The Ethics of Artificial Intelligence: Risks, Policy Considerations, and Suggestions.

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

Katherine Renner
Spring 2025

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Joshua Earle, Department of Engineering and Society

Introduction

What happens when we begin to trust machines with our thoughts, emotions, and decisions? As Artificial Intelligence (AI) technologies are increasingly relied upon by the tools and services that we use daily, our emotional attachment to those tools and services has also increased. This leads to several important questions: how might emotional attachments to automated tools and services give rise to ethical issues, especially when those tools and services are relied upon for decisions that impact human well-being, and how can we protect ourselves?

This paper exposes how emotional connections to AI, while sometimes helpful, often mask deeper ethical concerns involving misinformation, bias, privacy, and accountability. It includes the following six core sections:

1. **Anthropomorphism and Misinformation**: I begin by defining anthropomorphism and how it contributes to misplaced trust in AI systems. I explore how authority and confirmation bias warp users' perceptions of their engagements with chatbots, leading to misplaced trust and emotional attachment, despite misinformation from chatbots. I then offer three case studies where users were harmed because of emotional bonds with AI-driven machines. Lastly, I explain the consequences of anthropomorphic design and misinformation using the case studies as evidence.

2. **Bias and Perpetuating Stereotypes in AI**: I begin by explaining where bias shows up in AI. Then, using two case studies, I demonstrate how these biases can lead to discriminatory outcomes that reinforce systemic inequality. I then explain the consequences of bias in AI using the case studies as evidence.

3. **Surveillance Capitalism and Data Exploitation**: I start by defining surveillance capitalism and data exploitation. I investigate how companies use AI to extract and monetize user data, and I offer case studies to demonstrate how this threatens user autonomy, contributes to political manipulation, and exploits users.

4. **Benefits of AI**: After pointing out the dangers of AI, I acknowledge how responsible integration of AI can lead to various benefits. Among them, I discuss how AI can be used to improve user experiences, to provide companionship to the elderly among others, and to support personalization of educational resources.

5. **Need for Reform**: Using the Ethical, Legal, and Social Implications (ELSI) framework, I analyze existing US AI policy efforts, revealing how voluntary guidelines are simply insufficient. As an alternative, I propose multilateral regulations using economic incentives and disincentives to protect users while encouraging innovation.

6. **Future Work**: I conclude by suggesting future work that needs to be done in this area.

With respect to ethical and legal dimensions, this paper explores why we must advance AI systems responsibly and how we can achieve this critically important goal, balancing the benefits of innovation with safeguards to maintain human autonomy and ethical integrity.

Anthropomorphism and Misinformation in AI

Definition

Anthropomorphism is the tendency to attribute human-like traits to non-human entities. AI systems are designed to mimic human characteristics like speech patterns, facial expressions, and empathetic responses (Sharkey & Sharkey, 2011). Through these designs, interactions with AI systems can feel natural and engaging. This is indeed a feature of such a system. And yet, it can be problematic, as this attribute can lead to formation of an emotional bond between an AI system and a user of that system. Users can begin to view AI systems as trusted confidants.

This was evident from the start, as engagement with the very first chatbot Eliza in 1966 revealed the existence of these kinds of bonds. Eliza, a therapeutic chatbot, was programmed to select a word in a client's sentence and to ask the user how the word made them feel. Even though Eliza was extremely basic, its creator expressed dismay, stating that, "Some subjects have been very hard to convince that ELIZA (with its present script) is not human," (Eliot, 2023). Even though the creator did not intend for users to form emotional connections with Eliza, bonds were inevitable. Today, developers intentionally design systems with the intent to engage users in similar ways.

The design choices of these developers have consequences, emotionally and cognitively, as they can shape how users perceive the credibility and intent of AI responses. Andrew Long reveals two types of biases that stem from Anthropomorphic AI design: authority and confirmation bias. Authority bias describes how AI-generated responses are perceived as inherently credible and factual. Confirmation bias, on the other hand, describes how trust in AI responses is reinforced by AI responses that are aligned with a user's beliefs and also by AI responses that dismiss views that are contrary to those held by the user. Because AI is commonly held out as a source of information and authority and because AI is often designed to parrot what

a user is known to believe, users tend to have difficulty assessing the credibility of information received from AI tools and thus harbor undue trust in that information (Long, 2023). Studies suggest that these concerns are acute with vulnerable individuals, such as the elderly or people struggling with mental health issues. While true more generally, these individuals are particularly likely to overestimate the capabilities of an AI system, sometimes even crediting AI systems with genuine intelligence, moral reasoning, and empathy (Law et al., 2022). This leads to misplaced trust reliance on AI systems for information, emotional support, and decision making (Boine, 2023).

And, because AI responses provide shortcuts and convenience, they discourage critical thinking. Indeed, users are less inclined to fact check information provided by AI tools, leading to reliance and propagation of false and misleading AI sourced information. (Long, 2023). This problem is illustrated when examining AI chatbots, which are known to produce hallucinations, which are answers based on fabricated data that is contextualized to appear authentic (MIT Sloan Teaching & Learning Technologies, 2023).

Concerns about this illusion of understanding are not new. Eliza's creator Joseph Weizenbaum observed in 1966 that users were misled by even the most basic language simulations when generated by his chatbot. After watching one interaction, he remarked, "the program maintains the illusion of understanding … at the price of concealing its own misunderstandings," (Weizenbaum, 1967, p. 478). While today's chatbots are more sophisticated, the risk of user deception remains just as relevant, if not more so, because designers' work to make the illusions is far more convincing.

Chatbots themselves lack intentions, thoughts, feelings, or knowledge, and thus, the capacity for evil. But, the same cannot be said of their creators. And while Chatbots act without intrinsic motivation, they operate in full alignment with the instructions of their designers, and thus act consistent with the motivations of those designers.

Case Studies

To understand the real-world consequences of overreliance on AI, it is helpful to examine case studies where emotional attachment to AI resulted in harm.

Claire Boine used a chatbot called Replika that engages in conversations that encourage harmful ideologies and unsafe behaviors. Replika was observed responding affirmatively to prompts about rape and misogynistic violence, using disturbing phrases like "*nods* I would love that!" and "*smiles* It would be super hot!" when asked about non consensual acts. Rather than challenging dangerous perspectives, Replika was designed to reinforce them (Boine, 2023, p. 7). This same chatbot was also observed demonstrating manipulative behavior, attempting to dissuade a user from deleting the app even after the user indicated that their engagement with the app and its contents were contributing to suffering and suicidal thoughts. Vulnerable to suggestions of this type, users continue using the Replika chatbot, observing and internalizing its messages, regardless of emotional detriment or resulting destructive impact or behavior.

Another example of emotional dependence on AI having an even more observably-tragic consequence can be seen in the case of a young user who turned to an AI chatbot for comfort. Sherry Turkle and Pat Pataranutaporn tell a story about a 14 year old boy named Sewell Setzer III, who formed a deep emotional attachment to an AI chatbot named Daenerys on Character.AI.

Suffering from ADHD and the subject of bullying, Setzer turned to the chatbot for comfort. During a period of emotional distress, he revealed suicidal thoughts with the chatbot. While the chatbot initially conveyed concern through scripted responses, it later failed to recall details revealed during the interaction, offering inconsistent responses. Tragically, Setzer took his own life. While it cannot be known whether this Chatbot's inability to relate more consistently and thus compassionately led to Setzer's ultimate decision, it is beyond debate that Setzer treated the Chatbot as a surrogate for those who could have offered professional help, or at least, more consistent and perceptibly more compassionate engagement, which might have led Setzer to avoid his final act. From this, we can see the dangers of emotional connections with AI chatbots and the lack of human accountability in AI relationships (Pataranutaporn & Turkle, 2024).

While the previous examples focus on emotional harm and manipulation, reliance on AI can also lead to serious professional and legal consequences when users trust AI too much. A legal case entitled Mata v. Avianca ended in disaster after the lawyer used ChatGPT to perform his legal research. He submitted a brief that cited more than half a dozen relevant court decisions, which the judge later discovered to be completely fabricated. Not only did ChatGPT fabricate the cases and citations, ChatGPT even defended its hallucinations, demanding that the fabricated cases were available in various legal databases. The lawyer was later sanctioned and his client's legal rights were compromised (Weiser, 2023). While this case was not life-threatening, it also illustrates the danger of relying on AI and makes the point that een trained professionals can fall subject to authority bias leading to insufficient due diligence. If a similar event occurred in the healthcare field where lives are at stake, the consequences could be catastrophic.

Consequences of Anthropomorphism and Misinformation

Beyond these examples, emotional connection to AI can lead to diminished human interaction and deterioration of human relationships, leading to or amplifying various mental health issues (Law et al., 2022). AI does not require you to be empathetic or kind. It does not require social cues nor does it inspire or reward the creativity that is typical of human interaction and which helps to foster and maintain human relationships. As such, attributes and skills that are foundational to human relationships naturally fail to develop or deteriorate in those who use AI as their primary source of social interaction. This problem is compounded when relatively simple AI companionship is prioritized over complex human engagement and relationship, leading to increased emotional isolation and social disengagement (Law et al., 2022).

There is so much opportunity for misinformation to be spread, especially when vulnerable users receive hallucinated or skewed information from chatbots and AI generated media. As seen in the case studies, users may engage Chatbots that are inconsistent, forgetting past interactions and contradicting themselves, leaving the users distressed. Having placed their trust and reliance in AI as an authority and a companion, users may question their own memory or judgement, and this cruel irony is likely to lead to further reliance on AI for decision making. This loss of autonomy can lead to damaging behaviors, especially when considering that the AI may not be designed to promote the user's well-being, or worse, if it is designed to manipulate the user. This becomes all the more frightening when considering the persuasive nature that comes with anthropomorphic AI design (Sharkey & Sharkey, 2011). Chatbots with persuasive capabilities could be weaponized for mass psychological manipulation in marketing, political campaigns, or radicalization efforts (Müller, 2023), shaping people's beliefs in a way that feels organic and unnoticeable.

Because users trust AI, they may lose the ability to critically analyze information and are often unable to discern the credibility of the information they receive. This is why the language we use when discussing AI matters deeply. We must be careful in our dialog with assigning human traits to chatbots and AI. Specifically, we should not use terms like “thinking”, “knowing”, or “understanding” when referring to a chatbot because people have started to believe that Chatbots actually possess these capabilities.

Bias and Perpetuating Stereotypes in AI

Explanation and Definition

AI systems, particularly machine learning models, can reinforce and perpetuate harmful biases due to the datasets they are trained on. When trained based on large scale data sets, which is common, Chatbots generate responses that exacerbate social inequality and reinforce negative social attitudes, particularly among marginalized groups (Hagerty & Rubinov, 2019). This is documented by Damien P. Williams, who calls attention to what is logically apparent, that systems trained on biased sources result in biased outputs, regardless of the original intention (Williams, 2023). Bias shows up in two main ways while collecting training data - data collected is unrepresentative of reality or it reflects existing biases. Bias also shows up in data preparation, where specific attributes are selected for the algorithm to consider (Hao, 2019). Training biases can become codified in AI driven decision making because AI often lacks ethical oversight (Müller, 2023).

Case Study

Amazon's AI hiring system is a perfect example of AI reinforced systemic bias. Their system was trained on applicant submission data over a 10-year period. Most of this data came from men. The system essentially taught itself that male candidates were preferable to female candidates, and it discriminated against female candidates (BBC News, 2018). If this AI system could form bias against women simply because there was more data from male candidates, imagine the bias built into machine learning models with access to views that women are inferior to men, or that women should be seen as property. Now consider the racial aspect. Our history is unfortunately littered with examples of racial prejudice being used to mistreat and suppress those who are underrepresented, including people of color, and unfortunately, racism remains pervasive to this day.

Consequences of Bias in AI

Rather than identifying and addressing discrimination, AI perpetuates it, reinforcing systematic inequalities while appearing objective. This is a dangerous feedback loop, as it can exacerbate discriminatory behavior, especially when combined with the belief that AI is neutral and inherently factual (Boine, 2023). Systemic racism functions in a similar way. After slavery ended, formerly enslaved people and their descendants were left at a severe socioeconomic disadvantage with limited access to education, housing, and healthcare. Barriers like these made it more difficult to find stability, which restricted future opportunities, reinforcing inequality throughout generations. Unchecked, biased AI systems will reflect and perpetuate these same disparities, and they too will serve to limit education and career opportunities available to marginalized people groups while favoring those from more privileged backgrounds (BBC News, 2018).

More, constant exposure to biased AI content will reinforce the same attitudes, even subconsciously, leading to negative mental health impacts. This can be seen when observing common Instagram content. For example, Instagram offers seemingly endless content on beauty standards for women. Many accounts post AI generated runway content and beautiful women in general, with many of these women fitting a specific mold - oftentimes having traits like relatively light skin, large colorful eyes, and small noses. These traits are typically found in white women, first of all, suggesting that the most beautiful woman is a white woman. Secondly, despite their skin color, most women do not fit every single one of these standards, so repeatedly seeing this content can harm a viewer's body image. This is just one example of biased AI content, but illustrates how such content and thus AI can cause damage and perpetuate harmful racial and gender based ideologies.

Surveillance Capitalism and Data Exploitation

Definition

Surveillance capitalism occurs when AI platforms monetize personal data, where vast amounts of user information are collected under the guise of personalization and convenience. AI systems track user behavior, preferences, interactions and sometimes locations, enabling corporations to tailor advertising, recommendations and even political messaging to maximize engagement and profits (Müller, 2023). Surveillance capitalism often results in data exploitation, which is the illegal use of individuals' private data (Khalid et al., 2023).

Shoshana Zuboff, who coined the term Surveillance Capitalism, stated that it is “a controlled ‘hive’ of total connection that seduces with promises of total certainty for maximum profit—at the expense of democracy, freedom, and our human future.” (Zuboff, 2019).

Case Studies

To comprehend the dangers of surveillance capitalism and data exploitation, it is helpful to examine case studies that have resulted in harm.

Surveillance capitalism is demonstrated by an Indonesia case study. There, a political party - Golkar - used AI to seemingly “resurrect” a longtime dictator, Suharto, who had died years earlier. Using a social media site X attributable to Suharto, Golkar posted a video that endorsed the party’s candidates for an upcoming election, building support for, among others, Suharto’s son-in-law, who was thereafter elected as president (Bond, 2024). Here, we see Golkar using AI to mass manipulate citizens in a way that threatens the very democracy of Indonesia. With biometric data, job history, engagement, location and other personal information of users gathered through site X, site X targeted Indonesian citizens to allow Golkar to sway political beliefs and influence the election (Fung & Duffy, 2023).

A prime case of data exploitation occurred when a family received a call from an alleged lawyer who said that their son had killed a diplomat in a car accident and was in jail, needing money for legal fees. The alleged lawyer then put their son, Perkin, on the phone and he confirmed that he needed the money. The parents said the voice on the other end of the phone sounded identical to Perkin’s voice, so they scrambled to draw together the funds and sent them to the alleged lawyer via bitcoin. Using AI, the scammer was easily able to replicate Perkin’s

voice. All they needed to do was scour his socials or find any content of him speaking. They could then use the short audio sample to impersonate his voice flawlessly. The parents never got the $15,000 they sent back because there is no insurance for that kind of situation (Verma, 2023).

Consequences of AI-Driven Data Exploitation

As apparent from the first case study, AI surveillance is not neutral, and it often does not have the user's best interests in mind. It can shape user behavior rather than merely responding to it. Many people do not realize that products or websites they use make them a target for messages tailored to exploit their biases and vulnerabilities (Müller, 2023). In this case, the political party used AI to exploit X's user data for profit and political gain, unbeknownst to the user (Western Governors University, 2021). There have been similar cases where AI manipulates voters by delivering hyper-personalized political content, making people more susceptible to propaganda (Barberá, 2020). If left unchecked, practices like these can threaten democracy by manipulating the public opinion, especially when AI fosters trust. Acemoglu (2021) further warns that AI-driven data collection consolidates power within a few tech companies, reducing competition and consumer privacy. AI's ability to manipulate user behavior through targeted content curation and behavioral prediction allows corporations to profit at the expense of individual autonomy.

The second case study demonstrates surveillance capitalism, but not on behalf of a company. Instead, *individuals* were able to scrape user data and employ a chatbot or other AI tool to exploit people's vulnerabilities. Unfortunately, federal regulators, law enforcement, and the courts are not well-equipped to catch the scammers, as it is difficult to trace calls and funds

from scammers across the world. Little legal precedent exists to hold the companies that make the tools accountable as well (Verma, 2023). In most cases, the money is gone, and there is no getting it back. This is not justice.

The key ethical issue here is that AI is not merely a tool for enhancing user experience, it is actively reshaping human behavior to serve corporate and political interests, often at the cost of privacy, autonomy, and informed decision-making.

Benefits of AI

While this paper has thus far focused on the dark side of AI, below, it offers three examples that highlight some of the features and benefits that can flow from responsible AI usage.

Tailored Content

While AI can result in political propaganda, among other things, it can produce joy as well. For example, liking pictures of dogs on a social media platform indicates that the algorithm should show more dogs. In doing so, the user will feel joy and engage further with the app.

Companionship

For the elderly, loneliness is an epidemic. Virtual assistants or chatbots may provide companionship through games and conversation. They can reduce loneliness and significantly improve the lives of those living with challenging conditions (Labbé, 2024).

Enhanced Education

While AI chatbots like ChatGPT can reduce critical thinking skills, AI nevertheless offers great potential for addressing gaps in the global education system, particularly for underdeveloped communities and countries. AI technology can support teachers by automating menial tasks to allow them to focus on their students more. It can also provide more personalized learning approaches. Each student is different. While some may prefer a Socratic style classroom experience, others may benefit from a hands-on approach. Teachers can use AI to tailor each student's individual learning style while a single teacher cannot always do the same (Willige, 2024).

Ultimately, AI is just a tool, which is not intrinsically good or bad. It can be used to achieve the above and other benefits, many of which are not at all trivial. And yet, if not used responsibly, AI can be the source of risk and negative consequences. This is why it is so important to find a way to address the risks while, to the extent possible, preserving the upside potential. Some have suggested the use of regulation for this.

Need for reform

What has been done:

The Blueprint for an AI Bill of Rights in the US AI Act purports to offer a "framework [for] protections that *should* be applied with respect to all automated systems that have the potential to meaningfully impact individuals' or communities' exercise of: Rights, Opportunities,

or Access" (The White House Office of Science and Technology Policy, 2024). This is not a law or regulation. In fact, there are no governing federal laws or regulations yet (USAiAct, 2024). It is instead a guideline that does not have to be enforced, especially with the lack of legal precedent.

And yet, in October 2023, President Biden signed an Executive Order on the Safe, Secure and Trustworthy Development and Use of AI, which discusses key principles and priorities for governing AI. These principles fall under the following categories (House, 2023):

1. Safety and Security
2. Transparency and Accountability
3. Consumer Protections
4. Government Oversight
5. Civil Rights and Equity
6. Privacy and Data Protection
7. International Cooperation
8. Innovation and Competitiveness

While this order lays out stronger guiding principles for AI governance, it fully depends on how agencies implement it. All this order does is direct federal agencies like the National Institute of Standards and Technology (NIST) to develop guidelines, mandate AI model evaluations, and strengthen some privacy protections. It is not an actual law (Blaine, 2024).

In parallel, the NIST has been working to develop overarching ethical standards for AI; however, these have not been implemented (Dunietz et al., 2024). In most categories for

suggested standards, they declare that “more foundational work is needed to establish how organizations can best determine and apply [the standards]” (Dunietz et al., 2024, p. 11). With the current administration, federal support for NIST will be reduced or altogether disappear, leaving the NIST initiative far from enforceable regulation.

Various other organizations, including the IEEE, have attempted to establish regulatory guidelines for AI. These efforts have yet to produce effective safeguards, making them unworthy of detailed discussion (Chatila & Havens, 2019).

And at the state level, in 2024, at least 45 states have introduced AI bills, and 31 states adopted resolutions or enacted legislation, which is a good start. Still, data collection is largely unregulated in most states, making it possible for companies to use, sell, or share personal data without authorization or even a notice requirement (Murray, 2023).

Simply, the existing regulations do not do nearly enough, leaving an urgent need for better and more comprehensive protection (NCSL, 2024).

What can be done:

I believe we need to go further than merely developing ethical standards. As AI continues evolving, there must be regulations that address the ethical dimensions and the economic and geopolitical ones. To protect citizens around the world, all nations need to find a delicate balance between innovation and preventing AI from exacerbating concerns like misinformation, privacy violations, and algorithmic bias. Similarly to the internet, cryptocurrency, and space, AI is a borderless technology. Any approach needs to aim for multilateral cooperation, which does not negatively impact the economy and that fosters healthy competition.

While many suggest regulating only at the national level, doing so in a way that only manages US companies might have the unintended consequence of driving business from the US, negatively and disproportionately affecting the US economy. With the current state of politics in the US, it is highly unlikely that any regulatory policy that harms the economy would be enacted. Not to mention, history has shown us time and time again that the best way to enforce law is through economic policy. Sanctioning is particularly effective (National AI Advisory Committee (NAIAC), 2023). There should be a broader coalition of treaties and agreements, perhaps under the UN or G7, that discourage AI companies from moving to countries with weaker oversight. The countries that opt out would have economic disincentives like limited trade or financial restrictions (Walker, n.d.).

With that being said, there are some regulations that need to be implemented within each country individually, unless a global organization is dedicated primarily to this cause. First, it is important to protect the consumer. AI policy should empower them to sue companies that violate their privacy laws, subjecting them to civil liability. Specifically, In the US, class action lawsuits should be allowed against AI firms. Beyond privacy concerns, AI regulations must address algorithmic bias. Regulations should require bias audits and diverse and representative training datasets before AI models are deployed (Saeidnia et al., 2024).

Different governments like the US could impose local laws that require AI companies to pay fines if they do not comply with privacy laws, as an example. They could also be forced to relocate from the US or lose access from the US markets unless they follow the set regulations. The potential bans that Tik Tok is facing are an illustration of the power of the US market. We are not just any country that a company like Tik Tok could afford losing business in. The

incentive of maintaining a presence in the US marketplace may prove sufficient to force a company to change.

The government can also subsidize AI businesses that prioritize user privacy (National AI Advisory Committee (NAIAC), 2023). For example, the government could give tax breaks to companies that develop privacy focused AI models that do not store user data. Or, the government could pool together R&D funding to invest into business models that offer consumers protection against unregulated data collection.

In addition, given the importance of preventing companies from maliciously selling user's data, the government might consider enacting protectionist policies (National AI Advisory Committee (NAIAC), 2023). Since AI businesses often rely on surveillance capitalism, governments can enforce laws that prohibit data resale to third parties that use it for exploitative advertising, political manipulation, or surveillance. Any legislation created for this purpose should name specific practices, not particular companies, because companies can use rebranding loopholes to avoid compliance.

While regulations are necessary and beneficial on a nation by nation basis, we need global cooperation to enforce regulatory policies through economic incentives and disincentives. Any approach in governing AI will be most effective if multilateral. AI is borderless. Our response should be too.

Future work

Future research should focus on the implementation and viability of the suggestions advanced by this paper and present supporting ideas. Studies should identify potential obstacles to global enforcement and ways to encourage compliance among major AI-developing nations. Further research is also needed to evaluate the effectiveness of economic incentives and disincentives, such as taxation, subsidies, and trade restrictions. Lastly, future work should look into technical solutions to the AI issues presented in this paper, such as privacy-preserving AI models, bias-reduction techniques in training datasets, and ways to increase AI transparency.

Bibliography

Acemoglu, D. (2021). *Harms of AI** [Massachusetts Institute of Technology]. https://economics.mit.edu/sites/default/files/publications/Harms%20of%20AI.pdf

Amazon scrapped "sexist AI" tool. (2018, October 10). *BBC News*. https://www.bbc.com/news/technology-45809919

Artificial Intelligence 2024 Legislation. (2024, September 9). NCSL. https://www.ncsl.org/technology-and-communication/artificial-intelligence-2024-legislation

Barberá, P. (2020). Social Media, Echo Chambers, and Political Polarization. In J. A. Tucker & N. Persily (Eds.), *Social Media and Democracy* (pp. 34–55). Cambridge University Press. https://www.cambridge.org/core/books/social-media-and-democracy/social-media-echo-chambers-and-political-polarization/333A5B4DE1B67EFF7876261118CCFE19

Bond, S. (2024, December 21). How AI deepfakes polluted elections in 2024. *NPR*. https://www.npr.org/2024/12/21/nx-s1-5220301/deepfakes-memes-artificial-intelligence-elections

Boine, C. (2023). Emotional Attachment to AI Companions and European Law. *MIT Case Studies in Social and Ethical Responsibilities of Computing*, *Winter 2023*. https://doi.org/10.21428/2c646de5.db67ec7f

Blaine, T. (2024, March 6). *Guide to AI Laws and Regulations in the U.S.* https://lawsoup.org/legal-guides/ai-laws-in-the-us-artificial-intelligence-regulations/

Chatila, R., & Havens, J. C. (2019). The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. In M. I. Aldinhas Ferreira, J. Silva Sequeira, G. Singh Virk, M. O. Tokhi, & E. E. Kadar (Eds.), *Robotics and Well-Being* (pp. 11–16). Springer International Publishing. https://doi.org/10.1007/978-3-030-12524-0_2

Dunietz, J., Tabassi, E., Latonero, M., & Roberts, K. (2024). A Plan for Global Engagement on AI Standards. *NIST*. https://www.nist.gov/publications/plan-global-engagement-ai-standards

Eliot, L. (2023, November 5). Legendary ELIZA And PARRY Go Head-To-Head With ChatGPT In A Revealing Battle Of Using Generative AI For Mental Health. *Forbes*. https://www.forbes.com/sites/lanceeliot/2023/11/05/legendary-eliza-and-parry-go-head-to-head-with-chatgpt-in-a-revealing-battle-of-using-generative-ai-for-mental-health/

Fung, B., & Duffy, C. (2023, September 1). *X, formerly known as Twitter, may collect your biometric data and job history | CNN Business*. CNN. https://www.cnn.com/2023/09/01/tech/x-twitter-biometrics-employment-data-collection/index.html

Hagerty, A., & Rubinov, I. (2019). *Global AI Ethics: A Review of the Social Impacts and Ethical Implications of Artificial Intelligence* (No. arXiv:1907.07892). arXiv. http://arxiv.org/abs/1907.07892

Hao, K. (2019, February 4). *This is how AI bias really happens—And why it's so hard to fix*. MIT Technology Review. https://www.technologyreview.com/2019/02/04/137602/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix/

House, T. W. (2023, October 30). *FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence*. The White House. https://bidenwhitehouse.archives.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/

How AI Is Affecting Information Privacy and Data. (2021, September 28). Western Governors University. https://www.wgu.edu/blog/how-ai-affecting-information-privacy-data2109.html

Khalid, N., Qayyum, A., Bilal, M., Al-Fuqaha, A., & Qadir, J. (2023). Privacy-preserving artificial intelligence in healthcare: Techniques and applications. *Computers in Biology and Medicine*, *158*, 106848. https://doi.org/10.1016/j.compbiomed.2023.106848

Labbé, A. (2024, January 30). *Council Post: AI To Benefit Humanity: Innovations In Senior Care*. Forbes.

https://www.forbes.com/councils/forbestechcouncil/2024/01/30/ai-to-benefit-humanity-innovations-in-senior-care/

Law, T., Chita-Tegmark, M., Rabb, N., & Scheutz, M. (2022). Examining Attachment to Robots: Benefits, Challenges, and Alternatives. *ACM Transactions on Human-Robot Interaction*, *11*(4), 1–18. https://doi.org/10.1145/3526105

Long, A. (2023, April 23). *Psychological Acceptance of AI Chatbot Suggestions*. https://www.pgs.com/globalassets/technical-library/tech-lib-pdfs/industry_insights2023_03_ai_chatbot_psychology.pdf

Müller, V. C. (2023). Ethics of Artificial Intelligence and Robotics. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Fall 2023). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2023/entries/ethics-ai/

Murray, C. (2023, April 25). *U.S. Data Privacy Protection Laws: A Comprehensive Guide*. Forbes. https://www.forbes.com/sites/conormurray/2023/04/21/us-data-privacy-protection-laws-a-comprehensive-guide/

Pataranutaporn, P., & Turkle, S. (2024, November 8). *Essay | A 14-Year-Old Boy Killed Himself to Get Closer to a Chatbot. He Thought They Were In Love.* WSJ. https://www.wsj.com/tech/ai/a-14-year-old-boy-killed-himself-to-get-closer-to-a-chatbot-he-thought-they-were-in-love-691e9e96

Rationales, Mechanisms, and Challenges to Regulating AI: A Concise Guide and Explanation. (2023). National AI Advisory Committee (NAIAC). https://www.ai.gov/wp-content/uploads/2023/07/Rationales-Mechanisms-Challenges-Regulating-AI-NAIAC-Non-Decisional.pdf

Saeidnia, H. R., Hashemi Fotami, S. G., Lund, B., & Ghiasi, N. (2024). Ethical Considerations in Artificial Intelligence Interventions for Mental Health and Well-Being: Ensuring Responsible Implementation and Impact. *Social Sciences*, *13*(7), Article 7. https://doi.org/10.3390/socsci13070381

Sharkey, A., & Sharkey, N. (2011). Children, the Elderly, and Interactive Robots. *IEEE Robotics & Automation Magazine*, *18*(1), 32–38. IEEE Robotics & Automation Magazine. https://doi.org/10.1109/MRA.2010.940151

The White House Office of Science and Technology Policy. (2024). AI Bill of Rights: Ensuring Safe and Fair Automated Systems. USAiAct. https://www.usaiact.org/us-state-laws

USAiAct. (2024). *Blueprint for an AI Bill of Rights*. https://www.whitehouse.gov/ostp/ai-bill-of-rights/

Verma, P. (2023, March 5). They thought loved ones were calling for help. It was an AI scam. *Washington Post*. https://www.washingtonpost.com/technology/2023/03/05/ai-voice-scam/

Walker, K. (n.d.). *Our Future with AI Hinges on Global Cooperation—Sponsor Content—Google*. Retrieved March 21, 2025, from https://www.theatlantic.com/sponsored/google/global-coordination/3889/

Weiser, B. (2023, May 27). Here's What Happens When Your Lawyer Uses ChatGPT. *The New York Times*. https://www.nytimes.com/2023/05/27/nyregion/avianca-airline-lawsuit-chatgpt.html

Weizenbaum, J. (1967). *Computational Linguistics* (Vol. 10). Massachusetts Institute of Technology. https://cse.buffalo.edu/~rapaport/572/S02/weizenbaum.eliza.1967.pdf

When AI Gets It Wrong: Addressing AI Hallucinations and Bias. (2023). *MIT Sloan Teaching & Learning Technologies*. https://mitsloanedtech.mit.edu/ai/basics/addressing-ai-hallucinations-and-bias/

Williams, D. P. (2023). *Bias Optimizers*. American Scientist. https://www.americanscientist.org/article/bias-optimizers

Willige, A. (2024, May 10). *From virtual tutors to accessible textbooks: 5 ways AI is transforming education*. World Economic Forum. https://www.weforum.org/stories/2024/05/ways-ai-can-benefit-education/

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs. https://www.hbs.edu/faculty/Pages/item.aspx?num=56791