AI ETHICS ADVISORY PANEL
Minutes
Wednesday 3rd December 2025, 1500 - 1630

Attendees

Professor Tim Dafforn (Chief Scientific Adviser) – Chair

Dr Alex Blanchard (Stockholm International Peace Research Institute) – Guest Speaker

Professor Nick Colosimo, Head of Group Science & Technology & CTIO Engineering Fellow, BAE Systems and Visiting Professor Cranfield University (AI, Robotics & Space)

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

Dr Darrell Jaya-Ratnam, Managing Director, DIEM Analytics

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

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.

Dr Chris Moore-Bick, Head of Defence Science & Technology Policy and Head of Defence AI and Autonomy Unit (DAU)

AI Ethics Policy Adviser

Policy Adviser, DAU
Private Secretary for DG Transformation
Military Assistant for Director DD-SMD

1	**Introduction and Updates from the Chair** MOD's Chief Scientific Adviser (CSA) introduced himself as the chair (covering for DG Transformation) and welcomed members to the eleventh meeting of the Ethics Advisory Panel, emphasising the value of external challenge and academic advice to the MOD.
2	**Update on Defence Reform** The Head of the Defence AI and Autonomy Unit (DAU) noted that since the panel last met in February 2025 Defence has undergone significant changes. He referred to the Strategic Defence Review (SDR) and described the new Defence Reform operating model and explained the relationship between the different Areas.

	• The main changes have been enacted, but we are in a period when detailed mechanics and relationships are still being embedded and clarified. This includes confirming the future chair of this panel. The SDR contains several references to using technology and AI to strengthen our warfighting readiness and the panel's remains highly relevant in informing MOD's approach to AI adoption.
3	**Presentations and Discussion on Updating Responsible AI Policy** The Head of the Defence AI and Autonomy Unit (DAU) updated panel members on emerging thinking on the need to refresh our strategic approach to AI to align with the Departments broader strategic vision, noting that: • Considering the changed context since 2022, rapid technological evolution (including the 'ChatGPT' moment in November 2022) and the threats and opportunities set out in the SDR. • This provides an opportunity to reaffirm MOD's commitment to ethical principles and ambitious AI adoption, while exploring ways to strengthen Responsible AI (RAI) policy. Future RAI focus areas may include resilience, governance clarity, and alignment with emerging technological and strategic considerations, and input was sought on potential areas for further development. In discussion, panel members made the following points: *On context and key challenges* • An updated narrative should emphasise the changed context set out in the SDR and consider risk appetite for the use of AI – there should be a focus on the ethical risks of not using AI, whilst remaining cognisant of the challenges to AI adoption (i.e. automation bias, etc). • The updated narrative should highlight a list of key evolving challenges and risks of AI adoption, recognising that it is difficult to provide a list which doesn't become outdated quickly, given the rapid pace of technological evolution. • This list of challenges should include a focus on agentic AI and cyber security risks. *Opportunities for clarification and improved communication* • Recognising the AI hype cycle, an updated document could clarify definitions of AI to avoid 'AI washing'. Credible, outcome-driven technologies are essential to avoid user fatigue from overpromise and under delivery. • The document should also explain the terms 'AI Assurance' and 'Responsible AI' to ensure common understanding. • The narrative should recognise technological developments which point towards Artificial General Intelligence (AGI) and its potential implications for AI assurance and human control. • A refreshed document should be accompanied by a strong engagement and communications plan, including senior champions supporting the messaging. *Thoughts on Case Studies:* • Current proposals are lacking a focus on the concept of trust: A few years ago, documents worked to convince audiences of the potential of AI, now they need to focus on providing better understanding, for example, by focussing on explaining AI products and their benefits and limitations. • MOD could highlight how AI ethical principles have been implemented through different case studies as each raise different considerations and trade-offs. • Those case studies could highlight the breadth of people involved in AI development and use, and who makes decisions (beyond Responsible AI Senior Officers).

	• The updated document might emphasise the interoperability with the NATO Principles of Responsible Use, in the context of the SDR's objective for MOD to adopt a NATO-First policy. *On narratives and human machine teaming:* • Ethics should be seen as an enabler, providing clear boundaries that allow confident decision-making and faster progress, while managing trade-offs between principles, risks, and opportunities. • On human-machine teaming: There should be a greater focus on enabling the human autonomy. AI might risk narrowing the opportunities for a human to meaningfully influence an AI-system's output. To counter this, there must be emphasis on establishing ways for the human to disagree with the machine (recognising the relationships between Predictability, Autonomy and Control). • Related to the above, the role of the human as a legal and moral point/node should be explained and highlighted by drawing on some of the constructive thinking in international forums and recent research on human control. *Comments on the SDR:* • The SDR makes references to learning innovation lessons from Ukraine – the challenge here is that in the past the inertia of a large organisation like the MOD can make rapid reactions difficult. Defence Reform restructuring might improve this, and in this context, narratives on ethics as an enabler should be emphasised. • It was a missed opportunity that the SDR did not sufficiently recognise MOD's responsibility in shaping the normative and legal landscape with respect to employing AI technologies. Against the backdrop of societal expectations, the MOD can be more ambitious in this area. *Strengthening the work of the Panel* • Panel members highlighted that as part of credibly engaging with these issues, MOD could do more to facilitate time for the AI Ethics Advisory Panel to help inform policy positions and practical considerations, through an increased frequency of meetings like this. The chair noted the action that the secretariat will develop a forward look plan and rhythm of upcoming meetings to be shared with the panel.
4	**Presentation on AI-enabled Decision Support Systems (AI-DSS)** Dr Alex Blanchard (SIPRI) presented his 2025 paper *'Autonomous Weapon Systems and AI-enabled Decision Support Systems in Military Targeting: A Comparison and Recommended Policy Responses'*. Key points included: • Since the raise of LLMs (including the 'ChatGPT moment'), more examples of AI-enabled Decision Support Systems (AI-DSS) in the defence context have emerged. • Drivers for adoption of AI-DSS include the organisation of large volumes of information to improve the quality and speed of decision making. This could reduce the need for human analyst to interpret large quantities of data and provide humans with the opportunity for higher-order thinking. • AI-DSS can be used for a wide range of military purposes (e.g. resource optimisation, post-attack assessments, etc). This presentation focusses on AI-DDS use in military targeting which is seen as a high-risk military application of AI. • DSS can comprise a range of model-based procedures for processing data, e.g. rules-based programming in expert systems. • AI-DSS impact the human role in the use of force by altering the process of making a decision to use force.

	• AI-DSS may be intended to support military operations and improve targeting, however, the fact that they shape targeting decisions, coupled with technical limitations of AI, means that they present inherent risks. • Some well-known technical limitations include black box issues, reliance on quality data in large quantities, robustness issues (i.e. system failure when used for new tasks or environments, biases and susceptibility to adversarial behaviour). • All can lead to inaccurate, incomplete, misleading or false information, misrepresenting a non-threat as a threat (e.g. 'false positives' such as mistaking a civilian for a military operator). • Harm only materialises if the human both accepts that information to be true and acts upon it. AI-DSS risk mitigation therefore must focus on whether users can challenge, correct or disregard inaccurate outputs. • Risk determinants include the operating conditions, system affordances and control architectures. In discussion, the following points were made: • The discussion emphasises that technology does not simply assist human decision-making; over time, it can reshape the very role of humans as legal and moral actors and influence how knowledge and authority are understood on the battlefield. Current debates often focus on isolated examples of AI-DSS, but this snapshot approach misses the long-term trajectory. If normative and legal functions (such as compliance checks or ethical reviews) are increasingly managed through systems designed around machine requirements, there is a risk that human oversight becomes secondary to what the technology demands. This gradual shift could redefine standards of accountability and control, making human judgement less central and more constrained by system architecture. • Close attention is paid to AI-DSS in the targeting cycle as a high-risk military activity. However, AI-DSS use cases that lead up to targeting decisions can have 'invisible' but significant influence on later decisions. Critical review of other types of AI-DSS use cases should not be neglected. • Traceability is a key concept: Future doctrine could look at how operators/organisations track instances where different AI systems have been used to inform numerous smaller actions and how they have been checked.
5	**Any Other Business and Closing Remarks** No AOBs were raised, the chair summarised the actions: • Prof Nick Colosimo to share his keynote from a recent autonomy conference. • DAU to develop a forward look plan for a regular rhythm of EAP meetings and share this with the panel. • Panel members to share any additional thoughts on the Responsible AI policy narrative with the DAU.