Quantifying the Risks of Generative Artificial Intelligence

GenAI can revolutionize business processes and boost worker efficiency—but organizations must first mitigate potential risks.

Since Google's paper about its transformer model came out in 2017, the field of generative artificial intelligence (GenAI) has exploded. From the introduction of OpenAI's GPT-1 model in 2018 to the much-hyped launch of GPT-4 in 2023, the technology has quickly evolved from an innovation to a paradigm-shifting disruption. It seems like every week there's an exciting new AI announcement from a large player like OpenAI, Stability AI, Google, or Meta, or from a smaller AI startup like Anthropic, Hugging Face, or ElevenLabs.

Much of GenAI's early development had occurred quietly and away from public view. Prior to ChatGPT putting it into the hands of consumers in late 2022, it was something that Big Tech approached with caution. In 2016, when Microsoft's newly released GenAI chatbot began spewing racist messages,[1] companies decided to take a more careful approach. Big Tech's GenAI advancements (including better ways to mitigate ethical and safety issues) were primarily developed in private, but the launch of ChatGPT put pressure on other technology companies to introduce their own GenAI products.

Some fear that significant problems could arise as powerful but imperfect AI systems are made widely available despite lingering safety risks. In May 2023, a group of 350 researchers and technology leaders expressed a need for more caution around GenAI[2] in an open letter. Not long after, one of GenAI's pioneers, Dr. Geoffrey Hinton, quit his role at Google to focus on raising awareness of the dangers of the technology. In this paper, we'll explore the core risks that could affect broad adoption of generative AI, focusing on safety, privacy, and regulatory concerns.

While providing incorrect information during a business inquiry might potentially inconvenience a customer, other types of misinformation could be far more costly.

Readiness Challenges

Despite these concerns, private and public organizational leaders alike are feeling similar pressure to adopt GenAI to help solve operational problems, increase efficiency, and achieve competitive advantages. Yet very few are ready to do so according to a survey of senior executives across commercial and public sectors that Guidehouse conducted in partnership with CDO Magazine between November 2023 and January 2024.[3]

The resulting report, "The State of GenAI Today: The Early Stages of a Revolution," revealed that more than three-quarters (76%) of respondents said their organizations are not fully equipped to harness the power of GenAI. While about the same percentage (74%) indicated that they are likely to invest in GenAI projects over the next 12 months, their investments will be modest, with most allocating less than 5% of their IT budgets to GenAI initiatives in 2024.

That conservative approach is due in part to concerns about the inherent risks of this new technology. For GenAI to be broadly adopted across industries, a thorough focus on safety, privacy, and regulatory concerns is needed first.

Safety

Chief among most experts' concerns about GenAI are its safety implications. AI safety encompasses a broad range of issues including brand safety, bias, unethical or malicious use cases, and misinformation.

Brand safety

Organizational leaders responsible for protecting their respective brands have a number of reputational concerns when it comes to GenAI. All AI models can generate false information, which can lead to, damaging a company's reputation as a trusted brand. GenAI customer service agents speaking to customers in real time might pass along inaccurate information or even show bias. Some consumer-focused brands have already faced pushback for using AI. For example, customers have called out brands for featuring AI art in advertisements, citing ethical concerns related to pending lawsuits by artists[4] against companies that used their copyrighted art as training data without permission or licensing. The U.S. Federal Trade Commission (FTC) addressed concerns like this in its December 2023 report on GenAI and the creative economy.[5]

Bias

AI systems reproduce historic biases in their training data. If an AI model ingests texts using hateful language about a racial group, the system's output might reproduce those biases. In healthcare, the use of anonymized historical patient data or research in model training might reproduce that data's existing biases, leading to substandard patient care. For instance, an AI system designed to help with diagnostics might not flag common pregnancy risks in parts of the population that have been historically underdiagnosed due to racial bias.

Guidehouse

Organizations that decide to adopt GenAI should create risk management plans to address the technology's safety concerns.

Unethical or malicious use cases

Some fear that GenAI could be used for scams, deep fake videos, attacks, and mass political misinformation via social media bots. Unethical businesses could use the technology to spam their competitors' online reviews; scammers could use it to impersonate a company or a government institution; and foreign governments could use it in cyber warfare. Organizations will have to be vigilant to protect themselves, their employees, and their customers from these types of attacks.

Misinformation

One of the big risks of GenAI is the spreading of misinformation. While providing incorrect information during a business inquiry might potentially inconvenience a customer, other types of misinformation could be far more costly. For example, some companies are considering using GenAI to help with compliance and reporting documents. Mistakes in those documents could incur large fines in some sectors.

While these are serious risks, they are far from insurmountable. Organizations that decide to adopt GenAI should create risk management plans to address the technology's safety concerns. What those plans will look like will depend on the type of AI an organization is using, how it is using it, and the relevant risks. Mitigation plans could include fact-checking by external systems, human oversight, extensive fine-tuning to control biases, or custom guardrails.

Organizations looking to deploy GenAI technology face more risks because they will be doing so without a full understanding of how emerging regulations may affect implementation.

Privacy

A growing number of privacy challenges has spurred exploration and initiation of government regulation across the globe. In March 2023, Italy's data protection watchdog temporarily banned[6] ChatGPT in the country due to potential European Union General Data Protection Regulation (GDPR) violations and the technology's lack of age-gating. Since then, Canada's privacy commissioner has expressed concern over the company's privacy standards[7] and the U.S. Senate Judiciary Subcommittee on Privacy, Technology, and the Law held hearings on OpenAI and privacy.[8]

The FTC also opened an investigation[9] in July 2023 into whether the company violated consumer protection laws around illegal data collection, privacy violations, and the dissemination of false information about individuals. Probing how OpenAI's data leaks occurred is one more way[10] to better understand GenAI's impact on consumer privacy.

Organizations implementing GenAI should pay careful attention to privacy risks regarding training data, the use and storage of inputted prompt data, and their technology providers' history of data breaches and privacy protection. Ensuring that AI systems don't use or store user-provided data as training data is critical. Some sectors might also want to get permission, via waivers, from customers or patients when collecting personal data for use in AI systems. Organizations should adapt existing privacy policies to cover the additional privacy risks and emerging regulations pertaining to GenAI.

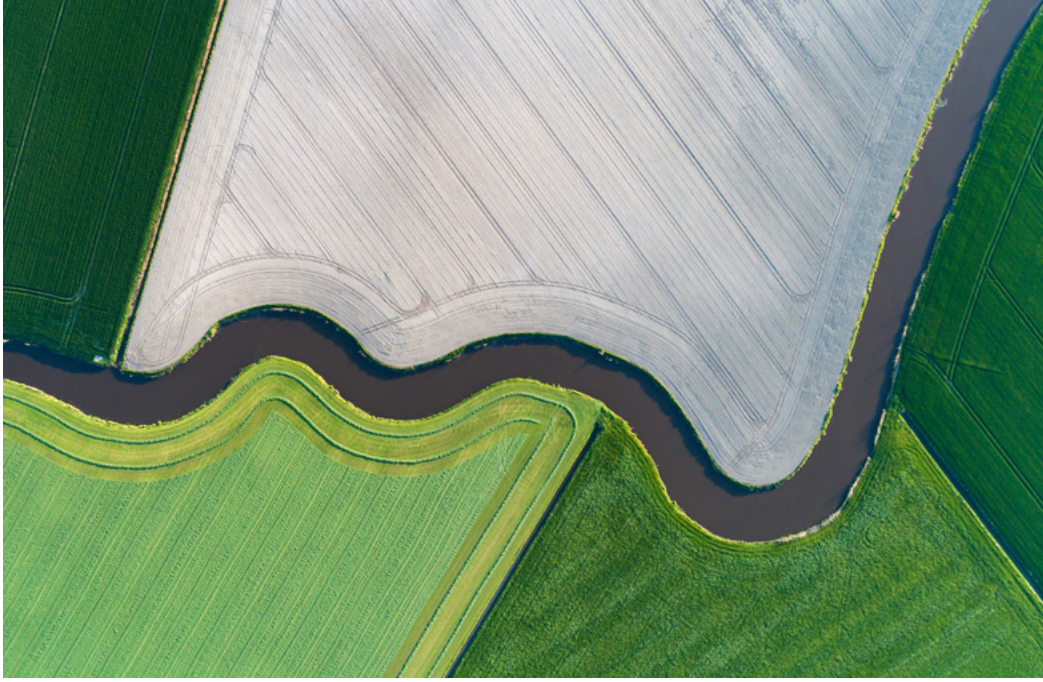
Regulatory moves

Governments and regulatory bodies around the world are rushing to regulate GenAI. The EU's AI Act[11] was one of the first comprehensive AI regulatory frameworks to be released. In the U.S., the AI Bill of Rights[12], while focused on the federal government's use of AI, hints at an approach that U.S. government agencies could enforce more broadly in the future. Multiple groups and agencies are working on AI regulations in the U.S. The FTC, for example, kicked off 2024 by holding an AI tech summit, proposing protections to combat AI impersonation, and launching an inquiry into Big Tech's GenAI investments and partnerships. [13, 14]

On the legislative side, a bipartisan group of U.S. Congressional leaders introduced a bill in June 2023[15] to create a commission that would focus on regulating AI by determining how to mitigate risks and harms while protecting innovation. Another aim of the legislation is for the commission to review the ways review the ways that agencies across the government currently regulate or provide oversight to AI to determine whether a new standalone AI regulatory agency is needed. Then on October 30, 2023, President Biden signed an executive order titled, "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence."[16] Part of the order requires the U.S. Secretary of Commerce to coordinate efforts among relevant agency heads to establish guidelines and best practices—with the goal of promoting consensus industry standards—for developing and deploying safe, secure, trustworthy AI systems within 270 days of the order.

Guidehouse

Organizations should make a risk management plan that anticipates some of the regulations likely to emerge in the coming years.



On a more granular level, some state and city governments have created AI regulations, such as New York City's regulations[17] on the use of AI in hiring and promotion decisions. Meanwhile, sector-specific regulatory agencies may create regulations of their own. While many proposed laws and regulations are being developed but not yet in place, organizations looking to deploy GenAI technology face more risks because they will be doing so without a full understanding of how emerging regulations may affect implementation.

Voluntary Measures

In July 2023, a number of tech companies, including OpenAI and Google, agreed to implement voluntary measures[18] for greater AI transparency and safety. This includes efforts such as watermarking content produced by AI models to guard against deepfakes and pledging to protect user privacy. Additional voluntary guidelines may emerge in the future.

Copyright Considerations

The U.S. Copyright Office[19] is also paying close attention to AI. In March 2023, it launched an initiative to examine issues around GenAI and copyright law. While content produced by AI models alone can't be copyrighted. Works that contain AI-generated materials and sufficient human authorship could potentially be eligible for copyright protection—but that protection would only cover the human-generated portions. For example, a company that generates a logo using AI wouldn't be able to copyright it, but a company that creates an eBook with some AI-generated images would be able to copyright the portions not generated by AI.

Next Steps

Organizations seeking to integrate AI into their operations face the challenge of doing so before the regulatory environment is transparent and defined. Until that materializes, companies might benefit from studying foreign regulations such as the EU's AI Act, a good template for regulatory action. Organizations should make a risk management plan that anticipates some of the regulations likely to emerge in the coming years.

Guidehouse is uniquely placed to help create a comprehensive GenAI strategy.

While GenAI's current and potential risks should be taken seriously, there are also significant benefits to its adoption. An effective GenAI plan should make risk management a central focus. Integrated GenAI risk management plans will help organizations ensure they're holistically assessing potential impacts of the technology before adopting it—and that they have clear plans for mitigating the risks that GenAI presents.

As an organization with significant proficiency in risk management and industry-specific compliance requirements, Guidehouse is uniquely placed to help create a comprehensive GenAI strategy. With deep experience in both corporations and governmental organizations, Our cross-functional experts help organizations manage risks while maximizing the benefits of new technologies like GenAI.

1 "Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day", March 24, 2016. (theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist).

2 "A.I. Poses 'Risk of Extinction,' Industry Leaders Warn," May 30, 2023. (nytimes.com/2023/05/30/technology/ai-threat-warning.html).

3 "The State of GenAI Today: The Early Stages of a Revolution," Guidehouse and CDO Magazine, April 2024 (https://guidehouse.com/insights/advanced-solutions/2024/the-state-of-genai-today).

4 "Artists Are Suing Artificial Intelligence Companies and the Lawsuit Could Upend Legal Precedents Around Art", May 5, 2023. (artnews.com/art-in-america/features/midjourney-ai-art-image-generators-lawsuit-1234665579/).

5 "Generative Artificial Intelligence and the Creative Economy Staff Report: Perspectives and Takeaways," December 15, 2023 (https://www.ftc.gov/system/files/ftc_gov/pdf/12-15-2023AICEStaffReport.pdf).

6 "Italy reverses ban on ChatGPT after OpenAI agrees to watchdog's demands", May 3, 2023. (foxbusiness.com/technology/italy-reverses-ban-chatgpt-openai-agrees-watchdogs-demands).

7 "Canada to launch probe into OpenAI over privacy concerns", May 26, 2023. (reuters.com/technology/canada-launch-probe-into-openai-over-privacy-concerns-2023-05-25/).

8 "OpenAI CEO embraces government regulation in Senate hearing", May 16, 2023. (nbcnews.com/tech/tech-news/openai-ceo-embraces-government-regulation-senate-hearing-rcna83931).

9 "F.T.C. Opens Investigation Into ChatGPT Maker Over Technology's Potential Harms", July 13, 2023. (nytimes.com/2023/07/13/technology/chatgpt-investigation-ftc-openai.html).

10 "March 20 ChatGPT outage: Here's what happened", March 24, 2023. (openai.com/blog/march-20-chatgpt-outage).

11 "The EU Artificial Intelligenc Act", June 14, 2023. (artificial-intelligence-act.com/).

12 "Blueprint for an AI Bill of Rights", n.d., (whitehouse.gov/ostp/ai-bill-of-rights/).

13 "FTC Proposes New Protections to Combat AI Impersonation of Individuals," February 15, 2024 (https://www.ftc.gov/news-events/news/press-releases/2024/02/ftc-proposes-new-protections-combat-ai-impersonation-individuals).

14 "FTC Launches Inquiry into Generative AI Investments and Partnerships," January 25, 2024 (https://www.ftc.gov/news-events/news/press-releases/2024/01/ftc-launches-inquiry-generative-ai-investments-partnerships).

15 "AI Regulation Is Coming To The U.S., Albeit Slowly", June 27, 2023. (forbes.com/sites/washingtonbytes/2023/06/27/ai-regulation-is-coming-to-the-us-albeit-slowly/?sh=38af42387ee1).

16 "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence," October 30, 2023 (https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/).

17 "How New York is Regulating AI", June 22, 2023. (nytimes.com/2023/06/22/nyregion/ai-regulation-nyc.html).

18 "OpenAI, Google, others pledge to watermark AI content for safety, White House says", July 21, 2023. (reuters.com/technology/openai-google-others-pledge-watermark-ai-content-safety-white-house-2023-07-21/).

19 "U.S. Copyright Office Provides Guidance on Registrations involving AI-Generated Works", March 22, 2023. U.S. Copyright Office Provides Guidance on Registrations involving AI-Generated Works | White & Case LLP (whitecase.com).

Contacts

Bob Dunmyer, Partner
Data, Analytics & Intelligence
bdunmyer@guidehouse.com

Bassel Haidar, Director
Data, Analytics & Intelligence
bhaidar@guidehouse.com

About Guidehouse

Guidehouse is a leading global provider of consulting services to the public sector and commercial markets, with broad capabilities in management, technology, and risk consulting. By combining our public and private sector expertise, we help clients address their most complex challenges and navigate significant regulatory pressures focusing on transformational change, business resiliency, and technology-driven innovation. Across a range of advisory, consulting, outsourcing, and digital services, we create scalable, innovative solutions that help our clients outwit complexity and position them for future growth and success. The company has more than 16,000 professionals in over 55 locations globally. Guidehouse is led by seasoned professionals with proven and diverse expertise in traditional and emerging technologies, markets, and agenda-setting issues driving national and global economies. For more information, please visit guidehouse.com.

guidehouse.com/services/data-analytics-intelligence

@GHTechSolutions linkedin.com/showcase/guidehouse-technology-solutions