What is Synthetic Data?
Synthetic data is annotated information that computer simulations or algorithms generate as an alternative to real-world data. Synthetic data is created in digital worlds rather than collected from or measured in the real world.

Why Generate Synthetic Data?
You need synthetic data to build high-quality, generalized AI models.

1. Control and Scalability
Create highly scalable training datasets with full control over their characteristics and dimensionality to build accurate AI models.

2. Address the Unexpected
Edge cases, rare events, or dangerous situations often can’t be collected safely or ethically with real-world data.

3. Cost Savings
Overcome the data gap and reduce overall cost of acquiring and labeling data required to train AI models.

4. Privacy
Address privacy issues and reduce bias by generating diverse datasets to represent the real world.

Who Uses Synthetic Data?

- Researchers
- Autonomous Vehicles
- Retail
- Healthcare
- Robotics
- Manufacturing
- Smart Cities
- Computer Vision Engineers
- AI Developers
- Retail
- Healthcare
- Robotics
- Manufacturing
- Smart Cities
- Computer Vision Engineers
- AI Developers

How Do You Generate Synthetic Data?
It is generated by a computer model that consists of algorithms or simulations, or both. Omniverse Replicator is one example. It’s a core extension of the Omniverse platform for generating synthetic data to train vision AI models. It allows users to easily import simulation-ready assets and generate new, diverse, physically accurate datasets.

- 3D Assets
- Scene Generation
- Procedural Scenario Generation
- Batch Generation of Annotated Synthetic Data
- AI Model Training & Optimization
- Inference

Why Generate Synthetic Data Using Omniverse Replicator?
1. Built on open source standards like Universal Scene Description (USD), PhysX, and Material Definition Language (MDL), Replicator easily integrates and connects to existing pipelines with the ability to import/export assets from 3D content creation tools.

2. Generate accurate, photoreal, physically-accurate 3D by varying lighting, object poses, positions and accurately scaled scenes to build perception models or 3D simulations.

3. Bootstrap your computer vision model training process by quickly prototyping with synthetic data across a multitude of domains, before training on real data.

4. Easily simulate custom sensors at scale, simultaneously in both visible and non-visible spectrum in the same scene (e.g. infrared).

5. Generate your training data quickly by scaling up and out on Multi-GPU and Multi-Node GPU systems - on-premise or in the cloud.

Ready to Get Started?
To learn more about how to integrate Omniverse Replicator into your solution, visit:
https://developer.nvidia.com/omniverse/replicator

Gartner, Use Generative AI in Applied Innovation to Drive Business Value, G00788659, May 2023

GARTNER is a registered trademark and service mark of Gartner, Inc. and/or its affiliates in the U.S. and internationally and is used herein with permission. All rights reserved.

© 2023 NVIDIA Corporation and affiliates. All rights reserved. NVIDIA and the NVIDIA logo are trademarks and/or registered trademarks of NVIDIA Corporation and affiliates in the U.S. and/or other countries.