MOD AI Ethics Advisory Panel

Friday 24th May 2024

Minutes

Attendees

Professor Nick Colosimo, Global Engineering Fellow & Technologist, BAE systems and Visiting Professor, Cranfield University (Centre for Autonomous & Cyber-Physical Systems).

Professor Mariarosaria Taddeo, Professor of Digital Ethics and Defence Technologies, Oxford Internet Institute, University of Oxford.

Professor Sarvapali (Gopal) Ramchurn (Responsible AI UK, CEO and UKRI Trustworthy Autonomous Systems (TAS) Hub, Director

Professor Peter Lee, Professor of Applied Ethics, University of Portsmouth.

Dr Darrell Jaya-Ratnam, Managing Director, DIEM Analytics

Paul Wyatt, Director General Security Policy

Lt Gen Tom Copinger-Symes, Deputy Commander, UK Strategic Command

Air Vice Marshall David Arthurton – Defence AI Centre

FMC Representatives

Army AI Centre (AAIC) Representatives

RAdm Jeremy Bailey, Submarine Delivery Agency (SDA)

Chief Data Officer, SDA

Dr Chris Moore-Bick, Defence Science & Technology Policy

Defence AI and Autonomy Unit Representatives

MOD Legal Adviser

Declined:

Richard Moyes, Managing Director and co-founder, Article 36

Tabitha Goldstaub, Exec Director of Innovate Cambridge, Chair of the AI Council

Dr Merel Ekelhof, Foreign Exchange Officer at the US DoD Chief Digital and Artificial Intelligence Office, attending the panel in her personal capacity.

Professor David Whetham, Professor of Ethics and the Military Profession, Kings College London

Alison Stevenson, DG Delivery & Strategy

MOD CSA – post gapped.

1 Introduction and Updates

MOD’s 2nd Permanent Secretary (2PUS) welcomed members of the panel, making the following points:

- We have published our previous panel minutes and as a matter of routine, we will continue to do so.
- We will discuss items on the practical implementation of our responsible AI policy and also consider how Defence should deal with high-risk, high-speed AI use cases, by looking at the current process for Urgent Capability Requirements and how this might need to be adapted if AI comes into play.

Dr Chris Moore-Bick presented an update on the MOD’s latest Responsible AI policy implementation:

- Within the context of *AI Readiness* we are looking at the key enablers (i.e. people, process (including policy), information, technology) that we need to mature to reach the tipping point where we can scale AI delivery at pace, from niche to enterprise and widespread adoption.
- We have recently internally published our Dependable AI JSP which sets out detailed directives on RAI across the AI-lifecycle. This is an important step in operationalising our ethical principles. We are working to publish the Dependable AI JSP 936 part 1 on gov.uk when appropriate.
- All Defence organisations are nominating an AI Ethics Senior Accountable Officer (AI ESAO) who will be responsible for embedding the right *culture* on RAI across their organisation.
- To support Defence teams we will publish guidance, including on AI ethical risk assessments, looking at international and industry best practice. We will continue tweaking our organisational governance so that we can provide the right and proportionate assurance that RAI is embedded to our ministers and, ultimately, Parliament.
- The initial response to the JSP has been positive, our people want to get this right now, rather than face blockers in the future. Speaking to operational teams we also know that communicating the policy externally, especially to our industry partners is key: Industry plays a supporting role helping us implement best practice, and also communicating JSP requirements back to Defence.
- Responsible AI is not free – we need to factor considerations around training, skills, technical assurance, and ongoing international dialogue into our budgets.

2 Practical Implementation of Dependable AI JSP (55 mins)

- Rather than taking feedback on the text of the JSP 936 part 1 now, we want to share our practical efforts implementing this policy, noting that we are on a journey with this. We seek the panel's feedback and advice on these emerging, practical approaches, helping our AI Ethics Senior Accountable Officers who are at the forefront of this effort. We will hear a briefing from the Army and Submarine Delivery Agency (SDA) on their implementation of this policy directive as they are among the frontrunners in Defence.

Army AI Centre (AAIC) Presentation:

- The Army is establishing the AAIC to work with the Defence AI Centre (DAIC), aligning its AI and data efforts with Defence regulations and best practices. The AAIC will support the Army's AIESAO in the oversight of AI risk management and ethical assurance, facilitating the use of good practices and building a comprehensive understanding of the Army's AI portfolio. Preparations for the Responsible AI compliance statements are also in progress.

Submarine Deliver Agency Presentation:

- We publicly committed to ensure that – regardless of any use of AI in our strategic systems – human control of our nuclear weapons is maintained at all times. We will continue to work internationally to reduce the risk of nuclear conflict and enhance mutual trust and security and will continue to promote and engage with international dialogue aimed at identifying and addressing crucial AI-related strategic risks.

- The SDA already has well-established safety and governance processes in place and with the arrival of AI additional considerations need to dovetail into existing structures. Risk balance cases and risk tolerances play a critical role in oversight, with extensive learning from our safety culture helping to determine the impact on our workforce.

In conversation, the following points were made:

- The work of both the Army AI Centre and the SDA is testament to how seriously we take our responsibility in assigning human accountability over AI and being clear on our risk assessment.
- AI Ethical Risk Assessment: AI ethical risk assessments should not become a mere tick-box exercise. Whilst the AI ethical risk assessment will build on traditional approaches of how MOD understands risk, when it comes to interpreting the AI ethical principles MOD needs to carefully arbitrate where risk manifests as this will be different in each AI use case, depending on the operational context. Implementing ethical approaches does not involve purely binary decision-making – Defence teams will need to find acceptable trade-offs between different considerations; this process cannot be simplified into a mere checklist.
- Identifying different risk thresholds is an ongoing process. A diverse range of relevant stakeholders need to be included in AI risk assessments. Safety cases might need to be dynamically reassessed.
- MOD should adopt an 'Ethical by Design' approach (on top of 'Secure by Design').
- Training and Building Trust: The axis of complexity potentially increases for commanders who are in charge of deploying AI-enabled systems – they will need to develop a deeper understanding on the use of such systems that goes beyond a basic training package.
- It can be difficult to build trust in AI systems – rather than providing binary systems, industry partners should build-in a dial for human involvement so users can set the right bounding parameters.

- Toolkits: When building ethical guidance and toolkits we need to ensure they are embedded within and influence existing procurement structures. Also, future AI guidance toolkits should include guidance on Article 36 Reviews.
- Next Steps: SDA and Army should share their approach to AI Governance across MOD and with other AI ESAOs so they can learn from good practice and prevent information from being stove piped.

3 **Briefing on Urgent Capability Requirements (UCR)**

- With the fast-paced technological advancements and increased use of AI-enabled capabilities in conflicts we want to think ahead and prepare for a time when Urgent Capability Requirements might involve AI-enabled systems. Clearly, we have a duty to develop and use such capabilities in a way that prevents any unwanted, unethical outcomes.

- We must also strike the right balance between having the right processes and structures in place and not undermining the tempo of operations. Discussing our risk-based approach now will help us be prepared for the technological inflection point rather than reacting to it cold.

- We would like the panel to provide advice about how we might need to adjust our processes and governance to approach UCRs (and other AI acquisition by extension), considering the balance between pace and assurance and the agility of the current system.

The presentation covered the following points:

- Background to the origin of UCRs and current context, noting that the war in Ukraine and the democratisation of AI technologies suggest that we need to get our MOD internal relationships and structures set up well for a time when AI might manifest in the UCR process, so that we can adopt safe capabilities at pace in order to outpace the adversary.
- Brief outline of the UCR Working Group Process and the role of different stakeholders, including the existing checks and balances in place.
- Considerations on our risk-based approach to AI-enabled capabilities: The competing forces of time vs performance and cost create a tension in how we manage risk: how do we balance tempo and rigour?

In conversation, the following points were made:

- Achieving strategic advantage as a small island nation will rely on us being able to have an agile procurement model, and the UCR process provides a useful blueprint.
- An integrated approach to proper testing and evaluation cannot be a niche activity but must be at the heart of what we do. The AI Ethics Advisory Panel remains an important sounding board as this becomes mainstream.
- Trust could be seen as a result of sufficient reliability, explainability and predictability. Air travel meets that trust threshold. When capabilities are put into service at speed, the ability to quantify predictability might be reduced because the time frame for rigorous testing has been constrained. This might have to be made up for by increasing the explainability of AI-systems.
- Legality must not be seen as a separate consideration as there are some non-mutable requirements.
- Building trust in an AI system takes time – it would be questionable to make trade-offs in robustness and reliability for the sake of speed – as such, a key consideration becomes the Human-Centricity principle and achieving robust human-machine teaming in this process.

4	**Any Other Business and Closing Remarks (10 mins)** The following AOB was raised: • Transparency: As part of our commitment to transparency, we previously discussed publishing any conflicts of interest of panel members. The DAU will reach out to compile these.