Deloitte.

The CISO's Guide to Generative AI

Opportunities, outcomes, and the urgency of now

True or false? Generative AI can help:

- **Unlock new opportunity and value in an organisation's cybersecurity approach.**
- **Reduce costs and supercharge generation of reporting and intelligence products.**
- **Rapidly protect against sophisticated phishing attacks.**
- **Guide organisations in identifying critical information based on past actions.**
- **Make sense of regulatory and compliance guidance.**
- **Build a cybersecurity road map for now and the future.**

If your answers were all true, then you're thinking correctly about this powerful technology.

Read on for more on how Generative AI can help transform your organisation's cybersecurity approach.

Generative AI is here. What can it do for you?

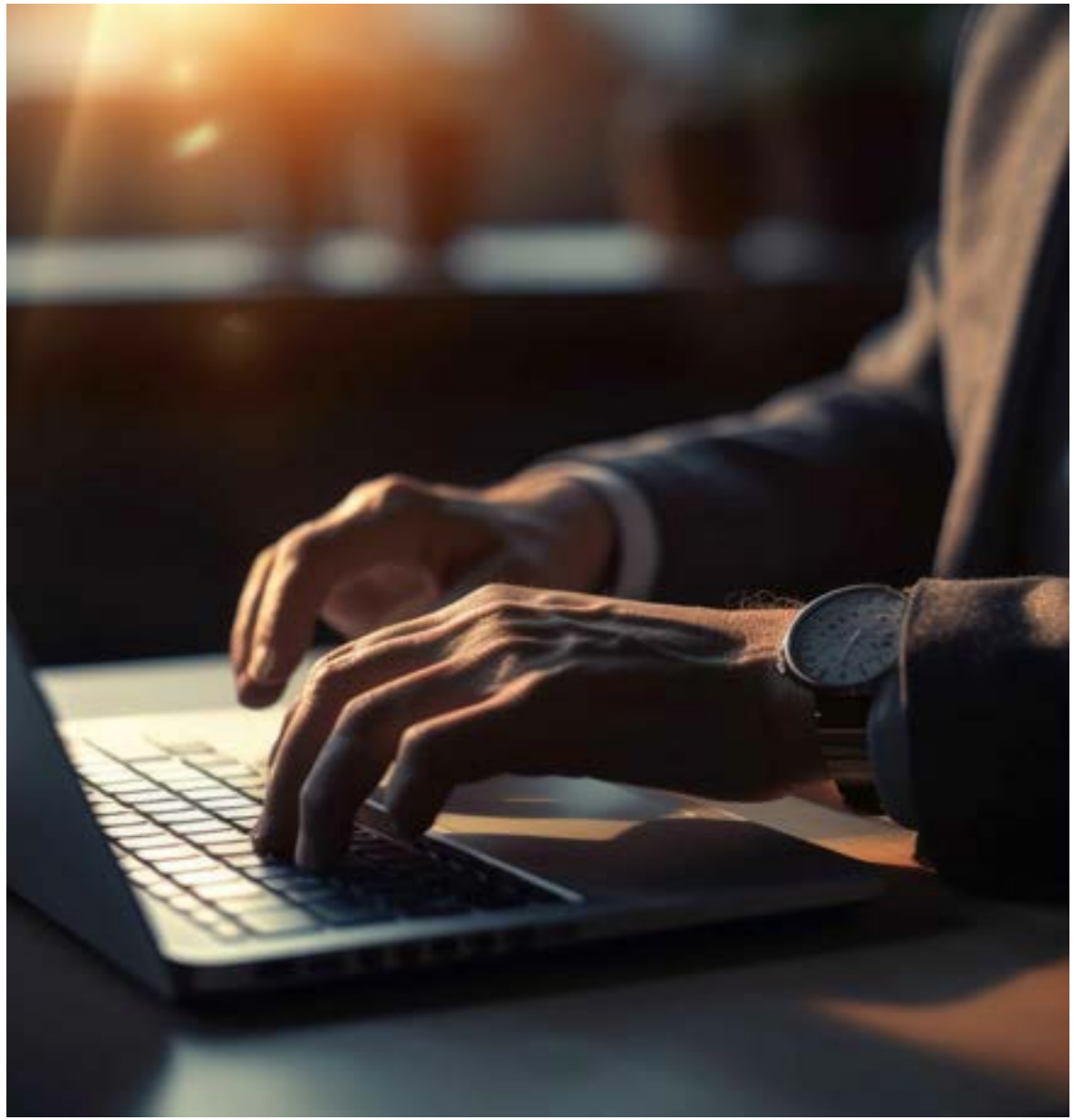
The Generative AI (Gen AI) buzz is everywhere. People are wondering what this new artificial intelligence (AI) can do for their organisations, their data, and their security. It's a complex question that defies an easy answer.

Gen AI is a subset of AI in which machines create new content in the form of text, code, voice, images, videos, and processes. The technology may truly revolutionise work and life. When it comes to cybersecurity, Gen AI holds promise for both organisations and governments that need to protect themselves, create tools to automate reporting and intelligence, reduce costs, grow more efficiently, sort through the varied and ever-changing regulatory atmosphere, and so much more.

Gen AI can also provide new tools for bad actors who want nothing more than to leverage these powerful technologies for negative outcomes and their own gain. Cyberattacks continue to increase in both volume and tactics: in fact, more than 90% of respondents to the Deloitte Global 2023 Future of Cyber survey reported at least one compromise.

But while cyber events have long since eclipsed the capabilities of a traditional human security operations center, AI is deeply impactful in enhancements to cyber infrastructure and detect-and-respond capabilities. Deep learning models are well suited to detecting attacks.

But cyber leaders may still wonder: While AI has increased our defense capabilities and postures, could Gen AI take us even further? How could it be used to limit blast radiuses of attacks, protect against data loss, and expand our threat response capabilities within budget and on time? In other words, can it help us get ahead—and stay ahead—of attackers?



GENERATIVE AI IS HERE. WHAT CAN IT DO FOR YOU?

Gen AI can do each of those things—and it holds so much promise for better cyber outcomes for organisations seeking out and defending against breaches. It's fast and reasoned and can process more knowledge than any one human can. It has the potential to reduce costs, supercharge security investigations, and speed up third-party risk assessments.

While more established AI capabilities (such as machine and deep learning) can identify patterns and make inferences, Gen AI can put it together while generating human-like responses and working at extraordinarily high speeds.
It can create a new type of threat intelligence that empowers security analysts with near real-time incident analysis to identify and help contain threats before they spread.

Cyber leaders are right to be concerned about how bad actors may use Gen AI—but they should be optimistic that, with the right approach and governance in place, Gen AI can help an organisation harden its cyber posture, overcome challenges in talent, and build new road maps for threat detection and response.

To unlock the potential of Gen AI, cyber leaders should first understand where it can help, the types of data it needs, and how to develop a plan of action that includes considerations for safety, resilience, and trustworthiness.

Two things to remember: This is an evolution of AI, not a net-new concept, and adoption plans and risk management constructs can be evolved accordingly. And like any true evolution, these are long-term transformation efforts. Adopting Gen AI for cybersecurity is a capability-building effort. Treat it that way.

This paper will explore how Gen AI can help and what the cyber considerations may be. It's important to remember that an organisation's success with using Gen AI to drive better outcomes rests on its ability to imagine a collaborative intelligence between humans and machines and to ask the right questions. Trusting that Gen AI can make a true impact in your organisation means first understanding its power and potential.

So let's get started.

Gen AI's immense value for cybersecurity

Gen AI is a force multiplier of value because it can do human-like work at hyper speeds that no human can match.

Machine learning has long been used to detect cyber vulnerabilities and perform threat monitoring at scale, but it takes a high degree of technical proficiency and investment to train an organisation's model to understand patterns and detect anomalies in the data. Rules-based AI, in other words, can find only known attacks and work in specific use cases.

But with Gen AI and large language models (LLMs), the game changes. Gen AI uses foundational neural network models that are powered by and trained on vast amounts of data, working across data silos and acting as a bridge between data sets. This can give analysts a more natural method for identifying, synthesising, and summarising insights.

Here's what we mean:

Predict: Analyse asset inventories, security logs, threat intelligence, etc., to help predict risk scores and recommend preventive measures.

Interpret: Summarise and process large volumes of textual data into coherent, actionable summaries; alert reception; and parsing. Generate logical analysis (inference, deduction, and/or explanation) given context or knowledge base.

Simulate: Extract information from a knowledge base to help generate responses to natural language questions; create test cases and sample scenarios.

Automate: Create incident response activities, including triaging alerts, correlating events, and guiding incident handlers with response playbooks.

Detect: Identify connections between alert data and threat intelligence reports to help determine the impact on infrastructure. Update specific responses that can guide security analysts in remediation and recovery activities.

Interact: Analyse governing documents, laws and regulations, data, and standards to quickly inform actions. Deliver personalised and targeted threat and crisis response trainings to employees based on roles, responsibilities, and job requirements.

Create: Generate content by converting it to a new format or style and for a variety of modalities based on a set of input data, examples, or specific themes or topics.

Looking for specifics? Gen AI can help transform cybersecurity activities like these.

Cyber risk management and compliance

Risk scoring and prioritisation
Analyse asset inventories, security logs, and threat intelligence to predict risk scores and recommend preventative measures

Third Party Risk Management
Analyse data in vendor submitted and external documentation to evaluate the security posture of third-party providers

Automated policy review & orchestration
Map current policies, standards and procedures against standard industry and regulatory frameworks to meet compliance requirements

Cybersecurity maturity assessments
Self-assess the organisation's cyber risk maturity; identify gaps in cyber strategy and generate relevant improvement recommendations

Threat detection and response

Actionable and precise threat intelligence
Generate summarised reports/ executive briefings for active threats from historic trends or publicly available data

Threat correlation and detection
Identify correlation between alert data and threat intelligence reports to determine impact on infrastructure

Security incident response
Automate incident response activities, including triaging alerts, correlating events, and guiding incident handlers with response playbooks

Enhanced recovery and remediation
Create specific responses that can guide security analysts in remediation and recovery activities

Gen AI-enabled phishing detection
Use Gen AI to detect threats and/or phishing attempts created by LLMs

Vulnerability management and security testing

Controls testing and automation
Create test cases/sample scenarios; expected outcomes; develop supporting documentation

Secure code generation
Develop application code and relevant supplementary test cases in line with the latest security considerations (backward integration of secure coding guidelines)

Enhanced vulnerability scanning
Correlate vulnerability data (scan data, external information and remediation plans) to prioritise action plans

Enhanced systems design/configuration
Augment system/security architecture design by drafting preliminary technical specification and/or recommending optimal configuration

Others

Role mining
Use Gen AI to recommend role assignments based on user attributes to ensure adaptive access control

Data classification and monitoring
Classify and monitor unstructured text-based data, which enables better protection against exfiltration

Training and awareness
Deliver personalised and targeted threat/crisis response trainings to employees based on roles, responsibilities, and job requirements

Note: This is not an exhaustive list. Feasibility of some of these use cases must be evaluated based on data availability and other constraints.

The power of pairing AI and Gen AI

An organisation using AI to detect and combat cyberthreats is already ahead of the game. Layering on Gen AI can add further complexity and power to its models.

While a traditional AI model can detect threats, adding Gen AI could allow it to summarise the incident, prepare documentation, and create a response action plan.

Gen AI can help an organisation move beyond rules-based analysis and expand into outputs of higher complexity and capabilities.

	Draft requirements	Indicators of Compromise (IOCs) signature generation	Co-pilot for incident response and SOC automations	Response process automation
Application of Generative AI	Simplify requirements-gathering phase by developing prototypes of complex applications. Provide more intuitive engagement between the analyst and the customer to better inform development.	Classify IoCs (e.g., information about a specific security breach that notifies security teams if an attack has taken place) using distinct signature generation.	Detect hidden patterns, harden defenses, and respond to incidents faster with triage signals and predictive guidance. Quickly synthesise data from multiple sources to provide actionable insights.	Automate cyber defense strategies, industry notifications, future mitigation strategies, etc., as part of the response process.
Benefits	Reduces the risks of miscommunication (i.e., the analyst and customer are able to align on the prototype before proceeding to the build phase).	Improves visibility of cyber attacks and streamlines the security team's response with expedited identification and triage.	Introduces robust and reliable approach to incident response, threat hunting, and security reporting	Improves organisational compliance with incident response plans and contingency plans through automation. In doing so, improves efficacy and streamlines execution.
force skills	• Customer engagement (e.g., review cycles) • Storyboarding	• Information gathering • Mission expertise/security clearances	• SOC • Threat detection and response	• Cybersecurity SOC • Threat response

The cyberthreat considerations for Gen AI

To understand Gen AI's power, an organisation should be fully aware of the considerations inherent to the technologies.

As we've said, Gen AI opens new opportunities for organisations to prepare for and defend against cyberattacks. But as with any new technology, Gen AI comes with risks and the potential to amplify existing ones as well.

Earlier AI systems were traceable, and it was possible to understand certain outputs via its data. But Gen AI is a different game with multiple parameters that can make it more challenging to trace output. Gen AI is also trained on much larger data sets than traditional AI, which can make it more difficult to know where and how the data may have been altered or where quality concerns may exist. Constantly evolving risk profiles demand a new perspective.

Growing concerns and global action

In the fall of 2023, the Biden administration announced an executive order on the safe and trustworthy use of AI that will likely create downstream effects for new regulations and standards, further complicating the regulatory atmosphere.

Meanwhile, the European Union (EU) is moving toward stringent rules around AI, even moving to ban its use in some cases. As the use of Gen AI becomes more prevalent, we expect governments to take more action to mitigate potential risks.

There's a lot to consider. We've broken out some current cyber risks for Gen AI:

Data breach

Sharing sensitive data with external Gen AI vendors for model training or through prompts may lead to leakage of confidential and/or personal information. Adversarial attacks can also be used to deceive the ML model by changing input data.

Unsecured integration

Improper integration of Gen AI tools with other organisational systems may lead to potential vulnerabilities (e.g., unsecured data channels) and back doors.

Reputational risk

Bad actors can leverage Gen AI tools to widely and rapidly spread misinformation and deepfakes, which can adversely influence public opinion, trust, and/or security.

Regulatory risk

Organisations using Gen AI may need to meet new compliance requirements as growing concerns influence new laws, regulations, and guidelines, such as National Institute of Standards and Technology's (NIST) proposed AI Risk Management Framework[1] and new EU regulations for General Purpose AI Systems[2]. (Read more)

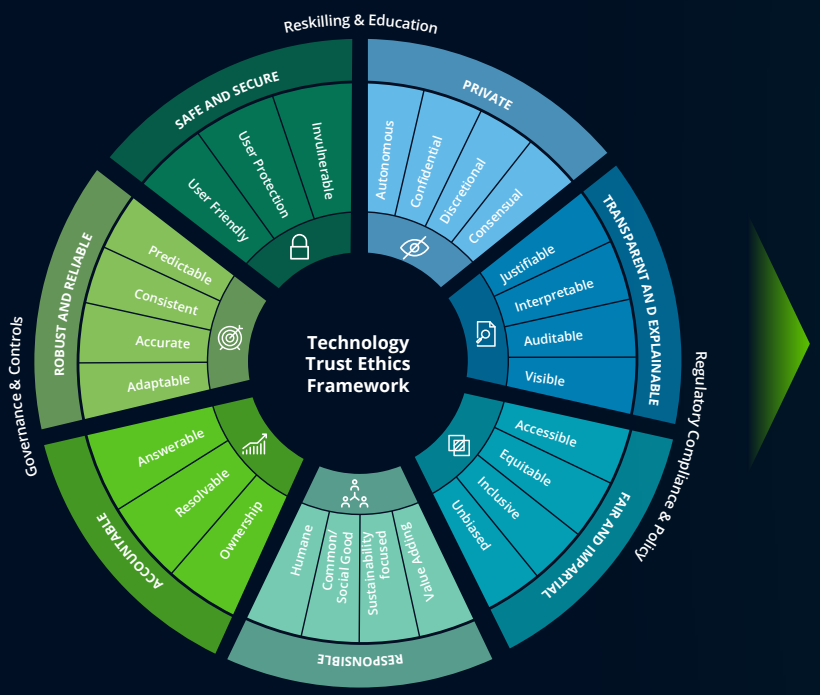
Don't just take our word for it.

The Open Worldwide Application Security Project (OWASP) has published its Top 10 risks for large language model applications, including trained data poisoning and supply chain vulnerabilities[3]

A framework for risks and limitations associated with Gen AI

While emerging tech has inherent risks, Deloitte's Technology Trust Ethics (TTE) framework can be leveraged to build, deploy, and commercialise AI applications

Deloitte's TTE framework



Deloitte's TTE framework may be leveraged for foundational and Gen AI specific capabilities

Foundational capabilities

AI strategy

- Define and implement an overarching AI strategy and management framework
- Regulatory compliance—review evolving regulatory landscape and prepare for new requirements

AI risk management

- Design and implementation of AI controls—design and implement controls to address AI-specific risks (e.g., bias) based on regulations and industry standards
- Monitor controls—assess controls effectiveness and initiate remediation

AI technology/cybersecurity

- Code assessment and model validation—provide independent testing of AI
- Threat monitoring and detection—monitor for specific technology threats (malicious and environmental) that are targeted at AI models and underlying technology

Gen AI-specific capabilities

Management of hallucinations and misinformation

- Identify and manage misinformation through Gen AI by implementing appropriate governance mechanisms (e.g., workforce upskilling, structured oversight, ubiquitous documentation)
- Regulatory compliance—review evolving regulatory landscape and prepare for new requirements

Attribution management

- Assess and validate attributions to the source information while ensuring that the model does not tread across lines of plagiarism and copyright violations

Accountability and explainability for Gen AI

- Focus on providing accessible, non-technical explanations of Gen AI, its limits, capabilities, and associated risks
- Derive viable methods of accountability, trust, and ethics when using Gen AI and underlying technology

How to prepare

With forethought and deliberation, cyber leaders can ready their organisations for the capabilities and risks of Gen AI.

Regardless of an organisation's particular needs, defining key outcomes and instituting guardrails can help leaders improve risk preparedness, promote resilience, and unlock new business opportunities around Gen AI.

Leaders should recognise that Gen AI requires new approaches to technology, training, and processes—but that said, this is an evolution of existing risks. Gen AI may not require net-new road maps and trainings. An organisation's risk management and cyber constructs may still work.

The key is to evolve those constructs to answer the nuanced risks and threats that may be targeted toward Gen AI or AI systems. An organisation's specific risks may depend on what adoption model it chooses, such as software-as-a-service or private LLMs.

When choosing adoption strategies, an organisation should recognise the power and necessity of end-to-end transformation rather than automating one or two activities.

Example checklist

- Update policies and controls for new types of bias, legal, regulatory, privacy, intellectual property, and data risks of Gen AI.
- Identify new compliance requirements and impacts on compliance activities with existing laws and regulations.
- Closely evaluate use cases for Gen AI for the organisation to help ensure impactful outcomes and overcome any resistance to adoption.
- Implement appropriate contractual obligations for Gen AI vendors around security and usage of any information shared, and monitor the data sharing channels used by them.
- Implement privacy and data protection standards and controls when developing and training models for Gen AI tools.
- Enhance existing code review processes to help test code created by Gen AI for back doors and vulnerabilities.
- Implement access controls and monitor use of Gen AI tools to help limit risks from inadvertent or inappropriate use.
- Establish secure channels and mechanisms to transfer data between enterprise and cloud-hosted Gen AI tools.
- Review third-party controls and establish contractual obligations to help protect sensitive data shared with Gen AI vendors.
- Monitor for novel attacks (e.g., prompt injection) and help ensure appropriate usage of Gen AI tools to prevent vulnerabilities.
- Define boundaries of where and when Gen AI technologies can be used within the organisation.
- Integrate secure-by-design principles during integration of Gen AI applications into enterprise architecture.
- Help protect the organisation's brand by monitoring for misinformation, and define communication strategies to counteract and decrease impact of disinformation campaigns.
- Take action immediately on the risks from adversarial and malicious Gen AI usage.

HOW TO PREPARE

Above all, remember: A road map for Gen AI adoption should include close, constant collaboration for risk stakeholders, including cyber leaders, chief resource officers, an organisation's legal team, and more, to help understand and anticipate the risks. (And don't forget to include testing and monitoring.)

Adoption of Gen AI by organisations will depend on six factors

1. **Cost and efficiency:** Ability to assess whether benefits of using Gen AI-based systems outweigh the associated expenses, as handling and storing large datasets can result in increased expenses related to infrastructure and computational resources.
2. **Knowledge and process-based work:** High degree of knowledge and process-based work vs. only field and physical work.
3. **High cloud adoption:** Medium-to-high level of cloud adoption, given infrastructure requirements.
4. **Low regulatory and privacy burden:** Functions or industries with high regulatory scrutiny, data privacy concerns, or ethics bias.
5. **Specialised talent:** Strong talent with technical knowledge and new capabilities, and ability to help transform workforce to adapt quickly.
6. **Intellectual property and licensing and usage agreements:** Ability to assess licensing/usage agreements and restrictions, establish and monitor related compliance requirements, and negotiate customised agreements with relevant vendors.

Cyberattacks won't stop. The good news is, Gen AI progress won't either.

Gen AI could accelerate both cyberattacks and threat response capabilities. Organisations need to recognise both sides of that equation.

The question is, how can cyber leaders steer their teams and organisations through the disruption while harnessing the capabilities of what is, to date, the most powerful artificial intelligence ever created? Many organisations are so busy fighting today's battle that it's hard to conceive of creating a new Gen AI ecosystem that may require development, operations, new talent, and evolved processes.

For any cyber leader, it's important to start the journey toward Gen AI with questions specific to the organisation. Gen AI is an unprecedented opportunity for a new kind of collaborative intelligence, one that can provide increased security and next-level collaboration. So where does a leader start?

With one question: "What if?" From there, it's all a new frontier.

With our deep bench of cyber experience, alliance relationships, and pragmatic perspective on the future, Deloitte can help organisations address their most pressing cybersecurity challenges—now, and for whatever is around the bend. Reach out to learn more.

Endnotes

1. AI Risk Management Framework | NIST
2. https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai
3. OWASP Top 10 for LLM Applications Version 1.1, October 16, 2023

Get started

Authors

David Caswell, Sabthagiri Saravanan Chandramohan, Deborshi Dutt, Chris Knackstedt, Vikram Reddy Kunchala, David Mapgaonkar, Mike Morris, Abdul Rahman, Kate Fusillo Schmidt, Niels van de Vorle

Contributors

Sanmitra Bhattacharya, Edward Bowen, Ben Bressler, Suzanne Denton, Eric Dull, Lena La, Sajin Mathew, Nirmala Pudota, Stephanie Salih, Colin Soutar

Contacts

Reto Haeni
Global AI Security Leader
Deloitte Switzerland
rhaeni@deloitte.ch
+41 58 279 7202

Florian Widmer
AI Risk Leader
Deloitte Switzerland
fwwidmer@deloitte.ch
+41 58 279 6910

Marc Beierschoder
Generative AI Lead
Deloitte Switzerland
msbeierschoder@deloitte.ch
+41 58 279 6778

Deloitte.