

First Prize Winner

Risks Emerging from Artificial Intelligence Widespread Use

Hanchen (Henry) Wang and Yongqi Liang

Any views and ideas expressed in the essays are the author's alone and may not reflect the views and ideas of the Society of Actuaries, the Society of Actuaries Research Institute, Society of Actuaries members, nor the author's employer.

ABSTRACT

This paper explores the risks posed by widespread artificial intelligence (AI) use on individual privacy and proposes a framework for integrating data into decision-making with a focus on fairness, accountability, and transparency. Through real-world examples and analysis, we examine how AI technologies can compromise privacy and highlight key challenges such as biases and the need for informed consent. The paper highlights the urgency for regulatory and ethical interventions to address these issues. Our framework is designed to support organizations in making responsible ethics-based data decisions with a priority on maintaining societal well-being. By proposing frameworks to preserve privacy while fostering AI innovation and promoting ethical data practices, this research contributes to the discourse on responsible AI governance and fosters data use safety within diverse organizations.

INTRODUCTION

The rapid development and widespread adoption of AI have transformed and promised significant benefits to various sectors from healthcare to finance. However, this transformation brings with it a wide range of risks that need careful consideration. This essay will explore the potential short-term and long-term risks associated with AI's widespread use. We will focus on privacy erosion, economic disruption, and biased decision-making, along with the broader existential threats that AI might pose.

SHORT-TERM RISKS

PRIVACY EROSION

AI systems rely heavily on vast amounts of data to function effectively. This often includes personal and sensitive information. The potential for data breaches and misuse of personal data is high, especially if AI systems are not adequately regulated. Privacy breaches can lead to significant financial losses, reputational damage, and a loss of trust among users.

Examples:

- *Camera Surveillance in China:* The extensive use of AI-powered surveillance cameras in China for monitoring citizens has raised significant privacy concerns. These cameras, often equipped with facial recognition technology, are deployed in public spaces to track individuals' movements and behaviors. This level of surveillance can lead to invasions of personal freedoms and privacy. This is due to the fact individuals may feel constantly watched and monitored by the state. The potential misuse of this data for political or social control further exacerbates the ethical issues surrounding AI surveillance (Yang, 2022). The pandemic has further accelerated the use of such technologies, justifying them under the guise of public health and safety. This cover-up has led to widespread acceptance among the population (Yang, 2022).
- *Human Tracking Devices in Self-Driving Cars:* The use of AI to track and collect data on individuals' movements and behaviors through human tracking devices in self-driving cars is another area of concern. While these devices can enhance navigation and safety features, they also collect vast amounts of personal data. U.S. corporations might even collect massive amounts of data on individuals to sell to the Chinese government for profit. The Chinese government even installed an unauthorized surveillance camera inside the Tesla Auto-Car made in China to track down individuals privately and monitor their daily activities (Schellekens, 2022).
- *Amazon Just Walk Out Technology:* Amazon's Just Walk Out technology allows customers to shop without going through the traditional checkout processes. They do this by using AI to track items picked up and placed down to automatically charge their accounts. While convenient, this technology raises privacy concerns as it involves constant monitoring of customers' shopping behaviors. The data collected can provide detailed insights into individuals' purchasing habits. This data could be used for targeted advertising and manipulation of consumer choices. This seamless shopping experience, while revolutionary, presents significant privacy challenges as every item selected by a customer is tracked and recorded. This detailed consumer profile could be exploited for commercial purposes without the consent of the consumers. In addition, Amazon would need its customers to accept terms and conditions to access its service. Once customers click the "accept" button, they grant Amazon extensive rights to their data, which allows Amazon to legally collect and invade customers' shopping preferences. Customers lose their privileges to demand justice once Amazon causes conflicts of interest with customers. Terms and Conditions clear Amazon when customers want to settle a lawsuit against them for violating their "privileged rights" (Ives et al., 2019).

ECONOMIC DISRUPTION

AI-driven automation has the potential to disrupt labor markets significantly. While AI can enhance productivity and efficiency, it also poses a risk of job displacement. This could lead to increased economic inequality and social instability. The transition to an AI-driven economy requires careful management to ensure that the benefits of AI are distributed equitably and that displaced workers are supported through retraining and social safety nets.

Examples:

- *Social Media Tracking:* Platforms like Instagram and TikTok use AI to track user behavior can potentially infringe on privacy and contribute to data misuse. AI algorithms analyze user interactions to personalize content, advertisements, and recommendations. While this enhances user experience, it also means that vast amounts of personal data are being collected and analyzed without users' explicit consent. The U.S. banned TikTok and integrated TikTok features in Instagram because any corporations outside of China can turn over all data directly to the Chinese

Communist Party. The main issue is access to individuals' entire phones: search history, microphone, geolocation, metadata, and camera. The Chinese government is more powerful than any multinational corporation in the mainland of China. This means they can use their political power to exploit all user data from other countries. This can lead to privacy breaches and exploitation of user data for the Chinese Communist Party's commercial gains. On the other hand, U.S. corporations are more powerful than the U.S. government, which indicates that social media platform users' data can be taken by U.S. corporations and sold to the U.S. government as well. The U.S. government could use data to track down individuals to access their records so they can find a way to change their values and beliefs. Many U.S. social media corporations will simply vote for those who push policies that benefit them the most (Perakakis et al., 2019). Ultimately, U.S. corporations other than social media companies use data to target audiences with AI. Companies leverage AI to analyze consumer data and tailor marketing strategies. However, oftentimes these companies do so without clear transparency or consent from users. This practice raises ethical concerns about data ownership, consent, and the potential for manipulative marketing tactics that exploit personal information (Svetlana et al., 2022).

BIASED DECISION-MAKING

AI algorithms are only as good as the data they are trained on. If the training data is biased, the AI system will likely perpetuate these biases, leading to discriminatory outcomes. This is particularly concerning in critical areas such as hiring, law enforcement, and access to financial services. Addressing algorithmic bias is essential to prevent AI from exacerbating existing societal inequalities.

Examples:

- *Education*: The use of AI in education, particularly tools like Zoom during the pandemic, highlights the potential dangers and changes in learning dynamics. While AI-powered platforms facilitated remote learning, they also introduced challenges such as unequal access to technology and data privacy concerns. Additionally, AI algorithms used for grading and student assessments can perpetuate biases and deprive certain groups of students based on their socioeconomic background or learning style. For example, Slimi and Carballido highlight that students with different skin colors struggle more academically due to their socioeconomic status, and women of color face significant barriers in STEM fields due to social isolation and biases (Slimi & Villarejo Carballido, 2023).
- *Online Proctoring Software*: Yoder-Himes conducted research on the bias in facial recognition technology used in online proctoring software to detect cheating. The study revealed that students with darker skin tones were more likely to be flagged for cheating, leading to increased scrutiny from instructors. This bias against students based on race and gender highlights the ethical concerns surrounding the use of AI in educational settings and the need for further research to understand and mitigate algorithmic biases (Yoder-Himes et al., 2022).
- *Use of ChatGPT*: The use of AI like ChatGPT may influence tendencies among university students. Muhammad, Ahmed, and Tariq conducted research on the causes and effects of general use of ChatGPT among university students. They compared and contrasted those who frequently use ChatGPT and those who do not to see if ChatGPT influences academic tendencies among university students. Their findings also suggest that students whose academic workload and time pressure is higher report higher frequent use of ChatGPT to cope with their stressful academic circumstances. Their findings also suggest that those who are sensitive to academic rewards report lower use of ChatGPT than those who aren't. Furthermore, excessive use of ChatGPT can cause mental damage to students and their academic outcomes. Those who always use ChatGPT

tend to procrastinate their work and report memory loss compared to those who barely use ChatGPT. This is because students who frequently rely on ChatGPT lose their critical thinking and problem-solving skills, which can harm their academic performance. It can be concluded that ChatGPT usage correlates with time pressure, academic workload, and sensitivity to rewards with students' academic outcomes (Abbas et al., 2024).

LONG-TERM RISKS

EXISTENTIAL THREATS

In the long term, AI could pose existential risks if it surpasses human control. The development of superintelligent AI systems could lead to scenarios where AI operates beyond human oversight, making decisions that could have catastrophic consequences. Ensuring that AI systems are aligned with human values and goals is crucial to mitigating these risks.

Examples:

- *Tesla Driverless Cars*: Recent news about driverless cars and potential data breaches or hacking could lead to uncontrolled scenarios. If AI systems in autonomous vehicles are compromised, it could result in widespread chaos and accidents. The potential for hackers to take control of driverless cars poses significant risks to public safety and highlights the need for robust cybersecurity measures in AI systems. The safety argument emphasizes the benefits of AI in driving accuracy but also highlights the need for extensive testing and reliable data to ensure safety (Blanco et al., 2016).
- *Statistical Data: "The Safety Argument"* describes the decisional phenomenon of autonomous vehicles (AV) by comparing and contrasting AI decisional capacity with real human drivers. Statistical driving data highlights the correlation between erroneous human driving decisions and road accident fatalities. The safety argument of AV technologies stresses the core safety benefits that can promote more accurate driving abilities and advocate advanced AI decision capacity to accurately navigate the road network. However, research studies such as from Blanco illustrate that by comparing and contrasting AV driving data to human driving data "The Safety Argument" cannot accurately determine the outcome of the AI decisions (Blanco et al., 2016). This is because "The Safety Argument" cannot be predicted and justified by data analysis in the AV system. The Virginia Tech Transportation Institute criticized the practice of using inaccurate data for safety analysis, claiming that "with a Poisson distribution and national mileage and crash estimates, automated vehicles would need to drive 725,000 miles on representative roadways without incident and without human assistance to say with 99% confidence they crash less frequently than vehicles with human drivers" (Goodall, 2014a). It seems that inaccurate statistical methods resulting in invalid data outcomes is why an AV occasionally goes wrong and presents different formats and critiques of AV crash report data. Furthermore, sensor error, programming bugs, unanticipated objects, classification errors, and hardware/software faults present further unsolved challenges to the AV safety argument in the future due to the statistical issue of data deficiencies. Blanco claims that future AI technologies cannot guarantee the safety of AV due to higher risks of crashes, stating, "The limited exposure of the self-driving car project to real-world driving increases statistical uncertainty in its crash rate. That uncertainty will decrease as it receives more on-road, in-traffic testing" (Blanco et al., 2016).

AI ALIGNMENT

One of the most significant challenges in AI development is ensuring that AI systems align with human values. Misaligned AI systems could make decisions that are harmful or unintended. Research in AI alignment aims to create systems that understand and prioritize human values, reducing the risk of catastrophic outcomes.

Examples:

- *Healthcare Technology:* AI's impact on healthcare and potential risks associated with data handling and privacy. AI applications in healthcare, such as diagnostic tools and personalized treatment plans, rely on sensitive patient data. Ensuring data privacy and security is paramount to prevent misuse and maintain patient trust. Additionally, AI systems must be designed to align with ethical principles in healthcare, prioritizing patient welfare and informed consent (Rahman et al., 2024).
- *Financial Systems:* AI's role in managing money on the internet, including banks, cryptocurrencies, and NFTs, with potential vulnerabilities to data breaches. The financial sector's increasing reliance on AI for transactions, fraud detection, and investment strategies necessitates stringent security measures to protect against cyber threats. The potential for AI-driven financial systems to be exploited for criminal activities underscores the need for regulatory oversight and robust cybersecurity frameworks (Hidayat et al., 2024).

GEOPOLITICAL RISKS

The widespread use of AI also has geopolitical implications. Nations that lead in AI development could gain significant strategic advantages, potentially leading to imbalances in global power dynamics. AI could be weaponized for cyberattacks, misinformation campaigns, and surveillance, all of which pose risks to national security and global stability. "Sub-conscious and personalized levels of algorithmic persuasion may have significant effects on the cognitive autonomy of individuals and their right to form opinions and take independent decisions" (Ashraf, 2021).

Examples:

- *Algorithmic censorship:* The rise of AI technology can threaten freedom of speech online. AI technologies can extract data to deliver customized content and influence users' thoughts and perceptions. This "algorithmic persuasion" can affect individuals' cognitive autonomy and their right to form independent opinions. The integration of AI in business models of "surveillance capitalism" can punish dissenters for not fitting into their own agenda. Zeynep Tufekci's concept of "algorithmic censorship" highlights how AI determines what can be seen online, shaping individuals' online environments, and potentially leading to biased content moderation (Ashraf, 2021).
- *Facebook-Cambridge Analytica Scandal Case Study:* The corporation Cambridge Analytica in 2018 harvested millions of Facebook users' private information and data without their consent for political campaign advertisements. The scandal underscores the ethical dilemmas surrounding data privacy and causes many controversies related to the ethical responsibilities of technology companies in safeguarding individuals' privacy rights. Algorithmic bias in criminal justice illustrates that all algorithmic tools utilized in criminal justice systems have been scrutinized for perpetuating discrimination against marginalized communities. Studies have concluded that algorithms in the criminal justice system could exhibit racial discrimination based on individuals' socioeconomic status in society. This could lead to inequalities in the legal system and disparities in sentencing outcomes. Facebook CEO Mark Zuckerberg acknowledged his failure to protect individuals' data and implemented measures to upgrade data protection and cybersecurity systems (Trout, 2016).

Researchers, regulators, and policymakers have demanded greater accountability and fairness in algorithmic decision-making in response to the scandal and concerns about algorithmic bias in the criminal justice system. Additionally, some jurisdictions have implemented bans and restrictions on the use of biased algorithms in the criminal justice system to enhance legal decision-making processes and reduce the risk of perpetuating discrimination. The Facebook-Cambridge Analytica Scandal can teach us that AI can play a vital role in shaping the future of the criminal justice system (Trout, 2016).

- *AI Content Moderation:* Online platforms have turned to AI for better content moderation. AI content moderation is responsible for reporting any negative content by removing offensive wordings, filtering inappropriate comments, and permanently deleting spam. This shows that content moderation generated by AI accumulates massive amounts of user data and applies data science tricks/techniques to identify certain patterns and correlations to hypothesize trends and outcomes to govern speech. Thus, AI can influence the way speech structures are structured by modifying content before being officially published. The use of AI content moderation illustrates biased stereotypes, racial discrimination, and hate speech online. This is because AI content moderation tools automatically target certain vulnerable minority groups and mark them as dangerous threats. AI hate speech detection tools demonstrate biases against African Americans. These systems are more likely to ban their tweets online due to their potential likelihood of physical violence demonstrations and verbal assaults online. AI technologies, like a double-edged sword, can create many opportunities and challenges for us. It is our responsibility to control it and fulfill its potential to benefit us in many aspects of our lives (Ashraf, 2021).

CONCLUSION

The rapid advancement of AI presents both opportunities and risks. While AI holds immense potential to improve various aspects of human life, it is crucial to address the associated risks proactively. Privacy erosion, economic disruption, biased decision-making, and existential threats are significant concerns that need careful consideration and regulation. By understanding and mitigating these risks, we can harness the benefits of AI while minimizing its potential harms. The Society of Actuaries Research Institute's call for this topic is a valuable initiative to stimulate discussion and promote further research in this critical area.

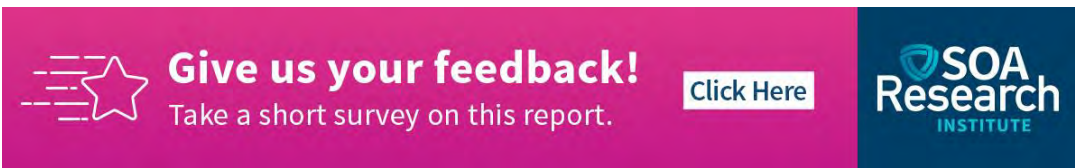
* * * * *



Hanchen (Henry) Wang graduated from the University of California, San Diego with degrees in Business Psychology and International Studies with focus on Economics. He enjoys collecting books, stamps, and different currencies. He enjoys working with data and inspires to become a licensed actuary. He can be reached at whc1996920@gmail.com.



Yongqi Liang is an undergraduate student at the University of California, San Diego majoring in Math-Econ with a minor in Finance. He is passionate about actuarial science and is working towards becoming a licensed actuary. He can be reached at yongqiliang2004@gmail.com.



Give us your feedback!
Take a short survey on this report.

[Click Here](#)

SOA
Research
INSTITUTE

REFERENCES

- Abbas, M., Jam, F. A., & Khan, T. I. (2024). Is it harmful or helpful? examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education*, 21(1). <https://doi.org/10.1186/s41239-024-00444-7>
- Ashraf, C. (2021). Exploring the impacts of artificial intelligence on freedom of religion or belief online. *The International Journal of Human Rights*, 26(5), 757–791. <https://doi.org/10.1080/13642987.2021.1968376>
- Blanco, M., Atwood, J., Russell, S. M., Trimble, T. E., McClafferty, J. A., & Perez, M. A. (2016, January 8). Automated Vehicle Crash Rate Comparison Using Naturalistic Data. *VTechWorks Repository*. <https://vtechworks.lib.vt.edu/items/d33dc4e1-58f3-4813-9677-4a0f425880c7>
- Goodall, N. J. (2014a). Ethical Decision Making during Automated Vehicle Crashes. *Transportation Research Record: Journal of the Transportation Research Board*, 2424(1), 58–65. <https://doi.org/10.3141/2424-07>
- Goodall, N. J. (2014b). Machine Ethics and Automated Vehicles. *Road Vehicle Automation*, 93–102. https://doi.org/10.1007/978-3-319-05990-7_9
- Hidayat, M., Defitri, S. Y., & Hilman, H. (2024). The Impact of Artificial Intelligence (AI) on Financial Management. *Management Studies and Business Journal (PRODUCTIVITY)*, 1(1), 123–129. <https://doi.org/10.62207/s298rx18>
- Ives, B., Cossick, K., & Adams, D. (2019). Amazon Go: Disrupting retail? *Journal of Information Technology Teaching Cases*, 9(1), 2–12. <https://doi.org/10.1177/2043886918819092>
- Perakakis, E., Mastorakis, G., & Kopanakis, I. (2019). Social Media Monitoring: An Innovative Intelligent Approach. *Designs*, 3(2), 24. <https://doi.org/10.3390/designs3020024>
- Rahman, Md. A., Victoros, E., Ernest, J., Davis, R., Shanjana, Y., & Islam, Md. R. (2024). Impact of Artificial Intelligence (AI) Technology in Healthcare Sector: A Critical Evaluation of Both Sides of the Coin. *Clinical Pathology*, 17. <https://doi.org/10.1177/2632010x241226887>
- Schellekens, M. (2022). Human–machine interaction in self-driving vehicles: a perspective on product liability. *International Journal of Law and Information Technology*, 30(2), 233–248. <https://doi.org/10.1093/ijlit/eaac010>
- Slimi, Z., & Villarejo Carballido, B. (2023). Navigating the Ethical Challenges of Artificial Intelligence in Higher Education: An Analysis of Seven Global AI Ethics Policies. *TEM Journal*, 590–602. <https://doi.org/10.18421/tem122-02>

Svetlana, N., Anna, N., Svetlana, M., Tatiana, G., & Olga, M. (2022). Artificial intelligence as a driver of business process transformation. *Procedia Computer Science*, 213, 276–284.
<https://doi.org/10.1016/j.procs.2022.11.067>

Trout, K. E. (2016). The Impact of Electronic Health Records on Healthcare Service Delivery, Patient Safety, and Quality. *DigitalCommons@UNMC*. <https://digitalcommons.unmc.edu/etd/173/>

Yang, Z. (2022, October 11). The Chinese surveillance state proves that the idea of privacy is more “malleable” than you’d expect. *MIT Technology Review*.
<https://www.technologyreview.com/2022/10/10/1060982/china-pandemic-cameras-surveillance-state-book/>

Yoder-Himes, D. R., Asif, A., Kinney, K., Brandt, T. J., Cecil, R. E., Himes, P. R., Cashon, C., Hopp, R. M., & Ross, E. (2022). Racial, skin tone, and sex disparities in automated proctoring software. *Frontiers in Education*, 7. <https://doi.org/10.3389/educ.2022.881449>