JUPITER PRIME

Consulting • Training • Advisory

جوبيتر برايم للاستشارات والتدريب

COURSE BROCHURE

Advanced Data Governance, Privacy and Integrity in AI Training Course

25 July – 5 August 2027
Kuala Lumpur

Advanced Data Governance, Privacy and Integrity in AI Training Course

Ref: 14_361 **Date:** 25 July – 5 August 2027 **Location:** Kuala Lumpur **Fees:** 8500 GBP

Data Governance, Privacy and Integrity in AI is a 2-week advanced training course for data governance leads, privacy officers, data owners and AI risk managers, who leave with an AI data governance framework covering the datasets behind their models. Every model inherits the quality, consent status and bias of its training data, so weak data controls now surface as flawed predictions and privacy breaches. Jupiter Prime Training runs this course on data governance, privacy and integrity in AI for the custodians of that data.

Why Data Governance, Privacy and Integrity in AI Matters Now

Three forces shape data governance, privacy and integrity in AI today: models trained on vast, mixed-origin datasets whose provenance is often unclear, data protection expectations that now extend to inference and model outputs, and data subjects and auditors who ask how personal records were used to train a system. Organisations that handle this well can trace every training set, justify its lawful use and prove its quality; those that handle it badly retrain models on tainted data or face complaints they cannot answer. Examples come from healthcare records and financial services transactions. The course rests on ISO/IEC 42001 for AI management systems and the NIST AI Risk Management Framework AI RMF.

Who Gains Most from Data Governance, Privacy and Integrity in AI

- Data governance leads who set ownership, stewardship and policy for enterprise datasets
- Privacy officers responsible for lawful basis, consent records and impact assessments
- Data owners and stewards accountable for quality, lineage and access to specific domains
- AI risk and programme managers who approve datasets for model training and release
- Information security managers who protect data pipelines, feature stores and model artefacts

The greatest gain comes when a privacy office, a data governance office and the AI product owners of one organisation attend together and agree shared dataset controls.

Data Governance, Privacy and Integrity in AI Capabilities

Participants learn to treat each training and inference dataset as a governed asset with an owner, a lawful purpose, a quality profile and a traceable history.

- Design a data governance operating model that assigns ownership of AI training data — the chief data officer
- Assess data quality, lineage and provenance evidence before model training — dataset approval boards
- Apply privacy by design, anonymisation and synthetic data techniques to AI datasets — privacy officers
- Calculate re-identification and bias exposure in candidate datasets — model risk reviewers
- Specify security controls for AI data pipelines and feature stores — information security leads
- Build an AI data governance framework with policies, roles and checkpoints — the executive data committee

Key Data Governance, Privacy and Integrity in AI Terms

- Data Lineage — a record of where data originated, how it was transformed and which systems and models consumed it.
- Data Provenance — documented evidence of the source, collection method and rights attached to a dataset used for AI.
- Privacy by Design — embedding data protection safeguards into AI systems from the first design decision, not after release.
- Pseudonymisation — replacing direct identifiers with codes so records cannot be linked to a person without separate key data.
- Data Protection Impact Assessment DPIA — a structured evaluation of privacy risks from a processing activity, with mitigating measures.
- Synthetic Data — artificially generated records that mirror the statistical properties of real data without copying actual individuals.

Data Governance, Privacy and Integrity in AI Day by Day

Day 1: Pressures on Data Behind AI

- AI Data Supply Chain Mapping — Sources, Brokers and Internal Systems
- Data Protection Expectations for Model Training — Purpose and Proportionality
- Integrity Failures in Deployed Models — Poisoned, Stale and Mislabelled Data
- Stakeholder Map for AI Datasets — Data Subjects, Regulators and Auditors
- Data Maturity Self-Assessment — Baseline for AI Readiness

Day 2: Data Governance Operating Models

- Centralised, Federated and Hybrid Governance Models — Decision Rights Choices
- Data Owner, Steward and Custodian Roles — Accountability for Training Data
- Data Governance Council Charter — Escalation and Approval Authority
- Data Policies and Standards Hierarchy — Enforceable Rules for AI Use
- Business Glossary and Metadata Catalogue — Shared Meaning Across Teams

Day 3: Data Quality and Lineage

- Data Quality Dimensions — Accuracy, Completeness, Timeliness and Consistency Targets
- Data Profiling and Validation Rules — Gatekeeping Before Model Training
- Automated Lineage Capture — Tracing Features Back to Source Systems
- Data Quality Scorecards — Remediation Priorities for Data Owners
- Master and Reference Data Alignment — Consistent Entities Across Models

Day 4: Training Data Provenance and Rights

- Datasheets for Datasets — Documenting Origin, Composition and Limits
- Licensing and Usage Rights for Third-Party Data — Acquisition Decisions
- Web-Scraped and Open Data Risks — Copyright and Consent Exposure
- Labelling Supply Chains — Annotation Quality and Worker Conditions
- Dataset Versioning and Retention Rules — Reproducibility of Training Runs

Day 5: Privacy by Design for AI

- Data Minimisation in Feature Selection — Necessity Tests for Attributes
- Purpose Limitation and Secondary Use — Reusing Records for Model Training
- Privacy-Enhancing Technologies — Federated Learning and Differential Privacy
- Model Inversion and Membership Inference Risks — Output Privacy Controls
- Privacy Requirements in AI Design Specifications — Engineering Sign-Off

Day 6: Anonymisation and Synthetic Data

- Anonymisation Versus Pseudonymisation — Legal Status of Transformed Data
- K-Anonymity, L-Diversity and Generalisation Techniques — Re-Identification Testing
- Synthetic Data Generation Methods — Fidelity and Privacy Trade-Offs
- Utility Testing of De-Identified Datasets — Model Accuracy Impact
- Release Decision Matrix for Shared Datasets — Internal, Partner or Public

Day 7: Consent, Lawful Basis and Data Protection Impact Assessments

- Lawful Basis Selection for AI Processing — Consent, Contract or Legitimate Interest
- Consent Records and Withdrawal Handling — Effects on Trained Models
- DPIA Method for AI Systems — Risk Scoring and Mitigation Plans
- Data Subject Rights Requests — Access, Erasure and Explanation Responses
- Cross-Border Data Transfers for Model Training — Transfer Safeguard Choices

Day 8: Bias and Representativeness in Datasets

- Sampling Bias and Coverage Gaps — Representativeness Testing
- Label Bias and Historical Discrimination — Ground Truth Review
- Proxy Variables for Protected Characteristics — Feature Exclusion Decisions
- Rebalancing, Reweighting and Augmentation — Dataset Correction Options
- Bias Documentation in Model Cards — Disclosure to Reviewers

Day 9: Security and Integrity of AI Pipelines

- Data Poisoning and Adversarial Input Threats — Pipeline Threat Models
- Access Control and Encryption for Feature Stores — Least-Privilege Design
- Integrity Checks With Hashing and Signed Datasets — Tamper Detection
- ISO/IEC 42001 Data Controls — AI Management System Evidence
- NIST AI RMF Map and Measure Functions — Data Risk Register Entries

Day 10: Practising the AI Data Governance Framework

- Operating Model and Role Assignments — Applying the Chosen Governance Model
- Dataset Approval Checklist — Quality, Provenance and Lawful Basis Gates
- Privacy and Bias Control Library — Linking Techniques to Risk Levels
- Monitoring Metrics and Audit Trail Design — Ongoing Assurance
- AI Data Governance Framework — Presentation to the Executive Data Committee

Applying Data Governance, Privacy and Integrity in AI After the Course

- Complete a datasheet and lineage trace for one dataset feeding a live model, and flag gaps in provenance or lawful basis to its data owner
- Run a DPIA for one planned AI use case, including re-identification testing of the proposed training extract

Data Governance, Privacy and Integrity in AI Questions FAQ

What is data governance, privacy and integrity in AI and why does it matter now?

Data governance, privacy and integrity in AI is the control of who owns, may use and must protect the data that trains and feeds AI systems. It matters now because models amplify any flaw, bias or unlawful use hidden in their datasets.

Which roles gain most from data governance, privacy and integrity in AI training?

Data governance leads, privacy officers, data owners and AI risk managers gain most. They approve datasets, answer data subject requests and evidence lawful use, so they need shared methods for quality, provenance and privacy testing.

How does advanced data governance, privacy and integrity in AI training differ from an AI governance course?

It concentrates on the datasets rather than on board oversight of AI as a whole. You work on lineage, provenance, anonymisation, consent and bias testing, at a depth suited to specialists who already run data or privacy functions.

What do you bring back from data governance, privacy and integrity in AI training?

You bring back an AI data governance framework. It sets the operating model, dataset approval gates, privacy and bias controls, pipeline security measures and monitoring metrics your executive data committee can adopt.

Data Governance, Privacy and Integrity in AI: The Way Forward

The AI data governance framework first informs which datasets may enter model training and under what conditions. Data committees, privacy officers and AI product owners use it to approve, reject or remediate datasets. After the first approval cycle, revisit the provenance and re-identification findings, since new external data sources often expose gaps the initial controls did not anticipate.