ABU DHABI FORUM FOR PEACE

AI FAITH & CIVIL SOCIETY COMMISSION

OVERVIEW OF WORKSHOPS

Workshop 1: Truthfulness

Led by Graham Budd, Faraday Institute for Science and Religion

AI ethics and regulatory frameworks seek to enable trustworthy AI that is lawful, ethical, robust and accountable. The practical focus is on AI safety, transparency, privacy and avoiding bias. However, truth-related issues with AI systems, such as hallucination, deepfakes, and disinformation are becoming increasingly important. Truthfulness is vital to enabling trustworthy AI in the longer term because it underpins trust within society, including our financial, legal and political systems. In this session we will explore issues around AI truthfulness and its impact on society, and consider how the value of truthfulness could be better incorporated into AI ethics frameworks and tools, to support the development of trustworthy AI that will have a positive impact on society and human flourishing.

Suggested Reading:

- Budd, G. ‘AI and Humanity: Navigating History’s next Great Wave’ https://www.faraday.cam.ac.uk/resources/articles/ – This includes a paragraph on truth and truthfulness, but also more generally covers why it is important for diverse religious voices to be heard in the debate about AI ethics, and the need for further research.
- OECD AI Principles – https://oecd.ai/en/ai-principles
- Verma, P. De Vynck, G. ‘AI is destabilizing ‘the concept of truth itself’ in 2024 election (Jan 2024) – Recent Washington Post article on AI destabilising the concept of truth – https://www.washingtonpost.com/technology/2024/01/22/ai-deepfake-elections-politicians/

Workshop 2: Accessibility
Led by Dr Kate Devlin, Humanists UK

1) Development: Is AI being developed inclusively? Who gets a say in what is being created?

- How can we understand the development landscape? Who holds the power globally?

2) Deployment: Is AI being deployed inclusively? Can everyone who wants to use AI get to use AI?
Who is being marginalised by this technology? National vs. international.

- From the December discussion, groups identified where accessibility is particularly important:
    - Global South, important AI benefits aren't concentrated in the western world;
    - People from low socio-economic backgrounds;
    - Other disadvantaged groups such as women and people with disabilities.
- Are there downsides to too much access? E.g., an overreliance on the tech?

3) Divides: Who has access to this technology and who is left out? There are those who gain from using AI; others are negatively impacted by it.

- There are already initiatives/organisations monitoring and campaigning for inclusion. Is there a comprehensive directory of these (e.g., by country)? Can we work with them?

4) Hard power vs. soft power: As mentioned in the December discussion, we can think of accessibility in "hard" and "soft" terms: "hard" things like technology, weapon systems, money; "soft" side: education, literacy, cultural norms and choices.
5) Representation: The UK's AI Safety Summit was restricted in terms of attendance. Next AI Safety Summit needs to be more accessible. The AI Fringe is one route to this, and we could be involved in that. What might we do? (Next AI Safety Summit is a virtual mini-summit in May, hosted by S. Korea. Next in-person Summit is France in the autumn.)

Suggested Reading:

- AI Governance Alliance Briefing Paper Series (2024). Generative AI Governance: Shaping a Collective Global Future, https://www3.weforum.org/docs/WEF_Generative_AI_Governance_2024.pdf – WEF report comparing national responses and highlighting the need for equitable access and inclusion, especially for the Global South. Focuses predominantly on governance.
- BBC Global News Podcast (15 Sept 2023). Special Edition - Artificial Intelligence - who cares? https://www.bbc.co.uk/programmes/p0gdvwj5 – AI's (global) impact on healthcare, the environment, the law and the arts in a special edition recorded at Science Gallery London.
- Devlin, K. (Oct 2023). Power in AI: Inequality Within and Without the Algorithm, https://kclpure.kcl.ac.uk/portal/en/publications/power-in-ai-inequality-within-and-without-the-algorithm – A handbook chapter that explores the history of AI and its impact from an intersectional perspective, highlighting the negative effects of bias both in the algorithms themselves and the tech workplace more broadly.
- Good Things Foundation (Dec 2023). AI and the Digital Divide, https://www.goodthingsfoundation.org/what-we-do/news/ai-and-the-digital-divide/ – UK-focused discussion on AI and inclusion from the AI Fringe last November.
- Ringer Morris, M. (2020). MAI and Accessibility: A Discussion of Ethical Considerations, https://arxiv.org/pdf/1908.08939.pdf – This analyses accessibility as it pertains to disability.
- WEF (Jan 2023). DAVOS AGENDA: The ‘AI divide’ between the Global North and the Global South https://www.weforum.org/agenda/2023/01/davos23-ai-divide-global-north-global-south/ – Overview of the disparity between Global North and Global South in the social benefits of AI.

Workshop 3: Transparency

Led by Rabbi Dr Harris Bor, London School of Jewish Studies

What is Transparency? WHy does it matter? Why is it so hard to achieve? The workshop on transparency will prompt participants to explore the concept's definition, significance, challenges, and practical implications. Participants will examine both technical solutions and its alignment with religious or philosophical values.

Attendees will be asked to consider the following topics of discussion:

1. Definition of Transparency
2. Importance of Transparency
3. Challenges of Transparency
4. Technical solutions
5. Transparency and the law
6. Faith and transparency
7. Practical implications

Workshop 4: Justice

Led by Silkie Carlo, Big Brother Watch

A discussion of the novel policy, legal and moral issues of importance to faith groups and civil society arising from uses of AI in criminal justice systems, how they can be mitigated, and the role faith groups and civil society should play in the debate. We will exchange information about AI-based processes entering justice systems and relevant current or expected outcomes.

- Consideration of the issues and risks arising both in AI-related procedures and outcomes (as discussed in our December Conference)
- Consideration of specific use cases as examples: AI surveillance (e.g. anomaly detection, facial recognition) and predictive analytics (e.g. recidivism scoring, predictive policing allocations)
- Consideration of how far human rights and other legal frameworks can/have mitigate/d risks
- Discussion of 'the case for humanity' and how faith-based and secular perspectives inform the discussion