Written evidence submitted by The Citizen-Centric AI Systems Team (University of Southampton)

About Us

The Citizen-Centric AI Systems (CCAIS) team[1] is a multidisciplinary group of academics and researchers at the School of Electronics and Computer Science (ECS) within the University of Southampton. The team is developing the fundamental science needed to build artificial intelligence (AI) systems that can be trusted by citizen end users. It is funded through a 5-year UK Research and Innovation (UKRI) Turing AI Acceleration Fellowship led by Professor Sebastian Stein. The team collaborates closely with Responsible AI UK (RAI UK), a network that brings researchers in the UK together to understand how we should shape the development of AI to benefit people, communities and society.

On behalf of their research group, Dr. Sarah Kiden, Dr. Vahid Yazdanpanah and Professor Sebastian Stein are keen on offering the Responsible Use of Citizen-Centric AI (RECA) framework to the Public Accounts Committee of the UK Parliament in response to the inquiry on the 'Use of artificial intelligence in government' https://committees.parliament.uk/work/8367/use-of-artificial-intelligence-in-government/. Drawing from extensive research and stakeholder workshops, the RECA framework provides evidence-based guidelines to promote the responsible, ethical and transparent development, deployment and use of AI for the benefit of all citizens.

[1] https://ccais.ac.uk

1. Introduction

In the digital age, the integration of AI into government processes holds immense potential to enhance efficiency, improve service delivery, support decision-making, and foster innovation. However, with this potential come significant ethical, societal, technical and legal considerations. As AI systems become more pervasive in government, it is imperative to ensure that they are designed, deployed, and governed responsibly, with a primary focus on benefiting citizens and society as a whole.

In order for AI systems to truly serve the interests of UK society, they need to be designed, developed, and deployed with a citizen-centric approach in mind. This entails prioritising the needs, preferences, and rights of individual citizens while upholding principles of fairness, transparency, explainability, and accountability. To address these challenges and opportunities, we propose the adoption of the Responsible use of Citizen-Centric AI (RECA) framework within the UK government and other institutions. This framework is informed by our established line of research on responsible and trustworthy AI systems [9], on citizen-centric AI systems [12], and our response to the United Nations (UN) report on governing AI for humanity [11], as well as various discussions with key stakeholders in the UK and international representatives [13]. The RECA framework is grounded in four key principles: (1) citizen awareness, (2) citizen beneficence, (3) citizen sensitivity, and (4) citizen auditability. By adhering to these principles, the government and other public institutions can ensure that AI systems contribute to societal wellbeing while mitigating potential risks and harms.

RECA can be used to evaluate AI systems that the UK government employs (that is, in using externally developed AI tools and applications) or to be considered during the design and development of those AI systems that government departments co-develop internally.

2. Responsible Use of Citizen-Centric AI (RECA) Principles

The RECA framework is not a standalone creation, but rather a product of research and collaborative refinement. It is built upon the foundation of the research roadmap on citizen-centric AI systems [12], which provides a comprehensive understanding of the principles and challenges inherent in designing AI systems with a citizen-centric focus. Moreover, the framework has been fine-tuned through stakeholder workshops organised by the University of Southampton and held at key venues such as the Royal Academy of Engineering (RAEng) in 2022 and the Royal Society in 2023. These workshops brought together representatives from diverse sectors including industry, academia, government departments, and non-governmental organisations (NGOs) to contribute their expertise and insights.

a. Citizen Awareness

AI systems deployed by the government must be designed to be fully aware of the preferences, needs, and constraints of individual citizens. This entails personalised and tailored services that respect privacy constraints and empower citizens with control over their data. To that end, the RECA framework encourages the development of AI systems that prioritise citizen preferences, ensure data privacy, and enable meaningful consent [3] in data usage. By incorporating citizen awareness into AI systems, the government can enhance user satisfaction, improve service delivery, and foster trust among citizens.

In order to achieve citizen awareness, AI systems can utilise various techniques such as preference learning, and context-aware decision-making. For example, government service portals can employ machine learning algorithms to analyse user behaviour and preferences [2], thereby customising the user experience and providing personalised recommendations in a diverse and inclusive manner.

Additionally, mechanisms such as privacy-preserving data aggregation [1] and differential privacy can be employed to ensure that individual privacy is protected while still enabling effective citizen awareness. These mechanisms leverage on the use of proxies (that is, an agent or intermediary) between end users and databases that use different forms of encryption and aggregating data inputs to enhance privacy for end users.

b. Citizen Beneficence

The RECA framework emphasises the importance of AI systems in maximising the utility and benefits for citizens and society as a whole. Government-deployed AI systems should incentivise socially beneficial behaviours, promote fairness, and support addressing societal challenges such as climate change and inequality [6]. By aligning AI systems with citizen welfare and societal good, the government can harness the transformative potential of AI technology to create a positive impact on people’s lives.

For promoting citizen beneficence, AI applications can be designed to prioritise outcomes that maximise societal welfare while considering the diverse needs and preferences of citizens. For instance, AI-powered healthcare systems (in the National Health Service (NHS), for instance) can be deployed to optimise treatment plans and improve patient outcomes while minimising healthcare costs. Similarly, AI-driven transportation systems can reduce traffic congestion and emissions, thereby benefiting both individual commuters and the environment [10].

c. Citizen Sensitivity

AI systems deployed by the government must be designed to make fair, inclusive, and equitable decisions that respect the diverse needs and perspectives of citizens. The RECA framework advocates for context-aware algorithms, transparent decision-making processes, and mechanisms for addressing biases and discrimination. By prioritising citizen sensitivity, the government can ensure that AI systems contribute to a more just and equitable society while avoiding harm or discrimination against marginalised communities.

To achieve citizen sensitivity, AI algorithms should be trained on diverse and representative datasets that encompass the full spectrum of demographic characteristics and cultural backgrounds in the UK. Additionally, algorithmic decision-making processes should be transparent and accountable, allowing citizens to understand how decisions are made and providing avenues for recalling issues in cases of unfair treatment or discrimination. This is key for fair ascription of responsibilities for unanticipated harm [14]. Moreover, evaluating AI systems before deployment, and ongoing monitoring and evaluation during and after deployment can help identify and address biases or disparities that may emerge over time.

d. Citizen Auditability

The RECA framework underscores the importance of transparency and accountability in AI systems deployed by government. AI algorithms should be explainable [4], allowing citizens to understand the rationale behind decisions and providing mechanisms for auditing and feedback. By enabling citizens, or citizen representatives, to monitor and maintain the ethical behaviour of AI systems, the government can build trust, enhance accountability, and mitigate the risks associated with AI deployment.

For citizen auditability, AI systems should be designed with built-in mechanisms for transparency, interpretability, and accountability. This may include AI co-development with involvement of citizen end-users (or focus groups), encouraging model explainability, and post hoc auditing of decisions. Additionally, citizens should have access to user-friendly interfaces and tools that enable them to interact with AI systems, provide feedback, and file complaints if necessary. By promoting citizen engagement and empowerment, the government can foster a culture of transparency and accountability in AI governance. Moreover, non-profit data and AI trusts can play the mediator role and represent communities during the process of AI development in, with and for government [7].

3. Implementation and Adoption

The successful integration of the RECA framework within the UK government necessitates a multifaceted approach that encompasses various domains such as policy development, regulatory frameworks, technical standards, and capacity building. To ensure effective implementation, several key steps need to be undertaken.

a. Policy Integration

Instead of providing an isolated set of guidelines, it is necessary to seamlessly integrate the RECA principles into existing government policies, guidelines, and procurement processes. This involves not only drafting ethical guidelines but also instituting robust data protection regulations and implementing comprehensive impact assessment requirements for all AI projects supported by the UK government (under various UK Research and Innovation (UKRI) Councils and funding streams). By embedding these principles into policy frameworks, the government can help ensure that AI systems adhere to responsible use standards, thereby safeguarding citizen rights and interests.

b. Regulatory Frameworks

Developing regulatory frameworks tailored to govern the use of AI in government is essential. This entails establishing mechanisms aligned with international efforts [11] and national interests [5] for auditing, ensuring accountability, and providing avenues for litigation and remedy in cases of harm or discrimination. The creation of oversight principles [8], regulatory sandboxes, and certification programmes for AI systems can facilitate compliance with ethical standards while fostering innovation and responsible deployment.

c. Technical Standards

Establishing technical standards, in collaboration with national institutes such as the Institute of Engineering and Technology (IET) and the British Computer Society (BCS), is crucial for guiding the development and deployment of AI systems. This includes formulating guidelines for data privacy, promoting algorithmic transparency, and facilitating ethical decision-making processes. Implementing best practices, certification schemes, and interoperability standards will help ensure that AI technologies operate ethically, transparently, and in alignment with societal values. To that end, national networks such as Responsible AI UK (https://rai.ac.uk), the Royal Society (https://royalsociety.org) and the Alan Turing Institute (https://www.turing.ac.uk) can be seen as some potential hubs to link different stakeholders from academia, industry, and government departments.

d. Capacity Building

Investing in continuous training and capacity building initiatives is essential for enhancing the competencies of government officials, policymakers, and AI developers in implementing the RECA framework. Participatory workshops, seminars, and educational programmes can play a pivotal role in raising awareness about responsible AI practices and equipping stakeholders with the necessary skills to navigate ethical challenges effectively. Such skills can include AI basics for end users, technical skills for developers, managerial and procurement support and training for those involved in procuring AI systems. Moreover, efforts can be coordinated towards ensuring that postgraduate research institutes and Centres for Doctoral Training (CDTs) integrate responsible AI training in their programmes so that AI researchers are trained with these principles in mind, as they prepare for the job market.

e. Stakeholder Engagement

Fostering collaboration and engagement with diverse stakeholders is fundamental to the success of AI integration in government practices. This involves actively involving citizens, civil society organisations, academia, manufacturers, and industry stakeholders in the decision-making process and co-development of AI tools in, with and for government. Public consultations, stakeholder forums, and multistakeholder partnerships offer valuable platforms for co-designing AI solutions that reflect the diverse perspectives and priorities of society, thereby enhancing their relevance and effectiveness.

4. Conclusions

The adoption of the RECA framework presents a step towards the responsible use of citizen-centric AI systems in government. By prioritising citizen awareness, beneficence, sensitivity, and audibility, the UK government can minimise risks, enhance trust, and maximise the societal benefits of AI deployment. As technology continues to evolve, it is crucial that the UK government remains committed to responsible AI practices and actively engages with various stakeholders to ensure that AI systems serve the public interest and contribute to a fair and inclusive society. Through concerted efforts and collaboration, we can harness the transformative potential of AI to build a better future for all citizens.

References

[1] Benny Applebaum, Haakon Ringberg, Michael J. Freedman, Matthew Caesar, and Jennifer Rexford. 2010. Collaborative, Privacy-Preserving Data Aggregation at Scale. In *Privacy Enhancing Technologies*, 2010. Springer, Berlin, Heidelberg, 56–74. https://doi.org/10.1007/978-3-642-14527-8_4

[2] Frederik Auffenberg, Sebastian Stein, and Alex Rogers. A Personalised Thermal Comfort Model using a Bayesian Network.

[3] Richard Gomer, m.c. schraefel, and Enrico Gerding. 2014. Consenting agents: semi-autonomous interactions for ubiquitous consent. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (*UbiComp '14 Adjunct*), September 13, 2014. Association for Computing Machinery, New York, NY, USA, 653–658. https://doi.org/10.1145/2638728.2641682

[4] David Gunning, Mark Stefik, Jaesik Choi, Timothy Miller, Simone Stumpf, and Guang-Zhong Yang. 2019. XAI-Explainable artificial intelligence. *Sci. Robot.* 4, 37 (December 2019), eaay7120. https://doi.org/10.1126/scirobotics.aay7120

[5] HM Government. 2021. National AI Strategy. (2021). Retrieved May 1, 2024 from https://www.gov.uk/government/publications/national-ai-strategy

[6] Behrad Koohy, Jan Buermann, Sebastian Stein, Vahid Yazdanpanah, Enrico Gerding, Paul Pschierer-Barnfather, and Pamela Briggs. 2024. Adaptive incentive engineering in citizen-centric AI. In *The 23rd International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2024). May 6-10*, 2024. International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS), Auckland, New Zealand.

[7] Kieron O'hara. 2019. *Data Trusts: Ethics, Architecture and Governance for Trustworthy Data Stewardship*. University of Southampton. https://doi.org/10.5258/SOTON/WSI-WP001

[8] Daria Onitiu, Vahid Yazdanpanah, Adriane Chapman, Enrico Gerding, Stuart E. Middleton, and Jennifer Williams. 2023. On the legal aspects of responsible AI: adaptive change, human oversight, and societal outcomes. In *International Conference on AI for People: Democratizing AI (24/11/23 - 26/11/23)*, November 24, 2023. .

[9] Sarvapali D. Ramchurn, Sebastian Stein, and Nicholas R. Jennings. 2021. Trustworthy human-AI partnerships. *iScience* 24, 8 (August 2021). https://doi.org/10.1016/j.isci.2021.102891

[10] Sarvapali Ramchurn, Mohammad Reza Mousavi, Seyed Mohammad Hossein Toliyat, Mark Kleinman, Justyna Lisinska, Diego Sempreboni, Sebastian Stein, Enrico Gerding, Richard Gomer, and Francesco D'Amore. 2021. The future of connected and automated mobility in the UK: call for evidence. https://doi.org/10.5258/SOTON/P0097

[11] Responsible AI UK. 2024. *Responsible AI governance: A response to UN interim report on governing AI for humanity*. Public Policy, University of Southampton. https://doi.org/10.5258/SOTON/PP0057

[12] Sebastian Stein and Vahid Yazdanpanah. 2023. Citizen-centric multiagent systems. In *The 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023). May 29 - June 2*, May 30, 2023. International Foundation for Autonomous Agents and Multiagent Systems, London, United Kingdom, 1802–1807. Retrieved March 27, 2024 from https://eprints.soton.ac.uk/475810/

[13] The Royal Society. 2024. *The United Nations' role in international AI governance*. The Royal Society. Retrieved March 27, 2024 from https://royalsociety.org/-/media/policy/publications/2024/un-role-in-international-ai-governance.pdf

[14] Vahid Yazdanpanah, Enrico Gerding, Sebastian Stein, Mehdi Dastani, Catholijn M. Jonker, Timothy Norman, and Sarvapali Ramchurn. 2022. Reasoning About Responsibility in Autonomous Systems: Challenges and Opportunities. *AI Soc.* (November 2022).

May 2024