ESTABLISHING A FUTURE-PROOF FRAMEWORK FOR AI REGULATION: BALANCING ETHICS, TRANSPARENCY, AND INNOVATION

Marcia Narine Weldon, Gabrielle Thomas, and Lauren Skidmore [1]

Imagine that you are sitting at home feeling sick. Instead of calling your doctor and waiting weeks or even months for an appointment or going to urgent care for faster service, you pick up your smartphone, which scans your face and body, diagnoses your illness, sends an autonomous ambulance to pick you up, alerts the appropriate medical staff at the hospital that you're on the way, and, by the time you get to the hospital, you're wheeled directly into surgery where the human staff, guided by artificial intelligence (AI), treats you. If you're not sick enough to go to the hospital, the AI-powered app on your smartphone determines the precise medication you need tailored to your personal genetic makeup, "writes" the prescription, and sends it to your 3D printer, which spits out your medication with the proper dosage for your illness, so you never need to go to urgent care or the pharmacy. Sounds

[1] Marcia Narine Weldon is the Director of the Transactional Skills Program at the University of Miami School of Law, a University of Miami President's Academic Fellow, a Lecturer in Law, the General Counsel of an artificial intelligence startup, and a former Deputy General Counsel, Chief Compliance Officer, and Chief Privacy Officer of a Fortune 500 Company. Coauthor Gabrielle Thomas is a 3L Soia Mentschikoff Scholar, Public Interest Scholar, and Human Rights Clinic Fellow and co-author Lauren Skidmore is a 2L at the University of Miami School of Law. We thank Stephanie Sylvestre, the founder of AI-powered Avatar Buddy, for her technical assistance and review of this Article.

crazy? According to at least one emergency room doctor,[2] this could be our future in a few years thanks to artificial intelligence.

Now imagine that you are a famous psychology professor who has a profitable speaking, consulting, and podcasting business. A former student sends you a chatbot that is a virtual replica of yourself with your voice, your ideas, and your intonations. Instead of paying you, people can access the chatbot, which sounds like you in substance and style, for free without compensation to, or permission from, you. You no longer receive lucrative invitations to speak and your income drops significantly. Lawyers tell you that you may not have legal recourse unless Congress passes the NO FAKES Act,[3] and even so, if the replicas were developed in countries outside of the United States' jurisdiction, there would be no remedy. This scenario is not as far-fetched as it sounds. In fact, influential American psychologist Martin Seligman and celebrity psychotherapist Esther Perel both had digital replicas of them made without their consent.[4] Although those individuals did not take

[2] Alan Thompson, AI + MEDICINE - WITH HARVEY CASTRO MD (GPT-4, MED-PALM 2, CARBON HEALTH, AMBIENCE, 311 CHATGPT CALL), YOUTUBE (Jul. 5, 2023), https://www.youtube.com/watch?v=jTmkiGjrgpA&t=737s&ab_channel=DrAlanD.Thompson.

[3] Mohar Chatterjee, *A New Kind of AI Copy Can Fully Replicate Famous People. The Law Is Powerless*, POLITICO (Dec. 30, 2023, 7:00 AM), https://www.politico.com/news/magazine/2023/12/30/ai-psychologist-chatbot-00132682.

[4] *Id.*

legal action, disputes over the potential use of AI replicas of actors was a central issue that led to the multibillion dollar strike in Hollywood in 2023.[5]

Hollywood and the film industry may have an even bigger problem. On February 15, 2024, the leader in generative AI technology, OpenAI, announced "Sora," a text to video application[6] that creates video so realistic that filmmaker and producer, Tyler Perry, halted a studio expansion project worth $800 million dollars so that he could assess the potential impact.[7] Even OpenAI recognized the potential dangers and limited the initial release to "red teamers—domain experts in areas like misinformation, hateful content, and bias—who will be adversarially testing the model" and to educators, policymakers, and artists to "identify positive use cases."[8]

OpenAI is likely concerned about Sora's ability to deceive the public, and this fear isn't far-fetched. In the same month that OpenAI released Sora, an employee in Hong Kong disbursed over $25 million USD to five fake

[5] Justin Hughes, *Opinion: Can Hollywood's New SAG-AFTRA Contract Hold AI at Bay?*, L.A. TIMES (Nov. 30, 2023, 3:00 AM), https://www.latimes.com/opinion/story/2023-11-30/ai-hollywood-sag-aftra-strike-streaming-residuals-digital-replacement [https://perma.cc/9H6H-9G3E].

[6] *Sora*, OPENAI, https://openai.com/sora (last visited Mar. 31, 2024) [https://perma.cc/VN95-P4J8].

[7] Katie Kilkenny, *Tyler Perry Puts $800m Studio Expansion on Hold After Seeing OpenAI's Sora: "Jobs Are Going to Be Lost"*, HOLLYWOOD REPORTER (Feb. 22, 2024, 4:07PM), https://www.hollywoodreporter.com/business/business-news/tyler-perry-ai-alarm-1235833276/ [https://perma.cc/65PE-WRFV].

[8] *Sora*, OPENAI, https://openai.com/sora (last visited Mar. 31, 2024) [https://perma.cc/VN95-P4J8].

accounts because he believed that had been instructed to do so by his CFO on a Zoom meeting with other colleagues.[9] The CFO and colleagues were in fact deepfakes—cloned video and voice versions—that looked so real that the skeptical employee chose to release the funds.[10]

We acknowledge that by the time you read this Article, much of the information may be obsolete. OpenAI's ChatGPT[11], Google's Gemini[12] and Bard (renamed Gemini)[13], Elon Musk's Grok[14], Anthropic's Claude 3[15], and other generative AI programs are almost household words now, but the technology's capabilities are growing at an exponential rate. Fear, hype, and hysteria about artificial intelligence is everywhere, particularly related to the concept of "singularity," where AI takes on and surpasses human-like

[9] Heather Chen & Kathleen Magramo, *Finance worker pays out $25 million after video call with deepfake "chief financial officer"*, CNN (Feb. 4, 2024, 02:31AM), https://www.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html [https://perma.cc/F8HR-FJ6D].

[10] *Id.*

[11] *About*, OPENAI, https://openai.com/about (last visited Sept. 9, 2023) [https://perma.cc/AM8Q-EWZ5].

[12] *Welcome to the Gemini Era*, GOOGLE DEEPMIND, https://deepmind.google/technologies/gemini/#introduction (last visited Dec. 29, 2023) [https://perma.cc/MZ5W-QJR9].

[13] *Meet Bard*, GOOGLE, https://bard.google.com (last visited Sept. 9, 2023) [https://perma.cc/8KGB-8R83].

[14] Shirin Ghaffary, *Elon Musk's Grok Represents a Serious Threat to ChatGPT*, BLOOMBERG, https://www.bloomberg.com/news/newsletters/2023-12-14/elon-musk-s-grok-represents-a-serious-threat-to-chatgpt (last visited Dec. 29, 2023) [https://perma.cc/689R-R9E2].

[15] *Anthropic: Making AI systems you can rely on*, ANTHROPIC, https://www.anthropic.com/company, (last visited Sept. 9, 2023) [https://perma.cc/T2EK-YC29].

capabilities.[16] This seemed like a wild science fiction fantasy when researchers proposed the concept over fifty years ago.[17] Now some futurists claim that the technology in general will advance as much in the next ten years as it has in the past one hundred and that the singularity with AI could occur as early as 2030, with the majority of AI experts believing that singularity will occur by 2050.[18]

When used properly, AI can amplify human capabilities.[19] The technology can yield endless benefits for economic growth, making professionals' lives easier by significantly reducing time to complete mundane tasks and even improving medical treatment.[20] For lawyers, AI

[16] For more essays on the concept of "singularity," *see generally, AI and the Growing Problem of Trust*, SINGULARITY2030 (Sept. 8, 2023), https://singularity2030.ch/ai-and-the-growing-problem-of-trust/ [https://perma.cc/3HS5-5YWV].

[17] Anthony M. Zador, *A critique of pure learning and what artificial neural networks can learn from animal brains*, 10, NATURE COMMUNICATIONS, 3770, 3772, https://doi.org/10.1038/s41467-019-11786-6, (Aug. 21, 2019) [https://perma.cc/EP75-CGBE].

[18] Zoë Corbyn, *Peter Diamandis: 'In the next 10 years, we'll reinvent every industry'*, GUARDIAN, (Jan. 25, 2020), https://www.theguardian.com/technology/2020/jan/25/peter-diamandis-future-faster-think-interview-ai-industry#:~:text=You%20say%20in%20the%20next,it%20also%20become%20more%20capable [https://perma.cc/B4ZP-QWLZ]; Cem Dilmegani, *When will singularity happen?1700 expert opinions of AGI [2023]*, AIMULTIPLE, (Aug. 13, 2023), https://research.aimultiple.com/artificial-general-intelligence-singularity-timing/ [https://perma.cc/66M9-GAVU].

[19] Ben Shneiderman, *Human-Centered Artificial Intelligence*, UNIVERSITY OF MARYLAND, https://hcil.umd.edu/human-centered-ai/ (last visited Dec. 29, 2023) [https://perma.cc/A75W-M8X5].

[20] Eve Gaumond & Catherine Régis, *Assessing Impacts of AI on Human Rights: It's Not Solely About Privacy and Nondiscrimination*, LAWFARE, https://www.lawfareblog.com/assessing-impacts-ai-human-rights-its-not-solely-about-privacy-and-nondiscrimination (last visited Dec. 29, 2023); Bernd Carsten Stahl, *Ethical Issues of AI*, SPRINGERLINK, https://link.springer.com/chapter/10.1007/978-3-030-69978-9_4 (last visited Dec. 29,

accelerates legal research, drafting legal documents, client communication, E-discovery, document review, and deposition review.[21] By transforming the legal industry, AI also accelerates efficiency.[22] Scaling artificial intelligence can create massive competitive advantage and create new opportunities for innovation[23] and growth and, when harnessed properly, can solve the greatest problems of our time from climate change to poverty, educational inequities, and health disparities.

In our view, the future of AI is one where humans and AI work together, not separately, to promote democracy, peace, and sustainable development.[24] Even so, the rapid advancement of AI capabilities over the past two years, which even surprised the companies developing them, sparks

2023); Dep't for Science, Innovation &Technology, Office for Artificial Intelligence, *Policy Paper: A pro-innovation approach to AI regulation*, GOV.UK, https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper (last visited Dec 29, 2023).

[21] DISCO, https://csdisco.com/discovery/us-disco (last visited Dec. 29, 2023) [https://perma.cc/3ETJ-AHCB].

[22] Stahl, *supra* note 20.

[23] *Artificial Intelligence and AI at Scale*, BOSTON CONSULTING GROUP, https://www.bcg.com/capabilities/digital-technology-data/artificial-intelligence?utm_source=search&utm_medium=cpc&utm_campaign=digital&utm_description=paid&utm_topic=ai&utm_geo=global&utm_content=future_of_ai_group&gclid=Cj0KCQjwzdOlBhCNARIsAPMwjbwjMA_l5Cbt2RxqlpCviPvz9wivcov2u97QnTeup7kjnySpJNY40SgaAqcYEALw_wcB (last visited Dec. 29, 2023) [https://perma.cc/9QH5-5VBT]; *What Is the Future of AI?*, ANALYTICS VIDHYA https://www.analyticsvidhya.com/blog/2023/04/future-of-ai/#:~:text=The%20future%20of%20AI%20in%202050%20is%20uncertain%2C%20but%20it,shaping%20the%20future%20of%20AI (last visited Nov. 2, 2023) [https://perma.cc/875U-7U56].

[24] A DECLARATION FOR THE FUTURE OF THE INTERNET, THE WHITE HOUSE, https://www.whitehouse.gov/wp-content/uploads/2022/04/Declaration-for-the-Future-for-the-Internet_Launch-Event-Signing-Version_FINAL.pdf (last visited Dec. 29, 2023) [https://perma.cc/88HZ-AZLZ].

concern for the future of humanity.[25] On March 22, 2023, the Future of Life Institute released an Open Letter asking AI companies to pause technological progress due to unknown and potentially existential risks.[26] It received more than 33,709 signatures as of December 2023, including signatures from Elon Musk, co-founder of OpenAI, and Steve Wozniak of Apple, who requested a pause on research on all AI systems more powerful than GPT-4.[27]

Ethical issues arising from AI include environmental impacts of innovation;[28] harm to physical integrity if algorithms go astray;[29] lack of trust;[30] wholesale unemployment because of worker displacement;[31] lack of

[25] Stahl, *supra* note 20; Kevin Collier, *AI Risks Leading Humanity To 'Extinction,' Experts Warn*, NBC NEWS, https://www.nbcnews.com/tech/tech-news/ai-risks-leading-humanity-extinction-experts-warn-rcna86791 (last visited Apr. 03, 2024) [https://perma.cc/3DGN-JBSJ].

[26] *Pause Giant AI Experiments: An Open Letter*, FUTURE OF LIFE INST., https://futureoflife.org/open-letter/pause-giant-ai-experiments/ (last visited Dec. 29, 2023) [https://perma.cc/3K5F-BFJH].

[27] Margo Anderson, *'AI Pause' Open Letter Stokes Fear and Controversy*, IEEE SPECTRUM, https://spectrum.ieee.org/ai-pause-letter-stokes-fear#toggle-gdpr (last visited Dec. 29, 2023) [https://perma.cc/SJ98-2BA6]; Emily Kitazawa, *The AI Development Pause: A Good or Foolish Idea?*, SHORTFORM, https://www.shortform.com/blog/ai-development-pause/#:~:text=In%20March%2C%20over%201%2C000%20tech,on%20the%20benefits%20of%20AI (last visited Dec. 29, 2023) [https://perma.cc/6G8F-AHKB].

[28] Tate Cantrell, *The True Cost of AI Innovation*, SCIENTIFIC COMPUTING WORLD, https://www.scientific-computing.com/analysis-opinion/true-cost-ai-innovation (last visited Dec. 29, 2023) [https://perma.cc/7ELM-WRCM].

[29] S. Matthew Liao, *Ethics of AI and Health Care: Towards a Substantive Human Rights Framework*, 42 TOPOI 857, 864 (2023).

[30] Richard Carufel, *As AI Permeates Digital Culture, Consumers Now Cite a Lack of Trust—and Fear of Malicious Intent*, AGILITY PR SOLUTIONS (Feb. 16, 2023), https://www.agilitypr.com/pr-news/public-relations/as-ai-permeates-digital-culture-consumers-now-cite-a-lack-of-trust-and-fear-of-malicious-intent/ [https://perma.cc/9CXV-LVRD].

[31] *Generative AI Could Raise Global GDP by 7%*, GOLDMAN SACHS (Apr. 5, 2023), https://www.goldmansachs.com /intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html [https://perma.cc/2ZHS-LQCA].

quality jobs;[32] disappearance of jobs;[33] lack of privacy;[34] potential systems failures in military use;[35] lack of informed consent in legal matters;[36] bias and discrimination;[37] exacerbation of inequality;[38] misuse of personal data;[39]

[32] Josie Cox, *AI Anxiety: The Workers Who Fear Losing Their Jobs to Artificial Intelligence*, BBC (July 23, 2023), https://www.bbc.com/worklife/article/20230418-ai-anxiety-artificial-intelligence-replace-jobs [https://perma.cc/ EB4N-LEE6].

[33] Calum McClelland, *The Impact of Artificial Intelligence - Widespread Job Losses*, IT FOR ALL (Jan. 30, 2023), https://www.iotforall.com/impact-of-artificial-intelligence-job-losses [https://perma.cc/R99H-4PWZ].

[34] Paul W. Grimm et al., *Artificial Justice: The Quandary of AI in the Courtroom*, JUDICATURE INT'L 2 (Sept. 2022), https://judicature.duke.edu/articles/artificial-justice-the-quandary-of-ai-in-the-courtroom/ [https://perma.cc/923U-3ZSL]; Jacob O. Arowosegbe, *Data Bias, Intelligent Systems and Criminal Justice Outcomes*, 31 INT'L J. L. & INFO. TECH. 22, 31 (2023).

[35] Daniel Hoadley et al., *Artificial Intelligence and National Security*, CONG. RSCH. SERV. (Apr. 26, 2018), https://a51.nl/sites/default/files/pdf/R45178.pdf; Wyatt Hoffman et al., *Reducing the Risks of Artificial Intelligence for Military Decision Advantage*, CTR. FOR SEC. AND EMERGING TECH. 1, 16 (Mar. 2023), ,https://cset.georgetown.edu/publication/reducing-the-risks-of-artificial-intelligence-for-military-decision-advantage/ [https://perma.cc/9KRB-YKGK].

[36] Jie Zhang et al., *Ethics and Governance of Trustworthy Medical Artificial Intelligence*, 23 BMC MED. INFORMATICS AND DECISION MAKING 1, 10 (July 23, 2023), https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-023-02103-9 [https://perma.cc/56MZ-VABP].

[37]Alexa Robertson & Max Maccarone, *AI Narratives and Unequal Conditions. Analyzing the discourse of liminal expert voices discursive communicative spaces*, 47 TELECOMM. POL'Y 1, 2, 7, 10 (Oct. 29, 2022), https://www.sciencedirect.com/science/article/pii/S0308596122001641?via%3Dihub/ [https://perma.cc/T4DU-7R9M]; Marissa Gerchick, *The Federal Government Should Not Waste the Opportunity to Address Algorithmic Discrimination*, ACLU (Dec. 6, 2022), https://www.aclu.org/news/racial-justice/the-federal-government-should-not-waste-the-opportunity-to-address-algorithmic-discrimination [https://perma.cc/C87V-XHCF].

[38] Robertson & Maccarone, *supra* note 37; Inga Ulnicane, *Power and Politics in Framing bias in Artificial Intelligence Policy*, 40 REV. OF POL'Y RSCH. 665, 681 (June 28, 2023), https://onlinelibrary.wiley.com/doi/full/ 10.1111/ropr.12567/.

[39] Van Rijmenam, *Privacy in the Age of AI: Risks, Challenges, and Solutions*, THE DIGITAL SPEAKER (Feb. 17, 2023) https://www.thedigitalspeaker.com/privacy-age-ai-risks-challenges-solutions/ [https://perma.cc/WRT6-8AXD]; *AI and Privacy: The Privacy Concerns Surrounding AI, Its Potential Impact On Personal Data*, ECON. TIMES (Apr. 25, 2023), https://economictimes.indiatimes.com/news/how-to/ai-and-privacy-the-privacy-concerns-surrounding-ai-its-potential-impact-on-personal-data/articleshow/99738234.cms?from=mdr [https://perma.cc/3XJL-Y7AN].

negative impact on democracy;[40] negative impact on the justice system;[41] potential for criminal and malicious use;[42] loss of freedom and individual autonomy;[43] contested ownership of data;[44] reduction of human contact;[45] problems of control and use of data systems;[46] lack of accuracy and predictive recommendations;[47] concentration of economic power;[48] violation of human rights in the supply chain;[49] violation of human rights of end users;[50]

[40] Mark Coeckelbergh, *Democracy, Epistemic Agency, and AI: Political Epistemology in Times of Artificial Intelligence*, 3 A.I. & ETHICS 1341, 1344 (Nov. 22, 2022), https://www.ncbi.nlm.nih.gov/pmc/articles /PMC9685050/.

[41] Grimm et al., *supra* note 34 at 2–3; Arowasegbe, *supra* note 34 at 28, 29–30.

[42] *The Criminal use of ChatGPT - A Cautionary Tale about Large Language Models*, EUROPOL, https://www. europol.europa.eu/media-press/newsroom/news/criminal-use-of-chatgpt-cautionary-tale-about-large-language-models (last visited Feb. 14, 2023) [https://perma.cc/J8AE-Y958].

[43] Bram Vaassen, *AI, Opacity, and Personal Autonomy*, 35 PHIL. & TECH. 87, 88 (Sept. 15, 2021), https://link.springer.com/article/10.1007/s13347-022-00577-5 [https://perma.cc/BCP6-QEMS].

[44] *Event Highlight: AI and Digital Inequities Summit*, NORRAG (Aug. 24, 2023), https://www.norrag.org/ai-and-digital-inequities-summit/ [https://perma.cc/7A7M-U6UC].

[45] THE IMPACT OF ARTIFICIAL INTELLIGENCE ON THE FUTURE OF WORKFORCES IN THE EUR. UNION AND THE U.S., WHITE HOUSE, https://www.whitehouse.gov/wp-content/uploads/2022/12/TTC-EC-CEA-AI-Report-12052022-1.pdf (last visited Jan. 8, 2024) [https://perma.cc/F3JE-BB32].

[46] Darrell M. West et al., *The Three Challenges Of AI Regulation*, BROOKINGS (June 15, 2023), https://www.brookings.edu/articles/the-three-challenges-of-ai-regulation/ (last visited Jan. 8, 2024).

[47] DEP'T OF COM., A.I. RISK MGMT. FRAMEWORK (AI RMF 1.0) 13 (2023), https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf (last visited Jan. 8, 2024).

[48] Martin Neil Bailey et al., *Machines of Mind*: *The Case for an AI-Powered Productivity Boom*, BROOKINGS (May 10, 2023), https://www.brookings.edu/articles/machines-of-mind-the-case-for-an-ai-powered-productivity-boom/ (last visited Jan. 8, 2024).

[49] Hannah Darnton & Lale Tekisalp, *AI and Human Rights in Retail*, BSR (Apr. 3, 2023), https://www.bsr.org/en/reports/ai-and-human-rights-in-retail [https://perma.cc/AG23-5PKS].

[50] VOLKER TÜRK*, A. I. Must Be Grounded in Human Rights, Says High Commissioner*, U.N. HUM. RTS. OFF. HIGH COMM'R (July 12, 2023), https://www.ohchr.org/en/statements/2023/07/artificial-intelligence-must-be-grounded-human-rights-says-high-commissioner.

unintended, unforeseeable adverse impacts;[51] negative impact on vulnerable groups;[52] and lack of accountability and liability. For these reasons, governments must act now.

However, the unprecedented rate of advancement coupled with the complexity of the technology makes it particularly difficult to regulate. On the one hand, there is almost universal agreement among the largest players in technology that governments must regulate AI. These leaders have consulted with governments around the world, often behind closed doors.[53] Smaller companies and some venture capitalists, in contrast, claim that the largest players seek to lock in their regulation preferences to stifle competition.[54] In 2023, at least twenty-five states, Puerto Rico, and the

[51] Syed Hamza Sohail, *Unintended Bias in AI-Driven Healthcare Applications,* HIT CONSULTANT (Feb. 14, 2023), https://hitconsultant.net/2023/02/14/unintended-bias-in-ai-driven-healthcare-applications/; Lorin Brennan, *AI Ethical Compliance is Undecidable*, 14 HASTINGS SCI. & TECH. L.J. 311, 313–14 (2023); Mary Louise Malig, *The Amazing Artificial Intelligence: The Urgency for Policies to Protect People,* SYSTEMIC ALTS., at 24 (2023), https://systemicalternatives.org/2023/06/06/the-urgency-for-policies-to-protect-people/ (last visited Jan. 9, 2024); *AI Impact Assessment*, DUTCH MINISTRY OF INFRASTRUCTURE AND WATER MGMT. (2023), https://www.government.nl/binaries/government/documenten/publications/2023/03/02/ai-impact-assessment/AI+Impact+Assessment.pdf (last visited Jan. 9, 2024).

[52] Bangul Khan et al., *Drawbacks of Artificial Intelligence and Their Potential Solutions in the Healthcare Sector,* BIOMEDICAL MATERIALS & DEVICES (Feb. 8, 2023), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9908503/ [https://perma.cc/5CXH-H6KS].

[53] Mary Clare Jalonick & Matt O'Brien, *Tech Industry Leaders Endorse Regulating Artificial Intelligence At Rare Summit in Washington*, AP NEWS (Sept. 13, 2023, 8:57 PM), https://apnews.com/article/schumer-artificial-intelligence-elon-musk-senate-efcfb1067d68ad2f595db7e92167943c [https://perma.cc/3ZQG-HDN8].

[54] Gerrit De Vynck, *Big Tech Wants AI Regulation. The Rest of Silicon Valley Is Skeptical*, WASH. POST (Nov. 9, 2023, 7:00 AM), https://www.washingtonpost.com/technology/2023/11/09/ai-regulation-silicon-valley-skeptics/ [https://perma.cc/9AH7-Z7UQ].

District of Columbia introduced artificial intelligence bills, while only eighteen states and Puerto Rico actually enacted artificial intelligence legislation.[55] But just like the privacy sphere, this piecemeal approach to regulation is unsustainable for both businesses and consumers.

In this Article, we propose a regulatory framework after examining the current landscape surrounding generative AI ("GAI"), including the EU's AI Act, which the EU Parliament enacted into law in March 2024,[56] as well as proposals from major powers such as the United States, China, and the UK. In Part I, we provide a basic, high-level introduction into the types of artificial intelligence and how they work. In Part II, we explore the "good," the "bad," and the "ugly" use cases, highlighting current and future possibilities and problems related to business operations, health care, climate change, access to justice, education, intellectual property, the media, and democratic life. In Part III, we review and critically examine current and proposed legislative frameworks in the United States, the European Union, the UK, Canada, China, and other nation states. In Part IV, we provide guidance to policymakers and regulators on AI regulations and governance

[55] *Artificial Intelligence 2023 Legislation*, NAT'L CONF. OF STATE LEGIS., https://www.ncsl.org/technology-and-communication/artificial-intelligence-2023-legislation [https://perma.cc/MHM7-VALN] (last updated Jan. 12, 2024).

[56] *Artificial Intelligence Act: MEPs adopt landmark law*, EUR. PARLIAMENT (Mar. 12, 2024, 12:25 PM), https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law (last visited Mar 31, 2024).

so that AI is "valid and reliable, safe, secure and resilient, accountable and transparent, explainable and interpretable, privacy-enhanced, and fair with harmful bias managed," in accordance with NIST standards of the U.S. Department of Commerce.[57] Any approach to regulation will require collaborative efforts between technologists, ethicists, industry experts, and the legal community to ensure that AI development is aligned with human values and societal goals. The objective is to harness AI's transformative power while steering it in a direction that upholds and enhances human dignity, equity, and well-being.

Part I: A Brief Introduction to Artificial Intelligence[58]

Artificial intelligence describes computing programs and models that "display human-like capabilities such as reasoning, learning, planning and creativity."[59] While there is no single accepted definition of AI, it is widely understood that "[i]t is a broad field focused on enhancing the ability of computers to make 'appropriate generalizations in a timely fashion based on

[57] U.S. DEPT. OF COM., NAT'L INST. OF STANDARDS & TECH., *3 AI Risks and Trustworthines*s, (Jan. 11, 2023), https://airc.nist.gov/AI_RMF_Knowledge_Base/AI_RMF/Foundational_Information/3-sec-characteristics# [https://perma.cc/2W5J-LAG2].

[58] For an easily digestible introduction to AI*, see* W*hat is Generative AI and how does it work? – the Turing lectures with Mirella Lapata*, YOUTUBE (2023), available at: https://youtu.be/_6R7Ym6Vy_I?si=MVhsftmVWoY1mOYA [https://perma.cc/H4EF-T95H].

[59] *What is artificial intelligence and how is it used?*, EUR. PARLIAMENT (Apr. 9, 2023), https://www.europarl.europa.eu/news/en/headlines/society/20200827STO85804/what-is-artificial-intelligence-and-how-is-it-used [https://perma.cc/6KVV-JJVC].

limited data.'"[60] Some have linked the difficulty in developing a definition of "artificial intelligence" to the difficulty in defining "intelligence," because there are no existing definitions of intelligence that are not descriptive or tied to human intelligence.[61] According to the National Security Commission on Artificial Intelligence, AI is not a single piece of hardware or software, but rather "a constellation of technologies that give a computer system the ability to solve problems and to perform tasks that would otherwise require human intelligence."[62]

There are different categories of AI, and the distinctions are important. Traditional, narrow, or weak AI are what most consumers have been using for years. These machines are designed to complete a certain task. This includes virtual assistants such as Siri or Alexa, recommendation systems that steer you toward products or build playlists based on your past activity, search algorithms such as Google that focus on finding information based on keywords, translation services, industrial robots that perform specific, repetitive tasks on an assembly line, spam filters, natural language

[60] Cameran Ashraf, *Artificial Intelligence and the Rights To Assembly and Association*, J. OF CYBER POL'Y, (2020), https://www.tandfonline.com/doi/full/10.1080/23738871.2020.1778760 [https://perma.cc/T94Q-7VGP].

[61] Rafael Dean Brown, *Property ownership and the legal personhood of artificial intelligence*, 30 INFO. & COMMC'N TECH. L., 208, 211–12 (2021), https://www.tandfonline.com/doi/full/10.1080/13600834.2020.1861714 [https://perma.cc/DLA4-GYCP].

[62] *Final Report, Artificial Intelligence in Context*, NAT'L SEC. COMM'N ON A.I., https://reports.nscai.gov/final-report/ [https://perma.cc/T2HQ-C4QZ] (last visited Feb. 14, 2023).

programs that help build chatbots, decision trees that determine whether a client qualifies for a loan, certain programmed autonomous vehicles, and facial recognition systems.[63] Traditional AI has also been used for fraud detection, data analytics, and to automate tasks.[64] Traditional or weak AI identifies patterns but does not have the power to create anything new.[65]

Generative AI, on the other hand, can create new texts, songs, and artwork, duplicate voices, write code, and synthesize data much more quickly than humans based on the data used to train it.[66] Although traditional AI does not have the same powers as generative AI, as we will discuss in Part II, misuse of tools such as surveillance and facial recognition technology can have devastating consequences. As of the time of this writing, "strong AI" or artificial general intelligence ("AGI"), does not exist.[67] Theoretically, AGI

[63] Bernard Marr, *The Difference between Generative AI and Traditional AI: An Easy Explanation for Anyone,* FORBES (July 24, 2023), https://www.forbes.com/sites/bernardmarr/2023/07/24/the-difference-between-generative-ai-and-traditional-ai-an-easy-explanation-for-anyone/?sh=4ed09d1e508a [https://perma.cc/E454-EDNC]; Emily Heaslip, *What's the Difference Between Traditional AI vs Generative AI?*, U.S. CHAMBER OF COM. (Oct. 16, 2023), https://www.uschamber.com /co/run/technology/traditional-ai-vs-generative-ai [https://perma.cc/TJW8-XT3H].

[64] Sunil Ramlochan, *The Yin and Yang of AI How Traditional and Generative Models Differ and Compliment Each Other,* PROMPT ENG'G INST. (Sept. 21, 2023), https://promptengineering.org/the-yin-and-yang-of-ai-how-traditional-and-generative-models-differ-and-complement-each-other/ [https://perma.cc/PYL4-JS4A].

[65] *Id.*

[66] *What is generative AI?*, MCKINSEY & CO. (Jan. 19, 2023), https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-generative-ai [https://perma.cc/D54A-FHLG].

[67] *What is strong AI?*, IBM, https://www.ibm.com/topics/strong-ai#:~:text=While%20there%20are%20no%20clear,in%20human%20intelligence%20and %20ability [https://perma.cc/822S-CAGZ] (last visited Jan. 11, 2023).

would reason like humans, learn from its mistakes, display emotional intelligence, communicate with the human it was interacting with, and would not need reprogramming.

AI attempts to simulate or replicate human intelligence in machines, a project that dates back to a 1950 paper[68] by mathematician Alan Turing, who helped to shorten World War II by cracking the German Enigma naval code.[69] Turing proposed the self-named "Turing Test" to determine whether the computer should be deemed intelligent based on how well it fools humans into thinking it's human.[70] Conceivably, "strong AI" would pass the Turing test. Turing was not alone in exploring these ideas. In the 1950s, computer scientist Arthur Samuel developed the concept of machine learning, which is "'the field of study that gives computers the ability to learn without explicitly being programmed."[71] The term comes from the idea that the program is "trained" on sets of data and "learns" to predict outcomes or generate an outcome based on inferences from the provided data.[72] Samuel's seminal 1959 article "Some Studies in Machine Learning Using the Game of

[68] A. M. Turing, I.-COMPUTING MACH. AND INTEL., 236 MIND 433, (Oct. 1950), https://academic.oup.com/mind/article/LIX/236/433/986238 [https://perma.cc/XGA5-K782].

[69] *Id.*

[70] Brown, *supra* note 61, at 211.

[71] Sara Brown, *Machine learning, explained*, MIT SLOAN (Apr. 21, 2021)), https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained [.https://perma.cc/3U5X-K3FQ].

[72] *Id.*

Checkers,"[73] was one of the earliest demonstrations of artificial intelligence. In 1966, Joseph Weizenbaum at MIT developed ELIZA, a program that marked the beginning of human-machine conversation.[74] ELIZA, named after Eliza Doolittle from the 1913 play *Pygmalion*, mimicked a Rogerian psychotherapist by turning user inputs into questions. Weizenbaum intended to demonstrate the program's limitations. However, the program's unexpected ability to engage "delusional" users led Weizenbaum to warn against excessive reliance on computers and AI.[75]

Artificial intelligence networks are inspired by the structure of the human brain and consist of layers made up of nodes, or artificial neurons.[76] These nodes process input using specific mathematical functions and pass the output to the next layer.[77] Neural networks are critical because they are the basis of GAI, AGI, and singularity. With the right configuration and

[73] Arthur L. Samuel, *Some Studies in Machine Learning Using the Game of Checkers*, 3 IBM J. 535 (July 1959), https://people.csail.mit.edu/brooks/idocs/Samuel.pdf [https://perma.cc/35ZQ-S64H].

[74] Minh Tran, *What Did a Chat Bot Made in 1966 Tell Us About Human Language?*, MEDIUM (Oct. 3, 2022), https://towardsdatascience.com/what-did-a-chat-bot-made-in-1966-tell-us-about-human-language-886613a16a7f [https://perma.cc/7LJF-7MWT].

[75] Oshan Jarow, *How the first chatbot predicted the dangers of AI more than 50 years ago*, VOX (Mar. March5, 2023, 6:01 AM), https://www.vox.com/future-perfect/23617185/ai-chatbots-eliza-chatgpt-bing-sydney-artificial-intelligence-history [https://perma.cc/B6CV-QJV7]; Karla Erickson, *What a Precursor To Chatgpt Taught Us About AI – in 1966*, SALON (April 10, 2023, 5:30 AM), https://www.salon.com/2023/04/10/what-a-precursor-to-chatgpt-taught-us-about-ai--in-1966/ [https://perma.cc/TJ3Q-PJUS].

[76] A.I. FOR LAW FIRMS: A PRACTICAL AND TACTICAL GUIDE, INFOTRACK 10, https://info.abovethelaw.com/hubfs/Infotrack%202023%20Content%20Syndication/AI_For_Lawfirms_ebook_infotrack.pdf (last visited Jan. 9, 2024).

[77] *Id.*

functions,[78] these networks enable deep learning and machine learning, offering remarkable flexibility and the ability to handle complex, nonlinear relationships.[79] Although the study of artificial neurons began in the mid-twentieth century, over the last few decades we have seen a surge in the use of artificial neural networks.[80] This increase aligns with advancements in computing power.

In machine learning, the computer assists in creating the model by learning to program itself through trial and error with data.[81] To create a model, programmers gather and prepare data sets, choose a machine learning algorithm, and let the computer take it from there to learn and program itself.[82] An algorithm is "a procedure for solving a mathematical problem in a finite number of steps that frequently involves repetition of an operation."[83] Lawyers may find it helpful to conceptualize "algorithms as the 'IRAC' of computers: when faced with an *Issue*, the computer finds *Rules* (or a set of rules), then *Applies* those rules to the problem and comes up with a

[78] Brown, *supra* note 61.

[79] Puru Rattan et. al., *Artificial Intelligence and Machine Learning: What You Always Wanted to Know but Were Afraid to Ask*, GASTRO HEP ADVANCES (2021), https://www.ghadvances.org/article/S2772-5723(21)00025-X/fulltext (last visited Jan. 9, 2024).

[80] *Id.*

[81] MCKINSEY & CO., *supra* note 66.

[82] *Id.*

[83] *Algorithm*, MERRIAM-WEBSTER, https://www.merriam-webster.com/dictionary/algorithm [https://perma.cc/P4ZX-8XWN] (last visited Jan. 8, 2024).

Conclusion."[84] Programmers can "tweak the model" to achieve more accurate outcomes or other desired results.[85] As the computer processes the data through the algorithm, it gradually "learns" and refines its programming.

Deep learning, a subset of machine learning, is akin to giving computers a level of understanding that closely resembles human thought processes such as the ability to recognize text, pictures, and sounds.[86] Unlike traditional machine learning, which requires more hands-on guidance, deep learning automates a significant portion of its learning process.[87] This means it can take in vast amounts of data and, much like the human brain, find patterns and make sense of them without much human intervention. Deep learning is better suited to complex tasks than machine learning.[88] This advanced processing is particularly useful in areas where understanding subtle differences or complex patterns is crucial, such as recognizing faces in photos or understanding spoken words in voice recognition software.

Whereas deep learning is an evolution of machine learning, a large language model ("LLM") is a machine learning model that has been pre-

[84] INFOTRACK, *supra* note 76, at 7.

[85] Brown, *supra* note 61.

[86] *What is Deep Learning?,* AMAZON, https://aws.amazon.com/what-is/deep-learning/#:~:text= Dee p%20learning%20is%20a%20method,produce%20accurate%20insights%20and%20predictions [https://perma.cc/ V4F7-55P7] (last visited Jan. 9, 2024).

[87] *Id.*

[88] *What's the Difference Between Machine Learning and Deep Learning?*, AMAZON, https://aws.amazon.com/compare/the-difference-between-machine-learning-and-deep-learning/ [https://perma.cc/W3C4-JVB6] (last visited Jan. 9, 2024).

trained on large amounts of data using deep learning.[89] Large language models are large-scale versions of a machine learning model called a transformer model.[90] The term GPT, used in ChatGPT, stands for generative pre-trained transformer.[91] LLMs do not require the same type of supervised training that machine learning requires; instead, LLMs train themselves.[92] The goal of LLMs is to produce text or images that could reasonably appear in the prompted context.[93]

LLMs such as ChatGPT are trained on enormous amounts of text data, including books, articles, websites, and other sources from the internet.[94] The training process entails the model learning the statistical relationships between words, phrases, and sentences and subsequently producing relevant responses when given a prompt by guessing the correct sequence of words.[95] Because of how the models are trained, if the model is

[89] *What are Large Language Models (LLM)?*, AMAZON, (July 20, 2023). https://aws.amazon.com/what-is/large-language-model/#:~:text=Large%20language%20models%20(LLM)%20are,decoder%20with%20self%2Dattention%20capabilities [https://perma.cc/PWC3-86JC].

[90] AMAZON, *supra* note 88.

[91] Noam Kolt, *Predicting Consumer Contracts*, 37 BERKELEY TECH. L. J. 71, 74 (2022).

[92] AMAZON*, supra* note 86.

[93] *Id.*

[94] AMAZON, *supra* note 89; Mark Riedl, A *Very Gentle Introduction to Large Language Models without The Hype*, MEDIUM (Apr. 13, 2023), https://mark-riedl.medium.com/a-very-gentle-introduction-to-large-language-models-without-the-hype-5f67941fa59e [https://perma.cc/TH78-6TGS0].

[95] AMAZON, *supra* note 89.

incorrect in its guesswork, the programmers can tweak the model bit by bit until it achieves the desired accuracy and responses.[96]

However, LLMs can "hallucinate," or simply make things up, because the models use data licensed (or not) from third parties; data from the internet, including PDF files, chat groups, and sites such as Reddit; and other user data,[97] which may have unverified and biased information. Understanding how a program utilizes input data to determine its output response is important because models can be fooled, undermined, or simply fail at any task, and knowing where those shortcomings originate allows users to be able to vet and confirm the AI output.[98]

Analyzing the process the program uses helps programmers and end users identify biases integrated in the model or problematic consequences of the output.[99] Understanding the relative accuracy of a model is also critical in effectively using AI because the utilization of some AI outputs requires much stronger accuracy than others.[100] There are tasks where a model does not need to achieve perfect accuracy, such as the model that recommends new shows to Netflix users or songs to add to a playlist on Spotify, but the same

[96] *Id.*
[97] Michael Schade, *How ChatGPT and Our Language Models Are Developed*, OPENAI, https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-language-models-are-developed (last visited Jan. 10, 2024).
[98] Brown, *supra* note 61.
[99] *Id.*
[100] *Id.*

cannot be said of the AI program where inaccuracies could produce extreme adverse effects, such as the program running a self-driving car or autonomous weapons.[101]

Discriminative v. Generative AI

Machine learning programs have two goals: to make inferences about a pattern or to predict the outcome of a pattern.[102] AI models that concentrate predominantly on predicting outcomes and are employed in comparative data classification fall under the category of discriminative AI.[103] These models are integral to supervised machine learning, where they are tasked with making informed predictions or decisions based on a set of input data that has been labeled or classified. The efficacy of these discriminative AI models hinges on their ability to accurately categorize and predict outcomes, which is crucial for their application in various practical and research scenarios.

AI models that prioritize pattern inference and are used to generate new outcomes or create new sets of data are called generative AI.[104] These

[101] *Id.*
[102] Rattan et al., *supra* note 79.
[103] *Id.*; *Generative Models vs Discriminative Models for Deep Learning*, TURINGLEARNING (2022), https://www.turing.com/kb/generative-models-vs-discriminative-models-for-deep-learning#discriminative-model [https://perma.cc/EB32-ZGWJ] (last visited Jan. 9, 2024).
[104] Fieldfisher Silicon Valley, *Generative AI: Privacy Risks & Challenges*, YOUTUBE (Mar. 9, 2023), https://www.youtube.com/watch?v=9xjFsy9_HBs [https://perma.cc/XM29-ADS3].

models focus on the overall general distribution outcomes without focusing on classifying them.[105] Some generative AIs deployed for public use are OpenAI's ChatGPT, Google's LaMDA, Google's Bard, and Lensa AI.[106] Generative models have data generation abilities, but discriminative models do not.[107] A model that classifies images is a discriminative model, whereas a model that creates a brand new image is a generative model.[108]

The Information Asymmetry

Scientists at Google's DeepMind have developed a taxonomy of AI starting with "emerging," which would include chatbots such as ChatGPT and Bard; and "competent," "expert," "virtuoso," and "superhuman," which could perform tasks that humans cannot, such as reading minds, talking to animals, and predicting the future.[109] Notably, OpenAI's mission statement is "to ensure that artificial general intelligence (AGI)—by which [we] mean highly autonomous systems that outperform humans at most economically valuable work—benefits all of humanity. [We] will attempt to directly build

[105] Rattan et. al., *supra* note 72.

[106] *Generative AI: Artificial Intelligence - Large Language Models*, UNIV. OF ME. AUGUSTA, https://umalibguides.uma.edu/c.php?g=1292431&p=9490680 [https://perma.cc/B5FJ-PYNL] (last visited Jan. 9, 2024).

[107] Shailendra Prajapati, *Understanding the Distinction: Generative Models vs. Discriminative Models*, LINKEDIN (June 17, 2023), https://www.linkedin.com/pulse/understanding-distinction-generative-models-vs-shailendra-prajapati/ [https://perma.cc/U49F-LSRN].

[108] *Id.*

[109] Will Douglas Heaven*, Google Deepmind Wants to Define What Counts as Artificial General Intelligence*, MIT TECH. REV. (Nov. 16, 2023), https://www.technologyreview.com/2023/11/16/1083498/google-deepmind-what-is-artificial-general-intelligence-agi/ [https://perma.cc/T7EP-PS8B].

safe and beneficial AGI, but will also consider our mission fulfilled if our work aids others to achieve this outcome." Notwithstanding the significant leaps in AI capabilities in 2022 and 2023, OpenAI's mission statement has not changed since 2018.[110]

As GAI evolves seemingly daily, its increasing complexity underscores a crucial challenge for regulators: the necessity of deeply understanding the technology they are tasked with overseeing. The advancement from basic machine learning to more intricate GAI and AGI systems highlights the dangers of the information asymmetry between regulators and tech companies and could significantly complicate regulatory efforts.

For instance, a critical aspect of neural networks is their capacity to emulate human cognitive processes and analyze extensive data rapidly. This dual capability raises concerns about AI potentially becoming uncontrollable, similar to the unpredictable behavior of a very bright child. Looking at AI as a highly intelligent yet immature entity, like an eleven-year-old, who can process data and provide a seemingly legitimate output within seconds, underscores the uncertainty of its developmental trajectory. While AI can significantly advance progress in the most critical aspects of our lives, as we

[110] *OpenAI Charter*, OPENAI, https://openai.com/charter [https://perma.cc/5NMC-ESGU] (last visited Jan. 9, 2024).

discuss in Part II, this situation is analogous to the difference between a knife wielded by a chef and one in the hands of a killer. Neglecting this focus risks creating laws that fail to tackle the root cause, merely addressing the surface symptoms. In the absence of adequate oversight, there is a danger that AI, like people, might exploit these solutions for malevolent purposes.

The disparity in technical understanding and the rapid advancement raises important questions about the capacity of regulatory bodies to not only grasp the full extent of AI's capabilities and limitations, but also to anticipate and mitigate potential risks and biases and to enact regulation that is "future proof." The onus falls on regulators to develop a nuanced, informed approach, balancing the innovative promise of AI with the imperative of safeguarding public interest in an environment where technological expertise is unevenly distributed. In Part II, we will explore the benefits and risks of some current and future use cases that demonstrate the urgency of regulating in this arena.

Part II. Use Cases: The Good, The Bad, and The Ugly

The Good

Although members of the general public have used generative AI for years when interacting with chatbots on their favorite shopping websites, the advent of ChatGPT, released to the public in November 2022 and used by over 100 million people a week, has transformed the way companies and entities of all types do business and how their clients interact with them. The following description of business use cases was written entirely by ChatGPT:[111]

> For entrepreneurs and small businesses, ChatGPT can be a game-changer, offering versatile support from drafting business plans and proposals to managing customer inquiries through AI-driven chatbots. It can automate social media management, creating engaging posts and responses, and provide efficient market analysis by collating and interpreting consumer data, thereby informing strategic decisions.
>
> Larger companies can leverage ChatGPT for complex tasks such as predictive analytics for market trends, AI-driven recruitment processes, and enhancing business intelligence. It can also streamline internal communication and manage large-scale customer interactions, providing comprehensive support for CRM systems.
>
> Nonprofits can utilize ChatGPT for grant writing assistance, crafting compelling narratives for fundraising campaigns, and managing donor communications effectively. It can also support volunteer coordination and provide insights into social impact metrics, enhancing outreach and operational efficiency.
>
> For governments, ChatGPT offers robust support in policy analysis, drafting and managing public communications, and automating routine administrative tasks. It can play a key role in citizen engagement platforms, facilitating real-time responses and feedback collection, thereby improving public service delivery and governance.

[111] After training ChatGPT in Professor Weldon's writing style, we used the following prompt (with several revisions and iterations): "Please prepare a summary of your capabilities for businesses. What are the common use cases and how could that help companies of all sizes be more profitable? Please write at least two paragraphs and be detailed and specific."

Although in this Article we focus on the need to regulate AI to prevent potentially catastrophic harms, we also recognize that AI has the potential to solve some of the world's greatest problems. We believe that AI's benefits can outweigh the risks if, and only if, governments and technologists work together now for a common goal. The United Nations, for example, has embraced AI to accelerate progress toward achieving the seventeen Sustainable Development Goals (SDGs) designed to end poverty, reduce inequality, increase peace, improve access to health and education, and mitigate the effects of climate change.[112] The SDGs were established to solve development in economic, social, and the environmental dimensions and realize sustainable development by 2030.[113] The major objectives of the SDGs include sharing technology and policy research; building open source infrastructures, cross organization working groups; evaluating the position and negative use of AI; ethics of governance of AI; joint workshops and annual reports; promotion of the use of infrastructures; and applications for underdeveloped regions and countries to ensure "no one is left behind."[114]

[112] *The 17 Goals*, UNITED NATIONS, https://sdgs.un.org/goals [https://perma.cc/K5ES-EU5M] (last visited Jan. 9, 2024).

[113] *Advancing UN Sustainable Development Goals and Digital Cooperation through AI Innovation and Partner Networks*, AI4SDGS COOPERATION NETWORK, https://www.ai-for-sdgs.academy/ai4sdgs-cooperation-network [https://perma.cc/PC5J-XYMZ] (last visited Jan. 9, 2024).

[114] *Id.*

Through the AI for Good program, the UN and over forty sister organizations have worked on hundreds of projects with a strong focus on SDGs 3 (Good Health and Wellbeing), 9 (Industry, Innovation and Infrastructure), 10 (Reduced Inequalities), 16 (Peace, Justice and Strong Institutions), and 17 (Partnership for the Goals) in 2022.[115]

AI and the Environment

Many scientists believe that AI will play a bigger role in predicting weather and climate disasters in the U.S. in 2024.[116] This should come as no surprise as scientists have been using climate prediction models based largely on the rules of physics and chemistry to forecast weather patterns for years.[117] AI can create accurate maps from aerial imagery to inform evacuation planning, classify building damages based on satellite images after natural disasters, and analyze social media data to summarize "kernels of insight."[118] Machine learning can analyze and synthesize climate data to model

[115] *United Nations Activities on Artificial Intelligence (AI)*, ITU PUBL'NS iv, vi (2022), https://s41721.pcdn.co/wp-content/uploads/2021/06/Executive-Summary-2022-Report.pdf [https://perma.cc/W4Q7-GQHR].

[116] Jude Coleman, *AI's Climate Impact Goes Beyond Its Emissions*, SCIENTIFIC AM. (2023), https://www.scientificamerican.com/article/ais-climate-impact-goes-beyond-its-emissions/#:~:text=Training%20and%20running%20an%20AI,way%20AI%20affects%20the%20climate [https://perma.cc/9U3Q-MUMK](last visited Jan. 9, 2024); Jude Coleman, AI'S CLIMATE IMPACT GOES BEYOND ITS EMISSIONS SCIENTIFIC AMERICAN (2023), https://www.scientificamerican.com/article/ais-climate-impact-goes-beyond-its-emissions/#:~:text=Training%20and%20running%20an%20AI,way%20AI%20affects%20the%20climate (last visited Jan 9, 2024).

[117] Coleman, *supra* note 116.

[118] Giulia Cirri, *AI for Climate Change: AI for flood adaptation plans and disaster relief*, OXFORD INSIGHTS (Dec. 11, 2023), https://oxfordinsights.com/insights/managing-floods-ai-for-adaptation-plans-and-disaster-relief/ [https://perma.cc/V2Z8-AAUU].

operational risk,[119] reduce energy consumption in 5G networks,[120] detect disease in plants, and monitor wildlife.[121] Autonomous robots even help first responders navigate in natural disasters by gathering data, traversing areas too dangerous for humans, identifying potential survivors, and providing real-time information to human personnel.[122]

On the environmental front, researchers are using geospatial AI to map cropland issues and predict susceptibility to landslides and air pollution.[123] The World Meteorological Organization, the International Telecommunication Union, and the United Nations Environment Programme (UNEP) formed the expert Focus Group on AI for Natural Disaster Management to create the framework for the U.S. for natural disaster management.[124]

Despite AI's potential to solve or mitigate the climate change crisis, AI has its own environmental footprint, and some estimate that the power

[119] *Discovery Channel,* AIFORGOOD.ITU.INT, (accessed Jan. 8, 2023), available at: https://aiforgood.itu.int/about-ai-for-good/discovery/.

[120] *AI-Driven Solutions for Climate Disasters provided by Zindi and ITU at AI for Good*, AI FOR GOOD (Sept. 27, 2023), https://aiforgood.itu.int/ai-driven-solutions-for-climate-disasters-provided-by-zindi-and-itu-at-ai-for-good/.

[121] *Powering Positive Change: Winning solutions form the 2023 tinyML Challenge Finale*, AI FOR GOOD , (Dec. 13, 2023), https://aiforgood.itu.int/powering-positive-change-winning-solutions-from-the-2023-tinyml-challenge-finale/.

[122] *Autonomous robots for disaster management*, AI FOR GOOD (Dec. 21, 2023), https://aiforgood.itu.int/autonomous-robots-for-disaster-management/.

[123] *AI for Good Perspectives: Harnessing AI to manage climate risk*, AI FOR GOOD, https://aiforgood.itu.int/event/harnessing-ai-to-manage-climate-risk/.

[124] A*I-Driven Solutions for Climate Disasters provided by Zindi and ITU at AI for Good*, AI FOR GOOD (Sept. 27, 2023), https://aiforgood.itu.int/ai-driven-solutions-for-climate-disasters-provided-by-zindi-and-itu-at-ai-for-good/ [https://perma.cc/5W3V-28ZC].

needed will consume 3.5% of the world's electricity.[125] While AI can be used for the development of disaster and climate hazard warning systems, its computing power creates a negative climate impact.[126] This new role comes at a cost.[127] AI tools powered by GPUs require a huge amount of energy and an effective cooling system,[128] which in turn requires a significant amount of water.[129] Unless scientists determine alternative methods, large scale adoption of AI will continue to lead to a huge increase in electricity consumption.[130] Additionally, electricity comes from fossil fuels predominantly in the U.S. and only 21% of the data centers in the U.S. are

[125] *Id*

[126] Kyungmee Kim et. al., *Artificial Intelligence for Climate Security: Possibilities and Challenges,* RELIEF WEB RELIEF(Dec. 8, 2023), https://reliefweb.int/report/world/artificial-intelligence-climate-security-possibilities-and-challenges#:~:text=AI%20can%2C%20for%20example%2C%20be,lead%20to%20insecurity%20and%20conflict [https://perma.cc/5CZB-3G4V]; Jude Coleman, *AI's Climate Impact Goes Beyond its Emissions*, SCIENTIFIC AM. (Dec. 7, 2023), https://www.scientificamerican.com/article/ais-climate-impact-goes-beyond-its-emissions/#:~:text=Training%20and%20running%20an%20AI,way%20AI%20affects%20the%20climate [https://perma.cc/Q5YP-GFP9].

[127] Natalaia Kotlowska-Wochna, *The Environmental Impact of AI, or the Climate Cost of Artificial Intelligence, KOCHANSKI,* (Dec. 15, 2023), https://www.kochanski.pl/en/the-environmental-impact-of-ai-or-the-climate-cost-of-artificial-intelligence/#:~:text=Although%20we%20associate%20AI%20mainly,for%20significant%20carbon%20dioxide%20emissions [https://perma.cc/64AW-2N7H].

[128] *Id.*

[129] *Id.*

[130] Wim Vanderbuawhede, *The Climate Cost of the AI Revolution*, RIPE LABS (May 12, 2023), https://labs.ripe.net/author/wim-vanderbauwhede/the-climate-cost-of-the-ai-revolution/ [https://perma.cc/LRG4-BJGR].

renewable.[131] Accordingly, technology giants have pledged to make their AI progress not lead to further climate degradation.[132]

AI and Healthcare

AI also holds great promise in the medical field. Consultants believe that AI can unlock one trillion dollars in unrealized value and improvement potential.[133] Through UN programs, autonomous mobile clinics "staffed" with AI doctors conduct diagnostic health screenings in underserved and hard to reach areas.[134] Long-term care facilities can now use socially-assistive robots to work with dementia patients on mental and emotional stimulation because they are designed, in part, to promote the release of the love or bonding hormone oxytocin.[135] Researchers at the National Institutes of Health have observed that AI can detect some cancers earlier and more

[131] *Id.*

[132] Oliver Milman, *Big tech vows to fight climate crisis but employs fossil fuel-linked lobbyists*, THE GUARDIAN (July 5, 2023), https://www.theguardian.com/us-news/2023/jul/05/big-tech-vows-to-fight-climate-crisis-but-employs-fossil-fuel-linked-lobbyists#:~:text=Apple%2C%20Google%2C%20Microsoft%20and%20Amazon,that%20are%20worsening%20global%20heating [https://perma.cc/G77E-XFSV].

[133] Shasank Bhasker et al*., Tackling healthcare's biggest burdens with generative AI*, MICKINSEY (July 10, 2023), available at: https://www.mckinsey.com/industries/healthcare/our-insights/tackling-healthcares-biggest-burdens-with-generative-ai [https://perma.cc/VG48-V3T6]..

[134] Elizaveta Argurbash, *Cognitively assisting robots for dementia care*, AI FOR GOOD, (Dec. 11, 2023), https://aiforgood.itu.int/cognitively-assistive-robots-for-dementia-care/ [https://perma.cc/2FAA-4Q34];/; *AI Clinics on Mobile (AICOM): Universal AI Doctors for the "Underserved" and "Hard-to-Reach,"* AI FOR GOOD (July 5, 2023), https://aiforgood.itu.int/ai-clinics-on-mobile-aicom-universal-ai-doctors-for-the-underserved-and-hard-to-reach/ [https://perma.cc/JC4X-HECH].

[135] *FTC Report Warns About Using Artificial Intelligence to Combat Online*, F.T.C. (June 16, 2022), https://www.ftc.gov/news-events/news/press-releases/2022/06/ftc-report-warns-about-using-artificial-intelligence-combat-online-problems [https://perma.cc/9KCW-WAGA].

accurately than clinicians.[136] Although up-front costs may be high for AI in some health practices, automating certain practices such as documenting patient visits, summarizing charts, and requesting insurance authorizations could cut healthcare costs by 25% and significantly reduce burnout for medical professionals.[137]

Using a hierarchical deep learning model, scientists are using AI to review patient transcripts using anonymous tagging to find the specific elements between the therapist-patient exchange that were effective and ineffective to improve overall accuracy.[138] This data will be used to determine the most efficacious treatment methods for the one hundred million people worldwide dealing with mental illness.[139] Nonetheless, experts note that clinicians must proceed with caution, particularly in the mental health space. While chatbots may help mitigate the effects of a shortage of mental health providers, can reach people who may not have access to or feel comfortable

[136] Bo Zhang et al., *Machine Learning and AI Cancer Prognosis, Prediction, and Treatment Selection: A Critical Approach*,16 J. OF MULTIDISCIPLINARY HEALTHCARE 1779, 1791 (June 26, 2023).

[137] Betha Coombs, *How hospitals are using A.I. to fight doctor burnout*, CNBC (Aug. 7, 2023), https://www.cnbc.com/2023/08/07/hospitals-use-ai-like-microsoft-nuances-dax-app-to-fight-burnout.html#:~:text=Hospitals%20like%20Baptist%20Health%20in,transcribe%20and%20document%20patient%20visits [https://perma.cc/54SA-5Z6J].

[138] Erin Kalejs, *How can AI Improve Mental Health for 100 Million People?*, AI FOR GOOD (June 22, 2022), https://aiforgood.itu.int/how-can-ai-improve-mental-health-for-100-million-people/ [https://perma.cc/2J73-VJKW].

[139] *Id.*

with a therapist, and can automate time consuming notetaking and other mundane tasks, they still pose serious privacy and bias risks.[140]

AI and Education

Additionally, AI has the potential to revolutionize education. Though many teachers and professors worry that students will cheat with AI and lose critical thinking skills, AI can increase teacher job satisfaction by automating mundane tasks and providing feedback loops[141] so they can focus on educating students.[142] With AI, educators are now providing tailored lesson plans to students with different language capabilities and learning styles.[143] Sal Khan, CEO of Khan Academy, has partnered with OpenAI to develop Khanmigo, which serves as a patient 1:1 tutor for students and as a teaching aide for instructors.[144]

More teachers are adopting AI in and out of the classroom. In a March 2023 survey, 51% of teachers surveyed indicated that they used

[140] Zara Abrams, *AI is changing every aspect of psychology. Here's what to watch for*, AM. PSYCH. ASS'N (July 1, 2023), https://www.apa.org/monitor/2023/07/psychology-embracing-ai [https://perma.cc/3EKK-ZNHU].

[141] Jiny Liu et al., *AI Can Make Education More Personal (Yes, Really)*, EDUC. WEEK (Aug. 14, 2023), https://www.edweek.org/leadership/opinion-ai-can-make-education-more-personal-yes-really/2023/08 [https://perma.cc/N77Y-QYE9].

[142] DEP'T OF EDUC. HANDOUT: AI AND THE FUTURE OF TEACHING AND LEARNING 2, https://tech.ed.gov/files/2023/05/ai-report-core-messaging-handout.pdf (last accessed Jan. 11, 2023).

[143] *Id.*

[144] *Supercharge Your Teaching Experience With Khanmigo*, KHAN ACADEMY, https://www.khanmigo.ai/ (last accessed Feb. 17, 2024).

ChatGPT, with a 69% adoption rate among Black and Latino teachers.[145] 40% of teachers reported using ChatGPT on a weekly basis and 10% use it almost daily.[146] From the student perspective, a significant portion of the 12-17 age group, about 33%, reports using ChatGPT for school-related purposes, with usage peaking at 47% among the middle school subgroup.[147] 88% of teachers and 79% of students had overwhelmingly positive feedback. In addition, notwithstanding the potential privacy and bias issues outlined by the FTC[148] and others, the U.S. Department of Education Office of Technology believes that AI can democratize how and what children learn.[149]

In light of the diverse and impactful applications of AI, as outlined above, policymakers must enact regulation that maximizes AI's potential benefits while cautiously addressing its risks. The United Nations' AI for Good initiative and other similar programs exemplify the positive influence AI can have on global challenges, such as climate change, healthcare, and education. However, as policymakers consider the path forward, they must balance this optimism with a careful consideration of the ethical and societal implications of AI, including addressing concerns related to intellectual

[145] *Teachers and Students Embrace ChatGPT for Education,* THE WALTON FAMILY FOUNDATION (Mar. 1, 2023), https://www.waltonfamilyfoundation.org/learning/teachers-and-students-embrace-chatgpt-for-education.

[146] *Id.*

[147] *Id.*

[148] F.T.C. , *supra* note 135.

[149] DEP'T OF EDUC., *supra* note 142.

property, privacy, bias, and the potential displacement of human roles in certain sectors, which we discuss below.

The Bad

Copyright Infringement, the Future of Creativity, and the Ability of AI Companies to Access Data[150]

As discussed in Part I, LLMs require vast amounts of data from public sources. This leads to inevitable concerns about infringement on intellectual property rights and will likely require a complete rethinking of what constitutes protectable work. The objective of the intellectual property and copyright legal regimes is "to incentivize and maximize creativity, diversity, technological processes and freedom of expression" by protecting creators' rights.[151] The universal principles underlying intellectual property and copyright law emphasize natural, moral, and economic rights and consist of "human authorship, subject matter such as literary, artistic and scientific works, original expression, a minimum of creative choices and ownership by a legal subject."[152] Copyright is tied to human creation because a copyright

[150] A comprehensive discussion of AI and intellectual property is beyond the scope of this Article. We will focus only on copyright here although there are significant issues related to patents as well.

[151] Mauritz Kop, *AI & Intellectual Property: Towards an Articulated Public Domain*, 28 TEX. INTELL. PROP. L. J. 297, 301 (2020).

[152] *Id.* at 301–02.

grants the author an exclusive right to compensation for the reproduction of the work and the right of prohibition.[153]

The extent of a copyright can be limited in a number of ways. Limiting methods include by contract and through voluntary alternative licensing frameworks like open source, copyleft,[154] and creative commons.[155] But generative AI rocked the music and entertainment industries in Spring 2023 when AI-generated songs featuring generated voices of Drake and The Weeknd,[156] Eminem,[157] as well as Rihanna and Bad Bunny[158] hit the internet. These works bring authorship, ownership, and protectability into question, and bring AI's role in copyright infringement into the limelight.[159]

The Copyright Act protects "original works of authorship fixed in any tangible medium of expression, now known or later developed, from

[153] *Id.* at 301.

[154] *What is Copyleft?*, GNU, https://www.gnu.org/licenses/copyleft.en.html [https://perma.cc/7H4W-SS3A] (last visited Jan. 10, 2024).https://www.gnu.org/licenses/copyleft.en.html.

[155] Kop, *supra* note 151 at 302.

[156] Joe Coscarelli, *An A.I. Hit of Fake 'Drake' and 'The Weeknd' Rattles the Music World,* N.Y. TIMES (Apr. 19, 2023), https://www.nytimes.com/2023/04/19/arts/music/ai-drake-the-weeknd-fake.html [https://perma.cc/SR55-PHWN].

[157] Thania Garcia, *David Guetta Replicated Eminem's Voice in a Song Using Artificial Intelligence*, VARIETY, (Feb. 8, 2023), https://variety.com/2023/music/news/david-guetta-eminem-artificial-intelligence-1235516924/ [https://perma.cc/BG4Y-U335].

[158] Gil Kaufman, *Bad Bunny and Rihanna AI Duet Drops – and the Fake Drake/ The Weeknd Creator Seems to Be Behind It*, BILLBOARD (Apr. 26, 2023), https://www.billboard.com/music/music-news/bad-bunny-rihanna-ai-song-fake-drake-weeknd-creator-1235315841/ [https://perma.cc/CCA6-2TSC].

[159] Gil Appel et al., *Generative AI Has an Intellectual Fillers*, HARV. BUS. REV. (Apr. 7, 2023), https://hbr.org/2023/04/generative-ai-has-an-intellectual-property-problem [https://perma.cc/EK4L-KDPV].

which they can be perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or device."[160] In determining the protectability of a work, the United States Patent and Trademark Office (USPTO) considers the author of the work, the independent creation of the work, the creative nature of the work, and the originality of the work.[161] The term "original," as used in copyright, means that the work is the independent creation of the author, and not a copy of another work.[162] As discussed above, AI models produce output based on the data they were trained on. This can complicate the "originality" aspect of protectability if the model is significantly borrowing from the works of others in its training set. The USPTO's creativity requirement carries a relatively low burden.[163] However, "[n]o AI is itself the wellspring of creativity. Rather, the creativity the AI displays flows either from the algorithm used to design and train it, or from the instructions provided by the users operating it."[164]

The authorship determination is more complicated. The copyright regimes protect "the fruits of intellectual labor."[165] Works protected by the

[160] U.S. COPYRIGHT OFF., COMPENDIUM OF U.S. COPYRIGHT OFF. PRACS. § 302 (3d ed. 2021), https://www.copyright.gov/comp3/chap300/ch300-copyrightable-authorship.pdf [https://perma.cc/4T2B-2V5T]; 17 U.S.C. § 102(a).
[161] U.S. COPYRIGHT OFF., *supra* note 160, § 302.
[162] Feist Publications, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 345 (1991).
[163] *Id.* at 358.
[164] Daryl Lim, *AI & IP Innovation & Creativity in an Age of Accelerated Change*, 52 AKRON L. REV. 813, 842 (2018).
[165] See *Trade-Mark Cases*, 100 U.S. 82, 94 (1879).

Copyright Act are "'original works of authorship.'"[166] To qualify, the work must have been created by a human:[167, 168] "[T]he Office [USPTO] will refuse to register a claim if it determines that a human being did not create the work."[169] AI generated works may not meet this standard if the copyright office cannot determine human authorship.[170] When creators use AI to supplement their work and creative process, the determination of authorship becomes more difficult.[171] The critical question to determine if the work is eligible for copyright protection is whether the work is a human creation that utilized technological assistance, or if the traditional, creative elements of authorship are attributable to the machine.[172] "With copyright law, however, authors need neither understand nor explain how the tools they use—cameras, computers, or AI—render their works of authorship. The touchstone is instead controlled."[173]

As previously discussed, for AI models to function, they must be trained on expansive amounts of data to be accurate and effective. To learn the patterns and predict outcomes, the AI has to read and consume the

[166] 17 U.S.C. § 102(a).
[167] U.S. COPYRIGHT OFF., *supra* note 160, § 313.2.
[168] *See* Burrow-Giles Lithographic Co. v. Sarony, 111 U.S. 53, 58 (1884).
[169] U.S. COPYRIGHT OFF., *supra* note 160, § 306. See *Burrow-Giles Lithographic Co.*, 111 U.S. at 58.
[170] See *Thaler v. Perlmutter*, 1:22-cv-01564-BAH, 1 (2023).
[171] Appel et al., *supra* note 159.
[172] U.S. COPYRIGHT OFF., *supra* note 160, § 313.2.
[173] Lim, *supra* note 164, at 843.

data.[174] Discriminative AI, such as facial recognition AI, does not depend on the artistic choices reflected in photographs, but focuses on matching facts and classifying data.[175] This kind of data use does not disrupt the market of expressive works.[176] Conversely, in the case of some generative AIs, that data consists of creative works.[177]

A generative adversarial network ("GAN") is a type of AI model used for creating new data that resembles a given set of data. It involves two parts: one part makes new data (such as images or music), and the other part evaluates it, deciding if it seems real or fake. They work against each other—hence "adversarial"—improving the quality of the generated data over time, aiming to make it increasingly difficult to distinguish from the original data.[178] A GAN was recently trained on 15,000 famous portraits.[179] In a similar vein, sophisticated AI platforms including Flow Machine, IBM Watson Beat, Google Magenta, and Spotify's Creator Technology Research Lab, are designed to digest a variety of music, from recent popular songs to classical pieces.[180] These platforms analyze key musical features like harmony, rhythm,

[174] *Id.* at 847.
[175] *Id.* at 850.
[176] *Id.* at 850.
[177] *Id.* at 847.
[178] *What is Gan?*, AMAZON WEB SERV., https://aws.amazon.com/what-is/gan/ [https://perma.cc/6GUB-BWZU] (last visited Feb. 14, 2023).
[179] Lim, *supra* note 164, at 847.
[180] *Id.*

and structure, identifying trends within this data.[181] Armed with this knowledge, they then generate their own unique musical compositions.[182]

The use of these potentially protected works creates an issue where the models may commit copyright infringement unless the use is agreed upon or allowed by fair use.[183] The doctrine of fair use permits the use of copyrighted material without the consent of the rights holder in certain contexts. These include criticism (encompassing satire), commentary, news reporting, educational purposes (such as creating multiple copies for classroom instruction), scholarship, or research.[184] Additionally, this doctrine extends to transformative uses of copyrighted content, specifically in ways not originally intended by the copyright holder.[185] Businesses or individuals that willfully use training data based on unlicensed work or work not covered by fair use could face damages of up to $150,000 for each instance of knowing use.[186]

The use of AI in the arts has thus led to several high-profile lawsuits. In late 2022, three artists joined together to sue several generative AI platforms for using their original works without authorization to train the AI

[181] *Id.*
[182] *Id.*
[183] *Id.*
[184] Appel et al, *supra* note 159.
[185] *Id.*
[186] *Id.*

on their styles and for allowing users to generate images with the program.[187] The plaintiffs alleged that the copyright violations occurred through Stability AI's program's use of an image generation procedure, 'Stable Diffusion.'[188] The plaintiff's complaint asserts that the output image produced by the model is a derivative work of the copyright-protected works originally produced by the plaintiffs "because it is generated exclusively from a combination of the conditioning data and the latent images, all of which are copies of copyrighted images."[189] If the court finds in favor of the plaintiffs, the defendants could face substantial infringement penalties.[190]

Getty Images also filed a lawsuit in February 2023 against Stability AI for unauthorized use of over 12,000 of its images to train the model, infringing on its copyright.[191] And, although OpenAI has already reached licensing agreements with several other news organizations, such as the Associated Press,[192] on December 27, 2023, the *New York Times* sued OpenAI

[187] *Id.*; Complaint at 1–6, Anderson v. Stability AI, et al., (N.D. Cal. 2023) (No. 3:23-cv-00201).
[188] Gabriel Karger, *AI-Generated Images: The First Lawsuit*, 24 COLUM. SCI. & TECH. L. REV. 1, 1 (Jan. 25, 2023), https://journals.library.columbia.edu/index.php/stlr/blog/view/479 [https://perma.cc/42AC-Y2HK].
[189] *Id.*
[190] Appel et al., *supra* note 159.
[191] *Id.*; Getty Images (US), Inc. v. Stability AI, Inc., (D. Del. 2023) (No. 1:23-cv-00135-UNA); Blake Betty, *Getty Images lawsuit says Stability AI misused photos to train AI*, REUTERS (July 13, 2023), https://www.reuters.com/legal/getty-images-lawsuit-says-stability-ai-misused-photos-train-ai-2023-02-06/.
[192] Matt O'Brien, *ChatGPT-maker OpenAI signs deal with Ap to license news stories*, APNEWS.COM, (July 13, 2023), https://apnews.com/article/openai-chatgpt-associated-press-ap-f86f84c5bcc2f3b98074b38521f5f75a [https://perma.cc/3CMS-KTEF].

and Microsoft (OpenAI's largest investor) for copyright infringement, arguing that "defendants should be held responsible for billions of dollars in statutory and actual damages related to the unlawful copying and use of the *Times*'s uniquely valuable works."[193] OpenAI has claimed that it was in negotiation with the *Times* up to a week before the suit was filed; the suit has no merit; any use of data is allowed under fair use; and the *Times* "manipulated" the data to produce the results it wanted for the lawsuit, going so far as to accuse the newspaper of "not telling the full story."[194]

Although there are entire articles devoted to proposing regulation for AI, and copyright and the U.S. Copyright Office has sought comments, we will not focus on proposing copyright rules here. Nonetheless, we do believe that there are some novel proposals that regulators should consider because the future of generative AI depends on access to training data. For example, Katherine Lee, A. Feder Cooper, and James Grimmelman have proposed a generative AI supply chain, a framework that breaks down the process from data collection to the creation of AI-generated outputs.[195] The initial stage of

[193] Michael M. Grynbaum, *The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work*, N.Y. TIMES (Dec. 27, 2023), https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html?smid=nytcore-ios-share&referringSource=articleShare.

[194] *Id.*; *OpenAI and journalism*, OPEN AI, (Jan. 8, 2024), https://openai.com/blog/openai-and-journalism [https://perma.cc/6DX5-PZ42].

[195] Katherine Lee et al., *Talkin' 'Bout AI Generation: Copyright and the Generative-AI Supply Chain*, J. COPYRIGHT SOC'Y, (forthcoming 2024) (available at: https://ssrn.com/abstract=4523551).

this supply chain involves the collection and selection of training data. The type and source of this data are critical as they set the foundation for what the AI can generate. For instance, an AI trained on a dataset of classical paintings will produce vastly different outputs compared to one trained on modern digital art. This stage raises important copyright questions, such as whether the use of certain data for training infringes on the original creators' rights. The model training phase is the next level, where the AI learns from the data, developing the capability to create new content. The nature of this content heavily depends on the training it receives, which in turn is influenced by the choices made in the data collection phase. The final output, although novel, may bear similarities to the training data, posing challenges in determining the originality and potential copyright infringement.

Of note, some venture capitalists raise concerns about overregulation in this area. For example, Marc Andreessen, an influential tech investor who has financed or invested in AirBnB, Facebook, and Twitter, argued in his comment to the Copyright Office that using copyrighted content for AI model training should be considered fair use, citing the impracticality and potential economic impact of alternative licensing models. He stressed that restricting AI development could harm U.S. competitiveness and stifle innovation, particularly for startups.[196] This is an important consideration that goes beyond the intellectual property arena and is a concern that others have raised regarding regulating AI in general.[197]

For context, imagine the fictional case of CreativeAI, a hypothetical AI company that has developed a model capable of generating digital artwork. This model, trained on a diverse array of images, including the copyrighted works of (fictional) renowned artist Narine Weldon, generates art pieces that closely mirror Weldon's style. This scenario, echoing recent high-profile legal disputes, spotlights the critical issue of whether AI-generated content that reflects the essence of human-created works constitutes an infringement of copyright. It brings into focus the fundamental questions surrounding the application of the fair use doctrine, the concept of transformative use, and the impact of AI on the market rights of original creators. Such cases challenge regulators to reevaluate the boundaries of copyright in an era where technology significantly contributes to the creative process. This hypothetical scenario dovetails with the broader implications of AI's role in intellectual property rights. The debate around

[196] *Comment from Andreessen Horwitz*, REGULATIONS.GOV, (Nov. 1, 2023), https://www.regulations.gov/comment/COLC-2023-0006-9057 [https://perma.cc/S5NR-B9ZN].

[197] Peter Henderson et al., *Foundation Models and Fair Use*, INST. HUMAN-CENTERED A.I. STAN. U. (Mar. 28, 2023), https://hai.stanford.edu/sites/default/files/2023-11/Foundation-Models-Copyright.pdf [https://perma.cc/KS8K-ZKCV].

CreativeAI's use of Weldon's work illustrates the tension between encouraging innovation and safeguarding the rights of creators.

The intersection of generative AI and intellectual property law, exemplified by the hypothetical case of CreativeAI, gets more complicated when considering recent real-world lawsuits filed by news agencies against AI companies. These lawsuits allege copyright infringement due to the unauthorized use of copyrighted material in training generative AI models. Such legal actions not only challenge the notion of authorship and originality, as seen in the CreativeAI scenario, but also raise a critical concern regarding the accessibility of training data. In the case of CreativeAI, the AI's ability to produce artwork reminiscent of Narine Weldon's style hinges on its access to a broad range of training data, including Weldon's works.

With news agencies pushing back against the use of their content without permission, there's a looming threat that could restrict the flow of information essential for training AI models. This could significantly impede the functionality and development of generative AI technologies. The lawsuits from news agencies and artists exemplify a broader trend where content creators are increasingly asserting their rights, potentially leading to a constrained environment for sourcing AI training data. The outcome of these legal battles could set precedents that restrict the types of data available for AI development, thereby impacting the scope and utility of generative AI models.

Job Displacement[198]

During the First Industrial Revolution in the mid-18th century, steam power, the use of coal, and mechanization transformed societies in Europe and the United States from artisanal, agrarian societies to ones in which people left the countryside to live and work in cities with factories. Middle- and upper-class citizens grew richer, but pollution, poor sanitation, disease, and overcrowding caused significant suffering for the majority of workers.[199] The concept of the division of labor was born during this period.

The use of steel, chemicals, electricity, the telegraph, and the automobile came with the Second Industrial Revolution, which also heralded the era of mass production.[200] Ford Motors, US Steel, General Electric, and

[198] *See generally* THE WHITE HOUSE, THE IMPACT OF A.I. ON THE FUTURE OF WORKFORCES IN THE EUR. UNION AND THE U.S. (2022), [https://perma.cc/G5A9-Q7EY]. We do not rely on this 2022 report because while it does focus on Chat GPT3, the exponential jump in capabilities of ChatCPT3.5 and 4 make its conclusions less meaningful, although still helpful. It did however, lead the AI Bill of Rights and highlighted issues of algorithmic bias, which will be addressed in the next section.

[199] *Industrial Revolution Key Facts*, BRITANNICA, https://www.britannica.com/summary/Industrial-Revolution-Key-Facts, [https://perma.cc/5HC4-38GN] (last visited Jan. 11, 2023); David Montgomery, *Chapter 3: Labor in the Industrial Era By David Montgomery*, U.S. DEP'T. OF LAB., [https://perma.cc/W3PM-PMWC] (last visited Jan. 11, 2024).

[200] Eric Niler, *How the Second Industrial Revolution Changes Americans' Lives*, HISTORY (July 25, 2023), https://perma.cc/L4DM-NSMG; Joel Mokyr, *The Second Industrial Revolution, 1870-1914*, NW. UNIV. (2003), https://perma.cc/U3AS-HAXW.

Coca-Cola are companies formed in that era that survive today.[201] The Third Industrial Revolution in the second half of the twentieth century ushered in the age of nuclear energy, renewable energy, mass communications, widespread use of computers, the dawn of the Internet, the Internet of Things (IoT), and the "digitization of everything."[202]

We are now in the Fourth Industrial Revolution, sometimes called Industry 4.0, which is marked by an increase in automation, digitization, augmented reality, blockchain use, cloud computing, wearable technology, and, of course, the rapid rise of AI capabilities.[203] Perhaps the best summary is that "Industry 4.0 leverages high volumes of data with advanced tools to improve business functions across the enterprise, from predicting failures and prescribing fixes before they occur, to helping generate new innovations and revenue streams."[204]

While many worry that AI will lead to the end of humanity, the more immediate and realistic concern for average citizens is the potential for

[201] *The Second Industrial Revolution: The Technological Revolution,* RICH. VALE ACAD. (May 16, 2022), https://richmondvale.org/second-industrial-revolution/ [https://perma.cc/RFL3-NZYG].

[202]Jeremy Rifkin, *Welcome to the Third Industrial Revolution*, WHARTON MAG., (Summer 2018), [https://perma.cc/5YGK-TVLH]; Jeremy Rifkin, *A New Economic Narrative: Industrial Revolution 3.0*, OUR WORLD (Mar. 7, 2012), https://ourworld.unu.edu/en/a-new-economic-narrative-industrial-revolution-3-0 [https://perma.cc/DH3E-RHQV].

[203] *What are Industry 4.0, the Fourth Industrial Revolution, and 4IR*, MCKINSEY & CO. (Aug. 17, 2022), https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-generative-ai.

[204]*Industry 4.0, the Fourth Revolution, Challenges, Benefits, Adoption and How to Begin*, FROST & SULLIVAN (2019), https://www.ibm.com/downloads/cas/QRKNQ4A8.

significant job displacement of the white-collar workers who had previously been spared during the Third Industrial Revolution, where blue-collar workers in factories fought for their wages and jobs as they worked side by side with robots or were replaced by machines on assembly lines. From 1990-2007, each robot used on the assembly line replaced 3.3 workers on average.[205] Now, the tables are turning, which could have a significant impact on higher education and the job market.

U.S. researchers have created a list of jobs most exposed to AI using the AI Occupational Exposure (AIOE) measure, a tool instrumental in understanding the impact of AI advancements on the workforce, particularly in language modeling (such as ChatGPT).[206] The AIOE provides a comprehensive analysis by mapping ten AI applications, such as playing abstract strategy games, recognizing images, and language modeling to a range of fifty-two human abilities, ranging from oral comprehension to inductive reasoning. The data driving this analysis comes from the Electronic Frontier Foundation (EFF), which monitors AI progress across various applications, and the Occupational Information Network (O*NET), a

[205] Peter Dizikes, *How many jobs do robots really replace,* MIT NEWS (May 4, 2020), https://news.mit.edu/2020/how-many-jobs-robots-replace-0504 [https://perma.cc/47XU-MFSK].

[206] Edward Felten et al., *How Will Language Modelers like ChatGPT Affect Occupations and Industries?*, U. PENN. WHARTON (Mar. 18, 2023), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4375268 [https://perma.cc/BM6J-Z7TA].

detailed repository of job-related skills and competencies from the U.S. Department of Labor.[207]

A comparative analysis of the original AIOE and its language modeling-focused version reveals both similarities and differences in industry exposure to AI. For instance, industries such as "Securities, Commodity Contracts, and Other Financial Investments and Related Activities" consistently appear as highly exposed in both versions. Similarly, legal services, insurance, accounting, tax preparation, and employee benefit funds remain among the top sectors affected. However, the language modeling-focused AIOE reveals a distinct trend: greater exposure within higher education and related industries. This includes junior colleges, grantmaking and giving services, and business schools, which are all featured in the top twenty most exposed industries. Most alarmingly, the researchers found a positive and statistically significant correlation between the average or median wage of an occupation and its level of exposure to AI language modeling.

The UK Department of Education used the same methodology and found that fields associated with clerical work and across finance, insurance, law business management, management consulting, accountants,

[207] *Id.*

psychologists, and teachers were particularly impacted.[208] The professional occupations least exposed to AI include veterinarians, medical radiographers, dental practitioners, physiotherapists, and senior police officers. Although neither the original nor the UK study made conclusions about job replacement, the information is still instructive for governments, educators, and business leaders.

There is no shortage of experts predicting how AI will lead to job displacement. Vice Chair of recruiting behemoth Korn Ferry indicated that approximately 19% of workers could be displaced, and those jobs requiring cognitive instead of manual skills would be most at risk in the short term.[209] The World Economic Forum's Future of Jobs Report looks at micro and macro trends from 2023-2027 and covers 803 companies—collectively employing more than 11.3 million workers—across 27 industry clusters and 45 economies from across the globe.[210] Its April 2023 report found that 75% of surveyed companies will adopt or have adopted AI, and 50% expect it to

[208] DEPT. OF EDUC. (U.K.), THE IMPACT OF AI ON UK JOBS AND TRAINING 5, https://assets.publishing.service.gov.uk/media/656856b8cc1ec500138eef49/Gov.UK_Impact_of_AI_on_UK_Jobs_and_Training.pdf [https://perma.cc/D8Y6-83PH] (last visited Mar. 9, 2024).

[209] *A.I. will challenge jobs 'where cognitive takas are value', says Korn Ferry's Alan Guarino*, CNBC (Oct. 17, 2023), available at: https://www.cnbc.com/video/2023/10/17/a-i-will-challenge-jobs-where-cognitive-tasks-are-valued-says-korn-ferrys-alan-guarino.html [https://perma.cc/2DR5-R66V].

[210] *The Future of Jobs Report 2023*, WEFORUM, (Apr. 20, 2023), https://www.weforum.org/publications/the-future-of-jobs-report-2023/digest/ [https://perma.cc/X6KE-MDC9].

create job growth, while 25% anticipate job losses, and 60% believe that workers will need reskilling.[211] Ironically, while many believe that the increased use of AI will lead to a decline in cognitive and creative thinking skills, those are the top two skills that employers will be seeking in the future. Technical literacy is in third place.[212] In a survey of more than 750 businesses published in December 2023, approximately 37% of employers indicated that they have replaced workers with AI in 2023, and 44% expect to lay employees off in 2024 because of efficiency gains.[213]

Regulators face a pressing need to address potential unemployment or underemployment, and here, we pose several key questions and considerations for policymakers. How should countries and companies develop and implement effective reskilling programs to equip the workforce for new roles in an AI-dominated landscape? What strategies can they adopt to ensure that AI integration in industries does not lead to significant job losses but rather complements human labor? In cases where AI displacement is inevitable, what forms of support and transition assistance can they provide to those who lose their jobs? How can they shape

[211] *Id.*

[212] Rachel Curry, *Recent data shows AI job losses are rising, but the numbers don't tell the full story*, CNBC (Dec. 16, 2023), https://www.cnbc.com/2023/12/16/ai-job-losses-are-rising-but-the-numbers-dont-tell-the-full-story.html#:~:text=More%20than%20one%2Dthird%20(37,Asana%20found%20in%20its%20surveying [https://perma.cc/7JTF-YJ2R].

[213] *Id.*

regulations to encourage the responsible use of AI, ensuring it boosts productivity without harming the job market? What long-term plans and policies do they need to anticipate and prepare for future changes in the job market due to ongoing technological advancements?

Access to Justice

One of the most promising use cases for AI relates to access to justice. 92% of low-income individuals in the U.S. currently have inadequate or unmet civil legal needs.[214] AI may help close the access to justice gap for marginalized communities who need help filing simple matters and understanding how legal proceedings work.[215] A chatbot can provide digestible, basic information to address legal problems throughout the world.[216] DoNotPay, for example, markets itself as a chatbot to "help you fight big corporations, protect your privacy, find hidden money, and beat bureaucracy."[217]

[214] Kristina Sonday, *Forum: There's potential for AI chatbots to increase access to justice,* THOMSON REUTERS, (May 25, 2023), https://www.thomsonreuters.com/en-us/posts/legal/forum-spring-2023-ai-chatbots/ [https://perma.cc/C7G9-JMPA].

[215] Hassan Kanu, *Artificial intelligence poised to hinder, not help, access to justice,* REUTERS (Apr. 25, 2023) https://www.reuters.com/legal/transactional/artificial-intelligence-poised-hinder-not-help-access-justice-2023-04-25/#:~:text=%22Fundamentally%2C%20problems%20of%20access%20to,categorically%20favor%20powerful%2C%20wealthy%20actors.

[216] Sonday, *supra* note 214.

[217] DONOTPAY, https://donotpay.com (last visited Jan. 11, 2023).

However, some legal chatbots or even the use of GAI for legal advice by laypeople can lead to incorrect, outdated, or misleading information.[218] Additionally, the data the algorithms rely on have biases embedded in their data sets.[219] While AI increases access to justice, who (or what) would be liable for AI's inaccurate information if a layperson relied on advice from a chatbot?[220] If a lawyer with over thirty years of experience relied on ChatGPT without knowing that the cases it produced were fabricated,[221] how could a regular citizen know whether to trust a legal chatbot? That lawyer paid a fine of $5,000 and became the butt of jokes and a cautionary tale.[222] And he is not alone. Michael Cohen, former lawyer for Donald Trump, who himself is facing legal jeopardy, provided cases generated by Google Bard without verifying them to his own lawyer, who also failed to verify them before filing a motion in federal court.[223]

[218] Brennan Donnellan, *Informational Session on Cleveland Legal Collaborative January 8th*, CLEV. METRO. BAR ASS'N (Jan. 3, 2023), https://www.clemetrobar.org/?pg=CMBABlog&blAction=showEntry&blogEntry=101280 [https://perma.cc/9K2V-NQUA].

[219] Heidi Wudrick & Robert Kwei, *Will AI revolutionize the legal profession? The jury is still out*, BEYOND (May 19, 2023), https://beyond.ubc.ca/will-ai-revolutionize-the-legal-profession-the-jury-is-still-out/ [https://perma.cc/5YSP-NDY9].

[220]Matthew Stepka, *Law Bots: How AI Is Reshaping the Legal Profession*, ABA (Feb. 21, 2022), https://businesslawtoday.org/2022/02/how-ai-is-reshaping-legal-profession/ [https://perma.cc/LB5F-PT4L].

[221] Ramishah Maruf, *Lawyer apologizes for fake court citations from ChatGPT*, CNN (May 28, 2023), https://www.cnn.com/2023/05/27/business/chat-gpt-avianca-mata-lawyers/index.html [https://perma.cc/32VT-TS3L].

[222] Mata v. Avianca, Inc., (No. 1:22-cv-01461-PKC), Document 54, at 34 (S.D.N.Y. 2023).

[223] Nicki Brown, *Michael Cohen says he unwittingly sent attorney non-existent case citations generated by AI*, CNN (Dec. 29, 2023), https://www.cnn.com/2023/12/29/politics/michael-cohen-attorney-generated-by-ai/index.html [https://perma.cc/8B8C-BR5Y].

The average citizen does not know how to conduct legal research or may not understand the legal jargon in cases and statutes. The consequences could be far more serious if an untrained citizen used a legal GPT or chatbot to advise her on a criminal, housing, trust and estates, family law, or employment matter. However, once the data sets improve in accuracy and trustworthiness, there may come a day when indigent criminal defendants have the right to AI assistance as counsel and when ordinary citizens have an AI lawyer in their pocket.[224]

The integration of AI in consumer-facing legal chatbots requires a comprehensive understanding of the ethical and practical implications. Ethically, the primary concern lies in ensuring confidentiality and upholding the attorney-client privilege through robust mechanisms within AI systems to safeguard sensitive information. Practically speaking, the dynamic nature of legal frameworks demands that AI systems be regularly updated to maintain accuracy and relevance, posing a significant operational challenge. In most states, new regulations come into effect in either January or July of a given year. Ensuring regulatory compliance of AI legal assistants across diverse jurisdictions presents an even more formidable challenge, given the varying legal landscapes.

[224] Stepka, *supra* note 220.

Further, the application of AI in complex legal issues underscores its current limitations in nuanced understanding and judgment. This requires clear communication regarding the capabilities and limitations of AI in legal contexts, particularly in cases that require deep legal acumen and discretion. Many lay people may not understand or appreciate the nuances, especially if they are explained by the same chatbot they are relying on. Moreover, the reliance on AI for legal assistance risks widening the digital divide, as those with better access to technology may disproportionately benefit from these advancements.

Addressing these issues requires a balanced approach, considering the innovative potential of AI with ethical standards and regulatory compliance. As U.S. Supreme Court Chief Justice John Roberts explained in the Court's 2023 Year-End Report, "Law professors report with both awe and angst that AI apparently can earn Bs on law school assignments and even pass the bar exam. Legal research may soon be unimaginable without it. AI obviously has great potential to dramatically increase access to key information for lawyers and non-lawyers alike. But just as obviously, it risks invading privacy interests and dehumanizing the law… These tools have the welcome potential to smooth out any mismatch between available resources and urgent needs in our court system. But any use of AI requires caution and humility."[225]

The intersection of AI with copyright law and its impact on job displacement and legal services demands a proactive approach from regulators. The legal challenges posed by AI-generated content, particularly in music and art, highlight the need for an evolution in copyright law, which is already underway. This evolution should not only clarify the status of AI-generated works but also reconsider the definition of creativity and authorship in the digital age. This may require recognizing AI as a tool in the creative process or introducing new categories of copyright specifically tailored to AI creations. Additionally, expanding the concept of fair use to explicitly include the use of copyrighted material for AI training purposes may be essential. Finally, regulators must strike a delicate balance between encouraging AI innovation and protecting individual and organizational rights. This requires collaborative efforts among AI developers, copyright holders, and policymakers. The aim should be to create frameworks that both incentivize technological advancement and respect the rights and interests of all stakeholders involved.

[225] John G. Roberts, Jr., *2023 Year-End Report on the Federal Judiciary* 5, SUPREME COURT OF THE U.S. (Dec. 31, 2023), https://www.supremecourt.gov/publicinfo/year-end/2023year-endreport.pdf[https://perma.cc/3PGF-UXQD].

The trend towards AI-induced job displacement, particularly in higher-wage, white-collar professions, underscores the urgency for strategies to manage this transition. Addressing this challenge calls for a concerted effort from policymakers and educational institutions to focus on reskilling and upskilling programs. These programs should emphasize areas where human workers excel (for now), such as emotional intelligence and complex problem-solving, thus complementing AI capabilities. Corporations, too, play a crucial role and should be encouraged to embrace a model of human-AI collaboration where AI tools enhance rather than replace human labor. Companies should encourage workers to use AI as a tool rather than a replacement.

In the realm of legal services, the integration of AI tools like chatbots has the potential to democratize access to justice but also raises legitimate concerns about accuracy, ethics, and the digital divide. To ensure the reliability and legal accuracy of these AI tools, a regulatory framework must mandate regular audits for bias and adherence to current laws. Additionally, ethical guidelines governing the use of AI in legal contexts, particularly concerning client confidentiality and attorney-client privilege, are imperative. Law students and legal professionals should be well-informed about the capabilities and limitations of AI tools to ensure their responsible use, should disclose the use of AI to generate legal filings and contracts, and should face stiff penalties and sanctions for failing to verify research results generated by GAI.

The evolving AI landscape requires legal and regulatory frameworks capable of addressing the unique challenges of protecting intellectual property, preparing the workforce for changes induced by AI, ensuring ethical use of AI in legal services, and fostering a balanced environment for AI innovation and protection. We now turn to "the ugly," which exemplifies the urgent need for cohesive, global regulation.

The Ugly

Businesses generally need four types of data: personal data, engagement data, behavioral data, and attitudinal data.[226] Personal data is identifying information about us such as biographical information and identifying numbers. Engagement data measures how consumers interact with the business and the world.[227] Behavioral data encompasses a wide range of actions companies collect data on, from consumers' transactional histories to mouse-movements.[228] Attitudinal data is information about customer

[226] Timothy R. Graeff & Susan Harmon, *Collecting And Using Personal Data: Consumers' Awareness And Concerns*, 19 J. OF CONSUMER MKTG. 302, 303 (2002).

[227] T. Tony Ke & K. Sudhir, *Privacy Rights and Data Security: GDPR and Personal Data Markets*, 69 J. OF MGMT. SCI. 43, 894389 (2022).

[228] Ángel Alexander Cabrera et al., *What Did My AI Learn? How Data Scientists Make Sense Of Model Behavior*, ACM TRANSACTIONS ON COMPUTER-HUMAN INTERACTION (Mar. 7, 2023), https://dl.acm.org/doi/abs/10.1145/3542921 [https://perma.cc/PZ65-N4NJ] (last visited Jan. 11, 2024).

satisfaction, product desirability, and other data concerning consumer's feelings.[229]

Companies collect consumer data by directly asking customers, indirectly tracking customers, and acquiring data through third parties,[230] as well as through other sources including loyalty card usage,[231] in-app behavior regarding game play,[232] satellite images,[233] employer databases,[234] email

[229] Evangelos Pournaras et al., *Collective Privacy Recovery: Data-Sharing Coordination Via Decentralized Artificial Intelligence*, ARXIV (2023), https://arxiv.org/abs/2301.05995 [https://perma.cc/4FVS-Q4Y5] (last visited Jan. 11, 2024).

[230] Forrester, *Third-Party Data Is Here To Stay*, FORBES (July 25, 2023, 8:27 AM), https://www.forbes.com/sites/forrester/2023/07/25/third-party-data-is-here-to-stay/) [https://perma.cc/X4E7-YD9C] (last visited Jan 11, 2024).

[231] *How do companies use my loyalty card data?*, BBC NEWS, https://www.bbc.com/news/technology-43483426 (last visited Jan. 11, 2024) [https://perma.cc/S66R-BYD4]; Rebecca Kowalewicz, *Council Post: The Role Of Loyalty Programs In Harnessing First-Party Data*, FORBES (June 13, 2023), https://www.forbes.com/sites/forbesagencycouncil/2023/06/13/the-role-of-loyalty-programs-in-harnessing-first-party-data/#:~:text=By%20asking%20customers%20to%20sign,and%20create%20personalized%20marketing%20campaigns. (last visited Jan. 11, 2024) [https://perma.cc/3ES3-X3CF].

[232] Jacob Leon Kröger, *Surveilling the Gamers: Privacy Impacts of the Video Game Industry Computing*, SCIENCE DIRECT, https://www.sciencedirect.com/science/article/abs/pii/S187595212200060X [https://perma.cc/MJN5-GGFR] (last visited Jan. 11, 2024).

[233] Sydney Shufelt, *Remote-Sensing Satellites and Privacy: Why Current Regulations Will Ultimately Fail*, AM. UNIV. BUS. L. REV. BLOG (2023), https://aublr.org/2020/03/remote-sensing-satellites-and-privacy-why-current-regulations-will-ultimately-fail/ [https://perma.cc/F56K-CGPH] (last visited Jan. 11, 2024); Rachel McAmis et al., *Over Fences and Into Yards: Privacy Threats and Concerns of Commercial Satellites*, available at: https://homes.cs.washington.edu/~rcmcamis/documents/SatellitesPrivacyCameraReady.pdf (last visited Jan. 11, 2024); Nicola Morini Bianzino et al., *Business Value Using Earth Observation And Satellite Data*, EY (2023), https://www.ey.com/en_gl/technology/how-can-the-vantage-of-space-give-you-strategic-advantage-on-earth [https://perma.cc/JX9M-ZVXW] (last visited Jan. 11, 2024).

[234] *Council Post: 11 Ways Businesses Can Use Customer Data The "Right" Way*, FORBES (June 1, 2023, 1:15 PM), https://www.forbes.com/sites/forbesbusinesscouncil/2023/06/01/11-ways-businesses-can-use-customer-data-the-right-way/ [https://perma.cc/6PK8-5FNE]); Michael Segalla & Dominique Rouziès, *The Ethics Of Managing People's Data*, HARV. BUS. REV., https://hbr.org/2023/07/the-ethics-of-managing-peoples-data (last visited Jan. 11, 2024).

inboxes, license plate images, cookies, heatmaps, GPS tracking, cell phone signal tracking, in-store WiFi activity, and facial recognition technology among many other methods. Some companies have even built their entire business models around using consumer data not only to tailor their own marketing, products, and services, but to have a significant, consistent revenue stream by selling the data they collect.[235] Data can also be used on an individual level to personalize a customer's experience.[236] Once a company has consumer data, it must process that data and increasingly, companies use AI and other machine learning models.[237]

Rise of Privacy and Cybersecurity Concerns with AI

It is nearly impossible to have a discussion about the collection and use of consumer data by companies without the accompanying conversation about data privacy. Data privacy, which is closely aligned with information privacy, is "the right to have control and knowledge about any personally identifiable information (PII) which is collected about an individual."[238] The National Institute of Science and Technology (NIST) defines PII as

[235] Max Freedman, *Businesses Are Collecting Data. How Are They Using It?*, BUS. NEWS DAILY, https://www.businessnewsdaily.com/10625-businesses-collecting-data.html html (last updated Oct. 20, 2023).

[236] *Id.*

[237] *Id.*

[238] Shaun G. Jamison, *Creating a National Data Privacy Law for the United States*, 10 CYBARIS 5 (2019) (available at: https://open.mitchellhamline.edu/cybaris/vol10/iss1/2) [https://perma.cc/XHB8-SN86].

"information which can be used to distinguish or trace the identity of an individual (e.g., name, social security number, biometric records, etc.) alone, or when combined with other personal or identifying information which is linked or linkable to a specific individual (e.g., date and place of birth, mother's maiden name, etc.)."[239] Privacy concerns the collection, storage, and dissemination of personal information, and is distinct from cybersecurity, which concerns protection from intrusion.[240] Cybersecurity may involve privacy protection, but not always.[241]

Privacy risks increase with AI compared to traditional software due to AI's enhanced data aggregation capability.[242] Under NIST's Artificial Intelligence Risk Management Framework, trustworthy AI must balance characteristics such as accountability, transparency, and privacy enhancement in order to be valid and reliable.[243] The past two decades have seen unprecedented innovation through the internet and associated technologies which is fueled by data about individuals that flow through intricate ecosystems.[244] Still, many individuals interacting with these systems do not

[239] *Id.*

[240] *Id.*

[241] *Id.*

[242] DEP'T OF COM., *supra* note 47, at 38.

[243] *Id.* at 12.

[244] NAT'L INST. OF STANDARDS & TECH., *Privacy Framework*, https://www.nist.gov/privacy-framework [https://perma.cc/L65U-5H8R] (last visited Mar. 10, 2024).

fully understand the implications.[245] AI systems rely on large amounts of personal data to learn and make predictions which raises concerns[246] regarding the source of the data and how the data is being used, the potential for data breaches, unwanted surveillance, unauthorized access to personal information, and unwelcomed AI profiling.[247] We discuss these issues later in Part II.

Cybersecurity risks may pose even greater concerns. Cyberattacks involve unauthorized access to or manipulation of computer systems, networks, or data, often with malicious intent.[248] Experts estimate that cybercrime may cause over ten trillion dollars in damage annually by 2025.[249] Cybercriminals use AI to commit crime more quickly, efficiently, and at a larger scale.[250] AI-powered cyberattacks consist of deep fakes, AI-powered password cracking, AI-assisted hacking, supply chain attacks, business email compromises (BEC), advanced persistent threats (APTs), ransomware

[245] *Id.*

[246] *AI and Privacy*AI*: The Privacy Concerns Surrounding AI, Its Potential Impact On Personal Data*, ECON. TIMES (Apr. 25, 2023, 8:31PM), https://economictimes.indiatimes.com/news/how-to/ai-and-privacy-the-privacy-concerns-surrounding-ai-its-potential-impact-on-personal-data/articleshow/99738234.cms?from=mdr [https://perma.cc/YW7M-6VSJ].

[247] *Id.*

[248] *What Is a Cyberattack?*, IBM, https://www.ibm.com/topics/cyber-attack (last visited Jan. 11, 2024) [https://perma.cc/669T-UWAB].

[249] Steve Morgan, *2022 Cybersecurity Almanac: 100 Facts, Figures, Predictions And Statistics*, CYBERSECURITY MAG. (2023), https://cybersecurityventures.com/cybersecurity-almanac-2022/ (last visited Jan 11, 2024) [https://perma.cc/T9S9-H8LV].

[250] Guarav Belani, *AI for Cybersecurity and Cybercrime: How Artificial Intelligence Is Battling Itself*, IEEE COMPUT. SOC'Y (Sept. (September 6, 2023), https://www.computer.org/publications/tech-news/trends/ai-fighting-ai/ [https://perma.cc/HR6L-D8KL].

attacks, fraudulent transactions, payment gateway fraud, distributed denial of service (DDoS) attacks, and intellectual property theft.[251] Additionally, criminals are using machine learning to improve algorithms for guessing user's passwords; automate and enhance various hacking activities; compromise the hardware and software supply chain of an organization by inserting malicious code or components into legitimate products and services; conduct phishing attacks at organizations to steal model and critical information from business executives and lawyers; breach business networks undetected; demand ransom for decryption codes; conduct payment gateway frauds; attack business websites and online services; steal valuable intellectual property; and enhance the intensity of DDoS attacks or malicious attempts to disrupt the normal traffic against the targeted server, service, or network by overwhelming the target for its surrounding infrastructure with a flood of internet traffic.[252]

LLMs enhance existing attacks by making it more difficult for antivirus software/spam filters to detect threats; creating new attacks by

[251] Rabiul Islam, *AI And Cybercrime Unleash A New Era Menacing Threats*, FORBES (June 23, 2023, 5:45 AM), https://www.forbes.com/sites/forbestechcouncil/2023/06/23/ai-and-cybercrime-unleash-a-new-era-of-menacing-threats/?sh=33283155162b [https://perma.cc/QP8M-PSWG].

[252] *What is a DDoS attack?*, CLOUDFLARE.COM, (accessed Jan. 8, 2023), available at: https://www.cloudflare.com/learning/ddos/what-is-a-ddos-attack/#:~:text=A%20distributed%20denial%2Dof%2Dservice%20(DDoS)%20attack%20is,a%20flood%20of%20Internet%20traffic; Islam, *supra* 251.

manipulating or creating fake data to create confusion or impersonate officials; and automating and scaling attacks.[253] These attacks can take various forms, from malware and phishing to more sophisticated methods like ransomware and DDoS attacks.[254] Malware injects harmful software into systems to steal or corrupt data,[255] while phishing uses deceptive communication to trick individuals into revealing sensitive information. In fact, phishing attacks have risen over 1,276% from the fourth quarter of 2022 to the fourth quarter of 2023.[256] ChatGPT's ability to draft highly realistic texts makes it a useful tool for phishing purposes or for the spread of propaganda and disinformation.[257] Ransomware locks users out of their systems, demanding payment for access restoration, and DDoS attacks overload servers, rendering them inoperable.[258] Some cybercriminals are even using AI to negotiate the ransomware.[259] With the advent of AI, these attacks

[253] Islam, *supra* note 251.

[254] IBM, *supra* note 248.

[255] *Malware and Ransomware*, UNIV. WASH. INFO. TECH., https://ciso.uw.edu/education/risk-advisories/malware-and-ransomware/ (last visited Jan. 11, 2024) [https://perma.cc/77NW-VN8A].

[256] Bob Violino, *AI Tools Such As Chatgpt Are Generating a Mammoth Increase in Malicious Phishing Emails*, CNBC (Nov. 28, 2023, 10:39 AM), https://www.cnbc.com/2023/11/28/ai-like-chatgpt-is-creating-huge-increase-in-malicious-phishing-email.html [https://perma.cc/N3Z3-F4A2].

[257] *Id.*

[258] *Understanding denial-of-service attacks*, CYBERSEC. AND INFRASTRUCTURE SEC. AGENCY (Feb. 1, 2021), https://www.cisa.gov/news-events/news/understanding-denial-service-attacks [https://perma.cc/84L7-9ZRS].

[259] Alex Holden, *Contending With Artificially Intelligent Ransomware*, ISACA (Sept. 1, 2023), https://www.isaca.org/resources/news-and-trends/isaca-now-blog/2023/contending-with-artificially-intelligent-ransomware [https://perma.cc/P67W-43H4].

have become more sophisticated because AI algorithms can analyze vast datasets to identify vulnerabilities, automate the creation of phishing messages that are incredibly convincing,[260] or optimize the spread of malware in ways that evade traditional detection methods. One of the primary legal challenges is the attribution of cybercrimes facilitated by AI, which may not involve humans, or the kind of intent required to prosecute. Who or what would be responsible if an AI system autonomously initiated a cyberattack: the programmer, user, or the AI itself?

AI-driven cybersecurity solutions, while effective in detecting and mitigating attacks, often require access to vast amounts of personal data.[261] This situation poses a risk of violating individual privacy rights, especially if data is handled without stringent safeguards and transparency. The potential for inherent biases in AI algorithms further complicates ethical considerations. For example, an AI system designed to detect fraudulent activities might inadvertently exhibit bias against certain demographic

[260] Michael Hill, *Generative AI Phishing Fears Realized As Model Develops "Highly Convincing" Emails In 5 Minutes*, CSO (Oct. 24, 2023), https://www.csoonline.com/article/656698/generative-ai-phishing-fears-realized-as-model-develops-highly-convincing-emails-in-5-minutes.html https://www.csoonline.com/article/656698/generative-ai-phishing-fears-realized-as-model-develops-highly-convincing-emails-in-5-minutes.html [https://perma.cc/USQ8-7C6P].

[261] Tamer Charife & Michael Mossad, *AI In Cybersecurity: A Double-Edged Sword*, DELOITTE MIDDLE E., https://www2.deloitte.com/xe/en/pages/about-deloitte/articles/securing-the-future/ai-in-cybersecurity.html [https://perma.cc/6FAE-NJ4J] (last visited Jan. 11, 2024).

groups, leading to unfair targeting or neglect as discussed below. Ensuring that AI systems in cybersecurity are unbiased, equitable, and respectful of privacy is therefore critical.

Companies and lawmakers must also consider the converse issue: cyberattacks on AI systems themselves. According to a January 2024 report from NIST, there are four main types of cyberattacks that AI systems can encounter: evasion, poisoning, privacy, and abuse attacks.[262] Evasion attacks occur post-deployment, where adversaries alter inputs to change the AI system's response, for instance, manipulating road signs to mislead an autonomous vehicle. Poisoning attacks happen during the training phase, where corrupted data is introduced, such as injecting inappropriate language into a chatbot's training dataset. Privacy attacks aim to extract sensitive information about the AI system or its training data, potentially leading to misuse. Abuse attacks involve feeding incorrect information from a compromised but legitimate source to change the purpose of the AI system.[263]

[262] *NIST Identifies Types of Cyberattacks That Manipulate Behavior of AI Systems*, NIST (Jan. 4, 2024), https://www.nist.gov/news-events/news/2024/01/nist-identifies-types-cyberattacks-manipulate-behavior-ai-systemshttps://www.nist.gov/news-events/news/2024/01/nist-identifies-types-cyberattacks-manipulate-behavior-ai-systems [https://perma.cc/9FNW-X2WP].

[263] *Id.*

To conclude, the intersection of privacy, cybersecurity, and AI is a rapidly evolving field. Privacy concerns, primarily about the collection, storage, and dissemination of personal information, have heightened with the advent of AI due to its capabilities for enhanced data aggregation. This raises critical issues around the handling and use of PII. Meanwhile, cybersecurity, distinct yet related to privacy, focuses on protecting systems from intrusion and data breaches. AI's role in cybersecurity is double-edged: while it can bolster defenses by identifying vulnerabilities and automating responses, it also presents novel risks. For instance, phishing attacks have seen a significant increase, partly due to AI's ability to generate convincing deceptive communications.

Moreover, AI poses unique legal challenges, especially in attributing liability for AI-facilitated cybercrimes, where traditional concepts of intent and human agency are insufficient. Regulators must ensure that AI systems not only adhere to stringent privacy and data protection standards but are also equipped to detect and mitigate sophisticated cyber threats. Lawmakers should also consider updating legal frameworks to address the unique challenges posed by AI, including redefining liability and attribution in the context of AI-driven cybercrimes. Simultaneously, companies must invest in robust AI security measures, continually monitor for biases, and maintain transparency in AI data usage to uphold public trust and comply with evolving legal standards.

Algorithmic Bias

AI is not always neutral or harmless[264] because the algorithms are only as good as the data they are trained on.[265] Algorithmic decision-making is susceptible to inaccuracies, discriminatory outcomes, and embedded or inserted bias.[266] When the data embedded in an AI algorithm is incomplete, it may not be able to identify misinformation, fake news, or inaccurate data.[267] Poor, incomplete, and faulty data create inaccurate predictions that reflect the "garbage in, garbage out" phenomenon.[268] The prejudices and cognitive

[264] Lian Parsons, *Ethical Concerns Mount as AI Takes Bigger Decision-Making Role*, HARV. UNIV. GAZETTE (Oct. 26, 2020), https://news.harvard.edu/gazette/story/2020/10/ethical-concerns-mount-as-ai-takes-bigger-decision-making-role/ [https://perma.cc/3H43-PC57]; FUTURE OF LIFE INST., *supra* note 26; Niral Sutaria, *Bias And Ethical Concerns In Machine Learning,* ISACA (Aug. 29, AND ETHICAL CONCERNS IN MACHINE LEARNING (2022), https://www.isaca.org/resources/isaca-journal/issues/2022/volume-4/bias-and-ethical-concerns-in-machine-learning [https://perma.cc/BY9E-3UYJ]; Tamlyn Hunt*, Here's Why AI May Be Extremely Dangerous--Whether It's Conscious Or Not*, SCIENTIFIC AM. (May 25, 2023), https://www.scientificamerican.com/article/heres-why-ai-may-be-extremely-dangerous-whether-its-conscious-or-not/ [https://perma.cc/AS88-7K3K].

[265] MacKenzie Sigalos & Ryan Browne, *A.I. Has A Discrimination Problem. In Banking, The Consequences Can Be Severe*, CNBC (June 23, 2023, 1:45 AM), https://www.cnbc.com/2023/06/23/ai-has-a-discrimination-problem-in-banking-that-can-be-devastating.html [https://perma.cc/KT2B-9CV3].

[266] *Artificial Intelligence: Examples of ethical dilemmas*, UNESCO, https://www.unesco.org/en/artificial-intelligence/recommendation-ethics/cases [https://perma.cc/J5V6-V7FB] (last updated Apr. 21, 2023).

[267] Abdulaziz Aldoseri et al., *Re-Thinking Data Strategy And Integration For Artificial Intelligence: Concepts, Opportunities, And Challenges*, MDPI (June 13, 2023), https://www.mdpi.com/2076-3417/13/12/7082 [https://perma.cc/RN9L-LEDJ].

[268] Monique Kilkenny & Kerin M. Robinson, *Data Quality: Garbage In Garbage Out*, 47 HEALTH INFO. MGT. J. 103, 104 (2018).

biases of our everyday life spill into machine learning.[269] These algorithms reflect stereotypes, bandwagon effects, priming, and confirmation bias.[270]

AI algorithms may perpetuate discrimination but can also cause more serious harm.[271] AI errors and bias may lead to privacy violations, misinformation and deep fakes, heightened criminalization of and discrimination against minorities, biased recruiting and consumer scores, algorithmic and discriminatory censorship, radicalization, and poor mental health.[272] Further, because so few AI companies reveal how they determine their conclusions, this exacerbates the "black box" problem, where there is no transparency.[273] This makes it nearly impossible for regulators and other observers to address and fix biases.

[269] *Shedding Light On AI Bias With Real World Examples*, IBM BLOG (Oct. 16, 2023), https://www.ibm.com/blog/shedding-light-on-ai-bias-with-real-world-examples/ [https://perma.cc/HEJ6-B3M9].

[270] Abigail Christina Fernandez, *Biased Data: For Better Or For Worse? A Comprehensive Case Study and Analysis in Machine Learning*, IGI GLOBAL (1970), https://www.igi-global.com/chapter/biased-data/256672 (last visited Jan 11, 2024).

[271] *See generally infra AI Bias, Censorship, and Social Media.*

[272] *See generally id.*

[273] Lou Blouin, *Ai's Mysterious "Black Box" Problem, Explained*, UNIV. MICH. DEARBORN (Mar. 6, 2023), https://umdearborn.edu/news/ais-mysterious-black-box-problem-explained [https://perma.cc/L4RA-JAJQ].

Racialized Algorithmic Bias

AI can produce racialized algorithmic biases[274] because digital technologies do not exist in a vacuum.[275] Facial recognition algorithms are less accurate with people with darker skin tones – especially women.[276] In December 2023, the Federal Trade Commission banned retailer Rite Aid from using AI-based facial recognition technology for five years after the agency determined that over a ten year period the company had hired two companies to build databases of "persons of interest" and falsely tagged women and people of color as potential shoplifters.[277] The FTC alleged that

[274] Kathleen Walch, *Training Data In Facial Recognition Use Cases Reveals Bias*, TECHTARGET ENTER. A.I. (Nov. 8, 2019), https://www.techtarget.com/searchenterpriseai/feature/Training-data-in-facial-recognition-use-cases-reveals-bias?Offer=abMeterCharCount_ctrl [https://perma.cc/J4LR-JNZ3]; Gian Volpicelli, *Forget Chatgpt: Facial Recognition Emerges As Ai Rulebook's Make-Or-Break Issue*, POLITICO (June 14, 2023, 5:07 PM), https://www.politico.eu/article/facial-recognition-artificial-intelligence-act-ai-issue-european-parliament/ [https://perma.cc/29TP-ZQLV].

[275] U.N. HUMAN RIGHTS OFF., OHCHR, https://www.ohchr.org/en/privacy-in-the-digital-age (last visited Jan. 11, 2024).

[276] *Companies Must Act Now To Ensure Responsible Development Of Artificial Intelligence,* AMNESTY INT'L (June 14, 2023, 2:50 PM), https://www.amnesty.ca/surveillance/racial-bias-in-facial-recognition-algorithms/#:~:text=Misidentification%20in%20facial%20recognition%20technology&text=Facial%20recognition%20is%20less%20accurate,police%20have%20wrongfully%20arrested%20people [https://perma.cc/E5YG-VCUT]; Hilary Homes, COMPANIES MUST ACT NOW TO ENSURE RESPONSIBLE DEVELOPMENT OF ARTIFICIAL INTELLIGENCE (2023), https://www.amnesty.ca/surveillance/racial-bias-in-facial-recognition-algorithms/#:~:text=Misidentification%20in%20facial%20recognition%20technology&text=Facial%20recognition%20is%20less%20accurate,police%20have%20wrongfully%20arrested%20people (last visited Jan 11, 2024).

[277] Amy Ritchie & John Newman, *Rite Aid banned from using AI facial recognition after FTC says retailer deployed technology without reasonable safeguards*, FED. TRADE COMM'N (Dec. 19, 2023), https://www.ftc.gov/news-events/news/press-releases/2023/12/rite-aid-banned-using-ai-facial-recognition-after-ftc-says-retailer-deployed-technology-without [https://perma.cc/WSJ5-HN3R].

Rite Aid employees, based on fuzzy pictures and wrong information, followed customers around the store, searched them, called the police, and caused humiliation and embarrassment.

AI Bias, Censorship, and Social Media

AI algorithms may also discriminate against marginalized communities in the virtual sphere through content moderation.[278] Social media companies use content moderation algorithms in part to cultivate safe spaces with accurate information for their users.[279] However, the personalization and customization of content minimizes users' exposure to diverse views.[280] It also interferes with user's options to seek and share ideas and opinions across ideological, political or societal divisions.[281] This type of moderation falls under the category of "algorithmic censorship."[282] Algorithmic censorship hinders marginalized communities from organizing protests and assembling for change because marginalized communities'

[278] Theara Coleman, *How Content Creators Cope With Discriminatory Algorithms*, THE WEEK (May 16, 2023), https://theweek.com/briefing/1023338/algorithm-ai-discrimination [https://perma.cc/9T2G-KK4U].

[279] Oliver L. Haimson et al., *Disproportionate Removals and Moderation Experiences Conservative for Conservative, Transgender* and *Black Social Media Users*, 5 PROC. ACM HUM. COMPUT. INTERACT. 466 (2021).

[280] Cameran Ashraf, *Artificial Intelligence and the Rights to Assembly and Association*, 5 J. OF CYBER POL'Y163, 169 (2020).

[281] *Id.*

[282] *Id.* at 171.

content is removed and moderated more often than their counterparts.[283] Shadow banning or making users' content invisible to others allows social media platforms to erase marginalized voices.[284] Unfortunately, in many cases, social media algorithms do not simply discriminate but actively perpetuate harmful disinformation and conspiracy theories, fake images, videos, and texts.[285]

Algorithms may also have negative mental health effects such as anxiety, depression, and body image concerns.[286] Beyond content moderation, some social media algorithms manipulate what people see to keep them addicted.[287] For example, Instagram and Facebook algorithms are intentionally designed to mimic addictive painkillers.[288] Even so, platforms continue to use algorithms to maximize engagement at the cost of increasing

[283]Haimson, *supra* note 279, at 1.; Hibby Thach et al., *(In)visible Moderation: A Digital Ethnography of Marginalized users and content moderation on Twitch and Reddit*, 22 SAGE J. NEWS MEDIA & SOC'Y. (2022).

[284] Thach, *supra* note 283, at 4.

[285] Miriam Fernandez & Harith Alani, *The Open University Monthly Downloads Bar Chart* (1970), https://oro.open.ac.uk/69799/ (last visited Mar 31, 2024).

[286] *Social Media and the internet*, AM. PSYCH. ASSOC. (Nov. 2022), https://www.apa.org/topics/social-media-internet [https://perma.cc/DS68-73SM].

[287] Christopher Cocchiarella, *Manipulative algorithms and addictive design on social media*, MINDFUL TECHNICS (Nov. 30, 2021), https://mindfultechnics.com/manipulative-algorithms-and-addictive-design-summing-up-whats-wrong-with-social-media/ [https://perma.cc/K377-H6D6].

[288] Hannah Schwär, *How Instagram And Facebook Are Intentionally Designed To Mimic Addictive Painkillers*, BUS. INSIDER (Aug. 11, 2021, 6:38 PM), https://www.businessinsider.com/facebook-has-been-deliberately-designed-to-mimic-addictive-painkillers-2018-12; Jessica Wulf, *Automated Decision-Making Systems And Discrimination*, ALGORITHM WATCH (June 2022), https://algorithmwatch.org/en/wp-content/uploads/2022/06/AutoCheck-Guidebook_ADM_Discrimination_EN-AlgorithmWatch_June_2022.pdf [https://perma.cc/4XWN-YPQV].

anxiety, depression, and suicide.[289] This is particularly dangerous for teens who have a heightened vulnerability to addiction.[290] In a 2022 lawsuit, 12,000 families sought to hold Meta's Instagram platform responsible for "causing and contributing to the burgeoning mental health crisis perpetrated upon children and teenagers of the United States."[291] Plaintiffs alleged that Facebook knew that its algorithms had dangerous designs and design defects and that these algorithms cause addiction, anxiety, depression, eating disorders, self-injury, and suicide to its users.[292]

Social media companies' actions may also invade privacy. A recent study found that social media algorithms can accurately predict depression and other mental health-related conditions based on Instagram photos and Tweets.[293] Users who posted bluer, darker, or greyer photos that received more comments but fewer likes were associated with depression.[294] A related

[289] *Social Media Addiction Lawyer*, CASEYGERRY, https://www.caseygerry.com/social-media-addiction-lawyer/ [https://perma.cc/H3CV-CJFV] (last visited Feb. 16, 2024).

[290] Michelle Llamas, *Social Media Lawsuits I Tech Giants Sued For Social Media Harm*, CONSUMER NOTICE, https://www.consumernotice.org/legal/social-media-harm-lawsuit/[https://perma.cc/Y7NV-T8MS] (last updated Mar. 1, 2024).

[291] Complaint For Personal Injuries, Spence v. Meta Platforms, Inc., No. 3:22-cv-03294, 2 (N.D. Cal. June 6, 2022) https://socialmediavictims.org/wp-content/uploads/2022/06/Spence-Complaint-6_6_22.pdf [https://perma.cc/ES3R-BN3P].

[292] *Id.*

[293] *Id.* at 7.

[294] *Id.*

study found that users who post negative words were also associated with depression.[295]

AI Bias and Housing

AI algorithms may also perpetuate discriminatory outcomes in housing for minorities based on unrepresentative, insufficient, and biased data.[296] In one study, a nonprofit news organization revealed that AI bias was responsible for 80% of Black mortgage applications being denied.[297] The UN Special Rapporteur on Racism has documented that Facebook used targeted advertising to prevent minorities from viewing certain content relating to housing opportunities.[298]

In 2022, plaintiffs filed a class action lawsuit arguing that SafeRent Solutions, LLC's scoring, based in part on information in their credit report, amounted to discrimination against Black and Hispanic renters in violation of the Fair Housing Act.[299] The lawsuit alleged that SafeRent's algorithm has

[295] *Id.*

[296] *Mining The Data: Algorithmic Bias In Housing Related*, NFHTA FORUM (Jan. 19, 2022, 2:00 PM), https://www.hudexchange.info/trainings/courses/nfhta-forum-mining-the-data-algorithmic-bias-in-housing-related-transactions/ [https://perma.cc/XUL3-52HB].

[297] Emmanuel Martinez et al., *The Secret Bias Hidden in Mortgage-Approval Algorithms*, THE MARKUP (Aug. 25, 2021), https://themarkup.org/denied/2021/08/25/the-secret-bias-hidden-in-mortgage-approval-algorithms [https://perma.cc/DWT7-VHA6].

[298] *Racial Discrimination and Emerging Digital Technologies: A Human Rights Analysis*, BUS. AND HU. RTS. RES. CTR. (June 18, 2020), https://www.business-humanrights.org/en/latest-news/un-report-on-racial-discrimination-in-emerging-digital-technologies-prompts-calls-for-structural-interventions/ [https://perma.cc/PBP5-XSYZ].

[299] Khari Johnson, *Algorithms Allegedly Penalized Black Renters. The US Government Is Watching*, WIRED (Jan. 16, 2023), https://www.wired.com/story/algorithms-allegedly-penalized-black-renters-the-us-government-is-watching/ [https://perma.cc/6527-L57D].

had disparate impact based on race and source of income in violation of federal and state laws.[300] Furthermore, the plaintiffs contended, "racial disparities in credit history and credit scores not only reflect historical disparities in wealth, but also perpetuate wealth inequalities through reduced financial opportunities and fewer financial safety nets, which hinder a consumer's ability to accumulate present or intergenerational wealth through homeownership or other financial investments."[301] Notably, the SafeRent Score algorithm "does not disclose all of the data it considers or how this data is weighted in this scoring modeling, thereby keeping its inner workings hidden."[302]

AI Bias and Employment

Algorithmic discrimination may also infringe on employee's rights because employers increasingly use AI for recruitment, hiring, and promotion.[303] Employment algorithms, used by over half of human resources

[300] Statement of Interest of the United States, Louis. v. SafeRent Solutions, LLC., No. 22-cv-10800-AK22cv10800-ak, 3 (D. Mass. Jan. 9, 2023) https://www.justice.gov/media/1268711/dl?inline https://www.justice.gov/media/1268711/dl?inline [https://perma.cc/HEV2-TKYV].

[301] *SafeRent Solutions Accused of Illegally Discriminating Against Black and Hispanic Rental Applicants*, COHEN MILSTEIN (May 25, 2023), https://www.cohenmilstein.com/update/saferent-solutions-accused-illegally-discriminating-against-black-and-hispanic-rental [https://perma.cc/7B3Q-5YJV].

[302] *Id.*

[303] David E. Schwartz et al., *AI and the Workplace: Employment Considerations: Insights,* SKADDENARPS, SLATE, MEAGHER & FLOM LLP (June 2023), https://www.skadden.com/insights/publications/2023/06/quarterly-insights/ai-and-the-workplace [https://perma.cc/6PMT-QPQB].

professionals, may increase unfair employment practices.[304] Employment algorithms fall under the category of algorithmic decision-making systems (ADMS) which may discriminate against individuals with protected characteristics such as disability, gender, and race.[305] In 2016, Amazon shut down its experimental artificial intelligence (AI) recruiting tool after realizing it discriminated against women because the AI tool skewed results in favor of White or male applicants because of criteria such as ratings, pay, and titles.[306] Furthermore, it replicated the company's existing disproportionately male workforce.[307]

Legislators have begun to focus on ADMS. In February 2023, the California legislature proposed employment AI regulation legislation requiring audits of AI tools used by employers and created by developers.[308] Additionally, in April 2023, New York City enacted a law that requires companies to show that the AI-powered tools they used in hiring and

[304] Alex Engler, *Auditing Employment Algorithms for Discrimination*, BROOKINGS (Mar. 12, 2021), https://www.brookings.edu/articles/auditing-employment-algorithms-for-discrimination [https://perma.cc/8S7E-7VXE].

[305] Laura Lapidus, *Using AI in Employment Decisions*, RISK MGMT MAG. (Apr. 3, 2023), https://www.rmmagazine.com/articles/article/2023/04/03/using-ai-in-employment-decisions [https://perma.cc/VWD7-S43S]; Jeremias Adams-Prassl et al., *Directly Discriminatory Algorithms*, 86 MOD. L. REV. 144 (2022).

[306] Schwartz, *supra* note 303.

[307] Cameron F. Kerry, *Protecting privacy in an AI-Driven World*, BROOKINGS (Feb. 10, 2020), https://www.brookings.edu/articles/protecting-privacy-in-an-ai-driven-world/ [https://perma.cc/554M-2W34].

[308] Jeffrey S. Bosley et al., *California Proposed Employment AI Regulations and Legislation*, DAVIS WRIGHT TREMAINE (May 2, 2023), https://www.dwt.com/blogs/employment-labor-and-benefits/2023/05/california-ai-employment-law-regulations [https://perma.cc/TTE8-8A4F].

promotion are not biased and to conduct bias audits.[309] As of this writing, there are also pending bills in Illinois that prevent the use of race in recruiting, hiring, and promotion algorithms and in New York state that requires employers and employment agencies to notify candidates for employment if machine learning technology is used to make hiring decisions.[310] In May 2023, the federal Equal Employment Opportunity Commission released a technical assistance document to aid employers in complying with Title VII and other civil rights laws under its purview.[311] In August 2023, the EEOC fined a company $365,000 for screening out 200 older applicants through its ADMS tool.[312]

AI Bias, Consumer Scores, and Healthcare

More recently, AI algorithms have been attacked for "consumer scores" which businesses such as hospitals and universities use to predict how consumers will behave in the future.[313] Consumer scores summarize relevant information for companies and corporations based on past behavior

[309] N.Y.C., N.Y., RULES OF THE CITY OF N.Y. tit. 6 § 5-300–04 (2022) (amended 2023).

[310] NAT'L CONF. OF STATE LEGIS., *supra* note 55.

[311] Press Release, *EEOC Releases New Resource on Artificial Intelligence and Title VII*, U.S. EQUAL EMP. OPPORTUNITY COMM'N (May 18, 2023), https://www.eeoc.gov/newsroom/eeoc-releases-new-resource-artificial-intelligence-and-title-vii [https://perma.cc/28CY-UM52].

[312] Press Release, *iTutorGroup to Pay $365,000 to Settle EEOC Discriminatory Hiring Suit*, U.S. EQUAL EMP. OPPORTUNITY COMM'N (Sept. 11, 2023), https://www.eeoc.gov/newsroom/itutorgroup-pay-365000-settle-eeoc-discriminatory-hiring-suit [https://perma.cc/4QS6-KMJZ].

[313] U.S. GOV'T ACCOUNTABILITY OFF., GAO-22-106096, CONSUMER DATA: INCREASING USE POSES RISKS TO PRIVACY (2022), https://www.gao.gov/assets/gao-22-106096.pdf [https://perma.cc/J7CA-YV5B].

and characteristics.[314] The data is collected from the web, mobile, and IoT devices that detail demographics, geography, and credit history.[315] This risks biased outcomes, inaccurate scores, and differential treatment in health care administration and higher education.[316]

Although, as stated earlier, AI can greatly increase access to medical treatment, there are downsides to its usage. In healthcare, consumer scores assess an individual's health status or history which hospitals and health companies use to help "assist with treatment triage services, patient payment strategies, and more."[317] However, these health consumer scores may contain racial biases.**[318] "[**F]or example, one study found that Black patients were assigned lower risk scores than White patients with the same health care needs, predicting less of a need for a care management program."[319] A 2019 study found that many health-care algorithms in the U.S. unintentionally yet systematically discriminated against Black patients.[320] As a result, Black

[314] *See* Sara M. Watson, *Predictive Analytics and Consumer Scoring*, INSIDER INTELL. (Nov. 30, 2020), https://www.insiderintelligence.com/content/predictive-analytics-consumer-scoring [https://perma.cc/278E-YKCL]; U.S. GOV'T ACCOUNTABILITY OFF., GAO-22-104527, CONSUMER PROTECTION: CONGRESS SHOULD CONSIDER ENHANCING PROTECTIONS AROUND SCORE USED TO RANK CONSUMERS (May 2022), https://www.gao.gov/assets/gao-22-104527.pdf [https://perma.cc/6GSP-PDRQ].

[315] Watson, *supra*, note 314, at 4.

[316] U.S. GOV'T ACCOUNTABILITY OFF., *supra* note 313.

[317] U.S. GOV'T ACCOUNTABILITY OFF., *supra* note 314.

[318] *Id.*

[319] *Id.*

[320] Crystal Grant, *Algorithms Are Making Decisions About Health Care, Which May Only Worsen Medical Racism*, ACLU (Oct. 3, 2022), https://www.aclu.org/news/privacy-technology/algorithms-in-health-care-may-worsen-medical-racism [https://perma.cc/3YV7-Q44Z].

patients were less likely to be referred to programs for interventions to improve their health.[321]

In 2022, a study found that an AI tool trained on medical images such as x-rays and CT scans, "had unexpectedly" learned to discern patients' self-reported race even though it was trained to help clinicians diagnose patient images.[322] Furthermore, it could discern a patient's race even if the doctor could not.[323] Additionally, another study published by *Nature Medicine* in 2022, found that these AI tools "consistently and selectively underdiagnosed under-served patient populations," who also do not have access to medical care.[324] Currently, there are gaps in the federal regulatory structure governing algorithms used in health care which leads to the use of unvetted technology based on biased data.[325]

AI Bias, Social Welfare Systems, and Immigration

[321] *Id.*

[322] *Id.*

[323] *Id.*; *AI Systems Can Detect Patient Race, Creating New Opportunities to Perpetuate Health Disparities*, EMORY UNIV. (May 27, 2022), https://news.emory.edu/stories/2022/05/hs_ai_systems_detect_patient_race_27-05-2022/story.html [https://perma.cc/3JRS-UE4B].

[324] Laleh Seyyed-Kalantari et al., *Underdiagnosis Bias of Artificial Intelligence Algorithms Applied to Chest Radiographs in Under-Served Patient Populations*, 27 NATURE MED. 2176, 2176 (2021), https://www.nature.com/articles/s41591-021-01595-0.pdf [https://perma.cc/XZK4-ZQD2].

[325] Crystal Grant, *ACLU White Paper: AI in Health Care May Worsen Medical Racism*, ACLU, https://www.aclu.org/sites/default/files/field_document/algo_health_white_paper_draft_final_v4.pdf [https://perma.cc/G3QG-GTVN].

Notwithstanding these known risks, governments continue to use AI to replace human decision-making,[326] increasingly using automated decision-making tools.[327] AI use in governmental social services and welfare is becoming more common.[328] For example, in mental health services, AI-powered chatbots and virtual therapists are being developed to offer counseling services.[329]

Additionally, public servants are using AI-powered fraud detection to track down large-scale corruption in benefit and welfare programs by identifying patterns in phone numbers and social media profiles.[330] In a major scandal over the use of algorithms to ferret out welfare fraud, an entire government resigned after investigators learned that 20more than 20,000 families, profiled due to ethnic sounding last names or birthplaces, were falsely accused of fraud.[331] Ten thousand of those families repaid benefits to

[326] EUROPEAN COMM'N., COMPETENCE CENTRE ON FORESIGHT (Feb, 7, 2023), https://knowledge4policy.ec.europa.eu/foresight/automated-decision-making-impacting-society_en [https://perma.cc/5HW2-MVD6].

[327] Jeffrey B. Welty, *Artificial Intelligence and the Practice of Criminal Law*, UNIV. N.C. SCH. OF GOV'T BLOG (Mar. 27, 2023), https://nccriminallaw.sog.unc.edu/artificial-intelligence-and-the-practice-of-criminal-law/.

[328] Marcin Frąckiewicz, *The Role of AI in Social Services and Welfare*, TECHNOSPACE2 (June 23, 2023), https://ts2.space/en/the-role-of-ai-in-social-services-and-welfare/.

[329] *Id.*

[330] Cem Dilmegani, *AI In Government: Examples, Challenges & Best Practices*, AI MULTIPLE (Jan. 12, 2024), https://research.aimultiple.com/ai-government/ [https://perma.cc/9ABH-ZRMH].

[331] Jon Henley, *Dutch Government Resigns Over Child Benefits Scandal*, THE GUARDIAN (Jan. 15, 2021), https://www.theguardian.com/world/2021/jan/15/dutch-government-resigns-over-child-benefits-scandal [https://perma.cc/8ZHD-XE77].

which they were entitled, and people lost jobs, homes, and marriages.[332] Some lost their lives to suicide.[333]

In 2022, the Customs and Border Portal implemented an AI Center for Innovation to implement AI in the immigration process.[334] The plan is to implement facial recognition systems and AI powered data processors to more quickly digest and process applicant information.[335] However, African and Haitian applicants at the US-Mexico border who are required to use the system have complained that the system cannot properly identify their faces because of their darker skin tones.[336] This worsens the border crisis by causing delays, and more alarmingly, could put migrants' lives at risk.

Predictive Policing

Predictive policing falls into four categories: predicting crimes, predicting offenders, predicting identities, and predicting victims.[337] These algorithms have allowed police to deploy more officers to minority

[332] *Id.*

[333] Melissa Heikkilä, *Dutch Scandal Serves As A Warning for Europe Over Risks Of Using Algorithms*, POLITICO (Mar. 29, 2022, 6:14 PM), https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms/ [https://perma.cc/5TB9-BVSR].

[334] *Artificial Intelligence to Harness Key Insights at CBP*, U.S. CUSTOMS & BORDER PROTECTION (Mar. 24, 2023), https://www.cbp.gov/newsroom/spotlights/artificial-intelligence-harness-key-insights-cbp [https://perma.cc/YR7K-YYB5].

[335] *Id.*

[336] Melissa del Bosque, *Facial Recognition Bias Frustrates Black Asylum Applicants to* US*, Advocates Say*, GUARDIAN (Feb. 8, 2023, 6:00 AM), https://www.theguardian.com/us-news/2023/feb/08/us-immigration-cbp-one-app-facial-recognition-bias [https://perma.cc/9842-MCJF].

[337] WALTER L. PERRY ET AL., PREDICTIVE POLICING: THE ROLE OF CRIME FORECASTING IN LAW ENFORCEMENT OPERATIONS XIV (2013).

neighborhoods.[338] Predictive policing was first developed in 2008 at the LAPD, but it is now used in many major cities in the United States.[339] The COMPAS algorithm is "one of the most widely used algorithms in the U.S. criminal justice system and it has been applied" in many of the largest states in the country, including Florida.[340] This system uses a 137-factor process[341] to perform a risk assessment to determine "the appropriate level of supervision for each client."[342] In 2013 and 2014, incomplete and biased data was used to predict future criminals in Broward County, Florida[343] and falsely flagged Black defendants as future criminals, incorrectly labeling them at "almost twice the rate as [W]hite defendants."[344] As of 2022, many cities including Los Angeles, New York, and Chicago have stopped using

[338] Martinez et al., *supra* note 297; Jonathan Li, *Pitfalls of Predictive Policing: An Ethical Analysis*, VCE (Feb. 17, 2022), https://vce.usc.edu/volume-5-issue-3/pitfalls-of-predictive-policing-an-ethical-analysis/#:~:text=This%20tactic%2C%20which%20has%20been,low%2Dincome%20neighborhoods%20and%20high%2D [https://perma.cc/5XFR-673A].

[339] Li, *supra* note 338.

[340] Alexandra "Mac" Taylor, *AI Prediction Tools Claim to Alleviate an Overcrowded American Justice System... but Should They Be Used?*, STAN. POL. (Sept. 13, 2020), https://stanfordpolitics.org/2020/09/13/ai-prediction-tools-claim-to-alleviate-an-overcrowded-american-justice-system-but-should-they-be-used// [https://perma.cc/V3TN-RTV2].

[341] Kathleen Moore, *Computer Systems that Help Judges Are 'Far' from Being Reliable, UAlbany Professor Says*, TIMES UNION (Feb. 24, 2023, 11:11 AM), https://www.timesunion.com/news/article/computer-systems-help-judges-far-reliable-17794755.php.

[342] THOMAS BLOMBERG ET AL., VFLA. STATE UNIV., VALIDATION OF THE COMPAS RISK ASSESSMENT 15 (2010).

[343] Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [https://perma.cc/BW58-BQAY}].

[344] *Id.*

predictive policing because it perpetuates systemic racism through the use of biased data.[345] However, as of 2024, police departments have continued to employ predictive policing throughout the United States.[346]

Facial recognition systems have also been attacked for algorithmic racial biases that fuel technologically-facilitated discrimination,[347] which exacerbates problems of already discriminatory policing.[348] Amnesty International noted that racist facial recognition technology is one of the most urgent threats to human rights and racial justice.[349] A study by MIT Media Lab found that "nearly 35% of [B]lack women . . . were misidentified by prominent facial recognition systems" that could result in incorrect arrests or denying minorities access to buildings and locations.[350]

AI algorithms were intended to improve the American justice system through "computer-driven calculations about risk, crime, and recidivism."[351] However, their deployment discriminates against racially marginalized

[345] Li, *supra* note 338.

[346] Matthew Guariglia, *Artificial Intelligence and Policing: Year in Review 2023*, EFF (Dec 23, 2023), https://www.eff.org/deeplinks/2023/12/artificial-intelligence-and-policing-year-review-2023 [https://perma.cc/J4P3-6FKW].

[347] Walch, *supra* note 274; Volpicelli, *supra* note 274; COMPANIES MUST ACT MUST ACT NOW TO ENSURE RESPONSIBLE ENSURE RESPONSIBLE DEVELOPMENT OF ARTIFICIAL INTELLIGENCE, AMNESTY INT'L CAN. (June 14, 2023, 2:50 PM), https://www.amnesty.ca/surveillance/racial-bias-in-facial-recognition-algorithms/ [https://perma.cc/T62W-KPHD].

[348] *Id.*

[349] *Id.*

[350] Walch, *supra* note 274.

[351] Molly Callahan, *Algorithms Were Supposed to Reduce Bias in Criminal Justice—Do They?*, THE BRINK (Feb. 23, 2023), https://www.bu.edu/articles/2023/do-algorithms-reduce-bias-in-criminal-justice/ [https://perma.cc/CFC4-8GPR].].

communities by amplifying bias.[352] These "predictive policing" tools are populated from crime reports, arrest records, and license plate images to create an algorithm "which is trained to look for patterns to predict where and when a certain type of crime will occur in the future."[353]

More troubling, in the predictive policing context, algorithms replicate systemic racism. As researchers from the Georgetown Law Center explain,

> [[t]hey are built with real-world data, which records and reflects the criminal legal system's biases and abuses. That means, for example, that police relying on the output of a crime forecasting algorithm will go to the same streets and target the same people they have in the past. And in most cases, that produces a feedback loop resulting in a persistent and disproportionate police presence in communities of color.[354]

Surveillance

Many types of surveillance use AI. Across the world, governments collect and analyze "social media posts and the private and professional networks built on publicly accessible communications platforms."[355] Law enforcement surveillance employs facial recognition to monitor and track

[352] *Id.*

[353] Pranshu Verma, *The Never-Ending Quest to Predict Crime Using AI*, WASH. POST (July 15, 2022, 7:00 AM), https://www.washingtonpost.com/technology/2022/07/15/predictive-policing-algorithms-fail/ [https://perma.cc/565L-2GEP].

[354] JAMESON SPIVACK, *Cop Out: Automation in the Criminal Legal System*, GEO. L. CTR. ON PRIV. & TECH., https://copout.tech/about/ [https://perma.cc/L3JX-DLW5].

[355] OFF. OF THE U.N. HIGH COMM'R FOR HUM. RTS., *The Right to Privacy in the Digital Age*, ¶ 35, U.N. Doc. A/HRC/51/17 (Aug. 4, 2022).

protestors.[356] The UN has noted that "surveillance cameras, deployed to monitor public streets, car parks, transportation hubs and other public places have become common in many countries."[357] "Smart cities" are growing and "focus on the collection and processing of data to inform the management of city facilities, enabled by ever more capable sensor technologies."[358] As a consequence, identifying individuals wherever they are located has become easier and easier.[359] This heightened monitoring occurs both online and offline.[360]

Unfortunately, unwanted surveillance leads to people being wrongfully accused of crime, especially people of color and women.[361] Human rights advocates have long sounded the alarm on potential human rights abuses with surveillance technology. This concern is not unfounded, as evidenced by instances in countries like Iran, where facial recognition is used to track women breaking hijab laws,[362] and China, where the technology was used to track Muslim-minority Uighurs, who were then held in forced

[356] *Id.* at ¶ 32.
[357] *Id.* at ¶ 30.
[358] *Id.* at ¶ 33.
[359] *Id.*
[360] *Id.* at ¶ 43.
[361] Maria Curi, *Inside Today's AI-Human Rights Hearing*, AXIOS PRO (June 13, 2023), https://www.axios.com/pro/tech-policy/2023/06/13/inside-todays-ai-human-rights-hearing.
[362] Khari Johnson, *Iran Says Face Recognition Will ID Women Breaking Hijab Laws*, WIRED (Jan. 10, 2023, 7:00 AM), https://www.wired.com/story/iran-says-face-recognition-will-id-women-breaking-hijab-laws/ [https://perma.cc/8X6S-UTVT].

labor camps and "re-education camps."[363] In the United States, this technology's deployment has raised concerns, notably in the context of identifying participants in Black Lives Matter demonstrations.[364]

Surveillance raises the risk of misidentification, which can have severe consequences. Notably, this risk disproportionately affects people of color and women, attributable to inherent biases in the algorithms' underpinning facial recognition systems. While federal legislative action in the United States has been slow to respond to these challenges, there is a growing movement at the state level. Over the past five years, more than a dozen states have introduced regulations to limit the use of facial recognition technology, reflecting a growing recognition of its potential implications and the need for oversight.[365] At a 2023 Senate Judiciary hearing, Center for Democracy & Technology CEO Alexandra Reeve Givens outlined federal policy solutions, including requiring law enforcement to obtain a warrant

[363] Paul Mozur, *One Month, 500,000 Face Scans How China is Using A.I. to Profile a Minority*, N.Y. TIMES (Apr. 14, 2019), https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html; Jane Wakefield, *AI Emotion-Detection Software Tested Detection Software Tested on Uyghurs*, BBC (May 25, 2021), https://www.bbc.com/news/technology-57101248 [https://perma.cc/A5GF-XJRK].

[364] *USA: NYPD Ordered To Hand Over Documents Detailing Surveillance of Black Lives Matter Protests Following Lawsuit*, AMNESTY INT'L (Aug. 1, 2022), https://www.amnesty.org/en/latest/news/2022/08/usa-nypd-black-lives-matter-protests-surveilliance/#:~:text=Facial%20recognition%20exacerbates%20discriminatory%20policing,greater%20risk%20of%20being%20targeted [https://perma.cc/3C66-GYTT].

[365] Carolina Rabinowicz, *Approaches to Regulating Government Use of Facial Recognition Technology*, JOLT DIGEST (May 4, 2023), https://perma.cc/P26V-FW7P [https://perma.cc/P26V-FW7P].

before using facial recognition, limiting use of the technology for serious offenses, and imposing software accuracy standards.[366]

Deepfakes

Algorithms provide the foundation for deepfakes, which can create convincing images; audio and video hoaxes; voice cloning; generative text; and misinformation.[367] The situation is so serious that the Federal Trade Commission issued a cloning challenge in January 2024, offering a $25,000 prize for the best solution to end voice cloning.[368]

Deepfakes have been used to perpetuate gender-based violence, thwart political campaigns; commit blackmail and reputational harm; create false evidence; conduct fraud; disseminate misinformation and political manipulation; and attempt stock manipulation.[369] Women face heightened

[366] Curi, *supra* note 361.

[367] *See generally* Todd C. Helmus, ARTIFICIAL INTELLIGENCE, DEEPFAKES, AND DISINFORMATION, RAND CORP., JULY 2022, at 2–6 (discussing "the technology undergirding deepfakes and associated AI-driven technologies that provide the foundation for deepfake videos, voice cloning, deepfake images, and generative text"); Nick Barney & Ivy Wigmore, *Deepfake AI*, TECHTARGET (MAR. 2023) https://www.techtarget.com/whatis/definition/deepfake?Offer=abt_pubpro_AI-Insider (last visited Jan 11, 2024)) [https://perma.cc/LC58-AMDC] (explaining that "[d]eepfakes uses two algorithms—a generator and a discriminator—to create and refine fake content").

[368] *The FTC Voice Cloning Challenge,* FTC, https://www.ftc.gov/news-events/contests/ftc-voice-cloning-challenge [https://perma.cc/75Y4-24Y2] (last visited Jan. 11, 2024). The Office of Technology, FTC has also proposed new protections to combat AI impersonation of individuals Federal Trade Commission. *FTC Proposes New Protections to Combat AI Impersonation of Individuals*, FTC, https://www.ftc.gov/news-events/news/press-releases/2024/02/ftc-proposes-new-protections-combat-ai-impersonation-individuals (last visited Mar 31, 2024).

[369] Barney, *supra* note 367.

technology-facilitated violence from deepfake images and videos.[370] A study in 2019 by Deeptrace Labs revealed that 96% of all deepfake videos were nonconsensual-consensual pornography.[371] These algorithmic productions carry gender-based violence from the offline world to the digital world with the intention of embarrassing, shaming, humiliating, and intimidating women.[372]

More importantly, GAI's ability to create convincing deepfakes poses a significant threat to the integrity of democratic processes and to national security.[373] These technological advancements have outpaced traditional methods of verification and fact-checking, creating a landscape where misinformation can rapidly influence public opinion and election outcomes.[374] For example, a 2022 deepfake video of Ukrainian President Volodymyr Zelenskyy asking his troops to surrender created understandable

[370] Laura Barrón-López et al., *Women Face New Sexual Harassment with Deepfake Pornography,* PBS (June 27, 2023, 6:30 PM), https://www.pbs.org/newshour/show/women-face-new-sexual-harassment-with-deepfake-pornography [https://perma.cc/9KAS-DEUV].

[371] *Id.*

[372] Jeff Hearn et al., *The Spread of Digital Intimate Partner Violence: Ethical Challenges for Business, Workplaces, Employers and Management*, 187 J. BUS. ETHICS 695, 697–98 (2023).

[373] *See generally* Daniel Pereira, DEEP FAKES AND NATIONAL SECURITY, OODA LOOP (July 11, 2023), https://www.oodaloop.com/archive/2023/07/11/deep-fakes-and-national-security/ [[https://perma.cc/45S8-4HPR] (discussing the public concern about bad actors nefariously using deepfakes to "erode public trust, negatively affect public discourse, or even sway an election").

[374] Daniel I. Weiner & Lawrence Norden, *Regulating AI Deepfakes and Synthetic Media in the Political Arena* (Dec. 5, 2023), https://www.brennancenter.org/our-work/research-reports/regulating-ai-deepfakes-and-synthetic-media-political-arena [https://perma.cc/GDY4-3PER].

confusion.[375] Deepfake audio was also used in Slovakia's 2023 election, falsely depicting a political leader.[376] It was conveniently released during the 48 hours before the election and under the law, the press was required to remain silent. This technological evolution necessitates an urgent and thoughtful regulatory response, especially because over 2 billion people in over 50countries will vote in elections in 2024.[377] It only took a few dollars and a few minutes for a democratic operative opposed to the use of AI in politics to create a deepfake robocall using the voice of President Joe Biden telling people not to vote in the New Hampshire presidential primary.[378]

While the need to protect the electorate from deceptive content is clear, it is equally important to ensure that regulatory measures do not infringe on First Amendment rights, particularly when manipulated content serves legitimate purposes such as satire or political commentary.[379] In

[375] *Id.*

[376] Morgan Meaker, *Slovakia's Election Deepfakes Show AI Is a Danger to Democracy*, WIRED (Mar. 18, 2023, 12:00 PM))), https://www.wired.co.uk/article/slovakia-election-deepfakes [https://perma.cc/37HE-XPBP] (reporting on the use of deepfake audio falsely depicting a Slovakian politician, Michal Šimečka, just two days before the country's elections).

[377] Heather Ashby, *Disinformation Casts Shadow Over Global Elections*, U.S. INST. OF PEACE (January 3, 2024), https://www.usip.org/publications/2024/01/disinformation-casts-shadow-over-global-elections [https://perma.cc/T2VC-PUFV].

[378] Alex Seitz-Wald, *Democratic operative admits to commissioning fake Biden Robocall that used AI*, NBC NEWS (2024), https://www.nbcnews.com/politics/2024-election/democratic-operative-admits-commissioning-fake-biden-robocall-used-ai-rcna140402 (last visited Mar 31, 2024).

[379] *See generally* Shannon Reid, *The Deepfake Dilemma: Reconciling Privacy and First Amendment Protections*, 23 U. PA. J. CONST. L. 209 (2021) (considering how federal courts could balance First Amendment protections of creators while concurrently allowing those harmed by deepfakes to defend their personality and reputation).

addition, some candidates are employing AI for legitimate purposes to replicate their own voices in different languages so that they can expand their reach.[380]

Regulators may want to consider mandatory labeling of AI-generated content. However, transparency alone may not be sufficient in all cases, and more stringent measures, like outright bans on certain forms of deceptive media, may be necessary. This approach is exemplified by laws in a handful of states, which ban certain types of deepfakes during election periods. Regulators should also consider holding candidates, PACs, and others who create and disseminate deepfakes and similar content related to elections liable, including online social media platforms.[381] Google and Meta have volunteered to adopt disclaimer requirements, but that may not suffice at a time when social media companies have laid off content moderators.[382] Further, even the US government has acknowledged that deepfake detectors may not be effective and that the deepfakes themselves may cause lasting damage.[383]

[380]Ashby, *supra* note 377.

[381] Weiner & Norden, *supra* note 374.

[382] Ali Swenson & Christine Fernando, *As social media guardrails Social Media Guardrails Fade and AI deepfakes go mainstream, experts Deepfakes Go Mainstream, Experts Warn of Impact on Alections*, PBS (Dec. 27, 2023, 5:27 PM), https://www.pbs.org/newshour/politics/as-social-media-guardrails-fade-and-ai-deepfakes-go-mainstream-experts-warn-of-impact-on-elections [https://perma.cc/RX2D-966A].

[383] U.S. GOV'T ACCOUNTABILITY OFF., *Science & Tech Spotlight: Combating Deepfakes*, U.S. GAO, https://www.gao.gov/products/gao-24-107292 (last visited Mar 31, 2024).

In this Part II, we have provided a high-level overview of the incredible potential of AI and its potential existential threat to our way of life. We will now turn to how governments are rising to the challenge of regulating in 2024.

III. Governmental Responses to the Rise of AI

We have come a long way from the days of ELIZA, the program which managed to fool its users in the 1960s into believing that it had emotions.[384] Mustafa Suleyman is an AI pioneer who left Google to form Inflection AI, which is creating an emotionally-intelligent personal assistant known as Pi, designed to "become a chief of staff for your life . . . to help you organize, prioritize, think, create, and plan y,our day."[385] Acknowledging that AI could help someone learn to make a bomb, Suleyman who in March 2024 joined Microsoft, OpenAI's largest investor,[386] nonetheless remains optimistic about AI's, opining that "this is going to end up being one of the

[384] *See* Caleb Sponheim, *The ELIZA Effect - Why We Love AIELIZA Effect: Why We Love AI*, NIELSEN NORMAN GRP. (Oct. 6, 2023), https://www.nngroup.com/articles/eliza-effect-ai/ [https://perma.cc/TFA9-AV33].

[385] Mark Sullivan, *Why DeepMind Cofounder Mustafa Suleyman Left Google to Start a Human-Focused AI Company*, FAST CO. (Sept. 29, 2023), https://www.fastcompany.com/90959853/mustafa-suleyman-inflection-pi [[https://perma.cc/KGT6-GDHU].

[386] Dan Milmo*, Mustafa Suleyman: The new head of Microsoft AI with concerns about his trade*, GUARDIAN (2024), https://www.theguardian.com/business/2024/mar/20/mustafa-suleyman-the-new-head-of-microsoft-ai-with-concerns-about-his-trade (last visited Mar 31, 2024).

most productive periods in the history of our species, if not the most productive."[387]

In the tech subculture there are two splinter schools of thought regarding AI. Effective accelerationists or "e/accs" want AI development and deployment to happen as quickly and with as few guardrails as possible.[388] To them, rigid regulation will stifle life-changing technological advances.[389] This contrasts with an earlier movement in AI known as effective altruism, which originally focused on data-driven approaches to philanthropy but which now shines the light on safety concerns and the very real threats to humanity if development is left unregulated.[390] Some believe that OpenAI CEO's Sam Altman's sudden termination and subsequent sudden rehire in November 2023, which shocked the AI world and almost led to billions in lost value for the company, was due to the effective altruists on the board

[387] Washington Post Live, *Transcript: The Path Forward: Artificial Intelligence with Mustafa Suleyman*, WASH. POST (Sept. 7, 2023, 1:41 PM), https://www.washingtonpost.com/washington-post-live/2023/09/07/transcript-path-forward-artificial-intelligence-with-mustafa-suleyman/ [https://perma.cc/7WCA-EVYF].

[388] Kevin Roose, *This A.I. Subculture's Motto: Go, Go, Go*, N.Y. TIMES (Dec. 10, 2023), https://www.nytimes.com/2023/12/10/technology/ai-acceleration.html [https://perma.cc/TZU8-THA2].

[389] *Id.*

[390] *Id.*; Eric Levitz, *Why Effective Altruists Fear the AI Apocalypse : A Conversation with the Philosopher William MacAskill*, INTELLIGENCER (Aug. 30, 2022), https://nymag.com/intelligencer/2022/08/why-effective-altruists-fear-the-ai-apocalypse.html [https://perma.cc/E2RH-J5JZ].

who had lost faith in the CEO for not being candid, likely about risks related to AI.[391]

Philosopher William MacAskill has coined the phrase "longtermism" to highlight the "moral obligation" to protect the future of humanity.[392] He warns that unchecked AI "could abet a global totalitarian dictatorship or decide to treat humanity like obsolete software — and delete us from the planet."[393] "Decels" or "doomers," a more pejorative term for the effective altruists, seek more regulation and guardrails and use the most explicit possible language about the technology's potential dangers.[394] E/accs such as influential venture capitalist Marc Andreesen have gone as far as to proclaim that "any deceleration of AI will cost lives . . . . Deaths that were preventable by the AI that was prevented from existing is a form of murder."[395]

[391] *See generally* Louise Matsakis & Reed Albergotti, *The AI Industry Turns Against Its Favorite Philosophy*, SEMAFOR (Nov. 21, 2023, 3:29 PM) (explaining that effective altruism played a role in the abrupt firing of Sam Altman—threatening OpenAI's entire existence), https://www.semafor.com/article/11/21/2023/how-effective-altruism-led-to-a-crisis-at-openai [https://perma.cc/4SEM-9S2W].
[392] Levitz, *supra* note 390.
[393] *Id.*
[394] Roose, *supra* note 388.
[395] Tara Suter, *Andreessen Warns Any AI 'Deceleration' Will 'Cost Lives'*, THE HILL (Sept. 10, 2023, 4:35 PM), https://thehill.com/policy/technology/4261105-andreessen-ai-warning-cost-lives/.

Former Chief Business Officer of Google X Mo Gawdat, who has warned that AI could be a bigger danger than climate change,[396] sums up the urgency to be thoughtful about AI this way:

> So that *kind* of intelligence that would take an infant 5 years to grasp and then 10 more years to make it effective in the real world, that would take a government agency or a regulator 5 years to recognizerecognise [*sic*] and then 10 years of meetings to talk about it and then 100 years to do something about it. In the case of a quantum computing powered AI, it might take a microsecond.[397]

Although we believe that the potential benefits outweigh the many risks involved in the development and deployment of AI, we also believe that robust, enforceable regulation across industries and geographies is vital and that the time to regulate is now.

Unfortunately, as of March 2024, the United States has not enacted comprehensive federal regulation and it is one of the dozens of nations undergoing a presidential election, which could mean a change in leadership or priorities depending on who is elected. The EU has outlined its legislation in broad strokes. The majority of the legislation will not come into force until 2026, although the prohibitions will apply after six months and the general-

[396] The Diary Of A CEO, *EMERGENCY EPISODE: Ex-Google Officer Finally Speaks Out on the Dangers of AI! – Mo Gawdat*, YOUTUBE (Jun. 1, 2023), https://www.youtube.com/watch?v=bk-nQ7HF6k4 [https://perma.cc/24XU-R5HV].

[397] Dr. Vikas Shah, *The Future of Artificial Intelligence, A Conversation with Mo Gawdat, Author of Scary Smart*, THOUGHT ECON. (Oct. 19, 2023), https://thoughteconomics.com/mo-gawdat/ [https://perma.cc/9XFB-QBXJ].

purpose AI rules will become effective after twelve months. China, the UK, Canada, and other nations are working on their own legislation. But without a common global framework, there are significant risks that will outweigh the benefits. Below, we will briefly examine the state of legislation as of the time of this writing.

European Union

In 2019, the High-Level Expert Group (HEGLHLEG) on Artificial Intelligence, a group established by the European Commission, published its Policy and Investment Recommendations for Trustworthy AI to the European Commission and Member States.[398] Along with over thirty recommendations, the HEGLHLEG identified eleven key takeaways to "guide European AI towards sustainability, growth and competitiveness."[399] Those takeaways are to "[e]mpower and protect humans and society;" "[t]ake up a tailored approach to the AI landscape;" "[s]ecure a Single European Market for Trustworthy AI;" "[e]nable AI ecosystems through Sectoral Multi-Stakeholder Alliances;" "[f]oster the European data economy;" "[e]xploit the multi-faceted role of the public sector;" "[s]trengthen and unite Europe's research capabilities;" "[n]urture education to the Fourth Power;" "[a]dopt a risk-based governance approach to AI and ensure an appropriate

[398] Mauritz Kop, *AI & Intellectual Property: Towards an Articulated Public Domain*, 28 TEX. INTELL. PROP. L. J. 297, 331– 37 (2020).
[399] *Id.* at 337.

regulatory framework;" "[s]timulate an open and lucrative investment environment;" and "[e]mbrace a holistic way of working, combing a 10-year vision with a rolling action plan."[400]

In early 2021, the EU Parliament proposed the AI Act, a regulatory framework for development, deployment, and the use of AI within the EU.[401] The EU Parliament's priority is to ensure that AI systems being used in the region are "safe, transparent, traceable, non-discriminatory and environmentally friendly."[402] The Act aims to make sure that AI systems are monitored by humans, rather than technology, to avoid detrimental outcomes.[403] Members of the European Parliament adopted Parliament's negotiating position in June 2023.[404] In December 2023, EU policymakers agreed to the AI Act,[405] and then ratified the Act.[406] The EU is now the standard bearer in artificial intelligence regulation. As with all EU-wide policies and programs, and as required by Article 2 of the Treaty of the

[400] *Id.*

[401] *EU AI* ACT: F*irst Regulation on Artificial Intelligence,* EUR. PARL. (Dec. 19, 2023, 11:45 AM), https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence [https://perma.cc/N4KR-XAWY]; Fieldfisher Silicon Valley, *Generative AI: Privacy Risks & Challenges*, YOUTUBE (Mar. 9, 2023), https://www.youtube.com/watch?v=9xjFsy9_HBs [https://perma.cc/P28M-YP8C].

[402] *EU AI Act: First Regulation on Artificial Intelligence*, *supra* note 401.

[403] *Id.*

[404] *Id.*

[405] Usman Wahid, *The EU AI Act Gets a Green Light: Deal-Readiness and Compliance Roadmap*, KPMG (Dec. 12, 2023), https://kpmg.com/uk/en/home/insights/2023/12/eu-ai-act-gets-a-green-light.html [https://perma.cc/Z6SA-6EUP].

[406] *Artificial Intelligence Act: MEPs adopt landmark law*, *supra* note 56.

European Union, the legislation was intentionally designed to respect the "fundamental" rights of EU citizens, which are: "respect for human dignity, freedom, democracy, equality, the rule of law, and respect for human rights, including the rights of persons belonging to minorities."[407] Article 6 of the Treaty on European Union requires that "[t]he Union recognize[]recognises [*sic*] the rights, freedoms and principles set out in the Charter of Fundamental Rights of the European Union . . .[. . .] which shall have the same legal value as the Treaties."[408] The Charter, which has been legally binding since 2009, is primary EU law; therefore, it is a standard "for examining the validity of secondary EU legislation and national measures."[409]

The EU AI Act will likely have the same "Brussels effect" as the GDPR privacy regulation.[410] The author who coined the term "Brussels effect" describes it as the

> EU's unilateral ability to regulate global markets by setting the standards in competition policy, environmental protection, food safety, the protection of privacy, or the regulation of hate speech in social media. Interestingly, the EU doesn't need to impose its standards coercively on anyone[—]market forces alone are sufficient. In fact, the EU is one of the largest and wealthiest consumer markets,

[407] *Protecting Fundamental Rights Within the Union*, EUR. PARL., https://www.europarl.europa.eu/about-parliament/en/democracy-and-human-rights/fundamental-rights-in-the-eu (last visited Feb. 15, 2024).

[408] *Fundamental Rights*, EUR. PARL. (alteration in original), https://www.europarl.europa.eu/committees/en/fundamental-rights/product-details/20160229CDT00541 [https://perma.cc/3LMK-3NNW] (last visited Feb. 15, 2024).

[409] *Id.*

[410] ANDREA RENDA, FOUND. FOR EUR. PROGRESSIVE STUD., BEYOND THE BRUSSELS EFFECT: LEVERAGING DIGITAL REGULATION FOR STRATEGIC AUTONOMY 4 (2022).

> supported by strong regulatory institutions. There are few global companies that can afford not to trade in the EU, and the price for accessing the single market is adjusting their conduct and production to EU standards, which are often the most stringent standards globally. Importantly, often these firms choose to abide by the same rules across other markets too, so as to avoid the cost of complying with different regulatory regimes.[411]

The Act categorizes AI systems into their levels of risk and associated regulation.[412] The most regulated category is deemed "unacceptable risk."[413] Systems included in this category are those that are "considered a threat to people and will be banned."[414] Such systems include those that have the ability to manipulate peoples' cognitive behavior, engage in social scoring ("classifying people based on [behavior], socio-economic status[,] or personal characteristics"), and "[r]eal-time and remote biometric identification systems, such as facial recognition."[415] There will be some exceptions to the blanket ban on these systems, such as biometric systems that can assist serious criminal prosecutions, if approved by a court.[416]

[411] Anu Bradord, *The European Union in the Globalized World: The "Brussels Effect",* GROUPE D'ÉTUDES GÉOPOLITIQUES, https://geopolitique.eu/en/articles/the-european-union-in-a-globalised-world-the-brussels-effect/ [https://perma.cc/7LHJ-9HBU] (last visited Feb. 15, 2024).

[412] *EU AI Act: First Regulation on Artificial Intelligence*, *supra* note 401.

[413] *Id.*; Wahid, *supra* note 405.

[414] EU AI ACT: FIRST REGULATION ON ARTIFICIAL INTELLIGENCE, *supra* note 401; Wahid, *supra* note 405.

[415] *EU AI Act: First Regulation on Artificial Intelligence*, *supra* note 401.

[416] *Id.*

Next is the high risk category.[417] Systems that fall into this category are those that affect fundamental rights and/or safety.[418] This category is further divided into two categories.[419] The first category consists of "AI systems that are used in products falling under the EU's product safety legislation."[420] Products in this category may include cars, medical devices, and toys among others.[421] The second consists of AI systems falling into seven specific areas that must be registered in an EU database: "[m]anagement and operation of critical infrastructure"; "[e]ducation and vocational training"; "[e]mployment, worker management and access to self-employment"; "[a]ccess to and enjoyment of essential private services and public services and benefits"; "[l]aw enforcement"; "[m]igration, asylum and border control management"; and "[a]ssistance in legal interpretation and application of the law."[422] High-risk AI systems "will be assessed before being put on the market and also throughout their lifecycle."[423]

The last risk category is limited risk (or low-risk).[424] The systems that fall into this category ought to "comply with minimal transparency

[417] *Id.*; Wahid, *supra* note 405.

[418] *EU AI Act: First Regulation on Artificial Intelligence*, *supra* note 401.

[419] *Id.*

[420] *Id.*

[421] *Id.*

[422] *Id.*

[423] *Id.*

[424] *Id.*; Wahid, *supra* note 405.

requirements that would allow users to make informed decisions."[425] The regulation of systems in this category depends on users making their own decisions about whether or not they want to choose to use the system.[426] However, one of those transparency requirements is that users must be made aware when they are interacting with an AI system.[427] "This includes AI systems that generate or manipulate image, audio, or video content, for example deepfakes."[428] Similarly, generative AI must comply with various transparency requirements, including "[d]isclosing that the content was generated by AI, "[d]esigning the model to prevent it from generating illegal content," and "[p]ublishing summaries of copyrighted data used for training."[429]

Certain aspects of AI are excluded from the Act, including: free and open-source software unless they are a "high-risk" system, prohibited application, or could cause manipulation; free and open-source models whose parameters are made publicly available, except for what concerns implementing a policy to comply with copyright law;[430] anything that would affect a member states' national security, military, or defense abilities; and AI

[425] *EU AI Act: First Regulation on Artificial Intelligence*, *supra* note 401.
[426] *Id.*
[427] *Id.*
[428] *Id.*
[429] *Id.*
[430] *Id.*

systems used for the sole purpose of research and innovation, or for people using AI for "non-professional" reasons.[431] Violators will face steep penalties of "€35 million or 7% of global turnover for violations of the banned AI applications, €15 million or 3% for violations of the AI Act's obligations and €7,5 million or 1.5% for the supply of incorrect information."[432] A "provisional agreement provides for more proportionate cap**s** on administrative fines for [small and medium sized enterprises] and start-ups.[433]

The provisional agreement on the EU AI law marks a significant step in addressing the ethical and rights-related concerns associated with high-risk AI systems. By mandating a fundamental rights impact assessment[434] before market deployment, the agreement acknowledges the potential risks these technologies pose and aims to mitigate them proactively. Enhanced transparency is another cornerstone of this agreement, especially with the amendment requiring certain public entities using high-risk AI systems to register in an EU-specific database.[435] This not only promotes accountability but also facilitates oversight of AI applications in public domains.

[431] Council of the European Union Press Release 986/23, Artificial Intelligence Act: Council and Parliament Strike a Deal on the First Rules for AI in the World (Feb. 2, 2024).
[432] *Id.*
[433] *Id.* (boldface type omitted).
[434]*Fundamental Rights and Algorithms Impact Assessment (FRAIA)*, OECD.AI (Apr. 5, 2023), https://oecd.ai/en/catalogue/tools/fundamental-rights-and-algorithms-impact-assessment-%28fraia%29 [https://perma.cc/3RUX-UL4M].
[435] Council of the European Union Press Release, *supra* note 431.

Furthermore, the introduction of provisions related to emotional recognition systems which mandates user disclosure to impacted individuals, underscores a commitment to personal privacy and informed consent.

At the heart of the AI Act's governance structure will be the AI Office, a central regulatory authority charged with the critical task of overseeing the Act's implementation, particularly focusing on advanced AI models.[436] This office will be pivotal in ensuring that the Act's provisions are enforced uniformly across member states, thereby guaranteeing consistency in both application and interpretation. This central oversight body will play a key role in maintaining accountability and ensuring strict adherence to the Act. However, the structure may face challenges in inter-agency coordination and overlap, requiring seamless collaboration and clear delineation of functions among the various entities to avoid inefficiencies and conflicting directives.

Complementing the AI Office will be a scientific panel of independent experts, a group endowed with the responsibility of providing specialized advice on General Purpose AI (GPAI) models.[437] Their expertise will extend to offering counsel, developing methodologies for evaluating foundational AI models, and vigilantly monitoring potential safety risks

[436] *Id.*

[437] *Id.*

associated with these technologies.[438] The inclusion of this panel is supposed to ensure that the Act's implementation is not only regulatory but also deeply rooted in scientific and technical expertise, offering a balanced and well-informed approach to AI governance. A potential gap here is the need for adaptability to rapid technological changes, as AI evolves swiftly and the regulatory framework must be flexible to keep pace.

Further enhancing this governance structure will be the AI Board, composed of representatives from member states.[439] This Board will serve as a vital "coordination platform and an advisory body to the Commission," facilitating the exchange of best practices and harmonizing AI regulations across different jurisdictions.[440] Its role is crucial in shaping the implementation of the Act, providing a collaborative space for member states to contribute insights on "the design of codes of practice for foundational models."[441]

Additionally, an advisory forum will be established, bringing together diverse stakeholders from businesses, small and medium-sized enterprises, start-ups, civil society, and academia. This Forum will be tasked with providing technical expertise to the AI Board, ensuring that the

[438] *Id.*
[439] *Id.*
[440] *Id.*
[441] *Id.*

implementation of the Act reflects a wide range of perspectives and experiences, thereby fostering a comprehensive and inclusive approach to AI regulation.[442] But despite these structures, there might be gaps in representation and diversity, with a need to ensure that all interests and perspectives, especially those of marginalized groups, are adequately represented.

Moreover, the governance framework might face challenges in transparency and accountability, global alignment and international cooperation, resource allocation and expertise, as well as in developing robust enforcement and compliance mechanisms. Addressing these potential gaps is essential for creating an effective governance structure that can keep pace with the advancements in AI technology while safeguarding ethical standards and societal interests.

The European Union's new AI law has been met with a chorus of criticism from various experts and organizations, highlighting significant concerns about its effectiveness and potential implications. Daniel Leufer of Access Now points out critical flaws in the legislation, emphasizing loopholes for law enforcement and insufficient protections in migration contexts.[443] He

[442] *Id.*

[443] Daniel Leufer, et al., *Human Rights Protections...with Exceptions: What's (Not) in the EU's AI Act Deal*, ACCESSNOW (Dec. 1415, 2023), https://www.accessnow.org/whats-not-in-the-eu-ai-act-deal/ [https://perma.cc/65UQ-XACE].

also notes that the law fails to adequately address opt-outs for developers and contains substantial gaps in the bans on the most dangerous AI systems.[444] Daniel Castro from the Information Technology and Innovation Foundation (ITIF) and Cecilia Bonefeld-Dahl of DigitalEurope express concerns about the legislative approach and its timing. Castro argues that the rapidly evolving nature of AI warranted a more cautious legislative approach, suggesting that the law might lead to unintended consequences that are harder to rectify than the technology itself. Bonefeld-Dahl criticizes the last-minute changes to the legislation, particularly regarding the regulation of foundation models, implying a departure from a risk-based approach centered on AI applications.[445] Ella Jakubowska from the European Digital Rights and Mher Hakobyan from Amnesty International further underscore these concerns.[446] They highlight the law's provisions on facial recognition technology, arguing that it fails to offer strong protections and could lead to significant human rights violations.[447] As Hakobyan observed, "[n]ot ensuring a full ban on facial recognition is therefore a hugely missed opportunity to stop and

[444] *Id.*

[445] *Id.*; Foo Yun Chee et. al., *Europe Agrees Landmark AI Regulations Deal*, REUTERS (Dec. 11, 2023, 11:29 AM), https://www.reuters.com/technology/stalled-eu-ai-act-talks-set-resume-2023-12-08/ [https://perma.cc/EQT5-33NM].

[446] *Id.*

[447] *Id.*; *EU: Bloc's Decision to Not Ban Public Mass Surveillance in AI Act Sets a Devastating Global Precedent*, AMNESTY INT'L(Dec. 9, 2023), https://www.amnesty.org/en/latest/news/2023/12/eu-blocs-decision-to-not-ban-public-mass-surveillance-in-ai-act-sets-a-devastating-global-precedent/ [https://perma.cc/TQ27-N6YY].

prevent colossal damage to human rights, civic space and rule of law that are already under threat throughout the EU.[448] Lawmakers also failed to ban the export of harmful AI technologies, including for social scoring, which would be illegal in the EU.[449] Allowing European companies to profit from technologies that the law recognizes impermissibly harms human rights in their home states and establishes a dangerous double standard."[450]

A scholar reviewing an earlier version of the EU AI Act raised some concerns of risk-based regulation, and their comments are instructive as the EU continues to refine the legislation.[451] They noted that it may not properly account for core underlying concepts in machine learning and distinct liability rules for those who develop and those who use AI systems.[452] This differentiation would address the varying degrees of control and understanding each party has over the AI's functioning. The author proposes strict standards for data quality, holding AI developers and users accountable for the data they use.[453] This includes mechanisms to ensure data accuracy and mechanisms to mitigate biases.[454] They further recommend developing clear guidelines for how humans should interact with and oversee AI systems,

[448] *Id.*
[449] *Id.*
[450] *Id.*
[451] Margot E. Kaminski, *Regulating the Risks of AI*, 103 B.U. L. REV. 1347, 1366–89 (2023).
[452] *See id.*
[453] *Id.* at 1376–77.
[454] *Id.*

balancing the need for human control with the benefits of AI autonomy. Finally, the author advocates for a regulatory framework that is regularly reviewed and adapted to keep pace with technological advancements in AI.[455]

As regulators finalize the legislation, they must consider whether any law can handle the complex and rapidly evolving landscape of AI and whether it will potentially lead to more harm than good in both the digital and human rights domains. The United States should consider these critiques when developing its own regulation.

United Kingdom

The United Kingdom (UK), which is not a member of the EU, is also proposing a regulatory framework for AI.[456] The UK has chosen not to implement codified regulation initially, in an effort to not place undue burden on businesses and instead is opting to empower regulators to provide expert guidance and encourage compliance.[457] The framework aims to "bring clarity and coherence to the AI regulatory landscape," and "to make responsible

[455] *Id.*

[456] *Establishing a Pro-Innovation Approach to Regulating AI*, GOV.UK (Jul. 20, 2022), https://www.gov.uk/government/publications/establishing-a-pro-innovation-approach-to-regulating-ai/establishing-a-pro-innovation-approach-to-regulating-ai-policy-statement [https://perma.cc/ZV86-RGRZ].

[457] Leigh Smith & Richard Ward, *The UK Government Has Published Its Response to Its Consultation on Artificial Intelligence and Intellectual Property*, JD SUPRA (Jul. 6, 2022), https://www.jdsupra.com/legalnews/the-uk-government-has-published-its-1249002/ [https://perma.cc/DJ76-GP6L].

innovation easier."[458] The patchwork framework involves guidance and working proposals from various regulatory bodies, including the Information Commissioner's Office (ICO), Competition and Markets Authority (CMA), and the Digital Regulation Cooperation Forum (DRCF), that assist companies that are using AI.[459] Due to the various regulatory bodies issuing patchworks of guidance, where a risk falls into a gap in the regulations, the regulatory bodies and the government will work together to identify a way to fill the gap, whether through legislative initiatives or changes to regulations.[460] Regulators are expected to collaborate proactively on behalf of society and the economy to achieve optimal outcomes.[461]

This framework aims to maximize the opportunities and benefits of AI, while promoting safety and risk mitigation, by building trust.[462] The regime is based on core principles: safety, security and robustness; appropriate transparency and explainability; fairness; accountability and governance; and contestability and redress.[463] The framework prioritizes innovation and coherence, asking regulators to interpret, implement, and

[458] *Id.*; *A Pro-Innovation Approach to AI Regulation*, GOV.UK (Aug. 3, 2023), https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper [https://perma.cc/P58X-RA3A].

[459] Wahid, *supra* note 405.

[460] *A Pro-Innovation Approach to AI Regulation*, *supra* note 458.

[461] *Id.*

[462] *Id.*; Wahid, *supra* note 405.

[463] *A Pro-Innovation Approach to AI Regulation*, *supra* note 458.

prioritize these principles.[464] The approach aims to be adaptable in a continually evolving landscape.[465] The framework also recognizes the international, borderless digital ecosystem, accounting for partnership and cooperation with other key players around the globe, including through the Council of Europe, OECD working groups, the Global Partnership on AI, and through global standards bodies such as ISO and IEC.[466] However, the responsibility ultimately falls on the government to ensure that the regulatory framework is operating in the way that it was designed to, prioritizing innovation and proportionate responses, and therefore retains the ability "to put mechanisms in place to coordinate, monitor and adapt the framework as a whole."[467]

One key element of the proposed framework is that it defines AI based on its functional capabilities, its adaptability, and how autonomous it is.[468] Using a definition of AI with reference to the functional capabilities and design of the program allows the framework to address the specific challenges of programs and to "future-proof" the framework against unforeseen challenges.[469]

[464] *Id.*
[465] *Id.*
[466] *Id.*
[467] *Id.*
[468] *Id.*
[469] *Id.*

The priority of the regulation is to "keep pace with and respond to the new and distinct challenges and opportunities posed by AI" as well as remain internationally competitive.[470] The proposal emphasizes innovation and is supported by cross-sectoral principles that are tailored to the specific challenges of AI.[471] The regulation has several key characteristics, including being "pro-innovation, proportionate, trustworthy, adaptable, clear, and collaborative."[472] The proposed regulation is context-specific, based on use and impact on individuals, groups, and businesses.[473] The framework focuses on combating issues where there is evident real risk which "ask[s] that regulators focus on high risk concerns rather than hypothetical or low risks."[474]

One interesting aspect to note about AI policy in the UK is that the UK is currently one of the only jurisdictions that does grant copyright protection to computer-generated works.[475] "The author of such works is deemed to be the person who made the arrangements necessary for the creation of that work," i.e. the programmer or prompter of the AI model.[476] The same is not true for patents, however. The UK does not grant patent

[470]*Id.; Establishing a Pro-Innovation Approach to Regulating AI*, *supra* note 458.

[471] *Id.*

[472] *Id.; A Pro-Innovation Approach to AI Regulation*, *supra* note 458 (boldface type omitted).

[473] *Id.*

[474] *Id; Establishing a Pro-Innovation Approach to Regulating AI*, *supra* note 458.

[475] Smith & Ward, *supra* note 457.

[476] *Id.*

ownership to AI generated inventions:[477] "[I]n the UK an invention can only be subject to patent protection where a human is identified as the [inventor] of the invention."[478]

The UK's AI Safety Summit, held in November 2023, was attended by twenty-nine countries that signed the Bletchley Declaration.[479] The Summit aimed to engage with the international community and all of the various stakeholders.[480] It had two main areas of focus——misuse risks and loss of control risks—and several objectives including a shared understanding of risks, international collaboration, increased user safety, research collaboration, and ensuring safe development.[481] The Declaration states that the signing counties recognized "that the protection of human rights, transparency and explainability, fairness, accountability, regulation, safety, appropriate human oversight, ethics, bias mitigation, privacy and data protection needs to be addressed. [They] also note[d] the potential for unforeseen risks stemming from the capability to manipulate content or generate deceptive content. All of these issues are critically important and

[477] *Id.*

[478] *Id.*

[479] *The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023*, GOV.UK (Nov. 1, 2023),), https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023 [http://perma.cc/2VBB-FGM3].

[480] *Id.*

[481] *AI Safety Summit: Introduction*, *(HTML)*, GOV.UK (Oct. 31, 2023), https://www.gov.uk/government/publications/ai-safety-summit-introduction/ai-safety-summit-introduction-html [https://perma.cc/3ECL-5Q7W].

[they] affirm[ed] the necessity and urgency of addressing them."[482] The risks arising from AI's borderless nature make them inherently international, and will continue to be best addressed via international cooperation.[483]

The United Kingdom's regulatory framework for AI, emphasizing flexibility and adaptability, contrasts with the EU's risk-based approach. The UK's strategy, focusing on guidance over stringent regulation, seeks to stimulate innovation. This approach recognizes the challenges in legislating a constantly evolving technology. However, it may lead to regulatory inconsistencies, particularly for multinational companies that operate in both the EU and the UK. While the UK model promotes technological advancement and adaptability, a risk-based approach provides clearer, more consistent guidelines, potentially offering greater predictability for businesses and consumers.

Hypothetical scenarios illustrate these differences. A U.S. firm developing an AI-driven diagnostic tool might find the UK's adaptable guidelines conducive to innovation but may struggle with the lack of clear directives, especially when considering the stringent FDA regulations, which are more aligned with the EU's risk-based approach. A fintech startup using AI for credit scoring would benefit from the UK's flexible framework,

[482] *The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023*, *supra* note 479.
[483] *Id.*

allowing for rapid adaptation to new data patterns. In contrast, a risk-based approach might impose strict compliance requirements, ensuring consumer protection but potentially slowing innovation. An American company developing autonomous vehicles might find the UK model's emphasis on innovation appealing but could face challenges in a risk-based regulatory environment that prioritizes specific safety standards and liability issues.

Thus, while the UK's approach offers a model that emphasizes flexibility and innovation, its application for multinational organizations necessitates careful consideration of consistency and clarity in regulation. U.S. legislators should weigh the benefits of an adaptable, innovation-friendly approach against the predictability and specificity offered by a risk-based model. As AI continues to evolve, finding the right balance in regulatory frameworks will be key to leveraging its benefits while ensuring safety and ethical considerations.

Canada

The draft Canadian Artificial Intelligence and Data Act (AIDA)[484] establishes a structured framework for managing the use of AI technologies, emphasizing the balance between innovation and public trust, and is not

[484] *Artificial Intelligence and Data Act (AIDA) – Companion Document*, GOV'T. OF CAN. (Mar. 13, 2023), B. C-27, 44th Parliament, 1st Sess. (2022) (Can.) https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document [https://perma.cc/6KGU-VCPD] [hereinafter AIDA Companion Document].

expected to come into law until 2025.[485] As of January 2024, there is a voluntary code of conduct in place.[486] The draft Act classifies AI systems into four distinct categories, each addressing specific concerns and potential risks.[487]

Screening systems for access to services or employment may lead to discriminatory outcomes, particularly affecting women and marginalized groups, because they may replicate or amplify existing biases in data or decision-making processes.[488] Biometric systems for identification and inference pose risks over mental health, personal autonomy, privacy invasion, and the potential for misinterpretation of biometric data leading to erroneous conclusions about individuals.[489] Systems influencing human behavior at scale involves AI-powered systems such as online content recommendation tools, which can extensively affect human behavior, expression, and emotion.[490] Systems critical to health and safety are integrated into crucial areas such as autonomous driving and healthcare triage.[491] The risks include

[485] *Id.*

[486] *Voluntary Code of Conduct on the Responsible Dev. and Mgmt.. of Advanced Generative A.I. Sys*, GOV'T OF CAN.. (September 2023), https://ised-isde.canada.ca/site/ised/en/voluntary-code-conduct-responsible-development-and-management-advanced-generative-ai-systems [https://perma.cc/Z5DT-CK9E].

[487] AIDA Companion Document, *supra* note 484.

[488] *Id.*

[489] *Id.*

[490] *Id.*

[491] *Id.*

direct physical harm from malfunctions or inadequate design, and biases leading to unequal or unsafe outcomes if not properly mitigated.[492]

AIDA also addresses two main types of adverse impacts of these high-impact AI systems: individual harms (physical, psychological, property damage, or economic loss) and systemic bias (collective harm impacting groups, often exacerbating severity for vulnerable populations like children).[493] The Act defines biased output as unjustified adverse impact based on prohibited discrimination grounds, as outlined in the Canadian Human Rights Act.[494] It mandates proactive bias risk assessment and mitigation, expanding on individual legal resources under current human rights legislation.[495]

Businesses under AIDA have specific regulatory requirements "to identify, assess, and mitigate risks" before deploying high-impact AI systems.[496] These requirements ensure compliance throughout the AI system's lifecycle.[497] Key guiding principles for high-impact AI systems include: human oversight and monitoring, transparency, fairness and equity, safety, accountability, and validity and robustness.[498] Responsibilities vary

[492] *Id.*
[493] *Id.*
[494] *Id.*
[495] *Id.*
[496] *Id.*
[497] *Id.*
[498] *Id.*

depending on the business role. Designers and developers must address risks related to harm and bias by documenting appropriate use and continually adjusting risk mitigation measures.[499] Distributors should consider potential deployment scenarios, informing users of limitations and restrictions.[500] Operational Managers are responsible for using the AI system as intended, continually assessing and mitigating risks, and ensuring ongoing monitoring.[501]

The AIDA recognizes different obligations based on business involvement in the AI system's lifecycle.[502] It separates the responsibilities of design, development, and operational management, ensuring a comprehensive approach to AI governance.[503] Additionally, the Act stipulates notification requirements for businesses in cases of material harm caused by high-impact AI systems, underlining the importance of proactive risk management and reporting.[504]

In sum, AIDA's effort to categorize AI systems into four distinct types is a good start, but this segmentation oversimplifies the risks inherent in AI technologies.[505] For instance, a biometric system used for employment

[499] *Id.*
[500] *Id.*
[501] *Id.*
[502] *Id.*
[503] *Id.*
[504] *Id.*
[505] *Id.*

screening might blur the lines between screening systems and biometric systems, thereby raising both discrimination and privacy concerns.[506] Regulators must be aware of these kinds of overlaps and ensure that the Act is flexible enough to address the multifaceted nature of AI systems.[507] Similarly, AIDA's emphasis on proactive bias assessment and mitigation addresses a critical need in AI regulation.[508] However, identifying and neutralizing biases in AI systems is a complex task, given that biases often originate from skewed data sets.[509] Imagine an AI-powered healthcare diagnostic tool that, due to historical data biases, is less accurate in diagnosing certain diseases in minority groups.[510] This scenario highlights the need for ongoing research and dynamic regulatory approaches to effectively manage such biases.[511]

The delineation of responsibilities among designers, developers, distributors, and operational managers under AIDA is a smart approach to governance but may be more difficult in practice.[512] For example, a distributed AI system for traffic management could face lapses in

[506] *Id.*
[507] *Id.*
[508] *Id.*
[509] *Id.*; James Manyika, Jake Silberg, & Brittany Presten, *What Do We Do About the Biases in A.I.?*, HARV. BUS. REV.: A.I. AND MACH. LEARNING (Oct. 25, 2019), https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai [https://perma.cc/6FMX-83PF].
[510] *Id.*
[511] *Id.*
[512] *Id.*

responsibility if the roles of designers, operators, and local authorities are not clearly defined.[513] Regulators should establish comprehensive guidelines and monitoring systems to prevent such oversights.[514] Another concern is that AI technology is evolving at a breathtaking pace, and there's a risk that AIDA could become outdated quickly. Regulators must stay ahead of the curve, continuously updating the framework to keep up with new developments.[515]

In the context of AI's global reach, it's crucial for Canadian regulators to align AIDA with international norms and standards.[516] This alignment is vital for ensuring that Canadian AI innovations are competitive and responsible on the global stage.[517] Imagine a Canadian AI firm developing autonomous drones for agricultural use, perhaps as part of the AI for Good initiatives previously discussed in Part II.[518] If AIDA's regulations are misaligned with those in key markets like the EU or US, it could hinder that firm's ability to compete internationally.[519]

Building public trust in AI goes beyond legislative measures.[520] Continuous dialogue with the public, experts, and stakeholders is essential

[513] *Id.*

[514] *Id.*

[515] *Id.*

[516] *Id.*

[517] *Id.*

[518] *See supra* Part II.

[519] AIDA Companion Document, *supra* note 484.

[520] *Id.*

for fostering transparency and trust in AI governance.[521] For instance, a public consultation process for an AI-driven urban planning project could help alleviate concerns, ensuring that the project is aligned with public interest and ethical standards.[522] In summary, while AIDA lays a foundational framework for AI regulation in Canada, its effective implementation requires addressing the complexities of AI categorization, evolving biases, ensuring compliance across the AI lifecycle, adapting to rapid technological changes, aligning with international standards, and fostering public engagement and transparency.[523]

China

China has developed three algorithm and AI regulations—the "2021 regulation on recommendation algorithms, the 2022 rules for deep synthesis (synthetically generated content), and the 2023 draft rules on generative AI."[524] These regulations were revised in July 2023 and went into effect in August of 2023.[525] One of the central goals of these measures is information

[521] *Id.*
[522] *Id.*
[523] *Id.*
[524] MATT SHEEHAN, CARNEGIE ENDOWMENT FOR INT'L PEACE, CHINA'S AI REGULATIONS AND HOW THEY GET MADE 1, 4 (July 2023).
[525] Huw Roberts & Emmie Hine, *The Future of AI Policy in China*, E. ASIA F. (Sept. 27, 2023), https://www.eastasiaforum.org/2023/09/27/the-future-of-ai-policy-in-china/ [https://perma.cc/C6TG-Y739]; Mark MacCarthy, *The US and Its Allies Should Engage with China on AI Law and Policy*, BROOKINGS (Oct. 19, 2023), https://www.brookings.edu/articles/the-us-and-its-allies-should-engage-with-china-on-ai-law-and-policy// [https://perma.cc/ZCT8-PU39].

control.[526] The recommendation algorithm rules safeguard the rights of workers dealing with algorithmic scheduling and prevent excessive price discrimination.[527] The deep synthesis regulation mandates clear labels on artificially generated content.[528] The draft generative AI regulation demands that "both training data and model outputs . . . be 'true and accurate,'" which could pose significant challenges for AI chatbots to meet.[529] Each of the three regulations require developers to file with the Chinese algorithm registry and to pass a security assessment.[530] By creating the algorithm registry and other reusable tools, the frameworks "can act as regulatory scaffolding that can ease the construction of each successive regulation."[531] China is expected to release a policy the size of the EU's AI Act sometime in the next few years.[532] Accordingly, understanding the Chinese regulations is important for other key global players.[533]

The first set of regulations responds to concerns about the use of AI in the dissemination of information online, which the Chinese Communist Party (CCP) viewed as a threat to the ability to regulate public discourse.[534]

[526] SHEEHAN, *supra* note 524, at 4.
[527] *Id.*
[528] *Id.*
[529] *Id.*
[530] *Id.*
[531] *Id.*
[532] *Id.* at 7.
[533] *Id.* at 8.
[534] *Id.* at 12.

The regulation includes more solidified tools for content control online, such as "requiring that platforms intervene in lists of hot topics on social media to ensure they reflect government priorities."[535] Further, as part of the algorithm registry, algorithm developers must provide details about the training and deployment of their algorithms, including the datasets used, and prepare a security assessment.[536] Once an algorithm is registered successfully, a limited version of the submission becomes publicly available.[537] Moreover, developers are obliged to register their algorithms under subsequent regulations related to deep synthesis and generative AI.[538]

The second set of regulations addresses deepfakes and deep synthesis.[539] The regulation covers "the use of algorithms to synthetically generate or alter content online, including voice, text, image, and video content."[540] The framework requires that such content conform to various information controls, be labeled as a synthetic generation, and that the algorithm developers have taken steps to mitigate risk and misuse.[541] The regulation contains several imprecise censorship guidelines like the requirement that deep synthesis content align with the approved political

[535] *Id.*

[536] *Id.* at 13.

[537] *Id.*

[538] *Id.* at 13–14.

[539] *Id.*

[540] *Id.*

[541] *Id.*

stance.[542] The model must refrain from disrupting economic and social stability, and not be employed for creating fake news.[543] If such content has the potential to confuse or misinform the public, it must prominently display a notice in a reasonable location, indicating that it was artificially generated.[544] The regulation also incorporates various measures to counter misuse, such as mandating that users of deep synthesis provide their real identities and that platforms prompt users to obtain consent from individuals whose personal information is being altered.[545] Deep synthesis providers are obligated to submit an entry to the algorithm registry.[546]

The third set of regulations is aimed at managing generative AI.[547] A draft regulation, released in April 2023, "reinforced many boilerplate content guidelines . . . and compelled providers to submit a filing to the existing algorithm registry.."[548] It also introduced several fresh requirements regarding training data and generated content that could pose significant challenges for providers.[549] For instance, it demands that providers "ensure the 'truth, accuracy, objectivity, and diversity' of their training data," "which may be an

[542] *Id.*
[543] *Id.*
[544] *Id.*
[545] *Id.*
[546] *Id.*
[547] *Id.* at 14.
[548] *Id.*
[549] *Id.*

exceedingly difficult standard to meet, particularly for large language models trained on extensive datasets scraped from numerous websites.[550] This standard also raises issues regarding the regulation's stipulation that training data must not infringe on intellectual property rights.[551] Additionally, "[t]he regulation mandates that generative AI not be discriminatory on the basis of race or sex and that generated content be "true and accurate," a solved technical problem for LLMs that are prone to 'hallucinating' inaccurate or baseless claims in their outputs."[552]

Another key aspect of the framework requires the creation of the National AI Office, a new agency "to coordinate and supervise the administration of AI technology.[553] The aim is to prevent the potential regulatory chaos, often likened to the mythical notion of '九龙治水,' (*jiǔlóngzhìshuǐ*, nine dragons governing water), where numerous regulatory bodies have jurisdictional crossover.[554] In some places, the law defers to standards and regulations already in force;[555] however, the regulation includes provisions that overlap with various regulations already in force to close

[550] *Id.*
[551] *Id.*
[552] *Id.*
[553] Graham Webster et al., *Forum: Analyzing an Expert Proposal for China's Artificial Intelligence Law*, DIGICHINA (Aug. 23, 2023), https://digichina.stanford.edu/work/forum-analyzing-an-expert-proposal-for-chinas-artificial-intelligence-law/ [https://perma.cc/PC8R-Z832].
[554] *Id.*
[555] *Id.*

necessary gaps, but may impose new challenges as to what rule is to be followed and to whom the enforcing party ought to be.[556]

In a December 2023 ruling, the Beijing Internet Court held that AI generated works could be covered under China's copyright regime.[557] China has also implemented a licensing regime for AI.[558] The licensing regime, however, is not applied to companies that have developed AI technology but have not yet deployed it for public use.[559] The Internet Court's decision reflects China's ambition to surpass the U.S. and EU as global AI leaders, as well as their cautious but tolerant approach to AI.[560]

In summary, China has instituted a series of comprehensive laws designed to shape the development and application of artificial intelligence. Each set aims to address specific aspects of AI technology, from protecting workers in the gig economy to ensuring the ethical use of deepfakes and maintaining the accuracy of generative AI outputs. For example, the 2021 regulation on recommendation algorithms safeguards against unfair algorithmic scheduling and extreme price discrimination, while the 2022 deep

[556] *Id.*

[557] Angela Huyue Zhang, *China's* SHORT-*Short-Sighted AI Regulation*, THE STRATEGIST (Dec. 12, 2023), https://www.aspistrategist.org.au/chinas-short-sighted-ai-regulation/ [https://perma.cc/BY97-Q7LY].

[558] MacCarthy, *supra* note 525.

[559] *Id.*

[560] Zhang, *supra* note 557; Roberts & Hine, *supra* note 525.

synthesis rules demand clear labeling of artificially generated content.[561] The 2023 draft for generative AI further intensifies the focus on training data and output accuracy, challenging developers to adhere to strict standards.[562]

These regulations have profound implications for multinational companies operating in China. Firms using AI for consumer recommendations or employee scheduling must adapt their algorithms to meet Chinese standards, a process that includes registering with the Chinese algorithm registry and undergoing rigorous security assessments. The demands of the deep synthesis regulation pose particular challenges for those in the media and advertising sectors, where AI-generated content must be clearly labeled and conform to Chinese political and social norms, which are often in direct contract to Western views. This could significantly restrict creative freedom and necessitate comprehensive content review processes. Similarly, the generative AI regulation compels multinational tech companies to ensure the authenticity and diversity of their training data, a difficult task given the expansive and varied nature of such data.

Moreover, the strict regulatory environment in China may impede the pace of AI innovation and research for global firms. Adhering to these regulations could mean additional bureaucratic layers and a slower time-to-

[561] SHEEHAN, *supra* note 524, at 4.
[562] *Id.*

market for AI-driven products and services, potentially making China a less attractive hub for certain types of AI research and development intended to have a global reach. This environment also necessitates a heightened focus on data privacy and protection, in line with China's broader emphasis on data sovereignty and U.S. and EU governmental concerns about the safety of information in Chinese government hands. Additionally, the recent ruling by the Beijing Internet Court, extending copyright protection to AI-generated works, adds another layer of complexity, particularly for companies in creative industries like gaming or software development. While these laws aim to set a global precedent, they also pose significant practical and ethical challenges for multinational companies who may have to comply with the conflicting regulations of the EU, Canada, and the UK. On a hopeful note, China did attend the Bletchely summit and did commit to working with other nations to address AI risks.[563]

Other Global Players

Southeast Asian countries are working on drafting ethics and governance guidance for AI.[564] Members of the Association of Southeast

[563] Paul Sandle & Martin Coulter, *AI Safety Summit: China, US and EU Agree to Work Together*, REUTERS (Nov. 1, 2023, 6:42PM42 PM), https://www.reuters.com/technology/britain-brings-together-political-tech-leaders-talk-ai-2023-11-01/.

[564] Fanny Potkin & Panu Wongcha-um, *Exclusive: Southeast Asia to Set "Guardrails" on AI with New Governance*, REUTERS (June 16, 2023, 4:13 AM), https://www.reuters.com/technology/southeast-asia-set-guardrails-ai-with-new-governance-code-sources-2023-06-16//.

Asian Nations (ASEAN) agreed in early 2023 on the importance of developing AI regulation.[565] The ASEAN Guide on AI Governance and Ethics aims to balance economic benefits and the risks that accompany emerging technology.[566] In contrast to the EU's AI Act, the ASEAN Guide on AI "asks companies to take countries' cultural differences into consideration and doesn't prescribe unacceptable risk categories."[567] Guidance that allows for flexibility and adaptability for cultural differences is important for this region as, "[w]ith almost 700 million people and over a thousand ethnic groups and cultures, Southeast Asian countries have widely divergent rules governing censorship, misinformation, public content and hate speech that would likely affect AI regulation."[568]

Like other ASEAN policies, it is meant to be guidance and is voluntary for member countries to abide by.[569] The guidance is also closely aligned with the U.S. framework.[570] Moreover, the Asia Society published the "Raising Standards on Data and AI report" which emphasizes "the need to bring in human factor considerations, adapting the technology in line with

[565] *Id.*; Singapore, Brunei, Cambodia, Indonesia, Laos, Malaysia, Myanmar, the Philippines, Singapore, Thailand and Vietnam are the ASEAN members.

[566] Fanny Potkin & Panu Wongcha-um, *supra* note 564.

[567] Fanny Potkin & Supantha Mukherjee, *Exclusive: Southeast Asia Eyes Hands-Asia Eyes Hands-Off AI Rules, Defying Rules, Defying EU Ambitionss*, REUTERS, (Oct. 11, 2023, 5:11 PM), https://www.reuters.com/technology/southeast-asia-eyes-hands-off-ai-rules-defying-eu-ambitions-2023-10-11/ [https://perma.cc/BAM3-G695].

[568] *Id.*

[569] *Id.*

[570] *Id.*

the local context, and introducing appropriate regulations for both data and AI."[571] This recommendation emphasizes the importance of "five key principles in regulations: agency, care, equity, inclusion, and reliability."[572] The key characteristic of the report is the promotion of human dignity through AI and technology.[573]

India is on the path to become a major player in the global economy and to have an impact on AI regulation.[574] However, India has oscillated between a hands-off approach and a more cautious approach.[575] India's culture and legal history makes it imperative for the country to adopt regulations that reflect its distinct qualities, rather than regulations that mirror other world players.[576] Moreover, India has a more labor-intensive economy than countries like the United States, UK, or members of the EU, and therefore must consider the labor-automation consequences of various AI policies and regulations.[577]

[571] Priyanka Sahoo, Commentary, *Understanding Data and AI Regulations in Southeast Asia*, TECH FOR GOOD INST., https://techforgoodinstitute.org/blog/commentary/understanding-data-and-ai-regulations-in-southeast-asia/ [https://perma.cc/38FA-RWVZ] (last visited Jan 10, 2024).

[572] *Id.*

[573] *Id.*

[574] Shaoshan Liu, *India's AI Regulation Dilemma*, THE DIPLOMAT (Oct. 27, 2023), https://thediplomat.com/2023/10/indias-ai-regulation-dilemma/ [https://perma.cc/P9CA-F2FG].

[575] *Id.*

[576] *Id.*

[577] *Id.*

Similarly to other regions, Latin America has a patchwork of regulatory frameworks and policies.[578] Seven Latin American countries have developed or are developing AI-specific regulations.[579] Eleven of twelve Latin American countries included in a Bloomberg report[580] have a regulatory scheme for issues adjacent to AI; however, gaps remain.[581] In October 2023, Chile hosted delegations from across the Caribbean and Latin America for an AI summit that resulted in the Santiago Declaration.[582] "A key outcome of this summit is the proposal to establish an intergovernmental Council on Artificial Intelligence for Latin America and the Caribbean."[583] Ad-hoc regional collaboration for technology and AI regulation has been demonstrated in a number of situations, such as IA-CKATÓN, and is likely

[578] Antonio Garrastazu & Beatriz de Anta, *The AI Revolution Is Coming ForAI Revolution Is Coming for Latin America. Is It Ready?*, AMERICAS Q. (Aug. 3, 2023), https://perma.cc/A57U-LKPP [https://perma.cc/A57U-LKPP].

[579] *Id.* (noting these countries are: Argentina, Brazil, Chile, Colombia, Mexico, Peru and Uruguay); *see also* OECD & CAF, THE STRATEGIC AND RESPONSIBLE USE OF ARTIFICIAL INTELLIGENCE IN THE PUBLIC SECTOR OF LATIN AMERICA AND THE CARIBBEAN THE STRATEGIC AND RESPONSIBLE USE OF ARTIFICIAL INTELLIGENCE IN THE PUBLIC SECTOR OF LATIN AMERICA AND THE CARIBBEAN, (2022),) https://doi.org/10.1787/1f334543-en [hereinafter RESPONSIBLE USE].

[580] Daniel Salazar Castellanos, *Which Latin American Countries Lead the Way in AI Regulation?*, BLOOMBERG LÍNEA (Aug. 30, 2023, 6:45 PM), https://www.bloomberglinea.com/english/which-latin-american-countries-lead-the-way-in-ai-regulation/ [https://perma.cc/WH8E-U6AF]. These twelve countries are Argentina, Bolivia, Brazil, Chile, Columbia, Costa Rica, Ecuador, Mexico, Panama, Paraguay, Peru, and Uruguay. *Id.*

[581] Ganiu Oloruntade & Faith Omoniyi, *Where is Africa in the global Global Conversation on Regulating AI?*, TECHCABAL (May 26, 2023), https://perma.cc/UMK7-JE9C;[https://perma.cc/UMK7-JE9C]; RESPONSIBLE USE, *supra* note 579.

[582] Daniel Rodriguez Maffioli, *AI Regulation in Latin America: Balancing Global Trends with Local Realities*, IAPP (Dec. 13, 2023), https://perma.cc/89NJ-J7X5. [https://perma.cc/89NJ-J7X5].

[583] *Id.*

to continue.[584] Like the other aforementioned nations and regions, Latin America is drawing insight and inspiration from the EU AI Act, but must be mindful to tailor the regulatory scheme the region and its countries develop to the needs and contexts present in the region.[585]

The African continent is very much a part of the global discussion on AI regulation and is bursting with potential, as it hosts more than 2,400 AI organizations already.[586] This is unsurprising, as the region has much to gain from the utilization of AI, especially in combating poverty, unemployment, and other socio-economic challenges.[587] For example, "Togo is [utilizing] AI systems to distribute social funds; Zambia to counter disinformation and misinformation during the electioneering periods, and Kenya harnessing machine learning in agriculture and education."[588] Unfortunately, however, there is evidence of the harmful ways that AI is being used in the region, such as the deployment of autonomous weapon systems in Libya, and in facial recognition surveillance systems in Zimbabwe.[589] With a long history of regional exploitation, African countries

[584] RESPONSIBLE USE, *supra* note 579, at 21.
[585] Maffioli, *supra* note 582.
[586] Oloruntade & Omoniyi, *supra* note 581; STRATHMORE UNIV., *The State of AI in Africa*, CTR. FOR INTELL. PROP. AND TECH. INFO. L. 8, [https://perma.cc/3QLX-T8XV] (last updated 2023) ("A sample of the many AI applications currently available in Africa can be found in an interactive dashboard on 'AI Applications' developed by the Center of Intellectual Property and Information Technology Law (CIPIT).").
[587] Oloruntade & Omoniyi, *supra* note 581.
[588] STRATHMORE UNIV., *supra* note 586, at 5.
[589] *Id.*

must be mindful of regulation and deployment of AI that "make a concerted effort to draft AI strategies and move towards enactment to protect vulnerable communities and enable responsible innovation."[590]

The widespread current practice of regulation in Africa is regulatory annexation targeting internet users using law and technology to address various issues, including speech and fake news.[591] An emerging regulatory practice is co-regulation between governments and platforms themselves.[592] Mauritius was the first African country to issue an AI strategy.[593] Several other countries have issued guidance and strategies or developed task forces or agencies to regulate AI.[594] Challenges that the region faces in developing comprehensive AI regulation include a lack of regulatory frameworks for other technology issues, such as privacy and a lack of a structured data ecosystem.[595]

The Middle East has embraced AI and recognized its potential value already. A virtual assistant used by the Dubai Electricity and Water Authority

[590] Chinasa T. Okolo, Commentary, *AI in the Global South: Opportunities and Challenges Towards More Inclusive Governance*, BROOKINGS (Nov. 1, 2023), https://www.brookings.edu/articles/ai-in-the-global-south-opportunities-and-challenges-towards-more-inclusive-governance/ [https://perma.cc/B5DA-6FXM].

[591] Vincent, *What* Obia, *What Can African Countries Do to Regulate Artificial Intelligence*, LSE (June 13, 2023), https://blogs.lse.ac.uk/medialse/2023/06/13/what-can-african-countries-do-to-regulate-artificial-intelligence/ [https://perma.cc/7CLF-FZQY].

[592] *Id.*

[593] Oloruntade & Omoniyi, *supra* note 581.

[594] *Id.*

[595] CTR. STRATHMORE UNIV., *supra* note 586, at 6–7.

has assisted in almost 7 million queries, and "the Fourth Industrial Revolution Center at Saudi Arabia's Aramco says it has reduced flare emissions by 50 percent since 2010 by using data and AI to monitor conditions and take preventative action."[596] A new draft IP law in Saudi Arabia may be one of the first in the region to address IP issues in regard to AI.[597] "The Law states that, '*In the event that the contribution of the natural person was insignificant or that the IP was generated by artificial intelligence independently of the person*', the IP will enter the public domain."[598] The law did not define "significant."[599] In conjunction with other Saudi Arabian laws, the policy scheme seeks to "support Saudi's vision notably those related to the Saudi Data and Artificial Intelligence Authority's . . . strategy objectives."[600]

The United Arab Emirates (UAE) has been a regional leader in AI.[601] The UAE appointed the world's first minister for AI in 2017.[602] The role

[596] Vinay Chandran et al., *The State of AI in GCC* COUNTRIES—*Countries—and How to* OVERCOME ADOPTION CHALLENGES, MCKINSEY*Overcome Adoption Challenges*, MCKINSEY DIGIT. (May 30, 2023), https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-state-of-ai-in-gcc-countries-and-how-to-overcome-adoption-challenges [https://perma.cc/44PV-AUTT].

[597] Melissa Murray et al., *Saudi Arabia Pioneers' Regulation of Artificial Intelligence in the Gulf Region with Proposed New Intellectual Property Law*, BIRD & BIRD (July 6, 2023), https://www.twobirds.com/en/insights/2023/global/saudi-arabia-pioneers-regulation-of-artificial-intelligence [https://perma.cc/459Z-KVCM].

[598] *Id.*

[599] *Id.*

[600] *Id.*

[601]Tarry Singh, THE STATE OF AI IN THE MIDDLE EAST IN 2023 LINKEDIN (2023), https://www.linkedin.com/pulse/state-ai-middle-east-2023-tarry-singh/ [https://perma.cc/SJE6-7MAW]].

[602] *Id.*

includes overseeing AI strategies.[603] The UAE launched a Strategy for Artificial Intelligence 2031, positioning themselves to be a global leader in AI, focusing on nine sectors where AI is predicted to be the most impactful: "transport, health, space, renewable energy, water, technology, education, environment, and traffic."[604]

In December 2023, Israel published its first comprehensive AI and ethics policy.[605] "The AI Policy identifies seven main challenges arising from private sector AI use: discrimination, human oversight, explainability, disclosure of AI interactions, safety, accountability, and privacy."[606] The policy is the result of collaboration between a multitude of stakeholders, both public and private.[607] The policy emphasizes "human-centric innovation, equality, transparency, reliability, accountability, and promoting sustainable development."[608]

[603] *Id.*

[604] *Id.*

[605] Press Release, Ministry of Innovation, Sci. and Tech. & Ministry of Just., Responsible Innovation: Israel's Policy on Artificial Intelligence Regulation and Ethics (Dec. 17, 2023), https://www.gov.il/en/departments/policies/ai_2023#:~:text=Jerusalem%2017th%20December%202023%20%E2%80%93%20Israel,of%20AI%20in%20various%20sectors [https://perma.cc/LN9D-V7YE].

[606] *Id.*

[607] *Id.*

[608] *Id.*

The U.S. Approach

Federal

In 2023, the White House launched a plan to advance democracy in the digital age and counter the misuse of technology and the rise of authoritarianism to shape emerging technologies to ensure respect for human rights and democratic principles.[609] In June 2023, dozens of civil rights organizations urged the Biden-Harris administration to make an AI Bill of Rights to ensure coordinated follow-through by federal administrations in light of Executive Order 14901 and the AI Bill Of Rights.[610] The same month, the U.S. implemented a binding regulation called the U.S. AI Disclosure Act, which requires disclaimers for consumers.[611] These disclaimers must state that the pertinent output has been generated by artificial intelligence.[612]

In 2023, the White House also implemented two executive orders. Executive Order ("E.O.") 14901 directs federal agencies to "root out bias in

[609] Press Release, The White House, FACT SHEET: Advancing Technology for Democracy (March 29, 2023), https://www.whitehouse.gov/briefing-room/statements-releases/2023/03/29/fact-sheet-advancing-technology-for-democracy-at-home-and-abroad/ [https://perma.cc/PP97-R78W] .

[610] Letter from The Leadership Conf. on Civ. and Hum. Rts. et al., to Neera Tanden, Dir. of Domestic Pol'y Council, Exec. Off. of the President et al. (June 13, 2023), https://perma.cc/4H3U-REG3.

[611] Bruce D. Sokler et al., *New AI Disclosure Bill and AI Strategic Plan Update — AI: The Washington Report*, MINTZ (June 13, 2023), https://www.mintz.com/insights-center/viewpoints/2191/2023-06-12-ai-washington-report-new-ai-disclosure-bill-and-ai [https://perma.cc/MTQ3-3CFT].

[612] *Id.*

the design and useof new technologies, such as artificial intelligence," to protect the public from algorithmic discrimination.[613] Then, on"[o]n October 3030, 2023, the Biden Administration released . . . [E.O.] 14110 on *Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*,," which establishes standards for AI safety and security.[614] This is the most comprehensive action implemented thus far in the U.S. to guard against AI risks.[615] It requires sharing of safety results for powerful AI systems, development of standards to ensure safety, security, and trustworthiness; creation of new standards for biological synthesis screening; protection against fraud and deception through implemented practices for spotting AI-generated content,; and establishment of advanced cybersecurity programs.[616] The October 2023 executive order also strives to protect citizen's privacy from the easy extraction, identification, and exploitation of data and to advance equity and civil rights to prevent discrimination, bias, and other abuses in justice, healthcare, and housing.[617] For consumers,

[613] Exec. Order No. 14091, 88 Fed. Reg. 10825 (Feb. 22, 2023). Exec. Order No. 14091, 88 Fed. Reg. 10825 (Feb. 16, 2023).
[614] LAURIE A. HARRIS & CHRIS JAIKARAN, CONG. RSCH. SERV., R47843, HIGHLIGHTS OF THE 2023 EXECUTIVE ORDER ON ARTIFICIAL INTELLIGENCE FOR CONGRESS (2023).
[615] Press Release, The White House, FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence (Oct. 30, 2023), https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/ [https://perma.cc/5XQ7-V9VA].
[616] *Id.*
[617] *Id.*

patients, and students, this executive order provides protection from misleading, injurious, harmful AI products and monitors how AI transforms education; while for workers, it aims to mitigate the risks of workplace surveillance, bias, and job displacement.[618] Lastly, like the previous executive order, it promotes innovation and competition while advancing American leadership abroad to ensure responsible and effective government use of AI.[619]

Congressional Action

Senator Ed Markey and Representative Doris Matsui introduced the Algorithmic Justice and Online Platform Transparency Act[620] in May 2021.[621] The legislation would make it illegal for someone to use AI on an online platform to "deprive[] an individual of a right or privilege under the Civil Rights Act of 1964."[622] The Act, in detail, aims to prevent discrimination on online platforms based on characteristics such as race, gender, age, and ability.[623] This encompasses various forms of discrimination, such as unfair treatment of users, discriminatory design features, biased advertising, and

[618] *Id.*

[619] *Id.*

[620] S. 1896, 117th Cong. (2023).

[621] *Id.*; MICHAEL BOPP ET AL., GIBSON DUNN, FEDERAL POLICYMAKERS' RECENT ACTIONS SEEK TO REGULATE AI 4 (2023).

[622] *Id.* at 4.

[623] *Id.*

handling personal data in ways that deliberately deny someone their voting rights in federal, state, or local elections.[624]

Additionally, the Act establishes a benchmark for the safety and effectiveness of algorithms, prohibiting online platforms from using automated processes that harm users or don't reasonably accomplish their intended purposes.[625] It requires online platforms using algorithms to disclose their processes and the data they use.[626] Furthermore, it mandates the annual publication of public reports on content moderation practices and empowers the FTC to enforce the Act, treating violations as "'unfair or deceptive act[s] or practice[s].'"[627] The FTC can also provide advisory opinions upon request.[628] Lastly, the Act establishes an inter-agency task force, consisting of several government departments, "to investigate discriminatory algorithmic processes employed in sectors across the economy."[629] Representative Matsui has called the bill "'an essential roadmap for digital justice to move us forward on the path to online equity," which addresses "'[b]iased artificial intelligence systems that have become embedded in the fabric of our digital society.'"[630]

[624] *Id.*
[625] *Id.*
[626] *Id.*
[627] *Id.* at 4–5.
[628] *Id.* at 5.
[629] *Id.*
[630] *Id.* at 4. (alteration in original).

In February 2022, Senator Ron Wyden introduced the Algorithmic Accountability Act (AAA).[631] One of the key features of the AAA is that it would call for the FTC—in collaboration with other key public and private players——to promulgate specific regulations, "requiring 'covered entit[ies]' to perform impact assessments of certain AI systems and 'augmented critical decision process[es].'"[632] These assessments "would include detailed documentation of consultations with relevant stakeholders, ongoing testing and evaluation efforts, employee training, and consumers' rights."[633] Covered entities would be required to submit an initial summary report of impacts for new AI systems to the FTC prior to deployment and an annual "summary report," which assesses "any likely negative impacts on consumers, as well as the need for any guardrails on the use of the AI system."[634] Non-covered entities are encouraged, but not required, to submit AI impact assessments.[635] Other key decisions that would be governed by this bill include those with "'legal, material, or similarly significant effect' on a person's access to a wide range of interests such as housing and employment."[636]

[631] S. 3572,117th Cong. (2022). Algorithmic Accountability Act of 2022, S. 3572,117th Cong. (2022).
[632] BOPP ET AL., *supra* note 621, at 3. (alteration in original).
[633] *Id.*
[634] *Id.* at 3–4
[635] *Id.*
[636] *Id.* at 4.

Congress took further steps in late Spring 2023 to consult with industry leaders to start formulating a comprehensive regulatory framework for AI.[637] Senate Majority Leader Schumer announced his intention to lead the Senate in this endeavor and put forth his own proposed regulatory regime.[638] The comprehensive framework "is expected to center around four 'guardrails' designed to guide the effective disclosure and testing of AI technologies . . . without stifling innovation. . . . [and] aim to regulate AI technology properly and align AI systems with American values."[639] The first three guardrails answer who, where, how questions.[640] First, the framework identifies who the program's intended audience is, and who developed and trained the program.[641] Second, the framework asks where the data the AI was trained on originated.[642] Third, the framework asks for an explanation of "'how [the AI system] arrives at its responses.'"[643] Lastly, the fourth guideline aims to ensure the protection of "transparent"[t]ransparent and strong ethical boundaries, focused on 'aligning AI systems with American values

[637] *Id.* at 1.
[638] *Id.*
[639] *Id.* at 2.
[640] *Id.*
[641] *Id.*
[642] *Id.*
[643] *Id.*

and ensuring that AI developers deliver on their promise to create a better world.'"[644] AI regulation has received bipartisan interest and support.[645]

Agency Action

The heads of many federal agencies, "including the Civil Rights Division of the Department of Justice (DOJ), Equal Employment Opportunity Commission (EEOC), . . .Federal Trade Commission (FTC,), and Consumer Financial Protection Bureau (CFPB),"),'' met on April 25, 2023, and "outlined their commitment to focus on mitigating potential discrimination arising from AI systems."[646] Two days later, on"[o]n April 27, 2023, the White House released a request for information (RFI) on how AI is 'being used to surveil, monitor, and manage workers.'"[647] The White House then held a meeting with the CEOs of multiple companies at the forefront of emerging technology and made a statement about the promotion of responsible use of AI innovation.[648]

President Biden's administration has underscored a people-centric approach in shaping AI policies, focusing on placing communities and individuals at the forefront of its AI strategy through a series of actions, including increased funding for AI research and development,

[644] *Id.*
[645] *Id.*
[646] *Id.* at 1.
[647] *Id.*
[648] *Id.*

comprehensive evaluations of existing generative AI systems, and the formulation of policies aimed at both mitigating risks and maximizing the potential benefits of AI within the federal government. Additionally, the White House Office of Science and Technology Policy has initiated a an RFI to gather insights on the application of AI in workplace surveillance and management.[649] The objective of this RFI is to leverage the collected data to develop new policies and promote best practices among employers and other relevant parties, ensuring responsible use of AI in workforce management.[650] This initiative reflects the administration's commitment to understanding and addressing the multifaceted implications of AI on labor and employment.[651]

Furthermore, the federal government has released non-binding pronouncements such as the Platform for Accountability and Transparency Act (PATA),[652] the AI Bill of Rights,[653] and the AI Risk Management Framework.[654] PATA, drafted in 2021, formally introduced in 2022, and

[649] *Id.* at 1, 5–6.

[650] *Id.*

[651] *Id.*

[652] John Perrino, *Platform Accountability and Transparency Act Reintroduced in Senate*, TECH POLICY PRESS (June 8, 2023), https://techpolicy.press/platform-accountability-and-transparency-act-reintroduced-in-senate/ [https://perma.cc/32Y6-G8BY].

[653] Off. of Sci. & Tech. Pol'y, *Blueprint for an AI Bill of Rights*, THE WHITE HOUSE, https://www.whitehouse.gov/ostp/ai-bill-of-rights/ [https://perma.cc/5RE6-TNT8] (last visited Jan. 10, 2024).

[654] *AI Risk Management Framework*, NIST, https://www.nist.gov/itl/ai-risk-management-framework [https://perma.cc/K4GV-TGLY] (last visited Jan. 10, 2024).

reintroduced in 2023, will require the FTC "to develop privacy and security protocols."[655] This bill is not yet law but makes important steps to aid scientists' study of social media by gaining access to data.[656] The AI Bill of Rights identifies five principles to guide the design, use, and deployment of automated systems to protect Americans against AI.[657] While not law, it serves as a guide to protect citizens from AI threats and aims to create "[s]afe and effective systems, algorithmic discrimination protections, data privacy, notice explanations, Effective Systems, Algorithmic Discrimination Protections, Data Privacy, Notice Explanation, Human Alternatives, Consideration, and Fallback."[658] The AI Risk Management Framework released in January 2023 aims to improve the ability to incorporate trustworthiness considerations into the design, development, use, and evaluation of AI products, services, and systems.[659] However, this framework is voluntary.[660]

State Action

There is an increase in state regulations in the context of AI. Since 2019, seventeen states including "California, Colorado, Connecticut,

[655] Perrino, *supra* note 652.
[656] *Id.*
[657] Off. of Sci. & Tech. Pol'y, *supra* note 653.
[658] *Id.*
[659] *AI Risk Management Framework*, *supra* note 654.
[660] *Id.*

Delaware, Illinois, Indiana, Iowa, Louisiana, Maryland, Montana, New York, Oregon, Tennessee, Texas, Vermont, Virginia, and Washington" have enacted twenty-nine bills to regulate AI.[661] The California legislature has passed an explicit right to privacy[662] and has proposed employment AI regulation legislation requiring audits of AI tools used by employers and created by developers.[663] However, Silicon Valley being the center of technology is changing, as many technology companies move to Florida and Texas.[664] Florida recently passed a Technology and Transparency bill that includes specific requirements for the collection of data and consumer data protection.[665]

Industry Recommendations for U.S. Regulation

On May 16, 2023, both the Senate Judiciary Committee's Subcommittee on Privacy, Technology, and the Law and the Senate Homeland Security and Governmental Affairs Committee held public

[661]Rachel Wright, *Artificial Intelligence in the States: Emerging Legislation*, CSG (Dec. 6, 2023), https://www.csg.org/2023/12/06/artificial-intelligence-in-the-states-emerging-legislation/ [https://perma.cc/FCQ3-AFAU].

[662] Natasha Singer, *Charting the 'California Effect' on Tech* Regulation, N.Y. TIMES (Oct. 12, 2022), https://www.nytimes.com/2022/10/12/us/california-tech-regulation.html [https://perma.cc/P3F4-K5TE].

[663] Jeffery S. Bosley, et al., *California Proposed Employment AI Regulation and Legislation*, DAVIS WRIGHT TREMAINE LLP (May 2, 2023), https://www.dwt.com/blogs/employment-labor-and-benefits/2023/05/california-ai-employment-law-regulations [https://perma.cc/A88T-KMAP].

[664] *Why Tech Companies Are Moving to Texas and* Florida, CEO N. AM., https://ceo-na.com/business/management-leadership/why-tech-billionaires-are-leaving-california-for-texas-and-florida/ [https://perma.cc/7HR9-W7UV] (last visited Jan 10, 2024).

[665] S. 262, 2023 Leg., Reg. Sess. (Fla. 2023). S. 262, 2023 Leg., Reg. Sess. (Fla. 2023).

hearings with testimony from OpenAI CEO Sam Altman, NYU Professor Emeritus Professor of Psychology and Neural Science, Gary Marcus, and IBM executive Christina Montgomery.[666] Senator Blumenthal explained that future AI legislation and regulation must demystify technology and hold accountable new technologies so as to avoid mistakes of the past.[667] Some of the initial suggestions from the Senators included requirements that AI companies disclose known risks associated with their programs and allow independent researcher access, as well as potential auto-bans on the use of AI where it can be harmful.[668]

After opening statements by the members of the Committees, Altman, Montgomery, and Marcus answered questions and gave their own recommendations for the regulation of AI.[669] When asked for his top three recommendations for an AI regulation framework, Altman listed a licensing agency, safety standards and test models before the technology can be

[666] Bopp et al., *supra* note 621, at 1; Neda. M. Sheheen et al., *Overview of the First U.S. Senate Hearing on the "Oversight of A.I.: Rules for Artificial Intelligence,"* CROWELL (May 19, 2023), https://www.crowell.com/en/insights/client-alerts/Overview-of-the-first-US-Senate-hearing-on-the-Oversight-of-AI-Rules-for-Artificial-Intelligence [https://perma.cc/PQT7-5P93].].

[667] CNBC Television, *OpenAI CEO Sam Altman Testifies During Senate Hearing on AI Oversight* – 05/16/23, YOUTUBE (2023), https://www.youtube.com/watch?v=fP5YdyjTfG0 [https://perma.cc/YTC3-DKV5] (last visited Jan 10, 2024).).

[668] *Id.*

[669] *Id.*

deployed into the world, and independent audits as his biggest priorities.[670] He also recommended a national privacy law regime.[671]

Montgomery from IBM suggested that the U.S. adopt a precision approach to governing AI by regulating deployment, rather than the technology itself.[672] Her three recommendations and priorities for AI legislation are transparency, explainability, and disclosure.[673] One notable aspect of Montgomery's testimony is that she was the only witness who stated that there should not be an agency created to deal with AI.[674] Marcus's top three recommendations and priorities are a safety review process, as with the FDA, prior to deployment of models; a nimble cabinet-level monitoring agency; and funding for safety research and promoting ethical AI.[675] Professor Marcus emphasized the influence AI can have on society, noting that AI is influenced heavily by its creators[676] and that "those that choose the data will make the rules."[677]

The United States and others on the world stage contemplating state, nationwide, regional, and global regulation have several complementary and

[670] *Id.*
[671] *Id.*
[672] *Id.*
[673] *Id.*
[674] *Id.*
[675] *Id.*
[676] *Id.*
[677] *Id.*

conflicting draft regulations to draw from and hopefully harmonize. In Part IV, we propose overarching considerations for any legislation.

Part IV: Considerations for Legislators and Regulators

Any legislation should take into account the NIST principles for trustworthy and responsible AI: 1) Validity and Reliability; 2) Safety; 3) Security and Resiliency; (3) Accountability and Transparency; (4) Explainability and Interpretability; (5) Privacy; and (6) Fairness with Mitigation of Harmful Bias.[678] Industry leaders' perspectives on AI regulation, as articulated in 2023 Senate hearings, highlights challenges with developing comprehensive yet balanced AI regulation that comport with the NIST principles.

Sam Altman, CEO of OpenAI, emphasized the importance of a licensing agency, safety standards, independent audits, and a national privacy law. These recommendations align with the need for accountability and transparency in AI deployment. Christina Montgomery from IBM advocated for a precision approach, focusing on the deployment of AI rather than the technology itself, with an emphasis on transparency, explainability, and disclosure.[679] This stance reflects a pragmatic view of regulation, aimed at

[678] *Trustworthy and Responsible AI: Overview,* NIST, https://www.nist.gov/trustworthy-and-responsible-ai [https://perma.cc/B9QY-QNGJ] (last visited Mar. 14, 2024)

[679] CNBC Television, *supra* note 667.

ensuring AI's responsible use without stifling innovation.[680] Professor Gary Marcus' recommendations included a safety review process similar to the FDA's, a monitoring agency, and funding for safety research and ethical AI promotion.[681] His focus on the social impact of AI highlights the importance of considering the broader societal implications of AI deployment.

These witnesses' views exhibit both alignment and divergence with the NIST standards. Altman's emphasis on safety standards and independent audits corresponds closely with NIST's focus on safety, security, and accountability. Montgomery's priority of transparency and explainability resonates with NIST's criteria on transparency, interpretability, and privacy. Marcus's advocacy for a rigorous safety review process and ethical AI aligns with NIST's emphasis on safety and fairness. However, there are potential conflicts in their approaches. Altman's suggestion of a licensing agency and national privacy law, while ensuring compliance, may not fully address the NIST criteria of fairness and mitigation of harmful bias. Montgomery's opposition to a dedicated AI agency could conflict with NIST's emphasis on resiliency and security, potentially limiting the government's ability to adaptively monitor and respond to AI developments. Marcus's proposal for a monitoring agency and safety research aligns with NIST standards but

[680] *Id.*
[681] *Id.*

could be perceived as potentially slowing down innovation if not implemented with agility.

In practical terms, implementing these recommendations will require a multi-pronged approach. For example, a licensing agency, as proposed by Altman, could ensure that AI systems deployed in healthcare adhere to stringent safety and privacy standards, enhancing patient trust and care quality. Montgomery's focus on transparency and explainability would be particularly relevant in finance, where AI-driven decision-making must be understandable to both regulators and consumers to maintain market integrity. Marcus' emphasis on safety reviews could inform the deployment of AI in autonomous vehicles, ensuring that these systems are thoroughly vetted for safety before public use. As the United States and other nations prepare to regulate, these views and the drafts from the EU, Canada, China, and the UK offer varied perspectives. We now turn to specific considerations for regulators aligned with these various points of view.

After considering the current regulatory landscape and the known and unknown risks related to AI, we make eleven big picture policy recommendations, all requiring a consideration of the NIST principles. In summary, we propose (1) AI Ethics Review Boards to evaluate AI projects based on NIST standards of accuracy, fairness, and transparency; (2) the mitigation of AI-induced job displacement with retraining programs aligned with NIST's reliability and fairness criteria; (3) energy consumption caps for AI training in accordance with NIST's safety and environmental impact standards; (4) protocols to combat deepfakes and misinformation; (5) regulation of Surveillance and Biometric AI; (6) enforcement of AI security standards based on NIST guidelines for security and resilience; (7) an international treaty on AI weaponization based on NIST principles of ethical use; (8) AGI development licensing using NIST guidelines on safety and ethical use; (9) an international AI regulatory body to enforce standards that mirror NIST's comprehensive guidelines; (10) meaningful and prohibitively expensive sanctions for regulatory violations based on the severity of deviation from NIST standards; and (11) regional tailoring of AI regulations while maintaining global NIST-based standards.

The recommendation to implement AI Ethics Review Boards across companies developing AI technologies is a critical step towards responsible AI development. These boards, comprising a mix of ethicists, legal experts, data scientists, and community representatives, should be tasked with scrutinizing and approving AI projects, particularly focusing on their ethical implications, potential biases, and legal compliance. For example, an AI system designed for hiring must be thoroughly reviewed for biases against minority groups to ensure fairness and inclusivity. The composition of these boards should reflect a diverse range of expertise and perspectives, with larger companies potentially establishing internal boards, while smaller entities might utilize external boards provided by consortia or industry associations.

These review boards should engage actively throughout the AI system's lifecycle, from inception to post-deployment, ensuring ongoing ethical compliance and societal impact assessment. Operating on standardized criteria that encompass privacy, bias, transparency, accountability, and societal impact, these boards should align their assessments with established ethical frameworks, like the EU's AI Ethics Guidelines. However, the implementation of review boards will face challenges. Smaller companies and startups may find the establishment and maintenance of these boards resource-intensive, lacking necessary expertise. Moreover, the application of uniform ethical standards across diverse company types and sizes may not always be practical or effective. For instance, the ethical considerations for an AI tool in education differ significantly from those in the financial sector.

Addressing these challenges requires a flexible approach, particularly for startups and small businesses. Regulators could consider options such as simplified review processes, extended timelines for compliance, and access to shared external ethics boards. Support mechanisms, including tax breaks, grants for ethics training, or shared resources, provided by governments and industry associations, could further ease the burden on smaller entities. These measures not only ensure compliance but also encourage innovation within a regulated framework. In addition, employees across all levels should receive training in AI ethics to cultivate a workplace culture that prioritizes ethical considerations in AI development. Public-private partnerships can play a significant role in this regard, pooling resources and expertise to aid smaller companies in meeting ethical oversight requirements affordably. We also recommend that employees specifically hired to focus on ethical and compliance issues receive additional legal protection from termination and have direct and unfettered access to the company's board of directors.

The second policy recommendation focuses on mitigating the impact of AI-induced job displacement, emphasizing the need for comprehensive support systems for workers. This includes the establishment of government funds dedicated to assisting those displaced by AI, with initiatives like retraining programs, unemployment benefits, and job placement services. An example could be a fund designed to aid retail workers who lose their jobs due to the adoption of automated checkout systems. A crucial step in addressing AI-induced job displacement involves identifying sectors and roles most vulnerable to this phenomenon using the frameworks established in Part II. This identification process requires a collaborative approach involving government agencies, industry experts, and labor organizations, using data-driven analysis. For example, manual roles in manufacturing or administrative data entry positions are likely candidates for displacement. Following this identification, retraining programs, developed in line with NIST's principles of fairness and reliability, should be implemented. These programs need to equip displaced workers with skills that are relevant in an AI-enhanced job market, such as training factory workers to operate and maintain AI-driven machinery.

Further, fostering public-private partnerships is essential in creating effective training initiatives. These partnerships could bring together tech companies, educational institutions, and government bodies to offer a blend of technical and soft skills training. For instance, a tech company specializing in AI might collaborate with a community college to offer courses in AI management and ethics. Comprehensive career transition services are also imperative, offering personalized assistance in career counseling, job placement, and interview preparation. However, implementing these strategies will be costly and will require careful budgeting to ensure effectiveness and sustainability. Moreover, the relevance of these training programs to future job market demands is crucial; they must be continuously adjusted to remain beneficial.

Adaptations of these strategies for companies of various sizes and types are also necessary. Smaller companies might need more substantial support, such as access to government-funded training programs or larger tax incentives. Training delivery should be flexible, with options such as online courses or part-time programs, to cater to the diverse schedules and capacities of workers. To protect jobs and address cost-cutting concerns, regulatory measures could be implemented to encourage or mandate retraining and redeployment over layoffs. This approach aligns with responsible and ethical workforce management. Moreover, engaging with labor unions and worker advocacy groups ensures that retraining programs and transition services are attuned to the actual needs of affected workers, fostering a cooperative approach to workforce transition.

The policy of implementing energy consumption caps for AI training is particularly critical to mitigate the environmental impact of AI development. This strategy involves setting specific energy consumption thresholds for AI training processes, which would be determined based on factors such as the complexity of the AI model and the industry of application. For example, the energy caps for training a sophisticated AI model in healthcare diagnostics could differ from those for a simpler AI application in retail customer service.

A tiered approach to these caps, taking into account the size and resources of different companies, is essential. Larger corporations might be assigned higher caps due to their broader operational scope, while smaller startups would have lower caps, reflective of their limited resources and smaller-scale operations. Incentives for adopting energy-efficient practices could include tax reductions, subsidies, or public recognition for companies that develop and utilize energy-efficient AI algorithms and hardware. Supporting companies in transitioning to renewable energy sources for AI operations could manifest in partnerships with renewable energy providers, government grants for renewable energy infrastructure, or preferential rates for green energy usage. Regular monitoring and reporting of energy consumption in AI training should be mandated to ensure transparency and adherence to the established caps.

However, we acknowledge that there are drawbacks to this approach. Strict energy caps might impede the development of advanced AI models, necessitating provisions for exceptions in cases where exceeding the cap is justified by significant technological or societal benefits. Smaller companies might struggle with the costs associated with adopting energy-efficient technologies. To address this, government-funded programs or partnerships with larger corporations could be implemented to support these companies in making the transition. Ensuring compliance and accurate reporting of energy usage is another challenge, which could be addressed through strict penalties for non-compliance and clear guidelines for energy reporting.

The implementation of this policy should be gradual, allowing companies time to adjust their operations and invest in energy-efficient technologies. Technical assistance and advisory services would be beneficial to help companies understand and meet these energy caps. This could include guidance on selecting energy-efficient AI algorithms and hardware, as well as options for renewable energy. Facilitating collaborations between AI companies and providers of energy-efficient technologies and renewable energy would also aid companies in finding cost-effective solutions to meet the energy caps.

There is a heightened sense of urgency to combat state-sponsored deepfakes and election misinformation due to the number of elections in 2024, but it is a complex challenge that requires a globally coordinated and multi-faceted approach. Key strategies include forming global cybersecurity alliances for intelligence sharing and establishing joint task forces to counter international misinformation campaigns. Investing in the development of AI tools for deepfake detection and utilizing blockchain technology for information verification are essential technological measures.

We also advocate for comprehensive international cybersecurity laws and implementation of cyber sanctions against offending states and rogue actors. Strengthening domestic cybersecurity infrastructure through national cyber defense centers and secure communication channels for election officials is vital, as is conducting media literacy programs and public awareness initiatives to educate the public in identifying misinformation and deepfakes.

Collaborating with tech companies and social media platforms for content monitoring, rapid response, and transparency reporting is necessary for online spaces. Additionally, preparedness through election security drills and developing crisis communication strategies ensures readiness for potential misinformation attacks. This comprehensive strategy, encompassing technology, law, public education, and prepared cyber defense, aims to protect electoral integrity and democratic processes against sophisticated digital threats, but we acknowledge this may be particularly difficult in an election year in which deepfakes are already being used and where the willingness to address these issues may depend on which leader or power is in party after this election season.

The proposal to regulate surveillance and biometric AI underscores the need for a comprehensive legal and ethical framework, designed to safeguard individual privacy and uphold ethical standards. This framework would include stringent laws governing the collection, use, and storage of biometric data, ensuring that such data is handled responsibly and transparently. For instance, biometric data protection laws would set clear guidelines on legal data acquisition and usage, while consent and transparency requirements would mandate explicit informed consent from individuals, ensuring they are fully aware of how their data is being used and processed.

We recommend the establishment of regulatory bodies or committees to oversee the use of biometric AI technologies that would ensure compliance with ethical standards and privacy laws, implementing strict criteria for any governmental use of facial recognition and other biometric AI in public spaces. Alongside these regulatory measures, ethical guidelines and regular impact assessments of biometric AI systems would evaluate their effects on privacy, civil liberties, and societal norms.

Public engagement and education are also essential, involving community consultations to understand public sentiment and concerns about biometric surveillance as well as launching educational campaigns to raise awareness about biometric data collection and its implications. Implementing rigorous data security standards and encouraging the use of anonymization techniques in situations where individual identification is unnecessary are key aspects of this strategy. Additionally, establishing clear mechanisms for individuals to challenge the misuse of their biometric data and seek redress, as well as imposing substantial penalties for entities that violate biometric data regulations, are critical for ensuring compliance and accountability.

This will not work without international cooperation and alignment with global standards. This may include regulating the cross-border sharing of biometric data and ensuring adherence to international human rights and privacy standards. This may be a challenge given the patchwork of privacy regulations in the U.S. Nonetheless, guidelines for the ethical research and development of biometric AI technologies should focus on minimizing risks to privacy and civil liberties, while public awareness campaigns and educational initiatives play a vital role in informing both the public and organizations about biometric data protection and individual rights.

Enhancing the cybersecurity of AI systems is a critical task that demands a comprehensive and evolving approach, especially in sensitive sectors like banking. This involves enforcing AI security standards aligned with NIST guidelines for robust security and resilience. For example, AI systems in banking should undergo rigorous testing to ensure they meet resilience criteria, particularly in fraud detection. Key strategies should include developing customized security protocols tailored to various AI applications, such as enhanced data encryption in banking and a focus on patient data privacy in healthcare.

Establishing baseline security standards include covering secure coding, regular vulnerability assessments, and robust data encryption. Rigorous security testing and validation throughout the AI system's lifecycle are crucial, including stress testing under simulated attacks and independent security audits by certified third-party organizations. Building resilient AI systems involves developing dynamic threat response mechanisms and robust fail-safe and recovery procedures to maintain functionality or safely shut down during cyber-attacks. Advanced data protection and privacy measures, such as enhanced data encryption techniques and fine-grained access controls, are vital. This includes using encryption and access controls to safeguard sensitive AI functionalities and data.

Specialized training and awareness programs should target AI teams with AI-specific security challenges and promote organization-wide cybersecurity awareness. Robust incident response and management, utilizing AI-driven tools for rapid detection and response, and establishing clear incident reporting protocols are fundamental aspects of a strong cybersecurity framework. Compliance with international standards and regular legal reviews ensure that AI systems adhere to regulations like GDPR and ISO/IEC standards for information security management. Finally, companies must use specialized measures to counter AI-specific threats, like adversarial AI attacks and ensure secure AI decision-making processes. By adopting and continually updating these measures to address emerging threats, organizations can significantly enhance the cybersecurity of their AI systems, ensuring a higher level of resilience and trust in AI technologies.

Nations must also prevent AI weaponization, with a focus on creating a comprehensive framework that balances innovation with ethical responsibility. Drawing a parallel with the Treaty on the Non-Proliferation of Nuclear Weapons,[682] world leaders should create a global standard that prevents the use of AI in military aggression while encouraging its peaceful and beneficial applications. This requires the establishment of an international oversight body that could play a role similar to the International Atomic Energy Agency,[683] monitoring the development and application of AI technologies, particularly those with potential military uses. Regular inspections, assessments, and reporting would be key functions, ensuring transparency and compliance with the treaty.

The treaty should also set forth clear and strict guidelines for AI research and development, akin to the provisions in the Biological Weapons Convention.[684] These guidelines would delineate acceptable and unacceptable paths of AI development, particularly focusing on preventing research that could lead to weaponization. For example, AI technologies that could be

[682] Treaty on the Non-Proliferation of Nuclear Weapons, *opened for signature* July 1, 1968, 21 U.S.T. 483, 729 U.N.T.S. 161.

[683] *See generally*, *Overview*, INT'L ATOMIC ENERGY AGENCY, https://www.iaea.org/about/overview [https://perma.cc/7DF2-ZWD6] (last visited Feb. 16, 2024).

[684] Convention on the Prohibition of the Development, Production and Stockpiling of Bacteriological (Biological) and Toxin Weapons and on Their Destruction, *opened for signature* Apr. 10, 1972, 26 U.S.T. 583, 1015 U.N.T.S. 163.

used in autonomous weapons or for mass surveillance should be subject to stringent controls and oversight. The treaty should not only prohibit the development of such technologies but also include measures for disarmament and non-proliferation, similar to those found in the Chemical Weapons Convention.[685] This would ensure that existing AI technologies developed for military purposes are dismantled and that their proliferation is prevented.

Furthermore, the treaty should include provisions for international collaboration and sharing of best practices in AI development, promoting the peaceful and beneficial use of AI. It should encourage signatory nations to engage in cooperative research programs, focusing on AI applications that can contribute to societal well-being, such as healthcare, environmental protection, and education. This cooperative aspect would not only reduce the risks of AI weaponization but would also help bridge the technological gap between nations, promoting a more equitable global landscape in AI development.

In terms of global AI governance, a structure comparable to the World Health Organization[686] could be effective. This body would oversee

[685] Convention on the Prohibition of the development, Production, Stockpiling and Use of Chemical Weapons and on Their Destruction, *opened for signature* Jan. 13, 1993, 1974 U.N.T.S. 45.

[686] *About WHO*, WHO, https://www.who.int/about [https://perma.cc/Z6SY-Q6YG] (last visited Feb. 16, 2024).

global AI standards and manage AI-related issues, akin to how WHO oversees global health standards and responses to health crises. It would develop and enforce global AI standards and best practices, ensuring worldwide adherence to ethical, safety, and security norms. The regulatory body could also regularly assess the "health" of global AI systems, evaluating them for ethical alignment, security vulnerabilities, and societal impact, mirroring the WHO's role in assessing global health trends and risks. Furthermore, in response to AI-related crises, such as security breaches or ethical violations in AI deployment, this body would coordinate rapid and effective responses. Finally, fostering international collaboration in AI research, focusing on ethical and secure development, would parallel collaborative research initiatives in global health.

To address enforcement and deter wrongdoing, we recommend a tiered system of sanctions. By imposing penalties based on the severity of deviations from NIST standards, this system introduces a level of precision and fairness in regulatory enforcement. It differentiates between unintentional errors and intentional non-compliance, promoting a culture of responsibility and ethical awareness. For example, an inadvertent use of a biased AI in hiring could attract a lesser penalty compared to a deliberate breach of ethical AI standards. This approach is especially effective in complex areas like algorithmic bias, ensuring AI applications in sensitive domains are equitable. However, the challenge lies in the subjective nature of assessing intent and context, which could lead to inconsistencies in penalty application.

We also favor regional tailoring of AI regulations, which allows for the creation of contextually relevant regulations that address the unique socio-economic challenges of different regions. For example, in areas facing high unemployment due to AI-driven automation, regulations could focus on AI's role in job creation or retraining programs. This localized approach ensures that AI benefits are more evenly distributed and tailored to specific needs. On the other hand, this could result in a fragmented regulatory landscape, potentially complicating compliance for multinational companies and hindering the seamless integration of AI technologies across borders.

To conclude, in developing future AI legislation, we argue for the integration of NIST principles of trustworthy and responsible AI, encompassing aspects like validity, reliability, safety, security, accountability, transparency, explainability, privacy, and fairness.[687] This legislative framework should include the establishment of AI Ethics Review Boards for ensuring ethical compliance, focus on mitigating AI-induced job displacement through retraining programs, and introduce energy

[687] *Trustworthy and Responsible AI: Overview*, *supra* note 678.

consumption caps for AI training to address environmental concerns. Additionally, protocols for combating deepfakes, misinformation, and surveillance AI are essential to maintain the integrity and responsible use of AI technologies while balancing human rights concerns. The legislative approach should also entail the creation of a licensing agency for AI systems in sensitive sectors such as healthcare and finance, ensuring adherence to safety and privacy standards.

Governments must prioritize the transparency and explainability of AI decision-making processes, particularly in sectors impacting public trust and market integrity. An international treaty to prevent AI weaponization, guided by NIST principles, along with the formation of a global AI governance body, will be pivotal in establishing a comprehensive international regulatory framework. We advise legislators to impose meaningful sanctions for regulatory violations and allow regional flexibility in AI regulation. This ensures that AI development is not hindered while maintaining global standards and addressing specific regional socio-economic conditions. As AI technology continues to advance rapidly, legislative measures must be flexible and responsive, aiming to harness AI's potential responsibly for societal benefit. This proposed roadmap offers guidance for legislators to create a regulatory environment that promotes ethical AI innovation, protects public interests, and ensures global competitiveness so that AI can indeed solve the world's greatest challenges.