



Ada Lovelace Institute-AHRC Network for AI & Ethics - Further Particulars

The Ada Lovelace Institute and the Arts & Humanities Research Council (AHRC, part of UK Research & Innovation) are seeking a part-time Director for a new Network for AI & Ethics, to be established as a one year project, in partnership with the two organisations.

Background

Recent years have seen growing interest from policy makers, civil society organisations, and the general public in the application of artificial intelligence (AI) and associated technologies. Researchers across all sectors, from academia to industry, have been working to develop increasingly sophisticated algorithms and models for application in a huge variety of important areas, from precision medicine to autonomous vehicles, and from cyber security to virtual home assistants. Simultaneously, the rise of misinformation and in particular the Cambridge Analytica scandal have shone a light on the questionable ethical codes of some individuals, groups, and organisations using AI. Most recently we have seen the emergence of the 'deepfake': fabricated, but visually convincing representations of living people, created using AI, which are then presented as evidence of individuals' actions or statements..

The House of Lords Select Committee report, *AI in the UK: ready, willing, and able?* (2018), recognised that "Consistent and widely-recognised ethical guidance, which companies and organisations deploying AI could sign up to, would be a welcome development". Much thought has been devoted recently to values, such as fairness, transparency, explainability etc., which need to be observed for the development and application of AI to meet appropriate ethical standards and thereby to win public trust. However, as the Nuffield Foundation's *Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research* (2019) stated, "most of the principles proposed for AI ethics are not specific enough to be action-guiding. While these principles do reflect agreement about which aims are important and desirable... they do not provide practical guidance to think through new and challenging situations". That is, we are still a long way from knowing how to translate abstract principle into practice – how to embody the relevant values not only in legislation but also in research and innovation culture. Not least because thinking about 'AI ethics' seems to be scattered across different academic disciplines – e.g. philosophy, ethnography, design, engineering and computer science, and no doubt others besides – which lack a common vocabulary and a common infrastructure for collaboration.

The Ada Lovelace Institute and the Arts and Humanities Research Council (AHRC) are now partnering to establish a network of researchers, policy makers, and practitioners to address this gap and turn values and aspirations into practical actions.

The establishment of the network corresponds to the Ada Lovelace Institute's remit to convene interdisciplinary research and foster diverse debate in order to ensure that data and AI work for people and society. It is envisaged that the network will be informed by, and will inform, Ada's own research, policy and public engagement activities, and will foster close collaboration between Ada and academic communities.

Purpose, scope and structure

The purpose of the network is to deliver a humanities-led collaborative approach to synthesising and building upon existing AI-focused activity across the disciplinary and sectoral landscape. The Network will provide thought leadership in this area, generating new insights through an innovative and interdisciplinary programme of activity, and will seek to influence the development of policy and practice.

Three interconnected and complementary themes have been identified that will guide the network's activity:

1. Ethics – including legal and philosophical approaches
2. Human-Computer Interaction
3. Designing Human-Centred AI-based Services

It is envisaged that each theme will correspond to a workstream operating concurrently in a 'lab' style, with members of the workstream working individually and collectively to engage with stakeholders (including the public, as appropriate), scope skills and expertise, produce briefing papers, working papers and think-pieces, initiate interdisciplinary and collaborative research projects, engage with wider disciplinary and sectoral stakeholders (including the public), and produce a final, shareable report. Details of the activities within these themes will be decided by the Director, in consultation with the theme leads and sponsor organisations. Ensuring integration and cross-fertilisation across workstreams will be fundamental to the success of the Network and a key role of the Director (see 'Role of the Director' below).

Led and coordinated by the Director, the core of the Network will comprise three workstream leads who will work together to independently and collectively deliver outputs and maximise impacts across the piece.

Governance

A small Steering Group, comprised of the Director and representatives of the AHRC and Ada Lovelace Institute, will oversee the Network and provide necessary assurance to the Executive Groups of the respective funders.

Role of the Director

The Director will be a credible and influential champion for humanities-led AI ethics activity across a wide range of stakeholders, in addition to coordinating the Network's activities to maximise uptake. They will provide leadership and direction to the Network, working closely with the sponsor organisations and ensuring open and responsive dialogue across workstreams.

Key to the success of the Network will be the integration and cross-fertilisation of workstream activities, and therefore a vital role of the Director will be to capitalise upon synergies, identify and propose ways to address gaps, and draw together the workstreams into a coherent whole. Responsible for setting the overall intellectual direction, the Director will ensure effective delivery of the theme labs and maximise the Network's outreach and impact. The Network will be developed across the themes, and in consultation with the sponsor organisations with the aim of building a truly interdisciplinary network that generates new insights and forges cross-disciplinary vocabulary and ways of working. Where the wider engagement activity reaches beyond the academy, the outputs of the Network will inform practitioners, policymakers and the public.

As a first activity, the Director will work with the sponsor organisations to develop the membership of the Network across all three themes. Membership will initially consist of theme/workstream leads, who will be experts with track records of coordinating and/or participating in collaborative research and engagement activities. Once the initial Network is established, it is anticipated that further associated members will be drawn in as appropriate.

The Director will also take a forward-looking approach, and will support the development of bids for further funding as required.

Funding

Up to £195k is available for a one-year period. The Director is likely to be appointed on a part time basis (secondments considered), and will be supported by a researcher/coordinator and a part-time administrator. Further eligible costs may include T&S, and other expenses as agreed with the sponsor organisations.

Further in-kind support will be provided by the Ada Lovelace Institute, including the possibility of workspace for the researcher/project co-ordinator at the Institute (headquartered at the Nuffield Foundation).

Eligibility

The Director must meet the following criteria:

Essential

- Either based at a UK Higher Education Institution or a UKRI recognised “Independent Research Organisation”, or associated with another leading centre in the field
- A scholar or practitioner from an appropriate disciplinary background and with a track-record of research leadership and of delivering high-profile collaborative research and/or engagement activities
- A commitment to collaboration across multiple disciplines both within and beyond the humanities and to wider public engagement
- An interest in one or more of the three themes identified above, ideally through understanding of the current landscape, including literature, stakeholders, policies and practices
- Commitment to upholding principles of Equality, Diversity and Inclusion

Desirable

- Existing partnerships or working relationships with key actors in the AI ethics landscape
- In-depth knowledge and understanding of emerging technologies

Timeline

The following indicative timeline is anticipated for this appointment and the creating of the network:

- Call for applications closes – 23 September 2019
- Interviews – w/c 07 October 2019