

# Comprehensive Review of Synthetic Data Generation Techniques and Their Applications in Healthcare, Finance, and Marketing

Prof. U. B. Bhadange<sup>1</sup>, Swamini Jadhav<sup>2</sup>, Bhagwati Jadhav<sup>3</sup>, Sneha Ghatol<sup>4</sup>, Piyush Kahale<sup>5</sup>

Guide, Department of Artificial Intelligence and Data Science<sup>1</sup>

Students, Department of Artificial Intelligence and Data Science<sup>2,3,4,5</sup>

Pune Vidyarthi Griha's College of Engineering and Shrikrushna S. Dhamankar Institute of Management, Nashik  
Savitribai Phule Pune University (SPPU)

swamineejadhav@gmail.com, bhagwatijadhav271@gmail.com,

snehaghatol01@gmail.com, pyush671@gmail.com, urmila.bhadange@pvgcoe.org

**Abstract:** *The demand for privacy-preserving, high-quality data has driven the rapid development of synthetic data generation techniques. Data scarcity, privacy regulations, and the need for large-scale datasets are some of the challenges these methods aim to address. Key methodologies for synthetic data generation include Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and rule-based systems. This review highlights the strengths, limitations, and practical applications of these techniques across various fields. It also explores ethical considerations related to privacy and fairness, focusing on privacy-preserving models such as differential privacy and federated learning. Despite the potential of synthetic data to overcome major barriers in data-driven industries, issues around data fidelity, fairness, and utility remain unresolved. Future research should prioritize the responsible use of synthetic data.*

**Keywords:** Synthetic data generation, Privacy-preserving data, Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), Differential privacy