Perceptions, attitudes and trust toward artificial intelligence — An assessment of the public opinion

5 authors, including:

Gian Luca Liehner
RWTH Aachen University
14 PUBLICATIONS **51** CITATIONS

SEE PROFILE

Hannah Biermann
RWTH Aachen University
30 PUBLICATIONS **290** CITATIONS

SEE PROFILE

Alexander Hick
3 PUBLICATIONS **108** CITATIONS

SEE PROFILE

Philipp Brauner
RWTH Aachen University
132 PUBLICATIONS **1,553** CITATIONS

SEE PROFILE

Perceptions, Attitudes and Trust Towards Artificial Intelligence — An Assessment of the Public Opinion

Gian Luca Liehner, Hannah Biermann, Alexander Hick, Philipp Brauner, and Martina Ziefle

Human-Computer Interaction Center, RWTH Aachen University, Germany

ABSTRACT

Over the last couple of years, artificial intelligence (AI)—namely machine learning algorithms—has rapidly entered our daily lives. Applications can be found in medicine, law, finance, production, education, mobility, and entertainment. To achieve this, a large amount of research has been undertaken, to optimize algorithms that by learning from data are able to process natural language, recognize objects through computer vision, interact with their environment with the help of robotics, or take autonomous decisions without the help of human input. With that, AI is acquiring core human capabilities raising the question of the impact of AI use on our society and its individuals. To form a basis for addressing those questions, it is crucial to investigate the public perception of artificial intelligence. This area of research is however often overlooked as with the fast development of AI technologies demands and wishes of individuals are often neglected. To counteract this, our study focuses on the public's perception, attitudes, and trust towards artificial intelligence. To that end, we followed a two-step research approach. We first conducted semi-structured interviews which laid the foundation for an online questionnaire. Building upon the interviews, we designed an online questionnaire (N = 124) in which in addition to user diversity factors such as belief in a dangerous world, sensitivity to threat, and technology optimism, we asked respondents to rate prejudices, myths, risks, and chances about AI. Our results show that in general respondents view AI as a tool that can act independently, adapt, and help them in their daily lives. With that being said, respondents also indicate that they are not able to understand the underlying mechanisms of AI, and with this doubt, the maturity of the technology, leading to privacy concerns, fear of misuse, and security issues. While respondents are willing to use AI nevertheless, they are less willing to place their trust in the technology. From a user diversity point of view, we found, that both trust and use intention are correlated to the belief in a dangerous world and technology optimism. In summary, our research shows that while respondents are willing to use AI in their everyday lives, still some concerns remain that can impact their trust in the technology. Further research should explore the mediation of concerns to include them in a responsible development process that ensures a positive impact of AI on individuals' lives and our society.

Keywords: Artificial intelligence, Public perception, Acceptance, Mental models

INTRODUCTION

What is AI? There are numerous answers to this question and the scientific literature is filled with attempts to come up with a definitive one. Finding it, however, is not as straightforward as it might seem. Simply defining the term serves a different purpose than assessing what people think AI is. The former, though necessary to set the scope, gives away, allowing for the latter to take the stage. After all, what is the point of defining something which people perceive differently. The present paper addresses the gap between scientific endeavor and real-world application. We will first look at the existing literature regarding perceptions, attitudes, and trust towards AI. Then, we will address the gap, as it were, and finally, present our contribution to the topic. Artificial Intelligence is a term summarizing multiple subgroups of computational solutions. Among the most common methods are machine learning models, deep learning models, and convolutional neural networks. These tools can analyze data to find patterns. Based on these patterns they can draw conclusions and suggest an output. It applies a multi-step process to get to a predefined goal or come up with a related, but not predefined, goal (Shapiro, 1992; Collins et al., 2021).

In the last five years, there have been numerous publications on the topic of AI in every context imaginable (198,807 results on PubMed, 2023; 184,136 results on Web of Science, 2023; 76,449 results on Scopus, 2023). This vast number of publications shows the momentum and importance that AI has produced in the world of academia. At the same time, there has been an equally substantial increase in public attention and media representation (Nader et al., 2022; Ouchchy, Coin & Dubljević, 2020). In both instances, AI has become a buzzword of our time and it continues to grow (Royal Society, 2016). The general public now faces the challenge of keeping up with the amount of information about the topic in order to manage their expectations about its potential. A major contributing factor in this management is the public perception of AI. It can shape people's attitudes and ultimately influence their trust in AI itself and AI-based technology.

Generally, there are mixed attitudes towards AI in the general public (Fast & Horvitz, 2017). Some consider it revolutionary while others fear its inexplicability, that is, its black-box nature (Kieslich, Keller, & Starke, 2022; Kolasinka, Lauriola, & Quadrio, 2019). Current studies support evidence for a growing interest in AI and AI-based technologies. This interest is not confined to the academic world, but also found in the general public. Smart-/-phones, -watches, and -TVs, are almost omnipresent in our modern western culture and the trend is moving up (Google, 2002; Google, 2022). However, some argue that this increase in technology is moving at a pace that is poorly aligned with our ethical, legal, and societal values and principles (Jobin, Ienca, & Vayena, 2019). In a recent study, Bochniarz and colleagues (2021) investigated how high-school students evaluate AI and measured to what degree these students had feelings of cynical hostility towards AI. Their results indicated that people from this sample perceived AI as being more hostile and had a greater distrust towards AI when it was perceived as a threat or as being subjective i.e., governed by emotions. In a large-scale study

with more than 10,000 participants, Kelley and colleagues (2021) collected data on public opinion of AI from eight countries. In their study, the researchers divided respondents by sentiment groups. These four groups represented attitudes towards AI about 1) Excitement, 2) Usefulness, 3) Worry, and 4) Futuristic perceptions of its development. About 23% of the respondents were sorted into the worrying sentiment group, while only 12% were sorted into the useful group. Although around 35% of the respondents were sorted into none of the four sentiment groups. These numbers suggest that people worry more about AI than they appreciate its potential. This is supported by multiple studies which underline people's fear of job loss, privacy issues, or ethical dilemmas (Dietterich & Horvitz, 2015; International Federation of Robotics, 2018; Kieslich, Keller, & Starke, 2022). A different study, focusing on medical AI, found that people, on the one hand, think that certain AI technologies are already being used when in fact they are not, while on the other hand, they base their knowledge on misinformation provided by, often, inaccurate media representations (Stai et al., 2020). This misalignment of knowledge and understanding can lead to a misrepresentation of AI's factual and actual abilities and ultimately to an over/underestimation of AI-based technology. We set up this study to understand what and why people hold certain attitudes about AI, and how their beliefs influence their perceptions. While there is a broad range of large-scale studies investigating themes of public opinion, there remains a gap between these quantitative results and the individual perceptions, attitudes, and beliefs about AI. This study addresses this gap using a mixed-method approach in which we asked participants to rate prejudices, myths, risks, and chances about AI and assessed how these ultimately influenced their intention to trust and use it.

METHODOLOGICAL APPROACH

To investigate the public perception of artificial intelligence we followed a two-step research approach: first qualitative (semi-structured interviews), second quantitative (online questionnaire). To lay the foundation for the online questionnaire and gather core factors of AI perception, we asked respondents of the interviews (N = 8) to answer questions in relation to their associations, prejudices, and attitudes towards AI. We also queried chances and risks resulting from new developments in AI technologies. Respondents were furthermore asked to identify application fields of AI (e.g., care robots, drones, housekeeping assistants, image recognition, etc.) which were later condensed to overarching domains (e.g. medicine, warfare, assisted living, autonomous driving, etc.). We also asked respondents about potential roles (e.g., companion, tool, support instance, or advisor) that AI could take or for that matter not take. For instance, we found that roles like psychologists, caretakers, artists, and safety officers were seen as skeptical, whereas vehicle drivers, researchers, medical diagnostics or surgeons were roles more likely to be filled by AI. The interviews concluded with a query of respondents' use intention and trust towards AI.

A qualitative content analysis (following Mayring, 2010) allowed us to design the online questionnaire on the basis of the extracted factors and to

generate their respective items. Apart from socio-demographic factors, such as age, gender, education, and employment, the first part of the questionnaire was dedicated to the query personality traits such as the belief in a dangerous world (Duckitt et al., 2002, $\alpha = .78$), the sensitivity to threat (adapted from Kramer et al., 2020, $\alpha = .74$) and technology optimism ($\alpha = .78$). Derived from the interviews, we also developed items to assess respondents' prejudices—or perceived myths—towards AI (e.g., "Strong artificial intelligence can become a threat to humanity and our planet."). Furthermore, respondents were tasked with evaluating chances and risks resulting from rapid advances in the development of AI. Based on the qualitative items we derived items such as "*Artificial intelligence could lead to espionage*" for the risk-related scale or "*I think the use of strong artificial intelligence can ensure a higher standard of living*" for the scale related to chances. In addition to this, respondents were asked to rate the retrieved application areas from the qualitative study with regard to *valence* (the personal perception of AI use for a specific application domain) and *expectancy* (the likelihood of AI being used in this specific domain). Lastly, we included a measure of projected use intention of AI and perceived trust in AI in the questionnaire. All of the above-presented constructs were assessed on a 6-point Likert scale from 1 = "*do not agree at all*" to 6 = "*fully agree*".

The questionnaire itself was hosted on Qualtrics and distributed over the natural social networks of the investigators in Germany. The median response time for the questionnaire was about 23 Minutes. After the collection, the responses underwent a multi-step data cleaning process. First, incomplete responses were removed from the dataset (i.e., respondents that had not fully completed the survey). Subsequently, "speeders" were filtered out, based on the median survey duration. Next, "straightliners" that responded repeatetly with identical answers to a single scale were removed. Lastly, entries with incomplete data for socio-demographic factors were removed to allow for a complete sample description.

The analysis of the resulting dataset ($N = 124$) was carried out in R version 4.2.0. A significance level of $\alpha = .05$ was set for inferential analysis. For the validated and self-developed scales indices were calculated we examined their internal reliability using Cronbach's α. In case, scales were not condensable into indices (i.e., prejudices and myths, chances and risks), distribution of agreements level were reported. In order to present the evaluations of the different application areas in a comprehensible way, they were transferred into a criticality map by calculating the means for the various application domains. To access the relationship between personality factors (sensitivity to threat, belief in a dangerous world, and technology optimism) and acceptance factors (use intention and trust) Pearson's correlation coefficients were calculated and examined for significance.

SAMPLE

The cleaned dataset ($N = 124$) consists of 86 male (69.4%) and 38 female (30.6%) respondents with a combined mean age of 38.2 (*SD = 16.7, min = 19, max = 80*). With regard to education, 77 (62.1%) respondents

stated that they were in possession of a university degree (e.g., bachelor, master, Ph.D., etc.) and 40 (32.3%) had successfully completed a high school degree. The remaining 7 respondents declared that they had obtained a general certificate of secondary education. Most respondents ($N = 59$, 47.6%) reported to be in an employer-employee relationship, followed by students who amounted to 36.3% ($N = 45$). The remaining part were either school students, self-employed, pensioners, stay-at-home spouses, unemployed, or in the process of completing an apprenticeship (all groups < 5%). When looking at personality traits, we found that the sample had a rather low belief in a dangerous world ($M = 2.72$, $SD = 0.68$), an close to neutral sensitivity to threat ($M = 3.24$, $SD = 0.75$), and a rather high technology optimism ($M = 4.59$, $SD = 0.86$) — all measured on a scale from 1 to 6.

RESULTS

In a first stage of analysis, we focused on prejudices, myths, roles, chances, and risks regarding fast developments in AI technologies. Figure 1 shows the general perception of AI in terms of associations, prejudices, myths, and beliefs. Generally, respondents perceived AI as a tool, that can make independent decisions, however, is rather difficult to understand. This is reflected in their assessment of AI as something "unknown". Additionally, half of our sample perceive AI as science fiction robots. Taking a closer look at extreme scenarios sometimes depicted by media and business magnates (e.g., "With artificial intelligence we are summoning the demon", Elon Musk, 2014), we found that our respondents are more likely to disagree with prospects of AI taking over humanity: Only 8.9% believe that AI will take control of the world. However, more respondents (30.6%) stated to perceive AI as a threat to humanity.
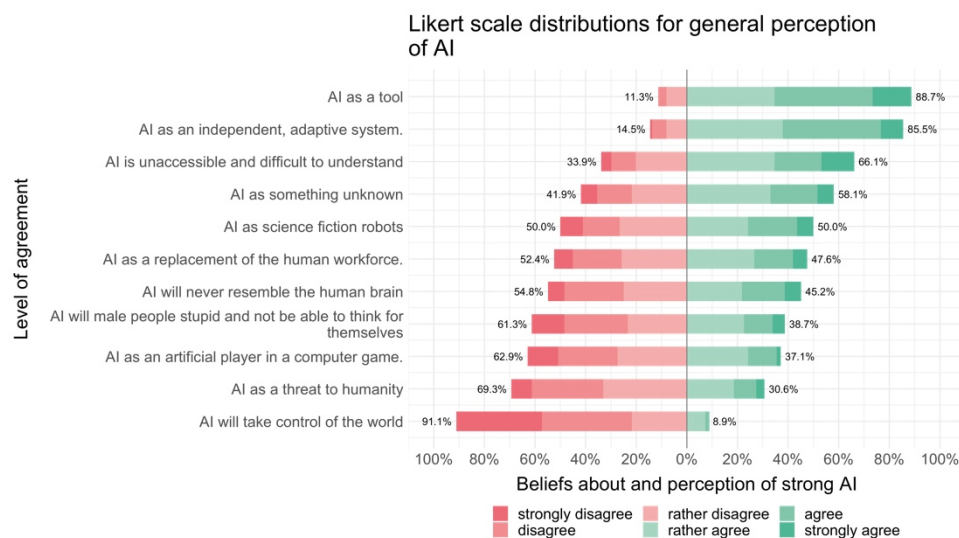
Figure 1: General perception of AI including associations, prejudices and myths.

Besides associations and prejudices, we investigated perceived chances and risks of AI. Firstly, we found that respondents considered chances in the use of AI to support people in their everyday life by, for instance, automating processes and thus reducing work time (> 90% agreement). Furthermore, respondents agreed (>70%) with the thought of AI promoting a higher standard of living and a general enhancement of life quality. Less agreement (51.6%) was recorded for the safe use of AI. Looking at risks, our respondents were mostly concerned with the possibility of AI abuse (i.e., to control people) with an agreement level of 85.5%. Likewise, the potential for unauthorized intrusion of AI systems (i.e., hacking) was also met with a high level of agreement (82.3%). Following a similar line of thought, we found high scores for privacy concerns, fear of espionage, and the irresponsible management of private data. On the other hand, respondents were also concerned with the lack of maturity of AI (79% agreement) and the, at this stage, only scarce legal basis for everyday use (74.2% agreement). Less agreement was found for overarching fears, such as the general risk for society, the threat for humanity or AI that takes a life of its own and causing harm in the process (< 45% agreement).

Besides the general perception of AI, we concluded our questionnaire by asking respondents to rate the likelihood (*expectancy*) and their personal evaluation (*valence*) of an AI technology to be used in a specific domain. We then contrasted both measurements in a criticality map. Figure 2 shows, that while respondents believe that AI technologies will appear in a series of domains, their evaluation of this development tends to be lower. Standing out are applications like Art & Music and Law Enforcements for which respondents don't appear to be able to forsee whether AI could be applied and rate
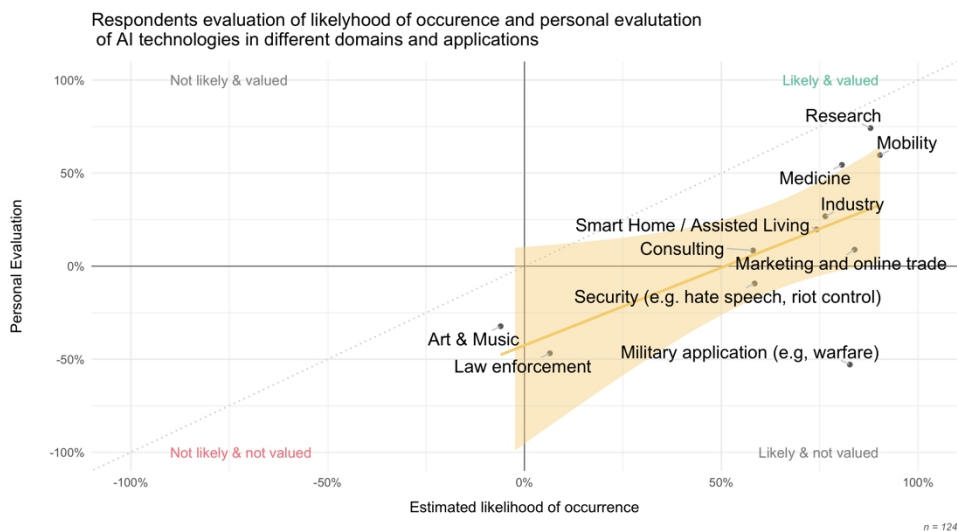
Figure 2: Evaluation of estimated likelihood of occurrence and perceived valence for the use of AI technologies in different domains.

them with a lower score on valence. Similar low ratings were found for military applications which are however expected to experience a high use of AI, similar to research, medicine, and industry. The highest rating in valence and expectancy were found for research and mobility.

In a last step of analysis, we examined the relationship between personality traits and intention to use AI on the one side, and trust in AI on the other. Here, we could identify a negative relationship between the belief in a dangerous world and both the intention to use AI ($r(122) = -.29, p < .001$) and the trust in AI ($r(122) = -.30, p < .001$). In other words, the higher people believe that our world is dangerous, the less they would trust or use AI technologies in general. A positive relationship was found between technology optimism and trust in AI ($r(122) = .47, p < .001$) as well as the intention to use AI ($r(122) = .34, p < .001$). Performing correlation analysis for the sensitivity to threat on the other hand did not yield any significant results.

DISCUSSION

In this paper, we investigated the perception, attitudes, and trust of the public towards artificial intelligence. We focused on individual mental models and used a mixed-method approach to capture user diversity and elaborate on the process of AI technology acceptance, that is, the intention to use said technology. To this end, our primary goal was to assess participants' belief in a dangerous world, their sensitivity to threat and technology optimism including prejudices, myths, risks, and chances about AI. While there exists a large body of research on the topic of public AI perception, individual attitudes, and beliefs have mostly been neglected for generalizability reasons. However, as the development of AI technology promises more individuality in e.g., medical care or recommendation systems, these beliefs lay the foundation of the adoption process.

Our results support the literature on the topic of public perception of AI that there is a lack of accurate understanding of the inner workings of AI. This lack might lead to an over/underestimation of its abilities and limitations, respectively. However, overall awareness of AI and contact with technologies based on it are prevalent in this sample and the scientific literature. This is no surprise given the fact that smartphones implement some form of AI e.g., machine learning for voice recognition or deep learning for face recognition software. These models, however, are typically hidden from the user and thus less prone to fearful attitudes or negative perceptions. In addition to this, most voice and face recognition software work surprisingly well and do not offer any opportunity for criticizing the component 'AI'.

The fact that half of our sample perceived AI as a science fiction robot, much like HAL 9000 or the Terminator (2001: A Space Odyssey, 1968; Terminator, 1984), is in line with some studies within the scientific literature (Dieter & Gessler, 2020). On the one hand, this kind of association makes sense because of the lack of accurate knowledge and understanding of technology itself. On the other hand, this raises the question of whether this part of our sample does, in fact, believe that a world depicted in the Terminator is probable. This would mean that believing in such a dangerous world

influences the overall likelihood estimation of negative outcomes regarding the development of AI. Although its introduction in the movie was after an apocalyptic event, AI seems to be seamlessly interwoven with the fabric of the movie, and thus, by extension reality. However, it remains unclear whether these depictions of so-called menacing AI result in actual fear rather than a quick association due to the popularity of the movies. In addition, the study conducted by Dieter and Gessler (2020) found that most movies depicting AI do so portraying it as friendly, helpful, and in close supportive interaction with humans. Support for this can also be found in the current study. In contrast to the finding that half of our participants associated AI with a science fiction robot, they thought it would be unlikely for an AI to take over the world. This means that even though they might have a subjectively negative association with the term AI, they do not believe that it truly can take over the world. From an intuitive standpoint, this sounds reasonable because there currently are no AI systems–let alone robots–that have the autonomy or learning transferability to take over any government or society. However, users might not look beyond this rather obvious fact. In the literature, AI experts are repeatedly found to hold more nuanced views on AI, while laypeople are typically less clear about how they think AI might affect them personally–as opposed to its impact on society itself–experts tend to be focused on specific tasks that AI can, or cannot, perform (O'Shaughnessy et al., 2022) and the given stakes of the situation (Kieslich, Lünich & Marcinkowski, 2021).

In terms of interpretability, our sample thought it is less obvious what AI does and how, in comparison to the question of what it is. Participants had a fair understanding of AI as such. This means they knew that it is a tool which can make independent decisions. However, it was less clear as to what was behind this ability. Furthermore, and in line with previous studies, this lack of interpretability had a less negative impact on the attitudes about AI when it was in high-stakes contexts, e.g., medical decision-making, where accuracy was a more relevant trade-off (Nussberger et al., 2022). In general, we can conclude that our study showed an association between negative attitudes about AI and trust towards it. With that in mind, there remains a fair amount of overlap between our results and existing studies. Overall, the current study supports the trend that people know about AI, that they form opinions pertaining to ethical, social, and personal dilemmas during high-stake contexts, and that there remains a gap between the factual understanding of what it is AI can and cannot do.

OUTLOOK

To close that gap, we suggest developing specifically designed communication strategies for AI as a concept and AI-based technologies. In these, the focus will lie on explaining to different stakeholders, e.g., end-users, policy makers, but also technology developers, which facts about the technology are important to accurately evaluate and positively implement it. As a starting point, we need to consider a more in-depth investigation into the adoption process and the acceptance of AI. For this purpose, future studies should focus on

perceived benefits and barriers of AI technology, user diversity, and personal stakes within the application context.

REFERENCES

Bochniarz, K. T., Czerwiński, S. K., Sawicki, A., & Atroszko, P. A. (2022). Attitudes to AI among high school students: Understanding distrust towards humans will not help us understand distrust towards AI. Personality and Individual Differences, 185, 111299.

Collins, C., Dennehy, D., Conboy, K., & Mikalef, P. (2021). Artificial intelligence in information systems research: A systematic literature review and research agenda. International Journal of Information Management, 60, 102383.

Dieter, D. G., & Gessler, E. C. (2021). A preferred reality: Film portrayals of robots and AI in popular science fiction. Journal of Science & Popular Culture, 4(1), 59–76.

Dietterich, T. G., & Horvitz, E. J. (2015). Rise of concerns about AI: reflections and directions. Communications of the ACM, 58(10), 38–40.

Duckitt, J., Wagner, C., Du Plessis, I., and Birum, I. (2002). The psychological bases of ideology and prejudice: Testing a dual process model. *Journal of Personality and Social Psychology* 83, 75–93. doi:10.1037/0022-3514.83.1.75

Fast, E., & Horvitz, E. (2017). Long-Term Trends in the Public Perception of Artificial Intelligence.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389–399.

Kelley, P. G., Yang, Y., Heldreth, C., Moessner, C., Sedley, A., Kramm, A., & Woodruff, A. (2021, July). Exciting, useful, worrying, futuristic: Public perception of artificial intelligence in 8 countries. In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (pp. 627–637).

Kieslich, K., Lünich, M., & Marcinkowski, F. (2021). The Threats of Artificial Intelligence Scale (TAI) Development, Measurement and Test Over Three ApplicationDomains. International Journal of Social Robotics, 13, 1563–1577.

Kieslich, K., Keller, B., & Starke, C. (2022). Artificial intelligence ethics by design. Evaluating public perception on the importance of ethical design principles of artificial intelligence. Big Data & Society, 9(1), 20539517221092956.

Kolasinska, A., Lauriola, I., & Quadrio, G. (2019, September). Do people believe in Artificial Intelligence? A cross-topic multicultural study. In Proceedings of the 5th EAI International Conference on Smart Objects and Technologies for Social Good (pp. 31–36).

Kramer, M. D., Patrick, C. J., Hettema, J. M., Moore, A. A., Sawyers, C. K., and Yancey, J. R. (2020). Quantifying Dispositional Fear as Threat Sensitivity: Development and Initial Validation of a Model- Based Scale Measure. *Assessment* 27, 533–546

Mayring, P. (2010). Qualitative Inhaltsanalyse. Grundlagen und Techniken. Weinheim und Basel: Beltz Juventa.

Mayring, P., & Fenzl, T. (2019). Qualitative inhaltsanalyse (pp. 633- 648). Springer Fachmedien Wiesbaden.

Nader, K., Toprac, P., Scott, S., & Baker, S. (2022). Public understanding of artificial intelligence through entertainment media. AI & society, 114.

Nussberger, A. M., Luo, L., Celis, L. E., & Crockett, M. J. (2022). Public attitudes value interpretability but prioritize accuracy in Artificial Intelligence. Nature Communications, 13(1), 5821.

O'Shaughnessy, M. R., Schiff, D. S., Varshney, L. R., Rozell, C. J., & Davenport, M. A. (2022). What governs attitudes toward artificial intelligence adoption and governance? Science and Public Policy, scac056.

Ouchchy, L., Coin, A., & Dubljević, V. (2020). AI in the headlines: the p ortrayal of the ethical issues of artificial intelligence in the media. AI & SOCIETY, 35, 927–936.

Shapiro, S. C. (1992). Encyclopedia of artificial intelligence second edition. New Jersey: A Wiley Interscience Publication.

Stai, B., Heller, N., McSweeney, S., Rickman, J., Blake, P., Vasdev, R., ... & Weight, C. (2020). Public perceptions of artificial intelligence and robotics in medicine. Journal of endourology, 34(10), 1041–1048.