White paper

Reversing the burden of proof in the context of (semi-)automated decision-making

Summary

This document formulates actionable suggestions to improve legal protection for citizens and consumers in the European Union in the context for (semi-)automated decision-making (ADM). The suggestions in this document are linked to an existing concept in EU non-discrimination law: the reversal of the burden of proof.

June 27th 2022

Algorithm Audit

White paper – Reversal of the burden of proof

This document formulates actionable suggestions to improve legal protection for citizens and consumers in the European Union in the context for (semi-)automated decision-making (ADM). The below suggestions are linked to an existing concept in EU non-discrimination law: the reversal of the burden of proof[1].

Background

1. EU non-discrimination law

In general, the following evidence needs to be collected to shift the burden of proof to the alleged offender when discrimination is suspected (high-level overview)[2]:

i. A particular harm has occurred or is likely to occur;
ii. The harm manifests or is likely to manifest significantly within a protected group of people;
iii. The harm is disproportionate when compared with others in a similar situation.

The precise evidential requirements differ per case and per national law of EU member states, e.g., employees, chronically sick and disabled people hold additional legal protection.

2. Hurdles to shift the burden of proof

In the context of (semi-)ADM the following hurdles occur in collecting the above-mentioned evidence (selection):

– Poor visibility of algorithms: Users are not aware of being subjected to (semi-)ADM;
– Intransparency: Even if users are aware of being subjected to (semi-)ADM, users are not informed about the rationale behind the made decision, for instance which user characteristics contributed in what degree to the model outcome;
– Complexity of digital ecosystem: Collecting the above-mentioned evidence in the context of (semi-)ADM requires significant effort, and digital, technical and legal expertise by individuals.

In part, these aspects have led to the fact that the reversal of the burden of proof has not yet been triggered in public or private sector law for (semi-)ADM in the entire European Union. Jurisprudence on this matter is therefore still non-existent.

[1] The reversal of the burden of proof is applicable to both direct and indirect discrimination. In some cases proxy variables are closely related to protected grounds, from which the questions arises whether such cases should be classified as direct or indirect discrimination. Such cases are beyond the case of this document

[2] S. Wachter, B. Mittelstadt, C. Russell, Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI p.15 (2020). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3547922

Suggestions

The following steps should augment individuals across the EU to call upon the reversal of the burden when discriminatory (semi-)ADM is suspected:

I. Information obligation for using ADM

Improve the visibility of algorithms by informing citizens and consumers when subjected to ADM. The Netherlands Authority for Consumers and Markets (ACM) deems it "essential" that consumers are "informed on personalized prices that are set by ADM"[3]. We believe the same should apply to all applications of ADM. For instance, customers should be informed about automated credit checks conducted before access is granted to products and services[4]. GDPR articles 13(2)(f), 14(2)(g), 15(1)(h), 22(1) already impose requirements on user profiling techniques, but compliance and enforcement of those articles is lacking[5]. GDPR article 22(2)(a) formulates exceptions when ADM is considered "necessary for entering into, or performance of, a contract". A newly formulated information obligation (in national legislation) is needed to improve visibility of algorithms for EU citizens and customers.

Such an information obligation for ADM can take various forms, e.g., notifications in the user interface of an automated decision, notifications about the rationale behind the decision, references to procedures of appeal, and references to the user's legal rights in case of ADM. For semi-ADM similar information obligations are feasible.

Stricter information obligations will facilitate (strategic) litigation invested in the reversal of the burden of proof. It is reasonable to expect that increased visibility of ADM will lead to a more critical attitude of users to the application and outcome of ADM: What is the rationale for the ADM process? How is the decision made? Upon enquiry by users, the explanation of the ADM process, to which the model owners are obliged under GDPR, can be fulfilled in various ways. GDPR case-law provides guidelines on which information needs to be communicated to data subjects. More effective information obligations for ADM galvanizes appeal procedures and will ultimately result in an uptake of (strategic) litigation aiming for a triggering of the reversal of the burden of proof. And more litigation results in new jurisprudence and will ultimately lead to clarified procedures to reverse the burden of proof in the context of (semi-)ADM.

[3] https://www.thuiswinkel.org/webshops/kennisbank/kennisartikel-en/richtlijn-acm-uitgelegd-personaliseren-van-prijs-en-aanbod/

[4] For the private sector (under Dutch law) the same legal basis can be used, which is currently used for information obligations pertaining to ADM-driven personalized pricing. See also the EU Directive 2019/2161 on better enforcement and modernisation of EU consumer protection rules. https://eur-lex.europa.eu/eli/dir/2019/2161/oj

[5] FPF report: Automated decision-making under the GDPR – A comprehensive case-law analysis https://fpf.org/blog/fpf-report-automated-decision-making-under-the-gdpr-a-comprehensive-case-law-analysis/

In Estonia[6] and Slovakia[7], national legislation imposes informational requirements to citizens on executive government bodies when ADM is used. For instance, citizens need be informed about the existence, scope and impact of the ADM tool at hand.

If adopted, the European AI Act is not a panacea to improve visibility of ADM. Interpretation of how Section 2.3 Proportionality of the Act[8], which pertains to the "provision of information to flag the use of an AI system when interacting with humans", applies to government organizations, remains a national responsibility. Furthermore, even for non-high-risk ADM, for which the act provides much less regulations, an information obligation might be desirable from a societal point of view.

II. Hotline for discriminatory (semi-)ADM

Algorithms are not like night clubs: One cannot see to whom access is also denied. The burden for individuals to unite as a disadvantaged group when discriminatory (semi-)ADM is suspected is high (point 2. from the above-mentioned evidential requirements) and practically infeasible. Third parties can play an important role in bringing together disadvantaged individuals.

Establish a central hotline where potential discriminatory (semi-)ADM can be reported and where legal and technical support can be provided. Support existing national discrimination helplines[9], Human Rights Bodies[10] and Data Protection Authorities, such that alleged discriminatory ADM can be reported adequately and followed up upon. In-house technical expertise for these organizations is a special point of attention. Adjust online registration forms such that alleged algorithmic discrimination can be reported easily. Aggregating submitted forms should lead to identification of otherwise undetected disadvantaged groups. This will enable further strategic litigation.

III. Foster public control of (semi-)ADM

The rapid adoption of (semi-)ADM in the public and private sector is challenging courts across the EU. There is a lack of techno-legal jurisprudence, a disconcerting fact that needs to change soon. Only jurisprudence will provide definite guidelines on when the burden of proof can be reversed in the context of (semi-)ADM. To this end, courts and other legal authorities need to familiarize themselves with the operations of (semi-)AMD and the qualitative interpretation of those operations. Train staff of courts, parliament, ministries, and other enforcement and controlling authorities in interpreting (semi-)ADM qualitatively. Quantitative methods, such as

[6] Unemployment Insurance Act §23(4)
https://www.ilo.org/dyn/natlex/docs/ELECTRONIC/62500/60979/F-1016941640/EST-62500.pdf
[7] e-Kasa system and the Slowakian constitution
https://www.slov-lex.sk/pravne-predpisy/SK/ZZ/2021/492/20211217
[8] https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206&from=EN
[9] https://www.government.nl/topics/discrimination/reporting-discrimination
[10] https://www.government.nl/topics/discrimination/reporting-discrimination/filing-a-complaint-about-discrimination-with-the-netherlands-institute-for-human-rights

fairness toolkits[11], are mentioned[12] as possible solutions to mitigate bias in (semi-)ADM. Quantitative methods should not be overestimated to realize fairer ADM however, since qualitative disputes necessarily re-emerge in the assessment of which quantitative metrics should be used to measure normative concepts like fairness and bias. The same holds for the 'correlation and proxy challenge'[13]. No single quantitative method exists to assess which (proxy)variables are legitimate to use as input data. Quantitative methods might prove useful, but qualitative evaluation remains indispensable.

The interwovenness and also the tensions between technical and legal-ethical assessments of (semi-)ADM will come to play an increasingly important role in case-law in the coming years. For society at large, but in particular for authorities responsible for public supervision, control, and enforcement, it must be a priority to strengthen expertise in this domain. "Invest in relevant internal, ethical, [technical] and legal expertise at public authorities", concluded the recently published report Weighing Algorithms[14] written by Dutch Rathenau Institute.

[11] IBM's AI Fairness 360 toolkit https://aif360.mybluemix.net
[12] Municipality of Amsterdam Algorithm Register
https://algoritmeregister.amsterdam.nl/en/illegal-holiday-rental-housing-risk/
[13] See also Section 1.4.3 of Algorithmic Discrimination in Europe
https://op.europa.eu/nl/publication-detail/-/publication/082f1dbc-821d-11eb-9ac9-01aa75ed71a1
[14] Weighing Algorithms, Rathenau Institute
https://www.rathenau.nl/sites/default/files/2022-05/Algoritmes_afwegen_Rapport_Rathenau_Instituut.pdf