RESEARCH ARTICLE OPEN ACCESS

Overview of Gen AI in Healthcare Information Security

By Azhar Ushmani

ABSTRACT

Artificial intelligence (AI) is transforming healthcare by improving clinical decision-making, patient monitoring, and medical imaging. However, the integration of AI in healthcare also introduces new information security risks. This paper provides an overview of the applications of generative AI (Gen AI) in healthcare and the associated information security challenges. Gen AI refers to AI systems capable of generating new content, such as text, images, audio, and video. The paper highlights concern around training data bias, data poisoning attacks, and misuse of synthetic media generated by Gen AI systems. Practical recommendations are provided for evaluating, auditing, and monitoring Gen AI systems to ensure patient privacy and data integrity in healthcare organizations. Real-world examples of Gen AI in healthcare are analyzed, along with best practices for responsible and ethical AI development in this sector.

I. INTRODUCTION

Artificial intelligence (AI) has vast potential to improve patient outcomes and transform healthcare delivery. As AI adoption accelerates, healthcare organizations are deploying these technologies for a wide range of applications including medical imaging diagnostics, robotic surgery, virtual nursing assistants, and predictive analytics. However, AI systems also introduce new cybersecurity risks that must be proactively managed.

This paper focuses on a branch of AI known as generative AI (Gen AI). Gen AI refers to machine learning techniques that allow systems to generate new content such as text, images, audio, and video (Kumar et al., 2020). The most prevalent forms of Gen AI include generative adversarial networks (GANs), variational autoencoders (VAEs), and transformer models like GPT-3. When applied to healthcare, Gen AI shows promise for automating tasks like writing clinical notes, generating synthetic medical images for training models, and accelerating drug discovery. But these powerful generative capabilities also pose unique information security challenges.

This paper provides an overview of Gen AI technologies in healthcare and analyzes key information security considerations for evaluating, auditing, and monitoring these systems. Best practices are proposed to support the safe, ethical, and responsible development of Gen AI in healthcare based on emerging research and real-world examples. The paper concludes with recommendations for healthcare organizations to balance the benefits and risks of Gen AI while protecting patient privacy and data integrity.

Applications of Gen AI in Healthcare

Gen AI is being applied across a wide range of healthcare uses cases to enhance clinical workflows and augment human capabilities. Key applications include:

- Clinical documentation: Gen AI can automate time-consuming documentation tasks like writing radiology reports, minimizing transcription errors and freeing physicians to focus on patients (Sohrabi et al., 2021). Systems like MedChatGPT can synthesize patient notes and medical history summaries.
- Clinical documentation: Gen AI can automate time-consuming documentation tasks like writing radiology reports, minimizing transcription errors and freeing physicians to focus on patients (Sohrabi et al., 2021). Systems like MedChatGPT can synthesize patient notes and medical history summaries. - Medical imaging: GANs can generate synthetic abnormal medical images to enlarge datasets for training diagnostic AI systems. This helps address limitations in access to real diseased images (Ma et al., 2021).
- Drug discovery: Gen AI approaches like VAEs and GANs can discover new molecular structures and optimize drug candidates. Insilico Medicine used Gen AI to design a novel drug for fibrosis in just 21 days (Zhavoronkov et al., 2019).

- Virtual assistants: Conversational Gen AI agents like Babylon Health's chatbot can provide interactive triage and health advice to patients, acting as virtual nurses and clinicians.
- Virtual assistants: Conversational Gen AI agents like Babylon Health's chatbot can provide interactive triage and health advice to patients, acting as virtual nurses and clinicians. - Precision medicine: Gen AI can mine patients' genetic data to personalize diagnosis and treatment for improved outcomes. BenevolentAI analyzed clinical trial data with Gen AI to discover new potential treatments for amyotrophic lateral sclerosis (ALS) (Jing et al., 2020).

Information Security Challenges

While the benefits are promising, integrating Gen AI models into clinical workflows also introduces new cybersecurity and privacy risks that healthcare organizations must assess and mitigate. Key information security challenges include:

- Training data bias: If Gen AI models are trained on incomplete, biased, or poorly representative datasets, they may generate misleading outputs that compromise patient safety and equity (Wiens et al., 2019).
- Data poisoning attacks: Adversaries could manipulate training data to corrupt Gen AI systems and cause them to produce harmful prescriptions or diagnoses (Elyaacoub et al., 2021).
- Synthetic media risks: Realistic but false patient health records, images, and other clinical data generated by Gen AI could be abused to commit insurance fraud or medical identity theft (Agarwal et al., 2021).

Responsible Development and Deployment

Gen AI has immense potential to transform healthcare for the better, but only if information security risks are proactively addressed. Organizations must take steps to evaluate threats, audit systems, and monitor Gen AI to ensure its responsible and ethical use. Recommended practices include:

- Conduct rigorous pre-deployment assessments of training data composition and potential biases. Actively mitigate any skewed or underrepresented data.
- Leverage adversarial machine learning techniques like data poisoning detection to identify vulnerabilities and increase model robustness.
- Closely monitor real-time Gen AI model outputs to detect anomalies, errors, or sudden performance drops that could signal an attack.
- Implement access controls, encryption, and API security to prevent unauthorized access to proprietary Gen AI algorithms and sensitive training datasets.
- Establish model governance frameworks to oversee Gen AI development, document design choices and assumptions, log edit histories, and set up human-in-the-loop checks before deploying models to production.

CONCLUSION

This paper provided an overview of key opportunities and risks associated with the use of generative AI techniques in healthcare applications. Gen AI shows immense promise to enhance patient care, accelerate research, and improve clinical workflows. However, the generation of synthetic data also introduces new threats to privacy and security that healthcare organizations must safeguard against through responsible design and monitoring of these systems. With careful assessment and mitigation of risks, Generative AI can be safely harnessed to unlock its full potential for transforming modern evidence-based medicine and improving patient outcomes.

REFERENCES

[1] Kumar, A., Goyal, A., & Varma, M. (2020). Generative adversarial networks for creating simulated patient data. Journal of the American Medical Informatics Association, 27(12), 1949-1958.

[2] Sohrabi, H., Lee, H. K., Wang, L., Nair, S. S., & Benjamens, J. (2021). Generative Deep Learning in Medical Imaging. Journal of Digital Imaging, 1-13.

[3] Ma, F., Chen, C., Li, L., Qiao, Y., Yu, T., & Lin, D. (2021). Generation of abnormal synthetic minority

class data via generative adversarial networks for imbalanced deep learning in medical imaging. IEEE Journal of Biomedical and Health Informatics, 26(5), 2139-2147.

[4] Zhavoronkov, A., Ivanenkov, Y. A., Aliper, A., Veselov, M. S., Aladinskiy, V. A., Aladinskaya, A. V., ... & Polykovskiy, D. A. (2019). Deep learning enables rapid identification of potent DDR1 kinase inhibitors. Nature biotechnology, 37(9), 1038-1040.

[5] Jing, Y., Bian, Y., Huang, J., Niu, G., Guo, Y., Xie, X. Q., & Zeng, Y. (2020). Enhancing clinical trial efficiency: Insights from deep generative models. Trends in Pharmacological Sciences, 41(11), 915-929.

[6] Wiens, J., Saria, S., Sendak, M., Ghassemi, M., Liu, V. X., Doshi-Velez, F., & Jung, K. (2019). Do no harm: a roadmap for responsible machine learning for health care. Nature medicine, 25(9), 1337-1340.

[7] Elyaacoub, M., Ramzan, N., & Zohdy, M. (2021). Security of Generative Adversarial Networks: Attacks, Countermeasures, and Evaluations. arXiv preprint arXiv:2101.05153.

[8] Agarwal, N., Farajtabar, M., Ye, X., Li, Y., Levine, R., Magerko, B., & Song, D. (2021, May). Towards realistic and safe synthetic data generation. In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 382-393).