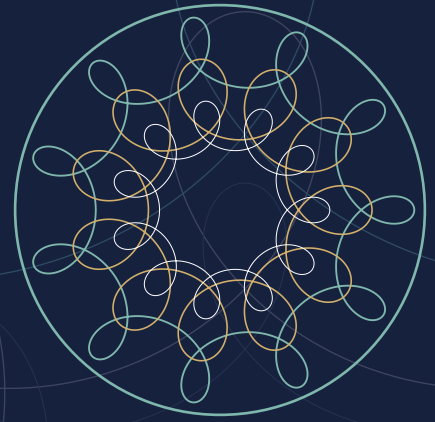


International Institute of Applied AI Technology



Course guide for AI-DGP

Certified AI Data Governance Professional

Govern the data that AI depends on: ownership, quality, lineage, metadata, privacy and the controls that let data be used for training and inference with confidence.

About this guide

Govern the data that AI depends on: ownership, quality, lineage, metadata, privacy and the controls that let data be used for training and inference with confidence.

This guide sets out what the AI-DGP examination covers and how to prepare for it. It assumes you already practice data management, and concentrates on what changes when AI is part of the work.

How to use it

Work through the modules in order. Each one sets out the ideas, lists the topics the exam can test, and ends with an exercise. Do the exercises on your own projects where you can: the exam questions are written around the same situations.

Who it is for	Data stewards, data governance leads, data architects and privacy officers.
Recommended experience	Two years in data management, governance or analytics.
Study time	About 30 hours
Exam	12 multiple-choice questions in 30 minutes, online. Pass mark 9 correct.
Certification	Valid for 3 years, with a printable certificate and a verifiable credential number.

Learning outcomes

When you have completed this program you will be able to:

- Set up ownership and stewardship for data used by AI
- Define and monitor data quality rules suited to model training
- Document lineage from source to model output
- Classify data and control its use for training and inference
- Manage training data, labels and datasets as governed assets
- Use AI itself to speed up governance tasks



Contents

Module 1. Governance foundations for AI

Module 2. Data quality for training and inference

Module 3. Metadata, catalogs and lineage

Module 4. Classification, privacy and permitted use

Module 5. Training data as a governed asset

Module 6. AI for governance

Preparing for the exam



Module 1

Governance foundations for AI

The governance disciplines that already exist, and what AI adds to each.

Topics covered

- Governance frameworks and their knowledge areas
- Policies, standards and procedures
- Owners, stewards and custodians
- Governance bodies and decision rights

Exercise

Assign an owner and steward to five datasets that feed a model and write their responsibilities.

Self-check

Before moving on, check that you can explain each topic above to a colleague in two or three sentences, and say how it would apply on a project you know.

Your notes



Module 2

Data quality for training and inference

Quality measured against the purpose: what is good enough for a report may not be good enough to train on.

Topics covered

- Fitness-for-purpose quality dimensions
- Profiling and rule definition
- Representativeness and class balance
- Quality monitoring in production

Exercise

Write ten quality rules for a customer dataset that will train a churn model.

Self-check

Before moving on, check that you can explain each topic above to a colleague in two or three sentences, and say how it would apply on a project you know.

Your notes



Module 3

Metadata, catalogs and lineage

Being able to say where data came from and what was done to it, all the way to a model's decision.

Topics covered

- Business, technical and operational metadata
- Data catalogs and glossaries
- End-to-end lineage including feature pipelines
- Datasheets for datasets and model cards

Exercise

Draw the lineage from source systems to one model output and note each transformation.

Self-check

Before moving on, check that you can explain each topic above to a colleague in two or three sentences, and say how it would apply on a project you know.

Your notes



Module 4

Classification, privacy and permitted use

Knowing what may be used for which purpose, and enforcing it.

Topics covered

- Sensitivity classification schemes
- Personal data, consent and purpose limitation
- De-identification and its limits
- Access controls for training environments

Exercise

Classify the fields in a dataset and mark which may be used for training without further approval.

Self-check

Before moving on, check that you can explain each topic above to a colleague in two or three sentences, and say how it would apply on a project you know.

Your notes



Module 5

Training data as a governed asset

Datasets, labels and versions recorded so that a model can be reproduced, audited and retrained.

Topics covered

- Dataset versioning
- Labeling standards and inter-rater agreement
- Synthetic data and its governance
- Retention of training data

Exercise

Write a labeling guideline for one classification task, including how disagreements are resolved.

Self-check

Before moving on, check that you can explain each topic above to a colleague in two or three sentences, and say how it would apply on a project you know.

Your notes



Module 6

AI for governance

Using AI to discover sensitive data, suggest glossary terms, detect quality issues and draft documentation.

Topics covered

- Automated sensitive-data discovery
- Suggested metadata and glossary mapping
- Anomaly detection for quality
- Keeping stewards accountable for AI suggestions

Exercise

Pick one governance task you do by hand and design how AI would assist it, with the steward's check.

Self-check

Before moving on, check that you can explain each topic above to a colleague in two or three sentences, and say how it would apply on a project you know.

Your notes



Preparing for the exam

The AI-DGP exam has 12 multiple-choice questions drawn at random from every module. You have 30 minutes and need 9 correct answers to pass.

Before you start

- Finish every module's exercise. Most questions describe a situation and ask what a practitioner should do.
- Re-read the module summaries: they state the main idea each question is built on.
- Set aside uninterrupted time. The timer keeps running if you leave the page.
- Have this guide to hand. The exam is open book, but you will not have time to look up every answer.

Taking the exam

Log in to your account on the Institute's website and choose Take the exam. Check that your name is shown exactly as it should appear on your certificate. Your result is shown as soon as you submit, and if you pass, your certificate and credential number are issued straight away.

If you don't pass, you can try again after 24 hours. Each attempt draws a different set of questions.

