PRIVACY
INTERNATIONAL

Thematic Consultation Submission

Comments to the Article 29 Working Party Guidelines on Automated Individual Decision Making and Profiling

November 2017

PRIVACY INTERNATIONAL

Comments to the Article 29 Working Party Guidelines on Automated Individual Decision-Making and Profiling

INTRODUCTION

Privacy international is a non-profit, non-governmental organization based in London, dedicated to defending the right to privacy around the world. Established in 1990, Privacy International undertakes research and investigations into government surveillance and data exploitation in the private sector with a focus on the technologies that enable these practices. To ensure universal respect for the right to privacy, Privacy International advocates for strong national, regional and international laws that protect privacy around the world. It has litigated or intervened in cases implicating the right to privacy in the courts of the United States, the UK, and Europe, including the European Court of Human Rights and the European Court of Justice. It also strengthens the capacity of partner organizations in developing countries to identify and defend against threats to privacy. Privacy International employs technologists, investigators, policy and advocacy experts, and lawyers, who work together to understand the technical underpinnings of emerging technology and to consider how existing legal definitions and frameworks map onto such technology.

COMMENTS ON DRAFT GUIDELINES

I – DEFINITIONS

A. Profiling

We appreciate that the Working Group adopted a broad definition of profiling that includes assessing or classifying individuals based on characteristics, regardless of any predictive purpose.

Part II – Specific provisions on automated decision-making as defined in Article 22

A. "Based solely on automated processing"

We welcome the attempt made by the Working Party to define the scope of solely automated decision making based on profiling. In particular, we support the position that "the controller cannot avoid the Article 22 provisions by fabricating human involvement."

We would recommend, however, that the guidance is further strengthened. In particular, with regards to the distinction between decision-making based on profiling (ii) and solely automated decision-making, including profiling (Article 22) (iii).

In the guidance, the Working Party introduces the following illustration:

> *(ii) A human decides whether to agree the loan based on a profile produced by purely automated means(ii);*
>
> *(iii) an algorithm decides whether the loan is agreed and the decision is automatically delivered to the individual, without any meaningful human input.*

Even purely automated processing is significantly shaped by human decisions. It is also important to recognize that human decision-making can be significantly influenced, shaped and prejudiced by profiles that are produced by purely automated means. The propensity for humans to favour suggestions from automated systems over contradictory information made without automation, even if correct, is well documented in the literature on automation bias.[1] A good example is the use of automated risk scores in the criminal justice system. Proprietary software, such as the COMPAS risk assessment that has been

[1] See for instance Slitka, L., Mosier, K., & Burdick, M. (1999). Does automation bias decision-making?

sanctioned by the Wisconsin Supreme Court in 2016 [2], calculates a score that predicts the likelihood for committing a future crime. Even though the final decision is formally made by a judge, the automated decision made by a programme can be decisive, especially if judges rely on them exclusively or have not receive warnings about the risks, including that the software produced inaccurate, discriminatory, or unfair decisions.

We agree that a controller should not be able to avoid the Article 22 provisions by fabricating human involvement and that human intervention must involve meaningful oversight. However, **we would like to see a more comprehensive explanation on what qualifies as human intervention, especially in light of complex and opaque forms of advanced processing** (see our response on 'meaningful information about the logic involved').

We agree with Veale and Edwards[3] that Data Protection Impact Assessments would be a natural place to assess whether a decision is indeed based on solely automated processing. However, given that the recent A29WP DPIA guidance does not mention Article 22, the guidance on profiling and automated decision-making should further clarify meaningful human involvement.

Building on this**, we would recommend that in order to qualify as meaningful human intervention, the individuals making such a decision** should be able to determine whether the profile that informs their decision is accurate, fair, and not discriminatory. This requires that decisions are reconsidered regularly. It also requires that the individual providing meaningful human intervention has sufficient level of technical understanding, particularity about the myriad of ways in which profiling and automated decision-making can lead to unfairness and inaccuracies. It also requires that the system used to make a decision is sufficiently interpretable, auditable, and explainable.

Considering all available input and output data is not always feasible in the context of big data analytics and machine learning. It is also insufficient to demonstrate *meaningful* human involvement.

[2] Citron, D. (2016). (Un)Fairness of Risk Scores in Criminal Sentencing.

[3] Veale, M., & Edwards, L. (n.d.). Clarity, Surprises, and Further Questions in the Article 29 Working Party Draft Guidance on Automated Decision-Making and Profiling (November 15, 2017

B. “Legal” or “similarity significant” effects

We agree with the Working Party’s interpretation of significant effects, specifically the reference to the fact that “a decision must have the potential potential to significantly influence the circumstances, behaviour or choices of the individuals concerned.”[4] **We would like to encourage the Working Party 29 to clarify that whether or not profiling result in “significant effects” should not place the burden of proof on the data subject**.

This is of particular significance in the case of targeted advertising, which does produce significant effects in many cases and frequently relies on highly intrusive profiling. Broad audiences such as “women in the Brussels region” which is given as an example in the guidance are not representative of current targeting practices. Facebook’s Ad Targeting options alone allows for much more granularity, such as the ability to use combinations of behaviours, demographics, and geolocation data to reduce an audience to as little as one person[5]. A recently published study reached over 3.5 million individuals with psychologically tailored advertising and showed that “matching the content of persuasive appeals to individuals’ psychological characteristics significantly altered their behaviour as measured by clicks and purchases”.[6]

Targeted online advertising also has the potential to lead to the exclusion or discrimination of individuals. A 2015 study by Carnegie Mellon University researchers, for instance, found that Google’s online advertising system showed an ad for high-income jobs to men much more often than it showed the ad to women[7]. The study suggests that such discrimination could either be the result of advertisers placing inappropriate bids, or an unexpected outcome of unpredictable large-scale machine learning. Intentional or not - **such discrimination is an inherent risk of targeted advertising and impossible for individuals to detect.**

The vast majority of targeted online advertisement exceeds consumer expectations. Most consumers still think about online privacy as being primarily concerned with the data they share, and not the data that is observed from their behaviour, inferred, or predicted. It is our experience that the general understanding of how profiling works and the kinds of information it can reveal is exceptionally low.

[4] Article 29 Data Protection Working Party. Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, p. 10

[5] Kim, L. (2017). 5 Ridiculously Powerful Facebook Ad Targeting Strategies.

[6] Matz, S., Kosinski, M., Nave, G., & Stillwell, D. (2016). Psychological targeting as an effective approach to digital mass persuasion.

[7] Datta, A., Tschantz, M. C., & Datta, A. (2015). Automated Experiments on Ad Privacy Settings.

At the same time, it is becoming more difficult for consumers to express their wishes. Most consumers don't even know that they are being profiled at all. As a result, consumers commonly don't understand why any particular ad has been targeted to them - an effect that has been coined "the uncanny valley of targeted advertisement"[8]. Industry initiatives like http://youronlinechoices.com are misleading in that they give the impression behavioural advertising relies on cookies that can be blocked or deleted, even though consumer tracking is no longer limited to browser cookies but has advanced to more sophisticated techniques, such as cross-device tracking and device fingerprinting, which are disproportionately harder to avoid.

For these reasons, we would like to encourage the Working Party to adopt a position on targeted advertising that also avoids a subjective interpretation of "legal" or "similarity significant" effects. Who defines whether a targeted data subject is vulnerable? An individual with financial difficulties and a gambling addiction is clearly vulnerable, but what about women who are concerned about their appearance and receive ads for diets and plastic surgery? Instead, **it should be the controllers who provide sufficient information to ensure that profiling does not significantly affect individuals.**

D – Rights of the Data Subject

Articles 13(2) (f) and 14(2) (g) – Right to be informed and articles 15(2) (h) – Right of access - in the context of part II (Article 22)

We notice that the Working Party has opted to interpret the right to "a meaningful information about the logic involved" as an ex ante right about system functionality. As a result, the right becomes the right to a general explanation, rather than a right that would allow individuals to obtain an explanation for a *particular* individual decision that affects them.

The guidance on this provision states that "the controller should have already given the data subject this information in line with their Article 13 obligations".[9] This interpretation assumes that notification duties by controllers are sufficient to meet data subjects' right of access. We think this is not sufficient.

Notification and access serve two distinct but interlinked purposes (see more below). They also create different obligations on data controllers. While a "more general form of

[8] Manjoo, F. (2012). The uncanny valley of Internet advertising: Why do creepy targeted ads follow me everywhere I go on the Web

[9] Article 29 Data Protection Working Party. Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679

oversight" is appropriate for notification duties, the right of access plays an important role in seeking redress.

According to the draft guidance, "the controller provides contact details for the data subject to request that any declined decision is reconsidered, in line with the provisions of Article 22(3)." In the absence of an ex post right to explanation, data subjects have to blindly trust that their decision is being reconsidered fairly. Given that Article 22 only applies to decisions that have a significant effect, this imbalance of power is deeply troubling, especially if either profiling or decision-making relies on machine learning. By definition, such system only ever produces probabilistic outcomes. In matters that are inherently subjective, such as evaluation of an individual's qualities or ability to perform a task, this makes it very difficult for individuals to challenge unfair outcomes based on knowledge about system functionality alone (see further below).

Article 22(1) – Right not to be subject to a decision based solely on automated decision-making

We welcome that Article 22 is interpreted and applied as a prohibition, since this protects data subjects by default. As a result of this interpretation, data controllers can only make automated decisions about data subjects, if based on their explicit consent, if necessary to enter or perform a contract, or if authorised by law (provided that suitable safeguards are in place). Since profiling and automated decision-making often occur without the awareness of those affected, we are concerned that data subjects would not be able to effectively exercise their right to object. A prohibition is also appropriate, given that automated decision-making increasingly relies on advanced processing, including the use of algorithms large amounts of data, and machine learning. Such processing can be complex, and as a result, difficult to interpret or audit, yet can still produce decisions that are inaccurate, unfair, or discriminatory.

Part III – General provisions on profiling and automated decision-making

D – Rights of the Data Subject

Even though the language of Article 15(1) (h) is identical to Articles 13(2)(f) and 14(2) (g), a data subject can request access at any point in time. This will predominately happen after a decision has been made, which suggests that data subjects should be able to obtain an ex-post explanation.

Some key expressions in articles 12-15, specifically "meaningful information about the logic involved" as well as "the significance and the envisaged consequences" (Article 13(2)(f)), need to be interpreted to provide data subjects with the information necessary to understand and challenge profiling and automated decision-making.

Meaningful information must be sufficient to answer questions that the data subject might have *before* they consent to the processing (notification) and *after* a decision has been made (right of access).

ex ante

Before consenting to automated decision-making, individuals need to be given sufficient information to judge whether profiling is safe and will be to their benefit. **Further, data subjects should be notified about the extent to which automated decisions will rely on data that has been derived or predicted through profiling.**

We welcome that the Working Party urges data controllers to provide advice on whether "credit scoring methods used are regularly tested to ensure they remain fair, effective and unbiased".[10]

To be meaningful, such information should include

- what data will be used as input;
- what categories of information data controller intent to derive or predict;
- how regularly input data are updated;
- whether the actions of others affect how data subjects are profiled;
- the presence of algorithms
- and what kinds of measures the data controller will take to address and eliminate bias, inaccuracies, and discrimination. Since misidentification, misclassification, and misjudgement are an inevitable risk associated to profiling, controllers should also notify the data subject about these risks and their rights to access and rectification.

ex post

After a decision has been made, data subjects need to be able to establish whether profiling has been either unlawful or unfair. For instance, 'why did I get this outcome rather than some other outcome?' ', or 'What would have to be different - either in my personal circumstances or attributes, or the design of the system - to result in a different outcome?'.

All of these questions can only be answered though a ex post explanation of an individual decision. We would suggest that information about "the logic involved" should include giving data subjects access to the data on which such decision was based, in combination with information about the way in which it was automatically processed. In addition, Data

[10] ibid.

Protection Authorities (or other external institutions) should be in a position to audit automated decisions to test for bias and unlawful discrimination.

Additional remark: lack of clarity on machine learning

We notice that the Working Party does not always address the specific opacity, interpretability, and auditability challenges arising from advances processing such as machine learning. We note that the Working Party intends to provide more detail about transparency in its forthcoming Guidelines.

We would welcome more specific guidance on the following:

- What constitutes interpretability and auditability, given that there are different technical definitions of these terms?
- If a model is not interpretable, does it always produce decisions that fall under Article 22, since there can be no meaningful human involvement?
- To what extent do data controllers need to make their systems more interpretable and auditable, even if this might involve trade-offs with other criteria, such as accuracy?
- While models can be made more interpretable, machine learning algorithms and the scale requires to apply them usefully inevitably results in outputs that are difficult to explain, and sometimes even to anticipate. In what circumstances is relying on machine learning to generate knowledge (profiling), or to make or inform a decision about individuals not GDPR-compliant?
- The Working Party advises against an "over-reliance on correlations" as one way to establish appropriate safeguards. Yet, profiling using machine-learning classifies data subjects based on correlations.
- Data controllers have an obligation to make sure that data is accurate. Profiling using machine learning, however, is inherently probabilistic. Profiling merely establishes correlation, and as a result, can merely determine that an individual is *highly likely* to be female, *likely* to be unworthy or credit, or *unlikely* to be married, heterosexual or an introvert. Even a high level of accuracy still creates false positives and false negatives. If data controllers cannot guarantee that profiling using machine learning produces accurate data, is it still appropriate?

REFERENCES

Article 29 Data Protection Working Party. Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, Pub. L. No. WP251 (2017).

Citron, D. (2016). (Un)Fairness of Risk Scores in Criminal Sentencing. *Forbes*. Retrieved from https://www.forbes.com/sites/daniellecitron/2016/07/13/unfairness-of-risk-scores-in-criminal-sentencing/#6074794b4ad2

Datta, A., Tschantz, M. C., & Datta, A. (2015). Automated Experiments on Ad Privacy Settings. *Proceedings on Privacy Enhancing Technologies*, *2015*(1), 92–112. https://doi.org/10.1515/popets-2015-0007

Kim, L. (2017). 5 Ridiculously Powerful Facebook Ad Targeting Strategies. Retrieved November 23, 2017, from http://www.wordstream.com/blog/ws/2015/01/28/facebook-ad-targeting

Manjoo, F. (2012). The uncanny valley of Internet advertising: Why do creepy targeted ads follow me everywhere I go on the Web? *Slate*. Retrieved from http://www.slate.com/articles/technology/technology/2012/08/the_uncanny_valley_of_internet_advertising_why_do_creepy_targeted_ads_follow_me_everywhere_i_go_on_the_web_.html

Matz, S., Kosinski, M., Nave, G., & Stillwell, D. (2016). Psychological targeting as an effective approach to digital mass persuasion. In *Proceedings of the National Academy of Sciences*. https://doi.org/201710966

Slitka, L., Mosier, K., & Burdick, M. (1999). Does automation bias decision-making? *International Journal of Human-Computer Studies*, *51*(5), 991–1006. https://doi.org/10.1006/IJHC.1999.0252

Veale, M., & Edwards, L. Clarity, Surprises, and Further Questions in the Article 29 Working Party Draft Guidance on Automated Decision-Making and Profiling (November 15, 2017). *SSRN*. Retrieved from https://ssrn.com/abstract=3071679