

Rebuilding and Enhancing Trust in Algorithms: Guidance for Policymakers

Algorithm-driven online systems are increasingly encountered in everyday life, making recommendations on what we might like, as well as critical decisions with significant life impacts in areas such as, job applications, case evaluations in the justice system, or students' examination results. A recent survey by the British Computer Society conducted in 2020 indicated that the majority of British adults do not trust algorithms.¹ The sense of distrust associated with these systems contributes not only to negative effects on people's online wellbeing, but also impedes the success of future innovations that may provide critical solutions to digital services of national importance. For example the UK Information Commissioner recently stated that "a lack of trust in the [NHSX contact tracing] app would have meant a lack of engagement with the app. And the benefits the service offered society would have been lost."²

The ReEnTrust project takes a human-centred approach to explore ways to rebuild trust in online algorithmic systems, drawing on 12 workshops, 300 responses to online surveys and 30 interviews, involving stakeholders across different sectors and user groups of different age groups (16-25 and 65+). Our study results offer policy recommendations for UK policy makers to enhance citizens' trust in digital services and platforms based in the UK through: 1) increasing systemic transparency; 2) engaging diverse user groups and considering a range of application contexts; and 3) increasing citizens' awareness of algorithms.

1. Increase systemic transparency

Explanations are an accepted element of existing policy guidelines. However, while explanations may be necessary, our research showed that they do not always provide sufficient transparency for users to trust the results they see. Users demand systemic transparency about how an algorithmic system came to its conclusions, the purpose of the system in an organisation and how the data will be used, and the underlying business model, with a balance between increased transparency and the protection of intellectual property rights and trade secrets.

Policy recommendations:

- The government should instruct appropriate bodies (e.g. ICO, CDEI) to produce detailed guidance on the types of systemic transparency required by citizen groups from specific sectors, similar to the transparency standard consultation set up for the Children's Code by the ICO.³
- Review the uptake of current guidance (such as Explaining Decisions Made with AI⁴) by stakeholders to identify barriers for adoptions.
- The government should set up specific efforts to support stakeholders, particularly SMEs, to gain support to assess their compliance of guidance, such as the one provided by the ICO for the compliance to the GDPR.⁵

2. Engage diverse user groups and consider a range of application contexts

There is no 'one-size-fits-all' approach to enhance trust. Trust in online systems is contextual and depends on many factors including the task to be completed and the relevance of the algorithmic decision to the user. Different age groups approach trust in different ways, with older people more likely to place their trust in established institutions that they are familiar with, while both young and old tend to expect that websites should behave in a trustworthy way.

Policy recommendations:

- Encourage innovators to develop co-creation approaches and engage broadly with a diverse user base during design and evaluation. This could take the form of citizen panels linked to innovation hubs or council provided co-creation tools/training.
- Establish a cross-disciplinary expert panel to assess how much different population groups are affected by distrust of algorithmic systems.
- Raise innovators' awareness and catalyse more responsible innovation.

3. Increase citizens' awareness of algorithms

Our research has shown that users largely have limited awareness of how algorithms are deeply embedded in our everyday life, especially for older citizens. Citizens need to be able to recognise the involvement of algorithms in digital services as a necessary first step to empower citizens to critically engage with these systems. Although sites are required to provide information about security and the use of cookies, for example, this does not provide support for users to judge whether an algorithm is fair or accurate.

Policy recommendations:

- Promote citizen awareness of the UK's algorithmic governance and strategy frameworks, as well as their related rights, by running a public awareness campaign, such as including algorithmic literacy as part of annual Safer Internet Day events.
- Develop and promote government endorsed algorithmic literacy programmes for the needs of different age groups, for example, expanding the inclusion of algorithmic literacy in existing digital literacy programmes, or developing public programmes like the Finland Toolbox.⁶
- Commission the development of means to communicate the use of recognised truly trustworthy innovation, such as trustmarks and standards/guidance from standards bodies like the BSI,⁷ potentially in coordination with similar efforts by the OECD⁸ and WEF.⁹

DISCLAIMER AND COPYRIGHT

This document is based on research outcomes from the Engineering and Physical Sciences Research Council (EPSRC) project "ReEnTrust: Rebuilding and Enhancing Trust in Algorithms" (EP/R033633/1; 2018-2021) and "From Human Data to Personal Experience" (EP/M02315X/1; 2015-2021). The content of the document is the sole responsibility of its author(s) and any opinions expressed herein should not be taken to represent an official position of their institutions. Reproduction and translation for non-commercial purposes are authorised, provided the source is acknowledged. For any further details please contact Prof. Marina Jirotko (marina.jirotko@cs.ox.ac.uk).