thereby emulating the cognitive capacities inherent to human beings (Russell & Norvig, 2016). This endeavor has precipitated a diverse spectrum of AI applications, ranging from rudimentary rule-based systems to sophisticated neural networks capable of intricate pattern recognition (LeCun et al., 2015; Goodfellow et al., 2016).

In the era of digital transformation, enterprises are amassing substantial volumes of novel data, commonly referred to as "big data." This data is characterized by its vast scale, diverse formats, and dynamic nature, undergoing constant changes. Traditional methods of data evaluation prove inadequate in harnessing its potential insights. Consequently, there is a growing reliance on artificial intelligence (AI) methods as elucidated by Fenske, Gutschmidt, and Grunert (2020). AI methodologies are recognized as pivotal tools for enhancing services, products and fostering innovations. This paper delves into the essence of AI, elucidating its significance when confronted with intricate data evaluation challenges. Unlike conventional programming paradigms, AI relinquishes the conventional approach wherein developers explicitly instruct each step of the data evaluation process. Instead, developers guide computer programs on how to autonomously discern processing rules from the data. The ensuing discourse provides a comprehensive examination of the term AI, with particular emphasis on the dichotomy between weak and strong AI.

In the discourse surrounding artificial intelligence (AI), scholars often draw a fundamental distinction between strong and weak AI, as articulated by Fenske et al. (2020). This categorization is pivotal in understanding the diverse manifestations of AI and its applications in contemporary society.

Weak And Strong Artificial Intelligence

Weak AI, as expounded by Fenske et al., finds ubiquitous application in various facets of daily life. Distinguished by its specialization in solving singular, well-defined problems, weak AI is prevalent in commonplace scenarios. Prominent examples include virtual assistant systems such as Apple's Siri or Amazon's Alexa or ChatGpt, which demonstrate proficiency in comprehending user inputs, particularly voice commands or text messages, and responding appropriately. Fenske et al. emphasize the imperative for these systems to accurately interpret and execute user instructions, underscoring the specificity inherent in weak AI applications. Weak AI does not act on its own initiative, but rather in order to solve a task set by humans using predetermined human solutions. Its ability to develop is correspondingly limited (Neuhöfer, 2023, p. 23).

Beyond personal assistants, weak AI manifests itself in the realm of vehicular technology, ranging from lane departure warning systems to fully autonomous driving. In these instances, automobiles are expected to navigate through road traffic independently, employing artificial intelligence methodologies. Moreover, within the framework of Industry 4.0, weak AI is actively employed in

quality monitoring during production processes and machinery maintenance. Recommendation systems, encompassing diverse domains such as music or product suggestions, further exemplify the application of weak AI. It is crucial to note that weak AI consistently revolves around tailored solutions for specific applications, as elucidated by Fenske et al (2020).

On the other hand, strong artificial intelligence (AI) transcends the limitations inherent in solving isolated problems and aspires to possess a multifaceted cognitive prowess. Strong AI is the term used to describe AI systems that develop new possible solutions beyond pre-programmability (Scherk et al., 2017). The attributes characterizing strong AI extend beyond mere problem-solving, encompassing logical reasoning, strategic application, puzzle resolution, decision-making in uncertain contexts, knowledge representation, planning, learning capabilities, and proficiency in natural language communication. A defining feature of strong AI, as posited by Schneider (2019), is the attainment of consciousness. The ongoing discourse among scientists revolves around the timeline for the potential realization of strong AI, with divergent views on whether such a milestone is attainable (ibid.). Presently, discussions about AI in practical applications predominantly pertain to weak AI. However, a nuanced understanding acknowledges that the classification of AI technologies extends beyond the binary categorization of strong or weak; instead, diverse sub-categories delineate the nuanced spectrum of AI technologies.

Subfields of Artificial Intelligence

Sub-fields of AI include machine learning as well as symbolic and probabilistic AI (Fenske et al., 2020). The differences and similarities of these fields are explained in more detail by Fenske et al. as follows:

1. Machine Learning

Artificial intelligence is often equated with machine learning. Although very important, machine learning is only a sub-field of AI. The essence of machine learning is that computer programs learn processing instructions based on data or training examples. In general, there are three different types of machine learning:

a. Supervised Learning

Supervised learning, constitutes a foundational paradigm within artificial intelligence (AI). This methodology necessitates the provision of training data to the AI, wherein the data includes predefined solutions, such as categories relevant to subsequent data processing tasks. For instance, when tasked with discerning customer loyalty or predicting the potential transition to another manufacturer, the AI requires training data comprising exemplars of historical customer behavior, such as data about prior purchases, as delineated by Fenske et al.. However, beyond mere examples, it is imperative to furnish information indicating the fidelity of these customers to the company. Through exposure to this comprehensive training dataset, the AI discerns decision rules, facilitating the subsequent resolution of analogous challenges presented by new datasets.

Supervised learning encompasses classical techniques, with classification and regression being primary exemplars. In classification, distinct categories are assigned to datasets, thereby enabling the differentiation between, for instance, loyal and churning customers. Conversely, regression within supervised learning entails the estimation of metric variables, such as weight, price, or time, for novel datasets (ibid.). This dichotomy of supervised learning encapsulates a methodical approach pivotal in enhancing the cognitive capabilities of AI systems, particularly in scenarios necessitating the classification or estimation of diverse data parameters.

b. Unsupervised Learning

Unsupervised learning represents a distinct paradigm within the realm of artificial intelligence, juxtaposed against the structured framework of supervised learning. A defining feature of unsupervised learning, articulated by Fenske et al., lies in the absence of predefined solution patterns within the dataset. In contrast to supervised learning, where training data includes explicit solutions, unsupervised learning is characterized by the pursuit of latent patterns inherent in the data, imperceptible to direct observation.

Various techniques constitute the arsenal of unsupervised learning, each tailored to unveil latent structures within datasets Fenske et al., 2020). Notably, association analysis stands out as a prominent method employed in diverse domains, exemplified by its application in shopping cart analysis. This technique endeavors to unveil patterns regarding which products are frequently purchased together, thereby providing insights into consumer behavior (ibid.). Additionally, clustering, another significant facet of unsupervised learning, is instrumental in delineating customer segments from voluminous datasets. This segmentation process is crucial for discerning inherent patterns and relationships within the data, contributing to a more nuanced understanding of complex datasets. The dynamic nature of unsupervised learning positions it as a pivotal tool in scenarios where the emphasis is on discovering intrinsic patterns and structures absent explicit guidance, thereby underscoring its indispensable role in contemporary artificial intelligence applications.

c. Reinforcement learning

According to Fenske et al., unlike supervised and unsupervised machine learning, reinforcement learning does not require an explicit data set to learn or extract new knowledge. Instead, AI learns through trial and error. It is at this stage that AI is expected to learn, through rewards and punishments, which behavior is desirable or undesirable in which situation. In other words, the AI generates experience values on its own and is then expected to use them productively. Furthermore, there are often so-called sequential problems that need to be learned or solved. This means that the individual steps to solve such a problem have a temporal relationship or dependency on each other. These are often referred to as sequential problems that need to be learned or solved (Fenske et al., 2020).

There are many different methods in the field of machine learning. One of the most prominent methods at the moment is (artificial) neural networks. Their name comes from the fact that they are based on the way biological neural networks work. Artificial neural networks can be used for many different problems. These include supervised, unsupervised, and reinforcement learning methods (Fenske et al., 2020). Applications of neural networks range from speech and video/image processing to the analysis of biological and chemical data sets. When these neural networks are particularly large or "deep", they are referred to as deep learning (ibid.). Especially deep neural networks are often not understandable to humans, which can limit trust in AI.

DIFFERENCES BETWEEN HUMAN AND MACHINE LEARNING

An essential facet of AI pertains to the process of learning, a fundamental mechanism underpinning adaptive behavior and intelligence (Russell & Norvig, 2016). Human learning, characterized by its capacity for abstraction, generalization, and creative synthesis (Kolb, 1984), stands in stark contrast to mechanized learning algorithms. While humans effortlessly assimilate knowledge from diverse sensory inputs and experiential interactions (Siegler, 1998), machines rely on meticulously crafted algorithms and vast datasets to discern patterns and derive insights (Bishop, 2006). Understanding the disparities between human and machine learning is paramount for delineating the capabilities and limitations of AI systems (Nilsson, 1998), as well as charting the trajectory of future advancements in artificial intelligence (Mitchell, 1997).

Harald Lesch's claim that, certainly, AI will not replace humans seems strange at first, since AI research was initially aimed at replicating human intelligence (Lesch, Schwartz, 2020). Moreover, AI was often used precisely where human thinking and decision-making needed to be supplemented, enhanced, or replaced. The term AI is as multifaceted as the term human intelligence. For example in psychology, AI is not uniformly defined and encompasses a wide range of cognitive abilities (thinking, problem-solving, speech, computation) that are defined and operationalized in different ways in different psychological approaches (Asendorpf, 2004).

Developers of earlier forms of artificial intelligence worked on from the 1960s onwards were aware that the technologies used would never succeed in replicating human intelligence exactly. However, major attempts, such as neural networks (artificial neuron networks as a sub-form of AI), turned towards the functioning of human intelligence and tried to simulate it using various methods such as learning systems or formalized cognitive models (Minsky, Papert, 1969, Boden, 2006). The latter aimed at

quantifying and specifying the mechanisms and processes of human intelligence. Based on this description, the models were then intended to be reproduced in computer programs. However, the massive availability of a wide variety of data provided, for example, by social media and increasing computing capacities have made another form of data-based AI possible (ibid.). Machine learning is not based on the application of assumed mechanisms and models of human cognitive processes (Michie & Spiegelhalter, 1994). Therefore, what and how the machine learns remains unclear to outsiders, and in many cases even to those who program it.

On the other hand, a typical example from the field of visual learning can be considered as follows: An algorithm has been developed that can distinguish between bottles and glasses in pictures and label them correctly. For this purpose, the machine is given a dataset *"graded"* by humans as a starting point for learning (Michie & Spiegelhalter, 1994). Pictures of glasses and pictures of bottles are labeled in this way so that the algorithm starts with human input and learns based on human categorizations. These include, for example, that shapes tapering towards the top are more likely to be bottles, while shapes expanding towards the top are more likely to be cups. According to Michie and Spiegelhalter, there are always borderline cases, making it necessary to relearn by correcting the machine classification and optimizing the learning outcome.

At this point, it remains an open question whether the machine makes (correct) classifications based on features similar to humans or otherwise. Furthermore, it can be ruled out that it has succeeded in labeling bottles and glasses correctly because the AI will not understand the linguistic meaning of "upward tapering shapes" or "upward expanding shapes" since it has no such understanding (Michie, Spiegelhalter, 1994). Therefore, what the machine has learned, if it has learned to sort the bottles and glasses correctly, remains unclear and the machine cannot express it descriptively. This is because the machine does not have the semantic vocabulary to say what the meaning of the criteria mentioned and used is. Moreover, even in tasks that are supposed to be so simple, an algorithm typically makes far more mistakes than a human - an indication that artificial systems are often not as intelligent as society assumes, according to scientists. This is particularly evident in the algorithms circulating on the internet that humans can easily solve, such as distinguishing between a dog's head and a cake, and how difficult they are for AI machines to solve (see Figure 1).

Figure 1. The image of a dog and a cupcake that Artificial Intelligence cannot yet distinguish from each other.

Many bot tests exploit this weakness of algorithms. These bot tests are designed to prevent AI machines from accessing online banking and personal data sites. Until now, AI machines have not been able to recognize these tests because they do not yet have cognitive thinking, and as a result, they have been blocked from accessing such sites. The fact that the abilities learned by machines are significantly different from what goes on in the human mind when recognizing objects has been demonstrated by the so-called Adversarial Examples (Gilpin, Bau, Yuan et al., 2018). This phenomenon explains that even changing individual image pixels can cause images to be misclassified, despite not affecting human semantically controlled perception and object recognition. For human perception, even if only one pixel is changed in a picture of an elephant, the elephant in the picture remains an elephant; whereas for an algorithm, even a small pixel change in a picture of an elephant can transform the elephant into a

completely different and even unrecognizable object. Studies on street sign recognition show that small changes that humans do not see can lead to a completely different - and objectively incorrect - classification for the machine (see Figure 2).

Figure 2. Different perceptions of street signs by humans and machines.

Machine Learning and Deep Learning

Today, smart solutions are mostly programmed manually. A smartphone, for example, contains more than ten million lines of code. Current developments in artificial intelligence point to a paradigm shift: instead of manually coding process steps, the ability to learn programs them. With the help of machine learning, AI can learn patterns from a large number of example situations and transfer them to new, similar situations. The greatest AI achievements are currently based on deep neural networks (deep learning), where a large number of artificial neurons process input information in several layers and provide the result in the output (Kersting, Tresp, 2019). Even with machine learning, humans still have to do programming, with one difference: ready-made solutions are no longer programmed by humans. Instead, artificial intelligence develops programs that learn the solution based on training data.

Deep learning, a sub-discipline of machine learning, has recently made significant breakthroughs in many fields (ibid.). According to Kersting and Tresp, these advances underpin many of the achievements in these application areas. For example, modern translation and image recognition systems are unthinkable without deep learning. Neural networks have a high expressive power, in simple terms the ability to approximate any continuous function with arbitrary accuracy. At the same time, however, it is often possible to adapt a network that has been extensively trained for a specific task to a new task with little effort (Kersting & Tresp, 2019).

As a result of the studies, it is claimed that the understanding of the theoretical foundations of deep learning is still partially incomplete (ibid.). In addition, optimal solutions imply a process of trial and error based on experience and heuristics. The terms artificial intelligence, machine learning, and deep learning are often used interchangeably (Kersting & Tresp, 2019). Artificial intelligence identifies challenges to be solved and develops solutions. Machine learning focuses on learning solutions. Deep learning currently offers some of the most powerful approaches to machine learning (see Figure 3).

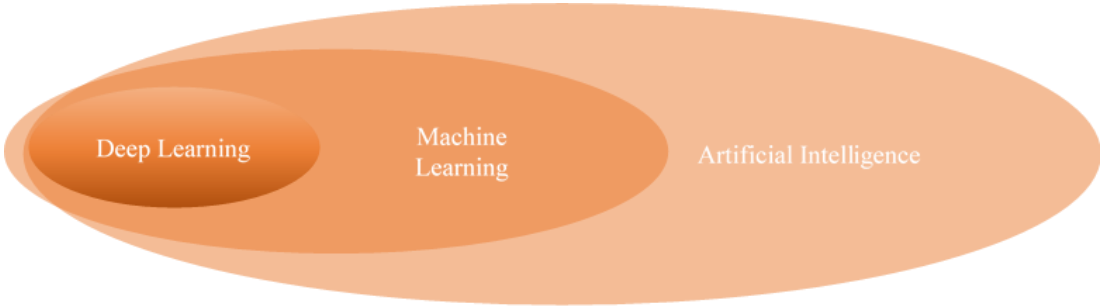

Figure 3. Relationship between Artificial Intelligence, Machine Learning, and Deep Learning.

Possibilities and Limits of Utilizing Machine Thinking

In light of all these findings, it is understood that human and machine thinking operate differently from each other. The question of to what extent humans are aware of this difference and to what extent they need to understand the algorithmic procedure sufficiently is important at this point. Regarding the second question, it can be argued that a certain understanding may be a prerequisite for acceptance and intention to use. Algorithms are used in decision support systems where human decisions are prepared by machine recommendations - for example, a pre-selection of suitable candidates in the context of filling a position, or a diagnostic recommendation based on the analysis of radiological data - or these decisions are made directly by the algorithm itself. In addition, however, to be able to trust the system and its recommendations, it is necessary, at best, to have some idea of how the system processes data and on what basis it makes decisions.

When interacting with other people, several innate mechanisms, referred to in various psychological sub-disciplines as common ground (Clark, 1996), mentalizing (Frith & Frith, 2006), or perspective-taking (Fussell & Krauss, 1992), allow us to understand the other person. Thanks to these mechanisms, which despite their different names ultimately function in a similar way, people can predict what others are thinking, how they feel, and how they will behave in many situations. These constructs describe how people can put themselves in others' shoes and understand what others know and how they come to their conclusions based on thinking patterns and brain structures that are similar in all people (Fussel & Krauss, 1992).

In particular, the Theory of Mind construct explains that people have an intuitive understanding of what others perceive, know, believe, and experience (Premack & Premack, 1995). On the one hand, this is made possible by the fact that we form theories and assumptions about other people's emotions through experience and socialization. On the other hand, people can simulate other people's experiences, i.e. directly imagine how other people feel (ibid.). This allows for a meaningful understanding of other people's behavior or decisions. For example, a person who goes into every room of the house several times in a row and goes to different corners is not considered to be crazy, but rather someone who is looking for something. But this kind of understanding relies heavily on the assumption that the other person's functioning is similar to one's own. In a non-human decision-making system, however, this is not the case. How a talking robot *"feels"* cannot be understood, and it is more difficult to attribute behavior patterns to potential causes.

There are two possibilities when interacting with a technical system (Ngo et al., 2020). Either existing basic assumptions about human thinking are transferred to the machine - which may necessarily lead to erroneous results, or one informs oneself about how an algorithm works and in this respect deviates from human processing mechanisms. However, many people seem to have only a vague idea of how intelligent algorithms work. Early studies show that some people have only developed a rudimentary understanding and know that a system only works based on collected data (Ngo, Kunkel, Ziegler, 2020). Other studies show that a wide range of ideas prevails, but these are mostly derived from media reports (DeVito, Birnholtz, Hancock et al., 2018).

However, a study by Horstmann and Krämer shows that information from the media can lead to serious misconceptions about artificial systems and algorithms (Horstmann & Krämer, 2019). According to this study, especially fictional media content leads to high expectations about the capabilities of social robots. A quantitative survey revealed that the depiction of robots in movies leads to a belief that the first robots also have high capabilities in real life. However, non-fiction media contributions can also convey false impressions, as anecdotal examples show. For example, the human-looking robot Sofia (Hanson Robotics) is regularly shown on talk shows interacting with an interlocutor without any explanation that the dialogue is predetermined and scripted. Conversations with Sofia are therefore by no means as autonomous as the presentation suggests.

Social cues are also a powerful influencing factor that can lead to incorrect assessments of an AI's intelligence. These include, for example, a human-like appearance, a similar language, the ability to

interact, the assumption of familiar roles in interpersonal communication, and other behavioral elements. If an AI is embodied in a human-like form or uses a human-like language (communicating robots, virtual assistants, voice assistants such as Alexa or Siri), people behave in a human-like manner in interaction. This has been extensively described and researched by Reeves and Nass (1996) within the framework of the so-called media equation assumptions *("media equals real life"*) (Reeves, Nass, 1996). Numerous studies have shown that when social cues are present, people unconsciously and automatically execute social scripts (e.g. polite behavior, self-presentation) that are otherwise only practiced in interaction with humans (Nass, Moon, 2000; Krämer, 2008.). This is true even if there is also a conscious level of denial that the artificial interaction partner requires social treatment.

Therefore, these unconscious mechanisms of interaction with a perceived human-like interaction partner can also lead to the intelligence of this partner being overestimated or falsely perceived as human-like. One way to counter this is through the efforts of the explainable AI community (Voosen, 2017). This community aims to open the black box, that is, to solve the riddle of the processes within AI. To do this, it is necessary to translate what learning algorithms produce in terms of contexts and decision patterns into a language that humans can understand. Explainable AI is aimed at developers themselves, who often cannot understand how systems learn. On the other hand, explainable AI is addressed to the end-users of technologies, so that they too can understand what inputs the system relies on during use and how it arrives at its outputs and recommendations (ibid.). For example, the counterfactual method is used in this sense for both groups of interlocutors. Here, relevant input data (language, text, or images) are systematically changed and it is observed how the output result changes as a result (Beck, Riggs & Burns, 2011). In this way, it is possible - even for laypeople - to understand how the system works, for example in the automatic calculation of insurance fees.

Disparities and Implications in the Distinctive Attributes of Human and Machine Intelligence

Substantial distinctions exist between human intelligence and artificial intelligence grounded in machine learning. These distinctions are often underestimated by human users in three distinct manners. Initially, influenced by media portrayals, individuals are prone to erroneously deduce that artificial systems possess elevated levels of intelligence and a more human-like cognitive capacity than is accurate. Secondly, due to evolutionary psychological predispositions, humans inherently tend to conceptualize intelligence in a human-like manner. Thirdly, based on social cues, there is a tendency to ascribe a level of human-like intelligence to intelligent systems in direct interaction, surpassing what may be deemed contextually appropriate.

What ramifications arise from the distinctiveness of machine intelligence and humans' constrained capacity to perceive beyond this distinctiveness? The implications of the inherent otherness of machine intelligence, coupled with the constrained human ability to surpass this otherness, warrant thorough examination. Primarily, the presumption of human resemblance to machine intelligence engenders heightened expectations regarding the capabilities of algorithmic systems. This elevation in expectations may evoke heightened apprehension, but, upon a discerning evaluation of reality, it can also lead to disillusionment, as discussed by Horstmann and Krämer (2019). Nevertheless, of greater significance is the observation that artificial intelligence (AI) often remains inscrutable to the majority of users, akin to the incomprehensibility attributed to the intelligence of extraterrestrial entities. This lack of comprehension assumes heightened significance, as it has the potential to pose risks. Specifically, the incomprehensibility of AI systems may render users unable to fathom the underlying principles guiding decisions affecting them or discern the extent to which their data is being collected, stored, and processed by the system. Concerning the initial consideration, recommendations provided by systems must possess transparency to enable users to assess the reliability of the recommendation. This issue is notably scrutinized in domains such as the medical field, where machine learning plays an expanding role in formulating radiological diagnoses (refer to Nensa et al., 2019). In this context, diagnosticians must enhance their comprehension of the system's underlying processes, either through augmenting their digital literacy or acquiring proficiency in artificial intelligence literacy. Alternatively, a system could elucidate its methodology by employing explainable AI principles. Enhanced comprehension of the underlying data processed by AI systems could contribute to a more elucidated perspective on the second aspect elucidated earlier. This understanding becomes imperative when users recognize that numerous

systems utilizing algorithms for intelligent decision-making store and assess their data. Such awareness empowers users to safeguard themselves as necessary. Consequently, while the prospective augmentation of AI holds manifold advantages—such as expeditious decision preparation and more dependable recommendations due to the rapid analysis of extensive datasets—there persists an unresolved question about the level of trust users can place in AI recommendations, and whether adherence to such recommendations genuinely results in improved decision outcomes.

The establishment of trust in artificial intelligence (AI) among individuals is deemed crucial. On one hand, this approach aids users in achieving a better understanding of the foundational principles of the system. Alternatively, trust in AI can be reinforced through other means, such as the formal testing and certification of systems. Examples enumerated, such as medical applications, not only underscore the facilitative role of intelligent algorithms and AI in our daily lives, rendering occupational tasks more reliable but also emphasize that greater acceptance of these procedures is contingent upon their comprehensibility. Proficiency in understanding the behaviors of others, as experienced in interpersonal interactions, is contingent upon possessing a robust skill set.

RISKS AND OPPORTUNITIES OF MACHINE LEARNING AND CONCLUSION

As with any technological advancement, there exist not only risks but also numerous opportunities. Technological developments create opportunities and possibilities in various categories, including daily life, the creative sector, learning skills, and the well-being of children and youth. For instance, chatbots like ChatGPT provide low-threshold access to a free and comprehensive tool that operates on all devices, catering to children, youth, and adults. Many obstacles encountered in daily life can be more easily overcome with technical support through such applications. For instance, individuals with dyslexia, a language-based learning disorder, can independently correct their written texts using these artificial intelligence applications. Additionally, artificial intelligence applications can be utilized for the completion of various petitions or official documents. Many children and young individuals exhibit reluctance to personally visit counseling services, possibly due to associated barriers such as shame, time constraints, or travel expenses (Sindermann, Albrich, 2023). In such instances, chatbots enhanced with artificial intelligence technologies can serve individuals with an anonymous, free, and time-efficient privilege. Leveraging data collected from real counseling services, these chatbots provide recommendations and thereby encourage those seeking advice to benefit from their services. Simultaneously, artificial intelligence technologies serve as a significant facilitator in overcoming language barriers. For example, with the assistance of Google Translate AI, even printed texts can now be translated in real-time using a smartphone camera with high reliability. Moreover, individuals with disabilities can potentially benefit from artificial intelligence technologies in the long term. Speech and text recognition AI systems, for instance, continue to improve in converting texts to speech or vice versa, enhancing accessibility. Consequently, it is conceivable that in the future, AI could translate spoken language into sign language.

On the other hand, anything adhering to clear and undisputed rules can be efficiently processed by artificial intelligence (AI) and consequently reproduced flawlessly. Evolving technologies and AI applications not only encourage independent learning but also create new opportunities for educators. For instance, AI can significantly streamline the creation of instructional materials. Whether it be textual exercises in mathematics, assignments encompassing science courses, or worksheets for art and physical education, everything can be generated within a matter of seconds. Admittedly, some editing and adaptation of outputs are necessary, but in any case, the information and documents produced by AI serve as a foundation—or at least a catalyst for individual ideas. Additionally, newly acquired competencies can later be utilized, for example, in the development of innovative didactic approaches. In the context of academic studies, AI technologies can provide benefits in the research phase of generating scientific studies. At this juncture, a crucial question arises: If AI can accomplish so much for users, does it foster creativity or hinder its development when it is still in its nascent stages?

This question needs to be further evaluated in the context of different research endeavors and, of course, in terms of different purposes for which artificial intelligence (AI) technologies are utilized. However, one undeniable fact is that users can swiftly produce results in their tasks with these technologies without

overly contemplating the process. Nevertheless, *"While machine creativity may not replace human creativity, the future will be about creative collaboration between humans and artificial intelligence. Encouraging creativity is imperative for the healthy progression of this man-machine collaboration, necessitating the development of applications that promote creativity. In addition to representatives from the business and scientific realms, the involvement of stakeholders from education, culture, art, and civil society becomes even more crucial for this collaborative endeavor, emphasizing the importance of developing practical application examples"* (Scharf & Tödte, 2020).

As with all the gains brought about by digitalization, individuals inevitably become concerned with the question of how the targeted use will evolve. The central inquiry here revolves around how media literacy, participation and inclusion, and media education processes can be initiated, thereby ensuring self-sufficiency. According to Dieter Baacke's concept of media education in the 1970s, there has been a profound shift in the pedagogical approach to media, as it previously predominantly focused on the idea of protection and the negative impacts of media. Baacke's approach views media users as active and self-influencing actors. According to Baacke, media literacy consists of four sub-domains: media criticism, media literacy, media usage, and media design; all of which can be similarly applied to the field of artificial intelligence (see Gross & Röllecke, 2022).

Media criticism refers to an individual's ability to reflect on and scrutinize his own knowledge about the media (ibid.). For this reason, the positioning of any content generated by artificial intelligence should be examined within the context of artificial intelligence. In particular, the potential risks and dangers of artificial intelligence for children and young individuals should not be underestimated. Specifically, in the context of cyberbullying, Deepfakes, created by artificial intelligence, can elevate cyberbullying to a new level by attacking individuals and producing demeaning expressions in artificially generated videos.

While the mentioned risks may appear distant at the moment, the potential of artificial intelligence is far from exhausted. Furthermore, children and young individuals need to be critical and sensitive to data protection when exposing their data. The amalgamation of big data and machine learning poses risks for children and young people if they disclose their data without critically considering the implications. From a media criticism perspective, young individuals should not only critically question their own media behaviors but also engage in discussions about the purpose of content published on the internet and the goals pursued with it.

However, to be able to be critical, one first needs media literacy, which involves knowledge about various systems, media, and technologies (ibid.). This can be challenging when it comes to artificial intelligence because it is not always clearly discernible whether a tool in question utilizes artificial intelligence. Looking at the legislative aspect, the European Commission called for the establishment of a legal framework last June for labeling and making the content of artificial intelligence explicitly understandable to consumers (see Krempl, 2023). From a media criticism perspective, such labeling requirements could be beneficial as this information may encourage young individuals to think critically.

Active media usage is another subdomain of Baacke's media literacy model, defining the active utilization of media (Baacke, 1998). There is a need for secure and educational spaces where children and young individuals can experiment with artificial intelligence and interact with it. To dispel the mysteries of these domains and acquire their own experiences in these areas, they require the guidance and support of expert media education personnel.

This is followed by the fourth domain of the media literacy model: Media design. Children and young individuals can use creative and innovative approaches to create new things and enhance their skills in the process. The focus here is not just on pure usage but on creating innovative content based on their interests. At this point, media education still faces challenges. Potential dangers, methods of use, and issues of social appropriateness hinder the creativity of media education. In this regard, different media pedagogical approaches need to be brought together and harmonized with each other (see Süss et al., 2010).

When intersected with action-oriented media education as seen in artificial intelligence, educational media protection for children and young individuals has the potential to bring forth concepts, impulses, and ethical foundations that inspire both worlds. Ultimately, artificial intelligence is now an integral part of daily life and should be profitably addressed in educational endeavors to empower children and young individuals for the future.

REFERENCES

Asendorpf, J. (2004). Psychologie der Persönlichkeit, Heidelberg

Baacke, D. (1998). Zum Konzept und zur Operationalisierung von Medienkompetenz, https://www.produktivemedienarbeit.de/ressourcen/bibliothek/fachartikel/baacke_operationalisierung.shtml.

Beck, S. R., Riggs, K. J. & Burns, P. (2011). Multiple developments in counterfactual thinking. Understanding counterfactuals, understanding causation, p. 110-122. Oxford Academics.

Bishop, C. M. (2006). Pattern Recognition And Machine Learning. Springer.

Clark, H. (1996). Using Language, Cambridge.

DeVito, M., Birnholtz, J. & Hancock et al. (2018). How People Form Folk Theories of Social Media Feeds and What It Means for How We Study Self, Proceedings of the ACM Conference on Human Factors in Computing Systems p. 1–12. https://socialmedia.northwestern.edu/wp-content/uploads/2018/01/FolkTheoryFormation_CHI2018.pdf.

Fenske, O., Gutschmidt, A. & Grunert, H. (2020). Was ist Künstliche Intelligenz?. Whitepaper-Serie des Zentrums für Künstliche Intelligenz in MV Ausgabe 1. Rostock.

Frith, Ch. & Frith, U. (2006). How we predict what other people are going to do. Brain Research. 1079/1, p. 36–46.

Fussell, S. & Krauss, M. (1992). Coordination of knowledge in communication: Effects of speakers' assumptions about others' knowledge, Journal of Personality and Social Psychology, 62/ 3, p. 378–391.

Gilpin, L., Bau, D., Yuan, et al. (2018). Explaining Explanations: An Overview of Interpretability of Machine Learning, IEE 5th International Conference on Data Science and Advanced Analytics (DSAA), https://doi.org/10.1109/DSAA.2018.00018.

Goodfellow, I., Bengio, Y. & Courville, A. (2016). Deep Learning. MIT Press.

Gross, F. & Röllecke, R. (2022). Dieter Baacke Preis Handbuch 17. Love, Hate & More. Gesellschaft für Medienpädagogik und Kommunikationskultur der Bundesrepublik Deutschland e. V. (GMK).

Horstmann, A. & Krämer, N. (2019). Great Expectations? Relation of Previous Experiences With Social Robots in Real Life or the Media and Expectancies Based on Qualitative and Quantitative Assessment. Frontiers in Psychology, 10, p. 939, https://doi.org/10.3389/fpsyg.2019.00939.

Kersting, K. & Tresp, V. (2019). Maschinelles und Tiefes Lernen. Digitale Welt 3. 32–34 (2019). https://doi.org/10.1007/s42354-019-0209-4.

Kolb, D. A. (1984). Experiential learning: Experience as the source of learning and development. Prentice-Hall.

Krempl, S. (2023). Manipulationsgefahr: EU-Kommission fordert rasch Kennzeichnung von KI-Inhalten in: heise online. https:// www.heise.de/news/Manipulationsgefahr-EU-Kommission-fordert-rasch-Kennzeichnung-von-KI-Inhalten-9179211.html, (2023, December 12).

Krämer, N., Artelt, A.Z Geminn et al. (2019). KI-basierte Sprachassistenten im Alltag: Forschungsbedarf aus informatischer, psychologischer, ethischer und rechtlicher Sicht. Universität Duisburg-Essen, https://doi.org/10.17185/duepublico/70571.

LeCun, Y. & Bengio, Y., Hinton, G. (2015). Deep learning. Nature, 521(7553), p. 436-444.

Lesch, H. & Schwartz, T. (2020). Unberechenbar. Das Leben ist mehr als eine Gleichung, Freiburg.

McCarthy, John. (1955). Proposal for the Dartmouth Summer Research Project on Artificial Intelligence in AI Magazine, 27/4, 2006.

Michie, D. & Spiegelhalter, D. (1994). Machine Learning, Neural and Statistical Classification. Ellis Horwood Series in Artificial Intelligence, New York.

Minsky, M., Papert, S. Perceptrons. (1969). An Introduction to Computational Geometry. Boston;

Margaret A. Boden, 2006. Mind as Machine. A History of Cognitive Science. Oxford.
Mitchell, T. M. (1997). Machine Learning. McGraw Hill.
Nass, C., Moon, Y. (2000). Machines and mindlessness: Social responses to computers. Journal of Social Issues, 56/1, p. 81–103; Krämer, N., 2008. Soziale Wirkungen von virtuellen Helfern. Stuttgart.
Nensa, F., Demircioglu, A., Rischpler, Ch. (2019). Artificial Intelligence in Nuclear Medicine, Journal of Nuclear Medicine 60/1, p. 1–9, https://doi.org/10.2967/jnumed.118.220590.
Neuhöfer, S. (2023). Grundrechtsfähigkeit Künstlicher Intelligenz. Duncker&Humblot. Berlin.
Ngo, T., Kunkel, J., Ziegler, J. (2020). Exploring Mental Models of Recommender Systems: A Qualitative Study. UMAP ’20: Proceedings of the 28th Conference on User Modeling, Adaptation and Personalization, p. 183–191.
Nilsson, N. J. (1998). Artificial Intelligence: A New Synthesis. Morgan Kaufmann.
Premack, D. & James Premack, A. (1995). Origins of human social competence, in Michael S. Gazzaniga (Ed.), The cognitive neurosciences, p. 205–218, Cambridge.
Reeves, B., Nass, C. (1996). The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. Cambridge.
Russell, S. J., & Norvig, P. (2016). Artificial Intelligence: A Modern Approach. Pearson.
Scharf, I. & Tödte, J. (2020). Digitalisierung mit Kultureller Bildung gestalten. In: Kulturelle Bildung .Online. https://www.kubi-online.de/ artikel/digitalisierung-kultureller-bildung-gestalten. 18.12.2023. Schneider, S., 2019. Artificial You. Princeton University Press, New Jersey.
Scherk, J., Pöchhacker, G. & Wagner, K. (2017). Künstliche Intelligenz, Artificial Intelligence. Pöchhacker Innovation Consulting. Linz.
Siegler, R. S. (1998). Children's thinking (3rd ed.). Prentice Hall.
Sindermann, M., Albrich, K. (2023). Chancen und Risiken: Künstliche Intelligenz im Spannungsfeld des Kinder- und Jugendmedienschutzes, BzKJAKTUELL 4/2023.
Süss, D.; Lampert, C. & Wijnen, C. (2010),. Medienpädagogische Ansätze: Grundhaltungen und ihre Konsequenzen. In: Medienpädagogik. VS Verlag für Sozialwissenschaften, ISBN 978-3-658-19823-7.
Turing, A. M. (1950). Cumputing Machinery and Intelligence. Mind A Quarterly Review Of Psychology And Philosophy.
Voosen, P. (2017). How AI detectives are cracking open the black box of deep learning. As neural nets push into science, researchers probe back. Science, https://www.science.org/content/article/how-ai-detectives-are-cracking-open-black-box-deep-learning.