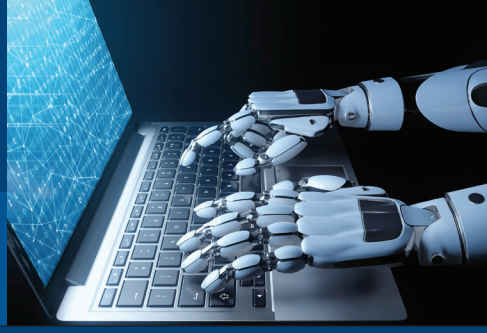
Association for Computing Machinery

acm Technology Policy Council

TechBriefs



GENERATIVE ARTIFICIAL INTELLIGENCE

by David Leslie and Francesca Rossi

PROBLEM

The rapid commercialization of generative AI (GenAI) poses multiple large-scale risks to individuals, society, and the planet that require a rapid, internationally coordinated response to mitigate.

POLICY IMPLICATIONS

- The absence of comprehensive and coherent guidelines for the development and deployment of GenAI systems, and their consequent proliferation, creates massive individual, societal, and socioeconomic risks.
- Swift and proportionate policy action is needed, at national and international levels,[1] to meet the challenges posed by the expanding scale and scope of GenAI-related risks.
- Wide disparities in the economic influence of GenAI system stakeholders have the potential, if not fully addressed by policy, to amplify inequality and thwart innovation and competition.

GENERATIVE AI: BY THE NUMBERS

8	Estimated value in billions of US dollars of the global GenAI market in 2023.[2]
65	Months it took Twitter to reach 100 million users after launch.[3]
2	Months it took ChatGPT to achieve that milestone.[2]
2	Rank of ChatGPT among all apps that have ever reached that mark.[2]
43	Estimated percentage of college students who have used ChatGPT or a similar app.[4]
80	Estimated percentage of U.S. workers who will have at least 10% of their work tasks affected by GenAI.[5]
19	Percentage of such workers who will have at least 50% of their tasks so affected.[4]
3.4	Gigawatt hours of electricity used to train Google's PaLM for two months.[6]
321	Number of U.S. homes whose needs could be fully met with that power for a full year.[7]
59	Millions of total words used in the 20-volume *Oxford English Dictionary* (2nd ed.).[9]
1,400,000	Millions of words (or word parts) in the dataset used to train Meta AI's Llama.[8]
76	Estimated percentage of U.S. consumers concerned about GenAI-produced misinformation.[10]

Overview

Generative AI (GenAI) refers to a broad set of computing technologies[11] and techniques that enable a computer to create content such as text, code, image, audio, speech, voice, music, and video.[2,12] Over the past decade, the development of sophisticated neural network architectures, combined with increasing computing capacity and access to vast bodies of training data, has led to a great expansion in their utility and uptake.

GenAI systems often rely on the use of so-called foundation models (FMs).[13] FMs are a subset of GenAI technologies that are pretrained on large quantities of data through "self-supervised learning" and that serve as base models. They can be fine-tuned, converted into diverse task-specific applications, and adapted to complete a wide range of downstream functions.[14]

GenAI, however, is not without serious inherent technical limitations that can make it unwise, unjust, and even unsafe to employ in certain spheres.[15] For example, language-based Gen AI systems like ChatGPT produce outputs by predicting the word sequences most likely to be appropriate given the statistical patterns contained in their training data.[16] Consequently, those outputs do not base their predictions on human perception.[17] In practical terms, they thus can "hallucinate" (i.e. , make up plausible-sounding but factually incorrect or baseless responses that may mislead or harm users).

GenAI's use can be unwise, unjust, and unsafe.

Moreover, because GenAI systems are so complex it is nearly impossible for humans to explain, validate, or even understand the rationale underlying their outputs. This "black box" character is especially problematic in high-impact or safety-critical domains like law, medicine, transportation, and human resources, among others.[18] It also makes it difficult to identify and control possible system vulnerabilities, faults, harmful behaviors, or biases.

GenAI poses risks to people, society, and the environment

The rapid growth and commercialization of GenAI systems is unprecedented. This has set off intense competition in GenAI with large tech companies quickly integrating these systems into their flagship services and hundreds of GenAI start-ups materializing across virtually every commercial sector.[19] The explosive commercialization of GenAI has introduced numerous large-scale risks to individuals and society. The environmental costs of these technologies, and of their supporting infrastructures, are also significant.

GenAI has introduced large-scale risks to individuals and society.

At the individual level, GenAI applications that interact with people, including minors, pose risks of discrimination, bias, exposure to harmful or hateful speech, violations of intellectual property, privacy and data protection rights, and behavioral manipulation.[20] By virtue of the uncurated nature of the datasets on which these systems are trained, GenAI applications can and do replicate social stereotyping, discriminatory and toxic language, imagery, and other learned representations in their outputs and produce biased decisions. Similarly, these applications can leak, infer, and expose sensitive personal data buried within their opaque architectures.

In addition, conversation agents like ChatGPT and Pi[21] that are trained to engage in humanlike dialogue can lead people to believe they are interacting with "sentient" or "intelligent" agents or other human beings, thus exposing users to potential manipulation and loss of agency. Meanwhile, intentionally malicious uses of GenAI systems can harm individuals by enabling large-scale fraud, facilitating cyberattacks, producing malware, and providing bad actors with data in support of bioterrorism, chemical warfare, or other hostile activities.

At the societal level, GenAI applications pose large-scale risks to the integrity of information ecosystems, the functioning of democratic society, and socioeconomic sustainability.[22] The irresponsible or malicious development and use of GenAI technologies could lead to the scaled production of disinformation, propaganda, and information that is false but sounds true, potentially flooding the digital public square with misleading and nonfactual content. This could undermine social trust in the information ecosystem and tear the fabric of reliable public communication. Moreover, the increasing volume of AI-generated text alone that will inevitably find its way into training data threatens to pollute later generations of AI.[23]

On the socioeconomic front, the breakneck commercialization of GenAI systems could lead to an unchecked consolidation of power by the few large tech firms that disproportionately control data, computing, and skills infrastructures. Even if there are also many small AI

companies and open-source efforts in the GenAI space, this centralization of power could shape the broader distribution of corresponding private and public benefits and risks.[24]

The irresponsible development and use of GenAI also could impose significant environmental costs. Such systems require large amounts of power for model development and training, system operation, and data storage. In addition, substantial carbon emissions may be generated by associated hardware manufacturing and infrastructure creation.[25]

GenAI risk management requires technically nuanced policymaking

Policymakers confronting this range of risks face complex challenges. AI law and policy thus should incorporate end-to-end governance approaches that address risks comprehensively and "by design."[26] Specifically, they must address how to govern the multiphase character of GenAI systems and the foundation models used to construct them.[27] For instance, liability and accountability for lawfully acquiring and using initial training data should be a focus of regulations tailored to the FM training phase.

AI policy should codify human responsibility across the GenAI workflow.

Guardrails in law and policy for the protection of stakeholders from harms attributable to operation of a GenAI system might better be geared to parties operating at the application phase. Important early design-stage mechanisms could include certification procedures,[28] context-based risk analysis, impact assessment, stakeholder engagement,[29] bias self-assessment, assurance processes for data quality, representativeness, privacy- and explainability-aware design,[30] watermarking regimes,[31] and reporting protocols that ensure all these governance mechanisms are properly documented and made transparent and accessible to relevant parties. Important system post-deployment mechanisms include clear and understandable explanations, traceability, auditability, correctability, accountability and responsibility requirements,[12] third-party oversight, and processes to ensure effective remedies.[32]

Policymakers must also confront the governance challenges presented by the complex and distributed supply chains that feed GenAI life cycles. Many of these systems are made up of parts or elements that derive from multiple suppliers, vendors, contractors, and open-source assets. This means that effective AI policy should codify multiparty, gapless, and end-to-end accountability and transparency mechanisms, which establish a continuous chain of human responsibility across the entire GenAI project workflow.[12]

Policymakers must pay close attention to potential power imbalances at the ecosystem level that could substantially affect the public interest. When large-scale private sector organizations control the critical digital infrastructures on which the production and use of FMs and GenAI applications depend, this may create distorted financial incentives to further centralize economic power and disadvantage smaller or less well-resourced commercial, academic, and public sector stakeholders.[33]

Effective AI policy should consider addressing such power asymmetries directly. Crucially, this would involve discouraging the largest FM and GenAI system innovators from embracing the "move fast and break things" credo of system design and instituting appropriate accountability mechanisms throughout the complex AI value chain.

KEY CONCLUSIONS

- AI policy should incorporate end-to-end governance approaches that address risks "by design" and regulate at all stages of the design-to-deployment life cycle of AI products.
- Governance mechanisms for GenAI technologies must address the entirety of their complex supply chains.
- Actors should be subject to controls that are proportionate to the scope and scale of the risks their products pose, with corresponding legal liability and other concrete consequences for irresponsible practices.

NOTES AND SOURCES

1. The EU, China, and the U.S. have all issued substantial policy guidance on generative AI. Nancy A. Fischer et al., "Unleashing the AI Imagination: A Global Overview of Generative AI Regulations," Pillsbury Law, August 11, 2023, https://www.pillsburylaw.com/en/news-and-insights/ai-regulations-us-eu-uk-china.html.
2. Manish Goyal, Shobhit Varshney, and Eniko Rozsa, "What Is Generative AI, What Are Foundation Models, and Why Do They Matter?," *IBM Blog*, March 8, 2023, https://www.ibm.com/blog/what-is-generative-ai-what-are-foundation-models-and-why-do-they-matter/.
3. Pallavi Rao, "How Long It Took for Popular Apps to Reach 100 Million Users," *VisualCapitalist*, July 13, 2023, https://www.visualcapitalist.com/threads-100-million-users/.
4. Lyss Welding, "Half of College Students Say Using AI on Schoolwork Is Cheating or Plagiarism," BestColleges, March 27, 2023, https://www.bestcolleges.com/research/college-students-ai-tools-survey/.
5. Tyna Eloundou et al., "GPTS Are GPTS: An Early Look at the Labor Market Impact Potential of Large Language Models," last updated August 21, 2023, arXiv:2303.10130.
6. Anil Ananthaswamy, "In AI, Is Bigger Always Better?" *Nature* 615, no. 7951 (March 2023): 202–205.
7. "How Much Electricity Does an American Home Use?," U.S. Energy Information Administration, last updated October 12, 2022, https://www.eia.gov/tools/faqs/faq.php?id=97&t=3.
8. "Introducing LLaMA: A Foundational, 65-Billion-Parameter Large Language Model," MetaAI, February 24, 2023, https://ai.facebook.com/blog/large-language-model-llama-meta-ai/.
9. "Dictionary Facts," *Oxford English Dictionary Online*, retrieved June 1, 2014, cited by https://en.wikipedia.org/wiki/Oxford_English_Dictionary#cite_note-facts2004-7.
10. Kathy Haan, "Over 75% of Consumers Are Concerned About Misinformation From Artificial Intelligence," *Forbes*, July 20, 2023, https://www.forbes.com/advisor/business/artificial-intelligence-consumer-sentiment/.
11. Because of their general-purpose character and knowledge transfer capability, FMs can be converted into diverse task-specific applications and adapted to complete a wide range of downstream functions. Sherry Yang et al., "Foundation Models for Decision Making: Problems, Methods, and Opportunities," submitted on March 7, 2023, https://arxiv.org/abs/2303.04129.
12. ACM Technology Policy Committee, "Principles for the Development, Deployment, and Use of Generative AI Technologies," June 27, 2023; Helen Toner, "What Are Generative AI, Large Language Models, and Foundation Models?," Center for Security and Emerging Technology, May 12, 2023, https://cset.georgetown.edu/article/what-are-generative-ai-large-language-models-and-foundation-models/.
13. The term "foundation model" was coined by Stanford researchers in 2021, though this nomenclature has faced criticism by some scholars for being vague and misleading. Others have criticized the use of the term as an attempt to assert epistemic hegemony and to siphon funding toward this "foundational" technology. For the original Stanford paper, see: Rishi Bommasani et al., "On the Opportunities and Risks of Foundation Models," August 16, 2021, https://arxiv.org/abs/2108.07258. For criticisms, see: Will Knight, "A Stanford Proposal Over AI's 'Foundations' Ignites Debate," *Wired*, September 14, 2021, https://www.wired.com/story/stanford-proposal-ai-foundations-ignites-debate/; Gary Marcus, "Has AI Found a New Foundation?," *The Gradient*, September 11, 2021, https://thegradient.pub/has-ai-found-a-new-foundation/; Brent Orrell, "How AI Is Being Transformed by 'Foundation Models,'" *The Bulwark*, May 2, 2022, https://www.thebulwark.com/how-ai-is-being-transformed-by-foundation-models/.
14. These include text summarization, conversation, multilingual translation, protein and chemical structure prediction, molecular property prediction, image classification, caption prediction, planning, robotic control, and generation of computer code. Xavier Amatriain et al., "Transformer Models: An Introduction and Catalog," last updated May 25, 2023, https://arxiv.org/abs/2302.07730. See also Alec Radford et al., "Learning Transferable Visual Models From Natural Language Supervision," in *Proceedings of the 38th International Conference on Machine Learning* 139, eds. Marina Meila and Tong Zhang, 8748–8763, http://proceedings.mlr.press/v139/radford21a.html; Aditya Ramesh et al., "Hierarchical Text-Conditional Image Generation with CLIP Latents," April 13, 2022, https://arxiv.org/abs/2204.06125.
15. Yihan Cao et al., "A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT," March 7, 2023, https://arxiv.org/abs/2303.04226.
16. Murray Shanahan, "Talking About Large Language Models," last updated February 16, 2023, https://arxiv.org/abs/2212.03551; Emily M. Bender and Alexander Koller, "Climbing Towards NLU: On Meaning, Form, and Understanding in the Age of Data," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, July 2020, https://doi.org/10.18653/v1/2020.acl-main.463.
17. Abeba Birhane et al., "Science in the Age of Large Language Models," *Nature Reviews Physics* 5 (2023), https://www.nature.com/articles/s42254-023-00581-4.
18. Luca Malinverno et al., "Explainable AI in Biomedical Research: A Systematic Review and Meta-Analysis," Social Science Research Network, January 24, 2023, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4335108.
19. Christophe Carugati, "The Age of Competition in Generative Artificial Intelligence Has Begun," Bruegel, May 11, 2023, https://www.bruegel.org/first-glance/age-competition-generative-artificial-intelligence-has-begun.
20. Emily M. Bender et al., "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?," in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, eds. Michael Ekstrand and Ana Paiva, March 2021, https://doi.org/10.1145/3442188.3445922; Rishi Bommasani et al., "On the Opportunities and Risks of Foundation Models," arXiv:2108.07258 [cs], (2021), https://arxiv.org/abs/2108.07258; Laura Weidinger et al., "Taxonomy of Risks Posed by Language Models," in *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, eds. Michael Ekstrand and Ana Paiva, July 2021, https://doi.org/10.1145/3461702.3462617.
21. Alex Konrad, "Inflection AI, Startup from Ex-DeepMind Leaders Launches Pi—A Chattier Chatbot," *Forbes*, May 2, 2023, https://www.forbes.com/sites/alexkonrad/2023/05/02/inflection-ai-ex-deepmind-launches-pi-chatbot/?sh=d0349073d6dd.
22. Leon Derczynski et al., "Assessing Language Model Deployment with Risk Cards," arXiv preprint arXiv:2303.18190 (2023), https://arxiv.org/pdf/2303.18190.pdf; Renee Shelby et al., "Sociotechnical Harms: Scoping a Taxonomy for Harm Reduction," arXiv:2210.05791 [cs](2022), https://arxiv.org/abs/2210.05791; Laura Weidinger et al., "Taxonomy of Risks Posed by Language Models."
23. Such a stream of false and questionable information could also propagate the distortion into downstream datasets as the automated production of nonfactual digital content that is indistinguishable from human creations becomes a major component by the volume of humanity's digital archive. Ilia Shumailov et al., "The Curse of Recursion: Training on Generated Data Makes Models Forget," arXiv preprint arxiv:2305.17493 (2023), https://arxiv.org/pdf/2305.17493.pdf
24. Along these lines, insofar as the near- and long-term future of the commercialization of GenAI is driven primarily by market imperatives, society will experience greater wealth polarization, global inequality, labor disruption, and the loss of vulnerable industries and vocational subpopulations (such as the creative professions).
25. Loïc Lannelongue, John Grealy, and Michael Inouye, "Green Algorithms: Quantifying the Carbon Footprint of Computation," *Advanced Science* 8, no. 12 (May 2021), https://doi.org/10.1002/advs.202100707; Roy Schwartz et al., "Green AI," *Communications of the ACM* 63, no. 12 (2020): 54–63, https://doi.org/10.1145/3418512; Emma Strubell, Ananya Ganesh, and Andrew McCallum, "Energy and Policy Considerations for Deep Learning in NLP," arXiv:1906.02243 [cs] (2019), https://arxiv.org/abs/1906.02243/.
26. David Leslie, "Understanding Artificial Intelligence Ethics and Safety: A Guide for the Responsible Design and Implementation of AI Systems in the Public Sector," Zenodo, June 11, 2019, https://doi.org/10.5281/zenodo.3240529.
27. Jakob Mökander et al., "Auditing Large Language Models: A Three-Layered Approach," *AI and Ethics* 3, no. 3 (August 2023): 1–31, https://link.springer.com/article/10.1007/s43681-023-00289-2.
28. Peter Cihon et al., "AI Certification: Advancing Ethical Practice by Reducing Information Asymmetries," *IEEE Transactions on Technology and Society* 2, no. 4 (2021): 200–209.
29. David Leslie et al., "Human Rights, Democracy, and the Rule of Law Assurance Framework for AI Systems: A Proposal," Zenodo, February 6, 2022, https://doi.org/10.5281/zenodo.5981676.
30. Information Commissioner's Office and the Alan Turing Institute, "Explaining Decisions Made with AI," 2020, https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/explaining-decisions-made-with-ai/.
31. Philipp Hacker, Andreas Engel, and Marco Mauer, "Regulating ChatGPT and Other Large Generative AI Models," in *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, June 2023: 1112–1123; John Kirchenbauer et al., "A Watermark for Large Language Models," last updated June 6, 2023, arXiv preprint arXiv:2301.10226.
32. "AI Language Models: Technological, Socio-Economic and Policy Considerations," OECD Digital Economy Papers, no. 352 (Paris: OECD Publishing, 2023), https://doi.org/10.1787/13d38f92-en.
33. Nur Ahmed, Muntasir Wahed, and Neil C. Thompson, "The Growing Influence of Industry in AI Research," *Science* 379, no. 6635 (March 3, 2023): 884–886, https://www.science.org/doi/10.1126/science.ade2420; Madhumita Murgia, "Risk of 'Industrial Capture' Looms over AI Revolution," *Financial Times*, March 22, 2023, https://www.ft.com/content/e9ebfb8d-428d-4802-8b27-a69314c421ce.

ADDITIONAL INFORMATION

With 100,000 members in 190 countries, the nonprofit **Association for Computing Machinery** is the world's largest and longest-established organization of professionals involved in all aspects of computing. Under the auspices of the global ACM Technology Policy Council, policy committees in the U.S. and Europe provide cutting-edge, apolitical, non-lobbying information about computing and its social impacts to policy makers at all levels of government in many forms, including briefings, testimony, consultation, and rulemaking comments, reports, and analyses.

To tap the deep expertise of ACM's global membership, please contact ACM's Global Policy Office at acmpo@acm.org or +1 202.580.6555. To receive *ACM TechBriefs* quarterly, in the body of a one-line email send *subscribe ACM-tpc-tech-briefs* followed by your first and last names to *listserv@listserv.acm.org*.

AUTHORSHIP

David Leslie is Director of Ethics and Responsible Innovation Research at the Alan Turing Institute and Professor of Ethics, Technology and Society, the Digital Environment Research Institute at Queen Mary University of London. Francesca Rossi is an IBM Fellow and the company's AI Ethics Global Leader. She is also a fellow of the Association for the Advancement of Artificial Intelligence (AAAI), the European Association for Artificial Intelligence (EurAI), a Radcliffe fellow, and the current president of AAAI. This brief was produced for the ACM Technology Policy Council and may be cited as "*ACM TechBrief: Generative Artificial Intelligence*, ACM Technology Policy Council (Issue 8, Summer 2023)."