Chapter 9
Generative AI for Cybersecurity
An Introduction

Kritika
https://orcid.org/0000-0002-1186-6032
Independent Researcher, India

ABSTRACT

The intersection of cybersecurity and generative artificial intelligence (AI) has become a crucial frontier as the digital landscape changes. By examining the interaction between AI-powered attacks and defence mechanisms and concentrating on applications like anomaly detection, synthetic data generation, automated incident response, and forensics, the chapter examines the potential of generative artificial intelligence (AI) in redefining conventional cybersecurity paradigms. In order to reduce the hazards associated with deepfakes and synthetic media, the chapter discusses the examination of adversarial machine learning techniques and strategies. Along with offering guidance on incorporating AI into security operations, encouraging human-AI cooperation, and building strong AI skills, it also discusses the ethical ramifications of AI-driven security procedures. It also serves as a comprehensive guide for security professionals, researchers, and decision-makers, offering a holistic understanding of the synergies between AI and cybersecurity.

1. INTRODUCTION

A class of machine learning models known as "generative artificial intelligence" (AI) is able to produce new types of data, including text, images, audio, and video, by using patterns and attributes that are discovered from training data (Radford

DOI: 10.4018/979-8-3693-8557-9.ch009

et al., 2022). These models replicate the underlying distribution of the training data by using deep neural networks and sophisticated algorithms to generate new, synthetic material. Generative AI has become a game-changing technology that is expanding the realms of art, entertainment, and science study, among other fields, beyond what was previously considered not feasible. One area of technology that is constantly developing and receiving a lot of interest is generative artificial intelligence (AI). This cutting-edge field includes a wide range of methods and models that let computers produce, develop, or synthesise original, meaningful material. Generative AI is pushing the envelope of what is feasible, from text generation and image synthesis to music composition and video creation, and it is bringing in a new age of creativity and invention. The fundamental idea of generative artificial intelligence is unsupervised learning, in which models are trained to recognize implicit correlations and patterns in incoming data without the need for explicit labels or targets (Goodfellow et al., 2016). This methodology enables the models to produce new instances that have resemblance to the training examples by teaching them about the intricate distributions and structures present in the data. One of generative AI's main benefits is its capacity to produce original content that has never been seen before, which may open up new avenues for problem-solving, creativity, and invention. Table 1 highlights the different model architectures and techniques of generative AI. Natural language generation (NLG) is a popular use of generative AI that entails machines producing text that is legible by humans. Generative AI systems are capable of producing coherent and contextually relevant content, such as news articles, fiction, poetry, and code, thanks to deep learning approaches like transformers and language models (Radford et al., 2019). By enabling the automation of numerous writing jobs, these models have the potential to revolutionise industries like journalism, content development, and customer service.

Table 1. Generative AI model architectures and techniques

Approach	Description
Generative Adversarial Networks (GANs)	Consist of two neural networks (generator and discriminator) trained adversarially. The generator creates synthetic data, and the discriminator tries to distinguish between real and generated data.
Variational Autoencoders (VAEs)	Learn to encode input data into a latent space representation and decode from that representation to reconstruct or generate new samples. Allow for efficient manipulation and exploration of the data distribution.
Autoregressive Models	Effective for generating sequential data like text or images. Predict the next element in a sequence based on previous elements, allowing for the generation of coherent and contextually relevant content. Examples: GPT, DALL-E.

continued on following page