

Make AI Adoption Safe

Six questions between an AI idea and a governed, evidenced use case.

J01 helps an organisation establish and sustain a proportionate, evidence-based approach to AI adoption, with uncertainty and authority boundaries visible. It does not promise universal safety, legal compliance or certainty.

A public learning product: the complete Knowledge Layer of J01. You may keep it and present it internally with attribution.

THE TRIGGER

AI adoption can outrun organisational capability

AI is being used before anyone has agreed how it should be used.

Adoption may appear active while the organisation cannot consistently show how use cases enter, how risk is sized, which controls apply, who owns them, how governance stays live, where evidence is held, or how incidents and surprises are handled.

WHO FEELS IT FIRST

Usually risk, security or governance leadership.

TYPICAL TRIGGERS

Regulation · an incident · an assurance gap · adoption that has already outrun its controls.

Use cases run ahead of the governance built to hold them.

M15 Use-case Intake & Gates

Activity is not the same as control

An organisation can look as if it is adopting AI well while being structurally unsafe. Activity is visible; governance is not.

Visible adoption

People are using AI and talking about it; usage is growing.

Controlled adoption

Each use case has been sized, carries proportionate controls, and those controls have named owners.

Evidenced adoption

The controls leave a trace as work happens, so the organisation can show what operated and when.

Sustainable capability

The organisation notices when something changes, knows who owns the call, and improves from what it learns.

From outside, the first three look almost identical until someone asks a question only controlled adoption can answer: how was this use case sized, which controls does it carry, who owns them, and what would we show if asked for proof?

Documentation is necessary and not sufficient. A control nobody owns is a wish. A size label without controls has no consequence. A control checked once is already out of date.

The EYW point of view

Safe AI adoption is an organisational capability problem, not simply a policy, technology or compliance problem.

Capability is present only when the organisation can do the work repeatedly in context, with authority, evidence, correction and learning visible.

STATED PLAINLY, SO YOU CAN DISAGREE

01

Size risk before choosing controls.

02

Controls must be owned and operable.

03

Governance must stay alive after launch.

04

Evidence is designed, not reconstructed.

05

Surprise needs a declared owner.

06

The six parts form one dependent system.

J01 does not promise universal safety, legal compliance or certainty.

The J01 safe-adoption system

Six connected questions between an AI idea and governed use. Each stage hands the next the context, decision or evidence it needs.

FRONT DOOR • M15 Use-case Intake & Gates • Where do AI use cases enter, and what bypasses that route?

SIZE

M10
AI Risk Sizing

How big is this, and would two reviewers agree?

CONTROL

M11
AI Risk Controls

Which controls are required, and who owns them?

KEEP-LIVE

M12
Living Governance

What changes after launch, and what wakes governance up?

PROVE

M13
AI Evidence & Audit-Readiness

If proof were requested tomorrow, where would it come from?

RESPONSE NET • M14 Incident & Surprise Response • Who declares an AI surprise, and who owns the first decision?

🔄 Learning returns to the Living Loop and may update Front Door entry criteria.

The route is iterative: change, evidence and incidents can reopen earlier judgements.

Front Door

Where do AI use cases enter, and what bypasses that route?

M15 • Use-case Intake & Gates

Decide whether and how a use case should enter the governed portfolio.

FRONT DOOR · M15

SIZE · M10



CONTROL · M11



KEEP-LIVE · M12



PROVE · M13

RESPONSE NET · M14

Where do AI use cases enter?

Where do AI use cases enter, and what bypasses that route?

Use cases enter through side doors, or before their scope and ownership are stable. They are discovered after build, near-duplicates arrive through different channels, sizing is used afterwards to justify work, and approval cannot be reconstructed.

SYMPTOMS OF AN UNGATED BACKLOG

- Use cases discovered live, already in production.
- Three teams shipping near-identical AI to one workflow.
- Risk sizing run after the build, to justify it.
- Audit asks "who approved this?" and the room goes quiet.
- Sunset never happens — nothing is anyone's to retire.

THE COST CURVE OF A SKIPPED GATE

- At intake — a thirty-minute conversation.
- At design — a redesign.
- At build — rework plus control retrofits.
- At deploy — a release block and a stakeholder call.
- In production — an incident, an audit finding, a headline.

Four ways to look at demand

Risk is necessary, not sufficient. A clear pass needs all four; any one lens can redirect or refuse.

RISK / REGULATORY	ARCHITECTURE / FEASIBILITY	VALUE / BUSINESS-CASE	PEOPLE / ORG-IMPACT
● Asks — how risky, and is it lawful? ● Why — the unsafe one becomes the headline. ● Trips on — data, autonomy, rights; Art. 5 / Annex III. ● Owner — Risk + Security / Privacy.	● Asks — can we build and run it here? ● Why — un-buildable burns budget, destabilises live systems. ● Trips on — missing data, integration, latency, coupling. ● Owner — Enterprise Architect + Platform.	● Asks — is it worth it, and how will we know? ● Why — unmeasured AI is cost you cannot defend. ● Trips on — no outcome, weak ROI, duplication. ● Owner — Sponsor + Value / Finance.	● Asks — what does it do to people and the org? ● Why — blindsiding the workforce fails on trust. ● Trips on — changed jobs, decisions on people, consultation. ● Owner — HR / Change + the affected function.

Risk-only intake is the commonest front-door failure.

Five gates before commitment

Five gates between a request and a live use case. One question and one outcome each — pass, redirect, or refuse.

GATE	THE ONE QUESTION	HANDS OFF TO
01 • INTAKE	Should this use case exist in our portfolio at all?	Four lenses • refusal patterns
02 • SIZE	How big is the risk — S, M, L or XL?	Module 10 Risk Sizing
03 • DESIGN	Which controls does this tier require?	Module 11 Risk Controls
04 • BUILD	Is the evidence trail audit-ready?	Module 13 Evidence & Audit
05 • DEPLOY	Is the live heartbeat live and owned?	Module 12 + Module 14

Five gates, three outcomes each — fifteen named decisions instead of one big bet.

Pass, Redirect or Refuse

Every gate ends in one of three decisions. Deferral pending a named prerequisite is not a fourth outcome.

THREE OUTCOMES • ONE DECISION

- PASS — all four lenses clear. Proceed to Gate 02.
- REDIRECT — one lens blocks the shape, not the need. Route to a non-AI fix or an existing use case.
- REFUSE — outside appetite, feasibility, value or competence. Record the reason so it is not re-pitched.

THREE OUTCOMES • NEVER PARK IT

- Pass, redirect or refuse at every gate — on any of the four lenses.
- "Park it and revisit" is a deferral, not a decision — and it costs more later.

Escalation is a governance path, not an outcome. When a call belongs above the gate, it goes to the appropriate governance authority. A Pass is recorded only once that authority has resolved the escalated issue and approved any recorded conditions; otherwise the outcome is Redirect or Refuse.



Wyreld enters through the Front Door

Wyreld Co. (mid-size building-products manufacturer). Same intake, four lenses each — and three different outcomes.

SITEINSPECTION • SIZE S

- Scene — a depot supervisor logs racking checks and near-misses on a tablet. No AI decisioning.
- Risk: clear • Architecture: clear.
- Value: clear • People: clear.
- AT THE DOOR — PASS. Sails through to sizing.

SHORTLISTASSIST • SIZE XL

- Scene — HR ranks seasonal applicants with a Copilot, to "save a day a week".
- People / Org — TRIPS: it decides who gets interviewed.
- Risk / Reg — TRIPS: EU AI Act Annex III employment.
- Architecture clear • Value real but modest.
- AT THE DOOR — escalated to the appropriate governance authority, then passed with recorded conditions.

CATALOGUEAGENT • SIZE L → XL

- Scene — a web FAQ agent, wired by sales into live pricing and stock to "help them order".
- Architecture — TRIPS: live pricing + stock.
- Value — TRIPS: the case shifts to ordering.
- Risk / Reg — rising: disclosure as it prices.
- AT THE DOOR — REDIRECT. Split FAQ from ordering.

What stronger Front Door practice looks like

STRONGER PRACTICE

Recent requests trace to entry, acknowledgement and owner.

Use cases carry intended user, action, data, benefit, affected party, boundary and non-goal.

Gate decisions carry criteria, conditions and rationale.

WHAT EACH GATE OWNS • MODULE 15

- The four-lens read: risk, architecture, value, people.
- The decision: pass, redirect or refuse.
- The named owner who makes the call.
- The record that travels with the use case.
- The forcing function to advance or sunset.

WHAT THE SPINE OWNS • THE DEPTH

- Module 10 AI Risk Sizing — the S/M/L/XL determination.
- Module 11 AI Risk Controls — the control library per tier.
- Module 13 AI Evidence & Audit-Readiness — the proof.
- Module 12 Living Governance + Module 14 Incident & Surprise Response — the live run.

Where do ideas enter, and what bypasses the route?

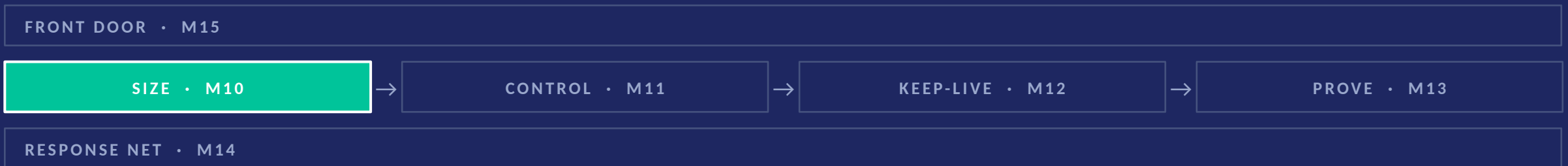
→ **Risk Size** A defined, triaged use case passes to Risk Size.

Risk Size

How big is this, and would two reviewers agree?

M10 • AI Risk Sizing

Establish an explainable, proportionate risk size before controls are chosen.



How big is the risk?

How big is this, and would two reviewers agree?

Intuitive labels and averages hide the decisive dimension. Teams size by instinct, average dimensions so the decisive risk disappears, or score a no-go into acceptance.

WHY GENERIC MODELS BREAK ON AI

AGENCY — it can act

- Risk scales with what it can do, not just what it says.

OPACITY — hard to explain

- Audit and explainability become first-class concerns.

DISTRIBUTION — one model, many uses

- Score per use case, not per model.

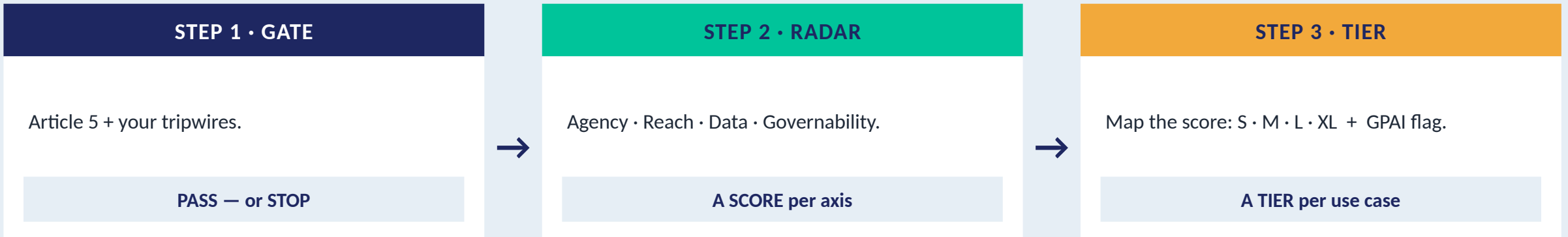
REGULATION — a legal floor

- The Act sets the floor; your organisation sets the rest.

Gate before you score. Size by pattern, not sum.

Gate → Radar → Tier

Gate the no-gos. Score what's left. Tier the response — every path resolves to one effective risk tier.



OUTPUT • EFFECTIVE RISK TIER

Each use case resolves to one effective size, S, M, L or XL, with a flag where a general-purpose model is involved. A use the law already treats as high-risk is not sized down.

Agency • Reach • Data • Governability

Four rungs per axis. The Radar shows the silhouette of a solution, not just how much risk it carries.

AGENCY — what can it do?

- Rungs: Assist • Augment • Act • Autonomous.
- Reversibility of its actions; informs or decides.
- How far past the human is it allowed to go?

REACH — what is the blast radius?

- Rungs: internal • cross-functional • customer-via-staff • direct.
- Action surface: read, write, or external comms.
- If it goes wrong, who feels it — and how many?

DATA — what is it touching?

- Rungs: public • confidential • PII • restricted.
- Egress — in-tenant, vendor, public model.
- Where does data go, and what could be injected?

GOVERNABILITY — can you steer it?

- Rungs: logged • reviewed • approved • audited.
- Auditability, explainability, recourse; runs as user or service principal.
- Could you stop it, undo it, or answer for it?

The rungs describe what each axis means. They are not a scoring scale: no cut-offs, anchors or calibration are given here. Platform, Model, Data, Policy and People are the Living Loop signal axes (Stage 4), not sizing axes.

Pattern, not sum

The two calls that keep sizing honest — and the two failure modes they prevent.

GATE BEFORE YOU SCORE

- Article 5 and your tripwires are pass / STOP — never a Radar input.
- A no-go cannot be "well-governed" into acceptance.
- Remove the prohibited element, then re-run the gate.
- WHY: scoring a no-go launders it into a number that looks debatable.

PATTERN, NOT SUM

- The highest single axis sets the tier floor.
- Two mid's never average into a low.
- Annex III overrides to XL, whatever the Radar says.
- WHY: averaging hides the one axis that actually carries the risk.

Judgements carry evidence and dissent. Where two reviewers read an axis differently, the disagreement is recorded, not averaged away.

CatalogueAgent lands at L because one axis — direct-customer Reach — carries it, not the average.

FIRM and FLEX

Firm and Flex separate the parts of risk you may debate from the parts you may not.

FIRM — pins the tier floor

- The Gate: Article 5 + your tripwires — absolute. A hit is STOP, no debate.
- Annex III overrides to XL, whatever the Radar score.
- The riskiest single Radar axis sets the floor — pattern, not sum.
- An override needs named, logged sign-off by the authority your organisation designates.

FLEX — adjusts up on evidence

- Where each axis boundary is drawn, calibrated to your organisation's own anchors.
- Flex up when multiple axes push the same way — pattern beats single signal.
- Flex never reduces below the floor the firm lines have set.

Firm pins the floor. Flex never reduces below it. Firm lines trace to a policy authority — not to the loudest voice in the room.

Wyreld: sizing ShortlistAssist

Wyreld Co.'s portfolio, each labelled Gate · Radar · Tier so the outcome level is explained. GPAI flags overlay; they do not change the tier.

S · MINIMAL — SiteInspection	L · ELEVATED — CatalogueAgent	XL · HIGH-RISK — ShortlistAssist	STOP · GATE FAIL — emotion inference
● A depot supervisor logs checks + near-misses on a tablet; no AI decisioning. ● GATE: PASS — not Art 5, no tripwire. ● RADAR: assist · internal · operational. ● TIER: S — light touch, standard hygiene.	● Web agent answers product questions for installers + merchants. ● GATE: PASS — limited-risk; discloses AI. ● RADAR: augment · direct-customer · pricing. ● TIER: L — direct-customer Reach is the floor.	● HR Copilot ranks CVs against a role profile, at volume. ● GATE: passes Art 5; trips employment tripwire. ● RADAR: looks moderate — but Annex III applies. ● TIER: XL — Annex III employment override.	● Tooling infers caller emotion to drive coaching. ● GATE: STOP — Article 5. ● RADAR: not reached — exits at the gate. ● TIER: none — forbidden at work.

A benign-looking Radar does not save you: ShortlistAssist is XL because the Act says employment AI is high-risk — the override beats the score.

What stronger Risk Sizing looks like

Context before sizing

The use case is defined before anyone sizes it.

Evidence and dissent kept

Judgements carry evidence and dissent; the class is explainable.

A named reassessment trigger

The class is revisited by named authority when a trigger invalidates the earlier judgement.

Would two reviewers agree, and where might they differ?

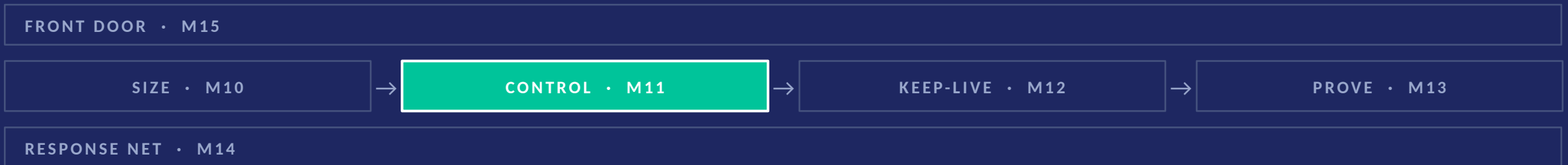
→ **Control Floor** Risk context, dimensions, class and trigger pass to Control Floor.

Control Floor

Which controls are required, and who owns them?

M11 • AI Risk Controls

Turn risk size into the smallest sufficient set of owned, operable controls.



A tier label is not a control plan

Which controls are required, and who owns them?

Generic or maximal control sets obscure proportionate ownership. Controls become abstract, uncalibrated, moment-in-time or unowned, so a tier label is mistaken for an operating plan.

FAILURE MODE	WHAT IT LOOKS LIKE	WHY IT FAILS
ABSTRACT	Policy says 'human oversight, audit, explainability'.	Nobody can operationalise it; two reviewers reach two answers. Principles do not ship — controls do.
UNCALIBRATED	A read-only summariser gets the same burden as an autonomous agent.	XL controls on an S case grind the org to a halt; S controls on an XL and the regulator finds out first.
MOMENT-IN-TIME	Heavy review at the gate, nothing once it is live.	The model drifts and integrations multiply; by month six the assessment describes a system that is gone.
UNOWNED	Policy names a control; no human owns evidence, cadence or escalation.	When it breaks nobody is on the hook. Review becomes theatre — sign the form, file it, move on.

The four-layer Control Floor

Four layers stack into one floor — the union sized to this use case, not the sum.

BASELINE • FROM THE GATE

- The legal + tripwire floor under every use case.
- Applies regardless of tier, axis or model.
- The safety net the whole stack rests on.

AXIS MODIFIER • FROM THE RADAR

- Targeted controls for the axis that sits high.
- Agency • Reach • Data • Governability.
- One axis high — its controls light up.

TIER DELTA • FROM THE TIER

- Extra controls that stack as the tier climbs.
- Sizes the response to the risk, S to XL.
- Each tier inherits the prior — it adds.

GPAI OVERLAY • FROM THE GPAI FLAG

- Provider-side controls for a GPAI model.
- The provider duties become your dependency.
- You evidence and inherit — you do not own.

Pattern, not sum: the controls floor is the UNION of the four layers — sized to this one use case.

What makes a control operational?

Seven fields turn a principle into something you can own, evidence and audit.

FIELD	WHAT IT CARRIES
NAME	Short, memorable, used in conversation.
LAYER	Baseline • Axis Modifier • Tier Delta • GPAI Overlay.
TRIGGER	When it switches on: always-on, axis-lit, tier-stacked, GPAI-overlaid.
OWNER	The named human accountable for evidence and escalation.
EVIDENCE	What proves it is in place: log, sample, document, attestation.
CADENCE	Weekly • monthly • quarterly • per-action • always-on.
ESCALATION	Who is paged, on what signal, and what they do next.

In practice, a material control also makes clear what it requires, how it operates in the work, how a justified exception is handled, and what proves it ran.

If a control cannot be named, owned, evidenced and escalated, it is not a control — it is a wish.

Smallest sufficient, not largest possible

Build the smallest sufficient floor. Controls that do not apply are excluded, with the reason recorded.

S MINIMAL	M LIMITED • STANDARD	L LIMITED • ELEVATED	XL HIGH-RISK
● Baseline only — no controls beyond the floor.	● + Pre-approval workflow on every action.	● + Sampled post-action review.	● + The legal high-risk obligations, evidenced.

Irrelevant controls, such as model-development controls for a use case that develops no model, are excluded with rationale rather than applied by default.

Each tier inherits the prior; the delta is what stacks — not a replacement, an addition. The natural lens of the Risk Spine.

Wyreld: from XL risk to proportionate control

Wyreld Co. — a mid-size building-products maker. The same controls model lands light, heavy, or climbing depending on the sized risk.

USE CASE (FROM M10)	TIER • EU AI ACT	THE FLOOR THAT LANDS
SiteInspection — H&S yard checks	S • Minimal	Light: a named owner and basic DLP. Baseline only — no axis, tier or GPAI controls beyond the floor.
ShortlistAssist — CV shortlisting	XL • High-risk (Annex III)	The heavy set: human sign-off, bias testing, candidate transparency, decision logging, a DPIA — plus GPAI evidence.
CatalogueAgent — customer Q&A	L → XL • Limited (transparency)	Transparency banner today; guardrails on price/stock commitments stack on as it reaches into live ordering.

For ShortlistAssist, Wyreld's resourcing lead is the named owner; a recruiter signs off every shortlist; CV data stays inside a data boundary and every action is logged.

Same model, three floors — sized to each use case, not to Wyreld. The tier label set the weight; the controls are the plan.

What stronger control practice looks like

Included and excluded, visibly

The floor shows which controls are included and which are excluded, and why.

Every material control is operable

Each control has a trigger, requirement, operation, exception and escalation.

Owners accept proof duties

The named owner accepts the duty to show the control operated.

Which controls operate and which are only policy?

→ **Living Loop • Evidence Chain** Owned controls pass to Living Loop and Evidence Chain.

Living Loop

What changes after launch, and what wakes governance up?

M12 · Living Governance

Keep governance live as platforms, models, data, policy and people change.



AI moves while conventional governance waits

What changes after launch, and what wakes governance up?

AI does not hold still. The riskiest posture is a static governance system sitting over a moving one.

LAYER	WHAT CLASSIC GRC ASSUMED	WHAT AI ACTUALLY DOES
PLATFORM	Stable — new versions on a quarterly release calendar.	Capability ships weekly; defaults change without you.
MODEL	A one-off build, validated once then frozen.	The provider swaps the engine under the same name.
DATA	Catalogued; labels rarely change after sign-off.	Sensitivity changes the moment a label is reapplied.
POLICY	Annual review cycle; changes are scheduled.	EU AI Act, NIST AI RMF, ISO 42001 — moving targets.
PEOPLE	Roles stable; access reviews quarterly.	Joiners, leavers, role changes — daily, not yearly.

Calendar-only review misses material change between meetings, and silence is mistaken for health.

Sense → Decide → Act

Sense the change. Decide who owns it. Act with evidence. Then loop back — the loop never closes.

SENSE • WATCH THE CHANGE

Catch every change event — platform, model, data, policy, people — and the silence of a signal that should have fired.

DECIDE • PICK THE RIGHT HANDS

Route each call to the right tier: agent-led, hybrid, or human-owned. Pattern work to the machine; judgement and weight to people.

ACT • MOVE, AND LEAVE A TRACE

Apply the control, page the human, or stop the agent. Every Line is a kill switch as well as a path, and every move leaves evidence that feeds the next Sense.

What wakes governance up?

Five axes, tech-agnostic — watch fewer than five and something moving is already unwatched.

PLATFORM — what ships under you	new model in default rotation; new tenant capability flag; defaults reset.
MODEL — what is generating	provider swaps the engine; eval scores drift; prompt template ships an update.
DATA — what it touches	sensitivity label upgraded; new source connected; DLP rule changes.
POLICY — the rules today	EU AI Act milestone; ISO 42001 audit; internal policy revision.
PEOPLE — who is in the loop	joiner with elevated access; role change; insider-risk flag fires.

PROPERTY

WHY IT MATTERS — WITHOUT IT, NO GOVERNANCE

NAMED SOURCE

If no one owns where it comes from, no one notices when it stops.

DEFINED SHAPE

Without it you drown in alerts — or miss the one that mattered.

EXPECTED CADENCE

Cadence is what makes SILENCE detectable — a quiet sensor is suspect.

ROUTED LINE

A signal with nowhere to go is logged once and forgotten.

THE HEARTBEAT RULE — The absence of an expected signal IS a signal.

Human authority remains part of the system

The two calls that keep governance alive between reviews.

EVENT, NOT CALENDAR

- The system moves between reviews — so governance fires on the change, not the date.
- Treat silence as a signal: a sensor gone quiet is suspect, not healthy.
- A scheduled review confirms; the event is what catches.

HOLD THE HUMAN FLOOR

- Three conditions force a named human, every time: non-delegable accountability, novel pattern, irreversible blast radius.
- Agents do the pattern work; humans hold the weight.
- Convenience never delegates a Tier 3 call to an agent.

Wyreld: when context changes

THE SCENE • SHORTLISTASSIST (XL • HIGH-RISK)		WHAT MOVED — TWICE	
ShortlistAssist (XL • high-risk): Wyreld's resourcing lead stood up a Copilot that ranks seasonal depot CVs against a role profile. Stage 2 sized it XL; Stage 3 fitted the heavy set: human sign-off, bias testing, candidate transparency, decision logging, a DPIA.		- Role profile: an H&S review makes an abrasive-wheels ticket mandatory; HR republishes it and the agent silently re-weights. - The law: an EU AI Act high-risk employment milestone tightens the logging and human-oversight evidence bar.	
STEP		WHAT LIVING GOVERNANCE DOES	
01 • SENSOR	Two events fire: HR publishes the revised role profile, and the policy watch-list flags the EU AI Act high-risk milestone.		
02 • TRIAGE	Both treated as material for a high-risk use case — they touch the bias baseline and the legal evidence bar.		
03 • DECISION	The human floor applies: the resourcing lead, with Legal and the DPO, owns the re-review. No agent auto-approves a high-risk change.		
04 • ACTION	Re-run bias testing on the new profile; refresh the candidate notice; raise logging to the new bar; re-confirm the DPIA. Hold the agent advisory-only until done.		
05 • EVIDENCE	Capture what changed, who signed, the new bias results and the DPIA date — and the next role change re-arms the trigger.		

The controls are re-fitted when the world moves, not when the calendar says: "we governed it" becomes "we are governing it".

What stronger Living Governance looks like

Signals trace to decisions

Signals trace to authorised decisions, actions and updated records.

The review is explicit

Review rationale, participants, inputs, thresholds and intervention routes are written down.

Silence is not health

A missing expected signal is treated as a signal, and change can trigger review between meetings.

What change or missing signal wakes governance up?

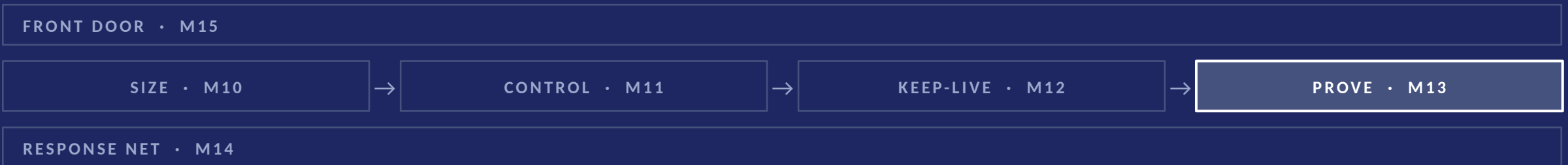
→ Evidence Chain • ↻ Risk Size Actions leave evidence; reassessment returns to Risk Size.

Evidence Chain

If proof were requested tomorrow, where would it come from?

M13 · AI Evidence & Audit-Readiness

Capture evidence while work happens so decisions, controls, signals and outcomes remain defensible.



Could you reconstruct last Tuesday?

If proof were requested tomorrow, where would it come from?

Screenshots, inboxes and recollection substitute for linked evidence captured during work. Evidence reconstructed at audit time is fragmented and weak in provenance and ownership.

SYMPTOMS • WHAT IT LOOKS LIKE

- Frantic email archaeology two weeks before the audit.
- Screenshots passed off as evidence; no chain of custody.
- Approvals that live in heads, not records.
- Logs over- or under-retained, with no policy applied.
- Five teams holding five fragments; no joined view.

WHY IT PERSISTS • WHAT TO FIX

- AI evidence is not named, so no one owns the discipline.
- Tooling captures activity, not decisions.
- Schema drift — every model team logs differently.
- Retention is assumed; rarely tested.
- Auditor expectations evolve faster than internal practice.

Evidence assembled at audit time is harder to produce and less defensible than evidence captured when the decision was made.

Four streams of evidence

Each stream answers a different audit question. Miss one and the record has a hole an auditor will find.

STREAM 01 • DECISIONS

- Answers: who chose what, and why.
- Captured at the approval gate, sign-off or exception.
- Decision records, RACI traces, approver identity.

STREAM 02 • CONTROLS

- Answers: did the safeguards run as designed.
- Captured at the control test, gate check or runtime guard.
- Test results, exception logs, override records.

STREAM 03 • SIGNALS

- Answers: what the system showed us.
- Captured at the heartbeat reading, axis move or alert.
- Telemetry snapshots, drift records, alert history.

STREAM 04 • OUTCOMES

- Answers: what actually happened in the world.
- Captured at the incident, near-miss or KPI report.
- Incident records, KPI snapshots, complaint log.

Capture at source

Not how much evidence you keep — WHEN you decide to capture it. That one choice sets your cost and your credibility.

CAPTURE AT SOURCE • THE DISCIPLINE

- WHAT — records fall out of how the system already runs; every decision, control, signal and outcome captured the moment it happens.
- WHY IT WINS — the cheapest evidence there is, and the most credible; the audit becomes a query against a structured store.
- HOW — wire capture into the runtime: the approval gate, the control test, the heartbeat, the incident — so no one has to remember.

RECONSTRUCT AT AUDIT • THE TRAP

- WHAT — the record is assembled only once an audit is announced, from whatever survived.
- WHY IT FAILS — it is harder, slower and less defensible, and still leaves a gap an auditor finds; approvals live in heads.
- THE TELL — email archaeology, screenshots, undated exports. Every audit is a project; every project misses something.

Create → Capture → Store → Retain → Retrieve → Defend

The weakest stage sets your audit-readiness, not the strongest. If a stage is unowned, the chain breaks and the evidence is unusable.

STAGE	WHAT HAPPENS HERE	WHO OWNS IT — AND WHAT BREAKS IF NO ONE DOES
01 • CREATE	The decision is made, the control runs, the signal fires, the outcome lands.	The asset owner. No trigger to capture means the event passes unrecorded.
02 • CAPTURE	A structured record is made — identity, timestamp and context attached.	The system owner. Capture activity but not the decision, and the record cannot answer "why".
03 • STORE	Held in a system that prevents silent edits and allows replay.	Platform / security. Editable storage means the auditor assumes you rewrote it.
04 • RETAIN	Kept for the period its type requires.	Compliance. Deleted early breaches the audit window; kept forever breaches minimisation.
05 • RETRIEVE	Findable on demand.	Compliance + platform. Not indexed means the evidence exists but cannot be found in time.
06 • DEFEND	The explanation stands up to scrutiny.	The accountable exec. No proven custody means the evidence is challenged and discounted.

Walk one artefact end-to-end before you scale. The stage with no owner is the hole the auditor walks straight through.

Wyreld: evidence without reconstruction

ShortlistAssist (XL · high-risk). A rejected applicant asks whether a human really reviewed the shortlist and whether the agent was bias-tested. Because Wyreld linked its records by use-case ID as the work happened, the answer is retrieved rather than rebuilt.

STREAM	WHAT WYRELD CAPTURED AT SOURCE	WHAT IT SHOWS
01 • DECISIONS	Per-shortlist sign-off: the resourcing lead's verified identity, criteria, evidence reviewed, rationale, timestamp.	Human oversight was real and named — not a rubber stamp.
02 • CONTROLS	Bias-test results on each role profile; the candidate-notice version shown; the live DPIA reference.	The high-risk control set actually ran, and when.
03 • SIGNALS	Heartbeat snapshots of score distributions by group.	Drift was watched.
04 • OUTCOMES	The down-ranking incident record, linked to the model version and the candidates affected.	What the agent did to real people, with root cause and fix.

Retrieval also shows what is missing. Known gaps are stated, not hidden.

→ **Response Net** Evidence supports every stage and later assurance questions.

Response Net

Who declares an AI surprise, and who owns the first decision?

M14 · Incident & Surprise Response

Contain proportionately, preserve evidence, learn and govern the lesson back into the system.



When the runbook does not fit

Who declares an AI surprise, and who owns the first decision?

Stale roles and familiar incident labels hide surprise and delay the first decision. Novel behaviour is forced into the nearest familiar runbook, and learning is not converted into owned change.

	WHAT CLASSICAL IR ASSUMES	HOW AI ACTUALLY BREAKS
FAILURE MODES	Enumerated in advance.	A novel output type nobody anticipated.
SEVERITY	Maps to a known runbook.	Behaviour drifts with no code change.
ROOT CAUSE	In code, config or infra.	In prompts, data, or the world.
CONTAINMENT	Restart, rollback or block.	May mean halting a class of use, not one server.
COMMS	A template exists for this scenario.	First time anyone has seen this — no template fits.

Incident versus surprise

Incidents are events you anticipated; surprises are events your playbook did not.

	INCIDENT • known failure mode	SURPRISE • novel failure mode	NAME THE SURPRISE OWNER FIRST
RECOGNISED BY	Pattern-matches a prior incident or the threat model.	Pattern does not fit — and that mismatch is the signal.	● A named seat owns declaring "this is a surprise" and naming it.
RESPONSE BASIS	Playbook plus a named runbook owner.	Surprise discipline plus a named Surprise Owner.	● It only fires on novelty — but it must exist before novelty arrives.
TIME TO ACT	Minutes — the paths are pre-wired.	Slower — first bound it, then decide.	● If the seat is empty when the surprise lands, the response is hoping.
LEARNING LOOP	Update the existing playbook; tune detection.	Write the new playbook entry the org never had.	

The most expensive mistake is forcing a surprise into the wrong playbook — so classify before you respond.

The Response Net

Same shape, different speed. Surprise stretches Detect and Diagnose; incidents move fast through both.

MODE	01 DETECT	02 CONTAIN	03 DIAGNOSE	04 REMEDIATE	05 LEARN
INCIDENT	Alert fires; known signature.	Trigger the pre-wired runbook.	Cause is on the known list.	Restore from the runbook fix.	Tune detection; refresh runbook.
SURPRISE	Something is off; define 'off' first.	Halt the smallest class that holds it.	Cause may be novel; cross-discipline triage.	Absorb, restrict or roll back; name the owner.	Write the entry the org didn't have.

When a control fails, or none existed, the Response Net catches it and bounds the blast.

Speed without bounding is panic. Bounding without speed is paralysis. Both fail.

Contain proportionately

Over-contain and you stop value; under-contain and it spreads. Pick the smallest ring that holds it, and escalate one ring at a time.

RING 01 • REQUEST

- Reject the one request.
- Hand off to a human.
- Tighten guardrails for this caller.
- USE FOR clear single-user harm.

RING 02 • USER GROUP

- Pause the feature for a cohort.
- Throttle a tenant.
- Restrict to safe inputs only.
- USE FOR an emerging pattern, scope unclear.

RING 03 • USE CASE

- Halt the use case end-to-end.
- Route to a human-only path.
- Pin to the prior model version.
- USE WHEN the use case itself is suspect.

RING 04 • CAPABILITY

- Disable the underlying capability.
- Block the tool or connector for all.
- Withdraw a model from rotation.
- USE ONLY when the capability is the root.

Move outward one ring at a time — and write down what made you escalate.

Preserve evidence, learn and govern back in

LEARN

- Name the surprise — give it a tag the org can refer back to, not a ticket number.
- Write the new playbook entry — next time this shape appears, it's an incident.
- Feed Module 11 controls and Module 12 heartbeats: new controls, new thresholds.
- Close the Module 13 evidence record — complete and queryable.

DOWNSTREAM • MODULE 12 LIVING GOVERNANCE

- Each response feeds new controls, heartbeats and thresholds.
- The loop refreshes, so next time the surprise is an incident.
- A response that teaches nothing leaves the spine no smarter.



Living Loop • Front Door Learning returns to the Living Loop and may update Front Door entry criteria.

Wyreld: an AI surprise

Sales wired the catalogue agent into live pricing and stock so it could "help them order" — pushing it from FAQ (L) toward XL.

STAGE	WHAT HAPPENED AT WYRELD CO.
DETECT	A Cork merchant rang to confirm an order: the agent had quoted a discontinued cladding panel at last year's price and taken it. No alert — a phone call. Class: surprise.
CONTAIN	The Surprise Owner switched the agent to read-only answers (Ring 03) — it could inform but not quote prices or place orders — while pricing was checked. Other use stayed live.
PRESERVE	The conversation transcript, stamped with model and catalogue version, was kept as evidence.
DIAGNOSE	The cause sat across DATA and WORLD — the catalogue feed had gone stale and the line was withdrawn, but the agent kept answering with old context.
REMEDiate	Re-pointed the agent at the live price/stock source; added a guardrail blocking any binding quote on discontinued or out-of-stock lines; restored ordering behind it.
GOVERN BACK IN	M10 re-sized it XL once it commits prices; M15 added an intake gate; M11 added a price/stock-commitment guardrail + AI-disclosure banner; M12 set a freshness heartbeat.

Detect, contain, preserve, then govern it back in — so the second time, it is an ordinary incident the spine handles by design.

Put the Journey back together

What does the whole system reveal once every stage is understood?

J01 · Synthesis

Read the pattern, reflect privately, understand the PUBLIC boundary and choose among neutral next routes.

FRONT DOOR · M15

SIZE · M10



CONTROL · M11



KEEP-LIVE · M12



PROVE · M13

RESPONSE NET · M14

The whole J01 system

Now each stage has been opened, the same map reads as one dependent system.

FRONT DOOR • M15 Use-case Intake & Gates • Where do AI use cases enter, and what bypasses that route?

SIZE

M10
AI Risk Sizing

How big is this, and would two reviewers agree?

CONTROL

M11
AI Risk Controls

Which controls are required, and who owns them?

KEEP-LIVE

M12
Living Governance

What changes after launch, and what wakes governance up?

PROVE

M13
AI Evidence & Audit-Readiness

If proof were requested tomorrow, where would it come from?

RESPONSE NET • M14 Incident & Surprise Response • Who declares an AI surprise, and who owns the first decision?

🔄 Learning returns to the Living Loop and may update Front Door entry criteria.

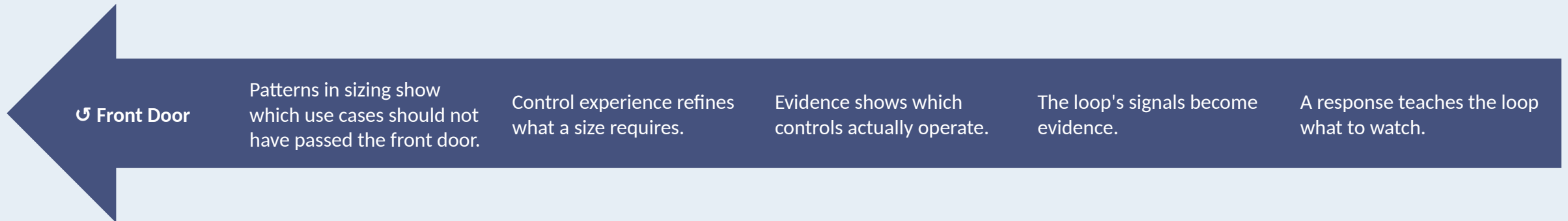
The route is iterative: change, evidence and incidents can reopen earlier judgements.

The Journey runs forwards and learns backwards

FORWARDS



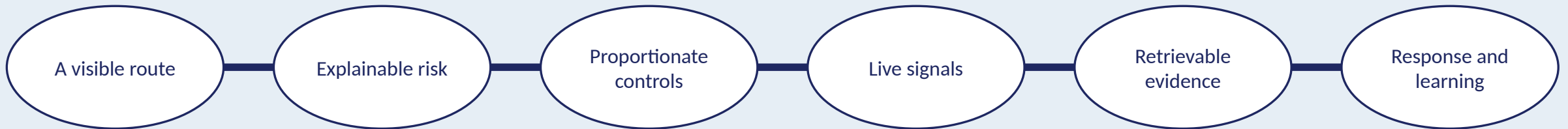
BACKWARDS



This describes how the system learns. It is not a routing rule and it selects nothing.

Read the pattern, do not calculate a score

Having read the six stages, you will have formed an impression. This is a way to think about it. It is educational, not an assessment: no score, no benchmark, no comparison and nothing that produces a rating.



Look for: Breaks between links. • Uncertainty about how a stage works in practice. • Weak hand-offs from one stage to the next.

Shared language: if conversations stall on what "governance", "control" or "oversight" mean, a shared picture may matter more than a new control.

Consistency and ownership: if the pieces exist but are used inconsistently, ask which would survive being tested.

Proof and learning: if use cases are sized and controlled but nobody could quickly show what operated, look at evidence and response.

Uncertainty: not knowing how a stage works in practice is worth recording honestly. Not knowing is different from doing badly.

One observer inside a system usually reports it as more consistent than it is. Comparing impressions with a colleague in a different function is often more informative.

A generous private reflection

Six private questions, one for each stage. Note what you notice. Nothing here is collected, scored or turned into a capability state.

FRONT DOOR

Where does AI enter?

RISK SIZE

How is risk understood?

CONTROL FLOOR

What control floor is visible?

LIVING LOOP

What wakes governance?

EVIDENCE CHAIN

What evidence could be retrieved?

RESPONSE NET

Who owns the first response decision?

Your notes stay with you. The LO Journey Check offers the same reflection as a short private document.

What J01 changes

STAGE	FROM	TOWARDS
Front Door	Use cases arrive by whatever route they find.	One declared route, with three honest outcomes and a recorded reason.
Risk Size	Sizing by individual judgement, or after the build.	Repeatable sizing that two reviewers can explain, before controls are chosen.
Control Floor	Generic controls, or a size label with no consequence.	Controls proportionate to the size, each named, owned and evidenced.
Living Loop	Governance set at launch.	Governance that responds to the change, with silence treated as a signal.
Evidence Chain	Evidence assembled when someone asks.	Evidence designed and captured as work happens.
Response Net	Improvisation by whoever notices.	Declared decision and response routes, including for failures nobody predicted, with learning fed upstream.

An intended direction of travel, not a guaranteed outcome.

What PUBLIC gives you

This deck is the complete J01 Knowledge Layer, with L0 orientation. That value is yours whether or not you do anything else.



Shared language for AI adoption risk.



Six canonical stage models, from the Modules that own them.



One worked-example universe, Wyreld Co., carried through every stage.



A qualitative way to read the pattern.



A private reflection you can use alone or with colleagues.



Five neutral continuation routes, including doing nothing.

What PUBLIC deliberately does not try to do

Some material does a different job: diagnosing, deciding, implementing, facilitating, assessing, assuring or authorising. It needs instruments, calibration and judgement that a public deck cannot responsibly supply on its own.

NOT IN PUBLIC

Formal diagnosis.

Implementation methods and working canvases.

L1 and L2 assessment records.

Facilitation methods.

Benchmarking.

Assurance authority.

PLAINLY

This is a learning resource. It is not a benchmark, assurance review, certification, or legal, regulatory or professional advice. Reading it confers no capability, accreditation, assurance or delivery authority.

Where to go next

Five neutral routes. None is a default, and none is chosen by a score.

ROUTE	WHEN IT FITS	WHAT TO RECORD
Sustain	Sufficient state is present and the main need is to maintain, monitor or renew it.	Record the persistence or review trigger.
Deepen J01	J01 remains the right problem frame and a material gap or uncertainty remains.	Record the gap, evidence and intended outcome.
Another Journey	A different admitted problem is materially more important or better matched.	State why. Do not default merely because J01 evidence is incomplete.
No further action	No proportionate action is justified now.	Record the reason, owner and reconsideration trigger.
Poor fit / redirect	J01 is not the right intervention, authority is missing, the need is outside EYW's boundary or a specialist route is required.	Record the safe redirect.

Routing is a human-owned decision. No score or weakest area selects the next product. Insufficient Evidence or Unable to Conclude is recorded before routing where applicable. Safety, legal, privacy, security, regulatory or authority concerns pause product routing. Commercial preference cannot override product fit.

If someone asked tomorrow how your organisation knows its AI adoption is safe enough for its context, what evidence could it show?

EmPOWER Your World · J01 Make AI Adoption Safe · Simon Owen, Microsoft MVP