AI ETHICS ADVISORY PANEL

Minute of the meeting held on Monday 1st March 2021

Attendees

Attendees

Sir Stephen Lovegrove
Permanent Secretary for Defence (*Chair*)

Lt Gen Doug Chalmers
Deputy Chief of Defence Staff for Military Strategy and Operations (DCDSMSO)

Dr Merel Ekelhof,
Foreign Exchange Office at the US Joint AI Center,

Tabitha Goldstaub
Founder of CognitionX and chair of the AI Council

Professor Peter Lee
Professor of applied ethics, University of Portsmouth

Richard Moyes
Managing Director and co-founder, Article 36

Professor Mariarosaria Taddeo
Deputy Director of the Oxford Digital Ethics Lab

Dominic Wilson
Director General for Security Policy (DG SecPol)

Professor Dapo Akande
Director of the Oxford Institute for Ethics, Law and Armed Conflict

Professor Nick Colosimo
Executive Manager Future Capabilities, BAE systems

Kata Escott
MOD Director of Strategy

Dr Darrell Jaya-Ratnam
CEO, DIEM Analytics

Professor Dame Angela McLean
MOD Chief Scientific Advisor (CSA)

Professor Gopal Ramchurn
Director at the Trusted Autonomous Systems Hub, University of Southampton

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

Also attending

Hd DST Policy: Dr Chris Moore-Bick
CDEI: [redacted]
CDEI: Sam Cannicott
CIO-Data-COE-Analytics Asst Hd: [redacted]
DST-Policy-EmTechPol-AHd: [redacted]
DST-DAU-StratPol2: [redacted]
CDEI: [redacted]
CIO-AI-COE-Engage-Con1: [redacted]
DST-SDIP-3: [redacted]

Introduction

1. The Permanent Secretary welcomed members of the panel, introduced the participants, and provided background to the convening of the panel. The Permanent Secretary made the following points:

    - The Ministry of Defence sees Artificial Intelligence and AI-driven technologies as a source of significant and ongoing benefit.
    - These technologies offer immense potential to speed up decision-making in warfare, reduce the cognitive burden on personnel, and better protect people from harm.
    - These benefits do not come without risks: there are real ethical challenges associated with the use of AI in Defence.
    - Defence must address these challenges to promote ethical AI use on the international stage, and retain the trust of society, but these must be balanced against the needs of a complex global security environment.
    - The panel marks over a year of consultation and policy development on this topic, and is intended to be a permanent and ongoing source of scrutiny.
    - That he will soon be stepping down as PUS to become the National Security Advisor, but wishes to continue to be part of discussions as an observer.

The MOD's work to date developing an AI Ethics Framework

2. Dr Chris Moore-Bick introduced this item, and made the following points:

    - The UK's National Security Capability Review in 2018 identified the rapid growth in AI and Autonomy as a major strategic issue for Defence.
    - Work on AI Ethics for Defence is linked to conversations on Lethal Autonomous Weapons Systems at the UN in Geneva, but the policy scope applies to all possible AI use cases in Defence.
    - Defence has partnered with the Centre for Data Ethics and Innovation (CDEI) as part of this policy development.
    - Defence has carried out over a year of consultation with over 120 individuals from over 80 organisations across Defence, academia, industry and civil society.
    - The DAU has created a set of ethical principles to guide MOD's approach to AI ethics.
    - Developing these principles is one part of getting ethics right: implementing and enacting them across Defence is the other key element.
    - The principles will be published as part of an AI Strategy for Defence in the coming months.
    - The panel is designed to provide the right level of scrutiny to this developing policy as it continues to evolve.

Panel Purpose and Objectives

3. Dr Chris Moore-Bick introduced this item, and made the following points:

    - The purpose of the panel is to be a source of honest and open advice and scrutiny of Defence's developing position on AI Ethics. The group should provide expert advice and challenge to policy documents developed in this area, and to raise issues and topics that Defence should consider.
    - The panel will meet frequently to begin with, then settle into a regular quarterly meeting cycle.
    - The panel will not have formal decision-making powers.

- Panel members should feel welcome to bring topics and issues to the panel for discussion.
- The DAU acts as the secretariat for the panel, with [redacted] the main point of contact for panel members.
- The panel's existence and membership will be publicly disclosed at a point in the future to be determined. Panel members will be fully consulted ahead of this moment of publication.
- Conflicts of interest may arise as the work of the panel continues: these will need to be stated upfront, or may require panellists to be absent for portions of future meetings.

4. In discussion, the panel noted:

- [GR] The Panel should articulate a clear definition for AI. Either to consider any system which automates as within scope (even if it doesn't include machine learning systems), or to focus narrowly on machine learning systems.
- [CMB response] The department adheres to the NATO definition: AI as a system which carries out a task normally requiring human intelligence.
- [TG] That the work of the panel should align with the recently published paper by GCHQ: *Ethics of AI: Pioneering a New National Security*. CMB responded that DAU were deeply involved in the consultation around that paper, and we are confident that our current approach aligns.
- [TG] The panel should involve more military voices and end users to ensure effective range of views are expressed.
- [DC] Currently Defence is developing policy, when work moves towards implementation then end users should be more involved.
- [MRT] The panel should take the NATO definition and build on it. The ethical challenges stem from when AI develops and learns beyond the direction given to it by its developers. Therefore a focus on the key ethical questions, rather than definitions is more productive. The panel should avoid sci-fi concerns or a simplification of AI, and focus on practical problems. PUS agreed.
- [DJR] Technology has to be developed before military end users begin to use new tools: Defence should have a mature view of these technologies from definitions to principles and implementation.
- [ME] [redacted]

Policy Review: Draft Ethical Principles

5. [redacted] (Defence AI & Autonomy Unit) introduced the latest ethical principles developed by the MOD, suggesting some questions for the panel to consider in giving its feedback:

- Does the language in the Humanity principle differ suitably from International Humanitarian Law principles?
- Does the language around context-specific levels of human control dilute the message of the Accountability principle?
- Does Explainability or Understanding offer a higher bar for ethical outcomes in the third principle?
- Whether the language in the principle of Bias and Harm Mitigation is too narrow to specific AI harms at the expense of wider harms from military action
- Whether the Reliability principle could be improved with language of safety or trust.

- [NC] The principles should articulate acceptable thresholds that Defence should accept in terms of an issue occurring and the impact of the effect, against the requirement for that use case to go ahead in terms of military necessity.

6. With regards to the Humanity principle, the panel noted:

- [MRT] The language of the humanity principle could be risky given that it is uncontroversial (all adoptions of AI must be humane), and could be distracting from the wide range of applications these principles seek to cover.
- RM and DJR agreed that this principle is problematic, overlaps with IHL considerations, and isn't clear.
- Instead, it may be better to focus on the justification to use or not use AI, in the context of military necessity for action.
- [ME] This principle should include concepts of military necessity, and avoid the language of IHL.
- [CSA] The language "where reasonably possible" is currently representing a lot of potential outcomes, and should be clarified.
- [DA] There should be no issue with conflating the language of IHL and the ethical principles – as long as this does not result in the MOD being held to lower standards than under IHL alone.

7. With regards to the Accountability principle, the panel noted:

- [MRT] Responsibility and Accountability are different concepts, and the principles should not conflate the two. Moral responsibility should underpin accountability.
- Accountability is related to Control, but the two concepts should be differentiated – appropriate human control leads to suitable levels of responsibility. Meaningful Control is also a suitably complex concept that it would benefit from being made its own principle.
- The principle should make a mention of transparency, as it is this that provides a mechanism to traceable responsibility.
- [DJR] There is a clear distinction in Defence between Responsibility and Accountability in current Defence tradition.
- [RM] Accountability should be decoupled from control.
- [ME] Agreed that two separate principles would befit Accountability and Control. However, control may not be a relevant concept beyond weapons systems: in all other tools there is a great deal of delegated or automated control.

8. With regards to the Explainability and Understanding principle, the panel noted:

- [NC] [redacted]
- [RM] The principles could say more on what the context-specific levels might look like for different use cases.

9. With regards to the Bias and Harm Mitigation principle, the panel noted:

- [NC] Any ethical approach should consider the risk of failure, balanced against a compelling need to automate. Where there is a drive to automate, mitigation of errors and harms is the most ethical approach possible.
- [PL] The MOD should avoid "function creep", as it may not translate well to a non-MOD audience.

- [RM] The concept of harm is currently too narrow, and needs to consider wider harms that could result from using AI.
- [ME] The principles are currently worded as limitations or burdens to Defence, and should be articulated as enabling effective and ethical AI. This principle could be rearticulated as “equitable and safe”.
- [GR] This principle currently emphasises bias too much: discriminatory outcomes are often the result of bad data management / training.

10. With regards to the Reliability principle, the panel noted:

    - [DJR] Trust is a difficult concept to quantify and demonstrate, so shouldn’t be part of this principle.

11. With regards to the principles as a whole, the panel noted:

    - [PL] The principles should further make clear that AI should be introduced within existing organisational context, and traditional structures of rigourous testing processes.
    - [ME] The overarching principle should make a reference to the system lifecycle, and the need to embed ethical practices and consideration of stakeholders at every stage of the process.
    - [GR] The principles should note the need for interoperational systems, and should be developed in consideration of how they might affect the UK’s international partners and our potential adversaries.
    - [RM] The principles should make clear where they are referring to weapons systems and where they are referring to more general use cases.
    - [PUS] There should be a clear taxonomy for where principles affect systems that produce kinetic effect.
    - [DC] [redacted]
    - [GR] The principles as a whole should further consider how AI will work within a system of systems, and how partnerships and human-machine relationships are built within those structures.
    - [ME] MOD should fully explain the scope of what it considers the ethics principles to cover (healthcare, logistics etc), to guarantee that the debate is focused more widely than the LAWS discussions.

12. The panel further discussed how the principles should be affected by wider international approaches to AI ethics, noting:

    - [DJR] [redacted]
    - [MRT] That the UK can show real leadership here, [redacted]
    - [RM] That MOD should avoid the consideration that lower ethical standards should translate to military advantage, and in many ways more ethical AI should be more effective.
    - [NC] [redacted]

Summary

13. The Permanent Secretary closed the meeting, thanking the participants for a valuable discussion. PUS reiterated:

    - That MOD must consider the exact language used in these principles carefully, as they will be scrutinised intently.
    - That the UK should take a leadership role and see the MODs work on this as a source of strength.
    - That the principles should include a careful taxonomy and scope for the use cases the ethical framework will cover.

14. The next session of the Ethics Advisory Panel will be on the 26th March 2021.

Action

- MOD to collate and consider comments raised in conversation, producing an updated set of ethical principles for consideration by the panel.