Institute for Youth in Policy

Ethics X AI: Exploring the Dangers

Policy Lead: Robert Nguyen[#1]

[#]*Technology Policy, Institute for Youth in Policy*
1700 Van Ness Ave #1143, San Francisco, CA 94109
[1]robert@fellow.yipinstitute.org

***Abstract*— The emerging Artificial Intelligence era contributes to a constant threat. The access to the powers of AI, from the private sector and government to anyone, has been continuously growing. In this highly controversial period of technological development, Autonomous and AI-enabled systems have shown to be used in malicious ways. This brief will dive into the ethicacy around the recent AI dangers, covering the important social issues brought by the development, providing ways to mitigate the challenges.**

Keywords*— *Artificial Intelligence and Machine Learning, Sextortion, Quantum Information Systems (QIS), Deep Fakes, OSINT, Neural Networks, National Security, Counter Influence, Cyber Warfare, Nation-States, Cybersecurity, Big Data Analytics, Adversaries, Russia, People's Republic of China, Geopolitics, Eurasia, GEOINT, Digital Privacy, Research and Development (R&D), Governance Risk & Compliance (GRC), Policy, Model

I. Executive summary

Ensuring the development of ethical AI practices is an important part in our strive for a better world. Despite the importance of comprehensive AI education, the majority of academic institutions do not cover the ethical considerations of AI. We are failing to train the future designers and operators of our intelligent systems. Curriculums offer advanced coursework in natural language processing, deep learning, and large-scale data sets, yet still do not provide the essential and practical knowledge to help our engineers safely but morally build our future networks. This leaves the next generation of engineers, unprepared to tackle the complex social issues that AI brings along with it.

II. Overview

Ethical AI practice is defined by UNESCO as "The protection of human rights and dignity… based on the advancement of fundamental principles such as transparency and fairness, always remembering the importance of human oversight of AI systems." As AI technology evolves, we must understand the dangers brought about by the growth. AI has been used in many applications. A modern example of the dangers is Deepfakes. Fabricated and manipulated content can be created through the use of AI. The AI models that generate the images, videos, recordings, etc are often unregulated. Deepfakes have been used in a variety of ways. Used to mislead constituents against their representatives, by generating false statements. Used in a variety of cyber crimes such as scamming concerned elderly parents into monetary sums, through a loved one's voice being cloned. Or used for cyberbullying to target insecure students with the horrors of fake 'AI generated' explicit images. These AI generated images have been used by international criminals– hundreds of miles away– to ruin the lives of so many young children and families. Financial criminals, individuals motivated to commit crimes to obtain money or property for gain, have used various forms of AI to aid them in financial sextortion scams, a type of cyber crime where a predator demands money from the victim to keep their sexual images private. The rise in AI is not limited to generated media and content, but nation-states have been using the technology in space weapons to counteract foreign ones. Governments and agencies have also been using AI to help flag satelite threats in high orbit. Along with AI, the revolution of quantum computing power has sparked outrage. More specifically QIS, or quantum information science, for the immense potential of greatness in our economy and national security. QIS is a new way to study data– the processing, analysis, and transmission of information. QIS is a new field of science and technology and harnesses the combinations across many disciplines of physical science, computer science, mathematics, and much more.

© 2024 Robert C. Nguyen, The Institute for Youth in Policy

A. Pointed Summary

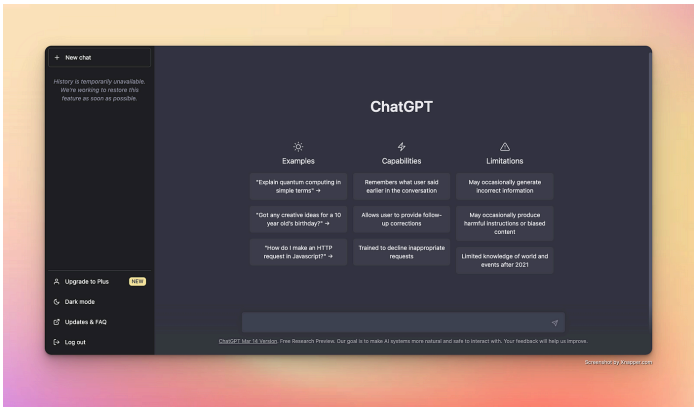
- Ethical AI Engagement and Education lacks exposure in the social sciences.
- This lack of education and training develops incompetent engineers, without the understanding of AI's social implications, who are building our unethical systems.
- Ethics education, within our education and engineering programs, will equip future engineers of our AI systems with the skills needed to build and design, implement, and manage the diverse needs of the ever changing global world.
- The creation, promotion, and funding of this 'ethics' education in educational institutions will decrease polarization, increase public understanding, and generally improve social welfare among American communities.

B. Relevance

As our society develops new ways to harness the capabilities of AI– especially in different capacities– it is crucial we promote the developers to look at the important social issues. Ethics, such as privacy and fairness, must remain the primary focus when scaling these new technologies. The systems implemented must cover how people will behave, recognize the patterns of communication, influence, connectivity, and behavior. Mainly, how will society be impacted and react to the developments.

A. Current Stances

AI generally is really new. The concept of AI has been around since the late 1950s, but has begun to flourish recently due to the 'more' accessible nature to the mainstream amount of internet users– allowing most to get their hands on AI. Machine learning algorithms are everywhere now. Throughout the day, we are interacting with various forms of artificial intelligence models. It could be Netflix's complex neural networks used in their content recommendation system or the deep downscaler to improve video quality. Designers may interact with Adobe's built in AI. Usually this is in the form of interacting with a large language model (LLM) through models like Meta's AI– Llama, Google's AI- Gemini, maybe Snapchat's AI assistant, or the famous OpenAI's ChatGPT, a chatbot which launched on November 30, 2022 with an easy user interface where you can type a question (input) and receive an output.



GPT-4's Interface | ***source***

B. Tried Policy

Several cases of specific tried policies have been enacted for AI ethics across various countries and organizations. The European Union's AI Ethics Guidelines, published in 2019 by the European Commission's High-Level Expert Group on AI, proposes a framework for the development and employment of AI. This was based off of seven key requirements:

- Human agency and oversight
- Technical robustness and safety
- Privacy and data governance
- Transparency
- Diversity, non-discrimination, and fairness
- Societal/Environmental well-being
- Accountability

The EU's framework, set the guidelines to how they approach integrating AI in all of their projects.

In April of 2018, United Kingdom's Office for AI and AI Council, established to oversee AI development ensuring it aligns with ethical standards and promotes public trust, released the "AI Sector Deal" as part of their Industrial Strategy.

© 2024 Robert Nguyen, The Institute for Youth in Policy

With the purpose of positioning the UK as a global leader in artificial intelligence, the AI Sector Deal aimed to leverage AI to improve productivity and address significant societal challenges. The key objectives were to invest in AI research and development, develop a strong pipeline of AI talent by enhancing education and training programs, improve access to data–while ensuring robust data governance and privacy protections–including initiatives making public data more accessible and usable for AI research and promoting a strong digital infrastructure. In addition, the deal allowed for better collaboration fostered between academia, industry, and government– accelerating the commercialization of AI technologies. While also providing regulation for the ethical and responsible use of AI. Lastly, it provided more discussion around the AI implications of inclusivity and ethicacy, through guidelines and standards promising AI systems remain fair, transparent, and accountable.

Singapore, in 2019, introduced the "Model AI Governance Framework", providing details and practical guidance on implementing AI governance practices. The implementation started as a pilot program involving business across various sectors to gather insights and refine the guidelines.

New York City enacted the "Automated Decision Systems (ADS) Task Force" policy in 2017 to examine and make recommendations on the use of ADS by city agencies. The purpose of the policy was to address the growing usage of automated decision systems by NYC agencies and to ensure these systems remain fair, safe, and equitable for all. In addition, the ADS policy aimed to develop and promote best practices for the use of ADS by city agencies. Through public engagement and input, the perspectives of those most affected by ADS were considered, specifically, in the policy-making process along with everyone. To involve public discussion, the task force sought to gather input from a wide range of stakeholders, including community groups, advocacy organizations, and experts. The rationale behind implementing this task force was the increased use of technology, concerns over bias and discrimination, need for regulatory oversight, and public trust/confidence. NYC, like many other municipalities, has increasingly relied on technology and automated systems to improve efficiency and decision-making in various areas–including but not limited to law enforcement, social services, and public health. Within this artificial intelligence and machine learning boom, there were growing concerns about the potential for ADS to perpetuate biases present in data they are trained on. This would lead to discriminatory outcomes, particularly affecting marginalized communities. Protecting NYC's diverse populations remained key, so the rapid adoption of automated decision systems also highlighted the urgent need for oversight. It is essential that ADS maintain public trust and confidence in government operations. The establishment of the ADS Task Force was a proactive step by New York City to ensure that the adoption of new technologies aligns with the city's values of equity, transparency, and accountability, ultimately aiming to protect and serve all the residents fairly and effectively.

These are some of the notable ways in which we, as a global community, have tried to safeguard the interests of AI while protecting the impacted society. The policies enacted all have the same objective: the goal to use AI to benefit the people.

III. Policy Problem

A. Stakeholders

Everyone is involved in improving social and ethics education within our education systems and programs from primary, secondary, and post high school. The accessibility is wide ranging allowing anyone to get their voice heard, so all levels of ages and education can and should get involved in the discourse. Whether that is lobbying to your local representative to vote in favor of your, and their constituents, opinions or sharing your ideas at your child's school district board meeting. It is a jointly collaborative effort to promote stronger and more robust STEM education. Creating a society, where everyone is well informed on how AI works and the

© 2024 Robert C. Nguyen, The Institute for Youth in Policy

intended benefits and unintended consequences brought by AI. Some of the most important stakeholders in improving ethics education around AI and engineering are The Federal Department of Education and other government agencies, companies, and everyone involved with undergraduate and graduation curriculum: Faculty members, students, advisory boards, community partners, employers, etc. Citizens must continue to pressure the education system, adapting more comprehensive programs. That be educational leadership at the local level, a school district's director of curriculum and instruction, or at the state level, the state's secretary of education. This is how a strong long-term education policy is initiated. Adaptations to historic systematically designed systems are difficult to challenge, but combined effort creates momentum driving policy changes.

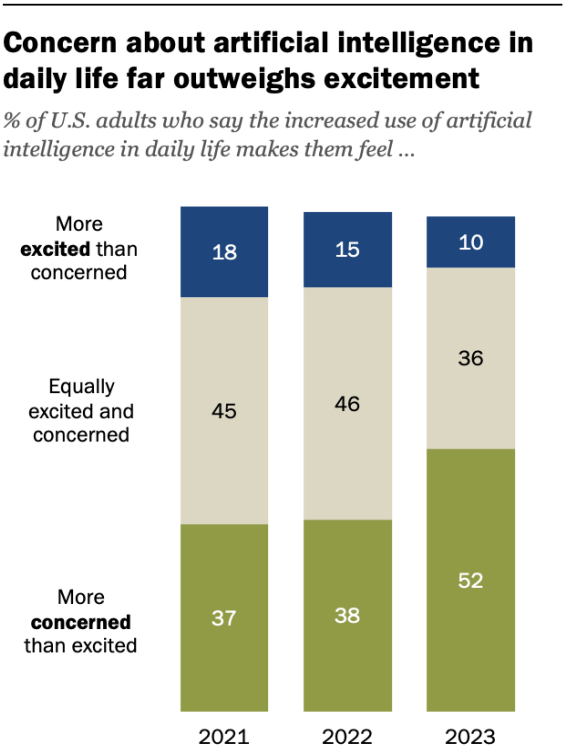
B. Risks of Indifference

According to the Pew Research surveys, the Partnership for Public Service and Freedman Consulting found that 56% of people do not or somewhat do not think that the federal government can be trusted to do what is right. In addition, 53% of people feel that the government has a negative impact on the U.S. Lastly, 67% feel that the government is not transparent, and 63% feel that the government does not listen to the public. These statistics show the scary distrust the American people have in their government. This reinforces the idea that we have to use our voices in the government to be heard. Citizens have to get more involved to shape AI ethics policy. People of all ages– young or old – should lobby for change.

Again, our future generations need to be equipped to design the fairest AI systems.

C. Nonpartisan Reasoning

Alec Tyson and Emma Kikuchi share results from researchers in a survey diving deep into the growing public concern about the role of AI in daily life. According to a new study conducted by the nonpartisan *Pew Research Center*, 52% of Americans say they feel more concerned than excited about the increased usage of AI. Only about 10% of Americans, from the diverse group sampled, say they feel more excited than concerned. The article starts, "A growing share of Americans express concern about the role artificial intelligence (AI) is playing in daily life" (Tyson and Kikuchi).
This is why ethics must be at the forefront of AI innovation. Keeping ethics in the conversations fosters trust. Whenever AI is integrated into society, everywhere around us, it is essential that the human-interaction component of AI is considered. Understanding how the AI system will impact the people and how the surrounding people will impact the AI system, is crucial in developing and placing these systems.



Since December 2022, when 38% of Americans had the opinion that AI in daily life outweighs excitement, it has increased by 14 percentage points. | *chart source*

IV. Policy Options

AI is used in so many ways from AI in military applications, like autonomous weapon systems, to AI in evidence collection and analysis, using big data analytics to monitor social media propaganda. While there is no universal way to combat all the malicious AI applications, below are several ways for ethical AI implementation.

Through transparency mandates, requiring organizations to disclose the methodologies behind their AI algorithms it will enhance trust and accountability. In addition, allowing external audits to ensure fairness and ethical standard. This can be done without compromising on the protection of proprietary information.

Establishing AI Ethics Certification Programs that focus on developing safe AI systems must be implemented, similar to those in energy and energy or information technology sectors (e.g., LEED certification (Leadership in Energy and Environmental Design), CISSP (Certified Information Systems Security Professional), etc. Industry wide program standards will provide a clear, standardized benchmark for ethical AI practices. In addition, it will encourage companies to adopt best practices to ensure safe development. Lastly, another way we can promote more groundbreaking ethical AI developments, is to allocate funding, government and private, specifically for research on the ethical implications of AI. This will advance our understanding of ethical challenges and solutions. It will also support the development of ethical AI technologies while encouraging interdisciplinary collaboration between social experts, engineers, and policymakers.

V. Conclusions and Recommendations

In the world's current rate of technology development, specifically in artificial intelligence, we can not continue progressing without establishing a strong AI curriculum that contains the importance of human factors. That dives into the complexity of morality when effectively scaling models. Everyone must get involved– in any way possible – the contribution may be big or small but it will ensure that the ethical implications of AI are discussed and focused on.

Acknowledgment

The Institute for Youth in Policy wishes to acknowledge Paul Kramer, Carlos Bindert, Gwen Singer, and other contributors for developing and maintaining the Programming Department within the Institute.

References

[1] "Ethics of Artificial Intelligence." *UNESCO.Org*, www.unesco.org/en/artificial-intelligence/recommendation-ethics. Accessed 19 June 2024.

[2] Johnson, Dave. "What Are Deepfakes? How Fake AI-Powered Audio and Video Warps Our Perception of Reality." *Business Insider*, Business Insider. www.businessinsider.com/guides/tech/what-is-deepfake. Accessed 19 June 2024.

[3] Bond, Shannon. "Ai Fakes Raise Election Risks as Lawmakers and Tech Companies Scramble to Catch Up." *NPR*, NPR, 8 Feb. 2024, www.npr.org/2024/02/08/1229641751/ai-deepfakes-election-risks-lawmakers-tech-companies-artificial-intelligence#:~:text=Artificial%20intelligence%20deepfakes%20are%20a%20threat%20to%20elections%20%3A%20NPR&text=Artificial%20intelligence%20deepfakes%20are%20a%20threat%20to%20elections%20As%20AI,are%20trying%20to%20catch%20up.

[4] "Voice Cloning Scams Are a Growing Threat. Here's How You Can Protect Yourself." *CBS News*, CBS Interactive, www.cbsnews.com/newyork/news/ai-voice-clone-scam/. Accessed 19 June 2024.

[5] "What Is Sextortion?" *FBI*, FBI, 7 July 2015, www.fbi.gov/video-repository/what-is-sextortion/view/.

[6] *Financially Motivated Sextortion - FBI*, www.fbi.gov/how-we-can-help-you/scams-and-safety/common-scams-and-crimes/sextortion/financially-motivated-sextortion. Accessed 22 June 2024.

[7] "DARPA Project Uses AI to Flag Space Weapons, Spy Satellites." *Missile Defense Advocacy Alliance*, missiledefenseadvocacy.org/space-news/darpa-project-uses-ai-to-flag-space-weapons-spy-satellites/. Accessed 19 June 2024.

[8] "What Is Quantum Computing?" *IBM*, 8 Sept. 2021, www.ibm.com/topics/quantum-computing.

[9] *Quantum Information Science*, www.nsf.gov/pubs/2000/nsf00101/nsf00101.htm. Accessed 22 June 2024.

[10] Jung, Eliot (2024). The Impact of Quantum Information Science and Technology on National Security. Purdue University Graduate School. Thesis. https://doi.org/10.25394/PGS.25669449.v1

© 2024 Robert C. Nguyen, The Institute for Youth in Policy

[11] "Politico: Senators Unveil Bill to Mandate AI Transparency within the Government." *Senator Mike Braun*, www.braun.senate.gov/news/in-the-news/politico-senators-unveil-bill-mandate-ai-transparency-within-government/. Accessed 21 June 2024.

[12] "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence." *The White House*, The United States Government, 30 Oct. 2023, www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/.

[13] "IEEE Certifaied – the Mark of Ai Ethics." *IEEE SA - The IEEE Standards Association - Home*, engagestandards.ieee.org/ieeecertifaied.html. Accessed 21 June 2024.

[14] *International Association of Privacy Professionals*, iapp.org/news/a/should-we-create-a-certification-for-ai-ethics. Accessed 21 June 2024.

[15] SITNFlash. "The History of Artificial Intelligence." *Science in the News*, 23 Apr. 2020, sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/.

[16] "Machine Learning Algorithms." *GeeksforGeeks*, 19 June 2024, www.geeksforgeeks.org/machine-learning-algorithms/.

[17] *What Are Large Language Models? - LLM AI Explained - AWS*, aws.amazon.com/what-is/large-language-model/. Accessed 22 June 2024.

[18] "Ethics Guidelines for Trustworthy Ai." *Shaping Europe's Digital Future*, digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai. Accessed 21 June 2024.

[19] LEMONNE, Eric. "Ethics Guidelines for Trustworthy Ai." *FUTURIUM - European Commission*, 5 Mar. 2021, ec.europa.eu/futurium/en/ai-alliance-consultation.1.html.

[20] AI Regulation in the Art and Media Industry: Safeguarding Creativity, Employment, and Ethical Integrity YIP, Oct 21 https://docs.google.com/document/d/1SDL03tfHDIbuZZhXaKJzj1mVdsqqvYroCeApTV9-M-g/edit?usp=sharing

[21] *National AI Strategy*, assets.publishing.service.gov.uk/media/614db4d1e90e077a2cbdf3c4/National_AI_Strategy_-_PDF_version.pdf. Accessed 22 June 2024.

[22] "Ai Sector Deal." *GOV.UK*, www.gov.uk/government/publications/artificial-intelligence-sector-deal/ai-sector-deal. Accessed 21 June 2024.

[23] "PDPC: Singapore's Approach to AI Governance." *Personal Data Protection Commission*, www.pdpc.gov.sg/help-and-resources/2020/01/model-ai-governance-framework. Accessed 21 June 2024.

[24] *New York City Automated Decision Systems Task Force* , www.nyc.gov/assets/adstaskforce/downloads/pdf/ADS-Report-11192019.pdf. Accessed 22 June 2024.

[25] *A Shadow Report of the New York City Automated ...*, ainowinstitute.org/wp-content/uploads/2023/04/ads-shadowreport-2019.pdf. Accessed 22 June 2024.

[26] *The Curriculum Approval Process*, ccleague.org/sites/default/files/trustees-resources/curriculum_approval_process_publication.pdf. Accessed 22 June 2024.

[27] Partnership for Public Service. "Trust in Government." *Partnership for Public Service*, 10 June 2024, ourpublicservice.org/publications/trust-in-government/.