November 28, 2017

To whom it may concern,

On behalf of the Center for Data Innovation (datainnovation.org), it is our pleasure to submit these comments to the Article 29 Working Party (WP29) on its guidelines regarding algorithmic decision-making and the General Data Protection Regulation (GDPR). The nonprofit, nonpartisan Center for Data Innovation is the leading think tank studying the intersection of data, technology, and public policy. With staff in Washington and Brussels, the Center promotes pragmatic policies designed to maximize the benefits of data-driven innovation.

The guidelines go beyond the requirements of the GDPR in ways that will further chill the development and use of artificial intelligence (AI) in the EU. The requirement that a human reviewing an algorithmic decision consider "all the available input and output data" will prompt those using AI to limit the sophistication of algorithmic decisions and the data they draw on, so as to minimize the labor costs of a human review, which are considerable. Similarly, the recommendation that companies be prepared to explain any decision, whether it has legal or significant effects or not, will encourage them to limit their use of AI in all cases.

The guidelines appear to rely on the flawed assumption that a decision made by an algorithm is more likely to be indecipherable, biased, unfair, or damaging than one made by a human, and that human decisions are therefore preferable, particularly when dealing with vulnerable individuals. Proper auditing can identify and control for bias in algorithms and the data they draw on. But human bias is much more difficult to track and prevent in an objective and systematic way, because human motivation is often indecipherable even from a subjective point of view. Algorithmic decisions are less prone to bias, and easier to rectify when biased, than human decisions. Far from being a threat, AI is a promising technology and policymakers should accelerate its adoption to help society.

Yours faithfully,

Daniel Castro
Director
Center for Data Innovation
dcastro@datainnovation.org

Nicholas Wallace
Senior Policy Analyst
Center for Data Innovation
nwallace@datainnovation.org

REQUIRING A HUMAN REVIEWER TO CONSIDER ALL INPUTS AND OUTPUTS WILL LIMIT THE USE OF AI

The GDPR, which was adopted in April 2016 and will enter into force in May 2018, imposes tight restrictions on many uses of data, especially algorithmic decision making. In a flawed attempt to address concerns about automation, policymakers included a right of individuals not to be subject to a decision based solely on automated processing that has legal or otherwise significant effects, under Article 22 of the GDPR. To comply with this provision, companies must allow human review of such automated decisions. WP29 is not responsible for the contents of the GDPR, but WP29's interpretation that a human reviewer "should consider all the available input and output data" goes beyond the text of the regulation.[1] Algorithmic decision-making can involve vast amounts of complex input data, making human review of all of this data costly and impractical, which will deter the use of advanced AI systems in Europe.

Having humans analyze large datasets is costlier, less accurate, and slower than using algorithms, which is one of the reasons for using AI in the first place. In many cases, a human could not practically replicate the complex calculations an algorithm makes to analyze a data set. This does not make the algorithm a "black box," it simply means the cost of having a human reproduce all of the calculations the algorithm is capable of can be prohibitive. Therefore, a legal requirement for a human to consider all of the data necessarily constrains the feasible complexity of the data the algorithm can be allowed to analyze, lest a human have to review it all. Constraining the complexity of the data, in turn, constrains the sophistication of the algorithm.

Commercial use of AI is still in its very early days, and requiring costlier human evaluation of all input data will also reduce the incentive to deploy these technologies in the EU. Given the future role that AI will have in improving productivity in many industries, lower levels of deployment of AI will hurt European competitiveness.[2] For example, AI can support faster processing of insurance claims, which is particularly important in the aftermath of natural disasters where many claims need to be processed. AI can also support fairer insurance decisions by using richer data. Though Europe has failed to become a leader in online platforms, it could yet prove competitive in the use of AI—especially when one considers how Europe's industrial base might benefit from the technology—but not if the EU over-regulates AI or regulates it too early.

Humans do not need to manually review absolutely all of the data in a decision-making process to determine whether an algorithmic decision is fair. They can look at the algorithm's aggregate behavior to check for evidence of bias: for example, by checking whether particular outcomes correlate to any protected characteristics (such as ethnicity, religion, or sexuality), and if so, by finding the data points related to those characteristics that lead to those outcomes, deciding whether they are fair, and if not, adjusting the algorithm accordingly. If necessary, a human auditor can also consider the most important or sensitive data involved in an individual case, which will vary depending on the nature of the decision—but aggregate auditing will likely remain a far more effective way of identifying bias. Therefore, WP29 should eliminate this recommendation for a human reviewer to consider all input and output data to review a decision.

THE RIGHT TO EXPLANATION WILL NOT PREVENT BIAS OR GUARANTEE FAIRNESS

The right to explanation in Articles 13 and 14 of the GDPR will not achieve its goals of ensuring accountability and rooting-out bias in algorithmic decisions, because it requires humans to review isolated decisions rather than its overall behavior, which makes bias harder to detect. The relationship between protected traits and specific markers could cause bias to be very subtle and imperceptible to a person not intimately familiar with the relevant sensitivities.

For example, researchers at the University of California, Los Angeles (UCLA) conducted two studies that identified bias with regard to names associated with African-American ethnicity, a subtle connection that many would fail to spot, particularly non-Americans not familiar with American naming customs and their cultural implications.[3] Similarly, many outsiders will be unaware of complex ethnic markers in various parts Europe, such as Northern Ireland, Belgium, the Baltic states, Transylvania, or the former Yugoslavia, and therefore would not notice biased decisions that took those markers into account.

In contrast, regardless of what markers cause the bias, one can search for evidence of bias with regard to any protected traits in the aggregate outcomes of many decisions, even if the person conducting the search has little awareness of what indicators of those traits may be in the data. If auditors find evidence of bias in this way, they can then work to identify the markers causing the bias. But asking them to do this in isolated cases is akin to asking them to find a needle in a haystack that may not even contain any needles at all. In addition, identifying bias through aggregate auditing makes it easier to adjust the algorithm and prevent the bias from recurring in

future, instead of waiting for individuals to demand post-factum explanations after the damage has already been done.

The notion that humans are less susceptible to bias than algorithms are is also patently false—it is humans who have the biggest problem with bias—so it is incorrect to assume that a human decision is preferable.[4] Indeed, subjecting human decisions, which may be tainted with implicit bias, to algorithmic review is likely to be a more effective countermeasure to eliminating bias in decision-making than the reverse, which may only introduce more of it. Unlike algorithms, humans are capricious, which means an absence of evidence of sustained bias over many human decisions does not rule out an unfair decision in any particular case. For example, a person with racist views might struggle to hold down a job unless he makes some effort to keep them to himself at least some of the time: at work, he may only treat people differently when he thinks he can get away with it, or when he is in a particularly bad mood and lets his usual inhibitions slip. Indeed, as the aforementioned UCLA studies demonstrate, he may not even be fully aware of his own biases. In contrast, algorithms are not afraid of getting fired and do not have bad days: an algorithmic model is either consistently and measurably biased or it is not biased at all. It also goes completely against companies' interests to develop AI systems that could succumb to primitive human biases like racism, not least because such an algorithm could hurt profits by turning away perfectly acceptable customers.

It is impossible to scrutinize objectively how a human being comes to any decision, which is not the case with algorithms. Indeed, even subjective analysis of one's own decisions can be extremely difficult. Not only can humans lie to others or remain silent, becoming "black boxes" to everyone else, they are also frequently unaware of their own biases and can lie even to themselves. A human caught making an unfair decision might go into denial or come clean and apologize, but in either case there is no way of being sure they will not make a similar mistake again. But data analysts can find algorithmic bias through objective analysis of the algorithm's behavior over time, and when they do find bias, it is much easier to fix and prevent in the future.

THE RECOMMENDATION THAT ALL ALGORITHMIC DECISIONS BE TREATED AS IF THEY HAVE SIGNIFICANT OR LEGAL EFFECTS WILL FURTHER CONSTRAIN AI BEYOND THE REQUIREMENTS OF THE GDPR

WP29 recommends that data controllers be prepared to meet GDPR Article 13 and 14 requirements for a right to explanation even for decisions that do not fall under Article 22, which limits the scope of the right to decisions with legal or significant effects.[5] This recommendation is in effect a recommendation that data controllers limit their algorithms even where there is zero risk of the decision doing any kind of harm. This will further constrain AI by encouraging companies to minimize the potential labor costs of a human review for all decisions, not just those with significant or legal effects, by limiting the complexity of the data the algorithm processes, and consequently the sophistication of the algorithm.

The right to explanation is already a flawed method for dealing with potential bias in algorithmic decisions, especially for those that do have significant or legal effects, because it deals with individual decisions in isolation instead of in aggregate, which is both costlier and less reliable. It is therefore a mistake to apply the right to explanation to even more decisions, beyond the requirements of the law. Additionally, Article 22 is explicit in only requiring explanations of decisions with significant or legal effects. Why bother with such a distinction in law if WP29 is going to recommend the same rules for all decisions regardless?

Furthermore, while the GDPR itself provides the reasonable example of money lending as a decision with legal or significant effects, the guidelines stretch the distinction to meaninglessness by adding tenuous examples such as advertisements for online gambling. A lending decision can have significant consequences on a person—such as by stopping them from buying the car they need to get a better job—but a decision about which advertisement to show an individual will be less consequential because people have the ability to decide what information to consume and how to respond to that information.

For example, if a person sees an ad for online poker and then loses a large sum of borrowed money while gambling, the significant decision is the one they have made in response to the ad, not the automated decision to show them the ad in the first place. That said, the owners of the poker website may share some responsibility for not implementing adequate safeguards when customers gamble large sums of money, and such safeguards could include automated decisions, provided regulation does not deter their use. Additionally, the placement of the online poker ad need not have even been automated for this person to have seen it, whereas the

penalties for failing to stop problem gamblers that exist in member states such as the UK create an incentive for gambling firms to use algorithms that exclude people who, for example, show up on self-exclusion lists.[6]

THE RECOMMENDATIONS THAT AUTOMATED DECISIONS BE LIMITED WHERE CHILDREN ARE INVOLVED WILL LIMIT SERVICES THAT COULD BENEFIT THEM, WILL NOT PROTECT THEM FROM UNSCRUPULOUS ADVERTISING, AND HAVE NO BASIS IN ANY ARTICLE OF THE GDPR

WP29 says that controllers should not rely on the exemptions in Article 22(2) for automated decisions involving children, and that organizations should refrain from profiling children for marketing purposes.[7] These recommendations seem to be based on the assumption that there is something inherently problematic about automated decision-making and that regulators should therefore minimize its use involving children. On the contrary, automated decision-making creates benefits that can help safeguard children, such as when it comes to controlling access to online content.

The first of these recommendations, that controllers should not rely on Article 22(2) exemptions when processing children's data, will limit services that could benefit children, such as age verification tools or smarter parental control tools in Internet-capable devices, by deterring developers from improving them with algorithms. If developers fear an algorithm that controls access to a service only suitable for adults—such as a violent videogame—may be subject to legal challenges for profiling children, then they may refrain from using it, making it easier for children to access content that could harm them psychologically.

The second recommendation, that organizations refrain from profiling children for marketing purposes, will fail to protect children from unscrupulous or predatory advertising because it focuses on the delivery method, rather than the advertisement itself. Ultimately, this is a matter for advertising standards regulation, not data protection: advertisers can produce ads aimed at children whether they use algorithmic targeting or not, as afternoon television makes quite clear. The public has a legitimate interest in controlling advertising to children, especially when it comes to products such as alcohol. Ergo, algorithmic advertising online should be subject to the same advertising standards rules as all other forms of advertising, but this is not a matter for a data protection body such as WP29. Indeed, advertisers could even use algorithms to ensure they are complying with advertising standards regulations.

Moreover, neither of these recommendations has a legal basis in the GDPR. No article of the GDPR distinguishes between adults and children, as WP29 points out. The argument in recital 38 that children may be less aware than adults of risks, consequences, and their rights is valid. But that rationale is the basis for consistent ways of dealing with that problem—such as by deferring to parents for consent, or regulating what can be shown in ads regardless of the delivery method—and not specific restrictions on algorithmic decisions.

CONCLUSION

Automated decision-making is not the threat that policymakers seem to believe it is. On the contrary, its use in important decisions—including decisions involving vulnerable people and people often subject to prejudice—will help to make public and commercial services fairer, faster, and better. Policymakers should be wary of regulating such a new technology, lest they stifle its development in Europe. Similarly, when interpreting new legislation, regulators should refrain from introducing new requirements that are not explicit in the law. The restrictions in the GDPR are already a threat to AI, and WP29's requirements and recommendations on automated decisions risk making the problem worse by calling for harmful and unnecessary limitations that are not even in the regulation.

---

[1] See section III A, *Guidelines on automated individual decision-making and profiling for the purposes of Regulation 2016/679, wp251* (Article 29 Working Party), October 3, 2017, p. 10, http://ec.europa.eu/newsroom/just/document.cfm?doc_id=47963.

[2] Daniel Castro and Joshua New, *The Promise of Artificial Intelligence* (Center for Data Innovation), October 10, 2016, http://www2.datainnovation.org/2016-promise-of-ai.pdf

[3] Colin Holbrooke, Daniel M. T. Fessler, and Carlos David Navarrete, "Looming Large in others' eyes: racial stereotypes illuminate dual adaptations for representing threat versus prestige as physical size," *Evolution and Human Behavior,* January 2016, Volume 37, Issue 1, Pages 67-68, http://www.ehbonline.org/article/S1090-5138(15)00079-3/fulltext;
S. Michael Gaddis, "How Black Are Lakisha and Jamal? Racial Perceptions from Names in Correspondence Audit Studies," *Sociological Science*, September 6, 2017, https://www.sociologicalscience.com/articles-v4-19-469/.

[4] Joshua New, "It's Humans, Not Algorithms, That Have a Bias Problem" (Center for Data Innovation, November 16, 2015), http://www.datainnovation.org/2015/11/its-humans-not-algorithms-that-have-a-bias-problem/.

[5] See section III D 1, A, *Guidelines on automated individual decision-making and profiling for the purposes of Regulation 2016/679, wp251* (Article 29 Working Party), October 3, 2017, p. 14, http://ec.europa.eu/newsroom/just/document.cfm?doc_id=47963.

[6] Rob Davies, "Gambling firm 888 penalised record £8.8m for failing vulnerable customers" *The Guardian*, August 31, 2017, https://www.theguardian.com/society/2017/aug/31/gambling-firm-888-fined-online-bookmaker-problem-gamblers.
[7] See section V, *Guidelines on automated individual decision-making and profiling for the purposes of Regulation 2016/679, wp251* (Article 29 Working Party), October 3, 2017, p.26-27, http://ec.europa.eu/newsroom/just/document.cfm?doc_id=47963.