Synthetic Data Generation for Privacy-Preserving Research Training Course

Professional Development Training Course Outline

Category	Research and Data Analysis	**Duration**	5 Days
Base Fee	$1,500 USD	**Delivery Mode**	Online / On-site Hybrid

Course Introduction

In today’s data-driven world, researchers often confront challenges when working with sensitive data involving personal, confidential, or regulated information. Synthetic Data Generation for Privacy-Preserving Research Training Course focuses on the critical intersection of sensitive data research and privacy-preserving methodologies, offering advanced techniques in synthetic data generation to uphold ethical standards and legal compliance. Through a blend of theory and hands-on exercises, participants will learn how to conduct secure research without compromising data fidelity or subject privacy.

This course provides essential skills to develop synthetic datasets, simulate real-world patterns, and preserve analytical value while mitigating risks such as re-identification. It also explores the latest advances in differential privacy, machine learning-based data synthesis, and federated learning, empowering professionals and researchers to ethically and confidently tackle sensitive research topics across sectors like healthcare, finance, and social sciences.

Programme Curriculum

Synthetic Data Generation for Privacy-Preserving Research Training Course

Introduction

In today’s data-driven world, researchers often confront challenges when working with sensitive data involving personal, confidential, or regulated information. Synthetic Data Generation for Privacy-Preserving Research Training Course focuses on the critical intersection of sensitive data research and privacy-preserving methodologies, offering advanced techniques in synthetic data generation to uphold ethical standards and legal compliance. Through a blend of theory and hands-on exercises, participants will learn how to conduct secure research without compromising data fidelity or subject privacy.

This course provides essential skills to develop synthetic datasets, simulate real-world patterns, and preserve analytical value while mitigating risks such as re-identification. It also explores the latest advances in differential privacy, machine learning-based data synthesis, and federated learning, empowering professionals and researchers to ethically and confidently tackle sensitive research topics across sectors like healthcare, finance, and social sciences.

Course Objectives

1. Understand the foundations of sensitive data research and regulatory frameworks (e.g., GDPR, HIPAA).
2. Learn the principles of synthetic data generation and its importance in privacy-focused research.
3. Explore privacy-preserving data analysis techniques for compliance and confidentiality.
4. Apply differential privacy algorithms for secure data transformation.
5. Gain hands-on experience in generating high-fidelity synthetic datasets.
6. Implement machine learning models for synthetic data creation.
7. Understand the implications of bias and fairness in synthetic data.
8. Evaluate the utility vs. privacy trade-off in synthetic datasets.
9. Use Python and open-source tools (e.g., SDV, Gretel.ai) for synthetic data workflows.
10. Examine real-world use cases from healthcare, finance, and social research.
11. Analyze the ethical and societal impact of synthetic data.
12. Learn how to audit and validate synthetic datasets for accuracy and compliance.
13. Develop best practices for secure data sharing and collaboration using synthetic data.

Target Audience

1. Academic Researchers
2. Data Scientists
3. Healthcare Analysts
4. Social Scientists
5. Government Researchers
6. Policy Analysts
7. Financial Data Analysts
8. Ethical AI Developers

Course Duration: 5 days

Course Modules

Module 1: Foundations of Sensitive Data Research

- Definition and types of sensitive data
- Regulatory landscape: GDPR, HIPAA, FERPA
- Risks of data breaches and misuse
- Overview of data anonymization vs. synthetic data
- Introduction to privacy-preserving methods
- **Case Study**: Sensitive survey data in mental health research

Module 2: Introduction to Synthetic Data

- What is synthetic data?
- Types: Fully synthetic, partially synthetic, hybrid
- Benefits of synthetic data in research
- Challenges in generation and use
- Tools and platforms for synthetic data
- **Case Study**: Creating synthetic patient records for public research

Module 3: Privacy-Preserving Techniques

- Differential privacy fundamentals
- K-anonymity, l-diversity, t-closeness
- Cryptographic approaches (e.g., homomorphic encryption)
- Comparison with data masking and redaction
- Risk assessment of privacy breaches
- **Case Study**: Applying differential privacy in education data sharing

Module 4: Tools & Technologies for Synthetic Data Generation

- Overview of SDV (Synthetic Data Vault)
- Using Gretel.ai and other AI platforms
- Building synthetic datasets with Python
- Evaluating synthetic data quality
- Integrating tools in existing pipelines
- **Case Study**: Automating synthetic data creation for fintech datasets

Module 5: Ethical and Legal Considerations

- Ethical use of synthetic data
- Informed consent and transparency
- Addressing bias in synthetic data
- Legal implications and compliance
- Mitigating unintended consequences
- **Case Study**: Ethics review of synthetic datasets in social science

Module 6: Bias, Fairness & Utility in Synthetic Data

- Detecting bias in original and synthetic data
- Ensuring fairness in AI-generated datasets
- Measuring data utility and quality
- Balancing privacy and usefulness
- Impact on marginalized communities
- **Case Study**: Evaluating synthetic data in hiring algorithms

Module 7: Validation, Auditing & Risk Mitigation

- Validation metrics for synthetic datasets
- Re-identification risk analysis
- Data quality auditing frameworks
- Reporting standards and documentation
- Post-processing techniques
- **Case Study**: Auditing synthetic healthcare data for clinical trials

Module 8: Future Trends & Applications

- Federated learning and synthetic data
- Use of generative AI (GANs, VAEs)
- Synthetic data in AI model training
- Cross-border data collaboration
- Emerging global standards
- **Case Study**: International research collaboration using synthetic census data

Training Methodology

- Instructor-led sessions with domain experts
- Interactive, hands-on labs and real-time demos

- Guided group discussions and peer learning
- Use of real-world data and open-source tools
- Capstone project: Build and validate your own synthetic dataset

Register as a group from 3 participants for a Discount

Send us an email: info@fineskilltrainingcenter.org or call *+254769199797*

Certification

Upon successful completion of this training, participants will be issued with a globally- recognized certificate.

Tailor-Made Course

We also offer tailor-made courses based on your needs.

Key Notes

a. The participant must be conversant with English.

b. Upon completion of training the participant will be issued with an Authorized Training Certificate

c. Course duration is flexible and the contents can be modified to fit any number of days.

d. The course fee includes facilitation training materials, 2 coffee breaks, buffet lunch and A Certificate upon successful completion of Training.

e. One-year post-training support Consultation and Coaching provided after the course.

f. Payment should be done at least a week before commence of the training, to Fineskill Training Center account, as indicated in the invoice so as to enable us prepare better for you.

Upcoming Scheduled Sessions

Start Date	End Date	Location	Price
23 Nov 2026	27 Nov 2026	Online	$1,000
23 Nov 2026	27 Nov 2026	Physical	$1,500
23 Nov 2026	27 Nov 2026	Physical	$3,000
23 Nov 2026	27 Nov 2026	Physical	$3,000
23 Nov 2026	27 Nov 2026	Physical	$6,000
23 Nov 2026	27 Nov 2026	Physical	$4,500
23 Nov 2026	27 Nov 2026	Physical	$6,000

Start Date	End Date	Location	Price
23 Nov 2026	27 Nov 2026	Physical	$6,000

Ready to enroll or request a customized program? Book online or contact us:

Register Online Now

Email: info@fineskilltrainingcenter.com | Website: www.fineskilltrainingcenter.com