The Urgent Need for Robust Trust

Cultivating an environment in which algorithmic decision-making serves society

Input paper by the Ethics of Algorithms project, Bertelsmann Stiftung, June 10, 2020

We welcome the European Commission's effort to harmonize AI regulation and create an ecosystem in which algorithmic decision-making systems work for people and become a force for good in society. In this paper we want to share our findings from the "Ethics of Algorithms project" as part of the consultation on the "White Paper on Artificial Intelligence – A European approach to excellence and trust."

Algorithmic decision-making (ADM) systems are having a profound impact on society. The use of ADM systems has the capacity to unlock enormous societal, political, economic and cultural potential. However, if not used in the right way, such systems could also exacerbate existing inequities or trigger unexpected new ethical issues with large-scale impact. In recent years, the Bertelsmann Stiftung's "Ethics of Algorithms" project has contributed to the debate on the increased use of so-called AI and other ADM processes. The project conducts research on these issues and promotes dialogue on the societal impact of ADM technology and regulatory needs; in so doing, it seeks levers able to shape the sociopolitical and economic environment in such a way as to optimize algorithmic decision-making's potential for social good, while mitigating its risks.

We believe that this issue cannot be addressed exclusively on a national level. We therefore welcome the European Commission's efforts to establish a single regulatory approach that guarantees a level playing field for all vendors, regardless of their country of origin or the member state in which the system is operating, and of whether the system is being operated from within or outside the EU, or involves decisions made regarding EU citizens. It seems essential for EU citizens and for the Digital Single Market policy to harmonize data-subjects' rights at the European level. This includes aspects such as the right to human intervention, the right to receive an explanation of a decision, and the right to challenge or contest a decision. Our recommendations 1-3 below comment on the nature and scope of this endeavor.

Discussions addressing the regulation of ADM systems often suggest that we are starting with a clean slate. In reality, existing legislation needs to be interpreted in a new light, and underlying principles have to be rearticulated and applied to new contexts. Even though this may take significant effort and expertise, it is necessary in order to avoid undercutting and delegitimizing existing regulations, and to properly focus the current discussions. Recommendations 4-7 therefore analyze relevant existing regulations while highlighting their deficiencies and necessary revisions.

The measures taken to ensure that ADM serves society must go beyond the establishment of new legal requirements and the revision of existing laws. We need a number of interlinked policy approaches that focus on the organizations involved in the development and deployment of AI, as well as on the mechanisms through which they interact. The EU should also help strengthen, promote and financially support approaches to ADM that demonstrate an exemplary best-practice implementation of the technology. Recommendations 8-12 thus present necessary policy measures that go beyond a revision of the legal framework.

Drawing on the findings of our research, we suggest the following steps for the creation of a European approach to AI, under which technology would work for people and become a force for good in society:

Nature and scope of the European approach to AI

(1) **Include rule-based algorithmic decision-making systems in the approach:** In addition to its focus on machine-learning technologies, the European approach to AI must also address the use of rule-based decision-making systems that do not rely on machine-learning methods. One of the Bertelsmann Stiftung's key findings is that the impact of ADM systems depends on more than the actual technology itself (i.e., deep learning, machine learning or statistical analysis). **Even technically simple ADM systems can have a strong impact on people's lives**, while technically sophisticated ADM may be used without having any impact on consumers and citizens (e.g., when used only for quality assurance purposes on production lines). Nevertheless, the specific characteristics of machine-learning systems (AI) should also be translated into additional requirements (i.e., an obligation that changes made in the code be documented, or specific requirements regarding the explainability of complex machine-learning systems).

(2) **Focus on societal impact:** A risk-based regulatory approach is the right way to go. Legal requirements should be based on the societal impact of an AI system within its specific application field. For example, ADM systems used on automated production lines or within similar environments may not require the same scrutiny as ADMs used in the public sector or by credit-inquiry agencies. Article 9 of the GDPR, which governs the processing of special categories of personal data, including particularly sensitive data, shows that relevant regulatory models are already in place that could serve as an example in this regard. However, simply identifying specific high-risk sectors poses the danger of failing to recognize all application cases with a substantial societal impact. Instead, **we recommend the use of a two-dimensional risk matrix for the classification of application cases** into four to five different classes, as proposed by Krafft and Zweig in chapter three of our working paper "From principles to practice". The first dimension of the risk matrix should express the intensity of potential harm for individuals and society as a whole (i.e., the potential negative impact on fundamental rights, equality or social justice; threats to democratic institutions; the number of people affected). The second dimension should reflect the degree to which potentially affected parties are dependent on the AI system (i.e., is there a human in the loop who could overrule the AI system's decision? Is there a possibility of switching out the AI system for another?). For cases that show a strong potential for total damage, and therefore fall into the highest risk class, regulators should altogether prohibit the use of any ADM component, whereas cases in the lowest risk class would not need any additional ADM-specific regulation. For cases in high risk classes, we would furthermore welcome the obligation to conduct technology impact assessments.

(3) **Consider how ADM is organizationally, socially and politically embedded:** ADM systems are sociotechnological frameworks. Their impact on society is influenced not only by the algorithm itself, but also by the ADM's underlying goals and decision-making models, the data used as input, the ways in which the algorithmic output is used, and the entire organizational and political environment surrounding its use. The European approach to AI should be developed from a holistic point of view that considers all of these factors together. Any regulation focusing solely on the technological aspects of an ADM should be accompanied by other measures that facilitate oversight over the entire sociotechnological system (see recommendation 8).

Current regulatory deficiencies and necessary revisions

(4) **Review the scope of GDPR to close legal loopholes:** Given that algorithmic systems rely strongly on the processing of personal data, we acknowledge that the GDPR contains a comprehensive framework for the regulation of ADM, for example by stipulating certain fairness, transparency and accountability requirements, including the need to conduct a data-protection impact assessment for high-risk applications. Nonetheless, the current regulation contained in GDPR article 22 defines ADM systems narrowly, and lacks necessary safeguards that would enable individuals affected by ADM to exercise their rights. A legal study by Professor Mario Martini, commissioned by the Bertelsmann Stiftung and published in January 2020 (available in German), examined five prominent examples of the use of ADM (e.g., allocation of study places at universities and predictive policing). The analysis showed that in many cases, the GDPR does set relevant boundaries for the use of ADM systems in the EU. However, substantial legal uncertainty stems from the legal loophole surrounding partial automation – that is, when

automated systems are being used to prepare and support human decision-making. The upcoming GDPR review should address this regulatory failure. The legal paper "What Are the Benefits of the General Data Protection Regulation for Automated Decision-Making Systems" by the Bertelsmann Stiftung gives practical suggestions for possible complementary approaches within and beyond the GDPR.

(5) **Require comprehensive, context-tailored transparency mechanisms towards parties affected:** If an ADM is deployed in an application case with a high level of societal impact, the provision of information on that system's functioning and underlying goals, as well as on the results of previously conducted assessments of the system's technology, must be required. Moreover, this information must be furnished in a context-tailored and proactive manner. Institutions deploying ADM systems must ensure that the technology is labeled as such and that affected parties receive all information necessary to exercise their legal rights. The subjects of ADM mechanisms must therefore also receive a comprehensible explanation of how each relevant decision was reached, and must be given the **possibility of requesting a legal examination of the decision made by the system**. Humans in the loop must be trained to function as mediators who can provide the necessary explanations to affected parties.

(6) **Clarify liabilities to reduce uncertainties:** To make ADM a safe option for European companies and organizations, while also ensuring accountability, clear liability rules are needed. In contrast to the product market, there is currently no legal framework regulating safety and liability issues for services. Today, any institution using automated decision-making technology is responsible for its use and abuse. To reduce uncertainties and promote the uptake of ADM systems, the future legal framework must clarify which burdens are carried specifically by ADM producers and vendors, and which burdens are by contrast shared by ADM producers/vendors and organizational users that implement the technology within a particular setting (additionally, see recommendation 7 on reversing the burden of proof for affected parties).

(7) **Consider and revisit the non-discrimination framework**: Given the centrality of the principle of non-discrimination to the values of the EU, and considering various technologies' potential to induce discrimination at a large scale, any regulation of ADM systems should take existing non-discrimination law into account. To be able to address new forms of data-driven discrimination, the fragmented framework for equal treatment needs to be revisited at European level. With regard to the provision of information about the functioning and purpose of ADM systems, non-discrimination law can serve as an example of how the burden of proof can be reversed from the claimant to the responding party (in the case of ADM, shifting the burden of proof from the subject of the decision to the vendor).

Necessary policy measures beyond the legal framework

(8) **Ensure enforcement of the legal framework by strengthening oversight mechanisms:** Algorithmic decision-making does not require the establishment of new fundamental rights. However, in order to ensure that existing principles, freedoms and rights and any new laws remain enforceable when ADM systems are used, we need to strengthen existing oversight bodies and civil society watchdog organizations through financial means and through expanded competence-building measures. The Bertelsmann Stiftung recommends the establishment of institutionalized fora that include all relevant oversight bodies; these fora should be tasked with addressing ADM-related issues and identifying any legal uncertainties or need for further action. Furthermore, support should be given to research projects focusing on the development of auditing processes, as well as to processes of knowledge exchange between scientific and private sector stakeholders. While we acknowledge the importance of algorithm auditing, negative impact from the use of ADM systems can also arise due to the way these systems are socially embedded, or due to the interaction between multiple ADMs. These are issues that cannot be captured in an input-output analysis of individual algorithms. Oversight bodies should thus establish mechanisms that allow affected communities to give an account of their situation, with a goal of assessing systems' overall impact rather than focusing solely on the functioning of the technology.

Concerning the regulation of digital services in the Single Market, our expert paper "Governance of Digitalization in Europe" suggests the implementation of a networked and decentralized governance model that would allow a set of independent sector-specific regulatory authorities (at the European, regional, national and subnational levels) to be convened and coordinated. In contrast, discussions around the Digital Services Act have raised the prospect of creating a centralized governance model around a new central regulatory authority. Such governance issues also need to be considered when discussing how to strengthen and coordinate ADM oversight bodies.

(9) **Use labeling to create incentives for ethical technology development:** The Bertelsmann Stiftung recommends the introduction of an AI-Ethics Label. Labeling can offer orientation to developers trying to create ethically sound AI systems, while also increasing the transparency and comparability of products for users, and providing a basis for better enforcement of legal standards by oversight and watchdog organizations. Different application cases have different requirements for issues such as transparency and robustness, and thus demand different measures. We thus recommend a nuanced labeling approach that can do justice to the diversity of application cases, modeled in part after the energy-efficiency label. More information on the design of such a label can be found in our working paper "From principles to practice."

(10)**Revise public procurement standards**: In order to prevent corporate secrecy from getting in the way of public sector accountability on the use of ADM, EU public procurement rules should be complemented by transparency requirements for these systems. Vendors and developers that create ADM systems for use in government should have a duty of care to assist users of ADM systems in ensuring transparency, accountability and effective auditing. They should thereby agree to waive trade-secrecy or other legal claims that might otherwise inhibit a proper and full audit of their software. Public procurement standards should also be examined in order to determine how they can contribute to increasing the diversity of the landscape of ADM providers and systems. Additionally, we recommend the establishment of a public register that lists all deployed public sector ADM systems that may have a significant impact on society, and additionally contains information about their providers, their underlying goals and the results of technology impact assessments.

(11)**Boost competence-building:** If ADM is to be used to promote the common good, greater technical competence is needed among the general population, among company executives, and particularly among policymakers and those working in civil service positions. Without a basic understanding of the functioning of AI systems and their limitations, there is a serious risk that the public sector's adoption of AI might actually cause more harm than good. The introduction of AI in public sector organizations should thus be accompanied by competence-building measures for any people who will be making decisions regarding the system's implementation or interacting with the system. We furthermore need systematic initiatives to strengthen broad-based algorithmic literacy; that is, irrespective of educational attainment level and occupation, every citizen should be aware of the relevance of algorithmic systems in their personal or professional lives, be in a position to deal with such systems as carefully as necessary, and be informed about mechanisms through which automated decisions can be challenged. The freely accessible Finnish online course "Elements of AI" (available in English) is an example of an effective measure in this sense that could be scaled EU-wide. However, competence-building among citizens cannot replace effective oversight mechanisms, and should not be used to shift responsibilities from system developers to affected parties.

(12)**Promote diversity and innovation for social good:** The EU and its member states should promote measures ensuring that society as a whole – and all of its individual parts – benefits from the use of AI. With data and top programming talent resting largely in the hands of a few big tech companies, we are witnessing a tendency toward monopolization in the field of AI. However, only a diversity of organizations and individuals developing technology can adequately represent social plurality, avoid discrimination and promote innovation. The EU should therefore support initiatives that promote diversity in the tech sector and in related academic fields, promote knowledge exchange between these initiatives, and ensure that all affected parties have a seat at the table when decisions about the use of AI technology are being made. The sharing of data for social good should be encouraged, and the infrastructure needed for this task should be built. Civil society and nonprofit stakeholders should be supported financially and through competence-building in order to foster AI innovation for social good.

Contact

Carla Hustedt, Project Lead Ethics of Algorithms
Phone: +49 5241 81-81216
E-Mail: carla.hustedt@bertelsmann-stiftung.de

Ralph Müller-Eiselt, Director Program Megatrends
Phone: +49 5241 81-81456
E-Mail: ralph.mueller-eiselt@bertelsmann-stiftung.de