Data is power: Towards additional guidance on profiling and automated decision-making in the GDPR

Abstract

In contrast to automated decision-making, profiling is a relatively novel concept in European data protection law. It is now explicitly defined in Article 4(4) of the EU General Data Protection Regulation (GDPR), and refers to the automated processing of data (personal and not) to derive, infer, predict or evaluate information about an individual (or group), in particular to analyse or predict an individual's identity, their attributes, interests or behaviour.

Through profiling, highly sensitive details can be inferred or predicted from seemingly uninteresting data, leading to detailed and comprehensive profiles that may or may not be accurate or fair. Increasingly, profiles are being used to make or inform consequential decisions, from credit scoring, to hiring, policing and national security.

Ever since the approval of the regulation in 2016, debates have focussed on the GDPR's potential to limit or offer protection against increasingly sophisticated means of processing data, in particular with regard to profiling and automated decision-making. While the GDPR offers new rights and protection, their scope and limits are open to debate, partly due to the clumsy syntax of the relevant articles and the lack of authoritative guidance concerning their interpretation.

The European Data Protection Board that will replace the Working Party on the Protection of Individuals with regard to the Processing of Personal Data is specifically tasked with publishing 'guidelines, recommendations and best practices' on the GDPR. In October 2017, the Working Party 29 has published draft guidance on profiling and automated decision-making. In this article we propose our suggestions to contribute to the development of guidelines which provide the strongest protections for data subjects.

Keywords artificial intelligence; algorithms; automated decision-making; data protection; discrimination; GDPR; privacy; profiling

Advanced profiling technologies answer questions we did not raise. They generate knowledge we did not anticipate, but are eager to apply. As knowledge is power, profiling changes the power relationships between the profilers and the profiled.[1]

1 - INTRODUCTION

Data, particularly when aggregated, can reveal a lot about a person. For example, when someone calls their best friend, visits a website of the National Unplanned Pregnancy Advisory Service, and then calls their doctor, we can assume that this person is probably thinking about an abortion, or is likely to have an abortion soon. Profiling automates such inferences and predictions by relying on an expanding pool of data sources, such as data about personal attributes, behaviour, location and contacts, as well as increasingly advanced data processing, such as machine learning. Once constructed, profiles can form the basis for decision-making.

In a world where everything we do becomes more and more traceable, profiling raises pressing policy questions: how can we protect people's privacy when intimate information can be predicted from seemingly mundane data? How do we ensure that profiling (and the decisions it informs) is legal, fair and non discriminatory? How can data subjects exercise their rights (in particular their right to object to automated decision-making) if the processing itself is opaque?

In contrast to automated decision-making, a hardly used right not to be subject to decision based solely on automated processing in the 1996 Data Protection Directive, profiling is a relatively novel concept in European data protection regulation. Now introduced into the EU General Data Protection Regulation (GDPR) by Articles 4 and 22, profiling refers to a form of automated processing of data to derive, infer, predict or evaluate certain attributes, demographic information, behaviour, or even the identity of a person.[2]

Even though the GDPR will apply from 25 May 2018, the implications of these changes, as well as the exact scope of new safeguards as they relate to new technologies, are already the subject of some debate. For instance, much debate has been directed at the question of whether a 'right to explanation' exists in the GDPR[3] and if it does, whether it is effective[4].

Such debates are complicated by the fact that some provisions of the GDPR lack "*precise language and explicit and well-defined rights and safeguards*".[5] A number of its provisions may thus lead to confusion, enforcement gaps or asymmetries. Some of these will have to be clarified once the regulation becomes effective. A key interpretative role will be played by the European Data Protection Board, which builds on the foundations of the Data Protection Directive's Article 29 Working Party, a body specifically tasked with publishing 'guidelines,

[1] Mireille Hildebrand, 'Profiling and the rule of law' [2008] 1(1) Identity in the Information Society 55-70
[2] General Data Protection Regulation [2016] OJ 2 119/33
[3] Bryce Goodman and Seth, 'EU regulations on algorithmic decision-making and a "right to explanation"' (*ICML workshop on human interpretability in machine learning 2016) <http://arxiv. org/abs/1606.08813>* accessed 23 November 2017; Sandra Wachter and others, 'Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation' [2017] 7(2) International Data Privacy Law 76-99
[4] Lilian Edwards and Michael Vale 'Slave to the algorithm? Why a "right to an explanation" is probably not the remedy you are looking for' [2017] Duke Law & Technology Review
[5] Wachter and others (n 1) 42.

recommendations and best practices'. In October 2017, the Working Party 29 has published draft guidance on profiling and automated decision-making.[6]

In this paper, we aim to offer suggestions as to how the GDPR can offer the strongest protections on profiling and automated decision-making for data subjects in Europe and beyond. After a brief introduction to profiling, automated decision-making and the harms it may create, we will provide a set of recommendations for additional guidance, with a focus on individual rights, as opposed to obligations placed on controllers.

2 - DEFINING PROFILING

Humans constantly categorise, generalise and classify the world around them to reduce complexity. Machines can be programmed to automatically process information in similar ways. Profiling practices create, discover or construct knowledge from large sets of data from a variety of sources. Such knowledge can be used to make or inform decisions that may or may not be automated.

2.1 – TYPES AND TECHNIQUES OF PROFILING

Valeria Ferraris et al. (2013) distinguish between group and individual profiling, as well as between direct and indirect profiling.[7] *Group profiling* identifies a group of individuals. Members of a group can either share a certain attribute (distributed profiling), or profiling can group people into a group without necessarily having the same attributes or without sharing all attributes (non-distributive profiling). An example for the former would be women who have visited an abortion clinic, while an example for the latter would be individuals with a higher risk for cardiovascular diseases as profiled by the occurrence of a certain number of risk factors. *Personalised or individual profiling* aggregates information about an individual and/or uses that information to derive, infer or predict unknown characteristics or future behaviour.

Both individual and group profiling may be conducted directly or indirectly.[8] *Direct profiling* uses data that has been provided by or observed about an individual or a group and uses that data to derive, infer or predict unknown attributes or future behaviour. *Indirect profiling* relies on data from a larger population and identified individuals on the basis of attributes that have emerged from the larger population. A good example are recommender systems that recommend music, videos, or books based on the purchasing history of others.
The knowledge that profiling creates can be generated from data using a wide range of automated processing techniques, from statistical deductions to advanced processing methods such as computational algorithms.

The reliance on complex computational algorithms, and machine learning in particular, may pose specific challenges with regards to opacity and the interpretability and auditability of the processing. Using machine learning methods, highly sensitive information can be inferred, or

[6] Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' [2017]
[7] Valeria Ferraris and others, 'Defining Profiling' (2013) <http://www.unicri.it/special_topics/citizen_profiling/WP1_final_version_9_gennaio.pdf> accessed 27 November 2017
[8] David-Olivier Jaquet-Chiffelle, 'Direct and indirect profiling in the light of virtual persons.' in Mireille Hildebrandt and Serge Gutwirth (eds), *Profiling the European citizen* (Springer, 2008)

predicted from non-sensitive forms of data. As a result of such profiling, databases that merely contain data about an individual's behaviour can be used to generate unknown data about their likely identity, attributes, interests, or demographic information. Such predictions may include information about health, political opinions, sexual orientation, or family life.

AI systems can be used to make or inform consequential decisions about people or their environment. Automated decision-making that relies on AI also plays a role in the personalisation of information and experiences, from news feeds to targeted advertising and recommendation systems. Such personalisation gears information towards individuals' presumed interests or identities, which are derived through profiling.

2.2 – PURPOSES OF PROFILING

Profiling occurs in a range of contexts and for a variety of purposes; from targeted advertising and healthcare screenings to predictive policing. Profiling is a way to collect, derive, infer or predict information about individuals and groups. Such knowledge can be used to make or inform decisions.

Profiling to infer or predict information

Through profiling, highly intimate information, including sensitive information, can be inferred, derived or predicted from personal and often non-sensitive data at varying degrees of accuracy. As a result, data about an individual's behaviour, can be used to generate unknown information about someone's likely identity, attributes, behaviour, interests, or personality.

- Personality traits, such as the Big-Five personality traits (extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience) can be predicted from standard mobile phone logs, such as call logs and contact data.[9]
- Publicly accessible data points (such as tweets) can be used to infer people's location, which in turn can be used to estimate someone's average income based on one's neighbourhood, average housing cost, debt, and other demographic information, such as political views.[10]
- Researchers were able to use cell phone usage history (call logs, contact data, and location) to predict users' socioeconomic status.[11]
- Emotional states, such as confidence, nervousness, sadness, and tiredness can be predicted from typing patterns on a computer keyboard.[12]

[9] Yves-Alexandre de Montjoye and others, 'Predicting Personality Using Novel Mobile Phone-Based Metrics' (2013) 7812 (4) SBP <https://doi.org/10.1007/978-3-642-37210-0_6> accessed 23 November 2017. 48, 55

[10] Ilaria Liccardi and others, 'I know where you live: Inferring details of people's lives by visualizing publicly shared location data' (Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems May 2016) <*http://people.csail.mit.edu/ilaria/papers/LiccardiCHI2016.pdf*> accessed 23 November 2017

[11] Joshua Blumenstock and others, 'Predicting poverty and wealth from mobile phone metadata' (2015) 350 (6264) Science <http://science.sciencemag.org/content/350/6264/1073.full > accessed 23 November 2017. 1073, 1076

[12] Clayton Epp and others, 'Identifying emotional states using keystroke dynamics' (Proceedings of the SIGCHI Conference on Human Factors in Computing Systems May 2011) <http://hci.usask.ca/uploads/203-p715-epp.pdf> accessed 23 November 2017. 715-724

- Social network profiles also predict traits such as impulsivity, depression and sensationalist interest, life satisfaction, emotional stability, drug use, sexual orientation and political views.[13]

Profiling to score, rank, evaluate and assess people

Profiling does not just result in descriptive profiles but through profiling individuals may also be measured against benchmarks of "*predefined patterns of normal behaviour*"[14] to establish whether and to what extent they deviate from such patterns. These benchmarks might be explicitly set, or might be implicitly calculated through anomaly detection technologies. Examples include:

- Religious groups in the US use profiling to identify unregistered Christians, they profile individuals and score them according to how seriously they take their faith.[15]
- A hiring software analyses an applicant's voice in order to identify applicants with "*energy and personality*" and evaluate "*language proficiency, fluency, critical thinking, and active listening*".[16]
- In 2016, IBM launched a tool that would help governments separate "real asylum seekers" from potential terrorists by assigning each refugee a score that would assess their likelihood to be an imposter.[17]

Profiling to make or inform a decision about an individual

Profiling generates information which may in turn be used to make or significantly inform decisions about individuals. Such decisions can be taken with varying degrees of human intervention and automation. Examples include:

- Hiring software automatically scores and sorts resumes and ranks applicants. The hiring company only considers applicants that score above a certain threshold.[18]
- The NSA reportedly uses web browsing data to predict an internet user's likely nationality, which allows the agency to distinguish between foreign and domestic communications.[19]
- In 2013, the Chicago Police Department conducted a pilot of a predictive policing program designed to reduce gun violence. The program included development of a

[13] Wu Youyou and others, 'Computer-based personality judgments are more accurate than those made by humans' (2015) 112(4) PNAS <http://www.pnas.org/content/112/4/1036> accessed 23 November 2017
[14] Fanny Coudert, 'When video cameras watch and screen: Privacy implications of pattern recognition technologies' (2010) 26 (4) CLSR 377, 384
[15] Bradley Hagerty, B., 2012, 'To Get Out The Vote, Evangelicals Try Data Mining' (2012) *New Hampshire Public Radio.* <http://nhpr.org/post/get-out-vote-evangelicals-try-data-mining> accessed 1 August 2017
[16] HireIQ 'Solutions' http://www.hireiqinc.com/solutions accessed 1 August 2017
[17] Patrick Tucker, 'Refugee or Terrorist? OBM Thinks Is Software Has the Answer' (*Defense One*, 27 January 2016) <http://www.defenseone.com/technology/2016/01/refugee-or-terrorist-ibm-thinks-its-software-has-answer/125484/> accessed 1 August 2017
[18] Alex Rosenblat and others, 'Networked Employment Discrimination' (Data and Society working paper, prepared for "Future of Work" project supported by Open Society Foundations, 2014) <https://www.datasociety.net/pubs/fow/EmploymentDiscrimination.pdf> accessed 23 November 2017
[19] John Cheney-Lippold, 'A new algorithmic identity: Soft biopolitics and the modulation of control' (2011) 28 (6) TCS 164,181

Strategic Subjects List (SSL) of people estimated to be at highest risk of gun violence. Research found that individuals on the SSL are not more or less likely to become a victim of a homicide or a shooting, but are more likely to be arrested for shooting.[20]

- A social networking site automatically flags some names as fake and suspends the respective accounts. As a result of this automated system, a disproportionate number of minorities' names are deleted.[21]

Profiling to make or inform a decision that personalises an individual's environment

Profiling is also used to automatically personalise experiences and information exposure, both online and increasingly offline. Real-time personalisation gears information towards an individual's presumed interests. Such automated decisions can even be based on someone's predicted vulnerability to persuasion or their inferred purchasing power. Examples include:

- Social media platforms tailor their services to their users' presumed tastes and interests, including what kinds of content, including news, users see in their news feeds, and in which order.[22]
- Billboards on the Tokyo Expressway—on one of Japan's busy expressways— detect and identify cars to then select and display content based on the types of cars.[23]
- Another study examined 16 major e-commerce sites and found search discrimination, i.e. differences in the products shown to users based on their click and purchase history as well as their operating system or browser or whether they were using a mobile device.[24]
- As we move towards 'smart' environments and 'persuasive computing' automatically modified choice architectures[25] can nudge the behaviour of data subjects in the real world.[26]

2.3 – WHEN IS PROFILING HARMFUL?

[20] Jessica Saunders and others, 'Predictions put into practice: a quasi-experimental evaluation of Chicago's predictive policing pilot' (2016) 12 (3) JEC 347, 371
[21] Dia Kayyali, 'Facebook's Name Policy Strikes Again, This Time at Native Americans'. (EFF, 13 February, 2015) < https://www.eff.org/deeplinks/2015/02/facebooks-name-policy-strikes-again-time-native-americans> accessed 23 November 2017
[22] Nicolas Holm, *Advertising and Consumer Society: A Critical Introduction* (1st edn, Palgrave Macmillan 2016)
[23] Intel and others 'Deep Learning Enables Intelligent Billboard for Dynamic, Targeted Advertising on Tokyo Expressway' (Case study, 2017) <https://builders.intel.com/docs/storagebuilders/deep_learning_enables_intelligent_billboard_for_dynamic_targeted_advertising_on_tokyo_expressway.pdf> accessed 1 August 2017
[24] Aniko Hannak and others, 2014, November. 'Measuring price discrimination and steering on e-commerce web sites' (Conference paper prepared for 'Proceedings of the 2014 conference on internet measurement conference', November 2014)
[25] Omer Tene, and Jules Polonetsky, 'Big data for all: Privacy and user control in the age of analytics.' NJTIP 239 (11) 27
[26] Jakub Mikians and others, 'Detecting price and search discrimination on the internet' (Conference paper prepared for *Proceedings of the 11th ACM Workshop on Hot Topics in Networks*, October 2012) 79, 84

Profiling practices are widespread and central to the way we experience products and services: recommender systems throughout the web rely on fine-grained profiles of what we might next want to read, watch, or listen to; dating apps rank possible partners according to our predicted mutual interest in each other; and social media feeds are automatically personalised to match our presumed interest, while online ads are targeted to show us what we might want to buy at a time when we are most likely to be perceptive.

At the same time, however, profiling poses three closely related risks. First, by virtue of generating new or unknown information, profiling is often highly privacy invasive. Second, it challenges common views about informed consent, but also raises issues around control, not just over personal data, but also one's identity. Third, since derived, inferred or predicted profiles may be inaccurate, or otherwise systematically biased, profiling may also lead to individuals being misidentified, misclassified or misjudged. When profiling is used to inform or feed into a decision that affects individuals, such inaccuracies may result in harm. In the words of the UN Human Rights Council:

> "*automatic processing of personal data for individual profiling may lead to discrimination or decisions that have the potential to affect the enjoyment of human rights, including economic, social and cultural rights.*"[27]

[27] UN General Assembly, *Human Rights Council: resolution / adopted by the General Assembly*, 22 March 2017, A/HRC/34/L.7/Rev.1

Power, privacy invasion and opacity

It is already difficult, if not impossible, for a data subject to understand or control how many entities hold what kinds of data about them, how they are linked, shared and aggregated.[28] Consumer tracking is no longer limited to browser cookies that individuals can block or delete, but has advanced to more sophisticated techniques, such as cross-device tracking and device fingerprinting, which are much harder to escape. At the same time, website, mobile applications, devices and smart objects routinely share data with unnamed "third parties" for purposes of advertisement.

In this current landscape, profiling using ever more advanced processing techniques, further shift the power relationship between data subjects and data controllers. Consumers are commonly unaware about the kinds of information that profiling can reveal about them.[29] For instance, consumers are aware of the sensitive nature of medical records, but not about the kinds of information that can be revealed from behavioural data, mobile phone records or smart meter electricity data. When sensitive information is latently present in a wide range of datasets, individuals have to trust that these will not be derived, inferred or predicted through profiling.

In addition, the process of profiling itself can be highly opaque, in particular if it is based on advanced processing, such as machine learning. Depending on the kinds of algorithms used, whether these are learning, and how they are trained, it can be difficult, even for the designers of such systems, to understand how or why an individual has been profiled in any particular way, or why a system has made a particular decision. Even in the absence of machine learning, profiling can often dynamic and evolving. For example, a data subject can be classified as likely to be homosexual today, and likely to be heterosexual tomorrow. Such changes may either be caused by new data about the data subject, by data about people that are not the data subject, or by changes in the way that profiling is being conducted.

Discrimination and unfairness

Both profiling and automated decision-making may lead to unfair, discriminatory or biased outcomes. The most apparent harm in this regard, is the ability of profiling to create uncannily personal insights, which may be used to the detriment of those who are already discriminated and marginalised. However, even if data controllers take measures to avoid the processing of sensitive data, this is not always effective.[28] In racially segregated cities, for instance, postcodes may be a proxy for race. Without explicitly identifying a data subject's race, profiling may therefore nonetheless identify attributes, or other information that would nonetheless lead to discriminatory outcomes, if they were to be used to inform or make a decision.

Uncannily accurate predictions may result in discrimination, yet, inaccuracies are another source of discrimination. On the one hand, inaccurate or systematically biased data can feed into profiles, which may lead to biased or discriminatory outcomes. At the same time, the process of profiling itself may generate data that is inaccurate.

---

[28] Max Van Kleek and others, 'Better the devil you know: Exposing the data sharing practices of smartphone apps' (Conference paper prepared for 'Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems', May 2017*)* 5208, 5220

[29] The Royal Society, 'Machine learning: The power and promise of computers that learn by example' (2017) <https://royalsociety.org/~/media/policy/projects/machine-learning/publications/machine-learning-report.pdf> accessed 1 August 2017

Individuals can be misclassified, misidentified or misjudged, and such errors may disproportionately affect certain groups of people. In fact, profiling creates a kind of knowledge that 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.

Societal implications: chilling effects, filter bubbles and autonomy

Some forms of profiling, such as risk scoring or the use of profiling for personalisation and recommendation may also have more widespread and long-term societal effects. Profiling sorts, scores, categorises, assesses, and ranks people. If ever-more data becomes the input of such evaluations, this might result in chilling effects. Individuals might pre-emptively self-censor their on-line behaviour, if the data it generates might be used against them.

Profiling also plays a role in personalisation of information, products and experiences. For example, by excluding content deemed irrelevant or contradictory to the user's beliefs or presumed interests, such forms of personalisation may reduce the diversity of information users encounter, resulting in filter bubbles[30] or echo chambers.

Personalisation of not just information but also our perception of the world around us will become increasingly important as we move towards connected spaces, like smart cities, but also in augmented, and virtual reality (VR). An environment that knows your preferences and adapts itself according to these presumed interests raises important questions around autonomy and the ethics of such manipulations.

3 - PROFILING AND AUTOMATED DECISION-MAKING IN THE GDPR

The GDPR introduces new provisions to address the risks arising from profiling and automated decision-making, notably, but not limited to, privacy.

Profiling is a relatively novel concept in European data protection regulation. The EU General Data Protection Regulation (GDPR) defines "*profiling*" in Article 4 as:

> "*Any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements.*"[31]

Profiling, as defined by the GDPR refers to both the *creation* and the *use* of profiles. By virtue of deriving, inferring or predicting information, practices of profiling generate personal and sensitive data.

3.1 - THE SCOPE OF GDPR PROTECTIONS

The limited scope of Article 22

[30] Eli Praiser, *The filter bubble: What the Internet is hiding from you* (1st edn, Penguin 2011)
[31] GDPR (n 2) 119/33

Article 22(1) of the GDPR provides additional safeguards against one specific application of profiling, namely the case of *automated individual decision-making* that fulfils is "*based solely on automated processing*" and produces "*legal effects concerning him or her or similarly significantly affects him or her*".[32]

Profiling can form the basis of decision-making that is both automated and produces significant effects. As we will argue in this paper, however, the wording of both "*based solely on automated processing*" and "*significant effects*" leaves room for interpretation and should be clarified to offer the strongest possible protection for data subjects. Even in its strongest possible interpretation, Article 22 will only ever address a limited range of automated decisions, including profiling.

In the absence of decision-making, profiling alone, therefore, does not give rise to safeguards under Article 22, but does still give rise to safeguards under Articles 13 to 15 (namely, information and access to personal data). To complicate matters further, the wording in Article 22(1) suggests that profiling is distinct from decision-making and is a form of automated processing while the wording in Articles 13(2)(f), 14(2)(g) and 15(1)(h) "*automated decision-making, including profiling*" suggests that profiling is itself a form of decision-making.

3.2 - OUR RECOMMENDATIONS FOR ADDITIONAL GUIDANCE

The European Data Protection Board that will replace the Working Party on the Protection of Individuals with regard to the Processing of Personal Data is specifically tasked with publishing 'guidelines, recommendations and best practices' on the GDPR.

So far, it must be highlighted that the understanding of profiling and automated decision-making is muddied by the clumsy syntax of the article and the lack of guidance concerning its interpretation. This is why we welcome the Working Party 29's guidance on profiling and automated decision-making. To contribute to the development of guidelines which provide the strongest protections for data subjects, we propose the following suggestions.

ARTICLE 22 – "AUTOMATED INDIVIDUAL DECISION-MAKING"

Article 22 of the GDPR is a welcome development: it is a significant right that addresses the growing reliance on automated decisions. However, there are numerous issues with the wording of Article 22 that can lead to asymmetrical interpretations and enforcement gaps.

A prohibition or right to object?

The wording of the "*right not to be subject to automated decision-making*" (Article 22 GDPR) can be interpreted as either a *prohibition* or a *right to object*.[33] Resolving this ambiguity is critical, since is greatly affects how strongly data subjects are protected. If interpreted as a *right to object*, data subjects could object to being subject to automated decision-making, unless the conditions in Article 22(2)(a)-(c) apply. If interpreted as a *prohibition*, data controllers would not be allowed to engage in automated individual decision-making, unless the conditions in Article 22(2)(a)-(c) are met (the conditions would have to be met before entering into or performing a contract, authorised by law, or explicit consent).

[32] GDPR (n 2) 119/46
[33] Wachter (n 1)

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.

Clarify definition of key terms

The safeguards laid out in Article 22(3) only apply to decisions that are "*solely based on automated processing*" and produce "*legal*" or "*similarly significant*" effects on the data subject.

"Legal" or "similarly significant"

Recital 71 of the GDPR provides very limited examples of activities that would have a significant effect. As a result, it is unclear whether the nature of "effects" depends on the subjective perception of the data subject or data controller, or whether some objective standards can be established to determine forms of automated decision-making which inherently produce significant effects.

We agree with the Working Party's interpretation of significant effects, specifically the reference to the fact that

> "*the effects of the processing must be more than trivial and must be sufficiently great or be important to be worthy of attention. In other words, the decision must have the potential to significantly influence the circumstances, behaviour or choices of the individuals concerned. At its most extreme, the decision may lead to the exclusion or discrimination of individuals.*"[34]

As it stands, the Working Party has opted for a nuanced subjective interpretation of "*significant effects*" that runs the risk of placing the burden of proof on the data subject. According to the draft guidance on profiling, *"processing that might have little impact on individuals generally may in fact have a significant effect on certain groups of society, such as minority groups or vulnerable adults.*"[35]

This guidance raises several important questions: 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 that are concerned about their appearance and receives ads for diets and plastic surgery? Arguably, it should be for the controllers to ensure that profiling does not significantly affect individuals according to an objective standard.

A subjective definition of "significant effects" has far-reaching consequences for targeted advertising, which increasingly and predominantly relies on automated individual decision-making. We appreciate that the Working Party defines conditions under which targeted advertising produces significant effects, including:

[34] Article 29 Data Protection Working Party (n 5) 10
[35] ibid.

- the intrusiveness of the profiling process;
- the expectations and wishes of the individuals concerned;
- the way the advert is delivered; or
- the particular vulnerabilities of the data subjects targeted.

Based on these criteria, however, we disagree with the conclusion that "*in many typical cases targeted advertising does not have significant effects on individuals*".[36] Targeted advertising 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 your audience to as little as one person.[37] 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*".[38]

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.

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. As a result, consumers commonly don't understand why any particular advert has been targeted at them - an effect that has been coined "*the uncanny valley of targeted advertisement*." [39] 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.

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.[40] 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.

---

[36] Ibid. page 11.
[37] Larry Kim, '5 Ridiculously Powerful Facebook Ad Targeting Strategies ' (*WordStream*, 20 November 2017) <http://www.wordstream.com/blog/ws/2015/01/28/facebook-ad-targeting> accessed 23 November 2017
[38] Sandra C. Matz and others, 'Psychological targeting as an effective approach to digital mass persuasion' [2016] Proceedings of the National Academy of Sciences 201710966
[39] Sara M. Watson, 'Data Doppelgängers and the Uncanny Valley of Personalization' (*The Atlantic* 16 June 2014) <https://www.theatlantic.com/technology/archive/2014/06/data-doppelgangers-and-the-uncanny-valley-of-personalization/372780/> accessed 23 November 2017
[40] Amit Datta and others, 'Automated Experiments on Ad Privacy Settings' [2015] 2015(1) Proceedings on Privacy Enhancing Technologies 92-112

For these reasons, we recommend that the Working Party adopts a position on targeted advertising that also avoids a subjective interpretation of "legal" or "similarity significant" effects.

Solely based on automated processing

In addition to decisions that produce "*legal*" or "*similarly significant*" effects, Article 22 only applies to decisions that are "*based solely"* on automated processing, including profiling.

Since "*based solely*" is not further defined in the regulation, the regulation allows for an interpretation that excludes any human involvement whatsoever. This offers a dangerous loophole that would render Article 22(1) inapplicable to many current practices of automated decision-making.

In light of these challenges, we welcome the attempt made by the Working Party to define the scope of solely automated decision-making based on profiling, by offering the following clarification:

(1) "*Based solely*" on automated processing *means that there is no human involvement in the decision process.*

(2) The controller cannot avoid the Article 22 provisions by fabricating human involvement.[41]

(3) *To qualify as human intervention, the controller must ensure that any oversight of the decision is meaningful, rather than just a token gesture. It should be carried out by someone who has the authority and competence to change the decision. As part of the analysis, they should consider all the available input and output data.*[42]

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. In addition, however, we would like to see a more comprehensive explanation on what qualifies as meaningful human intervention, especially in light of complex and opaque forms of advanced processing.

Meaningful human intervention or oversight is challenging to define. On the one hand, 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.[43] A good example is the use of automated risk scores in the criminal justice system. Proprietary software, such as the COMPAS risk assessment system, that has been sanctioned by the Wisconsin Supreme Court in 2016,[44] calculates a score that predicts the likelihood of committing a future crime. Even though the final decision is formally made by a judge, the

[41] Article 29 Data Protection Working Party (n 5) 10
[42] ibid.
[43]See for instance Linda Skitka and others 'Does automation bias decision-making?' (1999) 51 (5) IJHCS 999, 1006
[44] Danielle Citron, '(Un)Fairness of Risk Scores in Criminal Sentencing' (*Forbes* 13 July 2016) <https://www.forbes.com/sites/daniellecitron/2016/07/13/unfairness-of-risk-scores-in-criminal-sentencing/#6074794b4ad2> accessed 23 November 2017

automated decision made by a programme can be decisive, especially if judges rely on them exclusively or have not received warnings about the risks, including that the software produced inaccurate, discriminatory or unfair decisions.

On the other hand, oversight cannot be meaningful if the processing itself is opaque. This is especially important in the context of advanced processing that relies on computational algorithms, machine learning and large amounts of data. Such processing can be complex and opaque, and as a result, those who base their decisions on them, are not necessarily aware of its functions (and relative shortcomings). In this case, even if a human being makes the final decision, an automated process has *effectively* made the decision for them without the human being having the capacity to meaningfully query that decision.

In order to qualify as meaningful human intervention, the individuals making such decision should be able to determine whether the profile that informs their decision is accurate, fair and non discriminatory. This requires that the individual providing meaningful human oversight has sufficient level of technical understanding, particularity about the numerous ways in which profiling and automated decision-making can lead to unfairness, inaccuracies. It also requires that the system used to make or inform a decision is sufficiently interpretable, auditable and explainable. Considering all available input and output data, as the Working Party suggests, is not always feasible in the context of big data analytics and machine learning. It is also insufficient to demonstrate *meaningful* human involvement.

We agree with Veale and Edwards[45] that Data Protection Impact Assessments would be a natural place to assess whether a decision is indeed based on solely automated processing. One way to demonstrate actual oversight would be to document how often a human decision-maker actually intervenes in decisions and whether their intervention changes the result of the decision.

THE RIGHT TO BE INFORMED AND THE RIGHT OF ACCESS IN THE CONTEXT OF ARTICLE 22

Articles 13(2) (f) and 14(2) (g) require controllers to provide specific information about automated decision-making, based solely on automated processing, including profiling, that produces legal or similarly significant effects, namely:

- the existence of automated decision-making, including profiling;
- meaningful information about the logic involved; and
- the significance and envisaged consequences of such processing for the data subject.

Article 15(1) (h) uses identical language as Articles 13(2) (f) and 14(2) (g) and entitles data subjects the right of access to information about solely automated decision-making, including profiling.

However, some key expressions in articles 13-14, 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

[45] Michael Veale and Lilian Edwards, 'Clarity, Surprises, and Further Questions in the Article 29 Working Party Draft Guidance on Automated Decision-Making and Profiling' [2017]

understand and challenge profiling and automated individual decision-making. As a result, the "*right to explanation*" has been interpreted in two drastically different ways: as an ex-ante general explanation about system functionality, or as an post-ante explanation of a particular, individual decision.

The Working Party draft guidelines interpret "*meaningful information about the logic involved*" as an *ex ante* right about system functionality, for both, Articles 13(2) (f) and 14(2) (g), as well as Article 15(1)(h). 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. This interpretation also assumes that notification duties by controllers are sufficient to meet data subjects' right of access. On article 15(1)(h) the guidelines states that "*the controller should have already given the data subject this information in line with their Article 13 obligations*."[46]

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 taken, which suggests that data subjects should be able to obtain an *ex-post* explanation. Notification and access serve two distinct but interlinked purposes. They also create different obligations on data controllers. While a more general form of oversight is appropriate for notification duties, the right of access plays an important role in seeking redress.
In the interest of strong consumer protection, 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). For instance, in line with Article 22(3), data subjects may request that any declined decision is reconsidered. 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.

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*."[47]

To be meaningful, such information should include:

[46] Article 29 Data Protection Working Party (n 5) 15
[47] Article 29 Data Protection Working Party (n 5) 14

- what data will be used as input;
- what categories of information data controllers intend 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 lead to a different outcome?'.

All of these questions can only be answered though an *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 Protection Authorities (or other external institutions) should be in a position to audit automated decisions to test for bias and unlawful discrimination.

ARTICLES 17 AND 18

The rights to erasure and restriction of processing are useful and welcome forms of redress in the context of unlawful profiling techniques. In contrast to the portability rights established in Article 20, Articles 17 and 18 apply to all personal data, not just those that have been provided by the data subject. As a result, the data subjects' right to erasure and restriction of processing should apply to personal data that are being provided, observed, as well as derived, inferred and predicted. Further guidance is needed to clearly set out the Article's scope of application.

The way these Articles will operate in practice in a context where breaches are opaque and difficult to identify will require further clarification. This is particularly important in the context of automated processing that involves machine learning. In particular, there must be clear guidelines on how the rights to information and explanations can be strengthened and connected to these forms of redress.

4 - CONCLUSION

Profiling practices are common in a wide variety of contexts, from online advertising to policing, criminal justice, national security, immigration policy, supply chain management relying on RFID chips, or health care. While the techniques and technologies used can differ considerably, both the construction and the application of profiles have the potential to create significant harm to individuals.

The GDPR seeks to regulate most of the practices that we would normally label as "profiling". However, as we have demonstrated throughout this paper, a number of key provisions are either ambiguous or simply not defined. With this paper, we hope to make some suggestions as to how the existing Regulation can be supplemented with additional guidance to offer stronger protection for data subjects.

Finally, even after the GDPR will have come into force, individuals will not always be aware of their rights and of the forms of redress that they have available. Profiling is not only a practice generally carried out behind proprietary walls, it is also a complicated practice involving lots of data and opaque processing methods. Much effort will thus need to be invested between now and May 2018 but also after May 2018 to raise awareness and educate individuals about their new rights and the new safeguards provided under the GDPR.