Responsible AI UK response to the DSIT Public Consultation on the AI Management Essentials Tool

Date: 3 February 2025

We are submitting this response to the Department for Science, Innovation and Technology (DSIT) consultation on the new AI Management Essentials (AIME) tool[1] on behalf of Responsible AI UK (RAi UK), an open and multidisciplinary network that brings together researchers from across the four nations of the UK to understand how we should shape the development of AI to benefit people, communities and society. To arrive at this response, we sent out a call to Principal Investigators and Co-Investigators of RAi UK funded projects[2] to contribute to this consultation. What follows is a synthesis of the responses we received from our research community. Overall, RAi UK sees the AIME tool as a valuable first step for Small to Medium Sized Enterprises (SMEs) and Startups to adopt towards implementing robust and responsible AI governance practices. Our contributions below aim to improve the tool's usability and impact.

1. What are your general impressions of the AIME tool?

We welcome a UK Government developed and endorsed tool with the aim of producing unity in a complex environment of frameworks and guidelines. The AIME tool breaks down complex topics into manageable segments for people in companies and organisations. As experts in responsible AI, we would welcome the opportunity to evaluate AIME's scoring system and proposed recommendations. This will allow us to give insight into the complexities of evaluating the tool in its entirety. Provision of the second and third components of the full AIME tool to the UK's responsible AI research community would allow more comprehensive feedback.

The focus on AI-based systems could expand to cover terms such as Automated Decision Making (ADM), Automated Fraud Detection, Automated Driving, Automated Negotiation, Facial Recognition, and Algorithmic Trading. Companies already use many tools that incorporate AI, but often these tools are called something else. Such tools may employ advanced statistical techniques that are simpler forms of machine learning. Such terms will require updating as the technologies underlying AI-based systems evolve over time.

In general, the AIME tool is a useful starting point for people evaluating AI in their organisations because many of the questions that help organisations surface what policies or documents exist. We would propose that reflexive questions that explore the appropriateness of policies could improve the usefulness of the AIME tool. For example, a question on whether an AI policy helps users evaluate if the use of an AI is appropriate for a

[1] AIME tool https://www.gov.uk/government/consultations/ai-management-essentials-tool
[2] RAi UK research https://rai.ac.uk/research/

given function or task could lead to better practices and have a significant impact for responsible AI deployment and use across the UK.

The AIME tool covers various contexts. The challenge, therefore, is that the questions about fairness are at a fairly high level and do not go into any technical detail. For example, questions around accuracy, precision, sensitivity, and the consequences of these on the decisions being made with the tool stay at a level that may not help decision makers reflect on their choices.

Furthermore, the impact section is mostly about whether companies have conducted an impact assessment, not what the impacts are or might be. The tool's focus on internal management processes does not capture the consequences for using or not using the tool – a topic some researchers see as essential for achieving impact. In this sense, the tool could serve as a reminder of the need for these processes and documentation, but it does not include guidance about all the steps an organisation needs to consider to be properly functioning towards AI adoption.

2. Does the overall structure of the tool make sense? Why/why not?

The AIME tool is well structured as a self-assessment tool that businesses can use to establish management practices for the development and use of AI systems, especially when there are no instruments in place. While we recognise the need for improvements (as outlined in other sections of this contribution), the overall structure - comprising a self-assessment questionnaire, rating system, and action points - is logical and easy to follow. It provides institutions and businesses with a clear framework to evaluate their position and develop plans to address areas of concern.

We received suggestions from the RAi UK research network to streamline some sections of the tool and expand others, when considering the targeted audience of SMEs and Startups. Researchers in our network suggested making the document shorter, easier to navigate, and clarifying missing topics.

Streamlining

Some sections broadly cover the same issue or topic and could be merged. For example, Question 6 (Data Management) and Question 8 (Data Protection) could be merged into a single 'Data Practices' section.

One option to address the length of the tool could be to combine Question 3 (Fairness) and 7 (Bias Mitigation) into a broader category. For example, there is overlap in sub-question 3.4 'Do you have processes for monitoring and mitigating unfairness?' and sub-question 7.1 'Do you take action to mitigate bias?'. However, these concepts entail different social and technical aspects, both of which we consider important to be individually assessed.

Another option could be to define these concepts more specifically and in light of reflexive questions. RAi UK researchers pointed out that some concepts are presented vaguely and

broadly. As Bauer et. al. (2014) report,[3] unclear concepts in survey questions can lead to different interpretations, influencing respondents' answers and distorting empirical analysis. This would make the AIME tool easier to navigate without giving the impression of overlapping questions.

Expanding

Question 5 (Risk Assessment) would benefit from additional questions about how an AI system is evaluated. For example, what metrics, tasks, and processes are used? The answers to survey questions 5 and 7 of this consultation below further detail this point, which reinforces the already identified need for more specific guidance as alternatives to high level considerations.

3. Would you change the order of any of the sections/questions? If yes, which questions and why?

Throughout our response, we have made specific suggestions about the order of sections/questions, and proposals for new categories.

Question 2.5 on the frequency of updates to AI policy could ask about the process companies have in place to generate such updates instead. This is because the AI field is fast moving and companies will adopt AI tools with different capabilities over time. Hence, questions around what 'triggers' for policy changes are in place should be asked to understand governance capability.

4. We are planning to format the final version of the tool as an interactive decision tree (loosely based on the Cyber Essentials readiness tool). Do you agree that this format is intuitive/easy to use? Why/why not?

We understand that alignment of the AIME tool with an existing readiness tool will provide familiarity and consistency in format. The Cyber Essentials readiness tool is well presented and easy to navigate. In addition to the survey questions presented, we propose additional feedback for the landing page of the AIME tool. Before starting the self-assessment questionnaire, it may be helpful to provide some information about the ratings and how the action points will be provided. For example, the introduction section or the landing page could describe outcomes like 'a score of X' or 'a green rating' means that the company has good governance practices, or 'a red rating' indicates several areas of improvement. This will provide upfront expectations about survey outcomes. Additionally, at the end of the survey, action points should be provided as clear guidelines for what the company should do to improve their ratings the next time they complete the self-assessment. Finally, in order to foster genuine responses, it may be helpful to clarify that one of the purposes of the

[3] Bauer, P.C., Barbera, P., Ackermann, K. and Venetz, A., (2014). 'Vague concepts in survey questions: A general problem illustrated with the left-right scale'. *EPSA Conference*, April. Available at: http://pablobarbera.com/static/vague-concepts.pdf

self-assessment is meant to help SMEs, Startups and other institutions to self-evaluate and work towards improving their AI governance practices.

5. Are there any questions that you think are difficult to answer? If yes, what are they? Why are they difficult to answer?

As already highlighted above, a few questions appear vague or broad. This makes some questions difficult to answer. For example, it is difficult to answer Question 7 if an organisation does not know what bias is and how to measure it. As previous research has shown,[4] AI systems can be biased and in some cases cause harm. The sub-questions on bias mitigation are a good start for self-reflection. We would like to propose either an explanation of what bias is, or some examples of bias, like gender, race, age, location, socioeconomic standing, etc. As an illustration, was an algorithmic decision to deny someone a loan made based on an applicant's postcode?

Accordingly, the broad definition of fairness can be misleading. For instance, for the various meanings that fairness can assume, it is easy for respondents to mark 'Yes' as an answer. However, that does not necessarily imply that organisations fairness management processes are fair in practice. In sum, it can be hard for businesses to self-access solely with the use of the AIME tool, whether their understanding of fairness is appropriate and their actions reflect best efforts to maximise it.

More specifically on the wording of questions, researchers have identified the need to clarify the following to avoid vagueness.

- The difference between 'the majority' and 'some' in Q.1.2 as it is not clear what is the effect of respondents' answers.
- The meaning of 'legal position' of an individual in Q.4.1.1, which can also confuse respondents according to how they understand the term.
- The description or explanation of 'general failures and errors' in Q.5.3.1, which is open to multiple interpretations, and creates confusion about the meaning of the expression 'performing as expected' in Q.5.3.2. For example, are SMEs expected to keep a log of failure cases, or is the intention to recommend running some evaluations? Other sections have comments to clarify what these terms mean, we recommend adding a similar comment/description here.
- The difference and explanation of what are the *must haves* (to be legally compliant) and the *should haves* in Question 1 (AI System Record). This may be clarified by the scoring system and advice; however we are unable to review these elements of the tool at the moment since they have not yet been made public.

Concrete suggestions for rewording include:

[4] Buolamwini, J. and Gebru, T. (2018) 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification', in *Machine Learning Research*. *Conference on Fairness, Accountability, and Transparency*, pp. 77–91. Available at: https://proceedings.mlr.press/v81/buolamwini18a.html.

(a) Q.2.2 could incorporate 'and where appropriate, consumers and other relevant external stakeholders', and
(b) Q.8.3 could read 'affected data subjects and the data protection authority'.

6. Are there any questions that you think are superfluous/unnecessary? If yes, what are they? Why are they superfluous/not needed?

We do not believe any of the current questions are superfluous. As stated, we welcome the tool as a first step for SMEs and Startups to reflect on their AI adoption practices.

7. Are there any questions that you particularly liked or would find helpful for improving your internal processes? If yes, what are they? Why are they helpful/appealing?

Researchers within the RAi UK community welcome references to UK standards and legislation throughout the document. Examples include the link to the Information Commissioner's Office (ICO) guidance on Data Protection Impact Assessments (DPIAs) and the requirement to respond to concerns in a timely manner, explicitly linking to a 72-hour period under the General Data Protection Regulation (GDPR). These connections ground the tool and give clear expectations to users in alignment with familiar standards.

8. Are there any necessary conditions, statements, or processes that you feel are missing that organisations should be implementing? What are they?

When considering necessary conditions, statements, and questions that are missing in the AIME tool, the RAi UK community raised the following topics for DSIT's consideration in addition to the ones already presented above.

Integrating AIME to Auditing Processes: Considering that the tool functions as a subset of the broader audit process, researchers recommend emphasising the importance of integrating the tool's assessment into an organisation's existing internal and external audit framework. This is beneficial for two reasons. First, managers are already familiar with audit procedures and risk analyses, allowing them to transition from established practices to new AI-related assessments more seamlessly. Second, aligning the tool with existing audits reduces administrative burdens, prevents duplication or gaps, and reinforces that AI is not a standalone issue but an integral part of organisational governance**.**

Accountability and Responsibility: Multiple researchers saw the need for an additional section on accountability, which can address concerns such as: Who is accountable when things go wrong? Do they have processes set up to deal with that? Another possibility to address accountability is for the section on 'Internal Processes' to include questions that inquire about the person or office responsible for the outcomes of an AI tool or system. For

example, Yazdanpanah et al.[5] propose accountability reasoning as a form of responsibility for failure to deliver an allocated task. Accountability and responsibility should be enforceable as specified in the AI Equality by Design[6] framework. Therefore, such questions in the tool could cover clear outcomes of different segments of an AI system, how success is measured for each task, and who can provide guidance or a response should a particular task not be completed as specified.

Procurement and Ethical Trading: Researchers recommend that questions on procurement could also include ethical trading and a history of proposed AI partners. For example, Hikvision and Dahua, two companies that provide surveillance technology, have been banned in the UK[7]. The tool could point SMEs and Startups to such information.

LLMs and training data: The Centre of Excellence in Terrorism, Resilience, Intelligence and Organised Crime Research (CENTRIC)[8] is heavily involved in projects creating AI tools intended to be used by Law Enforcement Agencies within the European Union (EU). Based on recent experience, we recommend that a specific question is added relating to the use of Large Language Models (LLMs) and knowledge in the training of data. Technically, this is covered by existing questions within AIME, particularly sub-questions 7.2 and 7.3. However, a lack of transparency of training data of LLMs is emerging as a significant issue, preventing AI tools developed from LLMs demonstrating their lack of bias. Despite also being covered by a more generic question, both Accountability Principles for Artificial Intelligence (AP4AI) and AI Accountability for Policing and Security (AIPAS) are implementing a specific question about LLMs to explicitly highlight these issues (see also our answer Q10 below).

Risk Assessment and Management: The tool asks pertinent questions relating to known issues and risks in using AI systems. However, it would be beneficial to encourage organisations to consider the possible negative outcomes of NOT using AI. It is likely that AI will become an important part of cyber security. As AI driven cyber threats develop, AI driven cyber security may become essential to combat them. It may also be the case that businesses risk commercial disadvantage from competitors using AI, should they not adopt the technology. This consideration is important for effective risk assessment.

In a public sector context, the State has an unqualified duty to prevent certain harms under Article 3 of the European Convention on Human Rights (ECHR). It may be the case that not utilising a reasonably available AI capability results in a failure to meet this duty. It is plausible that AI may form part of future measures expected of various organisations, including SMEs. For example, this may be in preventing or detecting modern slavery within the supply chain. A failure to use AI may negatively impact on various critical issues, such as violence against women and girls (VAWG), child sexual exploitation and prevention of people trafficking.

---

[5] Yazdanpanah et al. (2022) Reasoning About Responsibility in Autonomous Systems: Challenges and Opportunities, AI & Society [Preprint].
[6] Amani, B. (2021) AI and 'Equality by Design. Rochester, NY: Social Science Research Network. Available at: https://doi.org/10.2139/ssrn.3734665.
[7] Responsible AI UK (2024) *Responsible AI Governance: A Response to UN Interim Report on Governing AI for Humanity*. Public Policy, University of Southampton. Available at: https://doi.org/10.5258/SOTON/PP0057.
[8] CENTRIC https://research.shu.ac.uk/centric/

Moreover, Question 5 could benefit from the inclusion of the following sub-inquiries:

- A question addressing the need to identify potential residual risks, despite meeting legal compliance and having procedures to address issues.
- A question regarding what kinds of recognised standards defined by professional bodies relevant to their sector are followed. These could be potentially given as a dropdown selection. The capability to also refer to domain specific regulations or standards would increase scope and specificity of the tool.
- A question about how they are informed about and share best practices in risk management and system evaluation.

Interoperability: Government publications clearly state that interoperability is a key aspect of their strategy.[9] Within the AIME tool, the following are explicitly referenced: ISO42001, the National Institute of standards and Technology (NIST) Risk Management Framework (RMF) and the EU AI Act. However, it could be beneficial to explicitly state with which of the UK's international partners interoperability focuses. Significant sections of the EU AI Act appear to be in relation to concerns regarding AI use in foreign states, and by foreign states. Similar concerns have been raised by the Head of the Security Service (MI5).[10]

The United States and the European Union have taken different regulatory approaches to AI. Aiming for convergence with both would explicitly benefit UK developers. Clearer guidance on the regulatory frameworks the AIME tool intends interoperability with is critical for progressing the UK AI sector at a competitive pace.

Issue Reporting: In Question 9, there could be a sub-question on how reporting works, what processes are in place in terms of how they can influence development, deployment and evaluation (AIME tool Question 5).

9. Is the tool overly burdensome or unrealistic for the target audience, (i.e., organisations with limited resources to extensively engage with AI governance frameworks, for example, start-ups and SMEs)

The tool does not seem overly burdensome or impractical, as many of its questions align with existing data protection risk assessments. Overall, it appears to be a valuable resource for fostering discussions within organisations, particularly SMEs, that may not have fully considered the risks associated with AI adoption.

10. We are exploring the possibility of embedding AIME in government procurement frameworks. In this model, organisations supplying government with AI products and services would be required to complete the tool to demonstrate baseline responsible AI management processes. Do you agree that this would incentivise organisations to implement responsible AI management systems?

Introduction of the tool will likely influence organisations to develop tools aligning with standards highlighted by AIME. However, we recommend aligning with pre-existing tools that

[9] E.g.: Assuring a Responsible Future for AI - GOV.UK
[10] E.g.: Five Eyes launch drive to secure innovation | MI5 - The Security Service

already have substantial usership, such as the AP4AI[11], specifically the UK-specific version, AIPAS[12], which is currently being finalised for use by UK policing and security agencies. Some of our researchers are involved with the developing AIPAS framework. The accountability considerations ingrained in these tools will help implement responsible AI management systems. Alignment with such frameworks will help to avoid overlaps as well as ensure consistency across different frameworks that are provided for guidance in the UK.

11. Do you believe that embedding AIME in government procurement processes could have an adverse effect on competition (e.g., add disproportionate burden on SMEs, who may have less resources/capacity to fill out a tool like this, compared to larger organisations)?

We do not believe that embedding AIME in government procurement processes will have an adverse effect on competition. Instead, we see it as an important step for self-reflection for SMEs and Startups in ensuring that their AI systems are robust and responsible. In fact, the pointers in AIME align with existing data protection requirements.

About Responsible AI UK

RAi UK brings together researchers from across the four nations of the UK to understand how we should shape the development of AI to benefit people, communities and society. It is an open, multidisciplinary network, drawing on a wide range of academic disciplines. This stems from our conviction that developing responsible AI will require as much focus on the human, and human societies, as it does on AI. Funded by the Technology Missions Fund, we convene researchers, industry professionals, policy makers and civil society organisations.

The organisations and research teams involved in this response to the the consultation include:

RAi UK Leadership and Researchers

- Professor Sarvapali (Gopal) Ramchurn, FIET, CEO, Responsible AI UK, and Professor of Artificial Intelligence at the University of Southampton
- Professor Gina Neff, Deputy CEO, Responsible AI UK, and Executive Director of the Minderoo Centre for Technology and Democracy at the University of Cambridge
- Isabela Parisio, Research Associate, King's College London
- Dr Sarah Kiden, Research Fellow, University of Southampton

[11] Accountability Principles for Artificial Intelligence (AP4AI) framework https://ap4ai.eu/sites/default/files/2023-09/AP4AI_Framework_Blueprint_22Feb2022.pdf
[12] AI Accountability for Policing and Security (AIPAS) https://aipas.co.uk/

RAi UK Researchers and Projects participating in this Consultation Response

- Professor Babak Akhgar, OBE, Director of the Centre of Excellence in Terrorism, Resilience, Intelligence and Organised Crime Research (CENTRIC), Sheffield Hallam University [Impact Accelerator Project: AI Accountability in Policing and Security (AIPAS)]

- Professor Karen Yeung, Interdisciplinary Professorial Fellow in Law, Ethics and Informatics, University of Birmingham [Impact Accelerator Project: AI Equality by Design, Deliberation, and Oversight]

- Professor Maria Liakata, Professor of Natural Language Processing, EPSRC/UKRI Turing Institute AI fellow, Queen Mary University of London [Keystone Project: Addressing socio-technical limitations of LLMs for medical and socio computing (AdSoLve)]

- Professor Marion Oswald, MBE, Professor of Law, Northumbria University [Keystone Project: PROBabLE Futures - Probabilistic AI Systems in Law Enforcement Futures]

- Dr Rafael Mestre, Lecturer, University of Southampton [International Partnership Project: Exploring Fairness and Bias of Multimodal Natural Language Processing for Mental Health]

Please send any queries and comments about this response to info@rai.ac.uk.