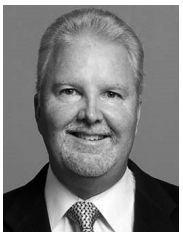
Navigating the GenAI frontier: The role of the CISO in GenAI governance

Received (in revised form): 29th October, 2024



James Christiansen
Vice President, Chief Security Officer, Netskope, USA

James Christiansen is the Vice President (VP), Chief Security Officer (CSO) at Netskope, where he drives the company's global strategy for cloud security. He works closely with global security leaders to develop and implement strategies for success. Before joining Netskope, James held a senior leadership role as VP Chief Information Security Officer (CISO) at Teradata, where he led global information and physical security teams. His extensive background also includes serving as CISO for Visa International, General Motors and Experian Americas. In addition to these roles, James has served as VP of Information Risk Management at Optiv and co-founded Evantix, where he was Chief Executive Officer (CEO). He is a well-regarded thought leader and frequent speaker at industry events such as the RSA Conference, Secure World, American Bar Association, American Banker, the RSA Conference, Secure World, BankInfoSecurity, Information Systems Security Association (ISSA), Information Systems Audit and Control Association (ISACA), Healthcare Information and Management Systems Society (HIMSS) and the Business Roundtable. His expertise has also been featured in prominent publications such as *The Wall Street Journal*, *USA Today*, *Cyber Security: A Peer Reviewed Journal* and *The New York Times*. James is a patent holder and has received multiple innovation awards in cyber security, governance, risk management and cloud computing. He is the author of the Internet Survival Series and numerous industry papers and has contributed to books such as *CISO Essentials* and *CISO Compass*. He holds an MBA with a focus on international management and a bachelor's degree in business management from Westminster College.

Tel: +1 714 615 5822; E-mail: james.christiansen@netskope.com

Abstract Artificial intelligence (AI) and large language models (LLMs) have rapidly evolved, becoming integral parts of modern enterprises. The exponential growth of generative AI (GenAI), spearheaded by organisations such as OpenAI, has introduced both remarkable opportunities and significant challenges. This paper explores the critical role of the chief information security officer (CISO) in navigating GenAI governance, highlighting the importance of responsible GenAI framework development, ethical use and regulatory compliance. It also lays some groundwork on the history of GenAI and the types of GenAI that the CISO may encounter.

KEYWORDS: GenAI governance, GenAI risk management, CISO's role in GenAI, GenAI data security, practices for GenAI governance, responsible GenAI model

DOI: 10.69554/HOJP4542

THE EXPANDING GENAI LANDSCAPE

ChatGPT reached 100 million monthly active users within just two months of its release, making it the fastest-growing consumer application in history.[1] Since late 2022, generative AI (GenAI) and large language model (LLM) technologies have become widely accessible. This rapid and widespread adoption necessitates the implementation of robust governance

frameworks to manage the extended attack surfaces and address the potential risks associated with GenAI use in enterprises. Key considerations include understanding the evolving GenAI landscape, addressing the ethical implications and ensuring compliance with regulatory requirements.

The rapid integration of GenAI into various sectors underscores the need for comprehensive governance structures. These structures are vital for maintaining the balance between leveraging GenAI's benefits and mitigating its risks. The landscape is constantly evolving, with new GenAI applications and models being introduced regularly, which requires ongoing vigilance and adaptability from organisations.

UNDERSTANDING DIFFERENT TYPES OF GenAI

Artificial intelligence (AI) is a broad field that encompasses various technologies designed to perform tasks that typically require human intelligence. These tasks include learning, reasoning, problem solving, perception and language understanding. AI can be classified into different types, each with specific capabilities and applications including artificial narrow intelligence (ANI), artificial general intelligence (AGI) and artificial superintelligence (ASI). Figure 1 illustrates the relationships of the AI categories.

ANI

Also known as 'weak AI', ANI is designed and trained for a specific task. Virtual assistants like Siri and Alexa are examples of ANI. These systems can perform predefined functions such as voice recognition, internet searches and scheduling but lack general intelligence.

AGI

Also known as 'strong AI', AGI refers to a machine's ability to understand, learn and apply knowledge in a way that is indistinguishable from human intelligence. AGI can perform any intellectual task that a human can. This level of AI remains largely theoretical and has not yet been achieved.

ASI

ASI surpasses human intelligence and capability. It would not only perform all intellectual tasks that a human can but would do so at a far superior level. ASI is a concept primarily explored in speculative fiction and futurism.

MACHINE LEARNING

Machine learning (ML) is a subset of AI focused on building systems that learn from data to improve their performance over time without being explicitly programmed. ML algorithms identify patterns within data and make decisions or predictions based on new data inputs. ML has been in used for many years and there are several types, including supervised learning, unsupervised learning and reinforcement learning.

Supervised learning

In supervised learning, the model is trained on a labelled dataset, which means that each training example is paired with an output label. An example is spam detection in e-mail, where the algorithm learns from a dataset of e-mails labelled as 'spam' or 'not spam'.

Unsupervised learning

Here, the model is given data without explicit instructions on what to do with it. The goal is to find hidden patterns or intrinsic structures in the input data. Clustering algorithms, such as K-means clustering used in customer segmentation, are examples of unsupervised learning.

Semi-supervised learning

A hybrid of supervised and unsupervised learning, combining a small amount of labelled data with a large amount of unlabelled data to improve model accuracy while reducing the need for extensive labelling. It is particularly useful in tasks like image classification and text analysis, where labelling is time-consuming but unlabelled data is abundant.

Reinforcement learning

This type of learning involves training an agent to make sequences of decisions by rewarding it for good decisions and penalising it for bad ones. It is commonly used in robotics, gaming and navigation systems.

GenAI

GenAI refers to algorithms that can create new content, such as text, images and music, by learning from existing data. These models generate data similar to the data they were trained on, making them powerful tools for creativity and innovation. Key examples include text generation, image generation and audio generation.

Text generation

Models like GPT-3 can generate coherent and contextually relevant text based on a given prompt. They are used in applications ranging from chatbots to content creation.

Image generation

Tools like generative adversarial networks (GANs) can create realistic images from textual descriptions or combine features from different images to produce new ones. These are used in design, art and entertainment.

Audio generation

GenAI models can generate music and other audio content. Applications include composing new music, creating sound effects and even mimicking human speech for virtual assistants.

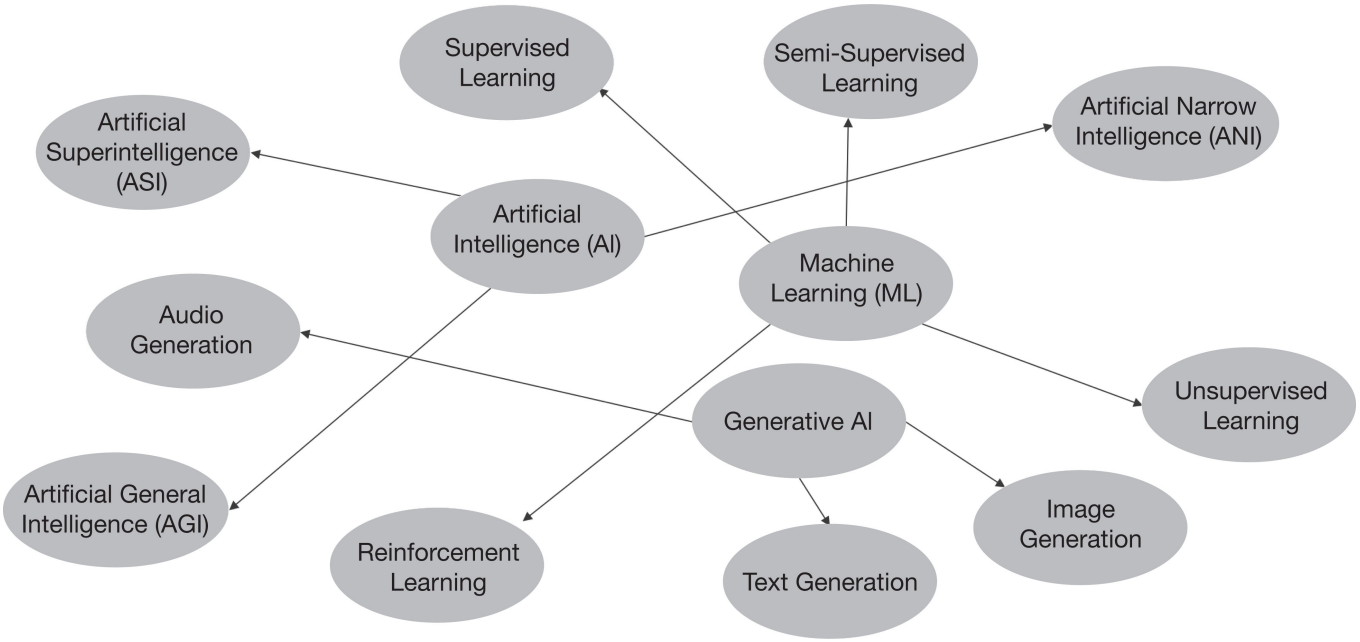


Figure 1: Relationships of the AI categories

THE ROLE OF THE CISO IN GenAI GOVERNANCE

The role of the chief information security officer (CISO) in GenAI governance is multifaceted and crucial for ensuring the secure and ethical use of GenAI technologies within an organisation. As AI governance is still a developing domain, however, there is ongoing debate about which executive function should hold ultimate accountability for GenAI oversight. The CISO's responsibilities span across establishing governance frameworks, managing risks and fostering ethical GenAI practices.

Traditional responsibilities of the CISO and new challenges

Traditionally, the CISO's responsibilities have centred on information security, data protection and mitigating risks related to cybersecurity. The rise of GenAI, however, has introduced complex ethical, legal and operational risks that fall outside of conventional security concerns. Specific responsibilities such as forming consensus, approving business cases, establishing training programmes, ensuring traceability and monitoring for ethics and bias are not typically within the CISO's historical purview.

This leads to a natural question: *Why should the CISO assume these responsibilities in the context of AI?*

Alternative executive roles for AI governance

Alternative executive roles for AI governance include:

- *IT governance, risk and compliance (IT GRC)*: In some organisations, the IT GRC function may seem better suited for overseeing AI governance, particularly when it comes to compliance, risk management and policy development. IT GRC professionals typically ensure that IT systems comply with regulatory and corporate policies, and they might already have frameworks in place to manage the risks introduced by AI. While IT GRCs may excel in governance, however, they may lack the technical expertise required to assess and mitigate GenAI-specific security threats, which involve both operational and cyber security risks.
- *Chief information officer (CIO)*: The CIO traditionally oversees the organisation's IT infrastructure and could argue for accountability over GenAI governance due to their broad purview over technology strategy. The CIO often balances innovation with risk, making them a potential candidate for GenAI oversight. The CIO's focus, however, is generally more on operational efficiency and technological implementation than on the intricate security and ethical concerns associated with AI, which may weaken their position as the sole authority for AI governance.
- *Chief technology officer (CTO)*: The CTO is frequently responsible for the technical direction and innovation within a company, often spearheading AI initiatives. While the CTO may excel in the development and deployment of AI technologies, however, their focus is typically on product and service innovation, leaving significant gaps in understanding and enforcing the privacy, security and compliance requirements that come with GenAI.
- *Chief legal officer (CLO) or chief risk officer (CRO)*: Both the CLO and CRO could make strong arguments for taking the lead on AI governance. The CLO focuses on regulatory compliance, intellectual property and legal risks, which are all crucial in AI oversight. The CRO traditionally handles risk at an organisational level, including operational and financial risks, which could encompass AI-related risks. But while these roles offer valuable insights into the legal and risk implications of AI, neither typically have the technical knowledge needed to manage the cyber security, data privacy and

technical nuances of AI deployment, which are essential aspects of AI governance.
- *Business strategy or chief data officer (CDO):* In some organisations, the CDO or business strategy leaders may also argue for their role in overseeing AI initiatives, as they are responsible for data strategy, insights and business use cases. While these executives play a vital role in utilising AI for innovation and business value, they may not have the security, compliance and governance expertise required to manage the risks associated with GenAI.

For effective GenAI governance, several organisational departments can serve as potential candidates to lead and oversee the governance programme. The key factor is that the chosen leader must be a high-level executive with extensive organisational insight and the ability to influence across departments. This leader must be skilled in cross-functional collaboration, leveraging insights from teams such as IT, legal, security and data privacy, to ensure comprehensive GenAI governance.

Ultimately, the executive in charge must be an experienced collaborator, capable of drawing on the expertise of multiple teams to set policies, assess risks and align GenAI initiatives with the organisation's strategic goals.

Is the CISO best placed to assume AI responsibilities?

Despite these alternatives, the CISO remains well suited for overseeing GenAI governance, particularly when considering the security risks and data privacy implications that come with AI systems. The core argument for the CISO's leadership is their deep understanding of data security, compliance and risk mitigation. As GenAI models rely heavily on sensitive corporate data, the CISO is already equipped to manage the confidentiality, integrity and availability of such information. Here are specific reasons why the CISO should lead GenAI governance:

- *Cyber security and data protection expertise*: The CISO already handles sensitive data protection and mitigating risks associated with data breaches. Given that GenAI models often rely on large amounts of sensitive corporate data, the CISO is best placed to prevent potential data exposure when using public models or ensuring proper data governance when training private models.
- *Risk management across IT and AI*: While other executives such as the CRO focus on broad organisational risks, the CISO specialises in IT-specific risks, including those associated with new technologies like AI. GenAI introduces novel cyber security risks, such as the potential for model manipulation, adversarial attacks and data leakage through AI prompts — all of which the CISO is already equipped to address.
- *Ethical and bias considerations*: Although ethics and bias might seem outside the traditional scope of the CISO, these concerns are increasingly becoming part of the security mandate, particularly when dealing with AI systems. AI bias can lead to reputational and operational damage, and the CISO's role in safeguarding data integrity positions them well to oversee bias mitigation processes, ensuring that AI models align with organisational values and compliance standards.
- *Interdepartmental collaboration*: In many organisations, the CISO already works closely with the business units and other departments, such as legal, compliance and IT. This collaborative approach is essential for AI governance, where cross-functional teams must come together to manage the ethical, legal and security aspects of AI. As a central figure in risk mitigation and governance, the CISO can lead these teams effectively, building consensus among stakeholders.
- *AI's expanding security concerns*: AI's capabilities introduce significant threats that CISOs are familiar with, such as

adversarial attacks and data poisoning. These security concerns align with the CISO's expertise and expanding their role to include AI governance can ensure that AI is integrated into an organisation's broader security strategy, providing a holistic approach to safeguarding AI use.

Expanding the CISO's role in AI governance

Although the responsibilities of consensus building, ethics and traceability may not traditionally fall within the CISO's remit, these functions are becoming inseparable from the role of a modern CISO in a GenAI-driven world. GenAI requires not only technical oversight but also ethical, legal and security governance — areas where the CISO is uniquely positioned to lead. While collaboration with other roles such as the CIO, CTO and CLO remains essential, the CISO is best equipped to own the core accountability for GenAI's secure and compliant deployment.

Regardless of who takes on the leadership role for GenAI governance, it is imperative that the chosen leader is equipped to collaborate effectively with other key stakeholders across the organisation. Whether the role is assumed by the CISO, CIO, CTO, CRO or another executive, the leader must leverage the specific expertise of each stakeholder to establish and maintain robust procedures for managing GenAI risk. This includes working with legal to ensure compliance with data privacy laws, collaborating with business strategy to align AI use with corporate goals, and engaging with data and IT teams to monitor AI performance and mitigate cyber security threats.

Ultimately, success in GenAI governance depends on a collective approach that taps into the strengths of multiple functions, ensuring that GenAI technologies not only mitigate risks but also enable the organisation to harness AI effectively to drive innovation and achieve business objectives.

ESTABLISHING GOVERNANCE FRAMEWORKS

A key responsibility of the CISO is to lead or assist in forming a cross-functional team dedicated to GenAI governance. This team collaborates to develop and understand business use cases for GenAI, ensuring that GenAI applications align with organisational goals and ethical standards. The CISO plays a pivotal role in publishing policies and training employees on the principles and practices related to the use of GenAI.

Risk assessment and mitigation

The CISO must discover and assess risks associated with GenAI applications. This involves real-time user coaching to remind employees of policies and responsibilities concerning GenAI. Additionally, the CISO employs data loss prevention (DLP) tools to enforce policies on sensitive data and GenAI usage, thereby mitigating the risk of data breaches and ensuring compliance with privacy regulations. More to come later in this paper.

Consensus building and policy development

To manage GenAI governance effectively, the CISO must maintain an unbiased perspective to facilitate consensus among stakeholders. Each enterprise will have unique business cases for using GenAI, and the CISO's role includes leading the consensus on business use cases, principles and policies. Tracking the usage of LLMs and ensuring traceability is also within the CISO's purview, enabling the organisation to monitor who is using GenAI and how it is being applied.

Data governance and security frameworks

Using conventional controls for data governance where feasible, the CISO ensures that GenAI systems operate within established security and privacy frameworks. This comprehensive approach supports the

secure deployment of GenAI and promotes its responsible and ethical use across the enterprise.

Core security capabilities for enablement

Effective GenAI governance necessitates a comprehensive approach to security that enables safe and responsible use of GenAI technologies. The core security capabilities for enablement include several critical components:

- *Application discovery and categorisation*: Understand which applications use GenAI and how they are utilised. Categorise these applications and evaluate their trust scores to determine associated risks.
- *Data discovery and classification*: To effectively assess risks and meet regulatory requirements, it is essential to have a comprehensive and accurate inventory of data intended for training LLMs. This process begins with thorough data discovery, identifying all relevant data sources. Next, data classification categorises information based on sensitivity, privacy level and regulatory implications, which is crucial for compliance. Proper classification enables organisations to align with data protection regulations and implement necessary safeguards, ensuring that only appropriate data types are used in GenAI training while minimising exposure to risk.
- *Instance-awareness and activity-level controls*: Implement controls aware of specific GenAI application instances and monitor their activity levels to identify anomalous behaviour or potential misuse.
- *AI/ML-powered data loss prevention*: Utilise GenAI and ML-powered DLP tools to enhance data security by filtering models and recording all transactions involving sensitive data. These tools help enforce data usage policies and prevent unauthorised access or leaks. The tool must be able to authorise specific use, and users of GenAI. The capability to control the upload and download of sensitive data is an absolute requirement.
- *Analytics for monitoring and reporting*: Develop robust analytics capabilities for continuous monitoring and reporting on GenAI applications, including tracking usage patterns, policy compliance and the overall security posture of GenAI systems.
- *Identity and access management (IAM)*: Ensure that only authorised individuals access GenAI systems. Implement strong IAM controls to prevent unauthorised access and effectively manage user roles and permissions.
- *Threat detection and response*: Deploy GenAI-specific threat detection systems to identify and mitigate potential security threats in real time, recognising unusual patterns that may indicate security breaches or malicious activities.
- *Regular security audits and compliance checks*: Conduct regular security audits and compliance checks to ensure GenAI systems adhere to established security policies and regulatory requirements. This practice helps identify and address vulnerabilities before exploitation.
- *Incident response planning*: Develop and maintain a comprehensive incident response plan outlining procedures for responding to security incidents involving GenAI systems, including containment, investigation and remediation steps.
- *Ethical and bias audits*: Regularly audit GenAI systems for ethical compliance and bias to maintain fairness and transparency, ensuring GenAI models do not perpetuate biases and operate within ethical guidelines.
- *Training and awareness programmes*: Establishing ongoing training and awareness initiatives is vital for equipping employees and stakeholders with the knowledge they need to engage responsibly with GenAI technologies. These programmes should emphasise GenAI security best practices, outline

potential risks and clarify individual roles and responsibilities in maintaining a secure GenAI environment.

To enhance these efforts, integrating real-time alerts can significantly improve awareness. These alerts, embedded within workflows, can notify users immediately when they may be exposing sensitive data to LLMs or other GenAI models. This proactive approach not only reinforces training but also fosters a culture of security awareness, enabling users to respond promptly to potential data exposure risks.

By integrating these core security capabilities, organisations can build a secure foundation for GenAI enablement, ensuring that GenAI technologies are used responsibly and effectively while mitigating associated risks.

THIRD-PARTY RISK GOVERNANCE FOR LLMs

The rapid proliferation of LLMs embedded within Software-as-a-Service (SaaS) applications has significantly increased the complexity of third-party risk governance. With hundreds of LLMs integrated into various platforms, enforcing robust controls at the foundational model level is essential. Enterprises must identify the use of LLMs embedded in SaaS applications and ensure that vendors are accountable for the results generated by these models. This includes LLMs that generate software, where the resilience and reliability of the outputs are crucial.

To mitigate these risks, SaaS vendors must be held accountable for the integration and use of LLMs, ensuring they comply with established governance policies and standards. Assessing third-party GenAI products using a maturity model such as NIST 600-1 provides a structured framework for evaluating the maturity and robustness of GenAI implementations. The Generative Artificial Intelligence Profile (NIST GenAI 600-1) helps organisations identify unique risks posed by GenAI and propose actions for GenAI risk management that align with their goals and priorities. By leveraging such maturity models, enterprises can enhance their third-party governance strategies and ensure that GenAI products are deployed responsibly and effectively.

AI GOVERNANCE COMMITTEE OBJECTIVES

The GenAI governance committee plays a crucial role in GenAI governance by providing strategic oversight and ensuring accountability. The objectives of the governance committee are categorised into four areas: purpose, responsibilities, authority and reporting.

- *Purpose*: The GenAI governance committee shapes the responsible and strategic use of generative AI within an organisation. This committee provides oversight by establishing policies, principles and actionable guidelines that align GenAI projects with the organisation's broader goals and ethical standards. To avoid potential misuse, the committee incorporates checks and balances and sets clear frameworks for data privacy, ethical considerations and transparency in AI practices. Rather than simply overseeing initiatives, successful GenAI governance committees often adopt a working group approach. This involves not only creating governance materials and frameworks but also actively engaging in the hands-on execution of the programme. This working model allows the committee to contribute directly to the development, rollout and monitoring of GenAI projects. By acting as an embedded team within the programme, the committee helps articulate specific business use cases and track their progress to ensure alignment with organisational objectives, resulting in more agile and effective governance.

- *Responsibilities*: The GenAI governance committee provides governance for all GenAI/ML initiatives, develops principles, practices and policies, and establishes an ethical compliance framework to ensure GenAI initiatives adhere to ethical standards.
- *Authority*: To be successful the GenAI governance committee must have the ability to put containment in place for GenAI-generated programmes, promote GenAI education of board and executive committees and provide oversight for accountability, compliance, tracking and/or monitoring of GenAI systems.
- *Reporting*: The GenAI governance committee will monitor the use of GenAI and associated LLMs across platforms and services, report on all GenAI programmes, the use of GenAI and legal and regulatory challenges, and report on key performance indicators (KPIs) as defined by programmes.
- *Membership*: The GenAI governance committee must define clear roles within the governance framework, including security, business operations, legal, privacy, technology, marketing decision makers and subject matter experts (SMEs). SMEs will play a pivotal role in the committee by helping to educate the rest of the committee on what is possible and practicable.

SAMPLE PRINCIPLES FOR GenAI GOVERNANCE

Establishing clear and robust principles is essential for effective GenAI governance. These principles should serve as the foundation for developing, deploying and managing GenAI systems within an organisation.

- *Fairness, transparency and accountability*: Key principles include fairness, which ensures that GenAI systems are designed and implemented without bias, promoting equal treatment and opportunities for all users. Transparency involves making GenAI processes and decisions understandable and accessible, allowing stakeholders to see how and why decisions are made. Accountability mandates that there are clear lines of responsibility for GenAI decisions and outcomes, ensuring that individuals or teams can be held responsible for the performance and ethical use of GenAI systems.
- *Privacy, security and integrity*: Privacy emphasises the protection of personal data, ensuring that GenAI applications comply with data protection laws and respect user consent. As part of this principle, customer data will be excluded from LLM prompts and LLM datasets to safeguard user privacy. Security focuses on safeguarding GenAI systems from threats and vulnerabilities, ensuring that data integrity and system reliability are maintained. The integrity of private or foundation LLM output is part of the ML system owner's responsibilities, ensuring that the generated content is reliable and trustworthy.
- *Sustainability, resilience and responsibility*: Sustainability encourages the development of GenAI technologies that consider long-term environmental and societal impacts, promoting practices that support sustainable growth and ethical use of GenAI innovations. Additionally, business process owners are accountable for the resilience of LLM output used to support that business process, ensuring that GenAI outputs align with business needs and withstand operational challenges. Data scientists that use LLMs are accountable for the resilience of the LLM output, taking responsibility for the quality and robustness of the GenAI-generated content.

These principles collectively guide organisations in building trust, mitigating risks and maximising the benefits of GenAI technologies.

SAMPLE PRACTICES FOR GenAI GOVERNANCE

Implementing effective GenAI governance practices is crucial for organisations to harness GenAI's potential responsibly. Here are some sample practices that can guide organisations in establishing a robust GenAI governance framework:

- *Encouraging employee engagement and learning*: Organisations should promote an environment where all employees are encouraged to explore and learn from the usage of LLMs within SaaS applications. This practice fosters a culture of continuous learning and innovation, enabling employees to understand the capabilities and limitations of GenAI technologies.
- *Clarifying GenAI capabilities*: It is essential to explain to employees and stakeholders that while LLMs may appear to apply judgment to content, they are not capable of human-like judgment by design. This distinction helps manage expectations and ensures a clear understanding of GenAI's role and capabilities within the organisation.
- *Documenting GenAI use cases*: Thoroughly recording and maintaining detailed descriptions of LLM use cases, along with SaaS applications that integrate these models, is essential for effective GenAI governance. This documentation not only supports transparency across the organisation but also enables a better understanding of the impact and utility of GenAI applications within various business processes.
- *Promoting accountability among engineers*: Engineers and developers involved in building ML systems should be held accountable for their work. Encouraging accountability ensures that GenAI systems are designed and implemented with a focus on reliability, security and ethical considerations.
- *Exploring GenAI security tools*: Organisations should explore the use of GenAI security tools that can filter models and record all transactions. These tools help safeguard GenAI systems against potential security threats and ensure the integrity of GenAI-generated outputs.
- *Publishing an accountability model*: Developing and publishing an accountability model for the enterprise is a vital practice. This model should outline the responsibilities of different stakeholders, including developers, business process owners and data scientists. For instance, business process owners are accountable for the resilience of LLM outputs used to support their processes, while data scientists are responsible for ensuring the reliability and accuracy of LLM outputs.
- *Ensuring data privacy and security*: Without proper consent, customer data should be excluded from LLM prompts and datasets to protect privacy and comply with data protection regulations. Additionally, maintaining the integrity of private or foundation LLM outputs is part of the responsibilities of ML system owners, ensuring that GenAI systems operate within the defined ethical and legal boundaries.
- *Training and awareness programmes*: Providing training programmes and resources to educate users about security best practices, regulatory requirements, potential risks and their roles in maintaining a secure GenAI environment promotes a culture of security awareness and accountability across the organisation.

By implementing these practices, organisations can build a robust GenAI governance framework that ensures the responsible and ethical use of GenAI technologies, fosters innovation and maintains trust with users and stakeholders.

SAMPLE RESPONSIBLE GenAI MODEL

Implementing a responsible GenAI model is crucial for ensuring that GenAI technologies

are developed and used ethically, securely and in compliance with regulatory standards. A comprehensive responsible GenAI model includes several key components to address various aspects of GenAI governance.

- *Build GenAI securely*: This involves incorporating secure software development life cycle (SDLC) principles to ensure that GenAI systems are built with security in mind from the ground up. It includes measures to protect against GenAI-specific attacks, such as data poisoning, and ensures the protection of models and training data in production environments. Regular monitoring for compliance with the regulatory landscape is also essential.
- *Security and privacy*: Ensuring GenAI systems are designed to protect sensitive data and maintain user privacy is paramount. Organisations must implement robust security measures to safeguard GenAI models against cyber threats. Protecting models and training data from unauthorised access and manipulation is crucial for maintaining the integrity and trustworthiness of GenAI outputs.
- *Accountability*: Establishing clear guidelines for accountability in GenAI decision making is essential. This includes defining who is responsible for the outcomes and potential negative impacts of GenAI systems. Regular audits and compliance checks ensure that GenAI initiatives align with ethical standards and legal requirements.
- *Transparency and auditability*: GenAI systems should be transparent, and their decision-making processes should be explainable. Implementing traceability mechanisms ensures that GenAI-generated decisions and outputs can be tracked and understood. This transparency helps build trust with users and stakeholders, making GenAI actions predictable and accountable.
- *Bias and fairness*: Addressing bias and ensuring fairness in GenAI systems is critical. This involves applying bias mitigation methods at various stages, including preprocessing (before training), in-processing (during training) and post-processing (after model deployment). Ensuring diverse and representative training datasets helps prevent the perpetuation of societal biases.
- *Inclusion and collaboration*: Engaging various stakeholders in the GenAI design and oversight process ensures that GenAI systems are developed and used ethically. This includes business teams, risk teams, legal and compliance experts, security professionals and public relations teams. Ethical use of GenAI is assessed across all areas to enhance the quality and trustworthiness of outcomes.

By incorporating these elements into a responsible GenAI model, organisations can ensure that their GenAI systems are not only effective but also ethical, secure and compliant with regulatory standards. This comprehensive approach to GenAI governance fosters innovation while safeguarding against potential risks and ethical challenges associated with GenAI technologies.

ADDRESSING GenAI RISKS

The rapid adoption of GenAI technologies presents a range of significant risks that organisations must address to ensure safe and responsible usage. One of the primary concerns is the shift in workforce skill sets, requiring new competencies to effectively leverage GenAI while managing associated risks. Developing the appropriate skills to ask the right questions can maximise the accuracy and completeness of GenAI outputs, minimising the chances of false information or 'hallucinations' generated by the GenAI.

One of the key risks associated with the use of GenAI is the inadvertent exposure of confidential or sensitive corporate data

when using public AI models. Organisations that leverage public generative models, such as GPT-4, may unknowingly provide proprietary or personal data during model training, potentially losing control over sensitive information. This data could then become part of the broader model, accessible to other users, posing a significant risk to corporate confidentiality and security. CISOs must ensure robust data governance policies that prevent such data from being used to train external models, maintaining strict oversight of what data is exposed to third-party platforms.

Risks include ensuring GenAI outputs are accurate and free from biases or false information, protecting sensitive data and ensuring compliance with data protection laws, and safeguarding GenAI models from manipulation and ensuring their integrity. To address these risks, organisations must implement comprehensive security measures, including robust data protection protocols, access controls and continuous monitoring of GenAI systems. Additionally, organisations should develop incident response plans to address potential GenAI-related breaches and issues promptly.

Data privacy and legal concerns are also critical, particularly regarding the potential misuse of GenAI-generated content and the handling of sensitive information. GenAI can inadvertently produce inaccurate or biased outputs, posing risks to fairness and integrity. Moreover, model manipulation and integrity concerns are prevalent, with threats such as those listed in the Open Worldwide Application Security Project (OWASP) Top 10 GenAI attacks, which highlight vulnerabilities that can be exploited to compromise GenAI systems.

The existential risk to human safety, although a more long-term concern, cannot be overlooked. The ability of GenAI to generate misleading or harmful content poses a serious threat to societal well-being if not properly managed. Bias and fairness issues in GenAI outputs can exacerbate existing societal inequalities, making it imperative to implement robust bias detection and mitigation strategies at various stages of GenAI development.

Additionally, the increased attack surface due to the use of GenAI in various applications heightens the risk of cyber security threats. Data leakage through GenAI prompts can extend the attack surface for malicious actors, making it easier for them to access and misuse sensitive information.

Organisations must implement comprehensive risk management frameworks to address these challenges, ensuring that GenAI technologies are used ethically and securely. This involves continuous monitoring, regular audits and adherence to best practices in GenAI governance to mitigate the potential risks associated with GenAI.

The risk of deepfakes

One of the emerging risks associated with GenAI is the creation and dissemination of deepfakes. Deepfakes are synthetic media where a person in an existing image or video is replaced with someone else's likeness. This technology can be used maliciously to spread misinformation, commit fraud or damage reputations. For instance, LinkedIn co-founder Reid Hoffman has discussed the profound implications of deepfakes, highlighting a video where GenAI was used to create a deepfake of him. Such instances underscore the urgent need for robust GenAI governance and ethical standards to combat the misuse of this technology.[2]

LEGAL AND REGULATORY CONSIDERATIONS

AI governance must also address legal and regulatory challenges, such as intellectual property rights, data privacy and compliance with industry-specific regulations. Key legal considerations include ensuring GenAI systems do not infringe intellectual property

rights, upholding user privacy by securing personal data and requiring explicit consent for data processing and aligning GenAI practices with laws such as the General Data Protection Regulation (GDPR) and sector-specific regulations. The legal landscape for GenAI is continually evolving, with new regulations and guidelines being introduced to address emerging challenges. CISOs must stay informed about these changes and adapt their governance strategies accordingly.

GenAI lawsuits

As GenAI technologies have become more prevalent, they have also given rise to various legal challenges and lawsuits. These legal disputes often revolve around issues such as copyright infringement, data privacy and the ethical use of GenAI-generated content. One notable example is the case of the *New York Times* suing OpenAI, arguing that their language models utilised copyrighted content from *Times* articles without permission, potentially allowing models to generate misleading summaries or verbatim text without users needing to access the original content. The case hinges on whether using this content to train language models constitutes copyright infringement, with potential ramifications for how generative AI models handle licensed material.[3]

Deepfake technology has broader legal challenges, especially as it pertains to cyber harassment and identity theft. In many cases, deepfakes are used maliciously, causing reputational harm or portraying individuals in damaging ways, often leading to defamation claims when the content harms someone's public image. Some states have begun enacting specific laws to criminalise these non-consensual AI-generated images and videos, recognising the emotional and reputational damage they can cause.[4]

OpenAI faces a lawsuit by a public interest law company, alleging misuse of private user data collected from the Internet without consent. The case claims OpenAI failed to obtain users' informed consent when collecting personal information such as login credentials and social media data to train its models, raising significant data privacy and ethical concerns.[5]

The case has far-reaching implications, especially for organisations using customer data to train their LLMs. A crucial question is whether these organisations have obtained explicit consent to use customer data for model training purposes. As LLMs require vast datasets, companies are pressured to ensure compliance with data protection regulations such as the GDPR or California's Consumer Privacy Act (CCPA), which mandate transparency and specific user consent for data processing. This consent requirement raises essential questions about data privacy, consent management and ethical AI practices that companies need to address proactively.

These lawsuits highlight the need for clear legal frameworks and ethical guidelines to govern the use of GenAI, ensuring that its development and deployment are conducted responsibly and in compliance with existing laws.

Regulatory landscape of GenAI

The regulatory landscape of AI is rapidly evolving as governments and international bodies strive to address the ethical, legal and societal implications of AI technologies. The European Union (EU) has taken a proactive stance with the proposed EU AI Act, which aims to create a comprehensive regulatory framework for AI applications. This act categorises AI systems into different risk levels — unacceptable, high, limited and minimal — and sets stringent requirements for high-risk AI systems, including rigorous testing, documentation and oversight to ensure they are safe, transparent and non-discriminatory.

In the US, the federal government has also been active in shaping AI regulations. A significant development is the executive

order aimed at promoting the safe, secure and trustworthy development and deployment of AI technologies. This executive order outlines principles for AI governance, emphasising the importance of public trust, ethical standards, data privacy and security. It encourages federal agencies to prioritise AI research that aligns with these principles and to collaborate with international partners to develop harmonised AI standards. These regulatory efforts reflect a growing recognition of the need to balance innovation with robust safeguards to protect individuals and society from the potential risks associated with AI.

LOOKING FORWARD: THE CONTINUED EVOLUTION OF LLM RISK

The understanding and perception of risks associated with LLMs have evolved significantly. Initially, data leakage through prompts of public LLMs was considered the highest risk to enterprises.[6] As time progressed, however, it became evident that the top risk from LLM usage is the generation of incorrect data.[7] This shift highlights the critical issue of data trustworthiness, where LLMs might produce misleading or false information, leading to recursive pollution of data sets and improper use of GenAI outputs.

Furthermore, the widespread integration of LLMs into software by engineers, who select from a range of available components, has expanded the use cases and complexity of LLM applications. Specialised LLMs are increasingly being used by software engineers, who choose from various available components that may have unknown threats, adding another layer of risk. By 2025, it is estimated that there will be 750 million apps using LLMs,[8] including 180.5 million users,[9] 15 per cent of whom are based in the US, and 92 per cent of Fortune 500 companies integrating these models.[10] This extensive adoption underscores the importance of implementing robust governance frameworks to manage the evolving risks.

The attack surface for GenAI continues to expand, with new vulnerabilities emerging regularly. Today, there are better resources available to understand and mitigate these risks. For instance, the OWASP Top 10 risks for LLMs and frameworks such as the Financial Services Information Sharing and Analysis Center (FS-ISAC) GenAI Control Framework provide valuable guidance for organisations to define and deploy controls. These resources help in conducting architectural risk analyses and implementing robust security measures to safeguard against GenAI-specific threats.

Overall, the evolution of LLM risk emphasises the need for continuous monitoring, rigorous validation and the deployment of comprehensive risk management strategies to ensure the responsible and secure use of GenAI.

The integration of LLMs in various applications has exposed significant security vulnerabilities that threat actors can exploit. One major risk is that LLMs make it easier for malicious entities to locate personal information, such as e-mails and social network postings, and craft highly convincing phishing attacks on a large scale. The ability of LLMs to generate realistic and personalised content makes phishing attempts more credible and harder to detect, increasing the likelihood of successful attacks.

Moreover, LLMs have been found to generate software code that includes malware, which can be used in ransomware attacks and other malicious activities. This capability allows cybercriminals to automate the creation of harmful software, lowering the barrier to entry for conducting cyberattacks. Additionally, data leakage through LLM prompts extends the attack surface for threat actors, as sensitive information may inadvertently be revealed through interactions with GenAI systems.

These risks highlight the necessity for stringent governance and robust security

measures to protect against the misuse of LLMs. Organisations must be vigilant in monitoring GenAI-generated content, implementing advanced security protocols and educating users about the potential dangers posed by GenAI technologies.

The use of LLMs is becoming integral to enhancing enterprise security. GenAI in security products is expanding the range of options available to enterprises, particularly in automating tedious and data-intensive tasks. While LLMs are not inherently designed to recognise malicious threat actor patterns, they excel in analysing pattern deviations in user behaviour. This capability allows organisations to generate behaviour trust scores, which can trigger orchestrated and automated workflows in response to identified anomalies.

LLMs help streamline the analysis of log files and documentation sources, significantly improving productivity and the quality of insights derived from data. By focusing on deviations from established behavioural patterns, LLMs can effectively identify potential security threats that might be missed by traditional security measures. For instance, cyber security solutions are beginning to use GenAI to detect unusual patterns that deviate from the norm, providing an additional layer of security.

Furthermore, leveraging ML)systems and LLMs to recognise behavioural patterns rather than building new threat actor profiles enhances the resilience of security frameworks. This approach allows enterprises to respond more dynamically to emerging threats. Overall, the integration of LLMs into security protocols represents a significant advancement in protecting enterprises from an increasingly complex threat landscape.

CONCLUSION

AI is a rapidly evolving technology with the potential to significantly transform enterprises; however, its adoption brings challenges that require meticulous governance. While it can be argued that the CISO plays a pivotal role in developing and implementing robust governance frameworks for GenAI, ensuring ethical use, regulatory compliance and risk management, it is important for each organisation to assess which leader can provide the best leadership for this important transition.

By fostering a culture of transparency, accountability and continuous monitoring, organisations can harness the power of GenAI responsibly and securely. CISOs should take immediate actions such as understanding and assessing the business use cases for GenAI, implementing interim controls to manage risks while developing comprehensive governance frameworks and establishing a cross-functional GenAI governance team to oversee initiatives and ensure compliance with governance guidelines.

In the next 90 days, CISOs should focus on monitoring the progress of GenAI adoption, establishing metrics for reporting, and refining their governance programmes to adapt to the evolving GenAI landscape. By taking these proactive steps, CISOs can position their organisations to leverage the benefits of GenAI while minimising its risks, ensuring that these technologies are used ethically, responsibly and effectively.

The role of the CISO in GenAI governance is crucial in steering the organisation through the complexities of this new frontier. By implementing core security capabilities and enhancing third-party risk governance strategies, CISOs can safeguard against potential threats and maximise the potential of GenAI innovations.

References

1. Malec, M. (January 2024), 'Generative AI Statistics: The 2024 Landscape – Emerging Trends, and Developer Insights', HatchWorks, available at https://hatchworks.com/blog/gen-ai/generative-ai-statistics/ (accessed 29th October, 2024).
2. Cook, J. (April 2024), 'In mind-bending chat with deepfake digital twin, Reid Hoffman discusses Microsoft's big AI hire', Geek Wire,

available at https://www.geekwire.com/2024/in-mind-bending-chat-with-deepfake-digital-twin-reid-hoffman-discusses-microsofts-big-ai-hire/ (accessed 29th October, 2024).
3. Pope, A. (April 10, 2024), 'NYT v. OpenAI: The Times's About-Face', Harvard Business Review, available at https://harvardlawreview.org/blog/2024/04/nyt-v-openai-the-timess-about-face/ (accessed 29th October, 2024).
4. Gerstner, E. (January 2020), 'Face/Off: "DeepFake" Face Swaps and Privacy Laws', International Association of Defense Counsel (IADC), available at https://www.iadclaw.org/defensecounseljournal/faceoff-deepfake-face-swaps-and-privacy-laws/ (accessed 29th October, 2024),
5. Wilowski, M. (June 2023), 'OpenAI Faces Lawsuit Alleging Misuse of Internet Users' Data', Investopedia, available at https://www.investopedia.com/openai-faces-lawsuit-alleging-misuse-of-internet-users-data-7555183 (accessed 29th October, 2024).
6. Sharma, R. (September 2024), 'LLM security: Top risks and best practices', Protecto, available at https://www.protecto.ai/blog/llm-security-risks-best-practices (accessed 29th October, 2024).
7. Hackney, H. (May 2024), 'AI Governance: Is There Too Much Focus on Data Leakage?', Architecture & Governance Magazine, available at https://www.architectureandgovernance.com/data/ai-governance-is-there-too-much-focus-on-data-leakage/ (accessed 29th October, 2024).
8. Uspenskyi, S. (September 2024), 'Large Language Model Statistics And Numbers (2024)', Springs, available at https://springsapps.com/knowledge/large-language-model-statistics-and-numbers-2024 (accessed 29th October, 2024).
9. Fern, J. (October 2024), 'A More-than-Human Ecology: Evolving Generative Artificial Intelligence in Higher Education', Multidisciplinary Digital Publishing Institute (MDPI), available at https://www.mdpi.com/2227-7102/14/10/1102 (accessed 29th October, 2024).
10. Fomenko, M. (July 2024), '50+ Eye-Opening ChatGPT Statistics: Tracing the Roots of Generative AI to Its Global Dominance', Master of Code, available at https://masterofcode.com/blog/chatgpt-statistics (accessed 29th October, 2024).