Infosys topaz

VIEW POINT

MANAGING SECURITY RISKS OF GENERATIVE AI ON AZURE OPENAI

Infosys®
Navigate your next

Table of Content

External Document © 2024 Infosys Limited

1 Introduction

1.1 The Rise of Generative AI

Generative AI, formerly considered a future notion, is rapidly becoming a reality. Its ability to generate amazingly accurate text, artwork, and even code has the potential to alter a wide range of industries, including manufacturing and entertainment. However, amidst these alluring prospects, the disquieting prospect of unparalleled security dangers lurks.

Consider this scenario: a news item written by generative AI that is so realistic that it causes international market panic. Consider a Deepfake video depicting a world leader making controversial statements, causing unrest and instability. These scenarios are no longer just science fiction; they represent the conceivable outcomes of generative AI's significant ability to manipulate and deceive.

The security threats connected with generative AI go far beyond disinformation. As generative AI gains proficiency in replicating human creativity, it opens Pandora's box of vulnerabilities. Malicious actors can use AI to create complex phishing emails, bypass biometric authentication systems, and even create malware that avoids traditional detection methods.

The challenge is not to vilify generative AI but to recognize its intrinsic duality. Just as fire can offer warmth in a comforting fireplace or devour a forest, generative AI's tremendous creative potential can either facilitate progress or cause destruction, depending on the intentions of those using it.

So, how can we maximize the potential of generative AI while mitigating its security risks? This article investigates these challenges and shows how Azure Open AI service may drastically reduce security vulnerabilities.

1.2 The significance of Azure OpenAI

In the complicated world of security risk management, Azure OpenAI offers a challenging problem. Its arsenal of powerful AI models delves into massive data, revealing previously unknown risks such as malware, fraud, and vulnerabilities. Businesses are witnessing a paradigm shift in which vulnerabilities are proactively foreseen, suspicious activity is quickly identified, and vital systems are protected from imminent harm. However, this formidable shield necessitates careful deployment. Inherent biases in its algorithms can obscure specific risks, resulting in a flood of false alarms that hamper decisive responses. Worse, if employed recklessly, its great power can change it from a protector to a deadly weapon.

For Azure OpenAI to safeguard the future, its vision must be free of prejudice, its decision-making procedures transparent and easily understood, and its authority wielded with the utmost ethical rigor. Only then it will be able to move beyond its position as a mere shield and become a guiding beacon, illuminating a route to a more secure and resilient future? The burden rests not only with technology but also with the human hand that guides its use. Prioritizing the values of justice, transparency, and responsible development may help Azure OpenAI achieve its transformative promise, ensuring digital assets and the very fabric of trust in an increasingly interconnected society.

1.3 Purpose of the Article

This study sheds light on the complex interplay between generative AI's brilliant potential and its dark security threats. While it creates words and images with incredible creativity, threats like Deepfakes, automated attacks, and sophisticated scams lurk within its tale. Malicious actors might use their skills to propagate false information, manipulate markets, and undermine trust in the digital sphere.

However, this story is not exclusively about darkness. Enter Azure OpenAI, a skilled architect of security solutions. Its AI-powered technologies act as a deterrent, detecting Deepfakes before they spread, anticipating intrusions, and correcting vulnerabilities before they become critical. Azure OpenAI can reinforce the fabric of digital security, allowing innovation to flourish without being dominated by fear.

Nonetheless, attentiveness is central to this scenario. Transparency and cooperation are critical components that must be incorporated into the development of Azure OpenAI to ensure that its power is used to protect rather than deceive. By accepting ethical principles and encouraging open debate, generative AI's trajectory can be steered toward positive outcomes, establishing a future in which its creative brilliance contributes to a better and safer society.

2 Understanding Generative AI

2.1 What is Generative AI

Generative AI, which emerges from the heart of artificial intelligence, has enormous potential to change the content creation landscape. powered by massive datasets, this emerging technology has the potential to emulate and transcend traditional modes of expression. Its brushstrokes paint colorful landscapes, its quill creates intriguing narratives, and its musical tapestry weaves elaborate symphonies created entirely from data threads. Generative AI promises unrivaled text, code, and audio innovation, from customized artistic ideas to automated workflows.

However, like with any powerful weapon, possible drawbacks lurk in the shadows. Biases buried in training data can reverberate and amplify the outputs, while the simplicity of producing seemingly authentic content creates a fertile ground for misinformation and fraud. Furthermore, the ever-expanding canvas of automation created by generative AI forces observers to address the possible displacement of human labor, necessitating both ethical stewardship and reskilling initiatives.

As a result, the future of generative AI resides not in the unconstrained pursuit of limitless creativity but in a careful balance of unrestrained innovation and ethical consciousness. Transparency and human oversight must be the guiding principles, ensuring that the powerful technology's outputs reflect a symphony of human values and aspirations rather than distorted echoes of bias. By cultivating a collaborative environment in which researchers, developers, and policymakers work together, we can ensure that generative AI becomes a luminous beacon illuminating a future in which technology empowers human creativity, enriches society, and paints a tapestry of progress woven with threads of both ingenuity and accountability.

2.2 Applications of Generative AI

The applications of generative AI are as varied as the human imagination, covering multiple industries, and continually pushing the limits of what is possible. Here are a few examples that spark one's curiosity:

- **Music and Writing**: Artificial intelligence can create individual soundtracks for films, lifelike voices for audiobooks, and even personalized poems and stories. One may see AI creating a breathtaking symphony for an independent film or spinning a riveting story as unique as a person's fingerprint.
- **Software Development**: AI could help build new software features based on user feedback, automate repetitive processes, and produce boilerplate code. It can free up developers' time to pursue more creative endeavors while accelerating the development cycle.
- **Healthcare**: AI can evaluate medical imaging, predict disease outbreaks, and tailor treatment strategies based on a patient's information. AI could potentially help with early cancer detection or drug discovery.
- **Material Science**: In material science, artificial intelligence demonstrates the potential to build innovative materials with desired properties such as heat resistance or superconductivity. Consider the scenario when AI innovates a revolutionary cloth designed to resist stains readily or produces a biocompatible substance suited for manufacturing artificial organs.
- **Marketing and Advertising**: AI can generate individualized ad copy, develop targeted marketing campaigns, and create visually appealing content. One may imagine AI creating the perfect Instagram caption to attract the right audience or creating personalized video advertising for each user.

These examples hint at the limitless potential of generative AI. As technology advances, we may expect even more incredible applications to transform human existence, productivity, and inventiveness. Nonetheless, it is critical to recognize the weight of responsibility that such powers entail. Ethical discussions and conscientious development procedures are crucial in steering generative AI toward positive outcomes, ultimately encouraging a future in which technology enhances human creativity and promotes a more illuminating world for all.

2.3 Potential Risks

In the near future, up to 2025 existing risks will likely dominate near-term threats arising from generative AI. The technology significantly amplifies their scale and speed, introduces new vulnerabilities, and poses novel challenges across three key domains:

- **Digital Dangers**: As cybercrime and hacking escalate, AI empowers both sides. Nevertheless, AI also strengthens cyber defenses, potentially mitigating some harm.
- **Societal and Political Manipulation**: As generative AI becomes more sophisticated and widespreads, the potential for manipulating and deceiving populations becomes more accessible, posing risks as significant as digital threats.
- **Physical Risks**: Embedding AI into critical infrastructure and physical systems such as buildings might introduce novel failure points and attack vulnerabilities if safety and security controls prove insufficient.

These risks do not operate in isolation. They are likely to intertwine and amplify each other, further complicating matters. Moreover, unexpected threats and risks stemming from the inherent unpredictability of AI systems almost invariably emerge.

3 Ensuring Security and Privacy

3.1 Data Security

Organizations embarking on artificial intelligence initiatives, exemplified by tools such as ChatGPT and Generative AI, must ensure the proper implementation of cybersecurity protocols to mitigate online threats.

- **Open Data Input**: Generative AI services, such as code-generating platforms like GitHub Copilot, permit open text input, frequently involving sensitive, private, or proprietary data. It exposes potential vulnerabilities, as the code may encompass confidential intellectual property or sensitive information, such as API keys that grant access to customer data.

- **Intellectual Property Leak and Shadow IT**: The ease of use of web- and app-based generative AI tools raises concerns about unintentional IP leakage and the creation of "shadow IT." Data transmission over the internet inherent in these services necessitates robust security measures, such as VPNs for IP masking and data encryption.

- **Training Data Bias and Privacy**: The massive data sets utilized for training generative AI models might contain sensitive information. Accidental data leakage throughout training presents notable privacy risks, prompting the necessity for meticulous data management and scrubbing procedures.

- **Data Storage and Third-Party Risks**: Data storage in third-party spaces for model training and improvement presents potential security challenges. Organizations should implement comprehensive data security strategies encompassing encryption, access controls, and thorough breach-prevention measures to safeguard sensitive business information.

- **Compliance Implications**: Sharing sensitive data, including Personally Identifiable Information (PII), with third-party AI providers like OpenAI can lead to non-compliance with data privacy regulations such as GDPR and CPRA. Organizations must navigate these intricate legal landscapes to ensure responsible data-handling practices.

- **Synthetic Data Identification Risks**: Synthetic data generated by AI mimics actual data with remarkable accuracy, raising concerns about re-identification. Subtle patterns or details within synthetic data might unintentionally reveal individuals or sensitive features, necessitating meticulous anonymization techniques.

- **Unintentional Data Leaks**: Text- or image-based generative models can inadvertently leak information from training data, thus potentially exposing personal information or confidential business data. Robust filtering and data security protocols are crucial to mitigate such risks.

- **Malicious Misuse and Cyberattacks**: The potential for malicious actors to exploit generative AI to create deepfakes or spread misinformation necessitates vigilance. Additionally, unsecured AI systems become vulnerable to cyberattacks, further amplifying concerns regarding data security.

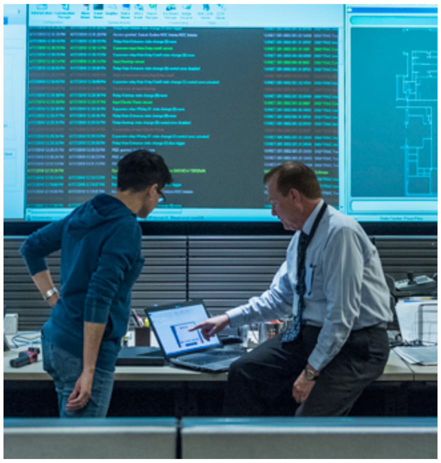
While generative AI presents enticing avenues for innovation across diverse fields, individuals must remain vigilant regarding the inherent security challenges. These concerns encompass unfiltered prompts, inadvertent disclosure of sensitive data, data storage vulnerabilities, complexities surrounding global legal frameworks, and risks of information leakage.

3.2 Privacy-Preserving Techniques

In the face of rising concerns about consumer privacy, a new wave of privacy-preserving technologies (PPTs) has emerged. These innovative solutions reconcile the seemingly conflicting demands for user privacy and effective data-driven marketing. By allowing users to control their personally identifiable information (PII) while enabling service providers and apps to gather valuable insights from aggregated data, PPTs break down the long-held belief that data-driven systems inherently compromise privacy. This technological leap forward leverages diverse approaches, including aggregation technologies like "Aggregated Advanced Privacy" and "Aggregated Conversion Modeling," advanced cryptographic mechanisms like "private set intersection" and "homomorphic encryption," and sophisticated machine learning techniques such as predictive analytics, incrementality measurement, and audience segmentation. Below are the two main models of PPTs:

- **Soft Privacy Technologies**: This model depends on trusted third-party partners that process data in compliance with strict regulations, respecting user consent, control, and robust auditing procedures. Examples encompass differentially private analysis, incorporating statistical noise to safeguard individual data points, and secure communication protocols such as SSL encryption.

- **Complex Privacy Technologies**: This model prioritizes the complete isolation of user data from any third-party involvement, eliminating the risk of privacy breaches even from potentially untrusted entities. Examples include implementing privacy-enhancing technologies (PETs) in secure electronic voting systems and utilizing decentralized data storage solutions.



3.3 Compliance and Data Regulations

In cybersecurity and data privacy laws, AI is rapidly emerging as a critical responsibility for companies seeking to solidify their position within financial services firms' risk, legal, and compliance frameworks. Comparable to cybersecurity and data privacy regulations, the governance of AI necessitates global, federal, state, and industry-specific attention, resulting in a complex, multi-layered landscape.

Europe has assumed a leading role on the global stage concerning AI governance, as evidenced by the EU's introduction of the EU AI Act on Dec. 8. The European Commission has additionally unveiled international guiding principles on AI and a voluntary code of conduct within the Hiroshima AI process, aligning with the EU's legally binding regulations.

In the United States, nearly a dozen states have implemented AI legislation, with more under consideration, often integrated into consumer privacy or sector-specific domains. At the federal level, the proposed American Data Protection and Privacy Act delineates AI regulations affecting companies developing and utilizing AI technologies. Despite the American Data Protection and Privacy Act's current legislative hurdles, the US government has issued guidance through the National Institute of Standards and Technology (NIST).

Various courts, such as the US Fifth Circuit Court and the US Court of International Trade, have addressed the use of AI, suggesting certifications for AI utilization in legal documentation. The White House has delineated principles and priorities within the blueprint for an AI Bill of Rights, with an Executive order issued on Oct. 30, 2023, directing government departments to assess and regulate AI adoption, releasing a streamlined summary to the public.

3.4 Leveraging Azure OpenAI

Azure OpenAI elevates security to a new level by leveraging Microsoft's unparalleled security infrastructure and stringent compliance standards. Powered by Microsoft's comprehensive responsible AI security framework, Azure OpenAI safeguards users' data. It operates within its allocated Azure environment, keeping its information entirely confidential. Additionally, the service utilizes isolated network security, creating a dedicated space for users' data and applications, preventing any intermingling with other clients' information.

User's data is not utilized for training or enhancing Microsoft products or any external services. Employing a pre-trained language model offers an additional security benefit by minimizing the need for extensive data manipulation. It reduces data exposure and supports the implementation of robust security protocols. Moreover, Azure OpenAI has earned a FedRAMP High Provisional Authorization to Operate (P-ATO) in US Commercial regions, satisfying the stringent security demands of US federal agencies and the Department of Defense (DoD) for particular data classifications. Government clients can establish a secure connection between Azure Government and Azure OpenAI's Commercial Service, enabling controlled access to its advanced capabilities.

4 Case Studies

4.1 Real-world Examples of Risks and Solutions

The immense potential of generative AI is accompanied by various security challenges spanning different areas. Below are some real-life examples of a few issues faced by famous individuals.:

Case Study 1: A recent controversy in American politics involved a digital fabrication targeting the esteemed Speaker of the House, Ms. Nancy Pelosi. The fabrication, a demonstrably edited version of her public address slowed down to 75% of its original pace, aimed to convey a false impression of Ms. Pelosi's state of mind through manipulation of her speech and physical appearance. This misleading content was subsequently disseminated across online platforms, finding resonance within segments of the political spectrum aligned with right-wing ideologies.

Case study 2: A Deepfake video featuring N. R. Narayana Murthy, founder of Infosys is circulating on Facebook. The video has been created by altering a conversation from the Business Today Mindrush event, during which Narayana Murthy discussed his views on the Indian economy. However, the original video, which is a year old, has been manipulated in the fake video; he was discussing the benefits of quantum computing software developed by his team. The video claims that the technology can help stock market investors earn approximately INR 2.50 lakh on the first day of using it. Nevertheless, this video has been flagged as false on Facebook. The initial video had been recorded a year earlier during the BT Mindrush event, during which Narayana Murthy imparted his insights to Indian businesspersons. In the Deepfake rendition, the video has focused closely on Narayana Murthy's face, presumably to circumvent any watermarks or distinguishing features in the original footage.

4.2 Lessons Learned

The Deepfake videos targeting Ms. Pelosi and Mr. Murthy offer alarming warnings in the age of misinformation. These fabricated clips, crafted by manipulating existing footage, aimed to erode trust in both public figures and the information itself. By weaponizing Deepfakes to spread false narratives, whether political or financial, these incidents highlight the vulnerability of prominent figures and the need for critical thinking. As technology evolves, individuals must enhance their media literacy to discern truth in a world where appearances can be deceiving. Only then can a resilient society be built that resists the manipulative power of deepfakes.

5 Future Trends and Considerations

5.1 The Evolution of Generative AI

The Evolution of Generative AI: The burgeoning field of generative AI has emerged as one of the most transformative forces in recent technology history. Its impact on the understanding of machine capabilities has been nothing short of profound, driven by both early theoretical groundwork and a recent surge in practical applications.

- **The Early Days**: The narrative generative AI unfolds not as a story of instant triumph but as a saga commencing in the mid-20th century with the pioneering efforts of AI visionaries such as Alan Turing and John McCarthy. This initial phase was marked by theoretical exploration and the development of fundamental models, establishing the critical foundation for subsequent breakthroughs.
- **The New Dawn**: In the 21st century, a notable transformation unfolds as machine learning and neural networks are integrated, introducing heightened sophistication to AI. During this era, AI transitions from basic rule-based systems to models capable of understanding and adaptation.
- **The Recent Explosion**: Generative AI is on an exciting trajectory, transitioning from academic research into the vibrant realm of practical, real-world applications. This shift defines a thrilling era and where theoretical concepts materialize into tangible innovations, reshaping daily experiences for users.

Major upcoming trends in Generative AI

- **The impact on Content creation**: The ramifications of these advancements are far-reaching. In content creation, AI empowers the generation of rich, context-aware, and engaging material, significantly augmenting the capabilities of human writers. Likewise, in customer service, AI-powered chatbots deliver increasingly coherent and helpful responses, enhancing customer experience and operational efficiency.
- **Application in Entertainment and Virtual Reality**: The entertainment industry harnesses these innovations to craft immersive and visually striking experiences. Customized dynamic and personalized content in marketing has opened unprecedented engagement opportunities. Virtual reality is also reaping the benefits as AI-generated environments evolve to be more realistic and interactive.
- **Growing Role in Drug Discovery**: In healthcare, a revolution is spearheaded by generative AI, particularly in drug discovery and personalized medicine. AI algorithms can forecast the interactions of diverse drugs with the human body, thereby expediting the drug development timeline. Within the domain of customized medicine, AI serves to facilitate treatments finely tuned to individual genetic profiles.

5.2 Anticipating Future Risks

These are the most pressing threats we face from generative AI:

- **Supercharged Cyberattacks**: AI significantly amplifies existing cyber threats, creating faster, more effective, large-scale attacks. Imagine personalized phishing emails so convincing that they are nearly impossible to detect or realistic replicas of malware that bypass traditional defenses. While complete automation of hacking seems unlikely by 2025, AI will undoubtedly make attackers more dangerous.
- **Vulnerable Infrastructure**: Integrating AI into critical systems creates new attack surfaces through "data poisoning" (manipulating training data), "prompt injection" (hijacking outputs), "model inversion" (extracting sensitive information), "perturbation" (misclassifying data), and targeted attacks on AI's computing power. These vulnerabilities could disrupt vital infrastructure and services.
- **Information Distortion**: Prepare for a flood of Deepfakes and hyper-realistic AI-generated content polluting the information landscape. It includes fake news, personalized disinformation, manipulation of financial markets, and even influencing the criminal justice system. By 2026, Deepfakes could make up a significant portion of online content, eroding public trust in institutions and fueling polarization and extremism. Current authentication solutions like watermarks are unreliable and need constant updates to keep pace with evolving AI.
- **Political Manipulation**: AI tools can sway public opinion on political issues and amplify the reach and impact of disinformation and misinformation. The ability to generate micro-targeted propaganda on a massive scale with unprecedented sophistication seriously threatens democratic processes and social cohesion.
- **Hidden Dangers in Critical Systems**: Integrating generative AI into crucial systems and infrastructure introduces vulnerabilities. Imagine data leaks, biased decision-making, or even "hallucinations" - AI glitches leading to compromised human judgment - due to poor security and opaque algorithms. Inappropriate use by large organizations could trigger cascading failures and amplify these risks. Additionally, relying on potentially fragile, opaque supply chains controlled by a few companies further exacerbates these dangers.
- **Weaponizing Knowledge**: In the wrong hands, generative AI can become a tool for assembling knowledge on physical attacks, like crafting chemical, biological, and radiological weapons. While leading AI firms develop safeguards against such dangerous outputs, their effectiveness remains uneven. Barriers like acquiring components and acquiring manufacturing expertise still exist, but they are steadily falling, and generative AI could accelerate this decline.

5.3 Preparing for Ethical and Regulatory Changes

Ensuring the ethical and responsible research and implementation of Generative AI (GenAI) remains a top priority. Ethical considerations, encompassing governance, privacy, and data ethics, demand meticulous attention, guiding organizations towards a robust ethical framework for the responsible advancement and utilization of AI. Managing AI involves establishing guidelines, standards, and norms governing the creation and utilization of AI systems. Explicit rules and regulations are crucial to ensuring AI's ethical and responsible utilization. Equally significant in the responsible creation and use of AI is data ethics. Given AI's heavy reliance on data, it is imperative to ensure that data collection and usage adhere to ethical and legal standards. Companies must ensure ethical and impartial data collection to train AI models, thereby mitigating the perpetuation of societal biases. Furthermore, they must maintain control over their data, ensuring privacy is upheld throughout the entire AI development process.

6 Conclusion

6.1 The Role of Azure OpenAI in Risk Management

In today's complex and interconnected world, organizations are placing a high priority on robust risk management strategies. This trend is highlighted by Gartner, Inc., which forecasts a 14.3% year-over-year surge in global end-user spending on security and risk management, which is projected to reach $215 billion in 2024. This substantial increase from the estimated $188.1 billion spent in 2023 reflects the clear commitment of organizations across various industries to proactively identify, assess, prioritize, and mitigate the impact of uncertain events. By strategically allocating resources and implementing systematic risk management processes, organizations can enhance their resilience, ensure business continuity, and navigate the volatile landscape of the 21st century with greater confidence and preparedness.

In today's dynamic business landscape, organizations worldwide embrace AI as a powerful tool for navigating risk and uncertainty. This versatile technology goes beyond prediction, empowering them to adapt and thrive in changing environments.:

- **Identify**: Accurately pinpoint internal and external factors that threaten goals, from market shifts to operational inefficiencies.
- **Prioritize**: Leverage data analysis to differentiate high-impact risks, allowing for strategic resource allocation and informed decision-making.
- **Mitigate**: Proactively implement safeguards, like safety protocols, backup plans, and security measures, to minimize potential damage from identified risks.
- **Stay Compliant**: Navigate the regulatory landscape effectively, avoiding fines and reputational harm through AI-powered compliance monitoring.
- **Ensure Business Continuity**: Develop robust contingency plans to bounce back quickly from disruptions, minimizing downtime and operational losses.
- **Achieve Financial Stability**: Protect assets, mitigate financial risks, and uphold stakeholder trust through AI-driven financial forecasting and risk assessment.
- **Make Strategic Decisions**: Navigate unpredictable environments confidently, using AI insights to inform crucial business decisions and capitalize on emerging opportunities.

By harnessing AI algorithms and advanced data analytics, Microsoft achieved a significant milestone in 2021 by thwarting over 70 billion email and identity threats. This accomplishment underscores the tangible influence of AI in mitigating cyber risks. This innovative solution transcends conventional risk management methods by analyzing IoT device and sensor data to anticipate and forestall equipment failures and operational disruptions. Such a proactive stance boosts efficiency, reduces downtime, and fortifies business continuity.

By adopting AI as a collaborative partner in risk management, organizations can transcend reactive strategies, cultivating resilience for the future. This transformation turns uncertainty into a catalyst for growth and success.

6.2 Final Thoughts on Responsible Generative AI Use

While generative AI boasts impressive capabilities, acknowledging its limitations is crucial. However, these advanced models remain tethered to their training data and algorithms. Users must exercise caution, refraining from solely relying on AI outputs for critical decisions. Human judgment remains irreplaceable in complementing AI-generated content.

- **Bias in Data**: Generative AI, trained on vast datasets, can unwittingly perpetuate existing biases. Mitigating these requires proactive measures like careful data selection, augmentation, and balancing to build diverse and representative datasets for equitable outcomes.
- **Transparency**: Clear communication about AI-generated content is essential to avoid confusion and deception. Disclosing its origin when shared publicly fosters trust and upholds ethical standards.
- **Sensitive Information**: Protecting privacy demands anonymizing or removing sensitive data during training. Robust security measures like encryption, access controls, and regular audits safeguard against unauthorized access and potential breaches.
- **Accountability and Explainability**: As AI advances, ensuring accountability and explainability becomes paramount. Documentation of the development and deployment process establishes a transparent chain of responsibility. Explainability in AI decision-making empowers users to understand the factors influencing AI outputs.
- **Continuous Monitoring and Iteration**: Regular evaluation, user feedback mechanisms, and ethical considerations enable iterative adaptation and refinement of AI systems. Establishing feedback loops empowers users to report problematic outputs or biases.
- **Educating and Empowering Users**: Effective risk management involves educating and empowering users to navigate potential pitfalls. Providing guidelines and best practices encourages critical assessment of AI-generated content. User education initiatives should raise awareness about limitations, biases, and risks associated with generative AI.
- **Legal and Ethical Concerns**: Generative AI presents legal and ethical concerns, including intellectual property rights and data ownership. Adhering to relevant legal frameworks, obtaining necessary rights and permissions, and ensuring

transparency in data processing are crucial to avoid legal repercussions and ethical dilemmas.

- **Collaboration with Experts and Stakeholders**: Collaboration with experts and stakeholders, including ethicists, legal professionals, data scientists, and domain experts, brings diverse perspectives to identify and address potential risks and ethical considerations, optimizing risk management.
- **Content Guidelines and Review**: Establishing content guidelines and review processes ensures AI-generated content aligns with organizational values, brand image, and legal requirements. Human oversight and editorial control remain essential for maintaining accuracy, quality, and relevance.
- **Regular Updates and Maintenance**: Regular updates and maintenance are essential to incorporate improvements, address emerging threats, and ensure compliance with changing ethical standards as generative AI technologies evolve.

By proactively addressing these limitations and risks, we can harness the power of generative AI responsibly and ethically, paving the way for a future where AI complements human judgment to achieve optimal outcomes.

7 References

7.1 Citing Sources and Additional Reading

- Annex B: Safety and Security Risks of Generative Artificial Intelligence to 2025
- 7 Deepfake Controversies That Rocked 2023
- Future proof—Navigating risk management with Azure OpenAI Service
- Responsible Use of Generative AI
- 8 Generative AI Security Risks That You Should Know
- What are privacy preserving technologies?
- Key Trends in Generative AI and Their Impact on the future
- 10 ways generative AI and Azure OpenAI Service are transforming businesses
- Security and Ethical Concerns of Generative AI for 2024

Authors

Chandan Malu
Principal Technology Architect,
Infosys Center for Emerging
Technology Solutions (iCETS)

Ritu Kumari Singh
Senior Associate Consultant,
Infosys Center for Emerging
Technology Solutions (iCETS)

The incubation center of Infosys called 'Infosys Center for Emerging Technology Solutions' (iCETS) focuses on incubation of NextGen services and offerings by identifying and building technology capabilities to accelerate innovation. The current areas of incubation include AI & ML, Blockchain, Computer Vision, Conversational interfaces, AR-VR, Deep Learning, Advanced analytics using video, speech, text and much more. For more information, please reach out to us at icets@infosys.com.

Infosys Topaz is an AI-first set of services, solutions and platforms using generative AI technologies. It amplifies the potential of humans, enterprises and communities to create value. With 12,000+ AI assets, 150+ pre-trained AI models, 10+ AI platforms steered by AI-first specialists and data strategists, and a 'responsible by design' approach, Infosys Topaz helps enterprises accelerate growth, unlock efficiencies at scale and build connected ecosystems. Connect with us at infosystopaz@infosys.com.

For more information, contact askus@infosys.com

Infosys®
Navigate your next

© 2024 Infosys Limited, Bengaluru, India. All Rights Reserved. Infosys believes the information in this document is accurate as of its publication date; such information is subject to change without notice. Infosys acknowledges the proprietary rights of other companies to the trademarks, product names and such other intellectual property rights mentioned in this document. Except as expressly permitted, neither this documentation nor any part of it may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, printing, photocopying, recording or otherwise, without the prior permission of Infosys Limited and/ or any named intellectual property rights holders under this document.