

# The Sovereign Machine

A blueprint for the AI-enabled IT department: sovereignty as architecture, ethics as controls, and the department that can answer for what AI does in its name.

ADRIAN HOLLISTER

FOUNDER, ALTLIBRE

- 1 The shift
- 2 The frame
- 3 The blueprint
- 4 Beyond
- 5 The working

## THE PRACTICE

15

## PEOPLE, FULLY REMOTE

The practice this blueprint is drawn from, serving a global organisation across more than forty countries.

40+

## COUNTRIES

One department, three planes, no first-line queue.

AI

## FIRST LINE AND TRIAGE

Substantially AI-operated. Development largely AI-generated and human-validated.

No. 3

## ALT-LIBRE PAPERS

October 2026. Released under CC BY-SA 4.0.

3

## WAYS TO RESPOND

Denial, abdication, reconstruction. Two of them are failures.

# The numbers first

## The practice in four numbers, before the argument.

Artificial intelligence is a force that dissolves the IT department's traditional boundaries and rebuilds them somewhere else, not a tool the department adopts. For twenty years the IT function justified itself through custody: of servers, of networks, of tickets, of the knowledge no one else could reach. AI removes the need for much of that custody, and with it the quiet monopoly that made the department indispensable on its own terms.

Three claims run through the paper. Sovereignty is an architectural property, not a marketing one. Ethics is a governance discipline, not a sentiment. The workforce question is a design question. What follows is built to be implemented, audited and measured, and it is drawn from practice rather than proposed in the abstract.

This document is written for CIOs, IT directors, security leaders and the boards that hold them to account. It is written from a European perspective, and it applies the same suspicion to all foreign dependencies, whatever their flag.

Part One documents the shift. Part Two defines the frame. Part Three is the blueprint itself. Part Four extends the argument beyond the department. The annexes are the working instruments: the dependency mapping template with a worked example, a sample register entry, the policy and DPIA skeletons, the board briefing template, the evaluation checklist and the maturity assessment checklists. A reader in a hurry should go to Chapters 3, 4, 7 and 13 first.

## 1. The department that deletes itself

Every technology transition begins with a lie told in the passive voice: "jobs will be automated." The passive voice conceals the agent. Someone decides. Someone designs the automation, scopes it, chooses which tasks it absorbs and which people it displaces. In the AI transition, that someone is, for the first time, frequently the IT department itself. The department is no longer merely the instrument of others' automation; it has become both the executioner and the executed.

Previous transitions did not work this way. When enterprise resource planning dissolved the purchasing clerk, IT built the system but did not staff the purchasing department. When cloud computing shrank the data centre team, the decision was taken jointly with finance, and the beneficiaries were legible on the balance sheet. AI reaches first into the IT organisation's own functions: incident triage, log analysis, code generation, documentation, service desk response, vulnerability management, configuration drift detection. The department that automates is the department being automated.

There are three ways to respond, and two of them are failures.

The first failure is denial. Some departments treat AI as a passing fashion and police its use rather than absorb it. This posture has a shelf life measurable in months. Shadow AI does not wait for policy: staff already use consumer models on corporate problems, pasting contracts, credentials and customer data into services over which the organisation exerts no control whatsoever. Denial does not prevent AI adoption; it prevents governed AI adoption. The department that forbids what it cannot see ends up responsible for what it never knew was happening, a position that satisfies neither the regulator nor the insurer.

|   |               |
|---|---------------|
| 1 | The shift     |
| 2 | The frame     |
| 3 | The blueprint |
| 4 | Beyond        |
| 5 | The working   |

The second failure is abdication. The department embraces AI as an efficiency programme, buys whatever the market pushes, and measures success in the number of tasks removed from human hands. Two outcomes follow predictably. Institutional knowledge leaves with the people whose tasks were automated, because tasks and knowledge travel together even when job descriptions claim they do not. And the organisation's decision-making steadily migrates into model providers whose priorities, jurisdictions and update cadences are not the organisation's own. Abdication is efficient in the way a fire sale is efficient.

The third response is reconstruction. The department takes the same capabilities, but treats them as objects of governance rather than instruments of efficiency: it decides where AI may operate, on what data, under whose oversight, with what fallback, and under which legal regime it answers. Reconstruction accepts that the operational core of IT will shrink. It invests the liberated capacity in the four functions that cannot be automated away because they are the functions that make automation legitimate: architecture, governance, security oversight and human accountability.

This paper is a blueprint for reconstruction. It assumes the reader is an IT or security leader who has already accepted that the change is real, and who now needs a defensible, implementable, ethically grounded plan.

One point of provenance matters to everything that follows. This blueprint is not a theoretical construct. It is drawn from the practice described in the executive summary: a department of roughly fifteen people, fully remote, serving a global organisation across more than forty countries, in which first line and triage are substantially AI-operated, development is largely AI-generated, internal teams build their own AI systems inside sanctioned guardrails, and every expert works with AI knowledge, speed and assistance at their shoulder. The consequences of that small scale deserve emphasis before the detailed argument, because they invert the conventional AI adoption narrative. The conventional narrative holds that AI threatens large operational teams; the practice demonstrated here is that AI threatens the premise of the large team itself. A department of fifteen people can now operate a service estate, a development pipeline and a security function at a standard that recently demanded multiples of that headcount, provided it is architected deliberately. Headcount is no longer the measure of an IT department's capability. Governance density per person is.

One further framing point before the detailed argument. The IT department of the coming decade is a political institution. This is not a metaphor. When a department chooses a model provider, it chooses which legal system will govern its organisation's most consequential processing. When it designs an AI-assisted decision path, it chooses whether a human retains the power to overrule it. When it selects training data, it chooses whose reality the system will recognise. These are political choices with technical implementations. Pretending they are merely procurement choices is how organisations sleepwalk into dependency.

## 60%

### OF EMPLOYMENT EXPOSED

IMF estimate for advanced economies. Plan at task level; do not argue with the forecasts.

## 2. What AI is already doing to IT work

The change is present tense, not a forecast. This chapter surveys the functions where AI has already altered the shape of IT work, in most organisations without any formal programme having been announced. The survey matters because the blueprint must be built on observed reality rather than vendor roadmap.

**Service desk and support.** Conversational agents now absorb the first tier of support: password resets, access requests, how-do-I queries, status questions. The remainder arrives at Tier 1 already summarised, categorised and, increasingly, with a suggested resolution attached. In the author's own

|   |               |
|---|---------------|
| 1 | The shift     |
| 2 | The frame     |
| 3 | The blueprint |
| 4 | Beyond        |
| 5 | The working   |

practice this is no longer a projection: first line support and ticket triage are substantially AI-operated, and the human role begins where the machine's confidence ends. The effect is demographic: the junior roles through which two generations of IT staff learned the trade are thinning, a problem addressed in Chapter 12.

**Development and engineering.** Code generation assistants now write a substantial fraction of routine code: boilerplate, tests, migrations, documentation, infrastructure definitions. In the author's practice, development is largely AI-generated: engineers specify, review, test and validate machine-written code, and the department's development throughput is a function of its specification and review discipline rather than its coding headcount. Engineers report gains on the routine and new failure modes on the complex, where generated code is confidently wrong in ways that review fatigue makes dangerous. The skill profile shifts from writing code to specifying, reviewing and validating it.

**Security operations.** AI performs triage in the security operations centre: alert clustering, false positive suppression, anomaly detection across telemetry volumes no human team could read. Detection engineering increasingly means curating model-driven detections rather than hand-writing signatures. The same tools arm the adversary: AI-assisted phishing is more convincing, AI-assisted vulnerability discovery is faster, and ransomware operators market their capability as a service to affiliates who lack technical skill.

**Operations and reliability.** AIOps platforms correlate events, predict capacity exhaustion, and propose or execute remediations. The operator's role moves from reading dashboards to interrogating explanations.

**Knowledge management.** The most profound change may be the least visible. Retrieval augmented generation over internal documentation turns the knowledge held in the heads of long-tenured staff into something external, queryable and uniform. The risk is symmetrical: staff who stop maintaining documentation because "the AI knows it" degrade the corpus the AI depends upon. Knowledge management becomes an operational discipline rather than a Friday afternoon chore.

**Administration and automation.** Script generation, policy drafting, licence reconciliation and configuration review are all AI-assisted in practice, whether or not the assistance is sanctioned. The work is being done with AI, and the only open question is whether the department governs that fact.

Three observations follow. First, AI's entry point is text-shaped work: anything expressible as a prompt, a ticket, a log line, a policy or a code comment is already touched, and the IT department is unusually text-shaped. Second, the tasks being absorbed are disproportionately the tasks on which junior staff learned; automation removes the legible, well-documented work that constituted the training ground, and in an AI-first department the absorption is near total, because there is no first line to learn on and no routine coding floor to learn from. Third, the aggregate employment effect in advanced economies is large by any serious estimate: the International Monetary Fund puts exposure at around 60 per cent of employment, with roughly half of that exposure more likely to complement the job than displace it. The lesson for an IT leader is not to argue with the forecasts but to plan at task level: which tasks disappear, which transform, which are created, and who decides.

- 1 The shift
- 2 The frame
- 3 The blueprint
- 4 Beyond
- 5 The working

## 4

### TESTABLE COMPONENTS

Jurisdiction, continuity, transparency, agency. Annex A scores each capability 0 to 8.

## 3. The dependency problem

This chapter makes a claim absent from most AI adoption papers: the choice of AI capability is a choice of jurisdiction, and jurisdiction is now a first-order security property.

Consider what a modern AI capability is. A model, certainly, but the model sits inside a stack: infrastructure that hosts inference, an update cadence that changes behaviour without notice, a legal entity that owns the model and its training data, a government with jurisdiction over that entity, and a chain of upstream suppliers whose own dependencies are invisible to the customer. When a European organisation adopts a hosted model from a United States provider, it is placing a component of its decision-making inside a legal regime that includes the United States CLOUD Act, which permits US authorities to compel production of data held by US-controlled providers regardless of where the data physically resides. That is not a software purchase.

The standard response is that the EU-US Data Privacy Framework cures the problem. It does not cure it; it manages part of it, and it has a history. The Framework's predecessors, Safe Harbor and the Privacy Shield, were both invalidated by the Court of Justice of the European Union, in 2015 and 2020 respectively. The current framework survived its initial challenge on 3 September 2025, when the General Court dismissed the annulment action brought by MEP Philippe Latombe, but an appeal is pending before the Court of Justice (case C-703/25 P, lodged on 31 October 2025, with Microsoft admitted as an intervener in June 2026). A framework resting on a foreign president's executive order, with its central provision back before the courts, is not a foundation for a twenty-year architecture. The CLOUD Act and the GDPR sit in unresolved tension: one permits compulsion, the other forbids transfer, and the reconciliation is political arrangement rather than law that either parliament controls.

The jurisdictional question is also not symmetric in the way the discourse assumes. Much of the European conversation frames dependency as a US-versus-China question, with the US positioned as the safe option. For a European organisation, both are foreign. The United States is a close ally with a shared intelligence tradition, and it is also the jurisdiction that pioneered the extraterritorial data access instruments described above and demonstrated in 2013 that its intelligence services bulk-collected allied leaders' communications. None of this is conspiratorial; it is how a sovereign state behaves in its own interest. The error is not in American behaviour but in European accounting, which prices convenience and does not price autonomy.

Dependency also compounds silently. The department adopts a hosted model for summarisation; the model provider's parent company owns the identity platform, the productivity suite and the code repository. Each adoption is individually rational, and the aggregate is a single point of political failure: an entire European organisation whose cognitive infrastructure, collaborative infrastructure and security visibility all answer, in the last instance, to one foreign legal regime. Sovereignty washing makes this worse: providers market European data residency while the control plane, the model weights, the telemetry and the legal entity remain outside European jurisdiction. Residency is not sovereignty. A data centre in Dublin operated under a foreign legal entity is a data centre in Dublin operated under a foreign legal entity.

The method this paper prescribes is simple and unforgiving: map the dependencies, name the jurisdictions, and require justification for each. Appendix A provides the template, with a worked example. For each AI capability in the estate, the department records the provider entity and its

|   |               |
|---|---------------|
| 1 | The shift     |
| 2 | The frame     |
| 3 | The blueprint |
| 4 | Beyond        |
| 5 | The working   |

jurisdiction, where inference occurs, where prompts and outputs are processed and stored, the applicable lawful access regimes, the model provenance and update cadence, the fallback if the service is withdrawn, degraded or legally compelled, and the classification of data the capability may touch. The completed map is the single most valuable artefact an IT department can put in front of its board this year, because it converts an abstraction, "AI risk", into a legible political fact: this is where our autonomy ends.

Two clarifications keep the method credible. First, sovereignty is a spectrum, not a binary. A department that moves from five unexamined foreign dependencies to two examined ones, with a European fallback for each, has made real progress even if it retains foreign components where the trade-off is justified. No organisation can be fully independent, and none needs to be. The goal is that every dependency is chosen, priced and reversible, rather than inherited, unpriced and permanent. Second, the sovereign alternative is now technically real: European inference providers offer hosted open-weight models under EU jurisdiction, and open weights mean the organisation holds the artefact rather than a licence to query it. The correct architecture for most departments is a portfolio: open-weight and self-hosted models for sensitive domains, EU-jurisdiction providers for general workloads, and a documented exception process for anything else.

One security point belongs here: the AI supply chain inherits every pathology of software supply chains, plus new ones. Model weights can be poisoned at training time; fine-tunes can be backdoored; inference endpoints concentrate access to the organisation's most sensitive text in a single service. The department that demanded software bills of materials from software vendors must now demand training data statements, evaluation results, update logs and incident histories from model providers. Chapter 9 covers what to put in the contract.

- 1 The shift
- 2 **The frame**
- 3 The blueprint
- 4 Beyond
- 5 The working

## o-8

### THE SOVEREIGNTY SCORE

7 to 8 sovereign; 4 to 6 conditional, justified; 0 to 3 dependent, fallback mandatory.

## 4. Sovereignty as architecture, not slogan

Sovereignty is the most abused word in the European technology discourse. It is used to mean data residency, though it is not. It is used to mean GDPR compliance, though it is not. It is used on the front pages of sovereign cloud brochures whose control planes sit in another jurisdiction, which is precisely the practice it should describe. This chapter defines the term operationally, because a blueprint needs a definition that can fail, be tested and be audited.

**Definition.** An organisation is sovereign over a capability to the degree that it can continue to exercise that capability, under its own law, on its own terms, if every external party to that capability withdraws, degrades, is compelled or fails. Sovereignty is a property of continuity and control, not of geography. It has four measurable components.

**Jurisdiction.** The capability answers, in the last instance, to which law? The test is not where the data sits but which authority can compel the entities that operate, update and secure the capability. A capability whose provider, its parent and its upstream dependencies are all within EU jurisdiction passes. A capability hosted in Frankfurt by a foreign entity whose home law permits extraterritorial compulsion does not, whatever the brochure says.

**Continuity.** If the provider ends the service, raises the price tenfold, degrades the quality or is legally prohibited from serving the organisation, what happens? A capability with a tested fallback passes; a capability whose loss stops the organisation's decision-making does not. Continuity is why open weights matter: the organisation that holds the weights holds the capability, and can re-host it. A subscription is permission; weights are property.

**Transparency.** Can the organisation inspect what the capability does, how it was built, and how it changes? Open weights, published training data statements, documented evaluations and change logs pass. Black-box services with unannounced model updates do not. Transparency is the precondition of every ethical control in Part Two: you cannot govern what you cannot observe.

**Agency.** Does the organisation control the capability's configuration, deployment boundaries and evolution, or does the provider? Agency fails when the provider can change behaviour, withdraw features or introduce dependencies unilaterally. It passes when the organisation decides where the capability runs, what it may touch, and when it is replaced.

Four components, each testable, together producing a single verdict per capability. The scoring method is in Appendix A, and the honest result for most departments will make poor reading: the majority of their AI estate will score as unexamined and non-sovereign, not because anyone chose dependency but because no one was asked to choose.

A necessary objection must be answered here, because it is made in good faith by competent people. Is sovereignty not expensive, second-best protectionism, a tax on efficiency that buys sentiment? The answer is that sovereignty is a risk management position, priced like any other. Every organisation already accepts inefficiency for resilience: RAID arrays, second suppliers, disaster recovery sites, insurance premiums. Sovereignty is the same category of spend applied to the cognitive layer, justified to the extent that the exposure to compulsion, withdrawal or degradation is real, which the dependency map demonstrates case by case. Where the exposure is trivial, the sovereign premium is not justified, and this paper says so plainly. Where the exposure is severe enough to stop the organisation functioning, the premium is among the cheapest insurance the organisation will ever buy. The error to avoid is failing to price the risk that makes the spend

- 1 The shift
- 2 **The frame**
- 3 The blueprint
- 4 Beyond
- 5 The working

## 5

### COMMITMENTS

Each a testable property; each implemented as a control in Chapter 13.

rational, not spending on sovereignty.

There is also a positive case the risk framing understates. Sovereign architecture is enabling architecture. Open-weight models under European control can be tuned to the organisation's own domain, evaluated against its own cases, and integrated without negotiating a vendor's roadmap. The capability gap between the best open-weight models and the best hosted models has narrowed to the point where, for the great majority of enterprise tasks, the trade-offs are about operations rather than quality. Chapter 10 turns this into concrete patterns.

## 5. The ethical frame

Compliance is the floor, not the frame. An AI-enabled IT department that satisfies every applicable statute can still be unethical, and the ways it manages to do so are predictable enough to design against. Five ethical commitments follow, each defined as a testable property rather than an aspiration, and each implemented as a control in Chapter 13.

**Commitment 1: Human accountability is preserved at the decision, not the process.** It is not sufficient that a human was involved somewhere in the pipeline. The question is whether a human with the authority and the competence to overrule an AI-assisted outcome can and does do so, for the decisions that matter. Accountability has an owner, a named person who answers for each consequential system; where no human could practicably intervene, the system must be classified accordingly, restricted to low-consequence domains, and stated as such to those it affects. What ethics forbids is the fiction of oversight: the human-in-the-loop who exists on the org chart and nowhere else.

**Commitment 2: Those affected are told.** Staff whose work is evaluated, prioritised or shaped by AI systems are informed of that fact, in plain language, before it affects them. Data subjects whose personal data feeds AI processing are covered by the GDPR's transparency duties in any case, but this commitment extends beyond data subjects to the workforce as subjects of algorithmic management. Departments that inform their own people last forfeit the internal legitimacy on which every later control depends.

**Commitment 3: Fairness is measured, not asserted.** Where an AI system differentiates between people, and more IT systems do than their designers admit, the department measures differentials across the groups the system touches. An automated access review that treats different joiners differently on the basis of inferred patterns is a differentiating system; a summarisation tool, in most deployments, is not. The question, "does this system differentiate between people?", is asked at intake, and if the answer is yes, evaluation for disparate impact is mandatory before deployment. The fairness literature is deep and sometimes contradictory between definitions; the department does not need to resolve those disputes, it needs to measure outcomes for the groups it affects and act on what it finds.

**Commitment 4: The environment is not externalised.** Every inference has an energy cost, and the department that measures its printing, its travel and its cloud spend while ignoring its token spend is measuring selectively. Model choices carry a carbon and cost line in the comparison, and reporting on AI use includes consumption alongside capability. Sustainability, in the operational sense this paper uses the term, has four components: models sized to the task rather than maximal by default, because smaller models are cheaper in cost, energy and latency; reuse of shared capabilities rather than duplicated deployments; consumption measured per capability and visible to builders; and lifecycle discipline, retired capabilities actually retired rather than accumulating. The mature position is accounting, not asceticism.

- 1 The shift
- 2 **The frame**
- 3 The blueprint
- 4 Beyond
- 5 The working

### 3

#### STATUTES, MAPPED

The AI Act, the GDPR and NIS2: from article to control to evidence in a single afternoon.

**Commitment 5: Failure is assumed and disclosed.** Models fail confidently. Retrieval systems fabricate provenance. Assistants drift as upstream models update. The ethical department assumes failure as the operating condition and designs for it: evaluation before deployment, monitoring after, a path to rollback, and an incident discipline that treats AI incidents with the same seriousness as security incidents. Internally, the department does not overstate capability to its own organisation; externally, it does not wait to be asked.

Three failure patterns are where AI ethics programmes die. The first is ethics as decoration: the principles document with no owner, no measurement and no veto. The test is simple: can the ethics function stop a deployment? If not, it is decoration. The second is ethics as delay: the review process so heavy that teams route around it, producing the ungoverned shadow AI the process existed to prevent; the remedy is proportionality, which Chapter 8 operationalises through risk tiers. The third is ethics as product feature: the assumption that buying a "responsible AI" branded service discharges the organisation's own duties. It does not. Under the EU AI Act, and under ordinary accountability, the deployer's obligations are the deployer's.

## 6. The governance stack

The European regulatory environment for AI is now dense enough that a department can construct its entire ethical frame from statute, which is both convenient and insufficient: convenient because the instruments below are well drafted and largely aligned, insufficient because law is a floor and Chapter 5 is the frame above it. This chapter maps the instruments to the controls the department must implement, so that Part Three can assign each control an owner and an evidence requirement.

**The EU AI Act.** The Act entered into force on 1 August 2024, with prohibitions and AI literacy obligations applicable from 2 February 2025, general-purpose AI obligations from 2 August 2025, and the main tranche of high-risk obligations from 2 August 2026. Most internal IT use of AI, summarisation, code assistance, service desk triage, sits outside the high-risk categories and instead attracts the Act's general obligations: transparency about AI interaction and, for deployers, the duty to ensure sufficient AI literacy among staff. Systems used in employment decisions, recruitment, evaluation and promotion are high-risk when they are the deployer's own, or where the department configures a provider's system in ways that make it the provider. The department's duties as a deployer of high-risk systems include human oversight by competent persons, use in accordance with instructions, input data relevance, logging, and information duties to affected workers. None of these is optional, and the penalties for the worst breaches scale to a percentage of global turnover.

**The GDPR.** The GDPR was regulating AI before the term had a marketing department. Article 22 restricts solely automated decisions with legal or similarly significant effects; Articles 13 to 15 require transparency; Article 25 requires data protection by design and by default; Article 35 requires a data protection impact assessment for high-risk processing, and the European Data Protection Board's 2024 opinion on AI models directs that DPIAs cover model use and training data. Every AI system touching personal data requires a lawful basis, a purpose specification, a retention rule for prompts and outputs, a transfer analysis where inference occurs outside the EEA, and a DPIA proportionate to risk. Appendix D provides the structure.

**NIS2.** The NIS2 Directive, in force since January 2023 with a transposition deadline of 17 October 2024, extends mandatory cyber risk management to a wide set of sectors, and its Article 20 supply chain duties bear directly on AI procurement: the department may not outsource the security accountability of an AI capability to its provider. Implementation across

|   |                  |
|---|------------------|
| 1 | The shift        |
| 2 | <b>The frame</b> |
| 3 | The blueprint    |
| 4 | Beyond           |
| 5 | The working      |

member states remains uneven. On 8 July 2026 the Commission referred Ireland, Spain and France to the Court of Justice for failing to notify full transposition; the Netherlands was referred in the same week, days before its own transposition law entered into force on 15 August 2026. The unevenness of enforcement does not soften the obligations of organisations already in scope. For AI, NIS2's most consequential effect is the obligation to treat model and inference suppliers as part of the assessed supply chain.

**Adjacent instruments.** The Cyber Resilience Act places security duties on products with digital elements from December 2027. The Data Act, applicable from September 2025, governs data access in connected products and cloud switching. The EUIDS regulation becomes relevant as AI systems authenticate users and agents. For UK readers, the UK GDPR and the Data Protection and Digital Information legislation occupy the privacy space, and the UK's pro-innovation framework operates through sector regulators rather than a single statute. A department operating across both jurisdictions should implement to the stricter instrument and document the mapping.

**From statute to control.** The mapping is deliberately mechanical. For each instrument, the department maintains a table of obligations, and for each obligation a control, an owner and an evidence artefact. The transparency duties of the AI Act become the disclosure register in Chapter 8. The AI literacy duty becomes the training matrix in Chapter 12. Article 22 of the GDPR becomes the decision classification scheme in Chapter 13. NIS2 supply chain duties become the procurement gates in Chapter 9. The value of the mechanical mapping is auditability: when the regulator arrives, the department can walk from the article to the control to the evidence in a single afternoon.

A final point on regulatory posture. The tempo of European AI regulation is accelerating, not settling. A department that builds to today's text alone will rebuild every eighteen months. The durable approach is to build to the principles the texts share, human oversight, transparency, proportionality, accountability, documentation, and treat each statute as a specific instantiation. Principles survive redrafts. Compliance programmes built to the letter do not.

- 1 The shift
- 2 The frame
- 3 **The blueprint**
- 4 Beyond
- 5 The working

## 3

## PLANES

Automation, supervision, governance. Logical, not organisational.

## 7. The target operating model

The blueprint begins with the shape of the department, because structure determines what governance can reach. The target operating model has three planes: the automation plane, where AI does the work; the supervision plane, where humans direct, review and overrule the automation; and the governance plane, which owns the rules, the records and the accountability. Most departments already have an implicit automation plane and no explicit supervision plane. The blueprint makes all three deliberate.

**The automation plane** is the estate of AI capabilities: the code assistant in the engineering toolchain, the triage models in the service desk, the summarisation and retrieval layer over documentation, the AIOps correlation in monitoring, the drafting support in security operations. Each capability is registered, tiered, bounded to a data classification, and equipped with the controls its tier requires. Nothing runs unregistered; Chapter 8 makes registration cheap enough that no one needs to evade it.

**The supervision plane** is the human layer that is being rebuilt, not retired. It has three roles that did not exist in most departments five years ago.

The AI operations engineer runs models the way a platform engineer runs infrastructure: hosting, upgrading, evaluating, monitoring, cost and capacity management. The skill set combines infrastructure operations with evaluation methods and enough statistics to know when a metric lies.

The AI risk and governance lead owns the register, the tiering, the DPIAs, the disclosure register and the relationship with the Data Protection Officer. This is a governance professional with technical literacy, not a data scientist, and the role reports sufficiently high that its veto is real.

The evaluator, often a part-time rotation rather than a post, builds and runs the evaluation sets that gate deployment and detect drift: senior engineers, service managers and security analysts who know what "good" means in the work, because evaluation cannot be entirely outsourced.

The governance plane is the subject of Chapter 8.

**Scale, and the enabler mandate.** Two further properties of the target model distinguish it from its predecessor, and both are demonstrated in the author's practice rather than merely proposed.

The first is scale. The model is calibrated for small teams, because AI has removed the correlation between team size and service capability. A department of fifteen people, fully remote, serving a global organisation across more than forty countries, can run the three planes if each plane is staffed by multiplicity rather than specialisation: the same senior engineer operates in the supervision plane on Tuesday and the automation plane on Thursday; the governance lead is a fraction of one person's week supported by automation rather than a committee supported by administrators. The planes are logical, not organisational. What cannot be shrunk is the discipline: a small team that skips the register, the tiering or the evaluation gates has no margin for the failures those controls prevent. Small scale makes the blueprint cheaper to run, not optional to run.

The second is the enabler mandate, and it is the largest single change in the department's relationship to the rest of the organisation. In the previous model, the department was a factory: requests arrived, the department built, and the queue was the interface. In the target model, the department is a platform and a guardrail service: internal business teams build their own AI systems, for their own domains, on the department's sanctioned patterns, tools and data boundaries, and the department's product is the safe space in which they do it. Empowering internal teams to create AI systems is not the abdication Chapter 1 condemned; it is the opposite, because the

- 1 The shift
- 2 The frame
- 3 **The blueprint**
- 4 Beyond
- 5 The working

empowerment is delivered inside the governance frame: registered capabilities, tiered assessment, sanctioned data boundaries, reusable patterns and, above all, the ethical and sustainability standards that make what internal teams build legitimate rather than merely functional. The department that tries to build every AI system itself becomes the bottleneck and loses to shadow AI. The department that makes governed building easier than ungoverned building becomes indispensable at a fraction of its former size, and its experts multiply through other people's work.

**Sustainability as a first-class constraint.** Because internal teams build on the department's platform, the department's design choices propagate. The four sustainability components of Chapter 5, right-sized models, reuse, consumption visibility and lifecycle discipline, are delivered as platform defaults rather than exhortations: the pattern library steers builders to the smallest sufficient model, the shared capability catalogue makes duplication visible, consumption is metered per capability and reported to the builder's owner, and the register review retires what nobody uses. The enabler mandate makes the department the custodian of these standards across everything the organisation builds, which is a larger environmental and economic footprint than the department's own estate would ever be.

**The functional map.** Service management moves from reactive ticket handling to experience design. Tier 0 is the governed assistant layer: instant, grounded in the knowledge corpus with provenance citations, handling the routine volume of access requests, how-do-I questions, status and account lifecycle. In the author's practice this is the settled state, not the ambition: first line and triage are substantially AI-operated, and the small team that remains does not stand behind a wall of tickets; it stands behind the assistant's quality. Escalations arrive at the human tier with structure: the assistant has gathered the context, categorised the request, proposed a resolution where it can, and recorded what it tried, so the expert does not start from nothing. Human work concentrates on exceptions, relationships and the cases the assistant mis-routed, which are precisely the cases that teach. The service model this produces deserves naming precisely, because it is the visible proof of the paper's central claim. A conventional department of this organisation's size and geographic spread would carry a first line team of meaningful headcount, a triage function, and the supervisory overhead that comes with both. The AI-first model carries neither. What it carries instead is a small group of senior experts, each supercharged: every one of them operates with AI knowledge at their shoulder, AI speed in their tooling and AI assistance in their workflow, so that the department's effective capability is the expert's judgement multiplied rather than the headcount added. The economics are not a cost reduction programme; they are a capability compounding programme, and the difference matters to the workforce, as Chapter 12 argues.

The assistant's quality ceiling is the corpus, so documentation inverts its traditional economics: the corpus is the model of the service, its curation is the most consequential investment in service quality, and every escalation ends in a corpus update decision. Engineering moves from writing code to specifying and validating it, with code review discipline, adversarial testing and domain architecture rising in status. Security engineering moves from signature authorship to detection curation, with humans owning the decision to act on what the models surface. Infrastructure and platform work absorbs the AI operations function; data management absorbs knowledge curation and the evaluation data pipelines. Security becomes larger, not smaller, because the department's AI estate is itself a target and its models are software dependencies with all the pathologies that category implies. Governance grows from a documentation function into a control function; in a department of any size it is at least one role, held within the governance lead even at a fraction of their week.

- 1 The shift
- 2 The frame
- 3 **The blueprint**
- 4 Beyond
- 5 The working

## 4

### RISK TIERS

Minimal, standard, consequential, restricted. Tier 4 is held empty on principle.

**Design principles.** Five principles govern the model. Proportionality: the weight of governance follows the consequence of the system, so low-risk tools move fast and high-risk ones move safely. Reversibility: every capability has a tested fallback, because an untested fallback is a hope, not a control. Least data: each capability is bounded to the minimum data classification that lets it work, and the default for anything touching personal or commercially sensitive data is the sovereign pattern of Chapter 10. Human decision rights: for each consequential decision class, the model states who may overrule, in what timeframe and on what evidence. One register: there is a single source of truth for the AI estate, and every governance artefact keys to it. Departments that allow a second, informal estate have given shadow AI a filing system.

**Measuring the new service.** Traditional metrics, first contact resolution, ticket volumes, break the model's meaning because the ticketing boundary has moved. The replacement set, defined in Chapter 16, centres on deflection quality and escalation contamination: the rate at which escalations contain assistant errors, the single most important number in the model.

**The transition discipline.** The service desk is where the workforce question is answered first, and the answer must be designed before the assistant is deployed, not after. The blueprint requires, for any capability that reduces workload in a team: a written statement of what happens to the liberated capacity, review with the affected staff before deployment, redeployment or development pathways defined, and the apprenticeship question answered explicitly: which roles will teach new staff the skills this team no longer uses daily, and how. Departments that skip this discipline keep the productivity and lose the credibility, and credibility, once spent on the workforce, does not return at the next announcement.

## 8. Governance structures and decision rights

Governance fails in two directions: it is either theatre, principles without power, or bureaucracy, power without proportion. The structure below avoids both by keeping the machinery small, tiered and attached to money.

**The AI register** is the spine. Each entry records: name and owner; purpose in plain language; tier; model and provider, with jurisdiction and hosting; data classifications permitted; decision classes affected; human oversight arrangement; evaluation status and date; fallback; DPIA reference; and lifecycle state. Appendix B provides a worked entry. The register is reviewable in minutes and comprehensible to an auditor in an hour. Every other artefact, DPIA, model card, evaluation report, contract clause, references its entry.

**Risk tiering.** Every capability is placed in one of four tiers at intake, and the tier determines everything that follows: the depth of assessment, the evidence required, the approval needed and the review cadence.

Tier 1, minimal: personal productivity assistance on non-sensitive data, summarisation of public or low-classification material, code assistance on non-production artefacts. Requirements: register entry, acceptable use acknowledgment, baseline training. Approval: line manager. Review: annual. The lightness of this tier is a security feature, not a convenience; if the fast lane moves slowly, the shadow estate grows.

Tier 2, standard: capabilities touching internal data or personal data below the high-risk threshold, knowledge retrieval over corporate documentation, service desk triage, AIOps analytics. Requirements: tiered data boundary, evaluation before deployment, monitoring, DPIA where personal data is

- 1 The shift
- 2 The frame
- 3 **The blueprint**
- 4 Beyond
- 5 The working

involved, named accountable owner, fallback documented. Approval: AI governance lead with the service owner. Review: semi-annual.

Tier 3, consequential: capabilities whose outputs influence decisions about people, security posture, access or money: automated access reviews, candidate screening support, AI-assisted change authorisation, agent systems with write access to production. Requirements: everything in Tier 2 plus disparate impact evaluation where the system differentiates between people, human override design with named decision owners, logging sufficient to reconstruct decisions, red-team exercise, board-visible risk acceptance. Approval: governance board. Review: quarterly.

Tier 4, restricted: capabilities that decide without practicable human intervention, process special category data, or operate in domains the organisation has declared off limits, including biometric categorisation, emotion inference in the workplace, and any social scoring application, which are prohibited or heavily restricted under the AI Act. Entry requires exceptional justification and executive approval, and most departments should hold the tier empty on principle.

**The governance board** is a small body, meeting monthly, chaired by the CIO or IT director, with the AI governance lead, the DPO, the security lead, HR where workforce systems are in scope, and one business representative whose job is to say when the process is strangling delivery. Its powers: tier assignment, Tier 3 approvals, suspension authority over any capability, and escalation to the executive board. Its discipline: no capability is discussed twice without a decision, and every decision is minuted against a register entry.

**Decision rights** are fixed explicitly, because ambiguity is where accountability dies. The AI governance lead may suspend any capability immediately, with appeal to the board. The security lead may isolate any capability that presents an active threat, with notification within one working day. The DPO holds a veto over any processing without a lawful basis. Owners may self-assess at Tier 1 but not promote. The executive board owns Tier 4 and the annual risk acceptance statement. And the register, not the meeting, is the record: if it is not in the register, it does not exist for governance purposes.

**The acceptable use policy** is the workforce-facing instrument, and the most read governance artefact the department will produce. It must answer, in one page: what staff may use AI for, what data may never enter any AI system, which tools are sanctioned at which tier, what disclosure is owed to colleagues and customers, and what to do when something goes wrong. Appendix C provides a skeleton. The policy is written to be enforced, which means it is written to be followed: any rule that staff cannot comply with in practice will be evaded in practice, and the evasion will be invisible.

## 5

### GATES

Jurisdiction, transparency, data boundaries, continuity, security. The score sits beside the price.

## 9. Procurement and vendor sovereignty

Procurement is where sovereignty is won or lost, because it is where the dependency map of Chapter 3 acquires its entries. The blueprint imposes five gates on any AI purchase, whether a model, a platform or a feature embedded in an existing tool. A purchase that passes all five gates may proceed; one that cannot is either rejected or escalated as an accepted risk with a fallback plan.

**Gate 1, jurisdiction and lawful access.** The provider discloses the operating entity, its jurisdiction, the parent chain, where inference occurs, where prompts, outputs and telemetry are processed, and which lawful access regimes apply. The department records this in the register, and for any non-EU component produces a written justification and a fallback. The test question is the one from Chapter 4: if this provider were compelled

- 1 The shift
- 2 The frame
- 3 **The blueprint**
- 4 Beyond
- 5 The working

tomorrow, what would we still be able to do?

**Gate 2, transparency of the model.** The provider discloses model provenance, whether weights are open, the training data statement at least at the level the AI Act requires of providers, the evaluation evidence, the update policy and the change notification terms. Silent model updates are a red flag: behaviour that can change overnight is behaviour the department has not assessed. Where the provider will not disclose, the capability is treated as unauditable, and unauditable capabilities do not touch sensitive data.

**Gate 3, data boundaries.** The contract must state, enforceably: what data is retained, for how long, whether it is used for training (the answer must be no for anything the organisation puts in), the subprocessor list, the transfer mechanism if outside the EEA, and the deletion terms at exit. The department runs prompt and output logging on its side for Tier 2 and above, because the provider's telemetry is not the department's evidence.

**Gate 4, continuity and exit.** The contract includes exit provisions: data export in usable formats, notice periods, and, for critical capabilities, the right to continue operating during transition. The architectural preference reduces the need for this gate: a capability built on open weights exits by re-hosting, not by renegotiation. This gate ensures the procurement path at least prices the risk.

**Gate 5, security and resilience.** The provider supplies security certifications, incident history, the software bill of materials for the inference stack where available, and commitments on vulnerability disclosure and incident notification within defined windows. NIS2's Article 20 makes clear that the department may not outsource this accountability: the supplier's certification is evidence, not discharge.

**Sovereignty scoring in procurement.** Each candidate capability is scored against the four components of Chapter 4, and the score appears on the same page as the price. Procurement that sees only price will buy dependency every time, because dependency is cheaper at the moment of purchase. The score makes the hidden cost a visible line.

The blueprint does not require that every capability be European or self-hosted. It requires that the sensitive ones are, that the others are justified, and that each critical capability has a tested sovereign fallback. A department that cannot list its AI providers without checking is not sovereign, whatever its cloud provider's marketing says.

## 7

### PATTERNS

Each stated as a problem, a pattern and a named failure mode.

## 10. Technical architecture patterns

This chapter provides the reference patterns. Each is stated as a problem, a pattern and a failure mode, because a pattern that cannot fail cannot be evaluated. The chapter closes with the security operations dimension, which inherits the same estate.

**Pattern 1: Sovereign inference tier.** Problem: the organisation needs model capability for sensitive domains: contracts, HR cases, security incidents, executive material. Pattern: a self-hosted or EU-provider-hosted open-weight model on infrastructure under EU jurisdiction, integrated under the organisation's own identity and logging. Data never leaves the boundary; the weights can be pinned to a version; evaluation runs against the organisation's own cases. Failure mode: capacity and cost. Inference is cheap at small scale and expensive at organisation-wide scale, so the tier is sized to the sensitive domains, not to general consumption.

**Pattern 2: Routed portfolio.** Problem: a single provider relationship recreates the concentration the dependency map condemns. Pattern: an abstraction layer that routes requests by task, data classification and cost to the

|   |                      |
|---|----------------------|
| 1 | The shift            |
| 2 | The frame            |
| 3 | <b>The blueprint</b> |
| 4 | Beyond               |
| 5 | The working          |

appropriate backend: the sovereign tier for sensitive data, an EU-jurisdiction general service for routine work, justified exceptions under Gate 1 for specialised capability. Routing policy is code, reviewable and testable. Failure mode: routing by convenience. If the default route is the best model rather than the appropriate one, classification boundaries erode.

**Pattern 3: Retrieval-grounded generation with provenance.** Problem: models fabricate, and a fabricated answer inside an IT workflow becomes a confident lie with an institutional voice. Pattern: every knowledge-facing assistant is grounded in a curated corpus with retrieval, and every answer carries citations to the source documents it used. The corpus becomes the governed artefact: owned, versioned, reviewed. Failure mode: corpus decay. If the assistant is tolerated because it is mostly right, quality erodes silently until it fails loudly; provenance citations let users check, and checking users are the best decay detector the department can deploy.

**Pattern 4: Agent containment.** Problem: agents that act, not merely answer, are the sharpest new risk surface; an agent with production write access and a prompt injection vulnerability is an attacker with legitimate credentials. Pattern: every agent operates with least privilege, scoped credentials, human approval gates for consequential actions, full action logging, and a kill switch owned by the supervision plane. Agents are treated as privileged accounts, with the same joiner-mover-leaver discipline as staff. Failure mode: scope creep, as "temporary" broad access becomes permanent; the register review catches this.

**Pattern 5: Evaluation as infrastructure.** Problem: model behaviour changes silently with updates, fine-tunes and data drift, and the department discovers regressions from incidents rather than from tests. Pattern: for every Tier 2 and above capability, a versioned evaluation set, real tasks and expected outcomes, measured automatically at every model change and on a schedule. The evaluation suite is CI for models, and its results gate deployment. Failure mode: vanity sets. A suite that only contains cases the model passes is a compliance artefact, not a control; it must contain the known hard cases and the previous failures.

**Pattern 6: Telemetry and audit fabric.** Problem: accountability requires reconstruction; the department must be able to answer, for a consequential AI-influenced decision, what the system was asked, what it produced, what the human did and when. Pattern: prompt and output logging at Tier 2 and above, stored under the organisation's own control, linked to the register entry, with retention set by the DPIA and the acceptable use policy. Failure mode: logging everything forever. Logs of prompts are personal data, and unbounded logging converts an accountability tool into a liability.

**Pattern 7: The human override channel.** Problem: the fiction of oversight described in Chapter 5. Pattern: for every consequential decision class, the override is designed, not assumed: the interface surfaces what the system proposed, the evidence, the confidence and the cost of delay, and the human's disposition is recorded. Override rates are monitored; a decision class where overrides never occur is not a well-calibrated system, it is an unaudited one, and it is investigated as such.

**Security operations in the AI-enabled department.** Security is the function where AI's dual nature is most visible: the same capability that suppresses false positives at machine scale also arms the adversary. Securing the estate means treating six asset classes, each with a distinct threat model. Prompts and outputs contain secrets and personal data, and they transit to wherever inference occurs; the controls are classification-aware routing from Pattern 2, data loss prevention on AI channels, and the logging discipline of Pattern 6. Models and weights are the software supply chain of the estate; the controls are provenance verification, signed releases, pinning, and scanning for known backdoor families. A model from an unvetted registry is an unvetted binary with a friendly interface. Inference

- 1 The shift
- 2 The frame
- 3 The blueprint
- 4 Beyond
- 5 The working

infrastructure is a concentration of the organisation's most sensitive text in one service, and it gets the same operations security as any Tier 1 system. Agents inherit Pattern 4 in full. Telemetry and logs are both a detection asset and a privacy liability. People remain the most attacked class: prompt injection, AI-generated lures, and the recruitment of insiders via synthetic personas; the control is training that shows real fabricated artefacts, and a culture that treats "the assistant told me to" as an escalation trigger, not an authorisation.

Operating security with AI follows the same discipline. Model-driven triage is now table stakes; the human role is curation, tuning and investigation. Staff-facing security analytics that differentiate between employees, anomaly scoring, insider risk ranking, are consequential systems, Tier 3, with measured impact and override design. AI assistance in containment drafting, evidence summarisation and playbook execution is valuable and safe when the human owns the execution decision. AI-assisted threat intelligence carries a provenance check, because fabricated intelligence is worse than none.

Two constraints deserve emphasis. The first is sovereign: the department that routes its security telemetry through the same foreign provider that supplies its models, identity and productivity suite has concentrated its visibility into that infrastructure inside one foreign jurisdiction. Security telemetry is sensitive-domain data; it belongs in the sovereign tier. An organisation that can be blinded by one provider's withdrawal, or one jurisdiction's compulsion, in the same week that it is being attacked has designed a target, not a defence. The second is adversarial: assume the adversary uses the best available models for reconnaissance, phishing generation, deepfake voice and video, and automated vulnerability discovery. Authentication moves toward phishing-resistant factors, because credential phishing becomes trivially convincing. Verification protocols move toward pre-agreed challenge phrases and out-of-band confirmation for high-consequence instructions, because voice and video are no longer evidence of identity. The department that updated its incident response plan for ransomware in 2021 must update it for synthetic impersonation now, or explain to the board why it did not.

---

## Art. 22

### GDPR

Solely automated decisions restricted. The DPIA structure is Annex D.

---

## II. Data protection by design and default

Every AI capability is a data processing operation, and the IT department is both a controller for its own processing and the operator of the estate through which the whole organisation's AI use flows. This chapter sets the data protection discipline for that dual position, and Appendix D provides the assessment structure that implements it.

**The inventory problem.** AI capabilities escape registration by the data protection function because their data flows are invisible: the service desk assistant reads tickets containing personal data; the code assistant ingests source code containing names, credentials and business logic; the AIOps platform processes the behavioural traces of every employee. The AI register of Chapter 8 solves this by design, because its fields are a processing inventory: what data, what purpose, what recipient, what retention, what transfer. The register and the record of processing activities should share identifiers, so that the data protection function and the AI governance function describe one estate, not two.

**Prompts are personal data.** A prompt containing a staff member's name and a request about their access rights is personal data about them; a log of prompts is a behavioural record of the prompter. Retention for prompt and output logs is a DPIA decision, not a storage default. Access to those logs is restricted on data protection grounds. Subjects can make access requests that reach AI logs, and the department must be able to answer them. And

- 1 The shift
- 2 The frame
- 3 **The blueprint**
- 4 Beyond
- 5 The working

the anonymisation claim, often made casually about telemetry, must withstand scrutiny: logs stripped of names but linked by account identifiers are not anonymous.

**Transfers.** Where inference occurs outside the EEA, the transfer analysis runs through the GDPR's mechanisms, and the honest version of it includes the Chapter 3 caveat: adequacy decisions rest partly on political instruments with a track record of collapse. The architectural position, sensitive domains on the sovereign tier, removes most of the analysis by removing most of the transfers. Where transfers persist, they are documented per capability in the register and re-reviewed when the political instruments change, not merely when the law does.

**DPIA discipline.** A DPIA is mandatory for AI processing likely to result in high risk to individuals, and it is good practice for every Tier 2 and above capability. The DPIA answers, in order: the processing and its necessity and proportionality; the risks to rights and freedoms, including the specific AI risks, fabrication, bias, re-identification, function creep, over-reliance; the mitigations, drawn from the pattern set of Chapter 10; and the residual risk decision, taken by a named accountable person. A DPIA that concludes "low residual risk" on every assessment is a signature, not an assessment.

**Confidentiality of business data.** The GDPR is not the only protection discipline. Client contracts, commercially sensitive terms, merger material and security findings carry obligations that are not personal data but are, in some ways, more exposed: there is no supervisory authority to complain to after a trade secret enters a provider's training corpus. The acceptable use policy states the classification boundaries, and the sovereign tier exists so that the statement can be true.

**The DPO relationship.** In the target operating model, the AI governance lead and the DPO are distinct roles with a defined interface: the register, the DPIA queue and the incident channel. This separation is deliberate. The DPO advises and monitors independently; the governance lead owns decisions within the risk appetite. Departments that merge the roles gain efficiency and lose the independence that makes the DPO's advice worth having.

## 5

### ROLE FAMILIES

Builders, enablers, supervisors, governors, translators. Hats, not posts.

## 12. Workforce: the reformation of roles

Every chapter so far constrains technology. This one constrains management, because the workforce transition is where AI programmes earn or destroy their legitimacy. The blueprint treats the following as requirements, not aspirations, and Chapter 15 sequences them.

**The role map.** Five role families define the target department, and in a small AI-first team they are worn as hats rather than held as posts: the same fifteen or twenty people occupy all five at once, moving between them by week and by task. This multiplicity is not a compromise forced by small scale; it is the natural shape of the AI-enabled department, in which the routine tier that used to absorb junior specialists no longer exists to be staffed.

**Builders:** engineers who specify, review and validate AI-generated work; AI operations engineers who run the model estate; data and knowledge curators who own the corpora. In the author's practice, development is largely AI-generated and the builder's craft has moved decisively to specification, review and adversarial validation. These roles absorb most of the growth.

**Enablers:** the newest family, and the one the conventional model lacks entirely. These are the experts who equip internal business teams to build their own AI systems: pattern authors who turn the Chapter 10 set into reusable, sanctioned templates; coaches who teach the literacy the matrix

|   |                      |
|---|----------------------|
| 1 | The shift            |
| 2 | The frame            |
| 3 | <b>The blueprint</b> |
| 4 | Beyond               |
| 5 | The working          |

below defines; reviewers who gate what internal teams build without building it for them; and curators of the shared capabilities that make sustainable reuse the default. The enabler's success metric is the number of governed systems the organisation runs that the department did not build.

Supervisors: the human oversight layer; decision owners in consequential classes, evaluators who build and run the gates, service experts who own the assistant layer's quality. These roles are staffed from the department's most experienced people, deliberately, because supervision without domain judgement is decorative.

Governors: the AI governance lead, the policy owner, the DPO interface, the audit and evidence function. Small in number, disproportionately large in consequence. In the small-team model this is a fraction of one senior person's week, made possible by the automation of the register, the tiering workflow and the evidence pipeline; governance that requires a bureaucracy is governance a small team will not sustain.

Translators: the people who carry AI capability into the business and the business's needs back into the department: prompt and workflow designers, trainers, the internal advocates who make the acceptable use policy a living instrument rather than a laminated one. In most departments this is a capability distributed across roles rather than a team.

The defining property of all five families is the supercharging relationship between the expert and the machine. The department does not employ AI to do the experts' work, nor experts to babysit the AI's work; it employs AI as a multiplier on each expert's judgement, knowledge and speed, so that one properly equipped senior person delivers what a small team used to, and a small team delivers what a department used to. Recruitment, development and retention all follow from that property: the department hires for judgement and domain depth, trains for AI fluency, and measures experts by the quality of what their machine-multiplied practice produces, not by hours of manual execution avoided.

**The apprenticeship problem, restated as a design.** Chapter 2 identified the mechanism: automation absorbs the junior tasks that constituted the training ground, and in the AI-first department the absorption is near total, because there is no first line to learn on and no routine coding floor to learn from. The answer has three parts. Entry roles are redesigned around the new training ground: evaluation, curation, assistant supervision and exception handling, which are junior-doable and senior-formative, because they teach how the systems fail. Deliberate rotation puts every early-career engineer through the supervision plane, so that by seniority they have judged, corrected and overridden systems they will later be trusted to govern. And the senior engineers who still hold the absorbed skills are recorded as knowledge assets before they leave, because corpus curation is the mechanism by which their expertise outlasts their departure. The department that hires juniors and gives them nothing but prompts to write has not solved the apprenticeship problem; it has outsourced it to the model provider, and the provider does not return the calls.

**The transition protocol.** For any capability that reduces workload in a team, the following applies: a written statement of what happens to the liberated capacity, agreed before deployment; consultation with affected staff before the change, not after the announcement; defined pathways, redeployment into supervision, curation or security roles being the natural ones in this model; and honest internal communication about intent. Departments that promise "no redundancies, only efficiency" and then restructure in month eighteen lose the workforce's cooperation for a decade, and they deserve to.

**AI literacy as a legal duty and a survival skill.** The AI Act obliges deployers to ensure a sufficient level of AI literacy among staff, and the blueprint treats this as a capability programme with a training matrix: baseline literacy for all,

- 1 The shift
- 2 The frame
- 3 **The blueprint**
- 4 Beyond
- 5 The working

## 12

### CONTROLS

Requirement, evidence, owner, tier. Deliberately few.

role-specific modules for builders, enablers, supervisors and governors, refresh at each major model or policy change, and evidence of completion held against the register. Literacy means knowing what the system does, where it fails, what data it may touch and how to escalate when it misbehaves. Departments that run literacy as a compliance video have satisfied neither the statute nor the risk.

**The human questions.** Three questions recur in every honest consultation, and the answers are stated plainly. Will this reduce headcount? In the operational core, over time, probably; the department owes its staff transparency about that and investment in their transition rather than euphemism. Will my skills still matter? Yes, differently: judgement, oversight and domain knowledge appreciate while routine execution depreciates. Who is accountable when the system errs? The named owner on the register, always, and the workforce's trust in the department will be built or broken by whether that answer survives the first serious incident.

## 13. Ethics in practice: the control set

Chapter 5 defined five commitments. This chapter defines the control set that implements them, in the format of every other control framework the department runs: control, requirement, evidence, owner, tier.

C1, register completeness: every AI capability in operational use appears in the register with current fields. Evidence: register extract; annual attestation by service owners; shadow AI discovery sweep of network and SaaS logs. Owner: AI governance lead. Tiers: all.

C2, tiered assessment: each capability is tiered before deployment and at each review. Evidence: tier record with rationale. Owner: governance board. Tiers: all.

C3, named accountability: each capability names an accountable owner; each consequential decision class names a decision owner with authority to overrule. Evidence: register fields; decision logs. Owner: governance board. Tiers: 2 and above.

C4, disclosure to those affected: staff and data subjects are informed, in plain language, when consequential AI systems affect them, before or at the time of effect. Evidence: disclosure register; communications records; privacy notices. Owner: AI governance lead with the DPO. Tiers: 2 and above; mandatory at 3.

C5, evaluation gating: no capability enters or changes in production without passing its evaluation set. Evidence: evaluation reports, versioned sets, pass thresholds. Owner: evaluator of record. Tiers: 2 and above.

C6, human override: consequential decision classes implement Pattern 7: proposal, evidence, confidence and disposition surfaced and recorded; override rates monitored; zero-override classes investigated. Evidence: decision logs; quarterly override review. Owner: decision owners. Tiers: 3.

C7, impact measurement: differentiating systems are measured for disparate outcomes across affected groups before deployment and semi-annually. Evidence: measurement reports; remediation records. Owner: AI governance lead with HR. Tiers: 3.

C8, incident discipline: AI incidents, errors with effect, misuse, data leakage, harmful outputs, are reported through the security incident channel, classified, investigated and fed back into evaluation sets and the corpus. Evidence: incident records; post-incident reviews. Owner: security lead. Tiers: all.

C9, model supply chain assurance: models and inference components are

- 1 The shift
- 2 The frame
- 3 **The blueprint**
- 4 Beyond
- 5 The working

## 5

### LEVELS

Unaware to sovereign and adaptive. Most who believe they are advanced are at Level 1.

acquired through the Chapter 9 gates, versioned, verified and pinned. Evidence: provenance records; checksums; update logs. Owner: AI operations engineer. Tiers: 2 and above.

C10, sovereignty maintenance: the dependency map is refreshed annually and on any material change; fallbacks are tested annually. Evidence: dependency map; fallback test records; board risk acceptance. Owner: CIO. Tiers: 2 and above.

C11, acceptable use enforcement: the policy is published, trained, acknowledged and technically supported: sanctioned tools are easier to use than unsanctioned ones. Evidence: acknowledgment records; tool availability; sweep findings. Owner: AI governance lead. Tiers: all.

C12, consumption accounting: cost and energy consumption per capability are recorded and reported with capability metrics. Evidence: consumption report. Owner: AI operations engineer. Tiers: 2 and above.

Twelve controls, deliberately few. A control set the department cannot run is a risk document, not a control set, and the maturity model in the next chapter measures whether the set is real. Appendix G turns each level of that model into an assessment checklist.

## 14. Maturity model and assessment

Departments need a yardstick. The maturity model has five levels, each defined by observable conditions rather than sentiment, so that two assessors reach the same answer.

Level 0, unaware: no register, no policy, no owner. AI use is individual and invisible. Evidence of the level: no one can answer how many AI capabilities operate in the estate.

Level 1, reactive: a policy exists; some tools are sanctioned; adoption is tolerated rather than governed. Shadow use is acknowledged but unmapped. The department will describe itself as "working on AI governance", which is a Level 1 signature.

Level 2, governed: the register exists and is mostly complete. Tiering operates. The acceptable use policy is published and trained. Tier 2 controls run: evaluation gates, DPIAs, named owners. A governance lead exists, at least part-time. The department can enumerate its AI estate and its data boundaries.

Level 3, supervised: the supervision plane is real: consequential classes have decision owners and override design; evaluation runs on schedule and at every change; incidents feed the sets; the dependency map exists and fallbacks are tested. The workforce programme runs: role map, transition protocol, literacy matrix.

Level 4, sovereign and adaptive: the routed portfolio operates; sensitive domains run on the sovereign tier; procurement gates are institutional; the board receives and accepts the annual dependency and risk statement. The department's AI estate is an asset the organisation could re-host, not a subscription it could lose.

**Using the model.** Assessment is annual, by the governance lead, evidenced against the control set, and reported to the board with the roadmap position. Most departments that believe they are advanced are at Level 1, and the gap between Level 1 and Level 2 is the largest single risk reduction in the entire model: it is the difference between an estate the department can see and one it cannot. Appendix G provides the per-level checklist an assessor works through.

- 1 The shift
- 2 The frame
- 3 The blueprint
- 4 Beyond
- 5 The working

24

MONTHS

Four phases, each with entry criteria and exit evidence.

# 15. The twenty-four month roadmap

The roadmap sequences the blueprint into four phases. Each phase has entry criteria, workstreams and exit evidence, because a roadmap without exit evidence is a wish list.

**Phase 1, months 0 to 3, seeing.** Establish the register: inventory all known AI use, sanctioned and suspected; run the first shadow AI sweep; draft the dependency map for the top capabilities by data exposure; publish an interim acceptable use policy with clear boundaries and a disclosure route; appoint the AI governance lead, even at a fraction of a role; stand up the governance board with its first agenda: the register, the tiers, the interim policy. Exit evidence: a register covering the known estate; an interim policy in force; a named lead; a board minute accepting the programme.

**Phase 2, months 3 to 9, governing.** Implement tiering and the control set for Tiers 1 and 2; run DPIAs on all Tier 2 capabilities touching personal data; build the first evaluation sets for the highest-volume capabilities, typically service desk and knowledge retrieval; publish the full acceptable use policy with the literacy programme behind it; begin the workforce programme: role map consultation, transition protocol, first rotations into supervision and enabler roles. Exit evidence: complete Tier 1 and 2 control operation; literacy delivered; first evaluation gates live; workforce consultation documented.

**Phase 3, months 9 to 18, building.** Deploy the sovereign inference tier for the sensitive domains; implement the routed portfolio; implement Patterns 4, 6 and 7 as agent, logging and override needs arise; apply the procurement gates to every renewal and new purchase; extend evaluation to all Tier 2 and above capabilities; first fallback tests; first disparate impact measurements on Tier 3 systems; stand up the enabler platform: pattern library, capability catalogue and builder onboarding for internal teams. Exit evidence: sovereign tier live for at least one sensitive domain; routing policy in code review; fallback test records; Tier 3 measurements on file; the first internal-team system registered and gated through the enabler path.

**Phase 4, months 18 to 24, institutionalising.** Run the first full maturity assessment; refresh the dependency map; present the board's annual risk acceptance statement, dependency map, maturity level, incident summary and workforce position; convert the roadmap into business as usual: register reviews, evaluation cadence, policy refresh, literacy refresh; publish the department's internal AI transparency statement to the workforce. Exit evidence: assessment at Level 3 or with a funded path to it; board acceptance minuted; the machinery running on cadence rather than on heroics.

Two scheduling notes. First, the phases overlap in practice; the sovereign tier can begin in Phase 2 for departments with capacity. Second, the regulatory clock sets the outer boundary: the AI Act's high-risk obligations, in application since August 2026, mean that any department beginning this roadmap today is starting late, and the roadmap is written to be compressed rather than admired.

I

PAGE

The board view: sovereignty ratio, maturity, incidents, coverage, accepted risks.

# 16. Metrics that matter

A small metric set, reported quarterly, each metric defined so it cannot be gamed by the systems it measures.

**Estate metrics.** Register coverage: capabilities registered versus discovered in sweeps; the gap is the shadow estate, and it is the department's most honest number. Tier distribution and change. Sovereignty ratio: proportion of Tier 2 and above capabilities meeting the Chapter 4 definition, with the trend, which is the metric a board should watch most closely. Enabled

- 1 The shift
- 2 The frame
- 3 **The blueprint**
- 4 Beyond
- 5 The working

estate: systems built by internal teams through the enabler platform, registered and gated, versus those discovered outside it.

**Quality metrics.** Evaluation pass rates by capability and version. Escalation contamination rate: assistant-assisted escalations containing material errors. Override rates by consequential class, with the zero-override investigation rule of Pattern 7. Corpus freshness: proportion of citations resolving to documents current within policy windows.

**Safety metrics.** AI incidents by class and severity; time from detection to register update; recurrence rate, which measures whether the learning loop of C8 is real.

**People metrics.** Literacy completion by role family. Transition protocol adherence: capabilities deployed with the statement, consultation and pathways complete. Workforce experience: the annual survey question that matters is whether staff believe the department tells the truth about AI, and the trend on that question is a better predictor of programme success than any technical metric.

**Cost and sustainability metrics.** Consumption per capability, cost and energy, with the C12 accounting; cost per resolved service interaction across the human-assistant boundary, used with care, because a falling cost with rising contamination is a service being degraded efficiently; reuse rate, the proportion of new systems built on shared capabilities rather than fresh deployments.

**The board view.** One page: sovereignty ratio and trend; maturity level; incidents and their dispositions; register coverage; the two or three accepted risks with owners and review dates. Boards do not need the detail; they need to know that the detail exists and who answers for it.

- 1 The shift
- 2 The frame
- 3 The blueprint
- 4 **Beyond**
- 5 The working

## 5

## THINGS

The minimum viable version: register, policy page, owner, boundary, annual review.

## 17. The wider setting

The blueprint is written for a department. Three groups of organisations face the same questions without the same machinery, and each deserves a short treatment of how the blueprint scales.

**Public sector and critical infrastructure.** Public bodies face the blueprint's questions in their sharpest form. A council, a health body or a university adopting AI-assisted decision systems in employment, benefits, safeguarding or admissions operates squarely in the AI Act's high-risk regime, with application dates running through 2026 and 2027. They are also attractive targets and constrained purchasers simultaneously: their budgets are public, their adversaries know it, and AI raises the capability of both sides while the defenders alone must run a governance programme on top of existing operations. Most importantly for this paper, they are the natural first movers of the sovereign pattern. Public data belongs, in a meaningful sense, to the public whose data it is, and routing a jurisdiction's public decision-making through foreign-hosted models is a policy question, not merely a technical one. The blueprint applies to them with two amendments: the dependency map of Chapter 3 should be published in summary form, because public accountability for public dependencies is the strongest discipline available; and the sovereignty score of Chapter 9 is a legitimate award criterion in public procurement where the rules permit qualitative criteria, with the lawyers engaged early, because the rules were largely drafted before model dependencies were a recognisable category.

Critical infrastructure operators inherit the same structure with NIS2 as the spine. For an operator of essential services, the sovereign constraint on security telemetry is the operational resilience argument, not a preference. An operator whose detection, triage and response capability can be degraded by a foreign legal instrument does not have resilience; it has an assumption of allied goodwill, which is a political variable rather than an engineering one.

**Small organisations and the fractional model.** Many organisations have no department: small enterprises, charities, micro-businesses, and the long tail of European commerce that employs the majority of its workforce. The blueprint scales by subtraction, not dilution, and the minimum viable version is this.

One register, even a spreadsheet, of every AI tool in use. One policy page: what data may never enter any system, which tools are sanctioned, who to tell when something goes wrong. One named owner, whatever their day job. One classification boundary between data that may go to any provider and data that may not, with the latter routed to a self-hosted or EU-jurisdiction tool. One annual review, prompted by a calendar, of the register and the dependency position. Five things; everything else in this paper is elaboration.

The fractional model addresses the capability gap, and the AI-first practice described throughout this paper is its proof of concept. The author's own department, roughly fifteen people serving a global organisation across more than forty countries, with first line and triage replaced by AI and development largely AI-generated, demonstrates the operating point the fractional model can now deliver: a small, senior, AI-supercharged team providing the enabler platform, the guardrails and the expert escalation path for an organisation many times its nominal capacity. A part-time technology director armed with this structure can deliver Level 2 maturity for organisations that could never fund a full-time function: the register, the policy, the tiering and the dependency map are all artefacts a competent fractional lead can operate across a small portfolio. This is, in the author's assessment, one of the fastest-growing segments of technology practice in

- 1 The shift
- 2 The frame
- 3 The blueprint
- 4 **Beyond**
- 5 The working

Europe, precisely because the obligations of the AI Act, GDPR and NIS2 arrive at every organisation simultaneously and with no small-business exemption from their risks. A note of interest: the author has chosen this fractional model for his own practice. The argument does not depend on that fact, but the reader should have it.

There is also an aggregating risk worth naming: the small organisations that defer the work entirely do not remain independent of AI; they adopt it individually, ungoverned, through consumer accounts and free tiers, and their collective exposure, across a region or a sector, becomes an economic risk rather than a private one.

**The regional dimension.** The sovereignty debate is usually conducted at the level of the Union, the state and the continent. It is decided, in practice, in places: where the infrastructure sits, where the skills live, and whether the benefits of the technology economy reach the regions that host its consequences. Three points connect that debate to this blueprint.

The first is talent. The supervision plane of Chapter 7 requires exactly the distributed, place-based, senior-but-not-relocated workforce that regions hold: evaluators, curators and decision owners who need domain experience rather than a desk in a hub city. Remote-first IT organisations can draw on this workforce deliberately, and regions that invest in the skills will find the demand arriving whether or not the investment was planned. The AI-enabled department is the first enterprise technology model in a generation in which the limiting factor is judgement rather than geography, and regions rich in capable people who choose not to relocate are its natural beneficiaries.

The second is infrastructure. The technology that makes the AI transition possible, data centres, renewable generation, connectivity, is capital investment in physical places, and the European model of AI sovereignty requires European sites, which are not all in Frankfurt, Dublin and Stockholm. A region with renewable energy resources, existing grid connections, a university producing graduates and a quality of life that retains them is a candidate location for the infrastructure of the cognitive layer, and the regions that understand this early will negotiate from strength.

The third is resilience. The concentration risk the dependency map condemns at the level of providers applies at the level of geography. A European AI estate sited entirely in three hubs is a strategic target, and regional distribution is the physical expression of the reversibility principle. The AI economy, like every infrastructure economy before it, will produce winners and losers by geography. Which is which is a policy choice, and it can still be influenced while the map is being drawn.

## I

### SENTENCE

The department that can answer for what AI does in its name.

## 18. Conclusion: the department that governs itself

The argument of this paper reduces to a single sentence: the AI-enabled IT department is not the department that uses AI best, it is the department that can answer for what AI does in its name.

Answering requires the machinery this paper has specified: the register, because accountability begins with enumeration; the tiers, because proportionality is what keeps governance alive; the controls, because ethics that cannot be evidenced are opinions; the dependency map, because sovereignty begins with honesty about where autonomy ends; the sovereign tier, because the honest answer must be followed by an architectural one; the evaluation suite, because competence that is not measured decays; the override channel, because human authority must be designed into the system rather than asserted over it; and the workforce programme, because

- 1 The shift
- 2 The frame
- 3 The blueprint
- 4 **Beyond**
- 5 The working

a department that automates its people out of their apprenticeship has consumed its own future for present efficiency.

The order matters, and the roadmap of Chapter 15 is the order: see the estate, then govern it, then build the sovereign core, then institutionalise. Departments that begin with the technology, a model, a platform, a pilot, will build something they cannot govern on infrastructure they cannot leave. Departments that begin with the register will build slower, ship later, and own what they ship. In a transition this large, ownership is the competitive position.

There is a final argument for acting now rather than eventually, and it is the quiet one underneath all the regulatory deadlines. Every month of ungoverned adoption adds entries to a shadow estate that must later be discovered, assessed, remediated or retired, and remediation is always more expensive than governance. The department that starts its register this quarter is buying the cheapest version of the future available to it. The department that waits is buying the same future with a late-payment premium, and the premium is rising.

The machine will not govern itself. That is the department's job, and it is a better job than the one it replaces.

Two closing observations follow from the practice that produced this blueprint. First, the small team is no longer the consolation case; it is the leading case. A department of fifteen people, with first line absorbed by governed assistants, development largely machine-generated, internal teams building within sanctioned guardrails and every expert multiplied by AI knowledge, speed and assistance, outperforms the department it replaced at a fraction of its cost and none of its queue. Scale has decoupled from capability, and the organisations that understand this first will hold the advantage for the length of the transition. Second, empowerment is the department's future shape. The IT function that remains a factory will be priced like one, by the ticket, and will lose that price war to software. The IT function that becomes the platform on which the whole organisation builds ethical, sovereign, sustainable AI will be priced like infrastructure: hard to replace, embedded in everything, and answerable for more than it operates. That is the department this blueprint has described, that practice has proven, and that the coming decade will select for.

## Annex A. Dependency mapping template, with a worked example

For each AI capability in the estate, complete one row. Score each of the four sovereignty components 0 to 2 and record the total against the capability in the register.

| Field                            | Question                                                                                                                  | Evidence                    |
|----------------------------------|---------------------------------------------------------------------------------------------------------------------------|-----------------------------|
| Capability name                  | What is it called operationally?                                                                                          | Register ID                 |
| Owner                            | Who is accountable?                                                                                                       | Register field C3           |
| Provider entity                  | Which legal entity provides the model and service?                                                                        | Contract, service terms     |
| Provider jurisdiction            | Under which law does the entity sit?                                                                                      | Corporate registry          |
| Parent chain                     | Who ultimately controls the entity?                                                                                       | Ownership records           |
| Inference location               | Where does inference physically occur?                                                                                    | Provider attestation, terms |
| Prompt and output storage        | Where are prompts, outputs and logs processed and retained?                                                               | Terms, telemetry review     |
| Applicable lawful access regimes | Which regimes can compel the entities above?                                                                              | Legal assessment            |
| Model provenance                 | Weights open or closed; source; version pinning?                                                                          | Model card, Gate 2          |
| Update policy                    | Can behaviour change without notice?                                                                                      | Contract terms              |
| Data classifications permitted   | What may the capability touch?                                                                                            | Tier assessment             |
| Fallback                         | What runs if this is withdrawn, degraded or compelled?                                                                    | Test record                 |
| Fallback tested                  | When was the fallback last exercised?                                                                                     | Test log                    |
| Jurisdiction score (0 to 2)      | 2: all entities and hosting within EU jurisdiction; 1: mixed or unclear; 0: foreign jurisdiction with compulsion exposure | Assessment note             |
| Continuity score (0 to 2)        | 2: tested fallback; 1: documented but untested; 0: none                                                                   | Test record                 |
| Transparency score (0 to 2)      | 2: open weights or full disclosure; 1: partial; 0: black box                                                              | Gate 2 evidence             |
| Agency score (0 to 2)            | 2: department controls deployment and evolution; 1: shared; 0: provider-controlled                                        | Contract review             |
| Sovereignty total (0 to 8)       | 7 to 8: sovereign; 4 to 6: conditional, justify; 0 to 3: dependent, fallback mandatory                                    | Board reporting             |

**Worked example.** A knowledge assistant over internal policy documentation, hosted with a large US-headquartered provider's "EU data boundary" region.

Provider entity: the provider's Irish subsidiary, contracting under Irish law; parent: a US-listed corporation. Inference: EU region data centres; prompts retained thirty days by default. Lawful access: the parent remains subject to the CLOUD Act; the EU residency arrangement is a contractual commitment, not a change in the parent's legal exposure. Model: closed weights, quarterly silent updates. Permitted data: internal documentation up to and including HR policy; no special category data, no security findings. Fallback: a pinned open-weight model on the sovereign tier, tested six months ago.

Scores: jurisdiction 1 (hosting in the EU, parent compelled elsewhere); continuity 2 (tested fallback); transparency 0 (closed weights, silent updates); agency 1 (region and retention configurable, model behaviour not). Total: 4 of 8, conditional. The register records the justification: the capability touches no sensitive classification, the fallback is proven, and the route will be re-scored at contract renewal. This is the method working as intended: not a veto, a visible, priced, reversible choice.

## Annex B. Sample AI register entry

---

A worked entry, in the field order of Chapter 8, for a Tier 2 service desk assistant.

Name: Service Desk Assistant. Register ID: AI-014. Owner: Head of Service Management. Purpose, plain language: answers routine IT queries from staff, resets account credentials within policy, escalates what it cannot resolve with full context attached. Tier: 2. Model and provider: pinned open-weight model, self-hosted on the sovereign inference tier, EU jurisdiction. Data classifications permitted: staff identity data for authentication purposes, ticket metadata; no special category data, no commercially confidential material. Decision classes affected: none consequential; all resolutions are routine or escalate. Human oversight: escalation to the on-call service expert; assistant proposals are advisory for any action with business effect. Evaluation: set of 240 real historical tickets with expected outcomes; pass threshold 95 per cent; last run at the current model version, three weeks ago, 97.1 per cent. Fallback: human-only queue with pre-AI routing rules; tested at last failover exercise. DPIA: DPIA-2026-03, covering prompt logging, retention of ninety days, and access restriction to the governance lead and the DPO. Lifecycle state: production. Consumption: 1.9 million tokens per month, cost per resolved interaction tracked in the Chapter 16 metric set.

An auditor reading this entry can answer, in under five minutes, what the system does, who answers for it, where it runs, what it may touch, how it is tested, what happens when it fails and what it costs. That is the standard every register entry should meet.

## Annex C. AI governance policy skeleton

---

Section 1, purpose and scope: the organisation's AI estate, definitions, relationship to the security policy set and the record of processing activities.

Section 2, principles: the five commitments of Chapter 5, restated in the organisation's language, with the accountability statement: the named owner answers for each system.

Section 3, the register and tiers: the four tiers, their requirements and their approval paths, condensed to one page with a pointer to the operating procedure.

Section 4, acceptable use: one page for staff: permitted uses; prohibited data classes; sanctioned tools by tier; disclosure duties; the escalation route; the sanction framework for evasion.

Section 5, data protection: classification boundaries, prompt logging and retention rules, transfer positions, DPIA triggers, the DPO interface.

Section 6, procurement: the five gates, the sovereignty score, the exception process with fallback requirements.

Section 7, security: the six asset classes of Chapter 10 and their controls; agent containment rules; the incident channel for AI events.

Section 8, workforce: AI literacy expectations by role family; the transition protocol; consultation commitments.

Section 9, governance bodies: the board, its powers, its cadence; decision rights; suspension and veto authorities.

Section 10, review: policy owner, review cadence, version history, and the annual board risk acceptance statement.

## Annex D. DPIA structure for AI systems

---

Part 1, processing description: capability, register ID, purpose, data categories, subjects, flows, providers, hosting and transfer positions.

Part 2, necessity and proportionality: why AI processing at all; why this model and tier; what alternatives were considered; the data minimisation decisions taken.

Part 3, risk identification: for each risk, likelihood and severity to individuals: fabrication with effect; bias or disparate outcomes; re-identification; function creep beyond the stated purpose; disclosure of sensitive material

through prompts; over-reliance by decision owners; loss of rights in challenge, appeal or explanation.

Part 4, mitigations: mapped to the pattern set: routing boundaries, retrieval grounding with provenance, logging with defined retention, override design, evaluation evidence, human decision support.

Part 5, residual risk and the decision: residual assessment per risk; the named person accepting it; the date and conditions of review; the fallback position.

Part 6, consultation record: DPO advice; affected staff consultation where applicable; governance board disposition.

## Annex E. Board briefing template

---

One page, five blocks. Block 1, the estate: capabilities by tier, register coverage, shadow AI findings. Block 2, sovereignty: the ratio and its trend, the dependency map summary, the accepted foreign dependencies with justifications and fallbacks. Block 3, risk: incidents and their dispositions, the accepted risks with owners and review dates, the maturity level and its evidence. Block 4, people: literacy completion, transition protocol adherence, the workforce trust question from Chapter 16. Block 5, asks: decisions required this quarter, with the recommendation and the consequence of deferral stated plainly.

## Annex F. Model evaluation checklist

---

Before deployment and at each model or prompt change: pass rate against the versioned evaluation set, with threshold; performance on the known hard cases and previous failures; disparate outcome measurement where the system differentiates between people; retrieval provenance accuracy for grounded systems; refusal and escalation behaviour on out-of-scope requests; injection resistance for agent systems; cost and energy per unit task; regression against the previous version; sign-off by the evaluator of record; and the register updated with the version and date.

## Annex G. Maturity assessment checklists

---

An assessor works down each list; the department's level is the highest level for which every item is evidenced.

**Level 0, unaware.** No inventory of AI use exists. No policy or guidance is published. No individual owns the topic. Discovery of AI use in the estate would be by incident.

**Level 1, reactive.** Some guidance exists and is known to some staff. One or more tools are sanctioned, informally. No register; no tiering; no evaluation. AI appears in the risk register only if an incident has already occurred. The department describes its posture as "aware and working on it".

**Level 2, governed.** The register exists and covers at least the known estate, with owners. The acceptable use policy is published, trained and acknowledged. Tiering operates, with approval paths in use. DPIAs exist for Tier 2 capabilities touching personal data. At least one evaluation set gates a production capability. A governance lead is named, at whatever fraction. A shadow AI sweep has been run at least once.

**Level 3, supervised.** Every consequential decision class has a named decision owner and a designed override. Evaluation runs on schedule and at every model change for all Tier 2 and above capabilities. AI incidents flow through the security incident channel and feed the evaluation sets. The dependency map exists; fallbacks are documented and tested. The workforce programme runs: role map, transition protocol, literacy matrix with evidence. Internal teams build through the enabler platform, and what they build is registered and gated.

**Level 4, sovereign and adaptive.** The routed portfolio operates in code. Sensitive domains run on the sovereign tier. Procurement gates are applied to every purchase and renewal, with sovereignty scores on file. The board receives and accepts the annual dependency and risk statement. Consumption and sustainability accounting is routine. Regulatory or model changes are absorbed by amending patterns and controls, not by restarting a programme.

## Annex H. Glossary

---

Agent: an AI system that takes actions in other systems, not merely produces text. Capability: one registered AI use in the estate; the unit of governance throughout this paper. Deployer: the organisation using an AI system in its own operations, with attendant duties under the AI Act. Enabler platform: the department's sanctioned patterns, tools, data boundaries and review path through which internal teams build their own AI systems. Evaluation set: a versioned collection of tasks and expected outcomes used to gate and monitor model behaviour. Escalation contamination: the rate at which human escalations contain material errors made by the assistant layer. Grounding: constraining model output to retrieved source material. Open weights: model parameters published for download and self-hosting. Prompt injection: causing a system to follow attacker-supplied instructions through content it processes. RAG: retrieval augmented generation. Register: the single inventory of the AI estate. Sovereignty: the tested ability to continue a capability under one's own law if external parties fail or are compelled. Sovereignty ratio: the proportion of Tier 2 and above capabilities meeting the Chapter 4 definition. Sovereignty washing: marketing residency or compliance as sovereignty while control remains elsewhere. Supercharging: the deliberate use of AI as a multiplier on an individual expert's judgement, knowledge and speed, in place of headcount growth. Tier: the one of four risk categories that determines a capability's assessment, approval and review regime.

## Annex I. References and further reading

---

Primary legal instruments: Regulation (EU) 2024/1689, the Artificial Intelligence Act; Regulation (EU) 2016/679, the GDPR; Directive (EU) 2022/2555, NIS2; Regulation (EU) 2024/2853, the Cyber Resilience Act; Regulation (EU) 2023/2854, the Data Act. Implementation timelines and obligations for the AI Act are maintained by the EU AI Act implementation tracker ([artificialintelligenceact.eu](https://artificialintelligenceact.eu)). Transposition status for NIS2 is tracked at the NIS2 Compliance resource ([nis2compliance.digital](https://nis2compliance.digital)).

European Data Protection Board, Opinion 28/2024 on AI models and GDPR obligations. General Court judgment of 3 September 2025 in Case T-553/23 (Latombe), appeal pending before the Court of Justice as Case C-703/25 P. International Monetary Fund, *Gen-AI: Artificial Intelligence and the Future of Work* (2024), on AI exposure in advanced economies.

Further reading on the themes of this paper: the AltLibre essays on dependency mapping, sovereignty washing and open weights as continuity strategy, available at the AltLibre publication library.

*Adrian Hollister is the founder of AltLibre and Director of IT and Cyber Security at Adam Smith International. This paper is published under the AltLibre digital sovereignty series. Comments, corrections and counter-arguments are welcome; the strongest objections to this blueprint will be answered in a future edition.*