

Agentic AI Pioneers Prize - Development Phase assessor guidance

This document contains the guidance and scoring used by the Assessors when reviewing your application. Assessors will review your answers for each scored question and mark each of them between 1 and 10 (1 being the lowest and 10 being the highest).

Any questions that are not scored will not be reviewed by the Assessor.

Question 1. Applicant location (not scored)

You must state the name and full registered address of your organisation and any partners working on your project.

We are collecting this information to understand more about the geographical location of all applicants.

Question 2. Animal testing (not scored)

Will you involve any trials with animals or animal testing?

You must select one option:

- Yes
- No

We will only support innovation projects conducted to the highest standards of animal welfare.

Further information for proposed solutions involving animal testing is available at the [UKRI Good Research Hub](#) and [NC3R's animal welfare guidance](#).

Question 3. Permits and licences (not scored)

Will you have the correct permits and licences in place to carry out your proposed solution?

We are unable to fund projects which do not have the correct permits or licences in place by your project start date.

You must select one option:

- Yes
- No
- In the process of being applied for
- Not applicable

Question 4. International Collaboration (not scored)

Does your proposed work involve any international collaboration or engagement?

You must provide details of any expected international collaboration or engagement. You must include a list of the names and the countries, any international project co-leads, project partners, visiting researchers, or other collaborators are based in. You must also include details of any service providers.

If your proposed work does not involve international collaboration or engagement, your answer must confirm this.

Question 5. Export licence (not scored)

You must indicate whether an export control license is required for this solution under the [academic export control guidance](#).

You must select one option:

- Yes
- No

Question 6. Trusted Research and Innovation (not scored)

You must explain if your proposed solution work relates to [UKRI's Trusted Research and Innovation \(TR&I\) Principles](#), including:

- a list of any dual-use (both military and non-military) applications to your research
- a list of the areas where your project is relevant to one or more of the [17 areas](#) of the UK National Security and Investment (NSI) Act
- whether an export control license is required for this project under the [academic export control guidance](#) and the status of any applications
- a list of any items or substances on the UK Strategic Export Control List

If your proposed work does not relate to UKRI's TR&I Principles, your answer must confirm this.

We may ask you to provide additional TR&I information at a later date, in line with UKRI TR&I Principles and funding terms and conditions.

Question 7. Challenge statement (not scored)

Which challenge statement does your solution align to?

- Advanced Manufacturing Challenge: Design Compliance Agents
- Advanced Manufacturing Challenge: Detailed Design for X Agents
- Advanced Manufacturing Challenge: System of System Concept Design Agents
- Health and Life Science Challenge: Digital Twin Design and Implementation Agents
- Health and Life Science Challenge: Multimodal Biomedical Insight and Decision Support Agents
- Health and Life Science Challenge: Adaptive Operational Intelligence and Readiness Forecasting Agents
- Creative Industries Challenge: Intelligent Creative Workflow and Asset Orchestration Agents
- Creative Industries Challenge: Real-Time Adaptive Media and Experience Design Agents
- Creative Industries Challenge: Rights, Attribution and Provenance Intelligence Agents

Question 8. Expression of Interest (EOI) application number (not scored)

What was your EOI application number?

Note, you must apply to this competition with the same lead and consortium that was named in your EOI.

Question 9. Consortium (not scored)

If Collaborative, who is in your consortium and what are each organisation's roles?

Explain the roles, skills and expertise of all organisations in the consortium that are relevant to the solution you are building.

If you are applying as a single organisation, you must answer 'Not applicable' in the text box.

You must submit a draft collaboration agreement as an appendix if your project includes partner organisations. The appendix must be a PDF no larger than 10MB, up to two A4 pages, and fully legible at 100% zoom.

If your project is collaborative and you do not provide a draft collaboration agreement, your application will be ineligible and will not be sent for assessment.

For collaborative projects, a final signed version of this agreement, signed by all partners, is required to receive a prize in this competition.

Question 10. High level technical approach**How does the system work and why is it credible?**

Explain:

- the final system architecture, agent roles, key components and data flows

- what is novel compared with the state of the art and what was hard in practice
- the key performance metrics, datasets and test conditions used to validate the system, including how these relate to the challenge objectives
- the tools, compute and environments used during development and validation

You must submit an appendix to support your answer. The appendix must be a PDF no larger than 10MB, up to two A4 pages, and fully legible at 100% zoom.

The appendix can include diagrams, tables or other visuals that help to explain your technical approach and system performance.

Question 10. High level technical approach – Assessor guidance and scoring

Scores 9 - 10

- The technical approach is clear, coherent, and well justified.
- The architecture, components and data flows are well explained and easy to understand.
- The approach is clearly and convincingly aligned with the selected challenge statement and its objectives.
- Where a multi-agent or system-to-system solution is claimed, agent or system roles and interactions are clearly described and evidenced.
- Novelty is credible and well demonstrated.
- Performance metrics are appropriate, validated and supported with convincing evidence, including how they show progress against the challenge objectives.
- The appendix significantly enhances clarity.

Scores 7 - 8

- The technical approach is clearly described but may lack depth in some areas.

- Architecture and data flows are understandable though less detailed.
- Alignment to the selected challenge statement and objectives is generally sound but not strongly evidenced.
- Where a multi-agent or system-to-system solution is claimed, roles and interactions are described but not fully substantiated.
- Novelty is present but less well substantiated.
- Performance metrics are provided with reasonable but not comprehensive evidence or link to objectives.

Scores 5 - 6

- The approach is described but lacks clarity, structure or detail.
- Alignment with the challenge statement and objectives is weak, generic or implied rather than clearly demonstrated.
- Claims of multi-agent or system-to-system behaviour are vague or lightly evidenced.
- Novelty claims are vague or weakly evidenced.
- Performance evidence is limited, illustrative or inconsistently presented, with limited connection to the stated objectives.

Scores 3 - 4

- Technical description is unclear or incomplete.
- Alignment with the selected challenge statement is poor or inconsistent.
- Minimal or no meaningful novelty is demonstrated.
- Little or no credible evidence that the technical approach can achieve the stated objectives.
- Performance evidence is weak, inappropriate or absent.
- Appendix adds little value or is poorly aligned.

Scores 1 - 2

- Technical approach is fundamentally unclear or incoherent.

- No credible evidence that the solution can meet the challenge objectives.
- Any claimed multi-agent or system-to-system aspects are absent or not substantiated.
- No credible performance evidence.
- Appendix missing or unusable.

Question 11. User and workflow fit

Who will use it, where does it fit in the workflow, and what does it improve?

Explain:

- the real user context, roles and tasks you are designing for, and any human-in-the-loop steps
- potential implications for time, quality, cost, safety or reliability in that workflow
- interoperability with existing tools, systems, data formats and governance constraints

You can submit an appendix to support your answer. If provided, the appendix must be a PDF no larger than 10MB, up to two A4 pages, and fully legible at 100% zoom.

The appendix can include visuals that clarify the user journey, workflow, or how your solution fits into existing processes.

Question 11. User and workflow fit – Assessor guidance and scoring

Scores 9 - 10

- The user context and workflow are described with strong domain understanding.
- The applicant shows clear insight into real tasks, constraints, dependencies and typical organisational practices.
- The proposed workflow integration is realistic, coherent and clearly linked to the selected challenge.
- Human-in-the-loop roles are identified and well justified where relevant.

- Interoperability with existing tools, systems and governance is addressed in a credible way.
- Any assumptions about adoption or operation are explicit and plausible.
- Appendix enhances clarity, if submitted.

Scores 7 - 8

- User context and workflow are well described, with good domain understanding in most areas.
- Workflow integration is plausible but some steps or constraints are not fully explored.
- Human-in-the-loop roles are mentioned but not fully developed.
- Interoperability is recognised with some relevant detail, though not comprehensive.
- Assumptions about adoption are reasonable but not deeply examined.

Scores 5 - 6

- User context is described but in a general or high-level way.
- Workflow integration is partially explained or relies on broad assumptions.
- Limited discussion of realistic constraints, dependencies or operational challenges.
- Human-in-the-loop roles are only lightly touched on or generic.
- Interoperability is mentioned but lacks useful detail.

Scores 3 - 4

- Limited or shallow understanding of users, roles or workflows.
- Proposed integration into real-world processes is unclear or unrealistic.
- Human-in-the-loop needs are ignored or poorly justified.
- Interoperability constraints are not recognised or are described inaccurately.

Scores 1 - 2

- No meaningful understanding of users or workflows is demonstrated.
- No credible explanation of how the solution would embed into real processes.
- Appendix adds no value or is irrelevant, if submitted.

Question 12. Minimal Viable Product (MVP) and integration readiness

Describe what aspects of your solution are ready for demonstration and how they can be deployed.

Explain:

- the MVP scope delivered, what can be demonstrated, and how this demonstrable capability aligns with your selected challenge statement and its objectives
- packaging, Application Programming Interfaces (API), interfaces and deployment model validated during the phase
- verification and validation results, acceptance thresholds met and known limitations
- what remains to reach production in a typical UK setting, and the plan to close gaps

You must submit an appendix to support your answer. The appendix must be a PDF no larger than 10MB, up to two A4 pages, and fully legible at 100% zoom.

The appendix can include diagrams, summaries or tables that help to show what your MVP does and how it can be deployed or integrated.

Question 12. Minimal Viable Product (MVP) and integration readiness - Assessor guidance and scoring

Scores 9 - 10

- The baseline capability and the progression to the current MVP are clearly and convincingly described.
- The MVP is well-defined, coherent, and strongly aligned with the selected challenge statement and its objectives.
- What can be demonstrated is clearly described and directly evidences progress against the challenge.
- Evidence of testing, demonstration and deployment is strong, relevant and robust.
- Packaging, interfaces and APIs are clearly described and technically credible.
- Verification and validation activities are substantial, appropriate and well evidenced, with clear acceptance thresholds and limitations.
- Remaining gaps to production are clearly identified with realistic and credible plans to close them in a UK context.
- Environmental and societal considerations relevant to deployment are recognised and addressed appropriately.
- The appendix is clear, well structured and materially enhances understanding.

Scores 7 - 8

- The baseline and progression to the MVP are described with generally good clarity.
- The MVP is clear and broadly aligned with the selected challenge statement, though links to objectives may be less explicit.
- Demonstrable capability is described and shows plausible progress, though evidence is less comprehensive.
- Testing, demonstration and deployment considerations are present but not fully developed.
- Verification and validation activities are reasonable but incomplete or less detailed.
- Remaining gaps are identified with broadly realistic plans.
- Environmental or societal considerations are mentioned but at a high level.

Scores 5 - 6

- The baseline and progression are mentioned but in a general or superficial way.
- The MVP is described but its alignment to the challenge statement or objectives is only loosely made or implied.
- Demonstrable capability is unclear or weakly evidenced.
- Evidence of deployment or integration is limited or high-level.
- Verification and validation activities are modest, illustrative or poorly connected to claims.
- Remaining gaps are noted but plans to address them are vague or unconvincing.
- Environmental or societal considerations are only briefly referenced or not clearly linked to deployment.

Scores 3 - 4

- The MVP is unclear, inconsistently described or not credible.
- Alignment with the selected challenge statement or its objectives is weak or not clearly established.
- Minimal useful evidence of testing, demonstration or deployment.
- Verification and validation are weak, absent or inappropriate.
- Remaining gaps are largely unrecognised or unrealistic.
- Environmental or societal considerations are not addressed where they would clearly be relevant.
- The appendix is limited, confusing or adds little value.

Scores 1 - 2

- No meaningful MVP is described, or description is incoherent.
- No credible link between what can be demonstrated and the selected challenge statement.
- No credible evidence of testing, demonstration or deployment.

- Required appendix is missing or unusable.

Question 13. Risks, assurance and explainability

What could go wrong and how will you control it responsibly?

Explain:

- the principal technical, data, safety, security and delivery risks encountered, and mitigations applied
- assurance activities completed, such as tests, red-teaming, evaluation protocols and audit trails
- how the system explains its outputs or actions to users, and any evidence that this is usable
- compliance status against any relevant regulations or standards, and remaining actions

You can submit an appendix to support your answer. If provided, the appendix must be a PDF no larger than 10MB, up to two A4 pages, and fully legible at 100% zoom.

The appendix can include materials that summarise key risks, assurance activities or how you present explainability to users.

Question 13. Risks, assurance and explainability - Assessor guidance and scoring

Scores 9 - 10

- Key risks are clearly identified, relevant and thoughtfully analysed.
- Mitigations are well justified, proportionate and clearly linked to the identified risks.
- Assurance activities (for example, testing, evaluation protocols, audit trails, red-teaming) are substantial, appropriate and credible.

- Data privacy, security, access control and logging are addressed in a clear and well-reasoned way, with realistic safeguards where sensitive or personal data may be involved.
- Potential misuse, unintended behaviours or harmful outputs are recognised with credible mitigation strategies, including any relevant bias or fairness considerations.
- Explainability approach is suitable for intended users and supported by some evidence of practicality or usability.
- Compliance with relevant regulations or standards is clearly described, with remaining actions well understood.

Scores 7 - 8

- Main risks are identified with generally reasonable mitigations.
- Assurance activities are present and broadly appropriate, though not fully developed or evidenced.
- Data privacy and security are recognised with some relevant safeguards, though not comprehensive.
- Misuse or harmful output risks are mentioned with some mitigation, but not deeply explored.
- Explainability approach is appropriate but only lightly evidenced or discussed.
- Regulatory or standards alignment is outlined but may lack detail.

Scores 5 - 6

- Risks are acknowledged but analysis is limited or generic.
- Mitigations are high-level or lack clear linkage to specific risks.
- Assurance activities are modest, incomplete or described at a high level.
- Data privacy and security are mentioned briefly or in generic terms, with limited detail on implementation.
- Misuse, harmful behaviours, bias or fairness are only lightly referenced, if at all.
- Explainability is described but without convincing detail on suitability for users.

- Regulatory or standards considerations are mentioned but unclear.

Scores 3 - 4

- Risks are weakly understood, incomplete or poorly prioritised.
- Mitigations are vague, unrealistic or largely absent.
- Little meaningful assurance activity is described.
- Data privacy and security are largely overlooked or described inaccurately.
- Misuse, harmful output or fairness risks are not recognised where they would clearly be relevant.
- Explainability approach is inappropriate, confusing or not aligned with users.
- Regulatory or standards aspects are not addressed or clearly misunderstood.

Scores 1 - 2

- Major risks are unaddressed or not identified.
- No credible mitigations or assurance activities are presented.
- No meaningful consideration of data privacy, security or misuse.
- No explainability approach is described.
- No recognition of relevant regulations or standards.

Question 14. Potential commercial impact and UK benefit

What impact might this project have outside the project team?

Explain:

- target customers, value proposition and evidence from this phase that supports ROI

- route to adoption, including partnerships, pricing, delivery and support models you validated or de-risked
- expected UK benefits such as productivity, cost reduction, quality, exports or resilience
- competitive positioning and how the work strengthens UK capabilities and IP

Question 14. Potential commercial impact and UK benefit - Assessor guidance and scoring

Scores 9 - 10

- Commercial proposition is clear, strong and well supported.
- Evidence of demand or ROI is compelling.
- Route to adoption is credible and informed by real partnerships or testing.
- UK benefits are clearly articulated, specific and meaningful.

Scores 7 - 8

- Commercial case is strong but some areas may lack detail.
- Evidence of demand reasonable though not comprehensive.
- UK benefits described but less fully developed.

Scores 5 - 6

- Commercial case present but generic or thinly evidenced.
- Adoption route unclear or high-level.
- UK benefits limited or vague.

Scores 3 - 4

- Weak or unrealistic commercial case.
- Little supporting evidence.
- UK benefits unclear or unconvincing.

Scores 1 - 2

- No credible commercial potential.
- No meaningful benefit to the UK.

Question 15. Future potential and scalability

How far do you think you can develop your solution?

Explain:

- the technical roadmap informed by the solution development phase
- generalisability to adjacent use cases or sectors, and limits to transfer
- scaling considerations for data, safety, compute, deployment and operations in UK settings
- partnership, IP and commercial milestones you will pursue post-competition

Question 15. Future potential and scalability - Assessor guidance and scoring

Scores 9 - 10

- Roadmap is clear, evidence-informed and realistic.
- Strong potential for wider application or generalisation.
- Scaling considerations are well understood and addressed.
- Next steps, partnerships or milestones are credible and aligned to opportunity.

Scores 7 - 8

- Roadmap reasonable with some evidence.
- Some discussion of generalisation or scale though less detailed.

- Next steps clear but modest.

Scores 5 - 6

- Roadmap present but speculative or high-level.
- Limited articulation of scalability or constraints.
- Next steps unclear or generic.

Scores 3 - 4

- Roadmap lacks credibility or detail.
- Scalability unexplored or unrealistic.
- No meaningful next steps.

Scores 1 - 2

- No viable future development described.
- No consideration of scale.