



AI ETHICS ADVISORY PANEL
Minutes
Thursday 26th September 1600 - 1800

Attendees

Paul Lincoln (2nd PUS) - Chair

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

Dr Darrell Jaya-Ratnam, Managing Director, DIEM Analytics

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

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

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

Professor Mariarosaria Taddeo, Associate Professor and Senior Research Fellow, Oxford Internet Institute, University of Oxford; Dstl Ethics Fellow, Alan Turing Institute

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

David Blackall (Director Digital Exploitation for Defence)

Cdre Rachel Singleton (Head of the Defence AI Centre - DAIC)

Dr Chris Moore-Bick, Defence Science & Technology Policy

AI Senior Scientist, Dstl

Principal Human Factors Specialist, Dstl

AI Ethics Policy Adviser, DAU

Summer Intern, DAU

Policy Adviser, Security Policy Operations

1	<p>Introduction and Updates from the Chair</p> <p>MOD's 2nd Permanent Secretary (2nd PUS) welcomed members to the ninth meeting of the Ethics Advisory Panel, making the following points:</p> <ul style="list-style-type: none">• Having published previous minutes, as a matter of routine, we will continue to do so.• Since the eighth meeting, the UK co-hosted the REAIM Summit 2024 in the Republic of Korea (RoK), completed work on the AI Ethics Risk Management Toolkit and formally launched our community of Responsible AI Senior Officers (formerly known as AI Ethics Accountable Senior Officers).
----------	--

2 Update on AI Ethics Training Tool

- In Spring 2024 Dstl published a tender for organisations to bid for a contract to develop an AI Ethics Training Tool – through a competitive selection process a King’s College London team led by Prof David Whetham, working with Compass Ethics, secured the contract and has been working with Defence stakeholders to create the tool.

Prof David Whetham’s presentation covered the following points:

- Context on the creation of the original [military Ethics Playing Cards was provided](#), noting the positive impact they have had so far - including significant global uptake.
- The military ethics cards were digitalised, translated into 12 languages and expanded to include a medical version of the cards. They are free to download and use via app.
- The development for the AI Ethics Training tool will follow a similar research methodology as that for the previous cards: The need for AI developers to understand AI ethics, and especially within the wider military ethics context was identified.
- The tool will present prompts and questions on AI ethical considerations to help the AI developer community have a conversation about the practical implications. These questions will be further expanded upon through a QR code link, allowing a verbal exploration of the different topics in addition to the other supporting material.
- It will focus on ethics foundations, military ethics, AI ethical principles and case studies.

In conversation, the following points were made:

- The training tool is being created for internal and external developers and these stakeholders were consulted as part of the research methodology. Enabling developers to engage with AI ethics in the early stage of the development process will save time later on in the process because it will help teams identify and mitigate risks proactively.
- The cards should be pitched as an enabler to developers to promote their adoption. The cards will help developers better understand what AI users are concerned about (by virtue of basing the card prompts on user community feedback).
- UK Ethics Specificity: The new tool will centre on UK AI ethics principles, but would enable comparison with the NATO and EU principles as part of their remit.
- Breadth and Scope: The tool will cover AI applications in the military domain as well as AI use cases for business operations which make up a significant part of MOD's AI use.
- Future sustainability: The use of an app would allow for easy updates following formal release. The cards could also link with the AI Ethics Risk Assessment Toolkit.
- The panel agreed the cards, and the principles that it supports more widely, must not be used as a mere compliance (box-ticking) exercise.
- **2nd PUS** noted the tool's potential value in upskilling MOD senior leaders, including the Responsible AI Senior Officers community.

3 Update on REAIM and RAISOs

Responsible AI in the Military Domain 2024:

- 2PUS attended the latest summit on Responsible AI in the Military Domain in Seoul on 9th and 10th September, alongside the Minister of State in the Lords, Lord Coaker. This senior representation demonstrated our strong commitment to the REAIM process.
- A key objective for the UK was to help broaden global discourse on military AI to consider important risks to strategic stability and escalation management, which appeared to be a shared objective with other attending partners. This included hosting a breakout session to discuss mechanisms toward achieving this.
- We also saw a lot more focus on application of Responsible AI (RAI) principles, norms and standards at the more tactical level.
- The broad attendance encouraged political buy-in but also made it challenging to tailor events and strike the right balance between high-level or technical discussions.

	<p>Responsible AI Senior Officers:</p> <ul style="list-style-type: none"> • 18 RAISOs have been appointed, one for each Defence organisation. We are building a strong community which supports and learns from each other as they are all in the foothills of this journey together. • RAISOs recognise the size of the challenge; whether it is working out how organisations most efficiently weave RAI considerations into their existing reporting and governance structures, or how RAISOs will make resources available to coordinate this effort. • The DAU and Defence AI Centre are developing guidance to help teams provide the right evidence on all aspects of Responsible AI to support policy compliance with the JSP 936. • The EAP's support has been very helpful and we expect further meetings to explore RAISO challenges in more detail.
4	<p>Briefing on AI Ethics Risk Assessment Toolkit</p> <ul style="list-style-type: none"> • Previous EAP meetings considered how to strengthen MOD's risk assessment process to properly take AI ethics risks into account. The Dependable AI JSP 936 mandates that AI Ethics risk assessments must be undertaken by every team that is using or developing AI. Members would be briefed on an initial toolkit that had been developed by Dstl to help teams to understand options to practically embed the 5 AI Ethics principles. <p>The Principle Human Factors Specialist (Dstl) presented the toolkit:</p> <ul style="list-style-type: none"> ○ The toolkit was designed to be proportionate, practical, clear, dynamic and beneficial. It is structured to map against the levels of approval set out in the Dependable AI JSP 936 to inform who needs to own the risk according to its level of impact and likelihood. Guidance on ethical risk complements existing risk management carried out by MOD projects. ○ Stage 1 required users to conduct an initial triage risk assessment to identify and assess potential benefits, ethical risks and mitigations – guidance tools include a five-question template, principles-based question sheets and a use case library. ○ At Stage 2, projects with ethical risks scoring medium or higher should engage relevant SMEs and stakeholders to better understand and mitigate risks. Projects with high or very high ethical risks require more senior approval with support from ethics experts. ○ The toolkit includes a log to document risks, including the context of the benefits which are to be achieved and also to log mitigations. ○ The products were still being refined, for incorporation into JSP 936 Part 2 in due course. <p>In discussion, the following points were made:</p> <ul style="list-style-type: none"> • Focus on whose responsibility it is to update ethics risk logs (team effort vs individual). • How would MOD access the expertise needed to support ethical risk assessment. • The inclusion of a user feedback mechanism for the toolkit. • It may be more consistent / efficient to focus on fewer process options at stage two. • Observations on terminology (AI ethics versus Responsible AI). • Involvement from DAIC seniors will be important to facilitate operational adoption.
5	<p>Any Other Business and Closing Remarks</p> <p>The following AOB was raised:</p> <ul style="list-style-type: none"> • A potential future agenda item was suggested by Prof Peter Lee: An academic critique on public perceptions of autonomous weapons systems and human dignity.