

# Synthetic Data: Advanced Concepts and Applications

GoSkills online course syllabus

**Skill level**

Beginner

**Lessons**

17

**Accredited by**

Verified by GoSkills

**Pre-requisites**

No prior experience needed

**Video duration**

38m

**Estimated study time**

38m 22s

**Instructor**

Michael Galarnyk

## Introduction

---

1

### Unlock the power of synthetic data

As technology continues to evolve, synthetic data has begun to replace real-world data, and its vast benefits are still in the early stages of being uncovered.

## Introduction to Synthetic Data

---

2

### Articulating synthetic data's value

In many domains, collecting, and especially labeling high quality, real-world data can be time consuming, difficult, expensive, dangerous, or even impossible.

3

### Required background knowledge

Generating and training models with synthetic data requires some basic knowledge of statistics and machine learning.

## Generating Synthetic Data

---

4

### Defining synthetic data

Synthetic data is data that is artificially generated rather than collected from the real-world.

5

### Generating synthetic data

Synthetic data has many use cases and it is not all generated in the same way.

6

### Defining domain gaps

A domain gap is the difference between two distinct but related datasets.

## 7 Reducing the domain gap

Reducing the domain gap between real and synthetic data can lead to improved machine learning performance.

## 8 What is generative AI

Generative AI represents a subset of AI algorithms that leverages machine learning, especially deep learning to produce new content.

## 9 Real data errors and solutions

Real datasets can have label errors.

## 10 Synthetic data for edge cases

A lot of machine learning use cases require datasets that are comprehensive, sufficiently large, high quality, diverse, and accurately representative of a problem space it's intended to model.

# Implementing Synthetic Data

---

## 11 Real-World Label Scarcity

Synthetic data.

## 12 Leveraging pre-training and fine-tuning

How do you incorporate synthetic data into your model training strategy?

## 13 Leveraging joint training

Once you have your real and synthetic data, how do you actually use them together to train a model?

## 14 Applying data sampling techniques

Data sampling can be defined as the process of selecting a subset of data for analysis.

## 15 Privacy with synthetic data

While synthetic data doesn't have the same privacy concerns, it is still something that needs to be considered.

## 16 Machine learning with synthetic data

The Machine Learning Development Cycle is a roadmap that guides you in creating and improving machine learning models.

# Conclusion

---

17

## Going further with synthetic data

Thank you for watching this course!

[Go to GoSkills.com](https://www.goskills.com)