BRIEFING

STOA Options Brief

A governance framework for algorithmic accountability and transparency

Algorithmic systems are increasingly used as part of decision-making processes in the public and private sectors, with potentially significant consequences for individuals, organisations and societies. However, the very properties of scale, capability to handle complex datasets, and autonomous learning that make these systems useful also make it difficult to provide clear explanations for the decisions they make. This lack of transparency risks undermining meaningful scrutiny and accountability, which is a significant concern when relating to decision-making processes that can have a considerable impact on fundamental human rights.

On the basis of a review of existing proposals for the governance of algorithmic systems, the study offers four sets of policy options, each addressing a different aspect of algorithmic transparency and accountability: i) awareness raising – education, journalism and whistleblowers; ii) accountability in public sector use of algorithmic systems; iii) regulatory oversight and legal liability; and iv) global coordination of algorithmic governance.

Awareness raising – education, journalism, whistleblowers

The general public are struggling to understand how algorithmic systems work, the impact they are having, and how to make a critical evaluation of their decisions. The same is true of many highly skilled non-technical professionals, such as judges and lawyers. A broad understanding of algorithmic systems, however, will do little to provide accountability unless there is a public debate about the types and properties of the algorithmic systems associated with the decisions concerned. Notifications should be standardised and short, akin to nutrition labels, while information should be limited to that which can impact a user's decisions, or wider public understanding.

Investigative journalism and whistleblowers play an important role in uncovering questionable uses and outcomes of algorithmic decision-making (e.g. Cambridge Analytica election manipulation). Whistleblowing by (ex-)employees is an important part of activism aimed at changing unethical company projects (e.g. Google Dragonfly).

Beyond their role as independent watchdogs, journalists help to present relevant aspects of algorithms in plain language with understandable narratives. Journalistic investigations have sparked broad public conversations and important normative debates, including triggering new academic studies (e.g. Propublica's report on 'Machine Bias' in the COMPAS algorithm triggered a series of studies into the meaning of 'fair' in algorithms).

To uncover cases of algorithmic 'malpractice' journalists are combining traditional investigation practices, with computationally intensive methods to reverse engineer algorithms (e.g. 'black box testing') so as to tease out the consequences of algorithmic system use. Reverse engineering can however involve violating of trade secrets and/or copyright rules.

Policy options

- Provision of 'algorithmic literacy' to teach core concepts of algorithmic selection/decisions
- Standardised mandatory notification to communicate algorithmic processing in decisions
- Provision of technical support for 'algorithmic accountability journalism'
- Whistleblower protection and protection against prosecution on grounds of breaching copyright or terms of service when doing so serves the public interest

Accountability in public sector use of algorithmic systems

Algorithmic systems are increasingly being used by public authorities to improve efficiency, implement complex processes and support evidence-based policy making. These uses can have far reaching impacts involving the weakest members of society and therefore require extra levels of transparency and accountability. Public sector procurement is a major source of business for many companies. Algorithmic impact assessments (AIAs) by the public sector for procurement purposes could therefore encourage the commercial development of transparent and accountable systems.

AIA is designed to help policymakers and their constituents understand where algorithmic systems are used within government, assess their intended use and proposed implementation, and allow community members and researchers to raise concerns. AIA draws on assessments in areas such as environmental policy, privacy law, and data protection. The framework requires public authorities to make their own assessments of the algorithmic systems they intend to use and will likely require additional information from vendors. In practice the exact AIA will depend on the context and the sensitivities of the public sector branch.

Policy options

Algorithmic impact assessment (AIA) for public sector uses of algorithmic systems consisting of:

- publication of the public authority's definition of an 'algorithmic system'
- public disclosure of the purpose, scope, intended use, associated policies/practices, self-assessment timeline/process, and potential implementation timeline of the algorithmic system
- performance and publication of the self-assessment of the system with a focus on inaccuracies, bias, harm to affected communities, and a description of mitigation plans for potential impacts
- publication of plan for meaningful, ongoing access to external researchers to review the system
- public participation period and publication of final AIA, once issues raised in public participation have been addressed
- renewal of AIAs on a regular basis
- opportunity for public to challenge failure to address issues raised in the public participation

Regulatory oversight and legal liability

The development and application of algorithmic systems is undergoing rapid growth with uncertain implications for citizens and society. Industry standards for best practice are all but non-existent. The interpretation of existing laws is sometimes uncertain when applied to algorithmic outcomes, and judicial experience is in short supply. While much of this is the result of rapid dynamic growth, it must not limit the rights and legal protection of citizens.

While requiring algorithmic impact assessments (AIA), similar to public sector uses of algorithmic systems, makes sense for high-impact commercial applications (e.g. for autonomous vehicles or political election-related services) for most private-sector applications the financial and administrative burden of AIAs would not be proportionate to the risks. For low-risk uses of algorithmic systems, it would be preferable to establish a legal liability framework to allow service providers to accept greater tort liability in exchange for reduced requirements vis à vis transparency and AIAs.

To facilitate a tiered regulatory regime of this kind, it would be necessary to establish a regulatory body with expertise in analysing algorithmic systems and a network of external expert advisors.

Policy options

- AIA requirement for commercial systems classified as causing a potentially severe non-reversible impact, similar to public sector applications
- Systems with medium severity impacts requiring providers to accept strict tort liability, with a possibility of reduced liability for systems certified as complying with best-practice standards
- Creation of a regulatory body for algorithmic systems tasked with:
  - conducting a risk assessment to classify algorithm types and application domains by impact on citizens
  - investigating the use of algorithmic systems where there is suspected infringement of human rights (e.g. evidence provided by a whistleblower)
  - advising other regulatory bodies regarding algorithmic systems
  - coordinating with standard-setting organisations, industry and civil society to identify relevant standards and best practices to use for third-party certification
  - auditing the AIAs of systems requiring high-level oversight, such as those used in highly sensitive and/or safety-critical application domains (e.g. private healthcare)
  - facilitating a tort liability mechanism to regulate the accountability of algorithmic systems by providing a contact point for citizens not familiar with legal procedures

Global coordination for algorithmic governance

As with much of the digital economy, services that use algorithmic systems are characterised by a high degree of cross-border and global reach. To govern algorithmic systems successfully therefore requires global dialogue and collaboration across borders. Without multilateral negotiation regulatory intervention for transparency and accountability risks being interpreted as protectionism or as an attempt to gain access to foreign trade secrets.

In order to respond effectively to the use of algorithmic systems to interfere in the informational integrity of national elections or to perform offensive cyber operations it is important to have a broad international community involved in establishing guidelines on the attribution of such attacks and the definition of proportionate responses.

Policy options

- Establish a permanent global algorithm governance forum (AGF) for multi-stakeholder dialogue and policy expertise related to algorithmic systems and associated technologies, to provide a forum for coordination and the exchange of governance best practices
- In international (trade) negotiations, protect ability to investigate algorithmic systems and hold parties accountable for violations of European laws and human rights

MAIN REFERENCES

Reisman D., Schultz J., Crawford K. and Whittaker M., 'Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability', *AI Now*, April 2018.

Erdelyi O. J. and Goldsmith J., Regulating Artificial Intelligence: Proposal for a Global Solution, 2018 AAAI/ACM Conference on AI, Ethics, and Society (AIES '18), February 2-3, 2018, New Orleans, LA, USA.

This document is based on the STOA study on 'A governance framework for algorithmic accountability and transparency' (PE 624.262) published in April 2019. The study was written by Ansgar Koene, main author/editor, University of Nottingham, at the request of the Panel for the Future or Science and Technology (STOA) and managed by the Scientific Foresight Unit within the Directorate-General for Parliamentary Research Services (DG EPRS) of the European Parliament. STOA administrator responsible: Mihalis Kritikos.

DISCLAIMER AND COPYRIGHT

stoa@ep.europa.eu (contact)

http://www.europarl.europa.eu/stoa/ (STOA website)

www.europarl.europa.eu/thinktank (internet)

http://epthinktank.eu (blog)