Rebuilding and Enhancing Trust in Algorithms: Policy Recommendations

DR. ANSGAR KOENE,
SENIOR RESEARCH FELLOW, HORIZON DER, UNIVERSITY OF NOTTINGHAM
GLOBAL AI ETHICS AND REGULATORY LEADER, EY
3RD JUNE 2021

The ReEnTrust project

ReEnTrust was an EPSRC research project (EP/R033633/1) funded under the Trust, Identity, Privacy and Security in the Digital Economy 2.0 (TIPS2) call.

The project was a collaboration between researchers from:

- Human Centred Computing Group at the, University of Oxford
- The Centre for Intelligent Systems and their Applications (CISA) at the University of Edinburgh
- HORIZON Digital Economy Research at the University of Nottingham

Project leader: Prof. Marina Jirotka

Project website: https://reentrust.org

The main objectives

Technologies for rebuilding trust

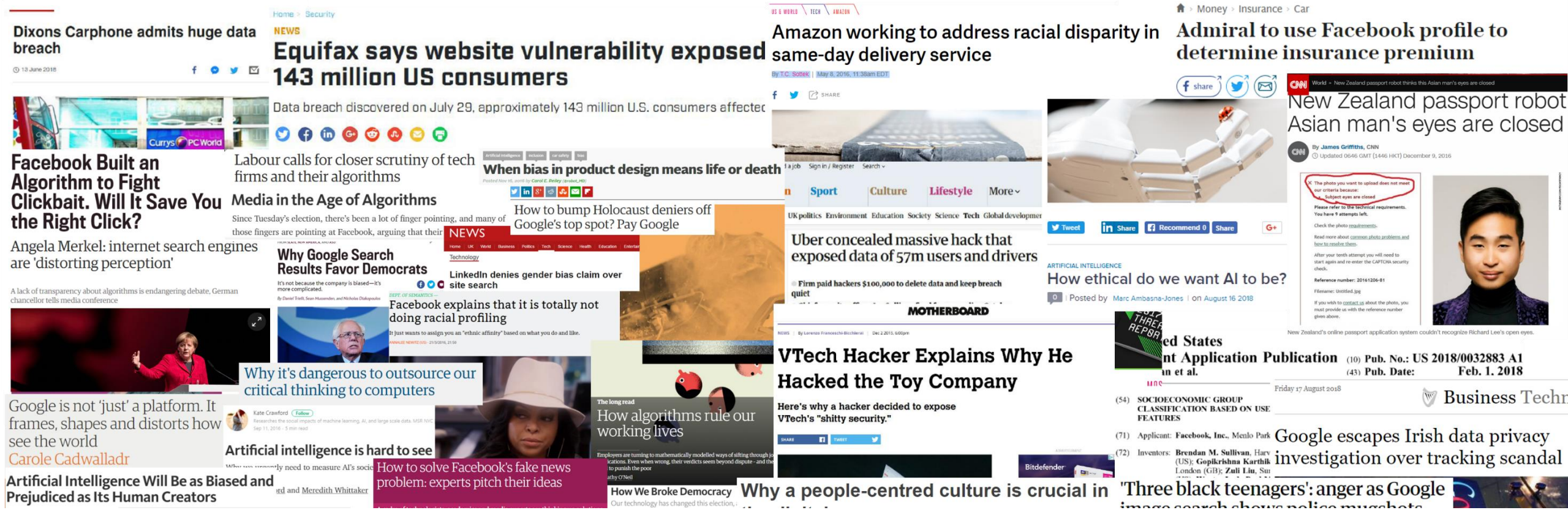
Design with partners and general public

Contribute to policy and regulations

Advance state of the art about AI and trust

Algorithmic systems and automated processes are often used for decision-making on online platforms and across our lives

There are significant tensions in user perspectives regarding how these algorithms are used on the Web.



This results in a breakdown of trust: Users do not know when to trust the outcomes of algorithmic processes and, consequently, the platforms that use them.

ReEnTrust explores new technological opportunities for platforms to rebuild user trust, in ways that are user-driven and responsible.

Workpackage 1: Responsible policy and practice

- Policy guidelines, cases studies, and engaging stakeholders in considering trust in the design, development, and use of algorithms.

Workpackage 2: User-centred trust

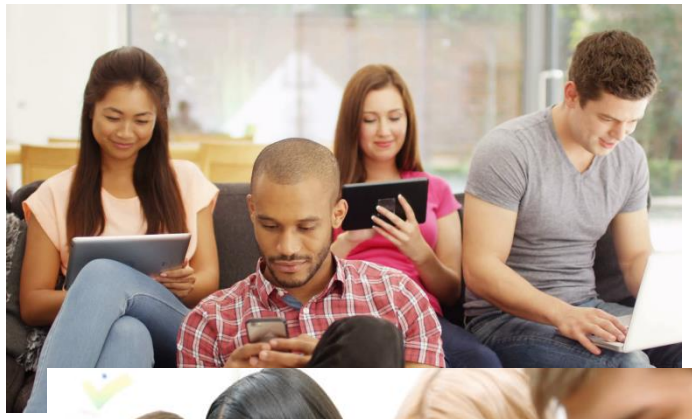
- Identifying the most important issues that affect trust in users' online service interactions, and how this affects wellbeing

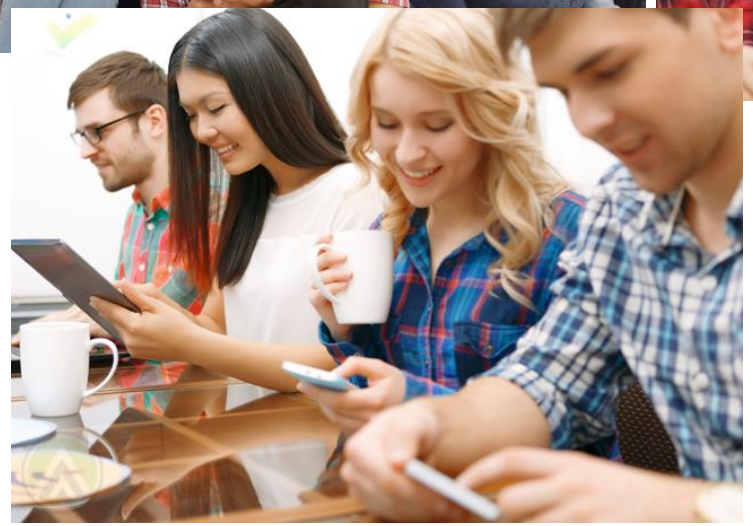
Workpackages 3/4: Computational methods for rebuilding trust

- Designing tools for online mediation and trust rebuilding

Two user groups: older adults aged 65 and over, and younger adults aged 16 to 25.

- Investigate differences in attitude and experience according to age





Trust factors

Competence or ability of the trustee to honour commitments

- correctness or accuracy

Honest intentions to act with full attention and without cheating

- intelligibility or transparency
- Fairness

Trust appeals to altruism and social norms

High-trust environments reduce the costs of intra- and inter-organisational transactions

The action of trusting requires some level of choice or free will.

- without choice, or with total control trust becomes irrelevant

Trust relates to (hypothetical) future events, and hence always involves risk.
A truster is in some kind of vulnerable position or has risk of loss in the interaction with the trustee.

Policy recommendations (1/3)

Increase systemic transparency

Explanations are accepted elements of policy guidelines. However, they do not always provide sufficient transparency for users to trust the results they see. Users demand systemic transparency:

- how an algorithmic system came to its conclusions,
- the purpose of the system in an organisation,
- how the data will be used,
- the underlying business model

- ➢ Detailed guidance on systemic transparency required by citizen groups from specific sectors
- ➢ Review of uptake of current guidance, e.g. ICO-ATI's guidance on Explaining AI decisions
- ➢ Support for stakeholders, particularly SMEs for achieving compliance with guidance

Policy recommendations (2/3)

Engage diverse user groups and consider application context

Trust in online systems is contextual. It 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:

- older people more likely to place their trust in established institutions that they are familiar with,
- both young and old tend to expect that websites should behave in a trustworthy way.

- Encourage co-creation approaches with diverse user base: Citizen panels linked to innovation hubs; council provided co-creation tools/training
- Cross-disciplinary expert panel to assess algorithmic system distrust among different population groups
- Promote uptake of “responsible innovation” frameworks, e.g. RRI AREA (Anticipate-Reflect-Engage-Act), BSI PAS440

Policy recommendations (3/3)

Increase citizen awareness of algorithms

Users largely have limited awareness of how algorithms are deeply embedded in everyday life, especially for older citizens.

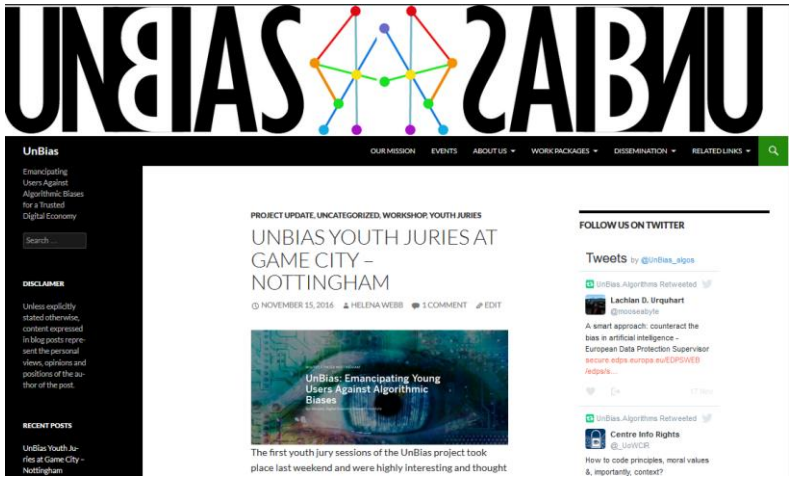
Citizens need to be able to recognise the involvement of algorithms in digital services to empower critical engagement with these systems.

- Promote awareness through campaigns such as including algorithmic literacy as part of Safer Internet Day events
- Develop and endorse algorithmic literacy programmes tailored to the needs of different groups
- Commission means to communicate trustworthy innovation by engaging with BSI and coordinating with OECD and WEF

Thank you

REENTRUST
Rebuilding and Enhancing Trust in Algorithms

http://reentrust.org/



https://unbias.wp.horizon.ac.uk/