University of Nottingham
Institute for Policy and Engagement

Policy Briefs:
Rebuilding and enhancing trust in algorithms

Research by ReEnTrust

Recommendations

Algorithms that have significant impact on people of society must be certified by a regulator or accredited third party.

Younger and older demographic groups may have different trust needs, meaning that there is no one-size fits all solution- this must be taken into account when building mechanisms for trust in algorithms.

Algorithms must be transparent and easily explained but give careful consideration to match the style of explanation to the intended recipients.

Users and other stakeholders must be involved in the decision making. Application of "human-in-the-loop" approach could engender trust and true buy in from users.

Policy Context

The public encounter algorithmic systems **indirectly** through applications such as predictive policing, sentencing and parole and medical decisions and diagnosis, as well as through credit scoring and financial decisions

The public encounter algorithmic systems **directly** through e-commerce services such as automated customer service assistants, hotel and holiday bookings, recommender systems and job recruitment.

This means that corporations and businesses must ensure the public have trust in these algorithms in order to benefit from the marketing and economic potential they possess.

This research encourages policy makers and regulators to support processes and development of algorithms that engender trust. The centrality of trust dynamics for the responsible growth of AI in society is reflected in a wide range of policy developments, including the Online Harms whitepaper, the work of the European Commission High-Level Expert Group on AI, and the UNESCO COMEST working group on Ethics of AI to name just a few.

The Research

The **ReEnTrust Project** explores new technological opportunities for platforms to regain user trust and aims to identify how this may be achieved in ways that are user-driven and responsible. Focusing on AI algorithms and large-scale platforms used by the general public, research questions include:

- What are user expectations and requirements regarding the rebuilding of trust in algorithmic systems once that trust has been lost?
- Is it possible to create technological solutions that rebuild trust by embedding values in recommendation, prediction, and information filtering algorithms, and by allowing for a productive debate on algorithm design between all stakeholders?
- To what extent can user trust be regained through technological solutions? What further trust rebuilding mechanisms might be necessary and appropriate, including policy, regulation, and education?

ReEnTrust's work includes quantification of trust as a metric. It has and will continue to engage with stakeholder workshops, comparing demographic groups' trust and use of algorithms, using surveys and sandbox experiments to compare different algorithms and use of a mediation tool to address issues of trust.

ReEnTrust has engaged with related projects and builds on its predecessor UnBias:

- The development of an IEEE Standard for Algorithmic Bias Considerations (P7003)
- The IEEE Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS)
- And for decision makers at organizations that develop or procure AI systems, in collaboration with EY and Proboscis[1]

Conclusions

Trust includes confidence that the other party will deliver, belief in the intentions of the other party and the fairness and reasonableness of the results.

Online interaction with algorithm-driven platforms is different from normal social trust. Online interaction is disembodied and distant, the action of the platform is underlaid by algorithms, users are trusting the algorithm as well as the organisations/website, and the algorithm is opaque.

Trust must be engendered in order to maximise the benefits that algorithms provide. Failure to produce a trustworthy platform will damage trust in the whole organisation.

ReEnTrust

ReEnTrust is an EPSRC[2] research project funded under the Trust, Identity, Privacy and Security in the Digital Economy 2.0 call. A collaborative project between the University of Oxford, University of Edinburgh and the University of Nottingham.

Contact the researchers

Dr Philip Inglesant

Research assistant in Human Centred Computing at the University of Oxford

Email:
philip.inglesant@cs.ox.ac.uk

Visit:
https://www.cs.ox.ac.uk/people/philip.inglesant/

Dr Ansgar Koene

Senior researcher fellow at the faculty of science at the University of Nottingham. Research co-investigator at Horizon Digital Economy Research.

Email:
Ansgar.Koene@nottingham.ac.uk

Visit:
https://www.nottingham.ac.uk/computerscience/people/ansgar.koene

Dr Jun Zhao

Senior researcher fellow in human centred computing at the University of Oxford.

Email:
jun.zhao@cs.ox.ac.uk

Visit:
https://www.cs.ox.ac.uk/people/jun.zhao/

[1] http://proboscis.org.uk/projects/ongoing/unbias/

[2] Engineering and Physical Sciences Research Council