Data Governance and Ethics

Exam revision brief · D4B Joint Professional Master's, University of Bologna · v3.3

www.houghton.consulting

Friday 7 August 2026 · 17:00 IST / 18:00 CET · 90 minutes · online, individual

30 MCQs × 2 marks (60) + 5 short responses × 8 marks (40) = 100 marks. Unsubmitted attempts auto-submit at the deadline.

Built from your twelve session decks and the Week 4 DGI (Data Governance Institute) companion guide, read in full. Where the slides differ from the published standard, this brief follows the slides — that is what is being marked.

Time budget: 35 min on the MCQs (~70 s each), 45 min on the written answers (~9 min each), 10 min to review and submit manually. Never leave an MCQ blank.

The Safe Exam Browser may be required — not confirmed yet. Watch for Mark's "more details to follow" email. Install and test it in advance regardless, and have your D4B login details to hand on paper either way — they do not autofill in the Bologna exam browser.

Version 3.3 · 3 August 2026 · Prepared by Joe Houghton, Houghton Consulting. What has changed since version 1 is on the next page; the full version history is on the last page.

Using version 1? What has changed

The practice questions are the substantial change. Version 1 asked a good deal about where something appeared — which week defined a term, which wording a slide used, how many items a list held. Those test recall of the deck rather than understanding of the subject. Every question has been rewritten to put you in a situation and ask you to diagnose it, tell two related concepts apart, reason about a consequence, or apply a framework to a dilemma.

Answers now carry a fuller explanation. Each names the misconception the wrong options are built on and says where the tempting choice breaks down, rather than giving one line of rationale. That accounts for most of the nine extra pages.

Every acronym is written out on first use in each section, so nothing depends on remembering what the initials stood for three pages earlier.

The Safe Exam Browser guidance was softened — it may be required, but that is not yet confirmed for this exam.

Unchanged: the twelve-session map, the counts table, the traps table, the definitions, the frameworks section, the General Data Protection Regulation summary, the ethics and artificial intelligence material, and the five short-response scaffolds.

The companion trainer at https://d4b-governance-revision-trainer.vercel.app shares this version number from v3.2 onwards, and draws a different twenty questions from a pool of 100 on each visit.

1. The twelve sessions and where the marks are

Titles are from the title slide of each deck — the "What we'll cover each week" list drifts between decks, so ignore it.

Wk	Title	Most examinable content
1	Introduction to Data Governance	TechTarget definition; governance-as-government analogy; the 7 named DG (data governance) frameworks; "data exhaust"; doubling every 18–24 months
2	Principles of Data Management and Governance	Governance vs management; DMBOK's 8-element DG scope; 4 data management principles; provenance vs lineage; MDM (master data management) and the "golden copy"
3	Data Risks, Integrity and Security	5-step risk process; 4 risk register headings; 12 identification methods; 5 integrity/security challenges; 6 general + 3 application controls
4	Data Governance Frameworks	Principles → Policies → Processes → Standards; 4 GAIP (Generally Accepted Information Principles) principles; CDO's 6 functions; 6 DMBOK steward types; DGI's 10 components
5	DG Operating Model 2, Architecture and Metadata	The 4 execution scenarios; escalation path; Ladley on architecture and metadata; data fabric (IBM)
6	Data Regulatory Compliance, Privacy and Data Protection	GDPR (General Data Protection Regulation): 7 principles, 6 lawful bases, consent, controller vs processor, Arts. 28–32, Art. 9, Art. 25, Chapter V transfers
7	Implementing Data Governance	Ladley's 8 phases; RACI (Responsible, Accountable, Consult, Inform); 3 maturity levels (Adler); 4 challenge quadrants; 4 objections and responses; monetisation
8	Ethical Concepts and Frameworks	3 branches of ethics; 4 normative families; Kant's two imperatives; Rawls's veil of ignorance; Asimov; the 7 information-age values
9	Ethics and AI	3 direct + 5 broader concerns; historical bias and feedback loops; interpretability; the case studies
10	Privacy, Analytics and Ethics	UDHR (Universal Declaration of Human Rights) Art. 12, ECHR (European Convention on Human Rights) Art. 8, Charter Art. 8; 4 causes of bias; fairness; privacy-preserving AI; AIF360 (AI Fairness 360)
11	Governance of AI and Advanced Analytics	AI Act tiers and timeline; Recital 71 and Arts. 13/15/22/35; FAT/ML's 5 principles; direct vs indirect auditing
12	Future Trends	4 trends; UN (United Nations) advisory group's 3 objectives; CRISP-DM (Cross-Industry Standard Process for Data Mining) adaptation; expanded CDO role

2. The counts MCQs test

This module runs on numbered lists and multiple-choice questions feed on them. If you learn nothing else by rote, learn this.

List	N	The items
DMBOK (Data Management Body of Knowledge) scope of a DG (data governance) programme	8	Strategy · Policy · Standards and quality · Oversight · Compliance · Issue management · Data management projects · Data asset valuation
Data management principles (as taught)	4	Data is valuable · DM (data management) requirements are business requirements · DM depends on diverse skills · DM requires leadership commitment
Data quality dimensions (as taught)	5	Accuracy · completeness · timeliness · consistency · relevance
Risk management process	5	Identification · Quantification · Assessment · Negation or mitigation · Reporting
Risk register headings	4	Financial · Reputational · Operational · Legal/Compliance
Risk identification methods	12	SWOT (strengths, weaknesses, opportunities, threats) · PESTLE (political, economic, social, technological, legal, environmental) · external review · internal review · DG programme · data audits · "white" hackers · regulatory attention · best practice review · brainstorming · scenario planning · disaster management planning
Integrity/security challenges	5	Human error · System failures · Cybersecurity attacks · Inconsistent data formats · Lack of proper data governance
Security controls	6+3	General: software, hardware, computer operations, data security, system development, administrative. Application: input, processing, output
GAIP (Generally Accepted Information Principles) data principles	4	Data content as assets · Data assets as having real value · Data as presenting risk · Asset value as determined by data quality
DMBOK steward types	6	Executive · Enterprise · Business · Data Owner · Technical · Coordinating
Chief Data Officer functions	6	Data strategy · aligning requirements with resources · standards and policies · advice to the business · evangelising · oversight of analytics and BI (business intelligence)
DGI (Data Governance Institute) framework components	10	Mission and Value · Beneficiaries · Data Products · Controls · Accountabilities · Decision Rights · Policy and Rules · Processes, Tools and Communication · Data Work Programme · Participants
Execution scenarios (Wk 5)	4	Decentralised single BU (business unit) · Decentralised multi BU · Centralised governance · Centralised governance with decentralised execution
GDPR (General Data Protection Regulation) principles (Art. 5)	7	Lawfulness, fairness and transparency · Purpose limitation · Data minimisation · Accuracy · Storage limitation · Integrity and confidentiality · Accountability
GDPR lawful bases (Art. 6)	6	Consent · Contract · Legal obligation · Vital interests · Public interest/official authority · Legitimate interests
Ladley implementation phases	8	Scope and initiation · Assess capability · Vision · Align data and business value · Functional design · Framework design · Roadmap · Roll-out and sustain
DG maturity levels (Adler 2016)	3	Reactive · Preemptive · Proactive — not a five-level CMM

List	N	The items
		(Capability Maturity Model)
Branches of ethics	3	Metaethics · Normative · Applied/Comparative
Normative families (as taught)	4	Virtue · Deontological · Consequentialist/Utilitarian · Social justice
Information-age values	7	Justice/equity · Freedom/agency · Care and compassion · Participation · Sharing · Sustainability · Responsibility
Sources of ethical concern in models	3+5	Direct: bias, transparency/interpretability, privacy. Broader: copyright, employment, energy, unequal access, education
Causes of bias in AI/ML models	4	Historical bias in datasets · Technical/statistical · Algorithmic objectives · Proxy attributes for sensitive attributes
FAT/ML (Fairness, Accountability and Transparency in Machine Learning) principles	5	Responsibility · Explainability · Accuracy · Auditability · Fairness
AI (artificial intelligence) Act risk tiers	4	Unacceptable · High · Transparency · Minimal or no risk

3. Traps: where the slides differ from the textbook

Answer to the slides. Each of these is a place where a well-read student can talk themselves into the wrong answer.

Topic	What the slides say	What you might wrongly assume
The third leg of CIA (confidentiality, integrity, availability)	Week 3 says confidentiality, integrity and accessibility — twice	Availability. It appears only in the Art. 32 GDPR (General Data Protection Regulation) list
Integrity and security	"Data integrity is a desired result of data security"	The reverse. Distractors flip the direction
DGI (Data Governance Institute) components	The value-driven v2.0 set: Mission and Value, Beneficiaries, Data Products…	The classic 2006 set: Mission and Vision, Data Rules and Definitions, Data Stakeholders, Data Governance Office, Data Stewards. Only Decision Rights, Accountabilities and Controls appear in both
Data Owner	A type of Business Data Steward, with approval authority	A separate top-level role alongside steward and custodian
Maturity levels	Three: reactive, preemptive, proactive	A five-level CMMI-style model. Do not import one
Rawls	"Social justice based ethics"	Contractarianism. That word never appears in the module
Asimov's laws	Numbered 1–4, humanity-level law placed fourth	The Zeroth Law taking precedence over the first
Data quality dimensions	Five, including relevance	The DAMA (Data Management Association) UK six, including validity and uniqueness
GDPR fine tiers	One figure: up to €20m or 4% of global turnover	The two-tier structure. The €10m/2% tier is never taught
Data mesh	Never taught — Dehghani is not cited anywhere	That it is examinable. The only architecture pattern named is data fabric (IBM)
DIKW	The letters never appear in the slide text	That it is a named framework here. It survives only as an unlabelled graphic

Not in the slides at all — do not spend revision time here

Data mesh and Dehghani. Data warehouse, lake, lakehouse. Federated governance as a named term.

GDPR breach notification and the 72-hour rule (Arts. 33–34). DPO (Data Protection Officer) designation and tasks (Arts. 37–39). The enumerated data subject rights — that slide is an image.

ISO/IEC (International Organization for Standardization / International Electrotechnical Commission) 42001 and 23894, IEEE (Institute of Electrical and Electronics Engineers) 7000, the NIST (National Institute of Standards and Technology) AI (artificial intelligence) RMF (Risk Management Framework), COBIT (Control Objectives for Information and Related Technologies). Floridi and AI4People. The trolley problem. k-anonymity and pseudonymisation as techniques.

If one of these turns up in an MCQ it came from a lab paper, not a deck — reason from first principles.

4. Definitions worth quoting

Term	As the module gives it
Data governance (TechTarget — the headline definition)	The process of managing the availability, usability, integrity and security of the data in enterprise systems, based on internal data standards and policies that also control data usage.
Data governance (DAMA-DMBOK)	Exercise of authority and control — planning, monitoring and enforcement — over the management of data assets.
Data governance (DGI (Data Governance Institute))	A logical structure for classifying, organizing, and communicating complex activities involved in making decisions about and taking action on enterprise data.
Governance vs management	Governance is "the management of data management" and envisages a separation between those governing and those managing. Governance sets the who, when and where; management does the daily actions. Business strategy ← data strategy → IT strategy.
Big G and little g (DGI)	"Little g" adds value to data itself — products, catalogues, definitions, metadata, controls. "Big G" creates clarity — decision rights, accountabilities, policies, guardrails.
Where governance lives (DGI)	"Management lives in the boxes of the organisation chart… governance lives in the white space between the boxes, where convened groups make decisions that flow across the chart."
Data risk	The potential for business loss due to poor data governance among other factors. Level of risk = probability of event × cost of event.
Data integrity	The assurance that information is accurate, consistent and reliable; it examines the consistency of methods used to create, retain, preserve, distribute and track information over the lifecycle.
Data security	The measures protecting data against unauthorised access, disclosure or corruption.
Negation vs mitigation	Negation avoids the risk altogether (not always possible). Mitigation reduces impact — by controls, alternative approaches, insurance, or business continuity planning.
Data architecture (Ladley)	A representation of the data management environment, its components and their interactions, interrelating framework, people, processes, projects, policies, technologies and procedures.
Metadata (Ladley)	Any tool that supports "data about our data" — data modelling, analytical models, algorithms, glossary, rules, documentation of metrics.
Provenance vs lineage	Provenance is origin — who created the data and why. Lineage is the pathway from point of origin to point of usage.
Data steward	Someone who manages property on behalf of others, representing the interests of all stakeholders. "The best data stewards are often found, not made."
Steward vs custodian (DGI)	Stewards sit in business and compliance functions — the eyes and ears of the programme. Custodians are their technical counterparts, applying controls and configurations.
Ethics	The branch of philosophy that involves systematising, defending and recommending concepts of right and wrong behaviour. It operates independently from legal systems but has a role in how laws are made.
Algorithm	A set of precise instructions or rules regarding actions to be taken in solving a predefined problem. An algorithmic system comprises one or more algorithms used in software to collect and analyse data and draw conclusions.

Term	As the module gives it
Bias (Olteanu et al. 2019)	A systematic distortion in the sampled data that compromises its representativeness.
Fairness (Duke Ch. 4)	The absence of any form of prejudice or favouritism toward an individual or group based on their intrinsic or acquired traits in the context of decision-making. To assess fairness we need transparency; to enforce it we need accountability.
Privacy	The right of a person to enjoy his own presence by himself and decide his boundaries; personal information is a form of personal property; the individual has a right to limit sharing.

5. Frameworks in one place

5.1 The four-level hierarchy (DMBOK (Data Management Body of Knowledge) 2017)

Principles → Policies → Processes and Procedures → Rules and Standards. Principles are statements of philosophy; policies turn them into fundamental rules and should be few, brief and direct; procedures prescribe how a process is done; standards are mandatory and codify a decision once so it need not be remade. Standards are drafted by data management professionals but reviewed, approved and adopted by the Data Governance Council.

Verb trap. Policies are communicated, monitored, enforced and periodically re-evaluated. Standards are communicated, monitored, reviewed and updated — "enforced" is absent.

5.2 The DGI (Data Governance Institute) framework — how the ten components group

Group	Components and purpose
Justify (1–2)	Mission and Value; Beneficiaries. Value comes three ways: adding value to products, services, processes, capabilities and assets; reducing cost, complexity, confusion and delay; reducing risk. A programme without named beneficiaries is a cost centre waiting to be cut.
Produce (3–7)	Data Products; Controls (preventative, detective, corrective); Accountabilities; Decision Rights; Policy and Rules. Whether to comply is an executive decision; how to comply needs multiple stakeholders in the room.
How (8)	Processes, Tools and Communication — twelve documented procedures, scored as documented, partially documented, or tribal knowledge.
Organise (9)	Data Work Programme — a portfolio of workstreams, each with focus, scope, goals, metrics, lifecycle, funding model and responsibilities. Value stated A-to-E: because governance intends to do A, we anticipate output B, with impact C for the beneficiary; otherwise they risk D, resulting in E.
People (10)	Participants — the Data Governance Office, decision bodies, and stewards plus custodians.

Worth quoting: the guide's claim that the first casualty of weak governance is nearly always Component 6 — nobody can say who decided the thing everyone is complying with — and the second is Component 2, because the programme never named who it was for.

5.3 The four execution scenarios (Week 5)

Scenario	Org size	Shared master data	Master data lifespan	Named weakness
1 · Decentralised, single BU (business unit)	Small	No	Shorter	Huge inconsistencies in data
2 · Decentralised, multiple BUs	Small–medium	Yes	Shorter	Duplicates; meaningless reporting
3 · Centralised governance	Large–medium	Yes	Longer	Delays; needs a formal, larger DG (data governance) organisation
4 · Centralised governance, decentralised execution	Large–medium	Yes	Longer	Needs proper controls; shared responsibility

Scenarios 3 and 4 are the two flagged as carrying potentially higher risks from data protection breaches. Scenario 1 is the only one with no shared master data.

5.4 Ladley's 8-phase implementation (Week 7)

Phase	The point of it
1 Scope and initiation	Decide span and depth. Set up a steering group led by an executive sponsor so it is not seen as just an IT project
2 Assess capability	Baseline survey. Does the organisation accept that data is an asset, assets need managing, management requires governance? Locate resistance
3 Establish vision	The elevator pitch. A vision is where you want to be; the mission is how to get there
4 Align data and business value	Build the business case. Show how DG contributes to the business, not to its components like data quality
5 Functional design	The what and how — principles, policies, processes. Principles provide the foundation of effective policies
6 Framework design	The who — organisation framework rather than org chart, because DG works best as a virtual organisation with minimal formal roles. RACI (Responsible, Accountable, Consult, Inform): Responsible, Accountable, Consult, Inform
7 Roadmap	Organisational change management = planning + executing + sustaining. "Start! Get busy with some DG activities"
8 Roll-out and sustain	The final step in a lifecycle, not the end of a linear process. The cycle may begin again at a higher maturity level

6. GDPR as taught

Item	As the slides give it
Basics	In effect 25 May 2018, replacing the Data Protection Directive 1995. A regulation with direct effect, applying to organisations handling personal data of EU residents regardless of physical location. Ireland's Data Protection Act 2018 sits alongside. 173 recitals; 99 articles in 11 chapters
Personal data	Any information belonging to an identified or identifiable natural person, directly or indirectly — name, identification number, location data, online identifier, or physical, physiological, genetic, mental, economic, cultural or social factors. Pseudonymised data is in scope; anonymised data is not
The seven principles (Art. 5)	Lawfulness, fairness and transparency · Purpose limitation · Data minimisation · Accuracy · Storage limitation · Integrity and confidentiality · Accountability
Consent (Art. 4(11))	Freely given, specific, informed and unambiguous; by statement or clear affirmative action. Silence, pre-ticked boxes and inactivity do not constitute consent (Recital 32). Withdrawal must be as easy as giving, does not affect prior lawfulness, and the right must be notified beforehand
Lawful bases (Art. 6)	Consent · contract · legal obligation · vital interests · public interest or official authority · legitimate interests
Special categories (Art. 9)	Racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, genetic data, biometric data for unique identification, health, sex life or sexual orientation — processing shall be prohibited
Controller vs processor	Controller determines the purposes and means, alone or jointly — defines the how and why. Processor processes on behalf of the controller and must maintain an audit trail of processing activities
Processor contract (Art. 28)	Documented instructions only; confidentiality; Art. 32 measures; conditions on sub-processors; assist the controller under Chapter III and Arts. 32–36; delete or return data; make everything available to demonstrate compliance
Security (Art. 32)	Pseudonymisation and encryption; ongoing confidentiality, integrity, availability and resilience; ability to restore availability in a timely manner after an incident; a process for regularly testing, assessing and evaluating effectiveness
By design and default (Art. 25)	Appropriate technical and organisational measures such as pseudonymisation to implement principles such as data minimisation; by default only data necessary for each specific purpose; by default not accessible to an indefinite number of persons without the individual's intervention
Transfers (Chapter V)	Art. 44 general principle · 45 adequacy · 46 appropriate safeguards · 47 binding corporate rules · 49 derogations. A tiered approach: if you cannot use the higher tier a lower one applies, with correspondingly increased responsibility on controllers. The standard is "essentially equivalent" (EDPB (European Data Protection Board) 2020)
Fines	Up to €20 million or 4% of global turnover

The short-response angle on transfers

The tiered logic is the most answerable applied question in Week 6. Adequacy first (Art. 45) — the transfer is then treated as if it were within the EU, assessed on rule of law and rights, an effective independent supervisory authority, and international commitments. Absent adequacy, appropriate safeguards (Art. 46): binding corporate rules, standard data protection clauses, or a legally binding instrument between public authorities, needing no authorisation; contractual clauses and administrative arrangements need supervisory authority authorisation. Absent both, derogations (Art. 49).

7. Ethics and AI

7.1 The four normative families

Family	Core claim	Applied to a data governance dilemma
Virtue	What does it mean to live a good life? Good behaviour is the golden mean — a virtue sits halfway between two vices (courage between cowardice and rashness). Plato, Aristotle, Aquinas, Confucius	There is a moral dimension to a data risk assessment: be informed, take a measured view of risks and benefits, have the courage to decide correctly. Criticised for ignoring differing views of virtue — cultural or moral relativism
Deontological	Moral duties come from universal laws. Kant: act only on a maxim you could will to be universal law; treat humanity always as an end, never simply as a means. Intent, not outcome	Good actions justify whatever happens. A deontologist keeps promises and follows the law. The challenge: would you do something wrong for a positive end?
Consequentialist / utilitarian	An action is right if it produces a greater quantity of good than any alternative — the greatest good for the greatest number. Consequences, not intent	The ends justify the means. But who decides what the greater good is? Is it fair to sacrifice the minority? Should data privacy be sacrificed for the greater good of progress?
Social justice	Rawls's veil of ignorance: what institutions would we want if we could not foretell our own position? Welfare is judged by the welfare of the weakest	Ask how a practice affects the least advantaged — by status, wealth or health. Those not disadvantaged should sacrifice some happiness to make the distribution just

The four levels: personal morality → group or professional ethics → societal morality → law. Note the parallel the slides draw — moral acts conform to what an individual believes is right; legal acts conform to the law or at least do not contravene it.

7.2 Bias and fairness

- **Historical human bias.** Humans remember trauma over positives, recall insults over praise and react more strongly to negative stimuli. Those experiences are reflected in training data, so models reproduce the same wrong judgements people make.
- **The feedback loop — likely short-response material.** Biases are reinforced without the user's knowledge. Low-income people targeted with high-interest credit card ads click them, so the algorithm never accumulates counter-factual suggestions — lower-interest options they might be eligible for and prefer.
- **Fairness trades off against accuracy,** and produces paradoxes: information on sensitive variables needs to be included in order to avoid discrimination.
- **"Equally right, unequally wrong."** ProPublica's 2016 COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) analysis found similar rates of correctly predicted recidivism but a much higher false positive rate for black defendants. The module's compressed statement of the competing-fairness-metrics problem.

7.3 The case studies — know four cold

Case	What happened	The governance failure
Togo (positive)	AI plus freely available satellite imagery identified impoverished villagers where survey infrastructure was absent or stale, allowing cash payments to be targeted	The counter-example — AI substituting for missing statistical infrastructure. Cite it whenever a question invites a balanced answer
Optum / Impact Pro	A tool meant to identify patients with gaps in care used prior spending to predict	Proxy-variable failure. Intention was good; spending is a biased proxy for need. Social

Case	What happened	The governance failure
	future need. African-Americans were underrepresented in the spending data, so it predicted less need and they lost preventative intervention	inequalities in the data were reproduced and exacerbated in operation
Amazon recruiting	Trained on ten years of résumés, predominantly from white males, in a workforce 60% male with men in 74% of managerial posts. It penalised any résumé containing "women's". Discontinued	Training data as a record of past discrimination. Note the tool was withdrawn — detection working
COMPAS / Rodríguez	Parole denied despite a decade without infractions, a job and accommodation waiting, four and a half years training service dogs and eleven volunteering. COMPAS ranked him high risk	Neither he nor the board knew how the score was calculated — Northpointe treats it as a trade secret. Opacity plus trade-secret protection removes contestability from the decision-maker as well as the subject

Also usable: Latanya Sweeney on arrest-record ads returned for African-American names; the Princeton 2.2-million-word association study; facial recognition training sets over 75% male and over 80% white, 99% accurate on white men; Joy Buolamwini, better recognised in a white mask — "default is not neutral".

8. AI governance

AI Act tier	Examples given	Consequence
Unacceptable risk	Social scoring, individual predictive policing	Prohibited
High risk	Recruitment, medical devices	Permitted, subject to requirements to make them safe and trustworthy
Transparency risk	Impersonation (chatbots), synthetic content	Permitted, with disclosure obligations such as labelling and watermarking
Minimal or no risk	—	Permitted with no restrictions

Currency warning — the timeline in your slides has been superseded

The slides give: 8/24 in force · 2/25 general provisions and prohibitions · 8/25 general-purpose AI and governance · 8/26 high-risk, transparency and innovation measures · 8/27 high-risk embedded in regulated products.

Regulation (EU) 2026/1744, the Digital Omnibus on AI, was adopted on 8 July 2026 and entered into force on 27 July 2026 — three days before this brief. Annex III standalone high-risk obligations move from 2 August 2026 to 2 December 2027; product-embedded high-risk moves from 2 August 2027 to 2 August 2028; regulatory sandboxes move to 2 August 2027. The deferral is an unconditional calendar date: the conditionality on harmonised standards in the Commission's original proposal was deleted by the co-legislators.

Tactically: for an MCQ, answer the slides — the exam is marked against taught material. For a short response, give the taught position and add one sentence noting the July 2026 amendment. That reads as currency, not contradiction.

8.1 GDPR (General Data Protection Regulation) as an algorithmic accountability instrument

- **Recital 71.** Controllers should use appropriate mathematical or statistical procedures for profiling, correct inaccuracies and minimise error risk, secure data proportionate to the risks, and prevent discriminatory effects on the basis of racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status, or sexual orientation.
- **Art. 22** — a right to contest an individual algorithmic decision. **Art. 13** — notice of solely automated decision-making. **Art. 15** — access to "meaningful information about the logic involved".
- **Art. 35 and the AIA argument.** Some commentators suggest the DPIA (Data Protection Impact Assessment) duty implies a duty to conduct an Algorithmic Impact Assessment. Kaminski and Malgieri (2019) argue the DPIA "serves as a central connection between its two approaches to regulating algorithms: individual rights and systemic governance". Learn this citation — it is the module's one substantial academic claim on AI accountability.

8.2 FAT/ML (Fairness, Accountability and Transparency in Machine Learning) and auditing

Principle	What it requires
Responsibility	Externally visible avenues of redress, and a designated internal role responsible for timely remedy. Steps include a sunset plan for the system
Explainability	Decisions and the data driving them explained to end-users and stakeholders in non-technical terms

Principle	What it requires
Accuracy	Identify, log and articulate sources of error and uncertainty so worst-case implications inform mitigation. Steps include sensitivity analysis and publicly reporting the error rate on a random sample
Auditability	Enable third parties to probe, understand and review behaviour, through documentation, suitable APIs and permissive terms of use
Fairness	Decisions must not create discriminatory or unjust impacts across demographics. Steps include a fairness-aware data mining algorithm and calculating error rates by sub-population

Two modes of audit. Direct auditing reads the code — feasible for decision trees and regression models auditors can deconstruct. Indirect auditing feeds widely varying datasets in and tests outputs for bias — essential for deep learning systems that write code autonomously. In practice: audit regularly, run your own "sniff test", commission a full audit.

Tools to name. IBM AI Fairness 360 (open source, 70 fairness metrics, 10 bias mitigation algorithms), Google's What-If Tool, LIME (Ribeiro, Singh and Guestrin 2016), Microsoft Fairlearn, Accenture's Teach and Test, PwC's Responsible AI Toolkit.

Privacy-preserving AI. Anonymisation and aggregation of training data; federated learning, keeping data decentralised on local devices and sharing only model parameters; differential privacy, continuously introducing statistical noise; homomorphic encryption, allowing operations on encrypted data without compromising the encryption.

9. Practice: 20 questions

These test whether you can apply the material, not whether you remember which slide it appeared on. Answers and explanations follow — work through these first with the answers covered. The companion trainer at https://d4b-governance-revision-trainer.vercel.app draws a different twenty from a pool of 100 each time you visit.

1. A manufacturer deals with each data problem as it appears — a breach produces a new password rule, a wrong figure in a board report produces a spreadsheet check by whoever spotted it. Each fix is sound in itself and none is connected to any other. What is the underlying weakness that data governance is meant to address?

a) The organisation has too few technical controls, and the answer is a longer list of rules until incidents stop recurring b) The organisation has failed to appoint a data protection officer, and no fix will hold until that post carries the work c) The organisation is collecting far more data than it can analyse, and the incidents are symptoms of that excess volume d) The organisation is reacting piecemeal to each problem, and nothing systematic connects the fixes it makes

2. Two departments disagree over which of them owns a customer field and whose definition should stand. The disagreement is escalated and formally settled. In the analogy that treats data governance as a system of government, which arm has just acted?

a) The legislature, because settling whose definition stands amounts to setting a standard for future use b) The executive, because administering data structures, roles and the daily record is an executive duty c) The judiciary, because issue management, escalation and dispute resolution belong to that arm d) The steering committee, because it sits above the three arms and takes the final decision itself

3. Which of the following is an act of governance rather than an act of data management?

a) Running the overnight load that refreshes the customer table, then reconciling the row counts against the source systems b) Correcting the 200 mistyped postcodes that a validation report flagged, and re-running the report to confirm the fix c) Deciding who holds the authority to approve a change to the customer record definition, and when they must exercise it d) Encrypting the database so that the stored records cannot be read if the underlying files are copied off the server

4. An information technology department selects a data platform, and the business is then asked to reshape its reporting needs around what the platform can do. Which relationship has been inverted?

a) Data quality assessment should have come before platform selection, so that the state of the existing data could shape which platform was chosen b) The data strategy should link business strategy to information technology strategy, so that data management requirements drive technology decisions rather than follow them c) Data governance should have sat with the technology department, which holds the expertise to match platform capability to the reporting the business needs d) The data lifecycle should have begun at storage rather than at planning, because the platform decides how the data can be held and for how long

5. A team has listed what could go wrong with its customer database and has estimated how likely each event is and what each would cost. A director then asks what it would actually mean for the business if one of them happened. Which step of the risk process is the director asking for?

a) Risk identification — establishing which events could go wrong with the customer database in the first place b) Risk quantification — putting a probability and a likely cost against each of the events that has been listed c) Risk assessment — working out what one of those events would actually mean for this business if it happened d) Risk reporting — reviewing the register's contents and implications with those who have to act on them

6. Risk A has a 90 per cent chance of occurring in a year and would cost 10,000 euro. Risk B has a 2 per cent chance and would cost 900,000 euro. Using the standard derivation of the level of a risk, which statement holds?

a) Risk A ranks higher, because a near-certain event has to be dealt with before one that may never occur b) Risk B ranks higher, because its expected annual loss of 18,000 euro is twice Risk A's 9,000 euro c) They rank equally, because A's higher probability offsets B's greater cost once both are multiplied out d) Neither can be ranked until the organisation has stated a risk appetite against which both can be judged

7. An organisation's board agrees the statement that all data is an asset with real value and must be managed accordingly. A year later nothing has changed in how teams handle data day to day. Which layer of the governance hierarchy has to come next, and what is that layer for?

a) A policy — it turns the statement into fundamental rules supporting standards and expected behaviours, and it binds globally across the organisation b) A procedure — it prescribes the particular way one named team carries out part of a process, which is where a principle has to land before anything changes c) A standard — it fixes an agreed measure of quantity, quality or naming, so a principle takes effect the moment that measure is published to teams d) A control — it stops the unwanted event before it happens, so a principle becomes real as soon as the controls implementing it are switched on

8. A data team holds two documents. One describes how duplicate customer records are identified, assessed and resolved anywhere in the group. The other is titled 'How the Dublin office clears a duplicate, step by step'. Which is the process and which the procedure?

a) Both are procedures — neither turns inputs into outputs, and the only difference between them is the level of detail at which the same work is written down b) The group-wide document is the process — related tasks turning inputs into outputs — and the Dublin document a procedure, a prescribed way of undertaking that process or part of it c) The Dublin document is the process because it turns inputs into outputs, and the group-wide document is a policy, since a statement of rule that applies globally is exactly what a policy is d) Neither is a process — a process has to be automated end to end, so both are procedures describing manual work that nobody has yet engineered into a process

9. In a data governance operating model, what determines how a particular data decision is handled?

a) The level at which the decision sits — strategic, tactical or operational — how decentralised decision-making is in that organisation, and whether it is a business or a technical matter b) The seniority of whoever raised the issue, since escalation follows rank and the operating model routes every question upward until someone with enough authority to settle it is reached c) Whether personal data is involved, since data protection considerations override the operating model and route any such decision to the data protection officer d) The volume of data affected, since the model scales the decision to the size of the dataset and sends the largest-volume matters to the Data Governance Council

10. A marketing platform holds no names — only device identifiers, approximate location and browsing history — yet it can single out the same individual across sessions and target them. Does this fall within the GDPR (General Data Protection Regulation) definition of personal data?

a) No — personal data means information about an identified natural person, and with no name, identification number or email address attached, nobody has been identified b) No — approximate location and browsing history are behavioural signals rather than personal data, which covers only the identifiers a person is registered under c) Yes — personal data is any information belonging to an identified or identifiable natural person, and identification may be indirect, including through an online identifier d) Yes, but only once the platform also holds a

direct identifier such as an email address, since indirect identification brings data into scope only in combination

11. A hospital replaces patient names in a research extract with codes, holding the code-to-name table in a separate locked system. A second extract has been aggregated so that no one can identify an individual, and the source records destroyed. How does data protection law treat each?

a) Both fall outside the law, since neither extract carries a name and the hospital cannot identify anyone from the codes or from the aggregate on its own b) The coded extract remains personal data and stays in scope for every obligation, while the genuinely anonymised extract falls outside the law entirely c) Both remain in scope, since anonymisation suppresses identifiers rather than removing them, so an aggregate can always be reversed to the individuals behind it d) The coded extract falls outside the law because the mapping is held separately, while the aggregated extract stays in scope because it derives from personal data

12. A bank collected transaction data in order to detect fraud. Its marketing team now wants to use the same data to target product offers. Which principle does this most directly engage, and what would make the further use lawful?

a) Data minimisation — data must be adequate, relevant and limited to what is necessary for the stated purpose, so the bank would need to cut the number of customer fields the marketing team is given access to b) Storage limitation — data may be kept no longer than the purpose it was collected for needs, so the bank would need to shorten its retention period before the marketing use begins c) Integrity and confidentiality — data must be processed with appropriate security, so the bank would need to encrypt the marketing extract before anyone in that team is allowed to use it d) Purpose limitation — data collected for specified purposes may not be further processed incompatibly with them, so the bank needs a compatible purpose or its own lawful basis, and must tell customers

13. A newly appointed data governance lead is told by the operations director that the programme's job is to clear the duplicate customer records out of the billing system and then stand down. What is wrong with that framing?

a) Nothing is wrong — a governance programme exists to remediate the defects it was funded to fix, so once the duplicate records are gone the programme has delivered its scope in full, and closing it releases budget and people back to the business b) The framing is wrong only in its choice of target — the programme should have started with reference and master data, where duplication does most damage and travels furthest, and reached the billing system afterwards once those domains were stable c) The framing is wrong because governance is a technical discipline that belongs permanently inside the technology function, which owns the systems where the defects arise, holds the specialist skills and already has reporting lines to escalate through, leaving the business simply to consume the standards it publishes for it d) A governance programme sets data strategy, enforces policy and standards, provides stewardship oversight, assures regulatory compliance, manages issue escalation and sets rules for valuing data assets — it directs remediation work rather than being that work, and does not stand down when one defect is fixed

14. A data governance programme launches with a temporary steering group drawn entirely from the technology function. Which risk does this most directly invite, and what was the steering group meant to achieve in the first place?

a) It invites the programme being seen as just an IT (information technology) project — the temporary steering group exists to bring the right balance of skills and politics at a level senior enough to be taken seriously, under an executive sponsor, so that the business owns the outcome b) It invites overspending on tooling, because technologists specify platforms before anyone has understood the requirement — the steering group exists to hold the programme budget and to approve or reject each purchase before money is committed to a vendor c) It invites poor documentation of what the programme decides — the steering group exists to build and maintain

the data catalogue and the metadata repository, so that scoping decisions taken early remain traceable once the people who took them have moved on d) It invites nothing in particular, since scope and initiation is a technical exercise concerned with span and depth — the steering group exists only to record that technical scope, with business involvement properly belonging to roll-out, when there is something concrete to react to

15. A data scientist is asked by her employer to build a scoring feature she believes is wrong. It breaks no law, and her professional body's code of conduct is silent on it. Which levels of the moral order are in play, and what does the case show about the relationship between ethics and law?

a) Only law is in play — where conduct is lawful there is no ethical question left to answer, since the legislature has already weighed the competing interests and reached a settled position that professional codes then restate in their own words, which is why her professional body's silence on the point settles the matter for her b) Personal morality is in tension with what her group or profession permits, with no societal consensus settling the matter and no legal rule engaged; ethics operate independently of legal systems, although they may have an important part in how those laws come to be made in the first place c) Societal morality has been breached, and it is that breach which makes the conduct unlawful even though no statute names it, because law does no more than codify the established social consensus about what is right at any given moment d) This is a metaethical problem, because what is really at stake is whether moral rules exist independently of the person who holds them, and no practical answer is available to her until that prior question about moral truth has been settled

16. Three questions are put to a seminar. (i) 'Are moral rules objective, or do they depend on who is asking?' (ii) 'What single criterion should we use to judge whether any action is right?' (iii) 'Is it wrong to share a licensed dataset with a colleague whose employer has not paid for it?' Which branch of moral philosophy does each belong to?

a) (i) normative ethics, since it asks which moral rules hold; (ii) applied ethics, since a criterion is meant for practical use; (iii) metaethics, since sharing a dataset raises the nature of obligation itself b) (i) applied ethics, since relativism is a live public controversy; (ii) metaethics, since a single criterion is a claim about moral truth; (iii) normative ethics, since it asks whether one act is right c) All three are applied ethics, because each concerns a decision somebody has to take in the real world, and the branches divide by field — business, medicine, data — rather than by the kind of question asked d) (i) metaethics, since it asks about the nature of moral claims; (ii) normative ethics, since it seeks one ultimate criterion of conduct; (iii) applied or comparative ethics, since it takes principles into a contested practice

17. A vendor tells a procurement panel that its product 'is only an algorithm, so governance obligations do not really apply'. Distinguish an algorithm from an algorithmic system, and say why the distinction matters here.

a) There is no useful distinction — the terms are used interchangeably in practice, so the vendor's point stands, and any obligation attaches to whoever operates the software rather than to the software itself or to the supplier that wrote and sold it b) An algorithm is trained from data while an algorithmic system is hand-coded to fixed rules, so only the first of the two can carry bias into the conclusions it produces, and only the first attracts the governance obligations the panel is asking about, since a rule set written by hand can be inspected line by line before it is bought c) An algorithmic system is any algorithm running in a cloud environment where data is pooled across customers, so code deployed on a customer's own premises falls outside the governance perimeter entirely, which is where this vendor's product happens to sit d) An algorithm is a set of precise instructions or rules for solving a predefined problem; an algorithmic system is software using one or more algorithms to collect and analyse data and draw conclusions within a process — governance attaches to the system, its data and its conclusions, not to the arithmetic alone

18. A compliance manager argues that privacy is purely a legal matter — meet the regulation and the ethical question disappears. On what grounds is privacy treated as an ethical matter in its own right?

a) Because privacy law is amended frequently and unevenly across jurisdictions, so ethics serves as a stopgap covering the ground until the next amendment catches up with whatever the technology has meanwhile started doing to people — and once the statute has caught up, the ethical argument has nothing further of its own to add to the position b) Because respecting personal autonomy, preventing harms such as trolling, identity theft and financial loss, requiring providers to be transparent and accountable in their use of data and to avoid discrimination, and sustaining the societal trust digital services depend on, are claims that hold whether or not a statute codifies them c) Because ethics binds individuals while law binds organisations, so the two are needed together to cover the whole field, which makes the compliance manager right about the organisation itself but wrong about the employees inside it, who remain bound by their own professional codes whatever the organisation's own compliance position happens to be d) Because privacy becomes an ethical matter only in jurisdictions that have no data protection legislation, where there is nothing else for a person to appeal to when they object to a particular use of their own data and no court or supervisory authority will hear the complaint

19. Under the risk-based approach of the EU (European Union) Artificial Intelligence Act, how would each of these be treated: a recruitment platform that scores job applicants, a customer service chatbot presenting itself as a person, and a municipal system that scores citizens on their general social behaviour?

a) All three are high risk, since all three process personal data about identifiable individuals, and the tier is set by the sensitivity of the data a system consumes rather than by the decisions it takes about people, so a scoring platform and a chatbot handling the same customer records sit in the same tier b) The recruitment platform is high risk — permitted, but subject to requirements to make sure it is safe and trustworthy; the chatbot falls in the transparency-risk tier, permitted but with disclosure obligations such as labelling and watermarking; social scoring is unacceptable risk and prohibited c) The recruitment platform carries disclosure obligations only, the chatbot is high risk because it interacts directly with the public, and social scoring is high risk because a public body operates it under statutory powers and remains answerable for it, so all three sit inside tiers that permit deployment subject to conditions d) All three are permitted with no restrictions, because the Act regulates how models are built and placed on the market rather than how deployers subsequently choose to use them, once a declaration of conformity has been filed for each of them

20. A firm plans to deploy AI (artificial intelligence) agents that operate autonomously, organise their own workflows and adapt continuously. Which governance problem does this most directly create?

a) None of substance — agents are simply faster software running the same underlying models, so the controls already applied to those models transfer to them unchanged and no new governance work is called for; the assessment, the documentation and the named accountability owner recorded for the model all carry across unaltered to every agent built on top of it b) A cost problem only, since autonomous operation consumes far more computing resource than a supervised process, so budgets rather than controls are what will constrain deployment, and a process that costs too much to run will not survive long enough to need governing, which makes the resourcing decision the only governance decision that actually binds c) Continuous adaptability and self-organised workflows mean the system assessed is not the system running, so point-in-time assessment, fixed documentation and static assignment of accountability stop matching reality — the location of decision-making, traceability and named human responsibility have to be re-established for a moving target d) A data quality problem only, since agents generate their own inputs from their own previous outputs and so compound their own errors over time, and no volume of monitoring can catch an error the system has already built further work upon, so the answer lies in stricter validation of the inputs rather than in any change to how accountability is assigned

10. Answers and explanations

Each answer gives the reasoning, then the misconception the wrong options are built on.

1. Answer d · Wk 1

Data governance is put forward precisely as the move from piecemeal responses to individual data problems towards a systematic approach that connects them. The weakness lies not in any single fix but in the absence of anything joining them up.

Where this trips people up. The tempting reading is that this is a controls problem — add enough rules and the incidents stop. That misses why data governance is treated as a discipline in its own right rather than as a longer list of controls. Isolated fixes cannot share definitions, cannot reveal a pattern across incidents, cannot allocate responsibility, and vanish when the person who invented them leaves. The argument runs from the growth of data and its falling collection cost, through the financial, reputational and societal risks that growth creates, to the conclusion that only a systematic response can hold together. Revise the case for the subject and the four changes data governance is expected to drive in an organisation.

2. Answer c · Wk 1

In the government analogy the judicial function covers issue management, escalation of disputes and dispute resolution — exactly what has happened here.

Where this trips people up. The near-miss is the legislature, because the outcome of the dispute will indeed be a definition, and definitions look like standards. The line to hold is between making a rule for the future and resolving a live conflict between parties — the first is legislative, the second judicial. The executive arm protects and serves data structures and roles and carries the administrative load. The steering committee coordinates the three functions rather than standing above them as a fourth authority, which is what makes the last option wrong. The value of the analogy is that it shows governance needs all three: policy without dispute resolution stalls, and dispute resolution without policy has nothing to reason from.

3. Answer c · Wk 2

Governance sets the ground rules — who acts, when and where — while management is the daily work that supports the data and, through it, the business strategy. Allocating decision authority is a governance act; the other three are daily work.

Where this trips people up. The temptation is to sort by importance rather than by kind, and encryption feels weightier than a decision about a definition. The test is whether the activity decides the rules or operates within them. This is why data governance is sometimes called the management of data management, and why it assumes a separation between those governing and those managing. That separation matters because if the same people set a standard and report on whether it was met, there is no independent check — the same reason corporate governance separates a board from an executive team. Revise the governance and management comparison, and how data strategy links business strategy to information technology strategy.

4. Answer b · Wk 2

Data strategy is the link between the business strategy and the supporting technology strategy, and one of the data management principles states plainly that data management requirements must drive information technology decisions.

Where this trips people up. The inversion is easy to miss because platform selection is a genuine decision somebody has to make, and the technology team holds the expertise to make it. What has gone wrong is the sequence: the business requirement should constrain the tool, because data management requirements are business requirements. The third option is the opposite of the lesson — moving governance into technology entrenches the inversion rather than correcting it. Governance defines who, when and where actions are taken; management carries out the daily actions that follow. Revise the data management principles, particularly that data is an asset, that its requirements are business requirements, and that those requirements drive technology choices.

5. Answer c · Wk 3

The process runs identification — what might happen; quantification — how bad can it be; assessment — how might it affect us; then negation or mitigation; then reporting and review. The team has completed the first two and the director is asking the third question.

Where this trips people up. Quantification and assessment collapse into one another easily, since both involve thinking about severity, and that collapse is the most common error here. Quantification produces figures — probability of occurrence during a year, potential loss, expected annual loss. Assessment interprets those figures for this particular organisation: whether the loss is survivable, which processes stop, who is affected, what has to be told to whom. Skipping assessment is how an organisation ends up with a well-populated risk register that changes no decisions. Reporting is the later discipline of regularly reviewing the register's contents and implications. Revise the five steps and what each one contributes.

6. Answer b · Wk 3

The level of a risk is the probability of the event multiplied by the cost of that event to the organisation, and a risk register ranks entries by that level — giving 18,000 euro for B against 9,000 euro for A.

Where this trips people up. The first option is the intuition that frequency equals urgency, and it is how untrained registers usually get sorted. The last imports risk appetite from the wider risk literature, whereas the derivation used here is the arithmetic one. Carry the limitation along with the formula: expected annual loss says nothing about whether a single 900,000 euro event would be survivable, and an organisation that could absorb ten instances of Risk A but not one of Risk B should not treat them as merely two-to-one apart. That is precisely why quantification is followed by assessment, and why the organisation has to decide an appropriate cut-off for mandating action rather than reading the ranking mechanically.

7. Answer a · Wk 4

The board has stated a principle — a statement of philosophy or core belief. Policies are the layer that converts principles into fundamental rules supporting standards and expected behaviours, and they are global within the organisation.

Where this trips people up. The misconception is that a well-worded principle changes behaviour on its own. It does not: the chain runs principles, then policies, then processes and procedures, then rules and standards, and each layer does a different job. Standards and procedures are tempting answers because that is where the visible detail lives, but neither can be written until a policy has set the rule they implement. Note also the guidance that there should be relatively few data policies, stated briefly and directly, and that they must be communicated, monitored, enforced and periodically re-evaluated. Revise the hierarchy as four layers with four distinct purposes.

8. Answer b · Wk 4

A process is a series of related tasks or methods that together turn inputs into outputs; a procedure is a prescribed way of undertaking a process or part of a process. The group-wide document describes the what, the Dublin document a particular how.

Where this trips people up. People collapse these two words together and then argue about which document is authoritative. The distinction is worth holding because it tells you where each artefact sits: policies are the rules of an organisation, and procedures are the processes used to enact those policies. option c is tempting because the detailed document feels more real, but detail does not make something a policy — policies are global statements of rule, not local instructions. option a loses the difference between a way of doing something and a particular way of doing it. Revise the process-versus-procedure slide alongside the definition of a policy.

9. Answer a · Wk 5

The operating model is defined by roles and responsibilities together with the processes that determine decisions, and those processes turn on the level of the decision, the degree of decentralisation, and whether the decision is a business or a technical one.

Where this trips people up. The tempting answer is rank, because that is how many organisations actually behave — but routing decisions by the seniority of whoever noticed the problem produces bottlenecks at the top and unresolved disputes at the bottom at the same time. The three factors matter because they are the variables an organisation can set deliberately: how much is devolved to business teams rather than held by a data governance office, and which questions are business questions dressed as technical ones. Remember also that one size does not fit all — operating models vary with an organisation's functions, geographical scope, culture, structure, data maturity and motivation for change. Revise the decision-making slide alongside the operating model diagram.

10. Answer c · Wk 6

Personal data covers information belonging to an identified or identifiable natural person, where the individual could be identified directly or indirectly — names, identification numbers, location data and online identifiers all count.

Where this trips people up. The misconception is that personal data means named data, and it is the most expensive mistake in advertising technology. A device identifier that lets you single out one person and follow them is an online identifier, which the definition lists explicitly, and physical, physiological, genetic, mental, economic, cultural or social factors can identify a person just as effectively. Scope decides everything downstream — lawful basis, data subject rights, security duties, transfer rules — so an organisation that gets this wrong gets all of it wrong at once. Note also that the regulation reaches organisations handling the personal data of EU (European Union) residents regardless of where the organisation is physically located. Revise the personal data definition together with territorial scope.

11. Answer b · Wk 6

Pseudonymised data is in scope; anonymised data is not. Holding the mapping elsewhere reduces risk but leaves the individual identifiable, so the coded extract is still personal data.

Where this trips people up. option d is the expensive error, and it is popular because separating the mapping genuinely does reduce risk. What it does not do is change the legal status of the data. Pseudonymisation appears in the regulation as a security measure and as a technique for building the principles into processing — it is a way of complying, not a way of leaving. The practical consequence is that everything continues to apply to the coded extract: lawful basis, minimisation, retention limits, security, data subject rights and the transfer rules. Anonymisation, by contrast, has to be genuine and irreversible before it takes you outside, and aggregation that still allows an individual to be singled out does not qualify. Revise the personal data definition alongside the security and design provisions.

12. Answer d · Wk 6

Purpose limitation is the principle that bites when data gathered for one stated purpose is repurposed. The reuse has to be compatible with the original purpose or rest on its own lawful basis, and customers have to be told.

Where this trips people up. Minimisation and storage limitation are tempting because all three principles are about restraint, but they answer different questions: minimisation governs how much data, storage limitation how long it is kept, and purpose limitation what it is for. Encryption addresses integrity and confidentiality and would do nothing to cure an unlawful purpose — a well-secured unlawful process is still unlawful. Notice that lawfulness, fairness and transparency runs alongside: even a compatible reuse fails if customers were never told it might happen, and fraud prevention notices rarely mention marketing. Treat the seven principles as seven separate questions to ask of any processing, and scenarios like this sort themselves out.

13. Answer d · Wk 7

The scope of a typical programme spans strategy, policies, quality and architecture standards, oversight, compliance assurance, issue management, project sponsorship and asset valuation. Cleaning duplicates is data management execution; governance is the decision-making, monitoring and enforcement layer above it.

Where this trips people up. The misconception is that governance is a remediation project with an end date. option a is tempting because a visible data quality defect is usually what triggers the programme in the first place, and fixing it feels like completion — but the defect is the symptom, and the programme exists to change how decisions about data get made so that class of defect stops recurring. option c trades on the same confusion in a different direction: putting governance inside the technology function is precisely what the scope-and-initiation phase warns against, because it strips the business of ownership. Hold the distinction between governance (deciding, monitoring, enforcing) and management (executing) — it recurs through every phase of the implementation model. Revise the scope-of-a-typical-programme material alongside the eight-phase process.

14. Answer a · Wk 7

The temporary steering group set up during scope and initiation is there to ensure the programme is not seen just as an IT project, that the right balance of skills and politics is present, and that it sits at a level senior enough to be taken seriously under an executive sponsor.

Where this trips people up. The misconception is that governance design is a technical competence, so technical people should staff it. option d is the most tempting because sequencing arguments sound disciplined — get the technical scoping done, bring the business in later — but by the time roll-out arrives the programme has already made the decisions the business needed to own, and resistance then arrives as a surprise. The steering group is doing political work as much as analytical work: deciding span (which business areas are covered) and depth (how far into the organisation the programme reaches), and approving scope while acknowledging constraints. This is also why governance is treated as a virtual organisation driven by everyday behaviour rather than as a technology department. Revise phase 1, scope and initiation, together with the change-management material on where resistance comes from.

15. Answer b · Wk 8

The four levels run personal morality, group or professional morality and ethical standards, societal morality, and law. Ethics operate independently from legal systems but may play a part in shaping how laws are made, so lawfulness does not settle an ethical question.

Where this trips people up. The misconception is that law exhausts ethics — if it is legal, it is permitted, and the discussion ends. option a is tempting because law is the level with institutions and enforcement behind it, and courts exist to resolve its contradictions and ambiguities, which makes it feel like the authoritative layer. But a group's ethics can bring an individual into conflict with their own personal morality, which is a perennial theme of literature and drama precisely because no external authority resolves it. Notice the direction of influence being insisted on: ethics feeds into law-making, so today's ethical argument is often tomorrow's statute — which is why practitioners cannot wait for legislation before acting. Revise the two morality-ethics-law slides and connect them to the observation that regulation tends to lag behind technological advance.

16. Answer d · Wk 8

Metaethics studies the meaning and nature of ethical theory itself and how truth values are determined; normative ethics studies moral duties and rules and asks what makes actions right or wrong; applied or comparative ethics takes principles into specific controversial issues in the real world.

Where this trips people up. The misconception is that any question with a moral flavour is 'applied', which is what makes option c tempting — every ethical question feels practical because a person is standing there waiting for an answer. The test is what the question is about rather than how urgent it feels: (i) asks about the status of moral claims, (ii) asks for a criterion that

would apply to all actions, (iii) asks about one contested practice. The distinction earns its keep because arguments talk past each other when the parties are in different branches — someone defending file-sharing on relativist grounds has changed the subject from applied ethics to metaethics. Note that normative ethics assumes there is only one ultimate criterion of moral conduct, whether a single rule such as the Golden Rule, a set of foundational principles, or a set of good character traits. Revise the three branches and the normative ethics material that follows.

17. Answer d · Wk 9

The definitions are deliberately different in scope: the algorithm is the rule set, the algorithmic system is the software that collects data, analyses it and draws conclusions within a process. Obligations follow the system because that is what touches data and produces consequences for people.

Where this trips people up. The misconception is that narrowing the description of a technology narrows the duties that attach to it, which is exactly the move the vendor is making. option b is tempting because bias is most often discussed in relation to learned models, and it is true that historical data is a major route by which bias enters — but choices of model, feature selection and which metric is optimised are all designer decisions that can carry bias into a system with no learning in it at all. Remember also the origin of the term, from the ninth-century Persian mathematician Al-Khwarizmi, and the older assumption that algorithms were independent of human prejudice — an assumption now displaced by an understanding of how bias and prejudice enter. Revise the algorithm and algorithmic system definitions, then look at the uses listed for algorithms, from mapping an online world and shaping a personal filter bubble to sentiment analysis and purchase prediction.

18. Answer b · Wk 10

Privacy is grounded in personal autonomy, in preventing tangible harms to individuals, in obligations of transparency, accountability and non-discrimination on service providers, and in the societal trust that underpins effective use of digital services, particularly finance and payments. None of those grounds depends on a statute existing.

Where this trips people up. The misconception is that compliance is the ceiling of obligation rather than its floor. option a is tempting because it grants ethics a role while keeping law in charge — but if ethics were only a stopgap, it could not explain why the same argument sometimes says the law itself is wrong, and it gets the direction of influence backwards, since ethical argument is part of how laws come to be made. Notice how wide the harm list runs: identity theft and financial loss are individual and material, while influencing public opinion and electoral systems through recommender systems is collective and structural, and privacy analysis has to hold both. This is also why the right to privacy appears in human rights instruments rather than only in sectoral statutes. Revise the material on privacy as an ethical issue, then read the human rights instruments as codification of an argument that already existed.

19. Answer b · Wk 11

The tiers run unacceptable risk and prohibited (social scoring, individual predictive policing), high risk and permitted subject to safety and trustworthiness requirements (recruitment, medical devices), transparency risk with disclosure obligations (impersonation by chatbots, synthetic content), and minimal or no risk permitted without restriction.

Where this trips people up. The misconception behind option a is that risk follows from the data a system processes rather than from what the system decides about people. A recruitment scorer and a spam filter can both process personal data, but only one of them allocates access to livelihoods, and that consequence is what places it in the high-risk tier with its accompanying requirements. option c inverts the chatbot and the recruiter, which is tempting because direct interaction feels more exposed than a background scoring process — but the transparency tier exists for a narrower harm, that of not knowing you are dealing with a machine or with synthetic content, and it is answered by labelling rather than by conformity requirements. Tiering is the first governance move to make with any new system, because it determines everything downstream:

documentation, human oversight, audit frequency and the resourcing needed to support them. Revise the risk pyramid and practise placing a novel system on it from a description of what it decides.

20. Answer c · Wk 12

Agentic systems are characterised by autonomous operation, self-organised workflows and continuous adaptability and self-improvement. Each of those attacks a different assumption behind conventional governance, which relies on a system staying as it was when it was assessed and documented.

Where this trips people up. The misconception is that autonomy is a difference of degree rather than of kind, which makes option a tempting — the controls are familiar, so surely they still apply. What breaks is the premise underneath the controls: a governance framework describes artefacts such as strategies, policies, processes and standards, roles and responsibilities, and the nature and location of decision-making, and a self-organising system moves the third of those without telling anyone. This is why greater attention to documentation, process and traceability is listed among the transparency measures expected to become more common, and why accountability is expected to be reasserted through assignment of roles and responsibilities and clarity about decision-making. It also sharpens the point that human agents remain responsible for ethical operation however capable the system becomes. Revise the rapid technological developments material and the data strategies, policies, processes and standards material that translates those trends into governance work.

11. Practice: five short responses with answer scaffolds

Eight marks each, nine minutes each. Each scaffold is the shape of a full-mark answer — define, apply a named framework, give a concrete example, then evaluate.

A. An organisation with four business units, each maintaining its own customer master data, is producing inconsistent management reporting. Recommend a governance operating model and justify your choice.

- Name the current state: Scenario 2, decentralised execution across multiple business units. Its named weaknesses are duplicate master data and inconsistent or meaningless reporting when multiple parties are involved.
- Recommend Scenario 4 — centralised governance with decentralised execution: a central body defines the framework of controls while businesses create their own parts of master data. It preserves agility while adding control, with shared responsibility between the DG (data governance) organisation and the business.
- Justify against the alternative: full centralisation (Scenario 3) delivers consistency but brings delays and requires a formal, larger DG organisation — disproportionate here.
- Flag the risk the slides attach to both centralised options: potentially higher risks from data protection breaches, because master data must be distributed across a larger system landscape.
- Close on implementation: Ladley phase 6 establishes framework design; use RACI (Responsible, Accountable, Consult, Inform) to assign responsibility and accountability, and let the degree of centralisation follow size, geography, structure, regulatory environment and culture.

B. A retail bank plans to use a machine learning model to score personal loan applications. Assess the governance requirements before deployment.

- Classify it: high risk under the AI (artificial intelligence) Act — permitted, but subject to requirements to make it safe and trustworthy.
- GDPR (General Data Protection Regulation) hooks: Art. 22 gives the applicant a right to contest a solely automated decision; Art. 13 requires notice of automated decision-making; Art. 15 requires meaningful information about the logic involved. Recital 71 requires appropriate statistical procedures and prevention of discriminatory effects.
- Impact assessment: Art. 35 requires a DPIA (Data Protection Impact Assessment) for high-risk processing using new technologies. Cite Kaminski and Malgieri (2019) — the DPIA is the GDPR's version of an Algorithmic Impact Assessment, connecting individual rights to systemic governance.
- Bias sources to test for: historical bias in lending data; proxy attributes standing in for sensitive attributes, such as postcode for ethnicity; algorithmic objectives bias from prioritising accuracy. Cite Sweeney on differential micro-targeting of higher-interest credit products.
- Controls: FAT/ML (Fairness, Accountability and Transparency in Machine Learning) gives the operational answer — designate a responsible person and a redress route, explain decisions in non-technical terms, log sources of error, enable third-party audit, calculate error rates by sub-population. Run indirect audits, since the model may not be code-readable.
- Evaluate: fairness trades off against accuracy, and avoiding discrimination may require collecting the sensitive attribute in order to test for disparity — the paradox the slides name.

C. "Ethics operate independently from legal systems." Discuss, using at least two normative frameworks.

- Start with the four-level structure: personal morality, group or professional ethics, societal morality, law. The parallel definitions are the hinge — moral acts conform to what an individual believes is right; legal acts conform to the law or at least do not contravene it.

- Argue the independence claim: something can be lawful and unethical. The Optum tool broke no law but reproduced and exacerbated social inequality. Conversely, compliance is not a defence — governance value must be claimed, not assumed.
- Deontological reading: duties come from universal law, not statute. Kant's imperative to treat humanity always as an end and never simply as a means condemns treating data subjects purely as model inputs, whatever the law permits.
- Consequentialist reading: assess by outcomes. This can license privacy trade-offs for the greater good of progress — and the slides' own critique applies: who decides what the greater good is, and is it fair to sacrifice the minority?
- Qualify the claim: ethics has a role in how laws are made in the first place, so independence is not separation. The GDPR was prepared for by the EU Charter; regulation tends to lag technological advance. Add Rawls if there is room — judge the practice by its effect on the least advantaged, who are least likely to be represented when the law is written.

D. Explain how historical bias enters an algorithmic system and what a data governance programme can do about it.

- Mechanism: models are trained on records of past human decisions. Humans remember trauma over positives and react more strongly to negative stimuli — those experiences are reflected in training data, so models reproduce the same wrong judgements people make.
- The four named causes: historical bias preexisting in datasets; technical or statistical bias including sampling and aggregation; algorithmic objectives bias from prioritising accuracy or poor feature selection; and proxy attributes standing in for sensitive attributes.
- The amplification mechanism: feedback loops reinforce bias without the user's knowledge. Targeted high-interest credit advertising is clicked, so the system never accumulates counter-factual lower-interest suggestions the consumer could be eligible for and would prefer.
- Evidence: Amazon's recruiting tool penalised résumés containing "women's"; facial recognition training sets are over 75% male and over 80% white; Buolamwini was better recognised in a white mask — "default is not neutral".
- Governance response: bias detection begins with careful handling of sensitive information, then comparing outcomes across groups, including by simulation before real-world application. Holstein et al. (2019) add fairness-aware data collection and curation, attention to blind spots in the team itself, awareness among those making labelling decisions, and more proactive and holistic auditing.
- Note the limit: groups may be harmed without being protected in anti-discrimination legislation — lower-income or rural applicants, for example. Legal compliance does not exhaust the governance obligation.

E. Your organisation is starting a data governance programme. Outline the first three phases and the main risks of failure.

- Phase 1, scope and initiation: decide span and depth. Set up a temporary steering group at a senior enough level, led by an executive sponsor, so the programme is not seen as just an IT project.
- Phase 2, assess organisational capability: baseline by survey and focus group. The test is whether the organisation identifies philosophically with the mantra that data is an asset, assets need to be managed, and management requires governance. Locate resistance points: poor communication, no business alignment, ineffective training, lack of sponsorship.
- Phase 3, establish vision: a vision is where you want to be, the mission is how to get there. Have an elevator pitch, identify business drivers, define what a day in the life would look like. Keep it simple.

- Failure modes: the four objections and their responses — we don't need more rules; you're slowing my project; how much is this costing; why can't I just call Mike. Answer with performance incentives, better metrics, baseline data showing value, and education.
- The structural risk: DG rarely becomes a stand-alone department; it is a virtual organisation of business and IT personnel, so it depends on behaviour rather than authority. The DGI's first casualty is decision rights, the second is naming the beneficiaries.
- Close with the challenge quadrants: technical (complexity, legacy, scalability), organisational (leadership, silos, inertia), cultural and ethical (culture, balancing usefulness against rights, public trust), regulatory (legislation, industry standards).

12. References

Module materials

Digital4Business / NCI Cloud Competency Centre. (2026). Data Governance and Ethics (DGE), Weeks 1–12 [Lecture slides]. University of Bologna D4B Joint Professional Master's. CC BY 4.0.

Data Governance Institute. (2025). The DGI (Data Governance Institute) data governance framework. https://datagovernance.com/the-dgi-data-governance-framework/ — as summarised in the Week 4 companion guide, 7 June 2026.

Cited within the module

DAMA (Data Management Association) International. (2017). DAMA-DMBOK: Data management body of knowledge (2nd ed.). Technics Publications.

Holstein, K., Wortman Vaughan, J., Daumé, H., Dudík, M., & Wallach, H. (2019). Improving fairness in machine learning systems: What do industry practitioners need? Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 1–16. https://doi.org/10.1145/3290605.3300830

Kaminski, M. E., & Malgieri, G. (2019). Algorithmic impact assessments under the GDPR (General Data Protection Regulation): Producing multi-layered explanations. SSRN Electronic Journal, 1–29. https://doi.org/10.2139/ssrn.3456224

Ladley, J. (2012). Data governance: How to design, deploy and sustain an effective data governance program. Morgan Kaufmann.

Olteanu, A., Castillo, C., Diaz, F., & Kıcıman, E. (2019). Social data: Biases, methodological pitfalls, and ethical boundaries. Frontiers in Big Data, 2, 13. https://doi.org/10.3389/fdata.2019.00013

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144. https://doi.org/10.1145/2939672.2939778

Legislation and current status

Regulation (EU) 2016/679 (General Data Protection Regulation). Official Journal L 119, 4.5.2016, pp. 1–88. https://eur-lex.europa.eu/eli/reg/2016/679/oj

Regulation (EU) 2022/868 on European data governance (Data Governance Act). Official Journal L 152, 3.6.2022, pp. 1–44. In force 23 June 2022; applies from 24 September 2023. https://eur-lex.europa.eu/eli/reg/2022/868/oj

Regulation (EU) 2024/1689 (Artificial Intelligence Act). In force 1 August 2024. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Council of the European Union. (2026, June 29). Artificial intelligence: Council gives final green light to simplify and streamline rules [Press release]. https://www.consilium.europa.eu/en/press/press-releases/2026/06/29/artificial-intelligence-council-gives-final-green-light-to-simplify-and-streamline-rules/

Regulation (EU) 2026/1744 (Digital Omnibus on AI (artificial intelligence)). Adopted 8 July 2026; Official Journal 24 July 2026; in force 27 July 2026.

Version history

Version	Date	Change
3.3	3 Aug 2026	Logo at the head of page 1 was rendering with its top clipped against the page margin; given headroom so it prints complete.
3.2	3 Aug 2026	Every numbered section now starts on a new page. Version numbering unified with the companion trainer, so this document and the web app share one number from here — which is why this jumps from 2.0 to 3.2.
2.0	3 Aug 2026	Practice questions rebuilt from a new 100-question bank and rewritten to test

Version	Date	Change
		applied understanding rather than recall of where material appeared. Answers now carry a fuller explanation naming the misconception behind each wrong option. Acronyms written out on first use throughout. Document grew from 17 to 26 pages.
1.2	3 Aug 2026	Safe Exam Browser guidance softened — it may be required but is not yet confirmed. Advice to have D4B login details on paper retained.
1.1	31 Jul 2026	Houghton Consulting branding applied: logo, brand teal sampled from the logo asset, footer on every page with page numbering and author attribution.
1.0	30 Jul 2026	First build from the twelve session decks and the Week 4 Data Governance Institute companion guide.

The companion interactive trainer at https://d4b-governance-revision-trainer.vercel.app shares this version number from v3.2 onwards — it carries a pool of 100 practice questions drawn at random, where this document carries a fixed 20.

Verification note

Verified: all session content above is quoted or closely paraphrased from your twelve slide decks and the Week 4 DGI companion guide, read in full. The AI Act deferral was confirmed independently against the Council of the EU press release of 29 June 2026 and the Official Journal publication of Regulation (EU) 2026/1744.

Inferred: the mapping of likely exam weighting to session content, the practice questions and the answer scaffolds are my construction, not the module coordinator's. Treat them as drill material, not as a prediction of the paper.

Not verifiable: four quiz slides per deck contain embedded interactive objects whose questions did not survive text extraction, and several slides — including the enumerated GDPR data subject rights, the DMBOK (Data Management Body of Knowledge) wheel labels and the Week 5 operating-model role labels — exist only as images. If the exam draws on those, this brief cannot cover them. Worth opening those specific slides in the original PDFs.