AIG0004

Written evidence submitted by Professor Keith Hyams and Dr Jessica Sutherland University of Warwick

Introduction

1. Professor Keith Hyams and Dr Jessica Sutherland are part of the Interdisciplinary Ethics Research Group (IERG) in the Department of Politics and International Studies at the University of Warwick. They are currently involved in a Horizon Europe project *Knowledge Technologies for Democracy*, which looks at the opportunities and risks for democracy arising from AI. The work of the IERG focuses on identifying ethical issues that arise in areas such as democracy, social media, big data, and AI.
2. AI tools such as natural language processing (NPL), computer vision, and generative AI can be understood as knowledge technologies due to the ways they support and contribution to the creation and dissemination of knowledge. Some key uses in government may include analytic uses to aid decision-making, or generative uses to create summaries of key information.
3. This evidence focuses on some of the ethical risks and opportunities can arise from the use of artificial intelligence (AI) in Government contexts, specifically in relation to the use of automated decision-making and profiling in decision-making contexts, generative AI, and wider public perception of the use of AI more generally.

Automated Decision-making and User Profiling

4. The rapid developments in AI tools have brought with them novel ways to improve efficiency in decision-making tasks that may have traditionally been performed by workers in areas such as public administration, and healthcare. These tools often use automated decision-making and user profiling to assist processes in areas such as administration, operations, and research. These tools are useful insofar as they create capabilities that would not otherwise be possible, as well as increasing the speed and accuracy of operations. Nevertheless, their use of personal data and the bias they may introduce without appropriate human oversight pose ethical risks that need to be addressed.
5. User profiling in decision-making AI tools is used to sort and analyse large sets of personal data and create digital representations of individual users, user groups or identify user types. This allows for instantaneous analysis of big datasets that would be almost impossible if left to human capacity alone.
6. For user profiling to become more precise it must rely on increasingly larger datasets. This data may include various biases either because it is historical data that includes historical structural inequalities, or simply because it is collected by humans who themselves introduce biases in how the data is collected and reported. Because of this,

bias may creep into various aspects of the user profiling process, including the hypotheses used to target profiling approaches, the encoding of the data itself, and the representativeness of training data. This is of particular importance for the use of AI in government given that historical data would likely be imported into these systems in areas such as healthcare, immigration, and administration. Additionally, user profiling may not be truly representative of the UK population if the data it is trained on does not sufficiently and proportionally represent all demographics. If underrepresented groups are not included in the training dataset for example, user profiling may lead to discriminatory decision-making when these omissions collide with existing social injustices.

7. Automated decision-making often relies on user profiling and is likewise typically trained on very large datasets that may include social inequalities within them. These inequalities and biases may be reproduced or even amplified within the decision-making system. This is of course especially problematic if there is bias introduced into government systems that deal with public services. Problems arising from technology are not easy to solve, especially when the proposed solutions do not take into account the social context. It is important to note that the shift from theory-based to data-led decision-making is not specific to the AI boom, and so problems with bias in datasets and data analysis are not new. Nevertheless, the scale on which such decision-making occurs, and the potential for increased lack of transparency behind AI-informed decisions, means that the problems can be significantly magnified.

Generative AI in the Government Context

8. Generative AI systems that utilise deep learning and neural networks can assist in the synthesis of information into various forms. Generative AI is currently regularly used to summarise, translate, or create new data and information, for example written reports, code, or even visual information. This can be useful in two main aspects: first it can assist research administration within government by making information processing faster and more efficient, and second it can provide novel ways for citizens to access information via chatbots, robocalls or other AI assistants. This can be particularly useful for citizen's who may require information in a different language, or in more accessible formats. The recent robocall campaign from New York Mayor Eric Adams in which thousands of robocalls were made in different languages is demonstrative of one use of generative AI to assist in citizens' access to information.
9. Generative AI however often suffers from hallucination issues in that it may generate information that is incorrect. Whilst guardrails can be adopted within AI systems created for specific tasks, more general-purpose AI such as ChatGPT or Bard do not have such specificities built in. Because generative AI such as Large Language Models (LLMs) rely on probability to predict the right pattern of words or data, they sometimes get this wrong, especially when there are no specific rules introduced. As such, the use of generative AI in government work may introduce false information and this may go undetected without sufficient human oversight. This may also be a concern in the use of chatbots which rely on generative AI to communicate information to the public. If incorrect or misleading information is communicated to

citizens, this can of course be damaging to public discourse, democratic participation, and public trust in government.

10. Another concern about increased use of generative AI arises if answers provided by the AI end up displacing appropriate reflection and deliberation about the salient facts and arguments. Generative AI, even when not reflecting bias, leans towards responses that are informed by aggregative tendencies rather than a careful assessment of the nuances and pros and cons of particular positions. While answers provided by generative AI can therefore provide a useful starting point, it is essential that users are required to think critically about the responses that they receive from the AI and assess facts and arguments on the basis of their own merits.

Public Perception and Trust

11. In what follows, we outline some risks to public perception that may arise from an increase in the use of AI in government. These include the issues of black-box decision-making in many advance AI systems, problems of undue trust created by a lack of public information about how AI systems work, risks of normalisation, and issues with the use of shadow AI.
12. One of the main arguments consistently levelled against the widespread use of AI in decision-making contexts is the lack of human oversight. Transparency and accountability are fundamental to the relationship between citizens and government. In contrast, AI systems are often complex and their decision-making processes opaque if not unknowable. Because of machine learning and deep learning techniques, it is not only that AI algorithms may be difficult to explain to a general audience, they may also be difficult to understand by technologists themselves. The use of AI in government may therefore undermine the relationship between institutions and citizens if algorithms are not transparent or explainable to the general public.
13. Transparency about the use of AI in public services more generally is also an important aspect of trust in government. Explaining to the public how, where, and why AI is used (particularly when personal or sensitive data may be involved) in ways that are easily understood is imperative to fostering trust being accountable to the public.
14. AI tools may also be afforded undue trust, both by those who use them and the citizens that their use affects. Because of the pace of advancements in AI and other emerging technologies, education and knowledge surrounding how these tools work and their capabilities is often slow to catch up. This can result in a gap between AI implementation and user understanding about the limits of the new tool. For example, a user may believe the system has a higher accuracy rate than in reality it does. Education and training for the users of AI systems, and information campaigns for the wider public may be crucial in tempering concerns about lack of trust, or too much trust, in AI systems.
15. The use of AI systems by government may contribute to the normalisation of the use of AI in other areas of society and the economy, in ways that bring both opportunities and risks. On the one hand, general unease within public discourse about the use of AI in society may be quelled by the government taking a leading role in AI adoption. The UK government being a key driver of technological implementation may allow the government to shape the values that are important in AI adoption more generally.

However, the normalisation of AI use poses an issue since normalisation can lead to complacency about potential risks that AI may pose more generally.

16. The problems outlined in this evidence may be of particular concern in the use of shadow AI within government. ‘Shadow AI’ is the notion that some employees may already be using AI tools without the oversight of their employers, and without adequate guidelines in place. This brings with it a variety of ethical issues such as company privacy and cybersecurity risks, and the problems of citizens’ personal data being inputted into AI systems without their consent or knowledge.

Conclusion

17. The use of AI in government can provide many benefits such as increases in efficiency and accuracy in areas such as research, administration, and public service provision. However, to mitigate against some of the risks outlined in this evidence, care must be taken to ensure that the use of AI is transparent to the public (both including the algorithms themselves, but also where, when and how AI is being used), that steps are taken to minimise bias and its discriminatory effects, that AI systems are not trusted unduly, and that there is sufficient human oversight in the development of AI tools and their outputs

April 2024