Introduction to the AIQI Consortium

A conformity assessment matrix for the AI System product lifecycle

Adam Leon Smith DEng FBCS

Chair, AIQI Consortium

October 2025

Objectives

The objectives of the Consortium are to:

- Establish a diverse forum of quality infrastructure bodies and other organisations to make connections, foster dialogue and ensure the TICC (Testing, Inspection, Certification, and Compliance Sector) sector plays a role in ensuring AI is deployed and developed in a safe, secure, and ethical manner;
- Identify, monitor, analyse and share information related to current and upcoming AI legislation;
- Advocate the use of the global quality ecosystem to policy developers, regulators and voluntary markets;
- Develop a comprehensive AI assurance framework to support the current and upcoming regulatory requirements and voluntary market solutions;
- Create AI certification, training programmes, and supporting skills development and capacity building.

QI stakeholders

Standards

- Develop essential standards that provide clear frameworks for responsible AI development and deployment.
- For instance, ISO/IEC 42001 on AI Management Systems offers guidelines that enable AI developers to demonstrate compliance with best practices, which can serve regulatory or industry-specific needs.

Metrology

- Provide the scientific foundation for measurements and data integrity, which is crucial in AI.
- By ensuring accurate and reliable measurement standards, metrology institutions support the robustness and reliability of AI technologies, which is essential for fair and consistent outcomes in AI applications.

Academia

- Contributes critical research and thought leadership, supporting advancements in both the QI ecosystem and AI.
- Academic input helps to refine AI standards and practices, ensuring that they are grounded in rigorous scientific understanding and adaptable to emerging innovations.

QI stakeholders

TIC

- Facilitate compliance by offering testing, inspection, and certification (TIC) services aligned with established AI standards.
- They also offer conformity assessment services that will be essential for AI providers in meeting future legislation, such as the EU AI Act, by serving as Notified Bodies.

Accreditation

- Provide independent verification that TIC companies are competent and impartial, ensuring their testing, inspection and certification of AI technologies meet strict quality requirements.
- This is critical in ensuring the validity of the certificates, testing calibration and inspection reports produced into which industry and consumers place their trust.

AIQI

Collectively, the QI ecosystem works to ensure that AI systems are safe, trustworthy, and ethical, with benefits extending to society as a whole. As AI continues to advance, the AIQI Consortium serves as a central forum for key players in AI quality and safety, actively engaging with global policymakers to emphasise the QI sector's essential role in AI policy development and implementation.

AIQI members

Figure: Example members

- 8 x Accreditation Bodies
- 17 x Conformity Assessment Bodies
- 7 x Research/Metrology
- 3 x Trade Representative Body[1]
- 5 x National or International Standards Body
- 3 x Independent members

[1]QI Trade, including the TIC Council — 100+ TIC organisations

History & Achievements

Figure: Policy paper

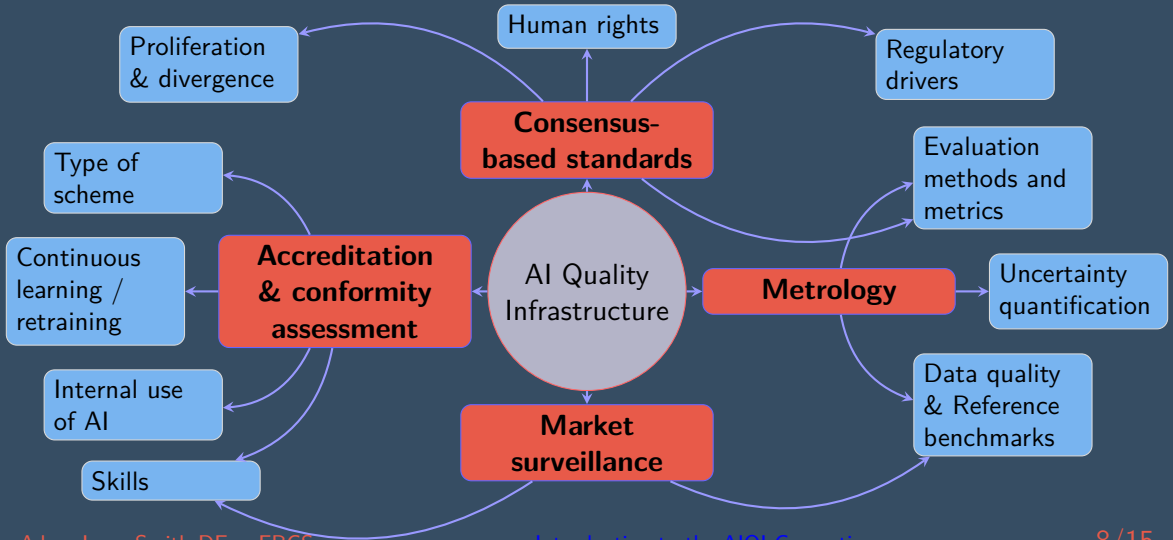
- First meeting September 2024

- Policy paper — Short paper about how QI enables responsible AI — link
- More than 3000 enrollments in our free, self-paced course provides a comprehensive understanding of ISO/IEC 42001 — link
- Presentations at the French AI Action Summit, Global AI Standards Hub and 3 x webinars
- Endorsement by UK minister in the context of AI assurance professionalisation

Upcoming highlights

- Short video explainers about the quality infrastructure
- Free training courses
  - ETSI TS 104 223 (Baseline requirements for cybersecurity of AI)
  - ISO/IEC 42006 (Requirements for bodies providing audit and certification of artificial intelligence management systems)

Key challenges for the AI quality infrastructure



A conformity assessment matrix for the AI system product lifecycle

- Our AI assurance workstream is building a matrix for the assurance of AI systems (not processes)
- Ultimately, we will open source our work
- Our matrix comprises:
  - Assessment dimension (e.g. performance)
  - Life cycle step
  - Task (e.g. classification)
  - Architecture (e.g. CNN)
  - Input type (e.g. image)
  - Domain (e.g. healthcare)
- Each of these cells/permutations is mapped to in progress or published standards

Focus on tasks

- Generic process-level standards provide valuable frameworks for development practices. Still, they cannot effectively measure functional correctness, bias, or robustness because, without a task-specific definition, these measurements become abstract and uninterpretable.
- An AI task is a specific problem to be solved by algorithmic means, e.g.Classification, regression, clustering, object detection, machine translation, automatic summarisation, automated speech recognition, pose estimation and image segmentation.
- Each of these tasks is associated with evaluation requirements and metrics.
- Accuracy, bias and robustness both represent critical dimensions, yet they are mainly measured using derivatives of task-level functional correctness metrics.
- This technical/metrology work is going extremely slowly in standards due to the limited number of practioners available.

Example 1: Named Entity Recognition

Named Entity Recognition is a fundamental natural language processing task that identifies and categorises specific elements in text that represent real-world objects, such as people, organisations, locations, dates, monetary values, and other named entities.
We can discuss it without mentioning any specific technology. It performs two key functions:

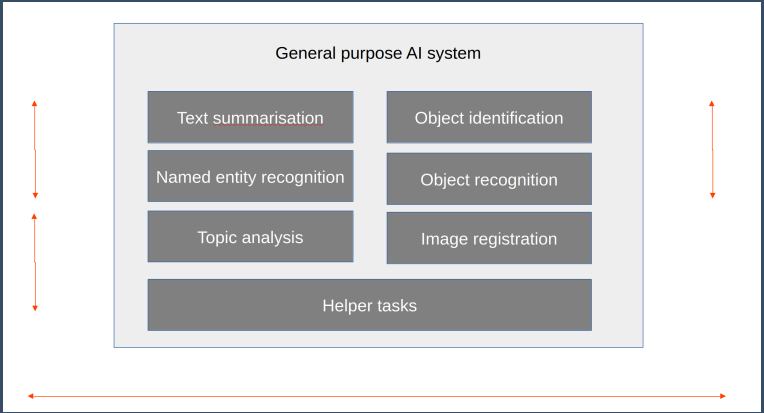
- Identifying the boundaries of named entities within text (detection)
- Assigning appropriate category labels to these entities (typing)

We can assess its functional correctness using a modified, named entity recognition-specific variant of classical metrics, such as precision, recall, and F1 scores, on identified entities. (Look out for future standards on how this differs from other forms of F1 score.)

Example 1: Named Entity Recognition

- **Bias measurement**: Comparing F1 scores between mentions of different demographic groups reveals potential bias.
  - For example, I can determine the F1 score on just the mentions of Western figures within the dataset, the F1 score on non-Western figures only, and the disparity between the two.
- **Robustness measurement**: I can evaluate how F1 scores change when entities are expressed in different ways using different spellings.
  - Perhaps I can determine that the F1 score on texts with standard spelling (for example, in news articles) is 0.90 and the F1 score with certain spelling variations (for example, in social media) is 0.65. This reveals the robustness of the AI system performing the named entity recognition task.

Combining tasks



Implications

- **Standards development**: Testing, bias and robustness standards must build upon task-specific functional correctness standards rather than being developed in isolation.
- **Testing frameworks**: Effective testing requires establishing task-appropriate evaluation metrics and extending these across relevant variations. When organisations use standardised task-specific metrics, comparisons between different AI systems become more meaningful and reliable through benchmarks.
- **Conformity assessment**: Assessments addressing accuracy, bias and robustness (such as in the EU AI Act) need to reference task-specific standards to be reproducible and repeatable.

Contact Information

adamleonsmith@protonmail.com
signal: adamleonsmith.01
linkedin.com/in/adamleonsmith

Thank you for your attention