AI Assurance Taxonomies

Summary report

Department for
Science, Innovation
& Technology

dg:cities

June 2024

Additional documents

This Summary Report provides an overview of the key findings of the study outlined in the main report **AI Assurance Taxonomies – Sector deep dive report.**

Acknowledgements

This independent research was written by Edward Houghton and Emily King of DG Cities ltd with additional analysis by Rudy Sooprayan of YouGov. The authors would like to thank the industry and expert research participants for their time and interest in sharing their insights for this work.

Introduction

Artificial intelligence (AI) has the potential to radically reshape the economy and wider society. A sizeable number of industries are predicted to be affected by its use, and with innovation occurring at a rapid pace, there is heightened interest in AI and the potential value it can bring.

Increased interest brings with it a vital need to ensure AI tools and services work and are used as intended. Described by the UK Government as 'AI assurance', this process is critical to ensuring AI is deployed safely and effectively. There is, however, limited evidence about how leaders and management teams in UK organisations understand AI assurance and the many associated terms that describe assurance strategies, processes and techniques.

This study was commissioned by the Responsible Technology Adoption Unit (RTA) to explore current understanding of AI assurance terminology in UK industry and how terms are used in practice.

Through an online survey of 1347 business leaders and managers across 7 industries and 30 deep-dive interviews with business leaders developing or deploying AI tools and services in HR, finance, and the connected and autonomous vehicle (CAV) industry, this study explores how business leaders understand AI assurance terminology; if and how alternative terminology is being used; and whether business leaders and their management teams see value in a terminology tool for AI assurance.

Findings

Increased appetite for AI, alongside concern over AI risks, highlights the importance of AI assurance

Our survey explored levels of confidence in compliance and the nature of AI risks with business leaders already making use of AI. Whilst the majority believe their organisation plans to continue to use AI tools in the future (86%), less than half (44%) are comfortable demonstrating compliance with UK regulations. There were several reasons noted for this: lack of understanding of the UK's regulatory approach (52%), limited understanding of AI assurance techniques (50%), and limited understanding of AI assurance more broadly (48%). Media/marketing and PR professionals were the least comfortable, whilst medical and manufacturing industries were more comfortable. This may be due to the medical industry being highly regulated and the relative maturity of AI tools used in this sector, compared to more emergent use cases in media and marketing.

We also explored perceptions of AI risk to better understand how different organisations conceptualise and define AI assurance. We found strong evidence that the perceived risks from AI differed across industries studied: leaders in the CAV industry regularly cited safety risks as a top priority, noting the need for additional regulation in the form of the forthcoming Automated Vehicles Bill and assurance practices that promote safety engineering. This differed from leaders in financial services and HR roles, where AI risks were conceived in terms of ethics and focused on issues like bias and fairness, whether in the nature of decisions related to consumer banking or in the delivery of recruitment and selection processes:

“Fairness is the key one. And that intersects with unwanted bias. And the reason I try and say ‘unwanted bias’ is that you naturally need some (bias). Any AI tool or any kind of decision-making tool needs some kind of bias, otherwise it doesn't produce anything. And so, I think front and centre is how does it work, does it work in the same way for all users?” HR, Developer and procurer, Private Sector

Overarching terminology differs across sectors – but there was little evidence of alternative terms beyond those tested.

We tested three key overarching terms- ‘Trustworthy AI’, ‘Responsible AI’ and ‘Ethical AI’- to understand business leader preferences for terms that describe AI systems that work and are used as intended. We found a broadly three-way split of preferences for the terms – ‘Trustworthy AI’ was preferred by almost 3 in 10 (28%) whilst ‘Responsible AI’ was the most preferred term by almost two-fifths (38%). A fifth (20%) of respondents preferred ‘Ethical AI’. A key industry difference to note is that the medical and health sciences were more likely than other industries to prefer ‘Ethical AI’ (38%) compared to the industry-wide average of 20%.

These preferences for different terms were also apparent and further elaborated on in our interview data– HR and recruitment leaders referenced transparency and visibility in relation to ‘Trustworthy AI’, with several connecting trust to important HR concepts such as equality, diversity and inclusion (EDI). In the financial services sector, several interviewees noted the subjective nature of ‘trust’ and the importance in understanding the context of deployments. Several financial services participants reflected on the recent history of their sector and the move to ‘responsible banking’ after the 2008 global financial crash - noting that ‘Responsible AI’ was more objective, referencing risks as well as intent and design.

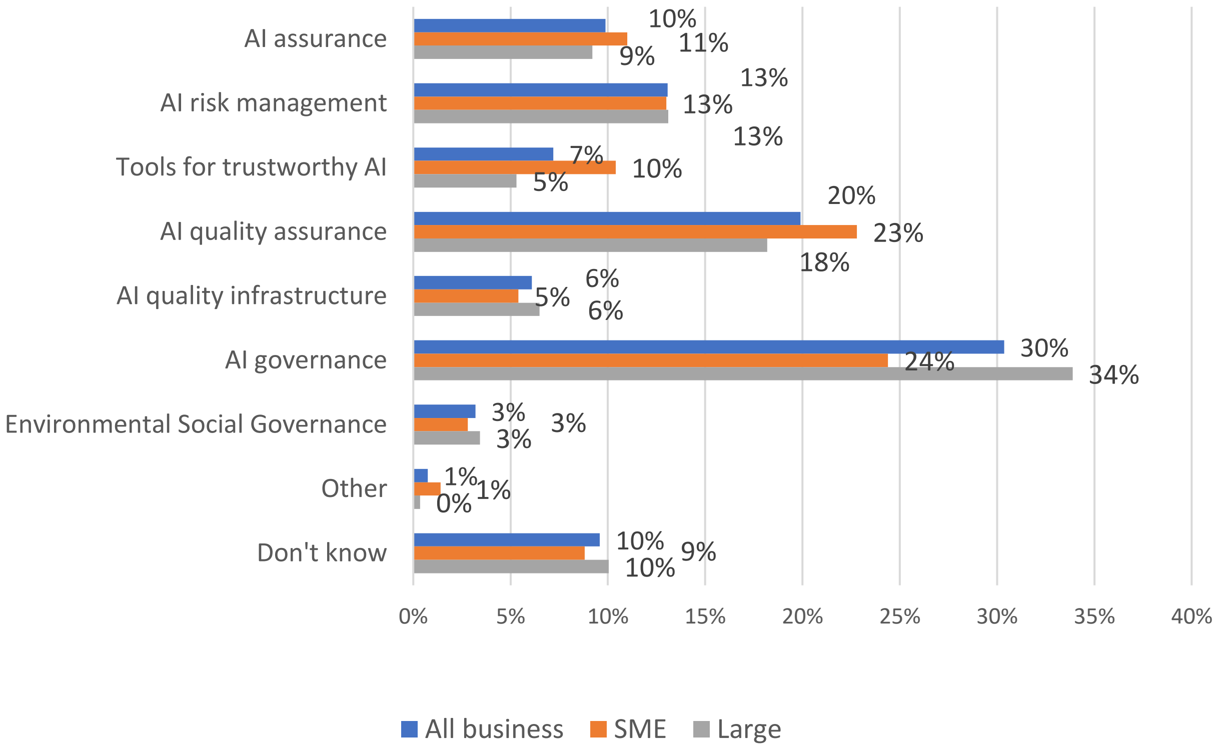
We found little evidence of alternative terminology beyond the three terms we tested. Terms such as ‘functional AI’ were referenced by a small number of respondents but this was not significant.

Whilst current terminology use is diffuse, ‘AI governance’ is the most commonly used and preferred term compared to terms like ‘AI assurance’ and ‘risk management’.

The use of terms to describe processes the ensure that AI works and is used as intended followed a similar trend to overarching terminology. ‘AI risk management’ (25%), ‘AI quality assurance’ (26%) and ‘AI governance’ (28%) were most used by survey respondents, whilst use of the UK Government’s preferred term ‘AI assurance’ differed across industries: manufacturing (21%) versus financials services (10%). We found that highly regulated, high-risk industries appeared to prefer ‘AI risk management’: financial services (30%) versus retail (19%) and media/PR/marketing (16%). We also note that the media/PR/marketing sector were the most likely to state they don’t know (44%) versus the industry-wide average of 33%, illustrating a potential difference in the maturity of AI and efforts to mitigate the risks associated with it across industries.

There were, however, clear terminology preferences across the sample. The term ‘AI governance’ was preferred by 3 in 10 (30%) of respondents (figure 1), with larger organisations more likely than SMEs to choose this term (34% vs 24%).

Figure 1: Preferred terms for describing the processes to ensure that AI works and is used as intended (n = 1347)



One finding to note is the association between some terms and existing and established business processes. ‘AI risk management’ appealed to several finance respondents, who referenced the value of augmenting existing processes or including AI within their established risk management processes: participants in the CAV industry alluded to risk as a practice to be mitigated and managed, requiring active engagement:

“AI risk management feels hugely right to us as a bank. Risk management is a big domain of which managing AI risks is one small part, whereas AI assurance is specific to AI. I don't particularly like ‘AI assurance’ – it’s not clear. It’s too woolly. AI assurance doesn't talk about risk. Responsible AI I find better.” Finance, Developer and Procurer, Private Sector

There are complex barriers to understanding of AI assurance terminology

We found multiple barriers limiting business leaders’ understanding of AI assurance terminology with only 6% of respondents stating they recognise no barriers. Lack of international standardisation (31%), lack of clear definitions in the UK ecosystem (26%) and limited organisation-wide understanding of AI assurance (26%) were all cited, with no clear consensus across respondents. Leaders we interviewed referenced the lack of commonly agreed, shared vocabulary as a major challenge:

“Is the language of AI assurance clear? I don't know whether it's the language per se, I think there's probably a lack of vocabulary… to me it's a question of ‘what are you assuring? What are you trying to show that you've achieved?’ And that all stems from: ‘what does the public want from the technology, what do they not want, what do regulators expect to see, how much evidence is enough evidence?” CAV industry, Developer, Private Sector

AI assurance techniques are broadly recognised, with those related to established business processes standing out to respondents.

Results from the survey showed variation in the level of familiarity towards AI assurance techniques. The vast majority of business leaders surveyed were familiar with risk assessment (98%), performance testing (94%), and compliance audit (94%) as AI assurance techniques. However, less than two-thirds of survey respondents were familiar with bias audit (62%) or conformity assessment (61%). However, respondents were even less familiar with more nascent mechanisms that could be used to assure AI systems, such as ‘Model cards’ (37%) and ‘Red Teaming’ (35%).

As with higher-level terms, respondents were familiar with terms associated with established and well-understood business processes to which leaders and managers are accustomed e.g risk management and governance mechanisms. This was further illustrated when business leaders were asked to consider the difference between ‘audit’, ‘assessment’ and ‘assurance – half (50%) of those surveyed stated that these terms had different meanings. Key semantic differences also stood out in reference to this language:

- **‘Audit’** is seen as a more formal, external process when compared to ‘Assessment’ and ‘Testing’, which can be conducted internally. Participants considered an audit to be more concerned with evaluating compliance to high-level principles rather than system performance: *“For me an audit is a form of test which is carried out by a third party, whereas testing and assessment can be done internally.”* Survey respondent
- **‘Testing**’ was perceived to be more closely related to the performance of an AI tool, and viewed as an ongoing process which occurs at the beginning of or throughout a development process rather than once a tool is released: “*Testing is upfront i.e. prior to release of a tool. Audit and assessment come after the fact, at a different stage of the AI lifecycle”* Survey respondent
- **‘Assessment’** was seen as a broader and possibly vaguer term to describe the process of evaluating a tool across a range of dimensions. Like testing, assessments were perceived the be conducted internally whilst a tool is in development or once it has been released. Assessment was viewed by some to be the outcome of testing, lying somewhere in-between the testing process and a formal audit: *“An audit ensures adherence to standards, Testing uncovers vulnerabilities, and Assessment provides a holistic evaluation of a subject, considering various dimensions.”* Survey respondent

There is broad support for an AI assurance terminology translation tool, but its purpose and value must be clearly defined.

A key part of the study was to understand respondent views on the tools and information they would need to support better understanding of AI assurance terminology. We proposed a translation tool to participants and explored their requirements, and views on the potential impact of a tool on their business practice:

- **A translation tool should define terminology and reference standards and regulatory principles:** there was broad support for a translation tool from interviewees, who recognised its value in ‘clarifying the terms of assurance frameworks’ and ‘bringing clarity to how terms relate to one another’. A number of participants believed it should include: terminology definitions (50%), standards terminology (49%) and regulatory principles (43%).
- **No clear preferences for tool format:** there was no clear preferred format for respondents to the survey, with 26% preferring an interactive tool based on regulatory principles, 21% preferring a glossary of terms, and 21% preferring a tool linked to the AI development lifecycle. The tension between accessibility and comprehensiveness of information was noted by participants, drawing out a need for different levels of information according to individual preferences of the intended users of the tool.
- **Potential impacts included shaping practices and building confidence in assurance:** Participants referenced the potential value of a tool that connected to existing frameworks, including the National Institute of Standards and Technology (NIST) Risk Management framework, and relevant standards to support building knowledge and capability. This was considered to be of real value to SMEs.

Some business leaders were concerned about how to clearly position the tool as informational rather than as setting standards on what is required in an assurance process. A risk here is that citing specific assurance techniques within a centralised terminology tool could be seen to indicate that these techniques are mandated/required, rather than being left to firms to decide or utilise to their own discretion.

Conclusions

AI assurance plays a central role in developing a safe and trusted AI ecosystem and is a vital practice for firms of all sizes and industries. This research highlighted that firms are actively considering AI risks to their business but that less than half were comfortable that they could demonstrate compliance, and of those over half (52%) cited a lack of understanding of the UK’s regulatory approach as a key barrier. This indicates that AI assurance has yet to mature or be fully understood by UK business leaders.

The diversity of terms used to describe AI assurance practices appears to compound what is already to many a complex field. At a high level, terminology such as ‘Trustworthy AI’ and ‘Responsible AI’ appear to resonate but are despite some perceived differences were often used interchangeably in practice. No clear preferences surfaced but we also found no evidence of alternative terms through our testing. Industry differences here pointed to potential challenges for an ecosystem-wide approach, making wide-scale adoption of the same term unlikely. Terminology will likely continue to emerge and adapt to meet the needs of specific sectors and industries.

Similarly, we found a diverse set of terms being used to describe AI assurance and an indication of preferences towards terms which reference established concepts such as ‘AI governance’ and ‘AI risk management’.

Of the concepts tested, terminology that describes AI assurance techniques was less well established. Terms that relate to existing practices and processes- such as risk assessment- appeared to resonate with participants, whilst some more nascent techniques such as ‘red teaming’ were less readily understood.

Our data highlights a diverse set of perspectives on AI assurance terminology preferences, with some industries- such as financial services- referencing active work in developing governance and risk management practices which use the terminology tested e.g ‘AI risk management’. The preference and familiarity of governance and risk-related terms here points to a potentially useful way of framing AI assurance that is likely to appeal to senior management teams.

The challenges related to learning and adopting new terms organisation-wide were clear but we found a desire in leaders to learn and better understand AI assurance terminology. For this reason, the concept of a translation tool was met with broad support. Industry specific searches and tools to support learning across all levels of knowledge were considered useful. Finally, we note that, throughout the study, leaders referenced the importance of leadership in the AI assurance ecosystem, and the value of tools, guidance and knowledge sharing that supports industry to utilise AI responsibly.

This publication is available from: www.gov.uk/dsit

If you need a version of this document in a more accessible format, please email alt.formats@dsit.gov.uk. Please tell us what format you need. It will help us if you say what assistive technology you use.