

[Home](#)[Tech Effect](#)[What is responsible AI and how can it help harness trusted generative AI?](#)

Summary

Generative AI can transform your business — if you apply responsible AI to help manage new risks and build trust.

Risks include cyber, privacy, legal, performance, bias and intellectual property risks.

To achieve responsible AI, every senior executive needs to understand their role.

7 minute read

April 24, 2023

For business leaders, there are plenty of reasons to be excited about generative AI, starting with its power and ease of use. But, as with any emerging technology, there are also potential new risks. Some of these risks may come from your company's use, others from malicious actors.

To manage both kinds of risks and harness generative AI's power to drive sustained outcomes and build trust, you'll need responsible AI. What exactly is responsible AI? It is a methodology designed to enable AI's trusted, ethical use. It has always been important, but it has become crucial in the dawning era of generative AI.



What generative AI can and can't do

With plain language commands, you can direct generative AI to create software code, data analysis, texts, videos, human-like voices, metaverse spaces and more. As the months pass, this power is expected to transform functions, business models and industries. Already, generative AI is being used by more people, not just data scientists, to deliver value at scale. It's boosting productivity, supporting human decision-making and reducing costs. Use cases that are scaling up today include:

Enhancing automation and personalization in customer service

Automating high-volume tasks such as processing claims or writing certain software code

Providing humans with supportive summaries and insightful analyses of business documents, meetings and customer feedback

Generative AI is far from perfect. It tries to predict the right response to a prompt or request, based on its algorithms and data sets. Its outputs can be highly useful, but they may resemble “first drafts.” You’ll likely need to verify these drafts’ output and analyze their quality, then modify them appropriately. At times, generative AI can also produce irrelevant, inaccurate, offensive or legally problematic outputs — in part because users may not give generative AI models the right prompts and in part because these models are by nature creative. New cyber risks are also emerging, including the ability for malicious actors to use these technologies to generate deep fakes and facilitate other cyber attacks at scale. Because more people than ever can use generative AI, some of these risks may also become more widespread: many generative AI users within your organization may be unfamiliar with how AI models work and how they should be used.



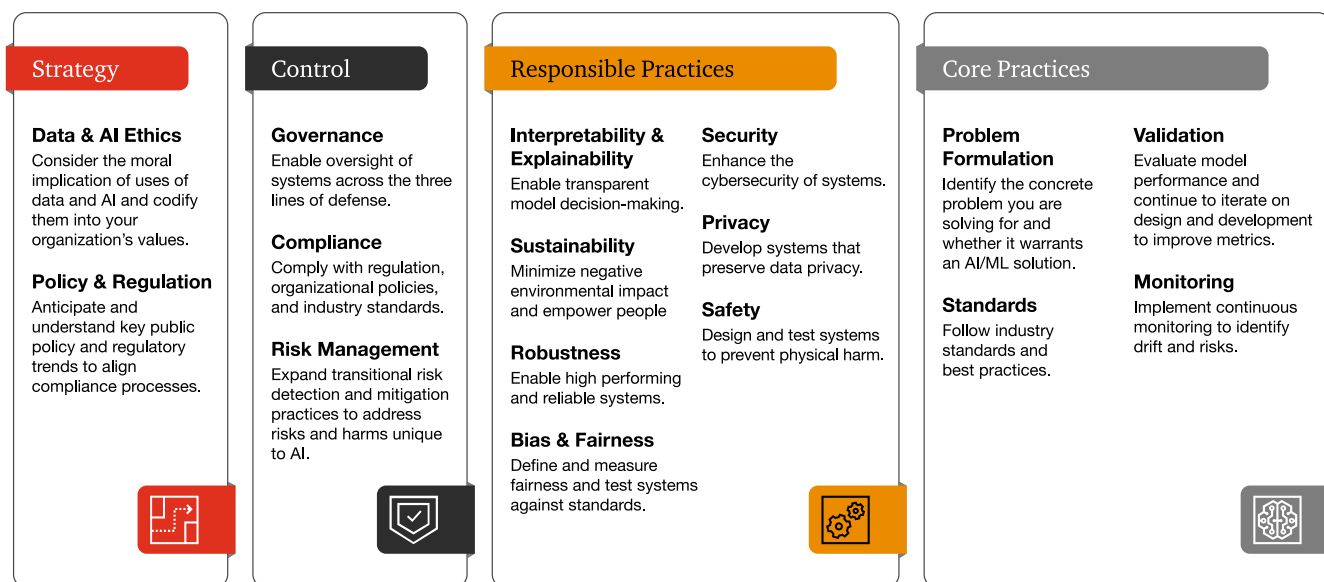
Generative AI’s new trust challenges and how responsible AI can help

Responsible AI can help to manage these risks and others too. It can grow trust in all the AI that you buy, build and use — including generative AI. When well deployed, it addresses both application-level risks, such as lapses in performance, security and control, and enterprise and national-level risks, such as compliance, potential hits to the balance sheet and brand, job displacement and misinformation. But to be effective, responsible AI should be a fundamental part of your AI strategy.

Our approach to responsible AI, PwC’s Responsible AI, delivers trust-by-design through the entire AI life cycle with frameworks, templates and code-based assets. It’s relevant to executives at every level: The CEO and board set the strategy, with special attention to public policy developments and to corporate purpose and

values. Chief risk and compliance officers are in charge of control, including governance, compliance and risk management. Chief information and information security officers take the lead on responsible practices, such as cybersecurity, privacy and performance. Data scientists and business domain specialists apply responsible core practices as they develop use cases, formulate problems and prompts and validate and monitor outputs.

PwC's Responsible AI Toolkit



As your business begins exploring generative AI use cases and how to apply responsible AI, here are some key risks to be mindful of:

Biased, offensive or misleading content. Like conventional AI, generative AI can display bias against people, largely due to bias in its data. But the risk here can be more intense, since it may also create — in your name — misinformation and abusive or offensive content.

New, darker black boxes. Generative AI usually runs on a “foundation model” that a specialized third party has built. Since you don’t own this model or have access to its inner workings, understanding why it produced a particular output

may be impossible. Some generative AI tools may not even disclose which third-party solution they are using.

Cyber threats. Just as generative AI can help you produce compelling content, it can help malicious actors to do the same. It could, for example, analyze your executives' email, social media posts and appearances on video to impersonate their writing style, manner of speech and facial expressions. Generative AI could then produce an email or video that appears to be from your company, spreading misinformation or urging stakeholders to share sensitive data.


Novel privacy breaches. Generative AI's ability to connect the data in its vast data sets (and to generate new data as needed) could break through your privacy controls. By identifying relationships among apparently disparate data points, it could identify stakeholders whom you've anonymized and piece together their sensitive information.

Hallucinations that threaten performance. Generative AI is good at coming up with convincing answers to almost any question you ask. But sometimes its answers are flat-out wrong yet presented authoritatively — what data scientists call “hallucinations.” Hallucinations occur in part because the models are often designed to generate content that seems reasonable, even if it's not always accurate.

Unsafe models, built on theft. With so much data underlying generative AI, you may not always know its source — or if you have permission to use it. Generative AI may, for example, reproduce copyrighted text, images or software code in the content that it produces in your name. That could be considered intellectual property theft and lead to fines, lawsuits and a hit to your brand.

Inadvertently sharing your intellectual property. Without care, you could find your proprietary data and insights helping your competitors generate content:

information you enter into a generative AI model could enter a database and become widely shared.



How to get started using generative AI responsibly

If you currently have a strong responsible AI program, your governance efforts have probably already flagged many of generative AI's new challenges. Still, there are key areas to pay close attention to and key steps to consider in applying responsible AI to a tech environment that's quickly evolving.

Set risk-based priorities. Some generative AI risks are more important to your stakeholders than others. Adjust or establish escalation frameworks so that governance, compliance, risk, internal audit and AI teams give the greatest attention to the greatest risks.

Revamp cyber, data and privacy protections. Update cybersecurity, data governance and privacy protocols to help mitigate the risks of malicious actors' generative AI inferring private data, unraveling identities or conducting cyber attacks.

Address opacity risk. With some generative AI systems, explainability is not an option. It's impossible to unravel "why" a certain system produced a certain output. Identify these systems, consider what practices can help support their fairness, accuracy and compliance, and tread carefully when oversight is impossible or impractical.

Equip stakeholders for responsible use and oversight. Teach employees who may need to use generative AI the basics of how it works, when and how to use it, and when and how to verify or modify outputs. Provide compliance and legal teams with skills and software to identify intellectual property violations and other related risks.

Monitor third parties. Know which of your vendors provide content or services that use generative AI, how they manage the related risks and what your possible exposure may be.

Watch the regulatory landscape. Policymakers around the world are issuing more and more guidance on AI development and usage. This guidance is still a patchwork, not a complete regulatory framework, but new rules are continually emerging — especially regarding AI's impact on privacy, AI bias and how AI should be governed.

Add automated oversight. With generative AI-created content ever more common, consider emerging software tools to identify AI-generated content, verify its output, assess it for bias or privacy violations and add citations (or warnings) as needed.

Design trust in from the start

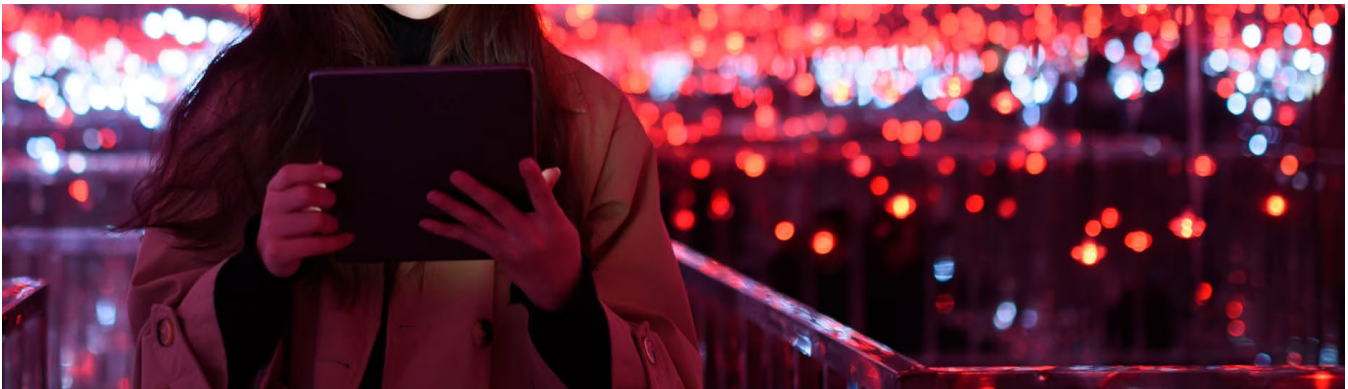
If there's a golden rule for responsible AI (and trusted technology in general), it's this: It's better to implement trust by design and ethics by design from the start rather than racing to close gaps after systems are up and running. That's why, whatever your company's maturity level with AI, including generative AI, it's wise to put responsible AI front and center as soon as you can — and keep it top of mind every step of the way.



PwC's Responsible AI

Helping you harness AI you can trust through frameworks, templates and code-based assets.

Learn more



Generative AI

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

Learn more

What can generative AI do for you?

Generative AI is here and already transforming business. Contact us to learn more about this rapidly evolving technology — and how you can begin putting it to work in a responsible way.

Get in touch



Ilana Golbin

Director and Responsible AI Lead, PwC US

Jennifer Kosar

Trust and Transparency Solutions Leader, PwC US

Rohan Sen

Principal, Data Risk and Responsible AI, PwC US

Related content

Generative AI

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

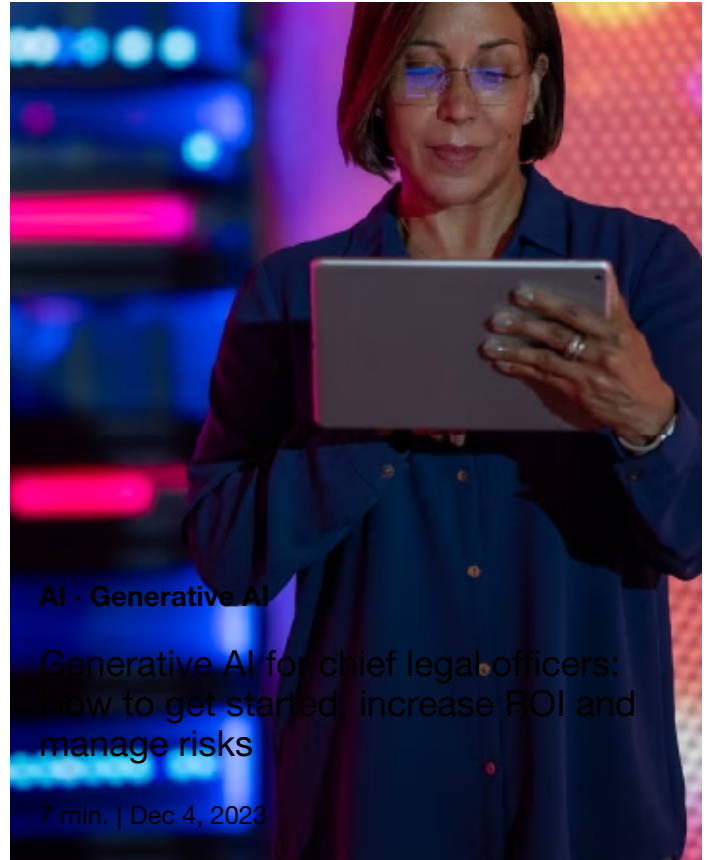
6 min. | Dec 11, 2023



Generative AI

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

16 min. | Dec 7, 2023



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 for further details.

[Privacy](#) [Data Privacy Framework](#) [Cookie info](#) [Legal](#) [Terms and conditions](#)
[Site provider](#) [Site map](#) [Your Privacy Choices](#)