Joint Paper

April 2020

Algorithm Inspection and Regulatory Access

Coordinated by Digital Action

Algorithms play a central role in social media platforms – and may contribute to systemic, structural challenges to democracy and human rights. As identified in the White Paper, through their decisions around the design of algorithms, the platforms have a significant impact on citizens' rights to information and free expression, our rights not to be discriminated against or abused, and core democratic values. As such, there is a need to establish new systems for transparency and oversight or auditing of algorithm design and algorithmic decision-making.

Options for algorithm inspection

In order to give effect to the stated aims of the Online Harms White Paper and to ensure true algorithm transparency and accountability, transparency must be improved for researchers and users, but ultimately any regulator should have powers to undertake algorithm inspections themselves. In this paper we are not endorsing any particular institutional arrangement for a regulator. Our goal here is to strengthen the case for the Government to establish the means and authority for algorithm accountability, by showing that there are a number of potential ways this could be achieved.

1. Information-gathering powers

The Online Harms regulator is already proposed to "have powers to require additional information, including about the impact of algorithms." Additionally, "the regulator will have the power to request explanations about the way algorithms operate." These information gathering powers should be considered the minimum any regulator in this space would need to be able to fulfil their functions.

2. Compulsory audit and inspection powers

Any regulator will need to go further than merely requesting information and instead need to develop the means to test the operation of algorithms and undertake inspections themselves.

A parallel could be drawn to the Information Commissioner's Office (ICO) who can undertake consensual audits to assess how data controllers or processors are complying with good practice in the processing of

personal data.[1] Should the company not agree to a consensual audit, the ICO can (should they decide that there are reasonable grounds for suspecting a data controller or processor is failing to comply with the Data Protection Act) seek a warrant to enter, search, inspect, examine and operate any equipment in order to determine whether a company is complying with the act.[2]

Alternatively, a model could be drawn from that used by the Investigatory Powers Commissioners Office (IPCO) who are responsible for keeping under review the use of investigatory powers by a number of public authorities including the security and intelligence agencies and law enforcement bodies. IPCO has powers[3] to conduct investigations, inspections, and audits as the Commissioner considers appropriate for the purpose of the Commissioner's functions, including access to apparatus, systems or other facilities or services.[4] In practice, this means IPCO are able to inspect on site the entire system used by the body they are auditing, including the underlying data, any technologies processing the data, and the output provided.

Any regulator will need a similar ability to carry out an algorithm inspection with the consent of the company; or if the company doesn't provide consent, and there are reasonable grounds to suspect they are failing to comply with the duty of care, to use compulsory audit powers to determine whether they are.

Such an ability is envisaged to some degree within the White Paper, which states that the regulator could, "require companies to demonstrate how algorithms select content for children, and to provide the means for testing the operation of these algorithms." Further detail would of course be needed on this.

3. Independent expert third party audit powers

Rather than the regulator themselves undertaking an audit or inspection, they could play a role instructing independent experts to undertake an audit on their behalf. This would help ensure that the correct expertise is acquired for the work as is needed, rather than the regulator needing to hold what might be a vast range of subject-matter expertise inhouse.

A model for this could be the Financial Conduct Authority's power to require reports from third parties; what they dub 'skilled persons reviews'.[5] It could be argued it is inherent in the regulators' powers to be able to delegate tasks to such third parties. However, it would be wise to include plain words to that effect, given the potential crucial role independent third-party auditors could play.

4. Access by academia to conduct research in the public interest

Finally, as recommended by the Centre for Data Ethics and Innovation,[6] academics need to have access to conduct research in the public interest. Existing efforts in this area are already underway, such as those between academia and the private sector to allow external researchers to analyse information amassed by companies to address societal issues. [7] These efforts have been challenging to set up, have yet to prove themselves, and are limited in scope.

So it is positive that the online harm regulator will adopt a role to "encourage and oversee the fulfilment of companies' commitments to improve the ability of independent researchers to access their data", but such language must go beyond mere encouragement.

[1] s129, Part 5, Data Protection Act 2018
[2] Schedule 15, Data Protection Act 2018
[3] s235(1), Chapter 1, Part 8, Investigatory Powers Act 2016
[4] s235(4), Chapter 1, Part 8, Investigatory Powers Act 2016
[5] https://www.fca.org.uk/about/supervision/skilled-persons-reviews
[6] www.gov.uk/government/publications/cdei-review-of-online-targeting
[7] https://socialscience.one/

The scheme should offer a mechanism for suitably-qualified and accredited academics to apply to access data for research, whether their research questions are within the current field of view of the regulator or not, so long as they pertain to content regulation and controls. This would balance concerns that the regulator may not always identify the correct problem set and allow academics to establish evidence relating to issues that are emerging or have not received sufficient care and attention.

This has happened to a degree with the (since restructured) Interception of Communications Commissioners Office who worked with an independent academic to assist in a particularly complex inquiry into the use of the Regulation of Investigatory Powers Act (RIPA) to identify journalistic sources. [8]

Essential characteristics of any algorithm inspection arrangement

Whichever method is to be used, the regulator should be able to:

- Examine the purpose, constitution, and policies of the systems, and to interview people who build and interact with different parts of that system, and observe how people use the system.
- Identify and assess what data was used to train the algorithm, how it was collected, and whether it is enriched with other data sources, and whether that data changed over time.
- Examine the model itself including considering the processing flow and the type of supervisory or monitoring mechanism used.
- Undertake a code review or "white-box testing" to analyse the source code, or the statistical models in use, including how different inputs are weighted.
- Maintain full compliance with GDPR and equivalent post-Brexit data protection laws.

The regulator should also be able to run controlled experiments over time and within ethical guidelines to determine if the algorithms subject to their review are producing unintended consequences that harm the public interest. Such experiments would be novel in this area, but such independent testing and experimentation are of course commonplace in other areas such as pharmaceuticals or food safety.

It is only by undertaking this sort of audit that a regulator, acting in the public interest, will be able to assess whether companies truly are acting responsibly, protecting the safety of their users and tackling how their platforms contribute to structural issues, such as impacting on an open and fair democracy.

In order to achieve this, the staffing of an online harms regulator would need to include a number of technologists who are able to advise to develop policies and procedures on how such an audit should take place as well as undertaking the audit

This work should be done openly and transparently, with continuing consultation with civil society, industry and academia to jointly pool expertise to establish best practice in this new area. It is envisaged that processes could be co-constructed with external stakeholders to ensure protection of fundamental rights such as privacy and free expression. Equally, it is envisaged that the approach would be continuously developing, applied proportionately, and responsive to the changing technological landscape.

[8] See Interception of Communications Commissioner's report into the use of Chapter 2 of Part 1 of the Regulation of Investigatory Powers Act to identify journalistic sources which was assisted by Professor Anne Flanagan, Professor of Communications Law at Queen Mary University of London. https://www.ipco.org.uk/docs/iocco/IOCCO%20Communications%20Data%20Journalist%20Inquiry%20Report%204Feb15.pdf

Drafted with assistance from AWO

Coordinated by Digital Action
For more information contact:
Nick Martlew
Digital Action

info@digitalaction.co
www.digitalaction.co