MOD AI Ethics Advisory Panel

Thursday 25 January 2024

Minute

Attendees

Paul Lincoln, 2nd Permanent Secretary (Chair)

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

Dr Merel Ekelhof, Foreign Exchange Officer at the US DoD Joint AI Centre and former Lead Researcher on AI and Autonomy at UNIDIR, attending the panel in her personal capacity

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

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

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

Alison Stevenson, DG Delivery Strategy

Paul Wyatt, DG SecPol

Lt Gen Tom Copinger-Symes, UKStratCom-DComd

Major General Adrian Reilly (in place of Deputy Chief of Defence Staff Military Strategy and Operations)

Additional Attendees

Dr Chris Moore-Bick, DST Policy

Defence AI Centre Representative

MODLA Legal Adviser

Absent:

Professor Gopal Ramchurn, Director, UKRI Trustworthy Autonomous Systems Hub and the University of Southampton

Dr Darrell Jaya-Ratnam, Managing Director, DIEM Analytics

Cdre Rachel Singleton, Head of DAIC

Tabitha Goldstaub, Founder of CognitionX and chair of the AI Council

Vernon Gibson CB FRS, Interim MOD CSA

Lt Gen Sir Roly Walker, DCDS MSO.

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

1	**Welcome and Updates** MOD's 2nd Permanent Secretary (2PUS) welcomed members of the panel, making the following points: • The seventh meeting of the Ethics Advisory Panel comes at a significant point for institutional and procedural change in Defence. Over the last year we have been developing the Dependable AI Joint Service Publication (JSP) which sets out directions on practical implementation of the Defence AI Ethics principles, along with supporting processes and governance. This will be published internally first and later externally to support transparency and communicate requirements to industry and other stakeholders. • We will bring the community of AI Ethics Senior Accountable Officers together to help set the right culture for Defence organisations. Guidance for these persons will be developed in 2024, covering topics like skills and training, AI ethics risk assessments, and TEV&V throughout the AI lifecycle. • As recommended by the panel, we continue to work with AI project teams across Defence to understand how they are implementing the AI ethical principles. We are developing a repository of case studies to inform policy development and share lessons. • The NATO's Data AI and Review Board has made progress towards a Responsible AI certification standard which will facilitate greater interoperability with allies, both technically and ethically.
2	**House of Lords (HoL) Report and Government Response Update** The Chair made the following points: • The HoL published its final report in December 2023. It covers recommendations on ethics, legal and governance, safeguarding against risks and technical enablers. • The MOD response is undergoing clearances and will shortly be published. • The department also received a call for evidence from the House of Commons Defence Sub-Committee as part of their Inquiry into *Developing AI capacity and expertise in UK Defence*. The Chair invited the panel to comment on the HoL Inquiry process and conclusions; the following points were made: • The focus on the EAP transparency was helpful – MOD should aim to publish as much detail as it can to promote understanding and confidence in our approach to Responsible AI, including through publishing the EAP minutes. • The Committee's engagement with the topic was impressive. "Proceed with caution", the overall message of the HoL report, mirrors the MOD's approach to AI adoption. • The EAP has a key role in ensuring that Defence (and by extension our stakeholders) takes a nuanced approach to the consideration of AI ethics issues; too often the public discussions can unhelpfully suggest that the issues are binary. • Further work is required to understand what context-appropriate meaningful human control looks like for different AI systems with different degrees of autonomy. • The US published its Responsible AI toolkit which is also informing the work of the NATO DARB.

3	**Presentation by Professor Peter Lee** During the last meeting Peter introduced his work on AI Triage on the Battlefield and was invited to present this case study so the panel can learn about how Peter's team have approached ethical considerations. Prof Peter Lee presented 'ATRACT: A Trustworthy Robotic Autonomous System to Support Battlefield Casualty Triage', making the following points: • The ATRACT project will develop and field-test a trustworthy drone-based Autonomous System to help frontline medics in decision-making in the first 'platinum ten minutes' following trauma. This 3-year project is UKRI funded and supported by a team of researchers at four universities. *This is a proof-of-concept project.* • A key strand of the project is to create an 'ethics checklist' for the project. At regular intervals the team will complete the comprehensive checklists and submit statements on the ethical considerations as an intrinsic part of the project. Ethical considerations run through each project development event and team meetings. • Technical challenges include data collection, data preparation, training and testing of machine learning models and deploying the model in server or edge. • Autonomous System control challenges in hypothetical scenarios include co-ordination of a system of drones which will be navigating contested environments to collect data and avoiding dynamic obstacles in a very uncertain environment. • Human-AI Collaboration considerations in hypothetical scenarios include understanding how the collected data will be processed for triaging wounded soldiers to assist the medic; avoiding information overload; and ensuring legal and ethical compliance during data-driven triaging of the wounded soldiers. In conversation, the following points were made: • There are several behavioural models to consider – those of the machine and those of the humans interacting with it and how it might help or hinder the cognitive capacity for medics working under pressure. • Considerations around how automation bias could be prevented – with training and education playing a key part. • Challenges around operating in a contested air space and communication degraded environments. • The implications for civil search and rescue missions and International Humanitarian Law considerations, including prioritisation and triage. • Performance assurance of complex models in complex environments (including how to assure the human-machine team) and human machine interface design considerations. The chair welcomed the presentation, noting that the discussion had helpfully brought to life a range of real-world ethical challenges. The DAU will work with Peter Lee to capture key lessons from the project as part of our emerging body of case studies.
4	**Future of EAP and Transparency Report**

	• Earlier last year, the Lords recommended to increase the transparency of advice provided by the EAP. In our letter to the committee, published in November 2023, we set out that transparency and challenge are central to our approach and that we are exploring options to be more proactive in communicating the work of the EAP, including through publishing the Panel's Terms of Reference, membership, and meeting minutes and, potentially, through an annual transparency report. Our immediate next step is to publish minutes of our meetings in the coming month. Dr Chris Moore-Bick shared the department's evolving thinking of the future of the panel: • The EAP is a key forum for the MOD to receive valuable challenge on our approach to AI Ethics implementation. Over the last year we have reviewed a lot of research which informs our thinking on how we shape our approach to AI Governance, including this panel. • This includes Prof Mariarosaria Taddeo's presentation on a potential methodology for an Ethics Board with greater decision-making powers than this panel currently has. Her paper set out a useful vision for this Board which could consider tolerance thresholds of the satisfaction of the ethical requirements according to the risk magnitude of specific AI use cases. We are considering the learnings from this paper as part of emerging thinking on the processes that Defence requires to effectively consider and manage AI ethical risks, including how we can bring the EAP into this work more closely. • We are considering options to enhance the transparency of the Panel's work, including options for a future transparency report, increasing the range of case studies reviewed by the Panel, and through opportunities for members to provide advice on the DAIC's AI activities. • Panel members agree that this would be a useful way forward.
5	**Any Other Business** The following AOB was raised: • The NATO legal White Paper is exploring the interaction between legal considerations and the NATO Principles of Responsible Use; workshops took place in December 2023. MODLA will keep the panel updated on progress in this space.