

Chapter 10


Securing the Imagination: Security and Privacy Challenges in Generative AI Systems

Nikhil Kumar Goyal

 <https://orcid.org/0009-0007-4532-8033>

Poornima University, Jaipur, India

Sumit Kumar Kapoor


 <https://orcid.org/0009-0005-6291-3176>

Poornima University, Jaipur, India

Kapil Gulati

Poornima University, Jaipur, India

Bright Keswani

 <https://orcid.org/0000-0003-1464-0431>


Poornima University, Jaipur, India

Ashish Avasthi

 <https://orcid.org/0000-0003-3069-1984>

Poornima University, Jaipur, India

Sangita Gupta

 <https://orcid.org/0009-0009-3751-8317>

Poornima University, Jaipur, India

ABSTRACT

Generative Artificial Intelligence (AI) has rapidly become a mainstream technology capable of autonomously producing realistic and diverse media content, including text, images, video, music, and code. As models like GPT, DALL·E, and Stable Diffusion transform industries such as entertainment, marketing, education, and software development, they also introduce a complex landscape of security and privacy risks. This chapter, per the authors, explores critical vulnerabilities such as prompt injection, adversarial attacks, data poisoning, and model inversion, which can lead to unethical outputs and exposure of sensitive training data. It further examines threats arising from large-scale data scraping, copyright violations, and lack of user consent, while also addressing ethical issues tied to misinformation, deepfakes, and surveillance. The chapter reviews technical safeguards including differential privacy, federated learning, watermarking, adversarial training, and secure deployment methods like sandboxing and real-time monitoring.

DOI: 10.4018/979-8-3373-5616-7.ch010

Copyright © 2026, IGI Global Scientific Publishing. Copying or distributing in print or electronic forms without written permission of IGI Global Scientific Publishing is prohibited. Use of this chapter to train generative artificial intelligence (AI) technologies is expressly prohibited. The publisher reserves all rights to license its use for generative AI training and machine learning model development.

1. INTRODUCTION

1.1 The Rise of Generative AI and Its Societal Impact

Generative Artificial Intelligence (AI) once was a new area of research; it is now an all-embracing and revolutionary technology, which has redesigned the process of content creation, consumption, and personalization. Generative AI systems have found their way to many applications such as text-based interactions with language models, synthetic images, deepfake media, AI-assisted music and code creation, and more, which penetrate society on a massive scale. Such models can create the equivalent results with human creativity and already open new horizons in the world of journalism and art, healthcare, and software engineering.

This authority is however, not free. Generative AI systems are also prone to potential uses that are not intended and potentially harmful exactly because of the nature that makes them compelling and makes them frightening: to autonomously generate new and realistic content. Cases of misinformation, user impersonation, and accelerated bias as AI occurs have already been recorded on various platforms. With the prominence of these systems, the necessity of securing that they are used in terms that do not infringe on the privacy of the users is becoming an issue of prime concern (Abadi et al., 2016).

Figure 1. Generative AI Use Cases & Applications



1.2 Why Security and Privacy Matter in Creative AI

In contrast to the traditional AI systems where a classification or prediction problem is addressed, generative AI works in the highly open-ended domain, therefore,

26 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/securing-the-imagination/395329

Related Content

Big Data, 3D Printing Technology, and Industry of the Future

Micheal Omotayo Alabi (2017). *International Journal of Big Data and Analytics in Healthcare* (pp. 1-20).

www.irma-international.org/article/big-data-3d-printing-technology-and-industry-of-the-future/204445

EMG-Based Mobile Assessment System for Neck and Shoulder Fatigue

Pei Lun Lai, Hsiu-Sen Chiang and Qi-An Huang (2017). *International Journal of Big Data and Analytics in Healthcare* (pp. 39-50).

www.irma-international.org/article/emg-based-mobile-assessment-system-for-neck-and-shoulder-fatigue/204447

Usage and Analysis of Big Data in E-Health Domain

Sushruta Mishra, Hrudaya Kumar Tripathy, Brojo Kishore Mishra and Soumya Sahoo (2022). *Research Anthology on Big Data Analytics, Architectures, and Applications* (pp. 417-430).

www.irma-international.org/chapter/usage-and-analysis-of-big-data-in-e-health-domain/290994

Voluntary Reporting of Performance Data: Should it Measure the Magnitude of Events and Change?

Vahé A. Kazandjian (2018). *International Journal of Big Data and Analytics in Healthcare* (pp. 27-37).

www.irma-international.org/article/voluntary-reporting-of-performance-data/209739

Improving Customer Experience Using Sentiment Analysis in E-Commerce

Vinay Kumar Jain and Shishir Kumar (2017). *Handbook of Research on Intelligent Techniques and Modeling Applications in Marketing Analytics* (pp. 216-224).

www.irma-international.org/chapter/improving-customer-experience-using-sentiment-analysis-in-e-commerce/170348