

Governance note — Problem Statement

A risk committee will ask: "Why AI?" before they ask anything else. Section 1 must make the problem case independently of the solution. If the problem statement only makes sense if AI is already the answer, rewrite it.

02 AI Suitability Assessment

Complete each question with a decision, evidence reference and owner. Any No or TBC is an open issue: identify its consequence and the evidence needed to resolve it. A material unresolved issue blocks Gate B; only the designated authority may accept a permissible condition under the firm's risk policy. Legal obligations and required validation cannot be waived by this form.

#	Gate Question	Yes	No	TBC
1	Is the problem well-defined with measurable outcomes?	☐	☐	☐
2	Is there sufficient, clean, labelled training or evaluation data available?	☐	☐	☐
3	Is AI materially better than a deterministic / rule-based solution?	☐	☐	☐
4	Can the system's outputs be explained to a risk committee or regulator?	☐	☐	☐
5	Is the failure mode acceptable — and have failure scenarios been documented?	☐	☐	☐
6	Have data privacy, bias, and regulatory constraints been assessed?	☐	☐	☐

2.1 Suitability Assessment Notes

For each No or TBC, record the evidence, risk, responsible owner, due date and gate disposition. Document the decision authority and link the firm's system of record. A checked Yes without evidence is incomplete.

2.2 Alternative approaches considered

What non-AI solutions were evaluated before recommending AI? Why were they insufficient?

Documents arrive via a monitored compliance inbox and an RSS feed aggregator. An analyst opens each document, reads the title and executive summary (approximately 8–12 pages for an FCA circular), applies a classification label from a 14-category taxonomy, and routes via SharePoint to the relevant team folder. No automated triage exists. Classification accuracy is estimated at 91% based on a manual audit of 200 documents in Q3 2025. Misclassified documents are typically caught in the receiving team's weekly review — an average of 3–4 days after initial routing. Three analysts cover this function across shift patterns; single points of failure exist during leave periods.

1.4 Why solve this now?

Why now (illustrative): document volume and missed internal deadlines have exposed capacity limits, and a 2026 technology budget is available. Whether operational resilience rules apply to the entity and any important business service requires a separate firm-specific assessment. The FCA rules came into force on 31 March 2022; the transition ended on 31 March 2025.

Outcome Dimension	Baseline	Target	Evidence / Owner
Classification accuracy	91% (manual audit Q3 2025)	97%+ (AI-assisted)	Compliance Ops lead
Time to route (per document)	35–45 minutes	Under 3 minutes	Compliance Ops lead
Deadline misses (per year)	2 confirmed, Q1–Q3 2025	Zero	Head of Compliance
Peak capacity (docs/week)	Degrades above 60	Stable at 120+	Compliance Ops lead

Governance note — Problem Statement

The risk committee will ask: "Why AI for this?" This problem statement demonstrates that the manual process has a documented failure rate (91% accuracy, two deadline misses) and a structural capacity constraint. Both are quantified, not asserted. Any spec without baseline numbers will be challenged here.

02 AI Suitability Assessment

Document known edge cases before testing. Eight illustrative scenarios are provided; add cases based on the use case and data. Each high-risk case must be tested before Gate C or explicitly block progression under firm policy.

ID	Edge Case Description	Trigger / Cause	Expected Behaviour	Actual Behaviour (post-test)	Risk Level	Test Status
EC-01	Document contains mixed regulatory regimes (e.g., FCA circular that references PRA standards)	Cross-regime regulatory publications increasing in frequency post-2024	Model assigns highest-confidence primary regime; flags as multi-regime with second suggestion shown to analyst	Pending testing	High	Pending — [date]
EC-02	New regulatory document format not in training data (e.g., unseen publication format)	Novel document types introduced by regulator post-training cutoff	Confidence score drops below 85% threshold; mandatory human review triggered; document not auto-suggested	Pending testing	High	Pending — [date]
EC-03	Document in language other than English (ESMA multilingual publications)	ESMA publishes in all EU languages; UK firm receives English version but metadata may be multilingual	Model processes English text correctly; multilingual metadata does not affect classification	Pending testing	Medium	Pending — [date]
EC-04	Duplicate document (same circular re-issued with minor amendment)	Regulators re-issue documents with version numbers or "amended" suffixes	Model treats as new document; analyst confirms or flags as duplicate during review	Pending testing	Low	Pending — [date]

1.3 What a correct output looks like

Describe — with an example if possible — what a successful response from this prompt looks like. Vague here means broken in production.

Governance note

If you cannot describe what a correct output looks like before writing the prompt, you are not ready to write the prompt. The acceptance criteria for a prompt must precede the prompt itself — exactly as acceptance criteria precede development in an Agile sprint.

02 Input Structure

2.1 What data or context does this prompt receive?

Input Element	Source	Format	Sensitive? (PII/confidential)	Validation Required?
Enter input element	Enter data source	Text / JSON / structured	Yes / No	Yes / No — describe validation

2.2 Input length and token budget

Maximum expected input tokens: _____ | Model context window: _____
Strategy for inputs that exceed context window: _____

2.3 Input sanitisation and injection risk

Identify untrusted sources, trust boundaries, tool permissions and expected attacks. Specify input handling, constrained tool access, output validation, human review and adversarial tests; prompt wording alone is not a security boundary.

03 Constraints

3.1 What this prompt must NOT do or produce

List all prohibited outputs, topics, tones, and behaviours. Be specific — "don't say anything bad" is not a constraint.

03 What Happens When It Gets It Wrong (3 minutes)

Delivery note

This is the most important section. A senior risk stakeholder does not expect zero errors — they expect that errors are contained, detected, and learned from. If you cannot answer all four questions below, you are not ready to present. "We are monitoring it" is not an answer.

Question the CRO will ask	Your answer (complete before the meeting)
How often does it get it wrong?	Enter: accuracy rate, error rate, and what the baseline (human error rate) is for comparison.
What happens when it gets it wrong?	Enter: the specific consequence, who is affected, what the recovery process is, and how quickly errors are typically detected.
What controls stop an error from causing harm?	Enter: the specific control at each point — confidence threshold, human review, routing audit, weekly sampling.
Who is accountable when it gets it wrong?	Enter: the specific named role (not "the system" or "the team") who is accountable for error investigation, remediation, and regulatory reporting if applicable.

Worst credible failure and containment: state the scenario, likelihood and impact evidence, preventive control, detection, recovery, owner and residual risk. Avoid promising it cannot happen.

Complete this sentence: "The worst credible failure scenario is [describe]. This is prevented by [specific control]. If that control fails, the fallback is [describe]. The accountability for this risk sits with [named role]."

04 Why We Can Trust It (3 minutes)

Evidence of quality before launch (replace example labels with actual reports and dates)

Evidence	Detail	Reference
Training data	Describe: volume, period, regime coverage, quality assessment	Reference: data audit report [date]
Test set accuracy	Enter: precision and recall on held-out test set	Reference: model validation report [date]
Edge case testing	Enter: number of edge cases tested, outcome	Reference: Edge Case & Bias Risk Register [date]
Regulatory alignment review	Enter: frameworks assessed, outcome	Reference: AI Feature Spec Section 3.1 [date]

#	Checklist Item	Status	Owner	Evidence Reference	Notes
A5	Accountable business owner is named and confirmed.				
A6	Materiality assessment: is this feature material to the business, to clients, or to regulators?				
A7	Initial regulatory triage completed: frameworks potentially in scope identified.				

Gate A Sign-off	Name	Role	Date	Decision
Programme Lead / Product Owner				Open / Conditional / Not open

Complete the required design evidence before build. Items dependent on test results must instead have an approved test plan, measure, owner and due gate. Risk and Compliance record their Gate B decision under firm policy.

#	Checklist Item	Status	Owner	Evidence Reference	Notes
B1	AI Feature Spec Sections 01–06 completed to the Gate B design standard, including proposed risks, metrics, controls and rollout. Section 07 records sign-off.				
B2	AI Suitability Assessment completed: all 6 gate questions answered with evidence.				
B3	Intended use and prohibited use are explicitly documented.				
B4	Data sources identified. Data owner confirmed. Data quality assessment completed.				
B5	Training data provenance documented. No unlicensed or privacy-breaching data in scope.				
B6	Key failure modes documented in Failure Mode Register (Section 04 of spec).				