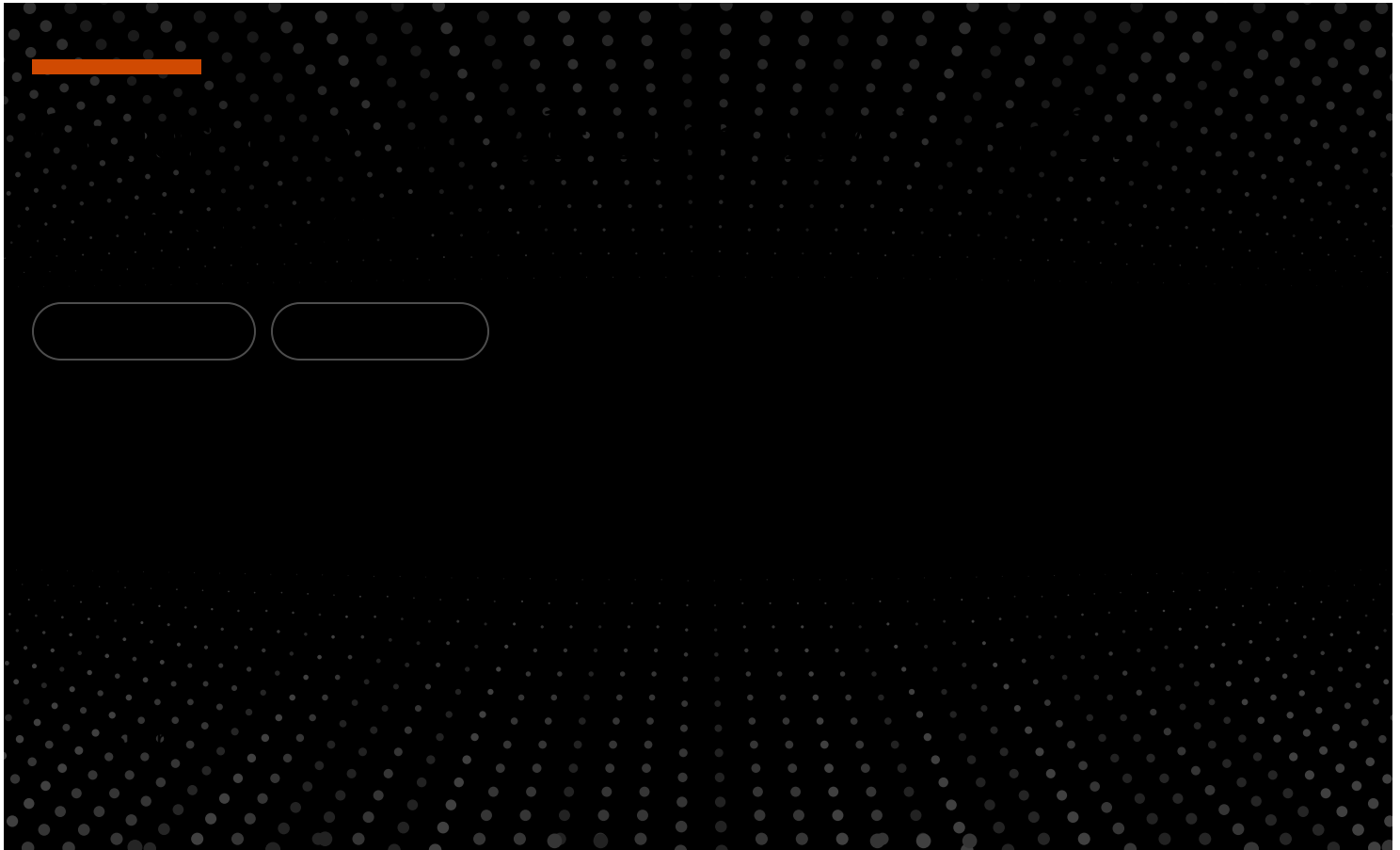


[Home](#)[Tech Effect](#)

Generative AI tools push new boundaries for responsible AI



## Summary

Expect regulatory proposals to address the specific risks and harms from widespread use of generative AI models.

Developers and platform owners can apply principles of responsible AI, which are now widely accepted.

Commercial users must employ generative AI ethically, including developing a policy around intellectual property related to products that are built from generative AI.

**4 minute read****February 2, 2023**

## The issue

The democratization of AI is upon us as new generative AI tools like OpenAI's ChatGPT and DALL-E put the power of AI into the hands of everyday users.

In the first five days following its release in November 2022, more than a million people logged into ChatGPT's platform to test its capabilities. Users are eager to experiment with how these generative AI tools can write code, craft essays, create art, design blueprints, sketch package designs, create virtual worlds and avatars in the metaverse, troubleshoot production errors and so on. They're also learning how to refine their prompts or instructions to the tool, in iterative cycles, to achieve better results.

While the positive use cases for generative AI are staggering, there's also potential for misuse and harm. As users began exploring the new tool, for instance, many discovered they could use it to generate malware, write phishing emails and spread propaganda. The same tools also were found to "hallucinate" facts and reinforce perspectives from misinformation campaigns.

As generative AI becomes increasingly popular and widespread, questions about who is responsible for mitigating the associated risks will become unavoidable.



## The regulators' take

More than 800 AI policy initiatives are pending in 69 countries, but the application to generative AI models is not settled. The Biden administration's blueprint for an AI Bill of Rights, for example, can help organizations and developers manage risks for consumer-facing AI, but it doesn't address unique

aspects of generative AI tools. The European Union announced its intention to regulate generative AI (included within general purpose AI systems), specifically around the data used to train them, under the **EU AI Act**.

Existing and proposed AI regulations today cover several **specific use cases** (e.g., data privacy, discrimination, surveillance), and specific decisions (e.g., **hiring**, lending, recommending on sites, public contracting), and most are enacted in response to the potentially harmful effects of AI on people and societies.

As users continue to discover new applications for these tools, new risks will likely emerge. The risks associated with generative AI are unprecedented and growing, but include:

Deepfakes

Phishing emails and social engineering

Malware and ransomware code

Plagiarism

Copyright infringement

Abusive or harmful content

Disinformation and propaganda

Even without regulations to guide them, some companies are voluntarily adopting **responsible AI** models, including OpenAI, the largest developer of generative AI today. For instance, when users reported that ChatGPT was generating discriminatory answers to prompts, the developers swiftly disabled prejudicial responses. OpenAI also employs teams dedicated to tagging harmful content to filter similar results from outputs.



## Your next move

Capitalizing on opportunities while managing risks will require action from three stakeholder groups.

### For developers and creators

Apply responsible AI principles and practices.

- Top platform owners **Microsoft**, **Google** and **OpenAI** have communicated their responsible AI strategies.
- Google recently updated its **search rater guidelines** from E-A-T (expertise, authority and trustworthiness) to E-E-A-T, adding experience to its parameters for evaluating whether the company's search ranking systems are providing "helpful, relevant information."
- OpenAI, Microsoft, Google, DeepMind and others all have dedicated research time and effort to responsible model development practices.

Prioritize nimble, iterative responses to events as they arise to complement robust internal practices. For example, observe OpenAI's approach to crowdsourcing use cases for ChatGPT to disable answers for potentially harmful prompts in addition to their internal efforts to block harmful outputs.

Educate users on how to use generative AI. These tools are useful, but they're far from perfect. Individual and commercial users should have access to all the information about how they work.

## For individual users

Apply a critical lens to generative AI. Now that generative AI-created content lives alongside human-created content, be even more circumspect about taking any content at face value. Using generative AI tools may require additional research to verify the information it generates. Use emerging tools to identify AI-generated content and add citations whenever possible.

Learn ways of maintaining integrity in the applications you work on. Generative models do not incorporate the concept of confidence in their responses and will attempt to make up an answer that better fits the prompt. In general, the default is to generate a response, true or not.

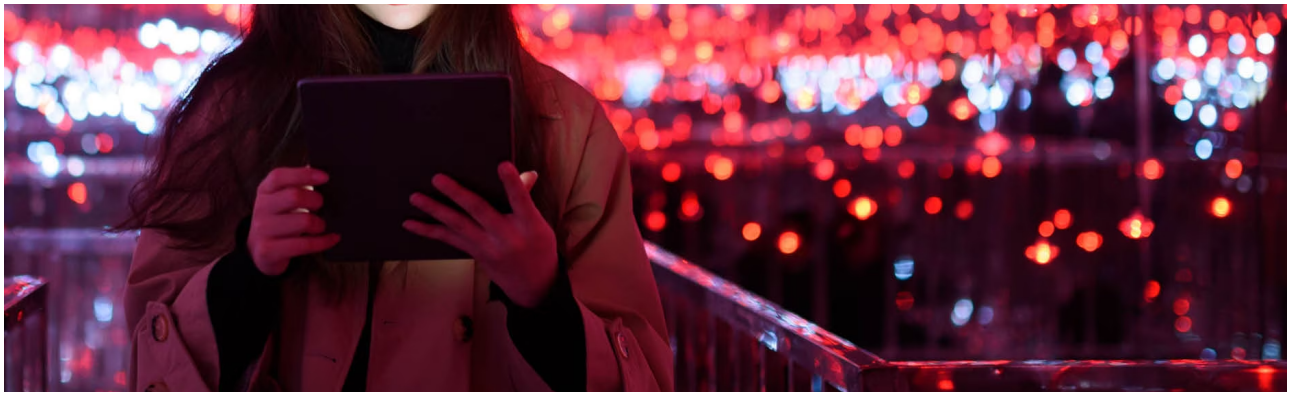
## For commercial users

Apply generative AI tools ethically but with the understanding that they come with potential risks. Lean on responsible AI frameworks to understand the potential shortcomings.

Develop a policy around intellectual property related to products that are built from generative AI. Some models “reproduce” content that was present within the body of work they draw from, creating a situation where a licensed piece of code is incorrectly “generated” as a novel solution. Copyright protections are an important legal concern, and ambiguities around AI generated content as well as the data used to train these systems may impact the ownership and rights to use creative works.

As with other AI, you may be using output from generative AI without even knowing it if your third-party vendors are doing so. Consider your third-party governance practices and how they can be augmented to provide additional guidance around creative applications of AI.

*This article first appeared in the January 2023 edition of **The Next Move**, PwC's insights on fast-moving policy and regulatory developments affecting technology.*



## Generative AI

Lead with trust to drive sustained outcomes and transform the future of your business

**Learn more**



## Cybersecurity, risk and regulatory

## A new equation to managing cyber, risk and regulation

**Learn more**



**Ilana Golbin**

Director and Responsible AI Lead, PwC US

**Joseph Voyles**

Principal, Advisory, Kentucky, PwC US

## Related content

### Generative AI

Manufacturers want to adopt generative AI. Where and how do they begin?

6 min. | Dec 11, 2023

### Generative AI

Do you have an “early days” generative AI strategy?

16 min. | Dec 7, 2023

### Cybersecurity

### Cybersecurity · Generative AI

## Tech Translated: Cyberphysical systems

4 min. | Dec 7, 2023

## For gen-AI-enabled threats, fight fire with fire

6 min. | Nov 28, 2023

---

[Trust solutions](#) [Consulting](#) [Tax services](#) [Newsroom](#) [Alumni](#)

[US offices](#) [Contact us](#)

---

© 2017 - 2024 PwC. All rights reserved. PwC refers to the PwC network and/or one or more of its member firms, each of which is a separate legal entity. Please see [www.pwc.com/structure](http://www.pwc.com/structure) for further details.

[Privacy](#) [Data Privacy Framework](#) [Cookie info](#) [Legal](#) [Terms and conditions](#)

[Site provider](#) [Site map](#) [Your Privacy Choices](#)