BRIEFING

STOA Options Brief

Understanding algorithmic decision-making: Opportunities and challenges

While algorithms are hardly a recent invention, they are nevertheless increasingly involved in systems used to support decision-making in society. These ADS (algorithmic decision systems) often rely on the analysis of large amounts of personal data to infer correlations or, more generally, to derive information deemed useful to make decisions. Human intervention in the decision-making may vary, and may even be completely out of the loop in entirely automated systems. In many situations, the impact of the decision on people can be significant, such as access to credit, employment, medical treatment, judicial sentences, among other things. Entrusting ADS to make or to influence such decisions raises a variety of ethical, political, legal, or technical issues,where great care must be taken to analyse and address them correctly. If they are neglected, the expected benefits of these systems may be negated by the variety of risks for individuals (discrimination, unfair practices, loss of autonomy, etc.), the economy (unfair practices, limited access to markets, etc.) and society as a whole (manipulation, threat to democracy, etc.). The purpose of this policy options briefing is to highlight the main challenges and suggested policy options to allow society to benefit from the tremendous possibilities of ADS while limiting the risks related to their use.

The Science and Technology Options Assessment (STOA) study on algorithms in decision-making analyses the benefits and risks related to the use of ADS respect to three categories of stakeholders: individuals, the private sector and the public sector. As far as individuals are concerned, ADS may undermine the fundamental principles of equality, privacy, dignity, autonomy and free will. They may also pose risks related to privacy, health, quality of life and even physical integrity. The opacity or lack of transparency of ADS is another primary source of risk for individuals, opening the door to all kinds of manipulation and making it difficult to challenge a decision based on the result of an ADS. ADS also create new security vulnerabilities in public services that can be exploited by people or organisations with malicious intent. Since ADS play a pivotal role in the workings of society, for example in nuclear power stations, smart grids, hospitals and cars, hackers who are able to compromise these systems can cause major damage. The opportunities presented by ADS for the private sector are endless, but there are also numerous risks. Any task that is repetitive, pressured by time, or that could benefit from the analysis of high volumes of data, is a prime target for ADS. Such tasks concern low-skilled as well as highly-skilled personnel, for example in sectors such as banking, insurance or justice. Certain types of jobs will change enormously, or be eliminated, while new ones will appear. The expression 'fourth industrial revolution' describes this dramatic societal change.

Some key requirements must be met in order to reduce the risks related to ADS. The STOA study analyses both intrinsic, such as fairness, absence of bias or non-discrimination, and extrinsic requirements, such as transparency, explainability and accountability. Transparency is defined as the availability of the ADS code with its design documentation, parameters and the learning dataset when the ADS relies on machine learning. Explainability is the availability of explanations about the ADS – in contrast to transparency, explainability requires the delivery of information beyond the ADS itself. Accountability can be seen as an overarching principle characterised by the obligation to justify one's actions and the risk of sanctions if justifications are inadequate.

The STOA study presents some of the many challenges to be addressed to reduce the risks related to ADS, classified according to the following three perspectives: (1) ethical and political, (2) legal and social, and (3) technical challenges.

Ethical and political: ADS exacerbate existing problems or force us to rethink issues such as discrimination, but also introduce new ethical questions that are very difficult to address. For example, the use of some ADS, such as evidence-based sentencing or autonomous weapons, have recently been heavily criticised. However, deciding which ADS are acceptable and which ones should be banned is far from straightforward. Fairness is another challenging issue: beyond existing criteria already identified in anti-discrimination laws, what types of treatment should be considered undesirable? Where should the line be drawn and under which principles? The risk of manipulation is difficult to address: how can online manipulation be characterised and be distinguished from (acceptable) influence or 'nudging'? In which cases should transparency, explainability or other forms of accountability be required and in relation to which underlying principles? Should certain types of ADS be forbidden when an acceptable level of transparency, explainability or accountability cannot be achieved (for example in court, or to support medical diagnosis)? All these questions deserve to be discussed and answered before the deployment of any ADS.

Legal and social: Before rushing to legislate on this matter, it is necessary to organize ethical and political debates." Assuming that a broad agreement has been reached on some of the issues discussed above, the next step is to decide what are the most appropriate instruments to implement that agreement The STOA study analyses different types of regulation (state regulation, self-regulation or co-regulation, hard law or soft law, general or sectorial regulation), and different modes of enforcement (regulatory agencies, dedicated oversight bodies, etc.).

Technical: Existing technical instruments are useful to meet the above requirements on ADS, but are still in their infancy, with a number of challenges that need to be addressed. Some of these challenges are 'conceptual', such as defining the best types of explanations depending on the different recipients, their level of expertise and objectives. Other challenges are 'operational', such as the implementation of explainability by design, fairness by design or privacy by design. These properties should be taken into consideration from the beginning of the conception of an ADS, as already required by the EU General Data Protection Regulation (GDPR). However, this phase requires a strong level of technical expertise that cannot be expected from all ADS developers. Providing guidance and assistance to designers and developers to help them implement these principles remains an open challenge.

To address these challenges, seven policy options can be drawn from the STOA study. Each policy option addresses one of the above challenges, but most of them are also mutually reinforcing, showing that more effective results can be obtained if they are implemented together.

Policy option 1: Incentivise interdisciplinary research on responsible ADS

ADS raise complex questions that are not entirely understood by experts, not to mention users or the people affected by them. It is therefore of prime importance to develop interdisciplinary research in ADS. More research is needed, for example, on ADS security, safety, privacy, fairness or explainability. In addition, philosophers, experts in ethics, social scientists, lawyers, computer scientists and AI experts should work together to develop further conceptual tools to analyse ethical issues raised by ADS. Both long-term and mid-term research projects should be funded, for example using existing mechanisms such as the framework programme that will succeed Horizon 2020, or the European Research Council.

A key condition to facilitate interdisciplinary research into ADS is the possibility to provide the research community with access, under specific conditions and the strictest confidentiality, to datasets held not only by public entities but also by private companies. This access right is justified by the fact that such large amounts of data can be considered 'data of public interest'. For the same reason, it should be made clear that reverse engineering for the purpose of analysing, explaining or detecting biases in ADS should be considered lawful and should not be limited by trade secret or more generally by intellectual property rights laws.

Policy option 2: Incentivise training to enhance expertise in responsible ADS

Training in responsible ADS should be developed. In particular, but not exclusively, all engineering schools should develop a curriculum to educate all students about ethics and existing approaches to address the ethical aspects of ADS. In addition to building a broader common knowledge in responsible ADS, the creation of a body of experts in responsible ADS is to be encouraged. Such experts are necessary to implement other measures, such as policy options 4, 5 and 6 below.

Policy option 3: Create a European ethical committee on ADS

Considering that ADS can have a major impact on society, they must be subject to public debate. Several conditions have to be met to ensure the quality of this debate. It must involve all stakeholders, opinions and interests, which means experts of all disciplines, policy-makers, professionals, NGOs and the general public. The debate must be conducted in a rigorous fashion and without overshadowing any of the key issues, including the preliminary question of the legitimacy of the use of an ADS in the context being examined. The role of a European ADS ethical committee would be to stimulate discussion, to organise public debates and to publish recommendations. European institutions could consult the committee, which would also decide on its own agenda. In addition, the reflections of such a European ethical committee would pave the way for future specific or general regulations on ADS (policy options 6 and 7)

Policy option 4: Create a framework for algorithmic impact assessments in Europe

The EU GDPR introduces an obligation for data controllers to conduct data protection impact assessments (DPIA) and encourages certification mechanisms. Considering the high stakes involved in ADS, there is no reason why they should not be subject to the same types of precautions. The STOA study recommends in particular that ADS should not be deployed without a prior algorithmic impact assessment (AIA) unless it is clear they have no significant impact on individuals lives. Conducting an AIA is not an easy task and models and tools should be proposed to make it easier. The STOA study presents some key issues which should be considered in an AIA: (1) the ADS' legitimacy, including legitimacy of purpose, techniques and parameters; (2) the ADS' qualities; and (3) integration of the ADS within the human environment. It should also be clear that AIA should not only focus on the risks of using an ADS – they should also assess the risks of not using an ADS. In other words, AIA should consider both benefits and risks.

Policy option 5: Support innovation on new accountability and auditability tools for ADS

Most ADS designers and developers are not experts in privacy, security, fairness or explainability. It is therefore important to develop tools and methodologies to help them reconcile the tensions that exist between accuracy, cost and explainability/fairness/privacy. Recommendation guides are not enough. Tools, methodologies and training that consider the entire development cycle of ADS should be developed and disseminated. Similarly, frameworks composed of metrics, methodologies and auditability tools that assess the impact of an ADS and test its desired properties should be developed. These frameworks could be used by designers to test their ADS, and by third-party entities, such as certification authorities, to validate them. As far as users are concerned, better explanation facilities are required, in particular, more interactive interfaces and dialogue models. In order to favour the development of these new tools and methodologies, innovation should be strongly supported on this topic, in particular to allow innovative SMEs to launch their products and services on the market.

Policy option 6: Adopt new sector specific regulations to enhance ADS accountability

The use of ADS in certain sectors can be considered as sensitive and should be subject to stringent accountability measures. These sectors, which could be defined more comprehensively by a European ethical committee (policy option 3) should include, for example, justice, police and healthcare. Certification agencies and oversight agencies should together provide a framework for the monitoring, certification and oversight of ADS in these sectors. Oversight agencies should also have the power to sanction operators of non-compliant ADS. Certifications and labels, if properly implemented, can be a way to enhance trust in ADS and to verify that they comply with certain rules (such as the absence of bias or discrimination). Certification requirements and obligations should be sectoral, because the needs and the risks vary greatly from one type of application to another. Therefore, sectoral supervisory authorities or agencies are in a better position to define reference evaluation criteria and to control their application. In other sectors, such as administration, media, and e-commerce, for which certification may not be required, ADS operators should still be subject to accountability requirements. In particular, the invocation of trade secrets to oppose audits by independent third parties (NGOs, journalists, experts, or the administration itself) should be forbidden.

Policy option 7: Prepare future general regulation on ADS

It is too early to adopt a general regulation on ADS. A European ethical committee (policy option 3) should first be established and address some of the important issues mentioned at the beginning of this policy options briefing. Nevertheless, it is important to prepare for a future general regulation on ADS. In the same spirit as the GDPR, such a regulation should, for example, define what constitutes a 'sensitive ADS' and specify how they should be handled. In particular, it is important to define criteria that can be used to differentiate acceptable ADS, ADS that should be subject to an algorithmic impact assessment (policy option 4), and ADS that should be prohibited. It could also define the obligations of ADS providers, such as, for example, the obligation to make their ADS auditable through an application programming interface. This regulation could also address the responsibility for informing the persons affected by an ADS and clarify the explainability requirements of the GDPR in the context of ADS.

This document is based on the STOA study on 'Understanding algorithmic decision making: Opportunities and challenges' (PE 624.241) published in March 2019. The study was written by Claude Castelluccia and Daniel Le Métayer (Institut national de recherche en informatique et en automatique - Inria), 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.

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)