

Call for Evidence: AI testing, evaluation and assurance in the UK

Part 1: About you

1. Are you responding on behalf of an organisation or in a personal capacity?

- In a personal capacity (move to question 5)
- On behalf of an organisation (move to question 2)

2. Which of the following describe your organisation? *Select all that apply.*

- Organisation that builds or develops AI systems
- Organisation that deploys, adopts or procures AI systems
- Provider of third-party AI assurance products or services
- Organisation with an in-house AI assurance capability
- Buyer or procurer of AI assurance services
- Advisory, consultancy or professional services (e.g. AI governance, legal or risk management)
- Academic or research organisation
- Trade association or professional body
- Standards development organisation
- Public sector organisation (non-regulatory)
- Regulatory body
- Charity, think tank, or other non-profit
- Other

3. What is the size of your organisation?

- 1-9
- 10-49
- 50-249
- 250+
- Don't know

4. What is the name of your organisation?

5. Which sector(s) do you work in or most represent? *Select all that apply.*

- Healthcare and life sciences
- Manufacturing
- Software and IT services
- Financial services

- Defence and security
- Critical national infrastructure/energy
- Public sector
- Telecoms
- Transport and automotive
- Legal and professional services
- Consumer and retail
- Cross-sector or sector-agnostic
- Other

6. Which perspective best represents your primary interest in this call for evidence? *You answer will be used to tailor questions you see in Part 2.*

- Primarily a supplier of assurance – developing and/or providing AI assurance products or services
- Primarily on the demand side – needing, buying, or relying on AI testing and evaluation for assurance
- Primarily an intermediary – for example a researcher, regulator, standards body, investor, funder, trade body or other ecosystem organisation

Part 2: Challenges, unmet needs and opportunities

This section gathers evidence about the underlying challenges and unmet needs in AI testing and evaluation. Part 2A is for all respondents and asks about challenges across the wider UK ecosystem. Part 2B is optional and asks about needs, opportunities or interests specific to you or your organisation.

2A: Ecosystem-wide challenges and capability gaps

- 1. Drawing on your experience to date, where do you see the most significant unmet needs or capability gaps in AI testing and evaluation for assurance?**
Please identify, where you can, the types of AI, sectors, use cases, risk domains or evaluation capabilities affected.
- 2. What is preventing the market or existing ecosystem from closing these gaps effectively?**
- 3. Where could public intervention add the greatest value, without duplicating or displacing activity the market is already delivering?**

2B: Your own organisation's needs or interests - AI testing and evaluation for assurance

This section is optional and asks about your own organisation's needs or interests. Questions are tailored to the primary perspective you selected in Part 1.

Supply side

For organisations developing or providing AI testing, evaluation or assurance products and services.

- 1. What specific AI testing and evaluation capability would you most like to develop, validate or scale with public support over the next 1–2 years?**

Please describe the capability, the AI systems or use cases it would apply to, and the unmet market or user need it would address.

2. What currently prevents you from developing, validating or scaling this capability?

Please identify the most significant barriers, such as access to data, models, expertise, real-world testing environments, standards, validation or funding.

3. What form of public support or collaboration would make the greatest difference, and what practical outcome would it enable?

Demand side

For respondents developing, deploying, procuring, buying or relying on AI testing, evaluation or assurance.

1. Describe an AI use case that you would like to pilot, procure, deploy or scale, but cannot currently proceed with confidently because of unmet AI testing, evaluation or assurance needs.

Please describe the use case, the type of AI system involved and its intended context of use.

2. What specifically are you unable to test, evaluate or verify with confidence?

3. What type of support would most help you address this gap and progress the use case?

Intermediary

For research, standards, regulatory or other ecosystem organisations.

1. What expertise, evidence, or capability could you or your organisation potentially contribute to addressing the testing and evaluation challenges identified above?

For example, this could include research expertise, datasets, standards expertise or programme delivery.

Part 3: Intervention design

We are exploring a range of ways a publicly funded programme could support the development and piloting of scientifically rigorous, repeatable and comparable AI testing and evaluation techniques underpinning AI assurance. The types of interventions below are illustrative examples to prompt your thinking, not a fixed or final set. We would like to understand which you think would be most useful, which you might take part in, and what conditions would need to be in place. We would also like to hear about approaches we haven't listed.

- a) **Collaborative research sprints:** a series of short, intensive events bringing assurance providers, deployers and researchers together to tackle a defined assurance challenge.
- b) **Matched testing pilot:** pairing an assurance provider with a deploying organisation to conduct testing and evaluation of a deployment-ready AI use case, contributing towards codification of best practice.

- c) **Innovation challenge fund:** a competitive grant where teams apply to solve a defined problem developed by a sponsor (e.g. governmental body or business) in testing and evaluation, with winning proposals funded to develop and trial their solution, contributing towards the codification of best practice.
- d) **AI assurance methods accelerator:** giving third-party AI assurance providers dedicated access to a network of technical experts from NPL and strategic partner organisations to workshop methodological challenges, refine and validate emerging assurance approaches and build evidence on their effectiveness, repeatability and limitations.

1. **Which approach do you think would be most valuable overall and why?**
2. **Is there a different approach, not listed above, that you think would more effectively strengthen AI testing and evaluation?**
Please describe it, including what it would produce.
3. **Which, if any, would you or your organisation be interested in participating in and in what capacity?**
For example, as an assurance provider, as a deploying organisation hosting a test, as a research or delivery partner, or as an advisor.
4. **What specific conditions or support would your organisation need in order to participate?**
5. **Are there particular areas you think we should focus on as part of this programme?**
For example, types of AI, AI testing and evaluation techniques, or risk domains.
6. **Is there anything else you would like to share with us on this topic?**

Part 4: Closing questions

NPL does not intend to publish individual responses to this Call for Evidence. We expect to publish an aggregated summary of the insights gathered, which may take the form of a report or a blog post. We will not attribute findings to individual respondents or organisations.

1. **May we include your organisation, or your name if you're responding in a personal capacity, in a published list of contributors to this call for evidence?**
This would acknowledge that you contributed to the evidence base only. We will not attribute any particular views or findings to you as a result of your inclusion in this list.
 - Yes
 - No
2. **May we contact you to clarify or discuss any part of your response?**
 - Yes
 - No
3. **Would you like to be kept informed about the outcomes of this Call for Evidence and relevant next steps or future opportunities to engage with the Centre for AI Measurement?**
 - Yes

- No

4. If you answered 'Yes' to either of the previous two questions, please provide your contact details.

Name:

Email address:

Your contact details will be used only for the purposes you have selected above and will be handled in accordance with the [NPL Privacy Notice](#).