SPANFORGE

AI Opportunity & Problem Qualification Assessment

Discover Phase · Gate 1 Pre-Qualification · Enterprise Governance Artefact

VERSION 1.0 | ALIGNED TO THE SPANFORGE WAY v3 | CONFIDENTIAL

Field	Detail
AI Initiative Name:	
Business Unit / Owner:	
Assessment Lead:	
Named Gate Authority (budget-accountable sponsor):	
Date of Assessment:	
Assessment Version:	
Stakeholders Present:	

Portfolio Context	
Relative Priority vs Other Initiatives:	☐ High — top-3 programme priority ☐ Medium — planned but not urgent ☐ Low — exploratory
Competing AI Initiatives (name any in flight):	
Portfolio Prioritisation Rationale:	
Resource Conflict Risk:	☐ No conflicts identified ☐ Shared team / budget with: _____

⛔ Gate Authority Precondition

SpanForge Non-Negotiable: A named Gate Authority with budget accountability must be identified before this assessment begins. Without this, any gate decision produced is advisory only and cannot be enforced. Record this name above before scoring.

⚠ Before You Start — Confirm Scope Fit

Scope Check: This tool is designed for AI initiatives with material business impact, production integration requirements, and cross-functional accountability. It is NOT for: exploratory R&D or horizon-scanning (apply time/budget boundary only); low-risk internal automation with limited blast radius (use lightweight two-gate model); technical spikes under six weeks (use as Gate 1 input, not a programme). Test: If this initiative could affect customers, regulated data, or require significant coordination — proceed.

Assessment History			
Version	**Date**	**Lead**	**Change / Reason**

FACILITATOR GUIDE

How to Run This Assessment

Read before starting · 10 minutes · Assessment Lead only

Who Should Attend

Role	Why Required	Must Attend?	Time Needed
Assessment Lead	Facilitates, documents scores and evidence	Yes — mandatory	Full session
Gate Authority / Executive Sponsor	Non-negotiable SpanForge precondition; makes binding gate decisions	Yes — mandatory	Full session
Business Domain Expert	Scores automation potential, human oversight, and value hypothesis	Yes — mandatory	Full session
AI / Technical Lead	Scores data, infrastructure, and skills dimensions	Yes — mandatory	Full session
Compliance / Legal	Scores regulatory and T.R.U.S.T.™ dimensions	Highly recommended	Dimensions 4 & 5
Data Steward / Owner	Confirms data access, quality, and governance status	Recommended	Dimensions 2 & 3

Session Structure & Timeboxing

Block	Activity	Time	Owner
Setup	Confirm attendees, confirm Gate Authority name, explain scoring rules, scope check	**10 min**	Lead
Phase 1 — D0	Problem Statement Quality — score + evidence	**10 min**	Business + Lead
Phase 1 — D1	Automation Potential — score + evidence	**10 min**	All
Phase 1 — D2	Data Availability — score + evidence	**10 min**	Technical
Phase 1 — D3	Data Risk & Fitness for AI — score + evidence	**10 min**	Technical + Data
Phase 1 — D4	Business Value & ROI — score + evidence	**15 min**	Business + Lead
Phase 1 — D5	Regulatory & Ethical Risk — score + evidence	**10 min**	Compliance
Phase 1 — D6	Human Oversight — score + evidence	**10 min**	Business + Technical
Phase 1 Gate	Tally score, gate decision, record rationale	**10 min**	Gate Authority
Phase 2 — D7–D11	Readiness dimensions (only if Phase 1 ≥ 45)	**35 min**	All
Phase 2 Gate	Tally score, gate decision, record rationale, commit next steps	**10 min**	Gate Authority

Close	Assign sign-off, schedule Gate 1 review if advanced	**5 min**	Lead

Typical Assessment Completion Time

Format	Participants	Total Time	Recommended When
Single session (standard)	4–6 people	**90–120 mins**	First-time assessment; all stakeholders available in one sitting
Two sessions	4–6 people	**2 × 60 mins**	Phase 1 first; Phase 2 after data and governance review
Async + review	4–6 people	**2–3 days**	Participants pre-score individually; Lead consolidates; 45-min alignment review

SpanForge Platform

This assessment can be executed via the SpanForge platform for automated scoring, real-time RAG tracking, and instant governance-ready reporting.

Platform features:

- Auto-calculated weighted scores
- Live blocker flagging
- Exportable audit-ready PDF
- Portfolio dashboard across initiatives
- Direct feed into Gate Readiness Score™

getspanforge.com/platform

Scoring Rules — Mandatory

The Scoring Scale		
Score	**Meaning**	**Evidence Rule**
0	**Not Present / Not Applicable**	No evidence required — state the reason for absence
1	**Minimal / Early Stage**	Describe current state; no formal evidence required
2	**Developing / Partial**	EVIDENCE REQUIRED — document what exists, what is missing, and the source
3	**Strong / Well-Established**	EVIDENCE REQUIRED — provide specific, verifiable proof with a named source and date

Facilitator Tip: Disagreements Are Data

Disagreements between participants are valuable signal — they reveal assumption gaps. Do not average them away. Record the disagreement, the positions held, and the reasoning. This is governance-grade evidence.

Common Facilitator Pitfalls

⚠ **Scoring by assumption**	Scores above 1 require documented evidence. If the team "believes" data is good but has not checked, score 1.
⚠ **Letting optimism advance an initiative**	If the Gate Authority cannot cite evidence for a gate condition, the condition is not met. Enthusiasm is not evidence.
⚠ **Skipping Phase 1 gate**	Phase 2 may not begin if Phase 1 score is below 45. This is non-negotiable.
⚠ **Leaving blockers undocumented**	Any score of 0 in a flagged blocker question must be recorded in the gate decision rationale and escalated.
⚠ **Averaging disagreements**	Record disagreements explicitly. They are evidence of assumption risk.

Scenario Guidance — Good vs Bad AI Problems

✓ Good AI Problem

"We process 5,000 insurance claims per day. 80% follow identical review rules. We have 3 years of labelled outcomes. Our compliance team has reviewed AI use in this context. The CFO is the sponsor with approved budget."

"Our customer support handles 2,000 tickets/day. 70% are classified into 12 known categories. We have 18 months of labelled tickets. Response time SLA is measurable. Head of CX is our Gate Authority."

✗ Bad AI Problem

"We want to use AI to improve decision-making." (No KPI, no data, no scope, no sponsor.)

"AI could help our analysts be more productive." (Vague benefit, no baseline, no data audit, no compliance review.)

"The CEO mentioned AI in the all-hands." (Not a problem statement. No Gate Authority. No budget.)

◈ PHASE 1 · QUALIFY

Is this worth solving with AI?

Complete all 7 dimensions · Gate threshold: Score ≥ 45 to proceed to Phase 2 · Estimated time: 85 minutes

D0 — Problem Statement Quality

Time	Weight
10 min	15%

⚠ **Why This Comes First**

SpanForge Gate 1 Non-Negotiable: Nothing progresses to Gate 1 without a signed Problem Statement Canvas. This dimension assesses the quality and rigour of the problem definition before scoring any other dimension. "Wrong problem defined" is the most common cause of Gate 1 failure — it is cheaper to identify it here than at Gate 3.

Evaluates whether the AI initiative is built on a clear, agreed, and bounded problem definition. High-quality problem statements prevent the most expensive failure mode in enterprise AI: building a technically correct solution to the wrong problem.

Low Score Means:

The problem is vague, contested, or too broadly defined to produce a testable success criterion. Any technical solution will be argued against by stakeholders who had a different problem in mind.

Typical Failure Pattern:

Teams spend months building, then discover at Gate 3 that different stakeholders had fundamentally different expectations of what the AI should do. Pilot declared a technical success but business rejected. Forced redesign or cancellation.

Q1 How clearly and specifically is the problem defined — in writing?

Score	Description	Evidence Required to Justify Score ≥ 2
0	☐ No written problem statement exists; the problem is described verbally and variably by different stakeholders	—
1	☐ A draft problem statement exists but is imprecise, contested, or describes a desired solution rather than a problem	—
2	☐ A written problem statement exists that most stakeholders agree on, but boundary conditions are not fully defined	*Attach or cite the draft statement — note version and author*
3	☐ A signed, precise problem statement exists with agreed scope, boundary conditions, and out-of-scope exclusions	*Cite the signed Problem Statement Canvas — include version number and sign-off date*

📋 **Evidence / Notes (mandatory for score ≥ 2):**

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Q2 Are all key stakeholders aligned on what the AI should and should not do?

Score	Description	Evidence Required to Justify Score ≥ 2
0	☐ Significant disagreement exists — stakeholders have materially different expectations of the outcome	—
1	☐ General alignment exists but there are known gaps or assumptions that have not been explicitly tested	—
2	☐ Stakeholders have been consulted and most are aligned; one or two outstanding concerns are documented	*Name the alignment exercise conducted and list any unresolved concerns*
3	☐ All key stakeholders have formally reviewed and signed off on the problem scope and expected AI behaviour	*Cite the sign-off record — list all stakeholders who formally confirmed scope*

📋 **Evidence / Notes (mandatory for score ≥ 2):**

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Q3 Are the boundaries of the AI initiative explicitly defined — what is in scope and out of scope?

Score	Description	Evidence Required to Justify Score ≥ 2
0	☐ No boundary definition exists; the scope expands or contracts depending on who is asked	—
1	☐ An informal boundary is understood but not documented; scope creep risk is high	—
2	☐ A boundary is documented but has not been formally reviewed or stress-tested against stakeholder expectations	*Cite the boundary document and note any open scope questions*
3	☐ A clear, documented, reviewed boundary definition exists; out-of-scope decisions are explicitly stated and agreed	*Cite the boundary document, the reviewing authority, and confirm out-of-scope examples are explicit*

📋 **Evidence / Notes (mandatory for score ≥ 2):**

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Dimension 0 Score

⚠ Score of 0 on Q1 = problem statement must be produced before any other scoring is meaningful

Failure mode triggered if score < 3 on any blocker question: **document & escalate**

D0 Raw Score

____ / 9

Weighted contribution: ____

D1	Automation Potential	Time	Weight
		10 min	18%

Assesses whether the task is structurally suited to AI automation. High-scoring tasks are pattern-based, high-volume, and have clear definitions of success. If a human cannot write rules for the task, AI will struggle too.

Low Score Means:
The task is too creative, context-dependent, or judgement-heavy for reliable automation. AI will produce inconsistent outputs with unacceptable error rates.

Typical Failure Pattern:
Teams build AI for complex reasoning tasks, achieve high demo performance on curated data, but fail in production on edge cases. Investment wasted; manual fallback reinstated.

Q4 How repetitive or pattern-based is the core task?

Score	Description	Evidence Required to Justify Score ≥ 2
0	☐ Highly creative or one-off — success depends on human judgement each time	—
1	☐ Partly structured; judgement needed but rules exist for portions of the task	—
2	☐ Mostly rule-driven with occasional exceptions requiring human review	*Description of rule structure and exception rate (approximate %)*
3	☐ Highly repetitive with consistent inputs, outputs, and decision rules	*Process documentation or SOP reference showing rule consistency and volume data*

📋 **Evidence / Notes (mandatory for score ≥ 2):**

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Q5 Can success be clearly defined and measured before build starts?

Score	Description	Evidence Required to Justify Score ≥ 2
0	☐ No clear metric exists; quality is entirely subjective	—
1	☐ Qualitative proxy metrics only (e.g. stakeholder satisfaction)	—
2	☐ Quantitative metrics exist but are imperfect or lagging indicators	*Name the metrics and current measurement method*
3	☐ Clear, measurable KPIs already tracked and trusted with a known baseline	*KPI name, baseline value, measurement system, and owner*

📋 **Evidence / Notes (mandatory for score ≥ 2):**

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Q6 What is the volume and frequency of this task?

Score	Description	Evidence Required to Justify Score ≥ 2
0	☐ Rare or ad-hoc (fewer than monthly occurrences)	—
1	☐ Monthly or quarterly — low volume	—
2	☐ Weekly or moderate volume (provide estimate)	*State approximate volume and source (system log, manual count, estimate)*
3	☐ Daily or high-volume — creates meaningful operational burden	*Provide volume data from source system or operational report*

📋 **Evidence / Notes (mandatory for score ≥ 2):**

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Dimension 1 Score

Failure mode triggered if score < 3 on any blocker question: **document & escalate**

D1 Raw Score

____ / 9

Weighted contribution: ____

D2	Data Availability	Time	Weight
		10 min	13%

Evaluates whether sufficient, accessible, quality data exists to ground or train an AI solution. This dimension focuses on existence and access. Dimension 3 focuses on fitness for AI use.

Low Score Means:
Insufficient or inaccessible data means any AI system will either be untrained, poorly trained, or trained on data that cannot legally or practically be used.

Typical Failure Pattern:
Teams proceed without data audit, discover access or volume problems mid-build, and either delay for months or ship with inadequate training data — causing post-production model failure.

Q7 How much relevant historical data is available for this task?

Score	Description	Evidence Required to Justify Score ≥ 2
0	☐ Little to none — data would need to be generated from scratch	—
1	☐ Some data exists but is inconsistently captured or siloed	—
2	☐ Good volume available with known quality issues that can be addressed	*State estimated volume (rows/records), time range, and storage location*
3	☐ Large, well-labelled, accessible dataset ready for use	*Provide dataset metadata: volume, date range, format, storage system, and labelling status*

📋 **Evidence / Notes (mandatory for score ≥ 2):**

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Q8 How clean and structured is the existing data?

Score	Description	Evidence Required to Justify Score ≥ 2
0	☐ Unstructured, scattered across systems with no governance	—
1	☐ Partially structured; significant cleansing effort required	—
2	☐ Mostly structured with documented quality issues	*Describe quality issues and state whether a remediation plan exists*
3	☐ Well-structured, validated, and actively maintained	*Reference data quality report or system audit confirming structure and maintenance regime*

📋 **Evidence / Notes (mandatory for score ≥ 2):**

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Q9 Is data access and ownership clearly established?

Score	Description	Evidence Required to Justify Score ≥ 2
0	☐ Ownership is unclear; access requires significant approvals or legal review	—
1	☐ Ownership identified but access processes are slow or incomplete	—
2	☐ Access is possible; minor permissions or privacy considerations to resolve	*Name the data owner and describe the remaining access steps*
3	☐ Full access confirmed; data governance and consent documented in place	*Provide the access approval reference, consent documentation, and data owner name*

📋 **Evidence / Notes (mandatory for score ≥ 2):**

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Dimension 2 Score	D2 Raw Score
Failure mode triggered if score < 3 on any blocker question: **document & escalate**	**_____ / 9** Weighted contribution: ____

D3 Data Risk & Fitness for AI

⏰ Time: 10 min | Weight: 13%

⍰ Why This Dimension Exists

This dimension is distinct from availability. Data can exist and be accessible but still be unfit for AI use due to bias, poor representativeness, or unreliable ground truth. All three questions must be scored before proceeding.

Moves beyond "does data exist" to "is this data safe and appropriate to train an AI on." Low scores here indicate model failure risk even when data appears abundant.

Low Score Means:

Even with abundant data, the AI will learn and amplify historical biases, perform poorly on underrepresented groups, or produce outputs that cannot be trusted because ground truth is contested.

Typical Failure Pattern:

Model achieves high accuracy on training set but performs poorly in production on edge cases, minority groups, or time periods not represented in training data. Compliance or fairness issues emerge post-deployment.

Q10 How representative is the data of the full population this AI will encounter in production?

Score	Description	Evidence Required to Justify Score ≥ 2
0	☐ Highly skewed — certain groups, time periods, or conditions are substantially underrepresented	—
1	☐ Some representativeness gaps identified; impact on model performance is unclear	—
2	☐ Reasonably representative with documented gaps and a mitigation plan	*Describe representativeness gaps and mitigation approach (e.g. oversampling, synthetic data)*
3	☐ Highly representative — coverage of population validated with demographic or distributional analysis	*Cite the analysis or audit — include methodology and date conducted*

📋 **Evidence / Notes (mandatory for score ≥ 2):**

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Q11 What is the risk of bias, discrimination, or unfair outcomes from this data?

Score	Description	Evidence Required to Justify Score ≥ 2
0 ⛔	☐ High — data reflects historical decisions that were biased; protected characteristics present or correlated	*BLOCKER*
1	☐ Moderate — potential for bias exists; no formal assessment has been conducted	—
2	☐ Low — bias risk assessed; mitigations identified and planned	*Attach the bias risk assessment and describe the mitigation approach*
3	☐ Negligible — formal fairness assessment completed; no material bias risk identified	*Reference fairness audit with methodology and findings*

📋 **Evidence / Notes (mandatory for score ≥ 2):**

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Q12 How reliable is the labelling or ground truth for this task?

Score	Description	Evidence Required to Justify Score ≥ 2
0	☐ No labelled data; ground truth does not exist or is actively disputed	—
1	☐ Limited labelled examples; labelling quality is inconsistent or inter-annotator agreement is low	—
2	☐ Partial labelling with reasonable confidence; inter-annotator agreement acceptable but not formally measured	*State labelling methodology and estimated agreement rate*
3	☐ Comprehensive, high-confidence labelling; inter-annotator agreement formally measured and acceptable	*Reference labelling methodology, volume, and agreement score*

📋 **Evidence / Notes (mandatory for score ≥ 2):**

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

⛔ **Automatic Blocker — Bias Risk**

A score of 0 on Q11 (bias risk) is an automatic blocker. No AI initiative should proceed without a formal bias risk assessment when the use case involves decisions affecting individuals. Escalate to compliance before Phase 2.

Dimension 3 Score

⚠ Score of 0 on Q11 = automatic blocker regardless of total score

Failure mode triggered if score < 3 on any blocker question: **document &**

D3 Raw Score

____ / 9

escalate	Weighted contribution: ____

D4 Business Value Hypothesis & ROI

Time: 15 min | Weight: 18%

> ⚠ **SpanForge Gate 1 Precondition**
>
> SpanForge Non-Negotiable: No initiative progresses to Gate 1 without a signed Problem Statement Canvas that includes a named value hypothesis, a baseline KPI, and a credible ROI model. This dimension assesses readiness to produce that artefact. Evidence of value must exist before build begins — not after.

Ensures the initiative has a credible, measurable business case. Aligns with SpanForge's cost model and value tracking framework. Without this, Gate 3 (Pilot KPI validation) becomes unmeasurable.

Low Score Means:

The initiative cannot demonstrate value at Gate 3 (Pilot). KPIs cannot be validated. Budget cannot be justified. ROI tracking is impossible — the initiative becomes a cost centre with no accountability mechanism.

Typical Failure Pattern:

Teams pilot AI successfully in technical terms but cannot show business impact. Sponsors lose confidence. Funding cut. Initiative quietly abandoned — the #1 enterprise AI failure mode.

Q13 Is there a clearly defined value hypothesis with a measurable baseline KPI?

Score	Description	Evidence Required to Justify Score ≥ 2
0	☐ No — the expected benefit has not been defined or quantified	—
1	☐ Directional — a general benefit is assumed but no baseline measurement exists	—
2	☐ Partial — a value hypothesis is documented with some supporting data; baseline is incomplete	*State the hypothesis, the KPI, and what baseline data currently exists*
3	☐ Defined — a specific, measurable hypothesis with a validated baseline KPI is documented and signed off	*Cite the Problem Statement Canvas — name the KPI and state the baseline value*

📋 **Evidence / Notes (mandatory for score ≥ 2):**

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Q14 Can a credible ROI model be constructed for this initiative?

Score	Description	Evidence Required to Justify Score ≥ 2
0	☐ No — cost and benefit structure cannot be estimated at this stage	—
1	☐ Weak — high-level estimates only; major assumptions have not been validated	—
2	☐ Developing — a draft ROI model exists with identified assumptions to validate in Phase 2	*Attach the draft ROI model — list the top 3 assumptions requiring validation*
3	☐ Strong — a reasoned ROI model with cost, benefit, and risk-adjusted scenarios is documented and reviewed	*Cite the ROI model and assumptions log — name the reviewing authority*

📋 **Evidence / Notes (mandatory for score ≥ 2):**

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Q15 What is the expected value type and has opportunity cost of delay been considered?

Score	Description	Evidence Required to Justify Score ≥ 2
0	☐ Value type undefined; no consideration of cost of inaction or delay	—
1	☐ Value type identified (cost saving / revenue / efficiency) but magnitude is estimated without supporting data	—
2	☐ Value type and rough magnitude estimated; opportunity cost of delay partially considered	*State value type, estimated magnitude, and time-to-value assumption*
3	☐ Value type, magnitude, and opportunity cost of delay formally modelled; reviewed by finance or senior stakeholder	*Cite the business case — name the finance reviewer and approval date*

📋 **Evidence / Notes (mandatory for score ≥ 2):**

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Q16 Has the ROI model been stress-tested across best-case, base-case, and worst-case scenarios?

Score	Description	Evidence Required to Justify Score ≥ 2
0	☐ No — only a single ROI estimate exists with no scenario testing	—
1	☐ Partial — a downside scenario has been considered informally but not modelled	—
2	☐ Developing — best/worst-case ranges have been discussed and documented; key sensitivities identified	*Cite the scenario analysis — identify the primary sensitivity drivers (e.g. adoption rate, data quality)*
3	☐ Stress-tested — formal sensitivity analysis completed across ≥3 scenarios; break-even conditions defined; reviewed by finance	*Cite the sensitivity model and scenario outputs — confirm break-even threshold is explicit and reviewed*

📋 **Evidence / Notes (mandatory for score ≥ 2):**

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Dimension 4 Score

Failure mode triggered if score < 3 on any blocker question: **document & escalate**

D4 Raw Score

____ / 9

Weighted contribution: ____

D5	Regulatory & Ethical Risk	⏰ Time	Weight
		10 min	13%

🚫 Inverse-Scored Dimension — 0 = Highest Risk

This dimension is inverse-scored. A score of 0 represents the highest risk. A score of 0 in any question here is an automatic blocker regardless of total score — document the risk, halt the assessment, and seek compliance review before proceeding.

Identifies compliance, legal, and ethical exposure before any architecture or build commitment. Aligned to SpanForge T.R.U.S.T.™ Framework dimensions: Transparency and User Rights.

Low Score Means:

The initiative operates in a high-risk regulatory environment without adequate controls. Post-deployment regulatory action, reputational damage, or individual harm becomes likely.

Typical Failure Pattern:

Teams build and deploy without completing regulatory mapping. A post-deployment compliance review identifies violations. Forced rollback or remediation at full production cost.

✦ T.R.U.S.T.™ Framework: Transparency · User Rights

T.R.U.S.T.™ Alignment — Transparency: Customers, regulators, and employees must understand how AI affects them. User Rights: Individuals affected by AI decisions must have the right to understand and seek redress. Both must be assessed here.

Q17 What is the regulatory complexity of this use case?

Score	Description	Evidence Required to Justify Score ≥ 2
0 🚫	☐ Highly regulated domain (healthcare, finance, legal, HR) with AI-specific rules in force	*BLOCKER — compliance review mandatory before proceeding*
1	☐ Regulated domain with general data privacy or sector rules that apply	—
2	☐ Low regulation but significant internal policy or reputational considerations	*Identify applicable policies and responsible owner*
3	☐ Minimal regulatory exposure; only general data handling standards apply	*Confirm with legal or compliance that no domain-specific rules apply*

📋 **Evidence / Notes (mandatory for score ≥ 2):**

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Q18 What is the potential for bias, discrimination, or unfair outcomes on affected individuals?

Score	Description	Evidence Required to Justify Score ≥ 2
0 ⊘	☐ High risk — decisions directly affect individuals' rights, access, employment, credit, or safety	*BLOCKER — fairness assessment mandatory*
1	☐ Moderate risk — outputs influence human decisions with downstream individual impact	—
2	☐ Low risk — outputs are informational or internal; limited direct individual impact	*Confirm that outputs do not feed into consequential decisions without human review*
3	☐ Negligible risk — no impact on individuals; fully internal or operational use only	*Confirm use case scope in writing*

📋 **Evidence / Notes (mandatory for score ≥ 2):**

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Q19 How explainable must the AI's decisions be to affected stakeholders?

Score	Description	Evidence Required to Justify Score ≥ 2
0	☐ Full explainability required by law or policy for each individual decision (e.g. GDPR Article 22, EU AI Act)	—
1	☐ High explainability expected — stakeholders and regulators will demand reasoning	—
2	☐ Audit trail sufficient — some transparency needed; individual-level explanation not required	*Confirm explainability requirement with compliance*
3	☐ Explainability is not a material requirement for this use case	*Confirm with legal and document rationale*

📋 **Evidence / Notes (mandatory for score ≥ 2):**

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Dimension 5 Score

⚠ Score of 0 on Q17 or Q18 = automatic blocker

Failure mode triggered if score < 3 on any blocker question: **document & escalate**

D5 Raw Score

_____ / 9

Weighted contribution: ____

D6	Human Oversight Requirements	⏰ Time	Weight
		10 min	10%

Determines the level of human-in-the-loop design required. High oversight requirements are not disqualifying but significantly affect solution architecture, operating costs, and timelines. SpanForge requires human-in-the-loop design to be specified before Gate 1. Aligned to T.R.U.S.T.™ Safety Guardrails dimension.

✦ T.R.U.S.T.™ Framework: Safety Guardrails

T.R.U.S.T.™ Alignment — Safety Guardrails: Technical constraints must be embedded in architecture, not left as policy aspirations. This dimension identifies what guardrails are needed and whether they can be built.

Low Score Means:

The process cannot tolerate AI errors at production volume, or meaningful human review is impossible. The AI cannot be safely deployed without unacceptable risk or the elimination of all efficiency gains.

Typical Failure Pattern:

Team builds a fast, high-volume AI system but fails to design human review into the architecture. Post-deployment, errors are caught too late or not at all. Regulatory breach or reputational incident triggers forced shutdown.

Q20 How tolerant is this process of AI errors?

Score	Description	Evidence Required to Justify Score ≥ 2
0 ⛔	☐ Zero tolerance — errors have legal, safety, or severe financial consequences	*BLOCKER — human-in-the-loop design is mandatory; document control architecture before Gate 1*
1	☐ Low tolerance — errors cause significant rework or reputational damage	—
2	☐ Moderate tolerance — errors are recoverable with acceptable cost	*Describe the error recovery process and cost estimate*
3	☐ High tolerance — errors are low-stakes and easily caught downstream	*Confirm error handling mechanism*

📋 **Evidence / Notes (mandatory for score ≥ 2):**

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Q21 Can humans effectively review AI outputs at the expected production volume?

Score	Description	Evidence Required to Justify Score ≥ 2
0	☐ No — volume and speed make meaningful human review impossible	—
1	☐ Unlikely — review is theoretically possible but would eliminate efficiency gains	—
2	☐ Yes, with process redesign — sampling or tiered review approach could work	*Describe the proposed review design and sampling methodology*
3	☐ Yes — human review is already built into the process and can be maintained at scale	*Reference existing review process and confirm capacity*

📋 **Evidence / Notes (mandatory for score ≥ 2):**

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Q22 Is there a clear override or escalation path when AI output is wrong?

Score	Description	Evidence Required to Justify Score ≥ 2
0	☐ No mechanism exists; there is no fallback if AI fails	—
1	☐ A fallback exists but is not documented or tested	—
2	☐ A fallback is documented but not regularly exercised	*Cite the fallback documentation — state when it was last tested*
3	☐ A tested, well-understood override process is already in place	*Reference override SOP, test log, and responsible team*

📋 **Evidence / Notes (mandatory for score ≥ 2):**

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Dimension 6 Score

⚠ Score of 0 on Q20 = mandatory human-in-the-loop design before Gate 1

Failure mode triggered if score < 3 on any blocker question: **document & escalate**

D6 Raw Score

____ / 9

Weighted contribution: ____

◈ PHASE 1 GATE

Qualify Score Summary & Gate Decision

Gate Authority signature required to advance

Failure Mode Mapping — What Low Scores Signal

Dimension	Low Score Failure Mode	Risk Category	SpanForge Risk Index
D0 — Problem Statement	Vague or contested problem definition → solution built for wrong problem → stakeholder rejection at Gate 3 regardless of technical quality	**Delivery Risk**	*Gate 1 failure (wrong problem)*
D1 — Automation Potential	AI applied to task not suited to automation → inconsistent outputs, high error rates, user rejection	**Delivery Risk**	*Gate 2 failure (PoC accuracy)*
D2 — Data Availability	Build begins without data → discovery mid-build, months of delay, possible cancellation	**Data Risk**	*Gate 2 failure (no training data)*
D3 — Data Risk & Fitness	Biased or unrepresentative training data → discriminatory outputs, regulatory breach, forced rollback	**Compliance Risk**	*Gate 3 failure (fairness audit)*
D4 — Business Value	No KPI baseline or stress-tested ROI → value cannot be demonstrated at Gate 3 → pilot abandoned despite technical success	**Business Risk**	*Gate 3 failure (no KPI validation)*
D5 — Regulatory Risk	Regulatory non-compliance → post-deployment enforcement action, reputational damage, forced shutdown	**Compliance Risk**	*Gate 3/4 failure (compliance sign-off)*
D6 — Human Oversight	No error recovery path → AI errors propagate unchecked in production → safety or quality incident	**Safety Risk**	*Gate 5 failure (governance)*

Dimension	Raw /9	Weight	Weighted	RAG
D0: Problem Statement Quality	____ / 9	**15%**	____	🔴 🟡 🟢
D1: Automation Potential	____ / 9	**18%**	____	🔴 🟡 🟢
D2: Data Availability	____ / 9	**13%**	____	🔴 🟡 🟢
D3: Data Risk & Fitness	____ / 9	**13%**	____	🔴 🟡 🟢
D4: Business Value & ROI	____ / 9	**18%**	____	🔴 🟡 🟢
D5: Regulatory & Ethical Risk	____ / 9	**13%**	____	🔴 🟡 🟢
D6: Human Oversight	____ / 9	**10%**	____	🔴 🟡 🟢
TOTAL			**____ / 100**	

Why These Weights

Weight Justification: D0 (Problem Statement, 15%) is scored first because a poor problem definition makes all other scores meaningless — building the wrong thing with perfect data is still wrong. D1 and D4 carry the next highest weight (18% each) because automation suitability determines technical feasibility and business value determines whether delivery is worth the investment. Data dimensions (D2+D3) collectively represent 26% because data failure is the single most common cause of AI project failure. Regulatory and oversight (D5+D6) represent 23% to ensure governance is weighted, not treated as a secondary concern.

Industry Benchmark Ranges — Phase 1 Qualify Score

These benchmarks are derived from SpanForge's AI portfolio diagnostic work across enterprise organisations. Use them to contextualise your score — not to lower the bar.

Score Range	Typical Profile	SpanForge Interpretation	Production Conversion Rate
80 – 100	Strong data foundation, clear KPIs, low regulatory risk, active sponsor	Rare in first assessment — indicates prior AI delivery experience. Proceed with confidence.	**> 50% (SpanForge target range)**
60 – 79	Good automation fit and value hypothesis; some data or governance gaps	Typical of experienced enterprise teams. Advance to Phase 2; address gaps in Design phase.	**30 – 50%**
40 – 59	Patchy readiness — one or two strong dimensions offset by material gaps	Most common first-assessment range. Conditional advance; gaps must be remediated before Gate 1.	**15 – 30%**
25 – 39	Multiple weak dimensions; problem or data fundamentals not established	Return. Invest in problem definition and data strategy before re-assessing.	**< 15%**
0 – 24	Systemic unreadiness across most dimensions	Recommend termination or fundamental redesign. Do not consume further resource on current framing.	**< 5%**

Industry average (S&P Global, 2025): enterprises scoring below 45 on first assessment have a > 80% probability of pilot failure or abandonment before Gate 3. The SpanForge target: initiatives scoring ≥70 at first assessment convert to Full Production within 18 months at 2–3× the industry rate.

Score	Decision	Action & SpanForge Next Step	Gate
70 – 100	**Advance**	Strong candidate. Proceed to Phase 2. Produce Problem Statement Canvas. Confirm Gate Authority.	**Gate 1**
45 – 69	**Conditional Advance**	Proceed to Phase 2 with conditions. Document remediation requirements, owner, and 30-day review date.	**Gate 1***
0 – 44	**Return**	Fundamental blockers present. Do not proceed. Revisit problem framing, data strategy, or value hypothesis. Re-assess with Gate Authority when remediated.	**Return**

Recommend Termination — Kill Fast Conditions

SpanForge Doctrine: "Failure must be documented, not apologised for." A formally closed initiative with captured learnings is more valuable than a pilot that quietly disappears. Recommend immediate termination — not Return — if ANY of the following are true:

☐ **D0 score = 0 AND no written problem statement can be produced within 2 weeks**	The initiative has no foundation. Continuing wastes resource on an undefined target.
☐ **D4 (Business Value) score = 0 AND sponsor cannot articulate a value hypothesis in the session**	If the person funding the initiative cannot state what value it creates, it should not exist.
☐ **Q11 (Bias risk) = 0 AND compliance team confirms no pathway to clearance**	Regulatory exposure with no remediation path. Risk to organisation exceeds any potential value.
☐ **Q32 (Gate Authority) = 0 AND no sponsor can be identified within 30 days**	Without a budget-accountable Gate Authority, the system cannot function as designed. Do not proceed.
☐ **Phase 1 score < 25 across all dimensions (not just one or two)**	Systemic unreadiness. Partial remediation will not produce a viable initiative.

If termination is recommended — mandatory documentation:

☐ Formal closure decision recorded with Gate Authority signature ☐ Learnings captured ☐ Rationale filed in governance record ☐ Team informed

Phase 1 Gate Decision

☐ Advance to Phase 2 ☐ Conditional Advance — remediation required ☐ Return — do not proceed

Named Gate Authority: _____

Decision rationale (evidence-based):

_____

_____

Top 3 risks identified:

1. _____
2. _____
3. _____

Committed next steps (owner + date):

_____

_____

PHASE 2 · DISCOVER

Are we ready to build?

Only complete if Phase 1 score ≥ 45 · Gate threshold: Score ≥ 45 to advance · Output feeds Gate 1 (Scoping) · Estimated time: 40 minutes

D7 Data Maturity

SpanForge Data Layer — Discover & Design phases

Time: 8 min | Weight: 20%

Assesses the organisation's overall data infrastructure, culture, and practices — beyond the specific dataset for this project. SpanForge's Data Layer framework is embedded across Discover and Design phases; this dimension confirms readiness to engage it.

Low Score Means:

The organisation lacks the data infrastructure to support ongoing AI delivery. Even if this project succeeds, data problems will recur on every future initiative.

Typical Failure Pattern:

Organisation invests in a point solution on a weak data foundation. The AI system degrades as data quality deteriorates post-launch. No monitoring catches it. Model fails silently.

Q23 How mature is the organisation's overall data infrastructure?

Score	Description	Evidence Required to Justify Score ≥ 2
0	☐ Ad-hoc — data stored in spreadsheets, email, or disconnected systems	—
1	☐ Developing — some databases and reporting tools but inconsistently used	—
2	☐ Managed — data warehouse or lake in place with defined ownership	*Name the data platform, the owner, and describe coverage scope*
3	☐ Optimised — real-time pipelines, data catalogue, and self-serve analytics are standard	*Cite the data catalogue and pipeline SLAs — attach governance documentation*

Evidence / Notes (mandatory for score ≥ 2):

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Q24 Is there an active data quality management practice?

Score	Description	Evidence Required to Justify Score ≥ 2
0	☐ No — data quality is not actively managed or measured	—
1	☐ Basic — quality issues are known but addressed reactively	—
2	☐ Defined — data quality standards exist and are partially enforced	*Cite the quality standards documentation — describe how they are enforced*
3	☐ Advanced — automated quality checks, SLAs, and data stewardship roles are active	*Name the monitoring system and cite SLA documentation and stewardship roles*

📋 **Evidence / Notes (mandatory for score ≥ 2):**

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Q25 Is training and evaluation data versioned, locked, and reproducible?

Score	Description	Evidence Required to Justify Score ≥ 2
0	☐ No — no versioning exists; data changes without a record	—
1	☐ Informal — some versioning by convention (e.g. file naming) but no system control	—
2	☐ Partial — version control planned or partially implemented for key datasets	*Describe the versioning approach and list the datasets in scope*
3	☐ Formal — data versioning system in place; holdout sets are locked and reproducible for audit	*Cite the versioning system — confirm the holdout set is locked and reproducible*

📋 **Evidence / Notes (mandatory for score ≥ 2):**

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Dimension 7 Score

Failure mode triggered if score < 3 on any blocker question: **document & escalate**

D7 Raw Score

____ / 9

Weighted contribution: ____

D8	Skills & Team Capacity	⏰ Time	Weight
		8 min	15%

Determines whether the organisation has, or can access, the human capability to build, deploy, and maintain an AI solution. Distinguishes between skills (what exists) and capacity (what is available for this initiative).

Low Score Means:

The team lacks the skills to build or the capacity to deliver alongside BAU commitments. The initiative will be under-resourced from day one.

Typical Failure Pattern:

Initiative begins but technical debt accumulates as under-skilled team members cut corners. Delivery slips. Senior stakeholders lose confidence. Scope reduced to something that can be shipped but not maintained.

Q26 Does the team have in-house AI/ML engineering capability?

Score	Description	Evidence Required to Justify Score ≥ 2
0	☐ No — no relevant technical skills exist internally	—
1	☐ Minimal — one or two individuals with foundational skills; no deployment track record	—
2	☐ Developing — a small team with some prior AI project experience	*Name team members with relevant experience and describe prior projects*
3	☐ Established — a dedicated AI/ML team with successful production deployments	*Reference deployed systems and team structure*

📋 **Evidence / Notes (mandatory for score ≥ 2):**

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Q27 Is domain expertise available to guide model development and validation?

Score	Description	Evidence Required to Justify Score ≥ 2
0	☐ No — domain experts are unavailable or unwilling to engage	—
1	☐ Limited — domain experts can provide only occasional input	—
2	☐ Moderate — domain experts available and willing to contribute regularly	*Name the domain experts and confirm their commitment and availability*
3	☐ Strong — domain experts are embedded in the team with formal time allocation	*Cite the resourcing agreement — state the allocated time percentage*

📋 **Evidence / Notes (mandatory for score ≥ 2):**

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Q28 Does sufficient dedicated capacity exist to deliver alongside BAU?

Score	Description	Evidence Required to Justify Score ≥ 2
0	☐ No — teams are at full capacity with no bandwidth for new initiatives	—
1	☐ Limited — some capacity possible with significant trade-offs	—
2	☐ Moderate — capacity exists with agreed prioritisation adjustments	*State the capacity commitment and what BAU work has been deprioritised*
3	☐ Strong — dedicated capacity is ring-fenced and protected from BAU pressure	*Cite the resourcing decision and name the approving authority*

📋 **Evidence / Notes (mandatory for score ≥ 2):**

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Dimension 8 Score

Failure mode triggered if score < 3 on any blocker question: **document & escalate**

D8 Raw Score

____ / 9

Weighted contribution: ____

D9	Governance Readiness *T.R.U.S.T.™ Framework: Traceability · Responsibility · Safety Guardrails*	Time 8 min	Weight 20%

✦ T.R.U.S.T.™ Framework: Traceability · Responsibility · Safety Guardrails

T.R.U.S.T.™ Alignment — This dimension maps to three of five T.R.U.S.T.™ dimensions: Traceability (audit trail and logging), Responsibility (named human accountability for AI outcomes), and Safety Guardrails (technical controls, not just policy). Low scores here indicate that governance must be built before any production deployment.

Evaluates whether the organisation has the policies, processes, and accountability structures to deploy AI responsibly. Low scores here are addressable in the Govern phase, but material gaps must be flagged at Gate 1.

Low Score Means:

The organisation cannot govern what it deploys. AI systems run without accountability, audit trails, or intervention mechanisms. Regulatory or board scrutiny will expose these gaps at the worst possible time.

Typical Failure Pattern:

AI system deployed without formal accountability. Incident occurs. Responsible party cannot be identified. No audit trail exists for regulators. Reputational and regulatory consequences follow.

Q29 Does an AI governance framework or policy exist? (T.R.U.S.T.™: Responsibility, Safety Guardrails)

Score	Description	Evidence Required to Justify Score ≥ 2
0	☐ None — no AI-specific governance exists	—
1	☐ In development — a framework is being drafted but not yet adopted	—
2	☐ Partial — a policy exists but coverage is incomplete or inconsistently applied	*Cite the policy document and list the primary coverage gaps*
3	☐ Established — a comprehensive AI governance framework is in active use across the organisation	*Cite the framework document, name the owner, and describe how consistent application is verified*

📋 **Evidence / Notes (mandatory for score ≥ 2):**

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Q30 Are processes in place for model monitoring and drift detection post-deployment? (T.R.U.S.T.™: Traceability)

Score	Description	Evidence Required to Justify Score ≥ 2
0	☐ No — no mechanism exists to track model performance post-deployment	—
1	☐ Ad-hoc — monitoring happens informally or only when issues are flagged	—
2	☐ Defined — monitoring processes documented but not yet operationalised	*Cite the monitoring design — confirm AgentOBS™ compatibility is in scope*
3	☐ Active — automated monitoring, drift detection, and alerting operational (or confirmed via AgentOBS™ roadmap)	*Name the monitoring system, state alert thresholds, and confirm the responsible team*

📋 **Evidence / Notes (mandatory for score ≥ 2):**

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Q31 Is there a clear accountability structure for AI decisions and outcomes? (T.R.U.S.T.™: Responsibility, User Rights)

Score	Description	Evidence Required to Justify Score ≥ 2
0	☐ No — accountability is undefined; no individual or team owns AI outcomes	—
1	☐ Informal — responsibility implied but not formally assigned	—
2	☐ Assigned — a named owner exists without formal authority or resources	*Name the owner and describe what authority and resources are missing*
3	☐ Established — named owner with formal authority, resources, reporting lines, and defined user recourse paths	*Cite the accountability structure and describe the user redress pathway*

📋 **Evidence / Notes (mandatory for score ≥ 2):**

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Dimension 9 Score

Failure mode triggered if score < 3 on any blocker question: **document & escalate**

D9 Raw Score

____ / 9

Weighted contribution: ____

D10	Executive Sponsorship & Gate Authority	Time	Weight
		8 min	20%

Non-Negotiable: Gate Authority Required

SpanForge Non-Negotiable: Every initiative must have a named Gate Authority with budget accountability before any gate review is scheduled. A score of 0 on Q32 is an automatic blocker — the assessment cannot produce a binding gate decision without this. A sponsor who cannot stop the initiative is not a Gate Authority.

Determines whether the initiative has sufficient leadership support and organisational priority to survive the challenges of an AI delivery programme. SpanForge requires a budget-accountable sponsor — not a delegate — at every gate.

Low Score Means:

Without a budget-accountable sponsor, gate decisions are advisory. The initiative advances by momentum rather than evidence. When problems arise, no one can make the call to halt or redirect.

Typical Failure Pattern:

Initiative continues past clear failure signals because no executive has the authority or incentive to stop it. Budget drains. Team demoralises. Initiative eventually abandoned without formal closure or captured learnings.

Q32 Is there a named, budget-accountable executive sponsor who will make binding gate decisions?

Score	Description	Evidence Required to Justify Score ≥ 2
0 ⊘	☐ No — no executive with budget authority is committed to this initiative	*AUTOMATIC BLOCKER — Gate review cannot be scheduled. No advance decision can be made.*
1	☐ Passive — an executive is nominally aware but has not committed budget or personal attention	—
2	☐ Engaged — a budget-accountable sponsor attends reviews and unblocks decisions	*Name the sponsor, confirm budget accountability, and confirm attendance at gate reviews*
3	☐ Active Gate Authority — sponsor is publicly committed, allocates personal time, and will make binding Advance/Return decisions	*Cite the Gate Authority appointment — confirm they understand Advance / Conditional Advance / Return*

Evidence / Notes (mandatory for score ≥ 2):

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Q33 Does this initiative align with stated organisational strategy?

Score	Description	Evidence Required to Justify Score ≥ 2
0	☐ Misaligned — this runs counter to current strategic priorities	—
1	☐ Neutral — no clear connection to strategic objectives	—
2	☐ Loosely aligned — can be linked to strategy with narrative effort	*Describe the strategic connection and reference the strategy document*
3	☐ Directly aligned — explicitly referenced in strategic plans, OKRs, or board-level AI agenda	*Reference the specific strategic document and statement*

📋 **Evidence / Notes (mandatory for score ≥ 2):**

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Q34 Is budget provisionally allocated or confirmed for this initiative?

Score	Description	Evidence Required to Justify Score ≥ 2
0	☐ No — no budget discussion has taken place	—
1	☐ Exploratory — budget conversations are happening but nothing is confirmed	—
2	☐ Provisional — budget identified subject to business case approval	*Reference the budget source and approval authority*
3	☐ Confirmed — budget approved, allocated, and controlled by the Gate Authority sponsor	*Reference the budget approval and the sponsor's control of allocation*

📋 **Evidence / Notes (mandatory for score ≥ 2):**

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Dimension 10 Score

⚠ Score of 0 on Q32 = automatic blocker — no gate decision can be made

Failure mode triggered if score < 3 on any blocker question: **document & escalate**

D10 Raw Score

_____ / 9

Weighted contribution: ____

D11	Observability & Infrastructure Readiness *SpanForge AgentOBS™ · T.R.U.S.T.™: Traceability*	Time 8 min	Weight 25%

AgentOBS™ Requirement

SpanForge Non-Negotiable: Every production AI system must have AgentOBS™ or equivalent observability active before the first production request. This dimension assesses infrastructure readiness to support that requirement. Even at the Discover stage, observability architecture must be considered — retrofitting monitoring post-deployment is a governance failure.

Evaluates whether the technical environment can support development, deployment, and ongoing operation of the AI solution. Observability is not optional — it is the technical enforcement of the T.R.U.S.T.™ Traceability dimension.

Low Score Means:

The organisation cannot deploy AI safely because it has no way to monitor, detect drift, or intervene when the system degrades. Governance becomes impossible without observability.

Typical Failure Pattern:

AI system deployed without monitoring. Model degrades silently. User-visible errors accumulate before anyone notices. No audit trail exists for the period of degradation. Regulatory and reputational exposure follows.

Q35 Does the existing technology stack support AI/ML workloads?

Score	Description	Evidence Required to Justify Score ≥ 2
0	☐ No — legacy stack; significant re-platforming required before any AI work can begin	—
1	☐ Limited — basic cloud or compute exists but is not configured for ML workloads	—
2	☐ Partial — cloud infrastructure in place; MLOps tooling is minimal	*Describe existing infrastructure and identify the primary gaps for ML workloads*
3	☐ Ready — modern ML platform with CI/CD, model registry, and serving capability confirmed	*Reference the platform, tooling stack, and responsible team*

Evidence / Notes (mandatory for score ≥ 2):

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Q36 Can the solution integrate with the systems where outputs will be used?

Score	Description	Evidence Required to Justify Score ≥ 2
0	☐ No — target systems are closed or use incompatible protocols	—
1	☐ Difficult — integration theoretically possible but requires significant engineering effort	—
2	☐ Manageable — APIs or integration patterns exist with known complexity	*Reference the integration approach and estimated engineering effort*
3	☐ Straightforward — well-documented APIs and integration patterns already in active use	*Reference the API documentation and any existing integrations*

📋 **Evidence / Notes (mandatory for score ≥ 2):**

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Q37 Is there a plan or capability for production monitoring and drift detection? (AgentOBS™ readiness)

Score	Description	Evidence Required to Justify Score ≥ 2
0	☐ No — no monitoring capability or plan exists; observability not yet discussed	—
1	☐ Partial — logging exists but no model-specific monitoring, drift detection, or alerting in scope	—
2	☐ Developing — monitoring approach being designed; tooling selection underway; AgentOBS™ compatibility under review	*Reference the monitoring design document and tooling shortlist*
3	☐ Ready — monitoring, drift detection, and alerting planned or operational; confirmed compatible with AgentOBS™ integration standard	*Reference the monitoring architecture and AgentOBS™ integration plan*

📋 **Evidence / Notes (mandatory for score ≥ 2):**

Score chosen: ____ Confidence: ☐ High ☐ Medium ☐ Low

Evidence: _____

_____

Source / Owner: _____

Dimension 11 Score

Failure mode triggered if score < 3 on any blocker question: **document & escalate**

D11 Raw Score

____ / 9

Weighted contribution: ____

PHASE 2 GATE

Discover Score Summary & Gate Decision

Gate Authority signature required · Output feeds Gate 1 (Scoping) of SpanForge Exit Gate System™

Failure Mode Mapping — Phase 2 Dimensions

Dimension	Low Score Failure Mode	Risk	SpanForge Gate
D7 — Data Maturity	Weak data infrastructure → model degrades post-deployment → no early warning system → silent failure	**Data Risk**	*Gate 5 failure*
D8 — Skills & Capacity	Under-resourced team → technical debt → delayed delivery → scope reduction → maintenance gaps post-launch	**Delivery Risk**	*Gate 2/3 failure*
D9 — Governance	No accountability structure → AI incidents without a named owner → regulatory exposure → no audit trail	**Compliance Risk**	*Gate 4/5 failure*
D10 — Sponsorship	No Gate Authority → gate decisions advisory → initiative advances on momentum → cannot be stopped when failing	**Delivery Risk**	*System failure*
D11 — Observability	No production monitoring → model degrades silently → user-visible failures → no governance possible at scale	**Safety Risk**	*Gate 5 failure*

Dimension	Raw /9	Weight	Weighted	RAG
D7: Data Maturity	____ / 9	**20%**	____	🔴 🟡 🟢
D8: Skills & Team Capacity	____ / 9	**15%**	____	🔴 🟡 🟢
D9: Governance Readiness (T.R.U.S.T.™)	____ / 9	**20%**	____	🔴 🟡 🟢
D10: Executive Sponsorship & Gate Authority	____ / 9	**20%**	____	🔴 🟡 🟢
D11: Observability & Infrastructure	____ / 9	**25%**	____	🔴 🟡 🟢
TOTAL			**____ / 100**	

⍰ Why These Phase 2 Weights

Weight Justification — Phase 2: Observability carries the highest Phase 2 weight (25%) because SpanForge's non-negotiable — production monitoring active before first production request — cannot be retrofitted. Governance and Sponsorship each carry 20% because without T.R.U.S.T.™-aligned governance and a real Gate Authority, the system cannot be sustained. Data Maturity (20%) because AI delivery depends on an organisational data foundation, not just a dataset.

Industry Benchmark Ranges — Phase 2 Discover Score

Score Range	Typical Profile	SpanForge Interpretation
75 – 100	Mature data infrastructure, strong governance, active Gate Authority with confirmed budget	Proceed with high confidence. Gate 1 Scoping review can be scheduled immediately.
55 – 74	Good sponsorship and skills; some governance or observability gaps	Typical of digitally mature organisations. Conditional advance; address gaps in Gate 1 risk register.
35 – 54	Weak data foundation or governance, or passive sponsor	Most common range. Return — remediate sponsorship, data maturity, or governance before Gate 1.
0 – 34	Multiple Phase 2 blockers; organisational readiness is not present	Do not proceed. Fundamental readiness investment required before any AI initiative of this scale.

Score	Decision	Action & SpanForge Next Step	Gate
70 – 100	Advance to Gate 1	Strong readiness profile. Proceed to Gate 1 Scoping review. Produce: signed Problem Statement Canvas, Gate Authority letter, Gate Readiness Score™ (target ≥70), confirmed business case.	Gate 1
45 – 69	Conditional Advance	Proceed to Gate 1 with conditions. Document all gaps, assign remediation owners and dates. Include gaps in Gate 1 risk register. Gate Authority must acknowledge conditions in writing.	Gate 1*
0 – 44	Return	Fundamental readiness blockers present. Do not schedule Gate 1. Revisit data strategy, governance structure, or sponsorship. Re-assess with Gate Authority when fundamentals are remediated.	Return

Phase 2 Gate Decision

☐ Advance to Gate 1 Scoping ☐ Conditional Advance — remediation required ☐ Return — revisit fundamentals

Named Gate Authority: _____

Decision rationale (evidence-based):

_____

_____

Top 3 risks identified:

1. _____
2. _____
3. _____

Committed next steps (owner + date):

_____

_____

EXECUTIVE SUMMARY

Assessment Output — Complete After Both Phases

Complete this page last · Present to Gate Authority · File as governance artefact

Purpose of This Page

This page is the governance artefact. It summarises the assessment for the Gate Authority, board, or audit record. It must be completed by the Assessment Lead and reviewed by the Gate Authority before any gate decision is recorded as binding.

Assessment Summary	
Initiative:	_____
Phase 1 Score:	____ / 100 RAG: 🔴 🟡 🟢 Gate Decision: ☐ Advance ☐ Conditional ☐ Return
Phase 2 Score:	____ / 100 RAG: 🔴 🟡 🟢 Gate Decision: ☐ Advance ☐ Conditional ☐ Return
Overall Recommendation:	☐ Advance to Gate 1 Scoping ☐ Conditional Advance ☐ Return
Gate Authority Name:	_____
Assessment Date:	_____

✓ Top 3 Strengths

1. _____
2. _____
3. _____

⚠ Top 3 Risks

1. _____
2. _____
3. _____

Automatic Blocker Status — Must be cleared before any Advance decision		
Q1 — Problem Statement (D0): No written problem statement exists	☐ Clear ☐ Blocker triggered	Remediation owner: _____
Q11 — Bias Risk (D3): High bias / discriminatory outcome risk	☐ Clear ☐ Blocker triggered	Remediation owner: _____
Q17 — Regulatory Complexity (D5): Highly regulated domain, AI-specific rules	☐ Clear ☐ Blocker triggered	Remediation owner: _____
Q18 — Individual Harm Risk (D5): High risk to individual rights or safety	☐ Clear ☐ Blocker triggered	Remediation owner: _____
Q20 — Error Tolerance (D6): Zero-tolerance process, human-in-loop design mandatory	☐ Clear ☐ Blocker triggered	Remediation owner: _____
Q32 — Gate Authority (D10): No budget-accountable sponsor identified	☐ Clear ☐ Blocker triggered	Remediation owner: _____

◈ GO / NO-GO NARRATIVE

Assessment Lead summary (max 5 sentences — evidence-based, no narrative optimism):

_____

_____

_____

_____

SpanForge Lifecycle — What Happens Next		
This Assessment	Discover Phase pre-qualification → feeds Gate 1	✓ **Complete**
Next Required Artefact	Signed Problem Statement Canvas (including KPI baseline and value hypothesis)	☐ **Pending**
Gate 1 — Scoping	Gate Readiness Score™ ≥70 required · Named Gate Authority constituted · Business case validated	☐ **Scheduled**
Gate 2 — PoC	Baseline accuracy on holdout set · Risk register reviewed · PoC report approved	☐ **Future**
Gate 3 — Pilot	KPI targets achieved · Operational integration confirmed · Compliance sign-off	☐ **Future**
Gate 4 — Limited Release	Adoption rate thresholds met · SLA compliance · Support model defined	☐ **Future**
Gate 5 — Full Production	AgentOBS™ active · Runbook approved · Owner accountabilities documented	☐ **Future**

🎯 Gate Readiness Score™ — What Comes Next

This assessment is the input to Gate 1. It does not replace the Gate Readiness Score™ — it produces the evidence base that the Gate Readiness Score™ draws on. Here is how they connect:

This Assessment	→	Gate Readiness Score™	→	Gate Decision
Phase 1 & 2 scoresEvidence capturedBlockers identifiedStakeholders aligned		**Gate Readiness Score™ — 5 dimensions:** **Technical Readiness** — Model performance, integration, infrastructure **Business Value Evidence** — KPIs defined, measured, validated (from D4) **Operational Integration** — Support model, runbooks, escalation paths **Compliance Status** — Regulatory mapped, data compliant (from D5) **Stakeholder Alignment** — Gate Authority active, sponsor committed (from D10) **Target: ≥70 before Gate 1 review is scheduled**		**Advance** All conditions met. Moves to next stage on confirmed schedule. **Conditional Advance** Conditions substantially met. Named remediation owner + review trigger. **Return** Conditions not met. Designed outcome — not failure. Document rationale, remediation, owner, date.

Contact SpanForge to obtain the Gate Readiness Score™ template and facilitation guide for Gate 1 Scoping: advisory@getspanforge.com | getspanforge.com/platform

Sign-Off — All three required before gate decision is binding

Assessment Lead	Gate Authority / Executive Sponsor	AI / Technical Lead
Name:	**Name:**	**Name:**
Role / Title:	**Role / Title:**	**Role / Title:**
Date:	**Date:**	**Date:**
Signature:	**Signature:**	**Signature:**

SPANFORGE

AI Assessment — Executive Lite Version

2-Page Summary · For executive review and gate discussion · Full assessment on preceding pages

Initiative:		**Gate Authority:**	
Assessment Date:		**Portfolio Priority:**	☐ High ☐ Medium ☐ Low

Score Dashboard

PHASE 1 — QUALIFY

D0	Problem Statement Quality	15%	__ / 9	
D1	Automation Potential	18%	__ / 9	
D2	Data Availability	13%	__ / 9	
D3	Data Risk & Fitness	13%	__ / 9	
D4	Business Value & ROI	18%	__ / 9	
D5	Regulatory & Ethical Risk	13%	__ / 9	
D6	Human Oversight	10%	__ / 9	
P1 TOTAL				__ / 100

PHASE 2 — DISCOVER

D7	Data Maturity	20%	__ / 9	
D8	Skills & Team Capacity	15%	__ / 9	
D9	Governance (T.R.U.S.T.™)	20%	__ / 9	
D10	Exec Sponsorship & Gate Auth	20%	__ / 9	
D11	Observability & Infrastructure	25%	__ / 9	
P2 TOTAL				__ / 100

< 25	25 – 44	45 – 69	70 – 100
Terminate	**Return**	**Conditional**	**Advance**

Automatic Blockers

☐ Clear ☐ Triggered Q1 — No problem statement	☐ Clear ☐ Triggered Q11 — Bias risk (D3)	☐ Clear ☐ Triggered Q17 — Highly regulated domain (D5)	☐ Clear ☐ Triggered Q18 — Individual harm risk (D5)	☐ Clear ☐ Triggered Q20 — Zero-tolerance errors (D6)	☐ Clear ☐ Triggered Q32 — No Gate Authority (D10)

✓ Top 3 Strengths	⚠ Top 3 Risks	Gate Decision
1.	1.	☐ Advance to Phase 2 / Gate 1
2.	2.	☐ Conditional Advance
3.	3.	☐ Return
		☐ Recommend Termination

Go / No-Go Narrative *(3 sentences max — evidence only, no optimism)*

Next Required Artefacts

- ☐ Signed Problem Statement Canvas
- ☐ Gate Authority letter confirmed
- ☐ Gate Readiness Score™ template initiated
- ☐ Gate 1 Scoping review scheduled

Sign-Off

Assessment Lead: _____ Date: _____

Gate Authority: _____ Date: _____

Technical Lead: _____ Date: _____

Full assessment on preceding pages · getspanforge.com