machine learning & knowledge extraction

Article

More Capable, Less Benevolent: Trust Perceptions of AI Systems across Societal Contexts

Ekaterina Novozhilova [1,*], Kate Mays [2], Sejin Paik [1] and James E. Katz [1]

1 College of Communications, Boston University, Boston, MA 02215, USA
2 Department of Community Development and Applied Economics, College of Agriculture and Life Sciences, University of Vermont, Burlington, VT 05405, USA
* Correspondence: ekaterin@bu.edu

Abstract: Modern AI applications have caused broad societal implications across key public domains. While previous research primarily focuses on individual user perspectives regarding AI systems, this study expands our understanding to encompass general public perceptions. Through a survey (N = 1506), we examined public trust across various tasks within education, healthcare, and creative arts domains. The results show that participants vary in their trust across domains. Notably, AI systems' abilities were evaluated higher than their benevolence across all domains. Demographic traits had less influence on trust in AI abilities and benevolence compared to technology-related factors. Specifically, participants with greater technological competence, AI familiarity, and knowledge viewed AI as more capable in all domains. These participants also perceived greater systems' benevolence in healthcare and creative arts but not in education. We discuss the importance of considering public trust and its determinants in AI adoption.

Keywords: artificial intelligence; trust; survey; generative AI; AI ethics; AI governance

check for updates

Citation: Novozhilova, E.; Mays, K.; Paik, S.; Katz, J.E. More Capable, Less Benevolent: Trust Perceptions of AI Systems across Societal Contexts. *Mach. Learn. Knowl. Extr.* **2024**, *6*, 342–366. https://doi.org/10.3390/make6010017

Academic Editors: Jianlong Zhou, Andreas Holzinger and Fang Chen

Received: 8 December 2023
Revised: 19 January 2024
Accepted: 1 February 2024
Published: 5 February 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The recent advancements in Artificial Intelligence (AI) have caught experts off guard. The release of ChatGPT (a large language-model-based chatbot) in late 2022 demonstrated that AI was approaching or, in some cases, surpassing humans in verbal-reasoning tasks like generating high-school-level essays, which had been projected to be automated only by 2026 [1]. ChatGPT's surprising success subsequently spurred other technology firms, including Meta, Google, Microsoft, and Baidu, to focus more of their resources on developing their own large language and multimodal models (LLMs and LMMs) and making them accessible to the wider public [2]

The public release of LLMs has resulted in adverse consequences evident across multiple societal domains. In the sphere of education, for example, ambiguities surrounding plagiarism protocols about the use of ChatGPT have led to instances of students being falsely accused of cheating [3]. Similarly, AI tools for image generation, such as DALL-E, have alarmed artists and creative professionals whose original works have been used for training datasets of these systems [4]. Public opinion surveys of American citizens indicate a steady rise in concerns over the influence of AI in daily life [5], as well as highlight the apprehensions regarding AI use in workplaces and healthcare [6]. While efforts have been made to establish regulatory oversight through national policies, such as the AI Bill of Rights [7] and the AI Act [8], as well as developers' solutions of incorporating guardrails for LLMs [9], the diverse range of ways users can implement these systems still presents a significant challenge to address.

AI's unintended consequences can stem from misaligned AI [10], which occurs when systems technically achieve their objective but exploit loopholes in the process that produces undesirable outcomes. For example, social media recommendation algorithms designed to generate user engagement can result in addiction and societal polarization [11].

Given the broader societal implications of these systems, it becomes imperative to extend AI systems' alignment beyond individual users and encompass the entire society. AI's "alignment problem" [10] originates from the prevailing technology-centered approach to AI development, which tends to overlook human needs and priorities [12]. To combat the harmful outcomes of misaligned AI, scholars and AI professionals have more recently pivoted to human-centered AI (HCAI), which emphasizes integrating human values and ethical considerations throughout the AI-development process [13].

A challenge for the human-centered approach is how to build AI that people will trust. Due to the autonomous nature of AI and its "black box" characteristics, trust has been one of the key factors in AI adoption [14]. Lee and See [15] defined trust as "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" (p. 54), mentioning that this definition should vary depending on the context. There are multiple approaches for categorizing the factors influencing users' trust in autonomous systems. Some concentrate on system capabilities, assessing variables such as ease of use and usability (e.g., [14,16]). Another approach stems from attempts to develop responsible AI, exploring how the explainability and fairness of systems impact user trust (e.g., [17,18]). Lastly, some adopt a broader societal perspective, considering factors like organizational trust that influence users' trust and adoption intentions (e.g., [19,20]). In this study, we adopt the latter approach, exploring trust within the framework established by Mayer et al. [21]. This framework evaluates trust based on three factors—the system's ability, benevolence, and integrity. This choice is supported by previous research indicating that the belief in an organization's benevolent intentions and its consideration of costs and benefits for stakeholders significantly shape user trust [22]. Additionally, this framework enables the examination of trust formation beyond the predictability of AI, encompassing AI's intention to align with users [23].

In this work, we use a representative sample of the U.S. population to measure people's trust in AI across different societal contexts and specific tasks. Specifically, we consider two dimensions of AI trust perceptions: how capable AI is at performing a task and how much AI performing a task is in one's best interest. We also explore whether individual traits, such as knowledge about AI, perceived technological competence, and socioeconomic status, influence people's trust. Our study diverges from the prevalent approach of assessing individual users' trust in the capabilities of AI systems or the quality of automated decisions in specific tasks. We argue that to ensure widespread AI adoption among the general public, it is imperative to consider people's trust in the AI system's contribution to societal well-being. Currently, there are few studies that delve into public perspectives on AI, with the majority of research placing more emphasis on expert opinions (e.g., [24,25]). As the demand grows for public inclusion in decision making about AI [26], our research contributes to the limited literature on the subject, offering insights for the future development and governance of AI technology.

2. Related Work

2.1. Defining Generative/Interactive AI

AI is a broad term that encompasses the ability of computer systems to perform complex tasks that are typically associated with human intelligence [27]. This involves a wide array of capabilities such as reasoning, decision making, learning, planning, and problem solving [8,28]. Various types of AI models and techniques exist, and in 2023, with the introduction of generative AI in the form of LLM and LMM, it is essential to distinguish between traditional machine learning (ML) AI and the more recent advent of generative, interactive AI.

Traditional ML AI primarily revolves around the analysis and interpretation of data, often deployed for predictive analytics, classification, or recognition tasks [29]. These systems typically function within predefined parameters, relying heavily on structured datasets to make decisions or provide outputs based on learned patterns (examples in Table 1) [30]. On the other hand, generative, interactive AI represents a more advanced and dynamic facet of AI technology. These systems not only analyze and learn from data but also possess the ability to generate new, original content or solutions [31]. This can include the creation of images, text, or music that did not previously exist, often driven by user inputs in real time (examples in Table 1) (e.g., [32–34]). The "interactive" aspect further distinguishes this type of AI, where the system collaborates with users, adapting and responding to their inputs, leading to a more dynamic, cocreative process [35–37]. In other words, generative AI represents the direct form of the output of using LLMs and LMMs. Interactive AI focuses on the human–computer interaction aspect of generative AI technology [30]. Concrete examples of generative, interactive AI encompass software such as chatbot platforms for customer service; content management systems that help summarize unstructured data [38]; synthetic media generation for creative work; and hardware such as voice assistants [39], automated vehicles, and robots within the realm of consumer electronics [40].

This paper addresses both generative and traditional ML AI technologies, which are presently employed in diverse capacities across various public sectors. Understanding the public's perception and trust in these systems is crucial to advancing human-centered AI design and offering targeted recommendations for technology policy.

Table 1. Public-domain-specific examples of traditional ML AI and generative, interactive AI.

Domain	Traditional Machine Learning AI	Generative, Interactive AI
Healthcare	Diagnostic tools for disease detection using ML algorithms (e.g., analyzing X-rays or MRI scans for tumors).	Personalized medicine platforms generating customized treatment plans based on patient data and feedback.
Education	AI-powered adaptive learning platforms adjusting educational content based on a student's learning pace and style.	AI systems for interactive learning scenarios, creating dynamic educational content like virtual lab experiments.
Creative Arts	Recommendation algorithms in music or video streaming services analyzing user habits to suggest content.	Generative music composition or digital-art-creation platforms, adapting outputs based on user inputs.

2.2. Embedding Public Perception into Human-Centered AI Design

The present development of AI systems follows a technology-centered design approach that gravitates toward building AI systems tailored for specific narrow tasks [12]. While this strategy enables developers to concentrate on refining algorithms for end users, it simultaneously overlooks the broader implications that these algorithms may have on the public at large. This approach has led to multiple instances of AI failures within societal contexts, such as the COMPAS parole system [41], biased hiring algorithms [42], and the exacerbation of social media polarization [43]. In a recent empirical study examining the detrimental effects of LLM-generated misinformation versus human-written misinformation, it was found that LLM-generated misinformation can be more challenging for humans and detection systems to identify when compared to human-written misinformation conveying the same message. This implies that LLM-generated misinformation may exhibit more deceptive styles and potentially pose a greater risk of causing harm [44].

The alternative, an HCAI approach, encompasses both technological and human factors [12,35]. The technological aspect pertains to the development of AI systems under human control that enhance people's capabilities, while human factors emphasize the consideration of human needs during the AI-development process. HCAI researchers also underscore the positioning of human users once the AI system has been implemented. Sowa et al. [45] outlines four levels of human involvement in AI systems, which include level 1—working separately or competing; level 2—supplementing each other's work; level 3—working interdependently; and level 4—full collaboration. The ethical employment of

AI systems is significantly influenced by the above level of human involvement. As illustrated in the examination of a medical AI system for image recognition, the inherent biases and stereotypes of users can create a self-reinforcing loop: users' biases are reproduced and amplified by algorithms, subsequently reinforcing existing biases. Tiron-Tudor and Deliu [46] further exemplify how variations in human-in-, on-, out-, and governing-the-loop impact administrative decisions in a business audit setting.

HCAI researchers have been addressing crucial issues in the design and implementation of AI systems, with a specific focus on the acceptance of these systems by individual users [47]. However, less attention is paid to normative considerations around the public's preferences for the widespread implementation of this technology. With AI systems being imposed on the public without the option to opt out, the acceptance and usability of these systems by individual users fail to accurately mirror the broader public's perceptions, especially in terms of the public's trust in such systems. The present study aims to address this gap in the existing literature by examining the perceptions of AI among individuals who may not have directly utilized the technology but experience the consequences of its implementation in various aspects of their everyday lives.

2.3. AI Trust

2.3.1. Engendering Trust through System Capabilities

One avenue of research delves into the examination of trust in AI by gauging users' confidence in a system's technical capabilities. For instance, studies guided by the technology acceptance model (TAM) have identified factors such as perceived usefulness, ease of use, and attitude leading to the acceptance of AI voice assistants [14]. Other research indicates that the characteristics of a system, including the design, service quality, and information quality, play a role in shaping users' trust [48]. Furthermore, given that trust has traditionally been a human-to-human trait, the anthropomorphism of systems has been shown to be crucial in fostering user acceptance and trust. Several studies illustrate that AI imbued with human-like qualities tends to be more trusted across various contexts, such as consumer settings [49], the banking sector [50], service industries [51,52], the travel industry [53], and personal use [54]. While this approach remains pertinent for evaluating users' trust in specific AI-powered tools like chatbots [16], voice assistants integrated into home speakers [55], or robots [56], its applicability may be constrained when applied to AI systems within a broader social context.

2.3.2. Engendering Trust through Ethical AI

In the literature on automated decision making, studies have explored the impact of fairness and explainability on users' trust in AI systems. For instance, Angerschmid et al. [17] discovered that in a medical setting, AI explainability increased users' trust while low introduced fairness decreased it. Another set of studies focused on trust in AI news and media recommendation systems revealed that not only explainability and fairness but also factors such as accountability and transparency played crucial roles in cultivating users' trust [57]. Some researchers paid attention to the relationship between explainability and fairness, demonstrating that emphasizing explanations of specific features of the AI system could either enhance or hinder its fairness [58]. Other studies also highlighted that explainability methods can be complex for the end users and can result in people overtrusting the system [56,59].

2.3.3. Engendering Trust beyond System Performance

In contrast to studies that primarily focus on evaluating individual users' trust in specific AI systems, this study seeks to broaden the concept of trust in AI as a sociotechnical system. With AI deployment by governments and corporations, trust is also contingent on people's confidence in these institutions' intentions [48]. One of the frameworks used to assess users' trust in AI, beyond individual trust in the systems' performance, originates from the organizational trust model proposed by Mayer et al. [21]. Within this framework, trust evaluation encompasses not only the system's ability but also its perceived benevolence and integrity, thus encompassing the societal dimensions of AI systems. In a qualitative study, Bedué and Fritzsche [20] underscored that these dimensions were crucial factors in establishing trust in AI as an emerging technology. Moreover, the significance of AI systems' ability, benevolence, and integrity in fostering trust was demonstrated in scenarios of public importance, such as contact tracing during the COVID-19 pandemic, big data analytics, and public health emergency research [19]. In this study, we adapt the Mayer et al. [21] framework to evaluate the public's trust in generative AI's ability and benevolence across various tasks in healthcare, education, and creative arts.

2.3.4. Individual Factors Influencing AI Trust

Bonnie Muir [60] was one of the first researchers to challenge the idea that users' trust consisted merely of systems' technical properties. Her earlier experiments with automated banking machines showed how participants' trust levels varied, even while the banking system properties remained constant. Further research demonstrated that individual differences contribute to variability in users' trust in AI systems [61].

Among those differences are demographic factors such as age, gender, and education level. Some studies indicate that older individuals are more inclined to express lower trust in algorithms [62], whereas the younger population tends to believe that AI systems exhibit less bias compared to human operators [63]. However, these observations are not consistently replicated across different contexts [64]. Moreover, other studies found that younger adults with higher education levels have greater trust in AI [65]. Regarding gender, research demonstrates that female participants tend to hold more negative attitudes toward AI in comparison to their male counterparts [64,66,67].

Beyond demographic traits, technological efficacy has a pivotal role in the adoption of new technology. In an organizational context, employees' technology competence acts as a catalyst for embracing new IT systems and autonomous machinery [68,69]. Within the wider population, research shows that individuals' perceived technological competence predicts their comfort with AI systems across various roles [66,67]. Similarly, those with computer science and engineering backgrounds tend to exhibit stronger support for AI development and implementation [70]. Conversely, a lack of AI knowledge can contribute to resistance in adopting new technology [71,72]. Finally, prior experience and familiarity with AI have been shown to positively influence people's perceptions of it, translating into enhanced trust [65,73].

2.4. AI Trust across Different Contexts

The current state-of-the-art AI systems that have a great impact on people's everyday lives are not domain- or task-specific. This underscores the necessity of assessing AI trust across diverse contexts. Moreover, previous research showed that trust in AI systems fluctuates depending on how impactful their outcomes are. For instance, in high-stakes decisions within healthcare and criminal justice domains, people evaluated automated decision-making systems as fairer than human experts, whereas no disparity in perceptions was observed in the media domain [64]. In another study, characteristics of trustworthy AI, such as accuracy, reliability, and objectivity, were more important in a high-risk context (e.g., a medical diagnostic system) and less important in a low-risk one (e.g., a music recommendation system) [74]. In the low-risk context, a satisfactory level of performance was sufficient for someone to trust the AI system. To engender trust in the high-risk context, however, the AI system had to meet substantially higher benchmarks in trustworthy AI characteristics [74].

This study explores AI perceptions across domains categorized by their associated risk levels: the creative arts, representing low risk; education, representing medium risk; and healthcare, representing high risk. Healthcare is presented as a high-risk domain due to the critical nature of decisions made in this field and the potential dire consequences of errors made by AI [75]. Education is identified as a medium-risk domain because, though the potential adverse effects of AI systems may not result in immediate life-threatening consequences, they can have a substantial impact on learning outcomes and the future of both students and teachers. The creative arts are classified as low risk due to the less severe impact of errors compared to the healthcare or education sectors. Creative arts involve subjective interpretation, and AI's role is often seen as augmentative rather than prescriptive. The following sections provide a brief overview of the present state of AI use within these domains.

2.4.1. AI in Healthcare

Today, AI technology has permeated the healthcare domain, spanning health-centric mobile applications [76], wearable devices [77], and robotic surgical procedures [78]. AI technology has unquestionably enhanced the domains of diagnosis and treatment and lowered the cost and time to discover and develop drugs [13]. It has augmented the efficiency of healthcare provision, leading to enhanced patient outcomes and mitigating the potential for selection discrimination [79]. ChatGPT specifically has been recognized as a valuable tool for helping the development of clinical practice guidelines for medical professionals [80], as well as an aid for patients in organizing information related to medications, lifestyle modifications, and treatment alternatives [79].

However, the medical field's integration of AI has also given rise to multiple biases. For instance, AI systems introduce information bias, originating from the over-representation of White and higher-income patients in electronic health record databases [81]. In medical AI image analysis systems, an interdisciplinary team of physicians, AI researchers, and regulators identified roughly thirty potential biases in each stage of the system development and deployment [82]. In public health settings, AI biases have been found in schedule-optimization systems, posing significant implications for cost benefits on one hand [83], and on the other hand, jeopardizing patients of specific racial groups [84].

Extant research on individuals' perceptions of medical AI systems has predominantly centered on those of domain experts. For instance, studies have evaluated how the design of medical AI systems contributes to either under- or over-trust among healthcare practitioners [24,59]. Nevertheless, it has been demonstrated that understanding individual experts' preferences does not guarantee the integration of AI technology into practice. Li et al. [85] presented numerous instances where the recommended AI models were either underutilized or used incorrectly, leading to additional workload for physicians and distress for patients. Research indicates that the design and implementation of AI systems in the medical domain should consider input from various stakeholders, including healthcare practitioners, social workers, policymakers, and patients [86].

The few studies examining the public's perception of AI-driven healthcare advice reveal that the general public tends to regard medical AI decisions as more reliable than human judgments (e.g., [87,88]). These findings contradict the recent public opinion survey on AI in healthcare, which indicated that 60% of the American population would feel uncomfortable if their healthcare provider relied on AI for their medical care [6]. Given that the harm resulting from medical errors committed by AI systems is felt more profoundly by patients than by their practitioners [75], our research aims to highlight areas where AI implementation faces the most significant resistance among the general public.

2.4.2. AI in Education

Traditionally characterized by slower adoption rates of technical innovations [89], the field of education has begun to pay greater attention to AI advancements. In particular, ChatGPT's release has catalyzed considerable disruptions to the learning process and prompted educational institutions to respond with various policies. While some schools persist in their prohibition of this emerging technology [90], others have embraced it and offered educators' courses on how to use ChatGPT [91].

The educational potential ushered in by tools like ChatGPT is difficult to overlook: its utility spans assisting in writing, research, and problem solving, bearing promising implications for personalized learning and aiding learners with disabilities [92]. For educators, ChatGPT proves to be a valuable ally in tasks such as crafting educational content, grading, and the development of test assignments [93]. Nonetheless, it is justifiable why certain educational institutions might hesitate to adopt these pioneering pedagogical approaches. Prior to the release of LLM-enabled tools like ChatGPT, AI systems' integration in the educational sector exhibited biases in decision-making processes for university admissions and in predictive analytics for identifying at-risk students [94]. In 2020, inaccurate algorithmic predictions of test scores led to U.K. students losing spots at universities, prompting student protests in London [95]. In the United States, enrollment-management algorithms have demonstrated a tendency to decrease the amount of scholarship funding extended to students and perpetuate unjust financial aid practices [96]. Given that American higher education is already grappling with low graduation rates, student debt, and persistent inequality for racial minorities, the introduction of AI systems may exacerbate these challenges. Currently, there are indications that the widespread use of LLMs in education is expected to introduce implicit bias, copyright concerns, the dissemination of inaccurate information, issues related to data privacy, and problems associated with plagiarism [97]. Analogous to the healthcare sector, there is a threat of teachers and learners becoming too reliant on AI tools coupled with the lack of trust in their usefulness for academic purposes [92].

Views on AI use in education diverge among different stakeholders. Educators' lack of trust in or excessive reliance on AI tools stems from an inability to comprehend the underlying rationale behind AI-generated decisions [71]. This is not the case for learners, however, whose trust and motivations are predominantly aligned with achieving good grades rather than understanding the AI system's decisions [98]. Furthermore, a study conducted among Chinese participants underscored that both students' and teachers' trust in AI systems is considerably reliant on the systems' performance, whereas parents' trust hinges on their confidence in the organizations deploying such systems [99]. A significant tension between the university administrative system and teaching staff emerged when implementing a motivational AI system [100]. While the tool held the potential to enhance the efficiency of professors' assessment, continuous monitoring was necessary to promptly address contradictions in teachers' opinions regarding the ideal assessment system and the forms of stimulation offered by university authorities.

2.4.3. AI in Creative Arts

Over the past few years, generative art has captivated a significantly broader audience than artists, tech geeks, and art enthusiasts. Systems such as ChatGPT for text generation and DALL-E for image creation have showcased their capacity to produce artworks that are often indistinguishable from, and in some cases surpass, human creative output [101]. To illustrate, in 2022, an AI-generated image won an art competition in Colorado, USA, setting intense debates concerning the work's authorship, originality, and the eligibility for submission to art contests [102]. Currently, a number of concerns surround the future of creative professions, extending not only to visual artists but also including music composers [103], game designers [104], and writers [105]. Beyond that, there is a discussion about how AI systems introduce biases rooted in demographic stereotypes [106], and how these biases might impact the realm of creative art [107].

Nevertheless, AI technology's growing capabilities generate new opportunities for individuals to engage in creative endeavors, thereby democratizing access to creative tools [108]. Generative AI tools are accessible not only to artists but also to individuals without an artistic background [109]. Further, collaboration between AI and humans has the potential to amplify creative outcomes [110]. However, successful human–AI creative collaboration also requires a new skill set in crafting effective prompts; otherwise, creative AI tools' output may fall short [111].

Research on attitudes about creative AI demonstrates nuance and different types of considerations. A study that analyzed discussions on subreddits dedicated to generative AI art, including r/aiArt, r/AIGeneratedArt, and r/DefendingAIArt, revealed that redditors both expounded on the remarkable technical capabilities of tools like MidJourney and DALL-E and also expressed concerns about these tools' socioeconomic implications [112]. These discussions about creative AI's potential impact on the labor market, various businesses, and industries are predominantly negative in tone [112]. In another study, Alves da Veiga [113] gathered feedback pertaining the distinction between the lay public and artists at the AI-generated art exhibition in Portugal. The general audience showed a lack of understanding of what "generative" means in generative art and displayed no awareness of the ethical implementations of these tools. In contrast, the artists were well-informed about copyright issues and controversies surrounding generative art.

Overall, the research summarized above demonstrated that there is not necessarily a one-size-fits-all approach to AI across domains. Different stakeholders (e.g., medical professionals, patients, teachers, students, parents, and artists) vary in their attitudes toward AI and have different criteria for their evaluations based on both the task at hand as well as their goals for the technology. In the present study, our objective is to expand the exploration of public trust in AI systems. We ask about attitudes not only across three domains but also among discrete tasks within each domain. To explore variations in these attitudes, we propose the following research questions:

RQ1: To what extent does public trust in AI's (a) ability and (b) benevolence vary across healthcare, education, and creative arts domains?

RQ2: Are there differences in the public's level of trust in AI's (a) ability and (b) benevolence across discrete tasks within healthcare, education, and creative arts domains?

Given the disparities in scope between this study and earlier research, we also evaluate how demographic and domain-specific characteristics influence AI trust. Thus, we put forth our third research question:

RQ3: To what degree is trust in AI systems influenced by (a) general demographic traits and (b) domain-specific traits?

3. Method

3.1. Design and Participants

In August 2023, we conducted a survey with U.S. participants (N = 1506) through an online questionnaire via the Qualtrics survey platform. The larger survey assessed attitudes and trust across different societal domains, alongside questions about demographic and individual traits. The main variables in this analysis are drawn from a section about perceptions of AI and were established from the outset of data collection. Qualtrics provides a survey technology platform and partners with over 20 online panel providers to supply a network of diverse, quality participants. Their recruitment strategies include purchasing mailing lists and advertisements on websites and social media networks. Participants accessed the survey via a survey link and completed it on the Qualtrics platform. Each participant received compensation upon survey completion. Sample quotas on gender, age, ethnicity, and education were specified to match those demographic distributions in the U.S. population (see Table 2). The sample was 52.6% female, 63.7% White/Caucasian, 54.4% had at least some college education, and the average age was 44.90 (SD = 17.78). A complete description of the scales and measure items used in the analysis can be found in Appendix A.

Table 2. Descriptive statistics of demographics.

Variable	Mean (SD)	%
Gender		
Male		47.9%
Female		52.1%
Age (18–95)	44.90 (17.78)	
Education		
Less than High School diploma		3.2%
High School diploma/GED		27.5%
Some college (no degree)		23.7%
Associate's degree		10.0%
Bachelor's degree		22.6%
Graduate degree		13.0%
Race/Ethnicity		
White/Caucasian		63.7%
Black/African American		11.1%
American Hispanic/Latino		17.9%
Asian or Pacific Islander		4.8%
American Indian or Alaska Native or Other		2.5%
Political Ideology		
Republican		23.8%
Democrat		43.6%
Independent or no affiliation		30.5%
Other		2.2%

3.2. Measurement

3.2.1. AI Trust

To measure participants' AI trust across different societal domains, we asked them to respond to different uses of generative AI in education, healthcare, and the creative arts. Based on Mayer et al. [21], we measured participants' trust in the AI system's ability to perform various tasks within a certain domain and the system's benevolence in performing those tasks. See Appendix A for the complete wording of vignettes and tasks. To measure perceptions of AI ability, participants were asked to provide a response on a five-point Likert-type scale, from "Not at all capable" to "Entirely capable". To capture perceptions of AI benevolence, participants were asked to indicate their agreement that an AI would perform a given task in the best interest of the task recipient, and responses were given on a five-point Likert-type scale from "Strongly disagree" to "Strongly agree". The following definition of AI was given preceding these questions: "Artificial Intelligence (AI) refers to computer systems that perform tasks or make decisions that usually require human intelligence. AI can perform these tasks or make these decisions without explicit human instructions".

We created indices for perceived AI ability and benevolence within each domain. To validate the scales, we conducted a principal components analysis (PCA) that treated AI ability and AI benevolence as separate unidimensional, five-item indices. Given that the items asking about AI abilities and benevolence across domains were novel, they were subjected to PCA with a varimax rotation. Across all six PCAs, the KMO measure of sampling adequacy was >0.85 and significant at $p < 0.001$. Only one component was extracted per PCA and explained between 67.57 and 81.39 percent of the variances. Most factor loadings exceeded 0.80, and all factor loadings exceeded 0.70. For the full statistics of each PCA, see Appendix D.

In addition, we asked participants for qualitative elaborations on AI applications in various domains. After the generative AI-related questions, we provided the following prompt: "OPTIONAL: Understanding your opinion is very important to us. If there is anything you would like to add about the use of AI in various applications, please elaborate in the text box below". From this prompt, we received 567 responses, 38 percent of the sample. Following data cleaning to remove off-topic or nonsensical responses, a total of 301 responses were analyzed. While we did not conduct a systematic qualitative analysis of the responses, we reviewed and included comments that illustrated quantitative findings in more detail.

3.2.2. Individual Traits

To measure AI knowledge, we assessed participants' perceived as well as demonstrated AI knowledge. For the former, we asked questions about whether participants are informed enough to explain what AI is and what it does; what best describes their AI knowledge; and whether they think they know enough information to make accurate judgments about AI ($a = 0.85$). Higher values corresponded to a stronger belief in having AI knowledge ($M = 3.10$, $SD = 0.75$).

For measuring demonstrated AI knowledge, we adopted the approach used by Zhang et al. [70]. Respondents were provided with five instances of technologies, randomly selected from a set of 14, and asked to indicate whether those technologies use AI. Scores were summed for each participant, ranging from 0 to 5, with a higher score indicating a greater demonstration of AI knowledge ($M = 3.36$; $SD = 1.39$).

Participants were also asked about their level of familiarity with AI with the answer options ranging from 1 = "I have never heard of AI" to 5 = "I have extensive experience in AI research and/or development" ($M = 2.74$, $SD = 1.10$). Finally, we asked participants whether they had used an online application or tool for generating text or images (ChatGPT/DALL-E) with answer options 0 = "No" and 1 = "Yes" (49.6% and 40.4% of participants, respectively). Perceived technological efficacy was adapted from Katz and Halpern [114] and updated to ask about digital technologies more broadly. Respondents were asked to state how much they agreed or not (1 = "Strongly disagree", 5 = "Strongly agree") with how sufficiently skilled they were at "Using technologies, like my phone, e-mail, social media, to interact with others", "Using the internet to find information online", "Trying to figure out and solve problems with my technology when they come up", and "Using technologies, like digital calendars, video conferencing, and word processors, to complete my work" ($a = 0.78$, $M = 4.16$, $SD = 0.74$).

3.3. Data Analysis

IBM SPSS Statistics (Version 29) was employed for all data analyses. Descriptive statistics were computed and provided for all the dependent variables (e.g., AI domains). The research questions were tested through ordinary least squares (OLS) regressions, with significant relationships determined by p-value less than 0.006 based on a Bonferroni correction. To ensure that assumptions for the OLS regression were not violated, we evaluated the independence of errors by using the Durbin–Watson test (the models' scores range from 1.72 to 1.86), collinearity statistics to determine a lack of multicollinearity (VIF < 2.4 for all variables across all models), and visually inspected the normal P–P plots and scatterplots of residuals to verify the normality of the residuals and homoscedasticity. The variables were sequentially introduced into the models in two stages: (1) demographics and (2) domain-specific traits.

4. Results

4.1. AI Trust

The descriptive analysis showed that, overall, participants trusted AI's ability more than they trusted its benevolence (see Figure 1). Qualitative responses underscore the participants' differentiation between a system's ability and its benevolence, highlighting the inherent complexity: "Sometimes it's not as straightforward—like, of course, AI could perform the tasks of a medical assistant—but in a SAFE and EFFECTIVE way?". Some respondents voiced concerns about the persona of the AI's programmers, as demonstrated by one participant's statement: "AI is very capable in 99 percent of cases, I believe; however, the person(s) who programmed the AI is the concern".

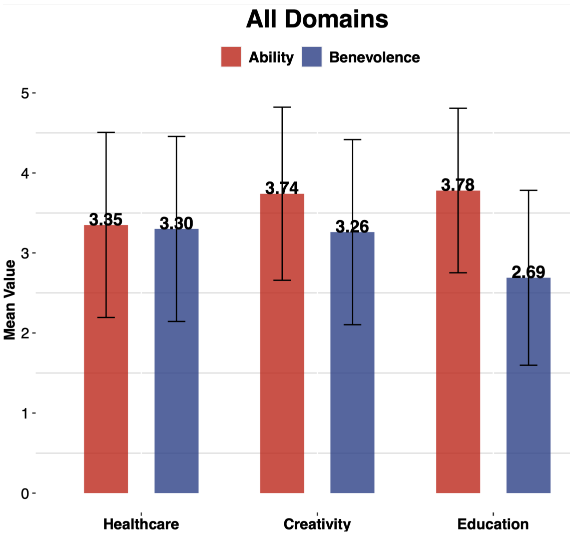

Figure 1. Mean distribution of participant's trust in AI's ability and benevolence across domains. Note: error bars indicate standard deviation.

This ability–benevolence distinction was most pronounced for trust in AI in education. Specifically, participants demonstrated greater trust in AI's ability and benevolence for tasks related to students, such as essay drafting or answering questions, compared to tasks typically associated with teachers, like providing learning support or creating lesson plans (see Figure 2). Participant comments shed light on this discrepancy, with some noting that AI tools are "useful for homework" or for "grading... given teacher shortages". However, concerns were raised that AI potentially "takes away student logic, thinking, and creativity" and the absence of "human interaction" in learning settings.

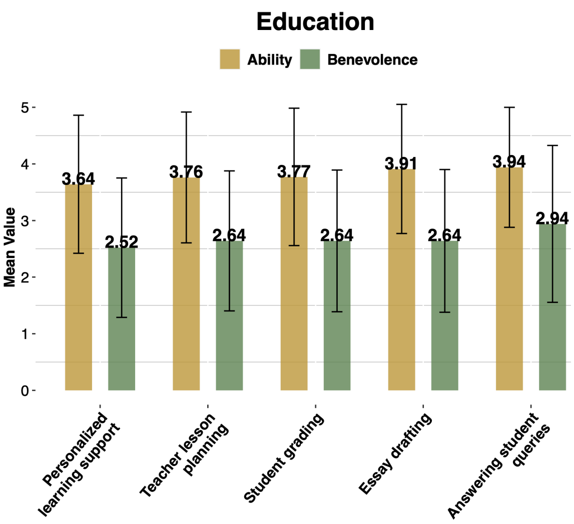

Figure 2. Mean distribution of responses across AI in education. Note: error bars indicate standard deviation.

Conversely, a different overall pattern emerged in the healthcare domain, with higher trust in AI's benevolence but lower trust in ability. AI tools designed to assist medical practitioners in disease diagnostics and medical research were generally seen as both more capable and more benevolent (see Figure 3). However, the use of AI in therapy settings was perceived as the least capable and benevolent across all tasks. Negative comments regarding AI applications in therapy were prevalent, with one comment encapsulating the general sentiment: "I don't know if therapy is the best use for this. As someone who has spent a lot of time in therapy, it requires a human touch. It is more than saying the right advice. Social interaction is important in recovery and sometimes the therapist is the only one a depressed or other mentally ill person interacts with."

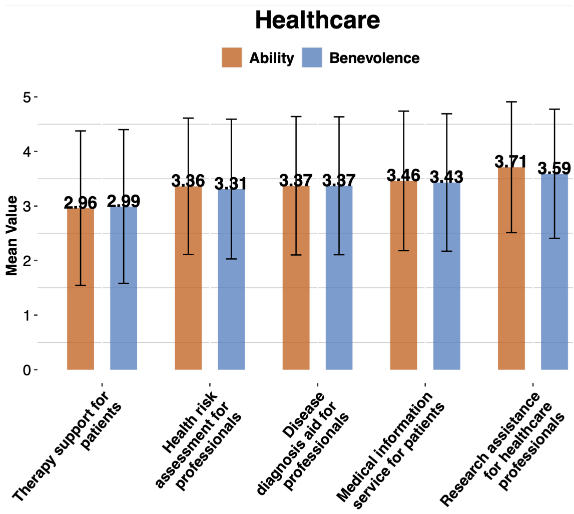

Figure 3. Mean distribution of responses across AI in healthcare. Note: error bars indicate standard deviation.

Finally, the least variance across tasks was observed in the domain of creativity (see Figure 4). Participants perceived AI applications in creative writing as less capable and benevolent; however, they expressed more trust in AI's ability to create videos. Notably, AI tools for generating art were perceived as highly capable but lacking in benevolence. As one participant remarked, "A lot of the AI that is used to make art has been using art from online artists without their consent". Nonetheless, qualitative comments revealed a range of perspectives. Some participants expressed great enthusiasm for AI tools in creativity, stating, "I really love generative AI art tools. Both text to image and text to video". In contrast, others were more skeptical of AI's role in art, emphasizing that AI is "good for producing but not for creating".

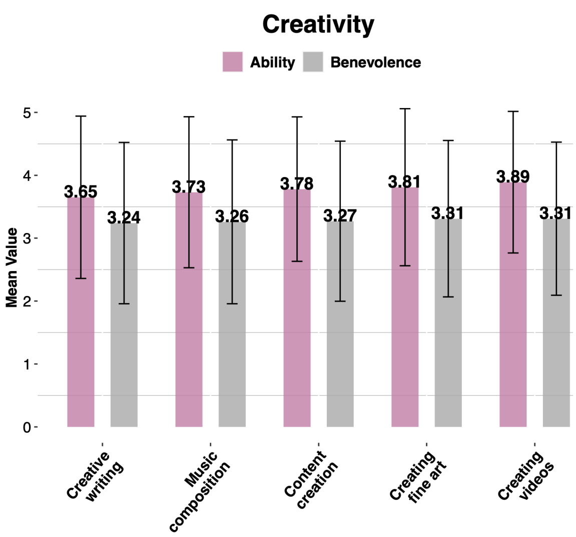

Figure 4. Mean distribution of responses across AI in creativity. Note: error bars indicate standard deviation.

4.2. Individual Traits

In order to examine the relative influence of individual traits on trust in AI's ability and benevolence, hierarchical regression models were run (see Tables 3 and 4). The dependent variables were newly constructed indices of perceived AI ability in the healthcare, education, and creative arts domains. Each model was composed of two blocks: (1) demographic traits (age, gender, and education level) and (2) technical knowledge and experience (perceived technology competence, perceived and objective knowledge of AI, familiarity with AI, and prior experience with LLMs).

4.2.1. Factors That Influence Trust in AI Ability across Domains

All three domain models for trust in AI's ability were significant (healthcare: $F(8, 1283)$ = 58.00, $p < 0.001$; education: $F(8, 1298) = 46.34$, $p < 0.001$; creative arts: $F(8, 1286) = 34.29$, $p < 0.001$) and overall explained 26.10 percent of the variance in healthcare ability, 21.7 percent of the variance in education ability, and 17.1 percent of the variance in creative ability (see Table 3).

Table 3. Influence of individual traits on trust in AI ability across domains.

	Healthcare β	Education β	Creative Arts β
Constant	0.42	1.44	1.46
Gender (male = 1, female = 2)	0.10 ***	0.04	−0.001
Age	0.04	−0.001	0.03
Education	0.04	0.03	0.02
R^2 change	12.3%	9.5%	5.8%
Perceived knowledge of AI	0.26 ***	0.17 ***	0.17 ***
Objective knowledge of AI	0.07 **	0.10 ***	0.11 ***
Perceived technology competence	0.07 **	0.14 ***	0.11 ***
Familiarity with AI	0.13 ***	0.12 **	0.13 **
Prior experience with LLMs	0.07 *	0.07 *	0.04
R^2 change	14.5%	12.7%	11.8%
Total adjusted R^2	26.1%	21.7%	17.1%

Note: β indicates standardized regression coefficient. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Correcting for the number of predictors across models using Bonferroni's method, the threshold for significance is $p < 0.006$, so significant results at $p > 0.001$ should be interpreted with caution.

Demographic traits had little influence on perceptions of AI's ability across three domains. Gender was the only significant factor, and only in the healthcare domain, wherein men were more likely to view AI as capable (β = 0.10, $p < 0.001$).

Technological experience and knowledge, on the other hand, contributed significantly to ability perceptions. Those who perceived themselves as more technologically competent were more likely to view AI as more capable in healthcare (β = 0.07, $p < 0.01$), education (β = 0.14, $p < 0.001$), and creativity (β = 0.11, $p < 0.001$). Subjective perceptions of AI knowledge were also significant: respondents with higher perceptions of AI knowledge and familiarity held higher perceptions of AI's ability in healthcare (β = 0.26, $p < 0.001$; β = 0.13, $p < 0.001$, respectively); education (β = 0.17, $p < 0.001$; β = 0.12, $p < 0.01$, respectively); and creative arts (β = 0.17, $p < 0.001$; β = 0.13, $p < 0.01$, respectively). Similarly, demonstrated knowledge of AI was positively related to beliefs about AI's ability in healthcare (β = 0.07, $p < 0.05$), education (β = 0.10, $p < 0.001$), and creative arts (β = 0.11, $p < 0.001$). Finally, those with prior LLM experience were more likely to see AI as more capable in healthcare (β = 0.07, $p < 0.05$) and education (β = 0.07, $p < 0.05$).

4.2.2. Factors That Influence Trust in AI Benevolence across Domains

All three models for trust in AI's benevolence were significant (healthcare: $F(8, 1275) =$ 42.88, $p < 0.001$; education: $F(8, 1280) = 35.10$, $p < 0.001$; creativity: $F(8, 1271) = 36.34$, $p < 0.001$) and overall explained 20.7% of the variance in healthcare benevolence, 17.5% of the variance in education benevolence, and 18.1% of the variance in creative benevolence (see Table 4).

Table 4. Influence of individual traits on trust in AI benevolence across domains.

	Healthcare β	Education β	Creative Arts β
Constant	1.07	4.17	1.44
Gender (male = 1, female = 2)	0.11 ***	−0.04	0.04
Age	−0.03	0.08 **	−0.04
Education	0.05	0.14 ***	0.05
R^2 change	12.2%	10.2%	9.3%
Perceived knowledge of AI	0.19 ***	−0.14 ***	0.08 *
Objective knowledge of AI	−0.03	0.01	−0.005
Perceived technology competence	0.08 **	−0.05	0.07 *
Familiarity with AI	0.014 ***	−0.19 ***	0.27 ***
Prior experience with LLMs	0.08 *	−0.04	0.02
R^2 change	9.0%	7.8%	9.3%
Total adjusted R^2	20.7%	17.5%	18.1%

Note: β indicates standardized regression coefficient. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Correcting for the number of predictors across models by using Bonferroni's method, the threshold for significance is $p < 0.006$, so significant results at $p > 0.001$ should be interpreted with caution.

Men were more inclined to think that AI used in healthcare would be in the best interest of the patient (β = 0.11, $p < 0.001$). Older respondents perceived AI used in education as more likely to be benevolent (β = 0.08, $p < 0.01$). Otherwise, demographic traits were not significantly related to views of AI's benevolence across the three domains. Perceived technological competence was positively related to perceptions of AI as benevolent in health (β = 0.08, $p < 0.01$) and creative (β = 0.07, $p < 0.05$) domains. Those who considered themselves to be more knowledgeable and familiar with AI were more likely to view it as benevolent in healthcare (β = 0.19, $p < 0.001$; β = 0.14, $p < 0.001$, respectively) and creativity (β = 0.08, $p < 0.05$; β = 0.27, $p < 0.001$, respectively). Interestingly, those with more knowledge (β = −0.14, $p < 0.001$) and familiarity (β = −0.19, $p < 0.001$) with AI were less likely to see it as benevolent in the education domain. Demonstrated knowledge had no significant relationship with benevolence perceptions in any domains. Prior experience with LLMs was positively related only to perceptions of benevolence for AI in healthcare (β = 0.08, $p < 0.05$).

5. Discussion

5.1. More Capable, Less Benevolent

This study investigates public trust in AI across different tasks in the domains of education, healthcare, and creativity. Our findings suggest that public trust in AI varies depending on the specific domain and the task within that domain. This corroborates prior research, which demonstrated that AI applications in higher- and lower-stakes domains elicit divergent attitudes [64]. Additionally, our results indicate that people have distinct perceptions of AI's abilities and benevolence. Across all domains, participants tend to trust AI's ability more than its benevolence, suggesting that while AI is seen as proficient, concerns linger about whether it always serves the public's best interests.

One plausible interpretation of these findings may be linked to the current decline in institutional trust among Americans [91]. Institutional trust has been shown to positively predict trust in AI [65]. More importantly, the low levels of trust in AI benevolence might be associated with a decline in trust in tech companies that deploy these systems. Recent public opinion surveys have indicated that the tech industry has fallen from favor compared to other industries and institutions [115]. The general public has expressed unease regarding data privacy, the extensive growth and influence of tech giants, and their impact on the U.S. economy. Previous research by Zhang et al. [70] revealed that fewer than one third of Americans believed that tech companies consistently act ethically. This sentiment is reflected in the qualitative comments of our study, where several participants voiced concerns about the "programmers" of AI technology. This illustrates people's lack of trust not in AI systems and their capabilities, but in the individuals who deploy them.

5.2. Preference for Humans over AI

The lower trust in AI systems' benevolence might be also rooted in people's perceptions of the impersonal nature of the technology. Our findings show that tasks involving a higher degree of human-to-human interaction tend to result in lower trust in AI's ability and benevolence. For instance, a promising AI application in healthcare—therapy chatbots—invoked the least trust by the participants. While other user-centered research highlights the great potential of this technology [116], our findings suggest that people have reservations about trusting this task to AI entities. In our study, participants were more inclined to trust AI as an assistive technology that aids practitioners in diagnosis or medical research rather than act as a substitute in doctor–patient interactions.

This pattern prevailed in the education domain: AI was trusted least in tasks typically handled by teachers, such as offering personalized learning support to individual students. This is despite the fact that customizing the learning environment for individuals is touted by both academics [92] and private companies [117] as a key advantage of systems like ChatGPT. In contrast, AI applications for tasks like essay drafting or providing answers to student queries were perceived as more capable and benevolent. However, it is important to highlight that perceptions of AI benevolence for all educational tasks were lower when compared to AI's ability. Our findings in both healthcare and educational domains align with previous public opinion surveys that indicate that people have reservations about AI taking on the roles of teachers and therapists, demonstrating that the advent of LLMs has not altered people's preferences for human involvement in these domains [66,67,118].

This aversion to AI's involvement in tasks requiring human interaction may be explained by the negative machine heuristics [119] people hold about AI. Negative heuristics come into play when AI is perceived as less adept in tasks that demand "human" qualities such as intuition, emotional understanding, or empathy. This could also account for why tasks in the creative domain did not see much fluctuation in levels of trust, as writing or composing are more typically independent tasks rather than interactive ones. This is not to say that AI is embraced in creative domains; other research has demonstrated how AI authorship negatively influences perceptions of creative work [120]. Compared to domains like healthcare and education, however, which arguably have a more direct, immediate, and material impact on individuals, people seem to be less wary of creative AI.

5.3. Policy Implications

Our findings suggest that factors such as technology proficiency, AI knowledge, and familiarity significantly influence trust in AI applications. In the qualitative section, several participants mentioned their lack of knowledge about AI but expressed interest in learning more. This interest stemmed from a combination of excitement and anxiety about this technology. Based on these results, we emphasize the pressing need for more comprehensive AI awareness and literacy programs.

Private companies are taking steps to offer guidance on the ethical utilization of AI technology, exemplified by OpenAI's release of guidelines for implementing ChatGPT in educational settings [121]. Universities have also contributed to public education about AI, with examples like the University of Helsinki offering a free online course titled Elements of AI, available in multiple languages [122]. Regarding government initiatives to enhance AI literacy among citizens, there are several reports and guidelines addressing AI's use in various sectors, including education [123] and healthcare [124], as well as general guidelines for generative AI [125]. Notably, we could not find guidelines specifically related to creative works generated by AI, except for the information that such works are not eligible for copyright protection [126].

In the United States, an attempt to establish federal regulations for the implementation of AI in social sectors was evident in the AI Bill of Rights. However, critics argue that the Bill is unenforceable [127] and lacks uniformity across sectors, with some domains receiving inadequate attention. Specifically, there is less clarity regarding regulations for AI in education compared to sectors such as healthcare, employment, and consumer protection. Outside of the United States, more progress has been made on enforceable AI regulations, with China enacting rules for public-facing generative AI in August 2023 [128] and the European Union passing its AI Act in December 2023 [129]. The EU's AI Act in particular emphasizes that the rules were designed to create trustworthy AI.

Another limitation of existing reports and guidelines is their focus on professionals developing AI proficiency for future work contexts, significantly narrowing the target audience from the average citizen. This emphasis on work-related AI ignores the widespread implementation of AI across sectors, with implications for people beyond their professional lives. Further, it does little to address potential concerns about how broader AI integration poses unique challenges to personal autonomy and self-determination [130], which may be more salient to AI resistance. Thus, AI literacy programs should encompass not only fundamental AI knowledge and its workplace applications but also guide individuals on adapting to an environment where AI may outperform humans in at least some tasks. This can include strategies for finding personal meaning in the face of the rapid obsolescence of individuals' skill sets or guidance on how to interact and collaborate with AI entities. Providing such roadmaps for personal development alongside AI integration can foster greater trust and acceptance of these intelligent systems.

6. Limitations and Future Research

Our study is primarily limited in its sampling and reliance on self-reported data, which may not accurately represent participants' actual experience, knowledge, and attitudes. The public is not particularly well informed about AI, and even though we provided a definition in the survey, it is uncertain what the participants specifically perceived as AI while responding. In order to mitigate participants' cognitive load when responding to the survey, we limited our operationalization of trust to beliefs about ability and benevolence; additional trust-related concepts such as integrity and fairness beliefs likely also contribute to overall trust in AI.

While the AI tasks were selected based on examples of AI used in real-world contexts, they were developed by the researchers and therefore may be limited in external validity. Future research should explore other societal domains influenced by AI, such as the legal system and journalism. Moreover, research on trust across AI domains could systematically evaluate tasks that vary in the level of human–AI interaction, stakes, near- and long-term risk, as well as human replacement versus human augmentation. While this paper looks at both generative and traditional ML AI, researchers may find it beneficial to focus specifically on trust in the context of tasks performed by the latter, as generative AI's ability to produce novel creations, rather than simply analyze or categorize existing data, might pose new questions and societal concerns about creativity, ownership, and the authenticity of AI-generated content. Further, generative AI often involves a higher level of interaction with users, who can influence the AI's output in real time. This cocreation process changes the user's role from a passive recipient of AI decisions to an active participant.

Contrary to previous studies, which found that demographic traits significantly contributed to AI attitudes (e.g., [67]), our study revealed that other individual traits related to domain knowledge and experience were more influential for AI trust. Future research should further explore the underlying causes of public distrust in AI technology. Additionally, in contrast to survey methods, qualitative approaches offer the opportunity to provide clarifications, identify and resolve misunderstandings, and elaborate on topics as they emerge. As such, qualitative investigations, akin to public assemblies [131], are critical for both educating the public about AI and better understanding the underlying mechanisms and beliefs that are driving AI attitudes.

Finally, this study only sampled U.S. participants, which limits the extent to which the findings can generalize and contribute to the global body of knowledge about AI. As a technology that spans borders, AI has a wide-reaching, cross-national impact, yet countries and regions are taking different approaches to AI implementation and regulation [132]. Comparative research would add more insight to the efficacy of these differing approaches and offer a better understanding of where there may be cross-cultural consensus about AI (see [133]), providing crucial information for global AI governance discussions. Thus, future research should include more cross-cultural analysis of attitudes about and drivers of trust and confidence in AI technologies.

7. Conclusions

The present study examines public trust in AI across various tasks within the domains of education, healthcare, and creativity. Using a representative sample of the U.S. population, this work goes beyond prevalent usability studies and provides empirical evidence of public perspectives on AI, thereby addressing a gap in the existing literature. Further, our study demonstrates the applicability of Mayer et al.'s [21] organizational trust framework to gauge trust in AI as a sociotechnical system. Our findings indicate that public trust in AI varies across domains and is influenced by individual differences, specifically technical competence and familiarity with AI technology. The results highlight a notable gap in trust between perceptions of system capabilities and its benevolence. Across all tasks in the three domains, individuals demonstrated higher trust in AI systems' capabilities compared to their perceived benevolence. As AI technology continues to advance in complexity and proficiency, its integration into the social context necessitates careful consideration. Our results carry significant implications for AI system design and policies, emphasizing the importance of incorporating public perspectives in the development and regulation of AI systems in public domains.

Author Contributions: Conceptualization, E.N., K.M., S.P., and J.E.K.; methodology, E.N., K.M., S.P., and J.E.K.; formal analysis, E.N. and K.M.; writing—original draft, E.N.; writing—review and editing, E.N., K.M., and S.P.; visualization E.N. and S.P.; supervision, J.E.K.; funding acquisition, J.E.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The present study was reviewed and approved as exempt by the Boston University Charles River Campus IRB.

Informed Consent Statement: Informed written consent to participate in the present study was obtained from all the participants.

Data Availability Statement: The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Scales and Measurements

AI Knowledge

In your opinion, do the following technologies use AI?
Note: Randomly show 5 out of the 14.

1. A website or application that translates languages (for example, Google Translate).
2. An application that identifies or categorizes people and objects in your photos or videos.
3. A self-driving car.
4. A chatbot that offers advice or customer support.
5. A digital personal assistant on your phone or a device in your home that can help schedule meetings, answer questions, and complete tasks (for example, Google Home).
6. A social robot that can interact with humans.

7. A website or application that recommends movies or television shows based on your prior viewing habits.
8. A website that suggests advertisements for you based on your browser history.
9. A search engine (for example, Google).
10. An industrial robot, such as those used in manufacturing.
11. A website or application where users can input text to generate images (e.g., DALL-E).
12. A chatbot that can generate essays, poems, and computer code based on users' input (e.g., ChatGPT).
13. A calculator that can do basic math (add, subtract, multiply, divide, etc.).
14. A set of rules that determines whether students receive a college scholarship based on their high school grades and SAT scores.

Answer choices:

- Yes.
- No.
- I don't know.

AI Familiarity

What best describes your experience with Artificial Intelligence (AI)?
Answer choices:

- I have never heard of AI.
- I have heard about AI in the news, from friends or family, etc.
- I closely follow AI-related news.
- I have some formal education or work experience relating to AI.
- I have extensive experience in AI research and/or development.

AI Experience

Have you ever used an online application or tool where you input text to generate new text (ChatGPT) or images (DALL-E, Midjourney, Stable Diffusion)?
Answer choices:

- Yes.
- No.
- I don't know.

AI Domains—Education

Imagine that a school wants to incorporate AI in various aspects of educational support for students and teachers. It is considering a number of AI uses.

Appendix A.1. Capability

Thinking about AI that is used for (option from field 1), how capable do you think this AI is for (option from field 2)?
Answer choices:

- Entirely capable.
- Mostly capable.
- Somewhat capable.
- Only a little capable.
- Not at all capable.
- I don't know.

Appendix A.2. Benevolence

When used for (option from field 1), how much do you agree or disagree that this use of AI will be in the best interest of the students?
Answer choices:

- Strongly disagree.
- Disagree.
- Neither disagree nor agree.
- Agree.
- Strongly agree.
- I don't know.

Table A1. Education tasks options.

Field 1	Field 2
Drafting lesson plans for teachers	Drafting lesson plans
Grading students' work	Grading
Drafting essays for students	Drafting essays
Providing answers to students' questions	Providing answers
Giving learning support based on individual students' needs	Giving learning support

Appendix B. Healthcare

Imagine a healthcare provider wants to incorporate AI in its medical practices. It is considering a number of AI uses.

Appendix B.1. Capability

Thinking about AI that is used for (option from field 1), how capable do you think this AI is for (option from field 2)?

Answer choices:

- Entirely capable.
- Mostly capable.
- Somewhat capable.
- Only a little capable.
- Not at all capable.
- I don't know.

Appendix B.2. Benevolence

When used for (option from field 1), how much do you agree or disagree that this use of AI will be in the best interest of the patients?

Answer choices:

- Strongly disagree.
- Disagree.
- Neither disagree nor agree.
- Agree.
- Strongly agree.

Table A2. Healthcare tasks options.

Field 1	Field 2
Providing answers to patients' medical questions	Answering patients' medical questions
Helping healthcare professionals diagnose diseases	Helping diagnose diseases
Assisting healthcare professionals in medical research	Assisting medical research
Conversing with patients in therapy settings	Conversing in therapy
Using patient data to determine health risks	Determining health risks

Appendix C. Creativity

We'd like to know your thoughts about how different creative industries are experimenting with AI.

Appendix C.1. Capability

Thinking about AI that is used for (option from field 1), how capable do you think this AI is for (option from field 2)?

Answer choices:

- Entirely capable.
- Mostly capable.
- Somewhat capable.
- Only a little capable.
- Not at all capable.
- I don't know.

Appendix C.2. Benevolence

When used for (option from field 1), how much do you agree or disagree that this use of AI will be in the best interest of the audience?

Answer choices:

- Strongly disagree.
- Disagree.
- Neither disagree nor agree.
- Agree.
- Strongly agree.
- I don't know.

Table A3. Creativity tasks options.

Field 1	Field 2
Content creation (e.g., generating blog posts)	Content creation
Creative writing (e.g., generating scripts or novels)	Creative writing
Music composition (e.g., generating lyrics or melodies)	Music composition
Creating art (e.g., generating digital paintings)	Creating art
Creating videos (e.g., generating video content)	Creating videos

OPTIONAL: Understanding your opinion is very important to us. If there's anything you'd like to add about the use of AI in various applications, please elaborate in the text box below.

Appendix D. Principal Components Analysis

Table A4. PCA.

	Kaiser–Meyer–Olkin Measure of Sampling Adequacy	Bartlett's Test of Sphericity χ^2(df)	Eigenvalue	Variance Explained	Factor Loadings (Range)
Capability					
Creativity	0.89	1127.87(10) ***	3.66	73.26%	0.84–0.86
Education	0.87	1099.49(10) ***	3.55	71.03%	0.82–0.86
Healthcare	0.89	1236.79(10) ***	3.69	73.75%	0.81–0.89
Benevolence					
Creativity	0.9	1619.71(10) ***	4.07	81.39%	0.88–0.93
Education	0.87	909.85(10) ***	3.38	67.57%	0.80–0.84
Healthcare	0.89	1338.87(10) ***	3.76	75.20%	0.81–0.91

*** $p < 0.001$.

References

1. Grace, K.; Salvatier, J.; Dafoe, A.; Zhang, B.; Evans, O. When Will AI Exceed Human Performance? Evidence from AI Experts. *J. Artif. Intell. Res.* **2018**, *62*, 729–754. [CrossRef]
2. Hsu, T.; Myers, S.L. Can We No Longer Believe Anything We See? *The New York Times*, 8 April 2023. Available online: https://www.nytimes.com/2023/04/08/business/media/ai-generated-images.html (accessed on 26 August 2023).
3. Lonas, L. Professor Attempts to Fail Students After Falsely Accusing Them of Using Chatgpt to Cheat. *The Hill*, 18 May 2023. Available online: https://thehill.com/homenews/education/4010647-professor-attempts-to-fail-students-after-falsely-accusing-them-of-using-chatgpt-to-cheat/ (accessed on 26 August 2023).
4. Shaffi, S. "It's the Opposite of Art": Why Illustrators Are Furious About AI. *The Guardian*, 8 April 2023. Available online: https://www.theguardian.com/artanddesign/2023/jan/23/its-the-opposite-of-art-why-illustrators-are-furious-about-ai (accessed on 26 August 2023).
5. Tyson, A.; Kikuchi, E. Growing Public Concern About the Role of Artificial Intelligence in Daily Life. *Pew Research Center*, 2023. Available online: https://policycommons.net/artifacts/4809713/growing-public-concern-about-the-role-of-artificial-intelligence-in-daily-life/5646039/ (accessed on 11 January 2024).
6. Faverio, M.; Tyson, A. What the Data Says About Americans' Views of Artificial Intelligence. 2023. Available online: https://www.pewresearch.org/short-reads/2023/11/21/what-the-data-says-about-americans-views-of-artificial-intelligence/ (accessed on 11 January 2024).
7. The White House. Blueprint for an AI Bill of Rights. Office of Science and Technology Policy. 2022. Available online: https://www.whitehouse.gov/ostp/ai-bill-of-rights/ (accessed on 11 January 2024).
8. The Act | EU Artificial Intelligence Act. European Commission. 2021. Available online: https://artificialintelligenceact.eu/the-act/ (accessed on 11 January 2024).
9. Rebedea, T.; Dinu, R.; Sreedhar, M.; Parisien, C.; Cohen, J. Nemo guardrails: A toolkit for controllable and safe llm applications with programmable rails. *arXiv* **2023**, arXiv:2310.10501. [CrossRef]
10. Christian, B. *The Alignment Problem: Machine Learning and Human Values*; WW Norton & Company: New York, NY, USA, 2020.
11. Bommasani, R.; Hudson, D.A.; Adeli, E.; Altman, R.; Arora, S.; von Arx, S.; Bernstein, M.S.; Bohg, J.; Bosselut, A.; Brunskill, E.; et al. On the opportunities and risks of foundation models. *arXiv* **2021**, arXiv:2108.07258. [CrossRef]
12. Xu, W. Toward human-centered AI. *Interactions* **2019**, *26*, 42–46. [CrossRef]
13. Ozmen Garibay, O.; Winslow, B.; Andolina, S.; Antona, M.; Bodenschatz, A.; Coursaris, C.; Xu, W. Six Human-Centered Artificial Intelligence Grand Challenges. *Int. J. Hum. Comput. Interact.* **2023**, *39*, 391–437. [CrossRef]
14. Choung, H.; Seberger, J.S.; David, P. When AI is Perceived to Be Fairer than a Human: Understanding Perceptions of Algorithmic Decisions in a Job Application Context. *SSRN Electron. J.* **2023**. [CrossRef]
15. Lee, J.D.; See, K.A. Trust in automation: Designing for appropriate reliance. *Hum. Factors* **2004**, *46*, 50–80. [CrossRef]
16. Cheng, X.; Zhang, X.; Cohen, J.; Mou, J. Human vs. AI: Understanding the impact of anthropomorphism on consumer response to chatbots from the perspective of trust and relationship norms. *Inf. Process. Manag.* **2022**, *59*, 102940. [CrossRef]
17. Angerschmid, A.; Zhou, J.; Theuermann, K.; Chen, F.; Holzinger, A. Fairness and explanation in AI-informed decision making. *Mach. Learn. Knowl. Extr.* **2022**, *4*, 556–579. [CrossRef]
18. Shin, D. The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *Int. J. Hum. Comput. Stud.* **2021**, *146*, 102551. [CrossRef]
19. Pickering, B. Trust, but verify: Informed consent, AI technologies, and public health emergencies. *Future Internet* **2021**, *13*, 132. [CrossRef]
20. Bedué, P.; Fritzsche, A. Can we trust AI? An empirical investigation of trust requirements and guide to successful AI adoption. *J. Enterp. Inf. Manag.* **2022**, *35*, 530–549. [CrossRef]
21. Mayer, R.C.; Davis, J.H.; Schoorman, F.D. An Integrative Model of Organizational Trust. *Acad. Manag. Rev.* **1995**, *20*, 709–734. [CrossRef]
22. Choung, H.; David, P.; Seberger, J.S. A multilevel framework for AI governance. *arXiv* **2023**, arXiv:2307.03198. [CrossRef]
23. Rheu, M.; Shin, J.Y.; Peng, W.; Huh-Yoo, J. Systematic review: Trust-building factors and implications for conversational agent design. *Int. J. Hum. Comput. Interact.* **2021**, *37*, 81–96. [CrossRef]
24. Agarwal, N.; Moehring, A.; Rajpurkar, P.; Salz, T. *Combining Human Expertise with Artificial Intelligence: Experimental Evidence from Radiology*; Technical Report; National Bureau of Economic Research: Cambridge, MA, USA, 2023. [CrossRef]
25. de Haan, Y.; van den Berg, E.; Goutier, N.; Kruikemeier, S.; Lecheler, S. Invisible friend or foe? How journalists use and perceive algorithmic-driven tools in their research process. *Digit. J.* **2022**, *10*, 1775–1793. [CrossRef]
26. Balaram, B.; Greenham, T.; Leonard, J. Artificial Intelligence: Real Public Engagement. 2018. Available online: https://www.thersa.org/reports/artificial-intelligence-real-public-engagement (accessed on 26 August 2023).
27. Copeland, B.J. Artificial Intelligence. Encyclopaedia Britannica. 2023. Available online: https://www.britannica.com/technology/artificial-intelligence/Reasoning (accessed on 26 August 2023).
28. West, D.M. *What Is Artificial Intelligence*; Brookings Institution: Washington, DC, USA, 2018. Available online: https://www.brookings.edu/articles/what-is-artificial-intelligence/ (accessed on 12 January 2024).
29. Russell, S.J.; Norvig, P. *Artificial Intelligence: A Modern Approach*; Pearson: London, UK, 2010.

30. Goodfellow, I.; Bengio, Y.B.; Courville, A. *Adaptive Computation and Machine Learning Series (Deep Learning)*; MIT Press: Cambridge, MA, USA, 2016. [CrossRef]
31. Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.D.; Dhariwal, P.; Sastry, G.; Askell, A.; Agarwal, S. Language models are few-shot learners. In Proceedings of the Advances in Neural Information Processing Systems, Virtual, 6–12 December 2020; Volume 33, pp. 1877–1901. [CrossRef]
32. Radford, A.; Kim, J.W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sutskever, I. Learning transferable visual models from natural language supervision. In Proceedings of the 2021 International Conference on Machine Learning, Virtual, 18–24 July 2021; pp. 8748–8763.
33. Brock, A.; Donahue, J.; Simonyan, K. Large scale GAN training for high fidelity natural image synthesis. *arXiv* **2018**, arXiv:1809.11096. [CrossRef]
34. Engel, J.; Agrawal, K.K.; Chen, S.; Gulrajani, I.; Donahue, C.; Roberts, A. Gansynth: Adversarial neural audio synthesis. *arXiv* **2019**, arXiv:1902.08710. [CrossRef]
35. Shneiderman, B. *Human-Centered AI*; Oxford University Press: Oxford, UK, 2022.
36. Heaven, W.D. DeepMind'S Cofounder: Generative AI Is Just a Phase. What'S Next Is Interactive AI. MIT Technology Review. 2023. Available online: https://www.technologyreview.com/2023/09/15/1079624/deepmind-inflection-generative-ai-whats-next-mustafa-suleyman/ (accessed on 26 August 2023).
37. Terry, M.; Kulkarni, C.; Wattenberg, M.; Dixon, L.; Morris, M.R. AI Alignment in the Design of Interactive AI: Specification Alignment, Process Alignment, and Evaluation Support. *arXiv* **2023**, arXiv:2311.00710. [CrossRef]
38. Generate Text, Images, Code, and More with Google Cloud AI. 2023. Available online: https://cloud.google.com/use-cases/generative-ai (accessed on 11 September 2023).
39. Kelly, S.M. So Long, Robotic Alexa. Amazon's Voice Assistant Gets More Human-Like with Generative AI. CNN Business. 2023. Available online: https://edition.cnn.com/2023/09/20/tech/amazon-alexa-human-like-generative-ai/index.html (accessed on 11 September 2023).
40. Syu, J.H.; Lin, J.C.W.; Srivastava, G.; Yu, K. A Comprehensive Survey on Artificial Intelligence Empowered Edge Computing on Consumer Electronics. *Proc. IEEE Trans. Consum. Electron.* **2023**. [CrossRef]
41. Chouldechova, A. Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments. *Big Data* **2017**, *5*, 153–163. [CrossRef]
42. Raghavan, M.; Barocas, S.; Kleinberg, J.; Levy, K. Mitigating bias in algorithmic hiring. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, Barcelona, Spain, 27–30 January 2020; ACM: New York, NY, USA, 2020. [CrossRef]
43. Balaji, T.; Annavarapu, C.S.R.; Bablani, A. Machine learning algorithms for social media analysis: A survey. *Comput. Sci. Rev.* **2021**, *40*, 100395. [CrossRef]
44. Chen, C.; Shu, K. Can llm-generated misinformation be detected? *arXiv* **2023**, arXiv:2309.13788. [CrossRef]
45. Sowa, K.; Przegalinska, A.; Ciechanowski, L. Cobots in knowledge work: Human–AI collaboration in managerial professions. *J. Bus. Res.* **2021**, *125*, 135–142. [CrossRef]
46. Tiron-Tudor, A.; Deliu, D. Reflections on the human-algorithm complex duality perspectives in the auditing process. *Qual. Res. Account. Manag.* **2022**, *19*, 255–285.
47. Xu, W.; Dainoff, M.J.; Ge, L.; Gao, Z. Transitioning to human interaction with AI systems: New challenges and opportunities for HCI professionals to enable human-centered AI. *Int. J. Hum. Comput. Interact.* **2023**, *39*, 494–518. [CrossRef]
48. Yang, R.; Wibowo, S. User trust in artificial intelligence: A comprehensive conceptual framework. *Electron. Mark.* **2022**, *32*, 2053–2077. [CrossRef]
49. Lv, X.; Yang, Y.; Qin, D.; Cao, X.; Xu, H. Artificial intelligence service recovery: The role of empathic response in hospitality customers' continuous usage intention. *Comput. Hum. Behav.* **2022**, *126*, 106993. [CrossRef]
50. Lee, J.C.; Chen, X. Exploring users' adoption intentions in the evolution of artificial intelligence mobile banking applications: The intelligent and anthropomorphic perspectives. *Int. J. Bank Mark.* **2022**, *40*, 631–658. [CrossRef]
51. Lu, L.; McDonald, C.; Kelleher, T.; Lee, S.; Chung, Y.J.; Mueller, S.; Yue, C.A. Measuring consumer-perceived humanness of online organizational agents. *Comput. Hum. Behav.* **2022**, *128*, 107092. [CrossRef]
52. Pelau, C.; Dabija, D.C.; Ene, I. What makes an AI device human-like? The role of interaction quality, empathy and perceived psychological anthropomorphic characteristics in the acceptance of artificial intelligence in the service industry. *Comput. Hum. Behav.* **2021**, *122*, 106855. [CrossRef]
53. Shi, S.; Gong, Y.; Gursoy, D. Antecedents of Trust and Adoption Intention Toward Artificially Intelligent Recommendation Systems in Travel Planning: A Heuristic–Systematic Model. *J. Travel Res.* **2021**, *60*, 1714–1734. [CrossRef]
54. Moussawi, S.; Koufaris, M.; Benbunan-Fich, R. How perceptions of intelligence and anthropomorphism affect adoption of personal intelligent agents. *Electron. Mark.* **2021**, *31*, 343–364. [CrossRef]
55. Al Shamsi, J.H.; Al-Emran, M.; Shaalan, K. Understanding key drivers affecting students' use of artificial intelligence-based voice assistants. *Educ. Inf. Technol.* **2022**, *27*, 8071–8091. [CrossRef]
56. He, H.; Gray, J.; Cangelosi, A.; Meng, Q.; McGinnity, T.M.; Mehnen, J. The Challenges and Opportunities of Human-Centered AI for Trustworthy Robots and Autonomous Systems. *IEEE Trans. Cogn. Dev. Syst.* **2022**, *14*, 1398–1412. [CrossRef]
57. Shin, D. User Perceptions of Algorithmic Decisions in the Personalized AI System: Perceptual Evaluation of Fairness, Accountability, Transparency, and Explainability. *J. Broadcast. Electron. Media* **2020**, *64*, 541–565. [CrossRef]

58. Schoeffer, J.; De-Arteaga, M.; Kuehl, N. On explanations, fairness, and appropriate reliance in human-AI decision-making. *arXiv* **2022**, arXiv:2209.11812. [CrossRef]
59. Ghassemi, M.; Oakden-Rayner, L.; Beam, A.L. The false hope of current approaches to explainable artificial intelligence in healthcare. *Lancet Digit. Health* **2021**, *3*, e745–e750. [CrossRef]
60. Muir, B.M. Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics* **1994**, *37*, 1905–1922. [CrossRef]
61. Hoff, K.A.; Bashir, M. Trust in Automation. *Hum. Factors J. Hum. Factors Ergon. Soc.* **2014**, *57*, 407–434. [CrossRef]
62. Thurman, N.; Moeller, J.; Helberger, N.; Trilling, D. My Friends, Editors, Algorithms, and I. *Digit. J.* **2018**, *7*, 447–469. [CrossRef]
63. Smith, A. Public Attitudes toward Computer Algorithms. Policy Commons. 2018. Available online: https://policycommons.net/artifacts/617047/public-attitudes-toward-computer-algorithms/1597791/ (accessed on 24 August 2023).
64. Araujo, T.; Helberger, N.; Kruikemeier, S.; de Vreese, C.H. In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI Soc.* **2020**, *35*, 611–623. [CrossRef]
65. Choung, H.; David, P.; Ross, A. Trust and ethics in AI. *AI Soc.* **2022**, *38*, 733–745. [CrossRef]
66. Mays, K.K.; Lei, Y.; Giovanetti, R.; Katz, J.E. AI as a boss? A national US survey of predispositions governing comfort with expanded AI roles in society. *AI Soc.* **2021**, *37*, 1587–1600. [CrossRef]
67. Novozhilova, E.; Mays, K.; Katz, J. Looking towards an automated future: U.S. attitudes towards future artificial intelligence instantiations and their effect. *Humanit. Soc. Sci. Commun.* **2024**. [CrossRef]
68. Shamout, M.; Ben-Abdallah, R.; Alshurideh, M.; Alzoubi, H.; Al Kurdi, B.; Hamadneh, S. A conceptual model for the adoption of autonomous robots in supply chain and logistics industry. *Uncertain Supply Chain. Manag.* **2022**, *10*, 577–592. [CrossRef]
69. Oliveira, T.; Martins, R.; Sarker, S.; Thomas, M.; Popovič, A. Understanding SaaS adoption: The moderating impact of the environment context. *Int. J. Inf. Manag.* **2019**, *49*, 1–12. [CrossRef]
70. Zhang, B.; Dafoe, A. U.S. Public Opinion on the Governance of Artificial Intelligence. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, Montreal, QC, Canada, 7–8 February 2020; ACM: New York, NY, USA, 2020. [CrossRef]
71. Nazaretsky, T.; Ariely, M.; Cukurova, M.; Alexandron, G. Teachers' trust in AI-powered educational technology and a professional development program to improve it. *Br. J. Educ. Technol.* **2022**, *53*, 914–931. [CrossRef]
72. Nazaretsky, T.; Cukurova, M.; Alexandron, G. An Instrument for Measuring Teachers' Trust in AI-Based Educational Technology. In Proceedings of the LAK22: 12th International Learning Analytics and Knowledge Conference, Online, 21–25 March 2022. [CrossRef]
73. Law, E.L.C.; van As, N.; Følstad, A. Effects of Prior Experience, Gender, and Age on Trust in a Banking Chatbot With(Out) Breakdown and Repair. In Proceedings of the Human-Computer Interaction—INTERACT 2023, Copenhagen, Denmark, 23–28 July 2023; pp. 277–296. [CrossRef]
74. Stanton, B.; Jensen, T. *Trust and Artificial Intelligence*; Technical Report; National Institute of Standards and Technology: Gaithersburg, MD, USA, 2021. [CrossRef]
75. Szalavitz, M.; Rigg, K.K.; Wakeman, S.E. Drug dependence is not addiction—And it matters. *Ann. Med.* **2021**, *53*, 1989–1992. [CrossRef]
76. Okaniwa, F.; Yoshida, H. Evaluation of Dietary Management Using Artificial Intelligence and Human Interventions: Nonrandomized Controlled Trial. *JMIR Form. Res.* **2021**, *6*, e30630. [CrossRef] [PubMed]
77. Zheng, Y.; Tang, N.; Omar, R.; Hu, Z.; Duong, T.; Wang, J.; Haick, H. Smart Materials Enabled with Artificial Intelligence for Healthcare Wearables. *Adv. Funct. Mater.* **2021**, *31*, 2105482. [CrossRef]
78. Huang, P.; Li, S.; Li, P.; Jia, B. The Learning Curve of Da Vinci Robot-Assisted Hemicolectomy for Colon Cancer: A Retrospective Study of 76 Cases at a Single Center. *Front. Surg.* **2022**, *9*, 897103. [CrossRef]
79. Tustumi, F.; Andreollo, N.A.; de Aguilar-Nascimento, J.E. Future of the Language Models in Healthcare: The Role of ChatGPT. *Arquivos Brasileiros de Cirurgia Digestiva* **2023**, *36*, e1727. [CrossRef]
80. Kung, T.H.; Cheatham, M.; Medenilla, A.; Sillos, C.; De Leon, L.; Elepaño, C.; Madriaga, M.; Aggabao, R.; Diaz-Candido, G.; Maningo, J.; et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLoS Digit. Health* **2023**, *2*, e0000198. [CrossRef]
81. West, D.M.; Allen, J.R. *Turning Point: Policymaking in the Era of Artificial Intelligence*; Brookings Institution Press: Washington, DC, USA, 2020.
82. Drukker, K.; Chen, W.; Gichoya, J.; Gruszauskas, N.; Kalpathy-Cramer, J.; Koyejo, S.; Myers, K.; Sá, R.C.; Sahiner, B.; Whitney, H.; et al. Toward fairness in artificial intelligence for medical image analysis: Identification and mitigation of potential biases in the roadmap from data collection to model deployment. *J. Med. Imaging* **2023**, *10*, 061104. [CrossRef]
83. Knight, D.; Aakre, C.A.; Anstine, C.V.; Munipalli, B.; Biazar, P.; Mitri, G.; Halamka, J.D. Artificial Intelligence for Patient Scheduling in the Real-World Health Care Setting: A Metanarrative Review. *Health Policy Technol.* **2023**, *12*, 100824. [CrossRef]
84. Samorani, M.; Harris, S.L.; Blount, L.G.; Lu, H.; Santoro, M.A. Overbooked and overlooked: Machine learning and racial bias in medical appointment scheduling. *Manuf. Serv. Oper. Manag.* **2022**, *24*, 2825–2842. [CrossRef]
85. Li, R.C.; Asch, S.M.; Shah, N.H. Developing a delivery science for artificial intelligence in healthcare. *NPJ Digit. Med.* **2020**, *3*, 107. [CrossRef] [PubMed]
86. Yu, K.H.; Beam, A.L.; Kohane, I.S. Artificial intelligence in healthcare. *Nat. Biomed. Eng.* **2018**, *2*, 719–731. [CrossRef] [PubMed]

87. Baldauf, M.; Fröehlich, P.; Endl, R. Trust Me, I'm a Doctor—User Perceptions of AI-Driven Apps for Mobile Health Diagnosis. In Proceedings of the 19th International Conference on Mobile and Ubiquitous Multimedia, Essen, Germany, 22–25 November 2020; ACM: New York, NY, USA, 2020. [CrossRef]
88. Ghafur, S.; Van Dael, J.; Leis, M.; Darzi, A.; Sheikh, A. Public perceptions on data sharing: Key insights from the UK and the USA. *Lancet Digit. Health* **2020**, *2*, e444–e446. [CrossRef] [PubMed]
89. Rienties, B. Understanding academics' resistance towards (online) student evaluation. *Assess. Eval. High. Educ.* **2014**, *39*, 987–1001. [CrossRef]
90. Johnson, A. Chatgpt in Schools: Here's Where It's Banned-and How It Could Potentially Help Students. *Forbes*, 2023. Available online: https://www.forbes.com/sites/ariannajohnson/2023/01/18/chatgpt-in-schools-heres-where-its-banned-and-how-it-could-potentially-help-students/ (accessed on 31 January 2024).
91. Jones, B.; Perez, J.; Touré, M. More Schools Want Your Kids to Use CHATGPT Really. *Politico*, 2023. Available online: https://www.politico.com/news/2023/08/23/chatgpt-ai-chatbots-in-classrooms-00111662 (accessed on 31 January 2024).
92. Kasneci, E.; Seßler, K.; Küchemann, S.; Bannert, M.; Dementieva, D.; Fischer, F.; Kasneci, G. ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education. *arXiv* **2023**. [CrossRef]
93. Hwang, G.J.; Chen, N.S. Editorial Position Paper. *Educ. Technol. Soc.* **2023**, *26*. Available online: https://www.jstor.org/stable/48720991 (accessed on 31 January 2024).
94. Williams, B.A.; Brooks, C.F.; Shmargad, Y. How algorithms discriminate based on data they lack: Challenges, solutions, and policy implications. *J. Inf. Policy* **2018**, *8*, 78–115. [CrossRef]
95. Hao, K. The UK Exam Debacle Reminds Us That Algorithms Can't Fix Broken Systems. MIT Technology Review. 2020. Available online: https://www.technologyreview.com/2020/08/20/1007502/uk-exam-algorithm-cant-fix-broken-system/ (accessed on 31 January 2024).
96. Engler, A. Enrollment Algorithms Are Contributing to the Crises of Higher Education. Brookings Institute. 2021. Available online: https://www.brookings.edu/articles/enrollment-algorithms-are-contributing-to-the-crises-of-higher-education/ (accessed on 31 January 2024).
97. Lo, C.K. What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Educ. Sci.* **2023**, *13*, 410. [CrossRef]
98. Conijn, R.; Kahr, P.; Snijders, C. The Effects of Explanations in Automated Essay Scoring Systems on Student Trust and Motivation. *J. Learn. Anal.* **2023**, *10*, 37–53. [CrossRef]
99. Qin, F.; Li, K.; Yan, J. Understanding user trust in artificial intelligence-based educational systems: Evidence from China. *Br. J. Educ. Technol.* **2020**, *51*, 1693–1710. [CrossRef]
100. Vinichenko, M.V.; Melnichuk, A.V.; Karácsony, P. Technologies of improving the university efficiency by using artificial intelligence: Motivational aspect. *Entrep. Sustain. Issues* **2020**, *7*, 2696. [CrossRef] [PubMed]
101. Ramesh, A.; Dhariwal, P.; Nichol, A.; Chu, C.; Chen, M. Hierarchical text-conditional image generation with clip latents. *arXiv* **2022**, arXiv:2204.06125. Available online: https://3dvar.com/Ramesh2022Hierarchical.pdf (accessed on 31 January 2024).
102. Roose, K. GPT-4 Is Exciting And Scary. *The New York Times*, 2023. Available online: https://www.nytimes.com/2023/03/15/technology/gpt-4-artificial-intelligence-openai.html (accessed on 31 January 2024).
103. Plut, C.; Pasquier, P. Generative music in video games: State of the art, challenges, and prospects. *Entertain. Comput.* **2020**, *33*, 100337. [CrossRef]
104. Zhou, V.; Dosunmu, D.; Maina, J.; Kumar, R. AI Is Already Taking Video Game Illustrators' Jobs in China. Rest of World. 2023. Available online: https://restofworld.org/2023/ai-image-china-video-game-layoffs/ (accessed on 31 January 2024).
105. Coyle, J.; Press, T.A. Chatgpt Is the "Terrifying" Subtext of the Writers' Strike That Is Reshaping Hollywood. *The Fortune*, 2023. Available online: https://fortune.com/2023/05/05/hollywood-writers-strike-wga-chatgpt-ai-terrifying-replace-writers/ (accessed on 31 January 2024).
106. Bianchi, F.; Kalluri, P.; Durmus, E.; Ladhak, F.; Cheng, M.; Nozza, D.; Hashimoto, T.; Jurafsky, D.; Zou, J.; Caliskan, A. Easily Accessible Text-to-Image Generation Amplifies Demographic Stereotypes at Large Scale. In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, Chicago, IL, USA, 12–15 June 2023; ACM: New York, NY, USA, 2023. [CrossRef]
107. Srinivasan, R.; Uchino, K. Biases in Generative Art. In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual, 3–10 March 2021; ACM: New York, NY, USA, 2021. [CrossRef]
108. Romero, A. How to Get the Most Out of Chatgpt. *The Algorithmic Bridge*, 2022. Available online: https://thealgorithmicbridge.substack.com/p/how-to-get-the-most-out-of-chatgpt (accessed on 31 January 2024).
109. Lyu, Y.; Wang, X.; Lin, R.; Wu, J. Communication in Human–AI Co-Creation: Perceptual Analysis of Paintings Generated by Text-to-Image System. *Appl. Sci.* **2022**, *12*, 11312. [CrossRef]
110. Mazzone, M.; Elgammal, A. Art, Creativity, and the Potential of Artificial Intelligence. *Arts* **2019**, *8*, 26. [CrossRef]
111. Rasrichai, K.; Chantarutai, T.; Kerdvibulvech, C. Recent Roles of Artificial Intelligence Artists in Art Circulation. *Digit. Soc.* **2023**, *2*, 15. [CrossRef]
112. Bosonogov, S.D.; Suvorova, A.V. Perception of AI-generated art: Text analysis of online discussions. *Sci. Semin. Pom* **2023**, *529*, 6–23. Available online: https://www.mathnet.ru/eng/znsl7416 (accessed on 31 January 2024).

113. Alves da Veiga, P. Generative Ominous Dataset: Testing the Current Public Perception of Generative Art. In Proceedings of the 20th International Conference on Culture and Computer Science: Code and Materiality, Lisbon, Portugal, 28–29 September 2023; pp. 1–10. [CrossRef]
114. Katz, J.E.; Halpern, D. Attitudes towards robots suitability for various jobs as affected robot appearance. *Behav. Inf. Technol.* **2013**, *33*, 941–953. [CrossRef]
115. Kates, S.; Ladd, J.; Tucker, J.A. How Americans' Confidence in Technology Firms Has Dropped: Evidence From the Second Wave of the American Institutional Confidence Poll. *Brookings Institute*, 2023. Available online: https://www.brookings.edu/articles/how-americans-confidence-in-technology-firms-has-dropped-evidence-from-the-second-wave-of-the-american-institutional-confidence-poll/ (accessed on 31 January 2024).
116. Yin, J.; Chen, Z.; Zhou, K.; Yu, C. A deep learning based chatbot for campus psychological therapy. *arXiv* **2019**, arXiv:1910.06707. [CrossRef]
117. Bidarian, N. Meet Khan Academy's Chatbot Tutor. *CNN*, 2023. Available online: https://www.cnn.com/2023/08/21/tech/khan-academy-ai-tutor/index.html (accessed on 31 January 2024).
118. Mays, K.K.; Katz, J.E.; Lei, Y.S. Opening education through emerging technology: What are the prospects? Public perceptions of Artificial Intelligence and Virtual Reality in the classroom. *Opus Educ.* **2021**, *8*, 28. [CrossRef]
119. Molina, M.D.; Sundar, S.S. Does distrust in humans predict greater trust in AI? Role of individual differences in user responses to content moderation. *New Media Soc.* **2022**. [CrossRef]
120. Raj, M.; Berg, J.; Seamans, R. Art-ificial Intelligence: The Effect of AI Disclosure on Evaluations of Creative Content. *arXiv* **2023**, arXiv:2303.06217. [CrossRef]
121. Heikkilä, M. AI Literacy Might Be CHATGPT'S Biggest Lesson for Schools. *MIT Technology Review*, 2023. Available online: https://www.technologyreview.com/2023/04/12/1071397/ai-literacy-might-be-chatgpts-biggest-lesson-for-schools/ (accessed on 31 January 2024).
122. Elements of AI. A Free Online Introduction to Artificial Intelligence for Non-Experts. 2023. Available online: https://www.elementsofai.com/ (accessed on 31 January 2024).
123. U.S. Department of Education. *Artificial Intelligence and the Future of Teaching and Learning*; Office of Educational Technology U.S. Department of Education: Washington, DC, USA, 2023. Available online: https://www2.ed.gov/documents/ai-report/ai-report.pdf (accessed on 31 January 2024).
124. U.S. Food and Drug Administration. *FDA Releases Artificial Intelligence/Machine Learning Action Plan*; U.S. Food and Drug Administration: Washington, DC, USA, 2023.
125. Office of the Chief Information Officer Washington State. *Generative AI Guidelines*; Office of the Chief Information Officer Washington State: Washington, DC, USA, 2023. Available online: https://ocio.wa.gov/policy/generative-ai-guidelines (accessed on 31 January 2024).
126. Holt, K. You Can't Copyright AI-Created Art, According to US Officials. *Engadget*, 2022. Available online: https://www.engadget.com/us-copyright-office-art-ai-creativity-machine-190722809.html (accessed on 31 January 2024).
127. Engler, A. The AI Bill of Rights Makes Uneven Progress on Algorithmic Protections. *Lawfare Blog*, 2022. Available online: https://www.lawfareblog.com/ai-bill-rights-makes-uneven-progress-algorithmic-protections (accessed on 31 January 2024).
128. Ye, J. China Says Generative AI Rules to Apply Only to Products for the Public. *Reuters*, 2023. Available online: https://www.reuters.com/technology/china-issues-temporary-rules-generative-ai-services-2023-07-13/ (accessed on 31 January 2024).
129. Meaker, M. The EU Just Passed Sweeping New Rules to Regulate AI. *Wired*, 2023. Available online: https://www.wired.com/story/eu-ai-act/ (accessed on 31 January 2024).
130. Ernst, C. Artificial Intelligence and Autonomy: Self-Determination in the Age of Automated Systems. In *Regulating Artificial Intelligence*; Springer International Publishing: Berlin/Heidelberg, Germany, 2019; pp. 53–73. [CrossRef]
131. van der Veer, S.N.; Riste, L.; Cheraghi-Sohi, S.; Phipps, D.L.; Tully, M.P.; Bozentko, K.; Peek, N. Trading off accuracy and explainability in AI decision-making: Findings from 2 citizens' juries. *J. Am. Med Informatics Assoc.* **2021**, *28*, 2128–2138. [CrossRef] [PubMed]
132. Engler, A. The EU and US Diverge on AI Regulation: A Transatlantic Comparison and Steps to Alignment. *Brookings Institute*, 2023. Available online: https://www.brookings.edu/articles/the-eu-and-us-diverge-on-ai-regulation-a-transatlantic-comparison-and-steps-to-alignment/ (accessed on 31 January 2024).
133. Dreksler, N.; McCaffary, D.; Kahn, L.; Mays, K.; Anderljung, M.; Dafoe, A.; Horowitz, M.; Zhang, B. Preliminary Survey Results: US and European Publics Overwhelmingly and Increasingly Agree That AI Needs to Be Managed Carefully. Centre for the Governance of AI. 2023. Available online: https://www.governance.ai/post/increasing-consensus-ai-requires-careful-management (accessed on 31 January 2024).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.