Is that your final decision? Multi-stage profiling, selective effects, and Article 22 of the GDPR

Reuben Binns* and Michael Veale**

Background

Data protection law provides a set of rights and obligations in relation to the processing of personal data. Amongst its substantive principles, such as lawfulness, fairness, transparency, and various procedural elements and risk-based measures that apply to personal data processing, it also addresses the use of automated decision-making systems. Provisions around automated decisions are not new, having been part of the data protection toolbox for several decades.[1] The Data Protection Directive 1995 regulated automated decision-making;[2] however in practice this was largely theoretical and the rights and obligations therein were rarely invoked. However, this previously dormant corner of data protection has begun to awaken in recent years, partly as a result of the rise of automation in the private and public sector, and partly due to new interest, novel concepts, and safeguards in the General Data Protection Regulation (GDPR).[3]

Article 22 of the GDPR contains provisions which restrict the use of automated decision making systems where they are 'solely automated' and have 'legal or similarly significant effects'. The exact remedies the provision provides[4] as well as their utility[5] are subject to a debate that this paper does not seek to engage in. Instead, we focus solely on the *scope* of Article 22. A similarly worded provision exists

Key Points

- Provisions in many data protection laws require a legal basis, or at the very least safeguards, for significant, solely automated decisions; Article 22 of the GDPR is the most notable.
- Little attention has been paid to Article 22 in light of decision-making processes with multiple stages, potentially both manual and automated, and which together might impact upon decision subjects in different ways.
- Using stylized examples grounded in real-world systems, we raise five distinct complications relating to interpreting Article 22 in the context of such multi-stage profiling systems.
- These are: the potential for selective automation on subsets of data subjects despite generally adequate human input; the ambiguity around where to locate the decision itself; whether 'significance' should be interpreted in terms of any 'potential' effects or only selectively in terms of 'realised' effects; the potential for upstream automation processes to

*Reuben Binns, Department of Computer Science, University of Oxford, Oxford, UK
**Michael Veale, Faculty of Laws, University College London, London, UK
R.B. was supported by the Engineering and Physical Sciences Research Council (EPSRC) grant EP/S035362/1. M.V. was supported by the Digital Charter Fellowship from the Alan Turing Institute and Department for Digital, Culture, Media and Sport, and receives funding from Fondation Botnar.

1 See generally Lee A Bygrave, 'Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling' (2001) 17 Computer Law & Security Review 17.

2 Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. OJ L 281 (1995) ('Data Protection Directive'), art 15.

3 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ 2016 L 119/1 ('GDPR').

4 Sandra Wachter and others, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 76; Andrew D Selbst and Julia Powles, 'Meaningful Information and the Right to Explanation' (2017) 7 International Data Privacy Law 233; Gianclaudio Malgieri and Giovanni Comandé, 'Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 243.

5 Lilian Edwards and Michael Veale, 'Slave to the Algorithm? Why a "Right to an Explanation" Is Probably Not the Remedy You Are Looking For' (2017) 16 Duke Law & Technology Review 18.

foreclose downstream outcomes despite human input; and that a focus on the final step may distract from the status and importance of upstream processes.

- We argue that the nature of these challenges will make it difficult for courts or regulators to distil a set of clear, fair, and consistent interpretations for many realistic contexts.

in the Law Enforcement Directive, the GDPR's sister instrument for police and criminal justice.[6] Furthermore, many other laws internationally share a similar or near identical construction to Article 22.[7] Whilst our analysis focuses on the GDPR to study this provision, the issues we raise are relevant more broadly.

The relevant section of Article 22, Article 22(1) reads

> 'The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.'

Recital 71, which comments on Article 22, reads (in its relevant section)

> 'The data subject should have the right not to be subject to a decision, which may include a measure, evaluating personal aspects relating to him or her which is based solely on automated processing and which produces legal effects concerning him or her or similarly significantly affects him or her, such as automatic refusal of an online credit application or e-recruiting practices without any human intervention. Such processing includes "profiling" that consists of any form of automated processing of personal data evaluating the personal aspects relating to a natural person, in particular to analyse or predict aspects concerning the data subject's performance at work, economic situation, health, personal preferences or interests, reliability or behaviour, location or movements, where it produces legal effects concerning him or her or similarly significantly affects him or her.'

Certain ambiguities in these provisions have already been addressed in national and international guidance. The Court of Justice of the European Union (CJEU) has not yet been asked to interpret any questions concerning either Article 22 or its predecessor in the Data Protection Directive.

The first ambiguity of relevance to introduce in this article is the nature of what it is to base a decision 'solely' on automated processing. In other words, what level of human oversight or input is necessary to render an automated decision not solely automated? The most significant guidance that exists on this comes from the European Data Protection Board (EDPB), who point to several factors, including that the human overseer should be in a position to independently evaluate the case and assess the outputs of the system, and not simply rubber-stamp them; they should have the authority to overturn the automated outputs; and they should consider additional information and mitigating factors.[8] Member States are generally uniform in this area, with the arguable exception of France outside of private decisions, in the specific cases of the judicial and administrative sectors, where the requirement for decisions to be 'based solely' on automated processing in order for safeguards to apply is relaxed.[9]

The second ambiguity concerns what kinds of decisions have a 'significant' effect on the data subject?[10] Beyond the examples of credit and recruitment given in Recital 71 of the GDPR, EDPB guidance has elaborated with several other examples of decisions whose effects would qualify as significant, including dynamic pricing

6 Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA, OJ 2016, L 119/89, art 11.

7 Similar constructions exist all over the world, for example in Dahir no 1-09-15 du 22 safar 1430 (18 février 2009) portant promulgation de la loi n° 09-08 relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel (Morocco) art 11; Data Protection and Privacy Act 2019 (Uganda) s 27; Law on Personal Data ZRU-547 2019 (Uzbekistan) art 24, some US states e.g. Virginia, VA ST § 59.1-571 *et seq* ('Consumer Data Protection Act'). See also, as an important international instrument, Council of Europe Convention 108: Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data 1981, ETS 108, as amended by the Protocol amending the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (opened for signature 10 October 2018) 228 CETS ('Convention 108+') art 9(1).

8 Article 29 Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679 (WP251rev.01)' (6 February 2018); See generally Michael Veale and Lilian Edwards, 'Clarity, Surprises, and Further Questions in the Article 29 Working Party Draft Guidance on Automat0ed Decision-Making and Profiling' (2018) 34 Computer Law & Security Review 398. Technically, this document is of the 'Article 29 Working Party', a group established under the Data Protection Directive 1995. However, the guidance has been endorsed by its successor, the EDPB, and we will refer to it as an EDPB document to avoid confusion.

9 Gianclaudio Malgieri, 'Automated Decision-Making in the EU Member States: The Right to Explanation and Other "Suitable Safeguards" in the National Legislations' (2019) 35 Computer Law & Security Review, 13.

10 The term is, in the GDPR, a bit wordier than just 'significant': technically a decision which 'produces legal effects concerning him or her or similarly significantly affects him or her.' We will use 'significant effects' as shorthand to describe this part of the provision and elaborate as required.

which effectively excludes certain people from buying certain goods or services.

This article is focused on a related but understudied set of ambiguities relating to human intervention and significance. These are ambiguities which arise when human intervention and/or a decision's significance can be stratified by stages or by particular decision outcomes. We call these situations 'multi-stage profiling systems'. The ambiguities they present are residual to those addressed in previous guidance and scholarship around the meaning of 'meaningful human input' and 'significance'; they would remain even if we had comprehensive working definitions of these terms. These ambiguities are particularly important as Article 22 is often interpreted, including by the EDPB in its aforementioned endorsed guidance, as prohibiting decisions within its scope lacking a legal basis, which in turn can be difficult to secure due to the absence of a 'legitimate interest' ground or equivalent.[11]

The article is organized as follows. We begin by introducing examples of automated decision-making which involve multiple stages and/or effects with different levels of legal or similar significance, to motivate the debate. We then outline three typical roles that automated systems play in multi-stage profiling systems: 'decision-support', 'triaging', and 'summarisation' of human decisions. We then outline challenges and complications that these in turn bring, suggesting some approaches for their resolution—although, as the reader will see, these issues often cannot be resolved in clear and simple ways. We then briefly conclude.

Structures for multi-stage profiling systems

The issues we focus on in this article arise only under certain conditions. These are: where there are multiple outcomes with different significance; where there are different levels of human involvement between outcomes or segments of the population; and/or where there are multiple stages at which either significance or human involvement differ. These problematic ambiguities would 'not' arise in the following scenarios:

- *Uniformly insignificant outcomes:* If every possible outcome in a decision-making context is insignificant, there is no issue, because Article 22 does not apply due to the absence of significance. For example, if a clothes retailer uses profiling to make decisions about which clothes to promote to customers.
- *Uniform meaningful human involvement:* If there is meaningful human involvement applied equally in every outcome at every stage, Article 22 does not apply because there are no fully automated decisions. This would be the case, for example, with a risk assessment algorithm which provides additional information to a human decision-maker, who in every case will apply their own judgment.
- *Single-step automated decision-making with significant effects*: If there is a decision to be made, and all outcomes involve legal or similarly significant effects with no meaningful human involvement, then Article 22 would apply uncontroversially. For instance, an automated system for assessing requests to increase a credit card limit increase which results in the account holder either being granted or denied credit with no human input in either case.

However, in many—perhaps most—practical situations, such conditions would not hold. Consider the following:

- *Anti-money-laundering (AML) detection system:* A bank uses an automated system to analyse transaction data, to determine how suspicious a bank account is. Accounts which are flagged as 'not suspicious' continue to make transactions as usual. Those which are sufficiently suspicious are flagged for human review, which may result in an account being frozen pending further manual investigation.
- *Loan application assessment process:* Loan applications are first assessed by an automated system which either approves the application automatically, or forwards it to a human reviewer, who in turn either approves or denies the application.
- *Automatic scoring from hand-written scoresheets.* Workplace or educational assessments might involve human assessors giving numerical marks to a set of criteria evaluating an employee or student. An automated system might be used to automatically scan these handwritten scores, encode them as structured data, and tally up the scores to determine an overall score for the individual, rank the individual against others, and ultimately automatically make a decision about them (eg select the employee for promotion, etc). There is substantial human input in the grading process, but an automated system is used to make the final decision.

11 Cf Luca Tosoni, 'The Right to Object to Automated Individual Decisions: Resolving the Ambiguity of Article 22(1) of the General Data Protection Regulation' [2021] *International Data Privacy Law*, ipaa024.

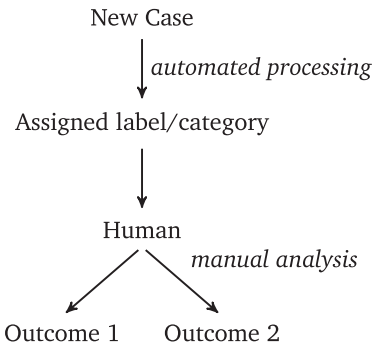

Figure 1. Decision support

To make this clearer, we now introduce several possible ways in which multi-stage profiling processes could typically be structured, and which form the basis for the discussions that follow. These centre around different roles that automation can play in assisting decision-making.

These roles are:

- Providing information to a decision-maker (Supporting);
- Determining which cases get to a human decision maker or passed to another automated process (Triaging);
- Consolidating decisions from one or more human decision maker(s) (Summarizing).

These roles can occur in the same system, but we will analyse them in isolation to provide some initial clarity. We do not consider the interaction between decisions over time, such as through the retraining or updating of a model, which may create important feedback effects.[12]

Supporting

Decision-support systems are arguably the most commonly discussed out of all multi-stage automated decision systems. They involve providing information to a human decision maker to help them make a decision about a case, but where they are just one source of information amongst others under consideration. The human decision maker is free to take the system's 'advice', but is not bound by it. This structure is illustrated in Figure 1.

Typically, new cases are profiled and assigned additional information as a result, which is passed to humans downstream. Humans in the process take the score, label, or category, and use it, alongside additional information (either used to generate the information above or not), to make a decision. These humans have all potential outcome categories available to them to assign cases to.

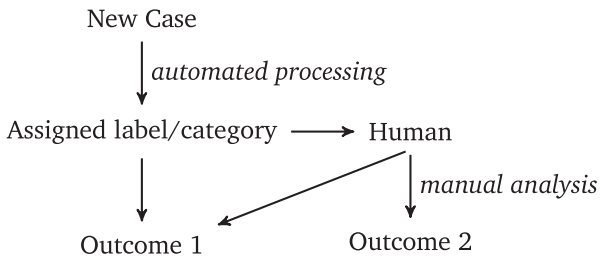

Figure 2. Triaging I

Examples of systems that would fall into this category include:

- Criminal risk assessment systems used to help guide judges determine bail conditions or parole, such as the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) system in use in the US.
- Systems used by employers to score candidates for job openings, where the score is not used to sift applications, but only to provide additional information to the human reviewer.

Triaging

Less considered in the literature are systems which result in some solely automated decisions, but where the status of the effects produced by those decisions (namely, whether they are legal or similarly significant) is ambiguous.

This is commonly found in anomaly detection systems. New cases are profiled and categorized. However, unlike in *Supporting*, the categorization determines the future decision pathway that the case continues along. For example, imagine (aided by Figure 2) that a case represents a credit card transaction, which is either assigned as non-suspicious (continues straight to outcome 1) or suspicious (sent for human review). In a 'pure' version of this case, the humans only receive cases which were classified in a certain manner by the automated system. They then look at information relating to these cases and sort them either into outcome 1 or outcome 2. These include:

- The Secondary Security Screening Selection list maintained by the US Transportation Security Administration (TSA), which aims to identify individuals who are a potential security risk, who are then typically subject to enhanced security screening when boarding commercial aircraft.

12 See eg Kristian Lum and William Isaac, 'To Predict and Serve?' (2016) 13 Significance 14.

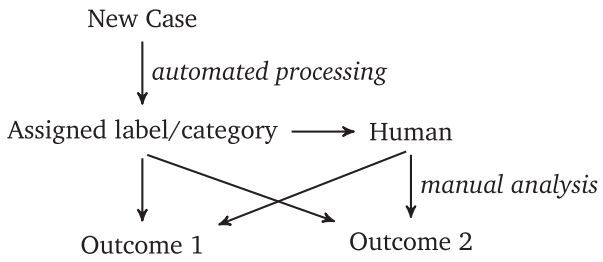

Figure 3. Triaging II

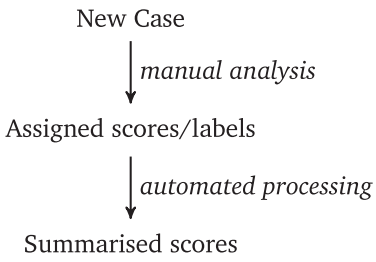

Figure 4. Automated summarization

- The 'Allegheny Family Screening Tool', a predictive-risk modelling tool created to support child welfare call centre operatives in the decision of whether to screen children into investigations upon information received on a hotline. The system scores the child's risk between 1 and 20, and children with a score of 20 are automatically screened in unless manually overridden by a supervisor. In other cases, the choice to screen in or out remains fully with the human operator.[13]
- The 'UK Visas and Immigration Directorate' deployed a (now decommissioned) immigration screening tool which automatically assigned red, amber, green, or 'super green' labels to applications to indicate risk. On the basis of these labels, the applications were forwarded to immigration officers of differing levels of seniority (in the case of red, amber, and green labels), or in the case of 'super green' labels, there may have been no human judgment at all.[14]
- The risk scoring systems in question in the CJEU cases *Accord PNR UE-Canada and La Quadrature du Net*, read in accordance with the requirements on those systems placed by the Court in light of the Charter, which ensure that positive flagged results be subject to 'individual re-examination by non-automated means'.[15]

In another configuration of triaging (Figure 3), there are three potential pathways that a case can go down after automatic processing: outcome 1, outcome 2, or a human reviewer (more outcomes can be added). This represents, among other situations, a genre of processes where the system also outputs a confidence level alongside its classification. Cases which are classified but with low confidence could be sent to a human reviewer.[16] A key difference between this and previous setups is that outcome 1 and outcome 2 are available to both the automated system, and to the human reviewer.

An example of this kind of structure is found in the content moderation field, and typified by Amazon's *Amazon Augmented AI (A2I)* set of products, which 'provides built-in human review workflows for common ML use cases like content moderation', where 'customers often need trained human experts to verify machine predictions that are lower confidence'. This system allows businesses to automatically send low confidence predictions either to crowdworkers on the *Amazon Mechanical Turk* platform, or their own employees or contractors.[17]

These two examples represent two canonical structures, but many more are imaginable, particularly in multi-class classification systems where cases are triaged to eg different specialists or a multitude of different downstream routes. However, these two suffice to describe the main tensions we wish to explore.

Automatic summarization

A final structure is one in which the human decision-making is upstream rather than downstream of the automated processing. One or more human decisions or assessments are recorded as structured data, and that data is summarized or consolidated automatically to generate an overall score or assessment which is used to make a decision. This is illustrated in Figure 4.

Examples of this kind of structure include:

- Handwritten assessments with numerical scores, such as exam scripts or employee assessments, produced

13 See generally 'The Allegheny Family Screening Tool' <https://www.alleghenycounty.us/Human-Services/News-Events/Accomplishments/Allegheny-Family-Screening-Tool.aspx> accessed 1 August 2021.

14 See Independent Chief Inspector of Borders and Immigration, 'An inspection of entry clearance processing operations in Croydon and Istanbul' <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/631520/An-inspection-of-entry-clearance-processing-operations-in-Croydon-and-Istanbul1.pdf> accessed 1 August 2021. This system was suspended by the Home Office following a legal challenge by civil society groups Foxglove and the Joint Council for the Welfare of Immigrants, viewable here <https://perma.cc/DQQ3-E9TD> accessed 1 August 2021.

15 Avis 1/15 *Accord PNR UE-Canada* ECLI:EU:C:2016:656 [173]; Joined Cases C–511/18, C–512/18 and C–520/18 *La Quadrature du Net and Others* ECLI:EU:C:2020:791 [182].

16 See generally NR Jennings and others, 'Human-Agent Collectives' (2014) 57 Communications of the ACM 80.

17 Amazon, 'Introducing Amazon Augmented AI (A2I) for Human Reviews of Machine Learning Predictions' (*Amazon Web Services, Inc.*, 24 April 2020) <https://perma.cc/M76G-U9W2> accessed 27 March 2021.

by examiners or HR managers, which are scanned and to which optical character recognition (OCR) or optical mark recognition (OMR) are applied. The automated process turns the handwritten recording of various human judgments into data and applies basic numerical operations to it (eg tallies the total scores, calculates an average, etc). This is then used to make a decision, possibly automatically, about an individual, such as giving a student an exam grade, or to select a candidate for a job.

- Vote counting based on optical recognition.
- Internet and gig economy platforms which summarize a set of user ratings about an individual in a single score. Here, the automated system simply aggregates multiple human judgments (although the function may not be a simple sum or average). For example, buyers and sellers on e-commerce sites like eBay and Etsy have overall scores which reflect the aggregate of multiple individual human ratings from other users. Drivers and riders on platforms like Uber have scores which are an average of ratings given to them by other drivers and riders, which are displayed to other users and in some cases used to restrict, suspend, or remove users from the platform.

In these cases, there is substantial human input and judgment in determining the initial data points which are later automatically aggregated. While the final decision is in some sense dependent on the automated process, and the aggregation method chosen, the main determinant is the collection of human judgments represented in the underlying data. The automated processes may be imperfect (for instance, they may make errors in transcription or in their basic numerical operations), but such imperfections are not driving the decisions. For any set of data, one would assume that when faced with the documents to summarize, a human would reach the same result, or that a human and machine would only differ insofar as they made random errors. Those deploying these systems therefore seek to automate processes rather than informationally 'augment' them beyond what a human would have added.[18] Some of the issues we cover below may also occur in systems that seek to both automate and augment results. For instance, a system that does not just seek to count marks, but also to normalize them within a class or sector,[19] or a gig work platform whose worker evaluations comprise more than just an aggregation of human ratings.[20] Where systems produce such outputs, that go beyond mere summarization of human inputs—such as medical diagnosis systems—these then may stray into the classic understanding of algorithmic systems under Article 22, and are not usefully characterized as multi-stage systems for our purposes, as all systems use input data originating in part from human judgment. We do not consider these a variant of summarization systems in this article in order not to conflate analysis of important, but distinct issues.

Challenges and complications

Having defined a range of ways in which automation can support human decision-making—supporting, triaging, and summarizing—we now address how these, individually and in combination, can create a range of complications for the scope of Article 22.

Selective automation

The 'pure' decision 'support' setup (where humans simply have regard to an applied profile) is *prima facie* the least controversial with regard to Article 22. Because no decision outcomes are reached without human input, it follows that decisions are not 'based solely' on automated processing and therefore are, in principle, not subject to Article 22.

There are a standard set of challenges identified in prior literature and in regulatory guidance to this. In short, if humans are effectively rubber-stamping the computer outputs, rarely or never overturning them, or lack the authority or competence to overturn them, it might be argued that the decisions remain 'based solely' on automated processing.[21] This is likely the issue in multi-stage decision-making under the automated decision-making provisions which has had the most public consideration to date.[22]

18 See generally Michael Veale and Irina Brass, 'Administration by Algorithm? Public Management Meets Public Sector Machine Learning' in Karen Yeung and Martin Lodge (eds), *Algorithmic Regulation* (OUP 2019) 123–25.

19 Such as Ofqual's 2020 exam results algorithm in England, abandoned under public pressure and a judicial review, including on Article 22 grounds, by civil society group Foxglove: Ofqual, 'Awarding GCSE, AS, A Level, Advanced Extension Awards and Extended Project Qualifications in Summer 2020: Interim Report' (*GOV.UK*, 13 August 2020) <https://www.gov.uk/government/publications/awarding-gcse-as-a-levels-in-summer-2020-interim-report> accessed 1 August 2021.

20 Such as Amazon Mechanical Turk's 'Master Worker' status <https://www.mturk.com/worker/help> accessed 1 August 2021, calculated using both ratings of requester customers and other 'marketplace data points'.

21 See generally Article 29 Working Party (n 8); Veale and Edwards (n 8).

22 For example, the Netherlands Scientific Council for Government Policy has warned of the risk of semi-automated decision-making ('semi-automatische besluitvorming'). See De Wetenschappelijke Raad voor het Regeringsbeleid (WRR), *Big Data in Een Vrije En Veilige Samenleving* (Amsterdam University Press 2016) 12. In other cases, faced down with amendments from opposition parties, the UK government had to clarify in parliamentary debate that they 'understood that mere human presence

Yet there is one implication that we do not believe has been highlighted so far. This is the situation where a subgroup in the data exists for whom all decisions are, effectively, more 'rubber-stamped' than the rest of the population, to a degree that it fails to meet whatever standard would be applied in the general case for this sub-group. In a manner of speaking, an implied 'triage' situation has been established before the moment of presumed decision support, by virtue of the way that different groups of individuals are treated by humans.

This could be present for both explicit groups and for latent groups constructed by the profiling system. The former would be the case if, for a subgroup of all cases sharing a certain input characteristic visible to the human decision-maker, eg all individuals with a certain occupation or protected characteristic, the human never overturned the judgment. Several recent studies have illustrated this using data from court decisions in jurisdictions in which the COMPAS system is used to help judges make bail decisions. They appear to show that some judges defer to the COMPAS score for black defendants but use their own judgment on white defendants.[23] Alternatively, in multi-class classification, what if all individuals classified by decision-support in a certain way (eg very high or low risk) never had the decision amended, whereas individuals classified in all other manners did? This appears to have been how the aforementioned *UK Visas and Immigration Directorate* case operated; applications in the lowest risk category of 'super green' were assessed as involving 'little or no human judgment'.

Would this mean that such a decision could be considered 'based solely on automated processing' for that subset? In a way, the human supposed to providing real overview has themselves become an automaton, at least in relation to the cases they are provided to examine. The conditions recommended by the European Data Protection Board (that decision makers must have the 'authority and competence' to overturn decisions; that decision-makers must not 'routinely apply' profiles; and that they must 'consider all the relevant data'[24]) could be selectively breached in this subset of cases. Individuals whose selective automation is feasibly predictable because, for example, decision maker bias is known to correlate with a particular input feature, they could perhaps be identified in advance, and a lawful basis for an Article 22 decision sought—although that the decision maker has such biases in the first place may be a deeper reason for concern. Considered at this level of granularity, non-Article 22 decisions could still be hypothetically made if this sub-group was automatically routed to a human decision maker. Yet, if overreliance was based on the predicted label rather than an observable input characteristic, by definition the individual could not be singled out in advance of generating a model output, and so Article 22 ground may be required for all individuals who feasibly could give rise to such a label.

Locating decisions

A further challenge is that of locating where a 'decision' takes place. A decision might be located in two potential places in decision-support systems. We might locate it in the moment of generation of a personality profile. This would be a decision to record a subject one way rather than another way. For example, the 'Secondary Security Screening Selection' list maintained by the US TSA aims to identify individuals who are a potential security risk, who are then typically subject to enhanced security screening when boarding commercial aircraft. The justification for locating it here could itself be justified in two ways. It could be argued that without being on such a screening list, an individual would not have been pulled aside and delayed such that they missed their flight. 'But for' the profiling at an earlier stage, the significant effect would not have occurred. This argument would likely be bolstered in a case such as the TSA case described here, where both profiling and searching are part of an expected, integrated system (rather than, say, a potential employer using some external reputation score in a way which was not expected by the data controller of that reputation scoring system). An individual might separately claim that a profiling process was intrinsically harmful to them. The issue of 'representational harm', related to the perpetuation of stereotypes, cultural denigration and the subordination of certain groups is an important structural consideration in automated systems,[25] although is unlikely to be seen as significant in data protection law given the high

or incidental human involvement is not sufficient' to avoid the automated decision provisions (HL Deb 13 December 2017, vol 787, col 1581).

23 See eg Bo Cowgill, who has used a quasi-experimental regression discontinuity design to assess whether judges using COMPAS were actually influenced by its risk scores or not. Bo Cowgill, *The Impact of Algorithms on Judicial Discretion: Evidence from Regression Discontinuities* (Working Paper, 2018). <http://www.columbia.edu/~bc2656/workingpapers.html> accessed 1 August 2021.

24 Article 29 Working Party (n 8) 21.

25 See Solon Barocas and others, 'The Problem With Bias: Allocative Versus Representational Harms in Machine Learning', Paper presented at the 9th Annual SIGCIS Conference, 29 October 2017. <http://meetings.sigcis.org/uploads/6/3/6/8/6368912/program.pdf> accessed 1 August 2021; Kate Crawford, 'The Trouble with Bias' (NIPS 2017 Keynote) <https://www.youtube.com/watch?v=fMym_BKWQzk> accessed 1 August 2021.

barrier of significance as being similarly significant to 'legal' effect, particularly due to the limitations of the wording of Article 22 insofar as harms relate to constructed groups.[26]

Alternatively, we might locate the decision in steps which only happen further 'downstream', and from which an effect is directly produced. For instance, imagine an individual whose flight risk has been automatically assigned being actually selected by a TSA official for further security screening. If TSA officials have the discretion to decide whether or not to select the individual for further screening, then arguably any relevant 'decision' has only been taken at this later point. Even if we might build a counterfactual showing that the decision would not have occurred but for the earlier applied profile, the potential of it being overturned at a later point might lead us to believe the decision is best located there.

Which of these two ways of locating the decision we adopt may have a bearing on whether or not Article 22 applies. If the initial automatic assignment itself is where the decision happened, and it produces the downstream effect of being subjected to further screening (which, we posit, is significant), then Article 22 applies. An escape from the prohibition, through consent, necessity for contract, or a provision in Member State law, as well as additional safeguards, would need to be identified before profiles were generated.

Alternatively, if the decision happened only later, when the human intervened, then Article 22 would 'not' apply—the decision would not be based solely on automated processing, unless there was a case of 'rubberstamping' as discussed in the previous section.

While it is clear that the initial profiling plays a role in producing the ultimate significant effect (of being subjected to further screening), where downstream discretion exists, it may not be sufficient by itself. However, in some cases it will likely be a *fait accompli*; in a system with significant effects and a profiling system, it will be inevitable that at some point, in some cases, a human will act in such a way as to bring about the significant effect through confirming the profile. This is one, core way in which even the most basic configurations of decision-support systems may present problems for determining the scope of Article 22. While there seem no easy answers to this quandary in case law or regulatory guidance, the first practical step this entails is that regulators may need to zoom out and seek empirical evidence on a much broader system than they initially thought they were investigating.

When is significance significant?

Take the challenges of anomaly detection, a common example of triaging. An anomaly detection system running over individual cases is likely to be designed in such a way as to be invisible to those whose activities are not considered anomalous. We typically do not know exactly when a bank has undertaken an assessment of whether a spending pattern is anomalous or not unless it has been flagged positively, at which case we might notice as our card is stopped, or potentially we are informed through a message. One decision outcome is invisible, and one is heavily visible.

This raises the first question of whether 'significance' is a condition that is determined where there is a 'potential' of a significant outcome, or not until a significant outcome has been 'realised'. If it is the former, the whole of an anomaly detection system might be in scope for Article 22. If it is only the realized decision, there are grounds for suggesting that only those with certain decision outcomes, such as a frozen account, should be in scope of Article 22, triggering the requirements for legal bases in Article 22(2) and specific information provisions in Articles 13–15, among other provisions. Significance would be a selective effect, associated with a particular decision subject rather than a system.

Confusingly, both the 'realisation' and the 'potential' approach have some backing in the law. Recital 71 of the GDPR gives examples of Article 22–qualifying decisions, specifying 'automatic *refusal* of an online credit application' [emphasis added]—rather than automatic assessment. Yet the information rights in Articles 13–14 must be carried out 'before' processing begins,[27] which requires an assessment of whether a decision falls within the definition of Article 22 to occur in advance of both the processing. The future looking language in the information rights articles, with a focussed on 'envisaged' consequences,[28] also support a reading of this obligation needing to be carried out at least in advance of processing.[29] Furthermore, the remedies in Article 22(3)—to seek human intervention, express a point of view and to

26 See generally Lilian Edwards and Michael Veale, 'Slave to the Algorithm? Why a "Right to an Explanation" Is Probably Not the Remedy You Are Looking For' (2017) 16 Duke Law & Technology Review 18. But see Lee A Bygrave, 'Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling' (2001) 17 Computer Law & Security Review 17, 19 ('a significant effect might lie merely in the insult to a data subject's integrity and dignity which is occasioned by the simple fact of being judged by a machine, at least in certain circumstances').

27 Case C-49/17 *Fashion ID GmbH & CoKG v Verbraucherzentrale NRW eV* [2019] ECLI:EU:C:2019:629 at [102]–[103].

28 GDPR arts 13(2)(f), 14(2)(g).

29 See generally Sandra Wachter and others, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 76.

contest a decision—only make sense 'once a decision has been taken'. This is particularly true for decisions relying on consent: if withholding consent would force human intervention (to avoid the decision being an Article 22 decision at all), then the safeguard of being able to obtain human intervention is redundant.

Other strange implications arise under the 'realisation' approach. A data controller could argue that they would rely on legitimate interests in order to justify the data processing for all non-significant decision-making based on automated processing, such as having your payments analysed and determined to be non-suspicious, 'but' seek a lawful basis (eg consent) before taking action on the basis of automated processing with an expected significant effect, such as being flagged as suspicious. Yet few people would be likely to consent to a negative effect where they know that a human review would be required were they not to. This would effectively mean that Article 22 forbids a controller from enabling a data subject to consent to an automated decision without already knowing the outcome, which itself seems quite contrary to the spirit of Article 22 as providing the opportunity for automated decision-making in those cases that safeguards exist. As EDPB guidance on the interaction between consent and other lawful grounds makes clear, it is unfair to 'swap from consent to other lawful basis' and therefore 'controllers must have decided in advance' what the lawful basis is.[30]

More practically, the 'potential' argument seems to make more sense than the 'realised' approach. Significance is a context-dependent concept. It is a function of what matters to some people: credit extended to some individuals would be a huge benefit to their standard of living; credit extended to others, potentially those in cycles of debt, might drive them further into financial problems; while for others, the effect may be marginal. Yet operationally defining significance on a case-by-case, realized basis will often not be practically possible in a manner that is not overly blunt. It will likely require significant surveillance to approach or achieve—a tension with the data protection regime more broadly.

In our view, the 'potential' argument is the only approach that makes sense. A decision mechanism should be considered 'significant' if it is reasonably foreseeable as such for some individuals who would be subject to it. There are no clear answers to the proportion of individuals that a foresight exercise must identify as potentially significantly impacted for it to 'contaminate' an entire automated decision system—is a small minority enough? If so, under what conditions? A data protection impact assessment is the likely venue where such decisions should be assessed and appraised by a data controller, and scrutiny by a supervisory authority over this document would likely be the first consideration of whether a controller had decided fairly or not.

While the 'potential' argument seems to be more sensible than the 'realisation' approach, interpreting significance in this manner, rather than in a more granular way for individual data subjects could have problematic consequences. Data controllers using anomaly detection systems, eg AML by banks, would certainly find it costly to interpret this as an Article 22 activity. Everyone who uses a bank account could potentially have their account frozen as a result of the routine automated processes applied to their account on a daily basis. In systems such as social networks where hundreds of automated 'non-decisions' are made over uploaded content to check it against varying standards such as terms of service, codes of conduct, legal obligations and the like[31], would a platform be expected to establish a lawful basis in relation to all of these? Given that necessity for contract is a very limited clause, short of creating a basis in Union or Member State law, platforms that wish to algorithmically curate or censor would be faced with the strong chance of needing to rely at least in part upon consent. The practical consequences of blanket refusal of this consent seem uncertain.

One way to lessen this issue might lie in adopting a two-fold approach: an interpretative strategy of treating 'positive' effects as non-significant, combined with an operational strategy of ensuring that no negative effects can happen automatically. In the AML case, this means that one's bank account continuing to work as usual is a 'positive' effect, and therefore not in scope and in need of a legal basis; meanwhile, merely having one's account flagged as suspicious for human review is not (yet) a significant effect, because no decision has been made at that point. While there may be potential for an automated decision with positive effect (business as usual), there is 'no' potential for an automated decision with a negative effect (accounts are only frozen after a human review, which is outside the scope of Article 22). In content moderation, it would imply that all flagged content was passed to a human to examine, while all non-flagged content was not.

30 EDPB, 'Guidelines 05/2020 on consent under Regulation 2016/679' (4 May 2020)

31 See generally Daithí Mac Síthigh, 'The Road to Responsibilities: New Attitudes Towards Internet Intermediaries' (2020) 29 Information & Communications Technology Law 1; Robert Gorwa and others, 'Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance' (2020) 7 Big Data & Society 2053951719897945.

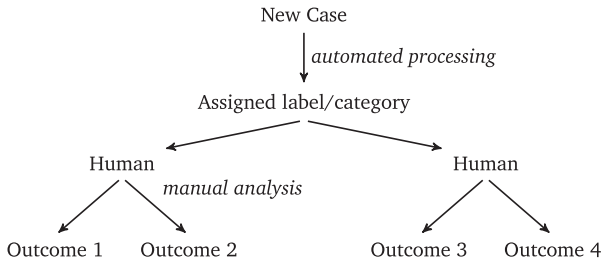

Figure 5. Upstream automation

Indeed, the *traveaux preparatoire* for the precursor to Article 22, Article 15 of the Data Protection Directive 1995 (DPD), restricted the scope to decisions which produced an 'adverse' legal effect or similarly adverse effect[32]—in line with the French administrative law which inspired this provision. However, such language never made it into either the DPD or the GDPR. In their first draft of their guidance, the Article 29 Working Party clarified that significant effects could be positive or negative.[33] In a later version, the Working Party removed all mention of their previous statement, and now express no view on the matter. Yet contrasting the GDPR with its sister instrument for policing, the Law Enforcement Directive, which in contrast to the GDPR explicitly mentions that only 'adverse' (legal) effects would apply, as well as the explicit removal of 'adverse' in the drafting of the Data Protection Directive, all serves as indicators that the legislator intended the provision to apply to significant effects regardless of their valence.[34]

The challenge that the valence of significance raises in relation to triaging can be framed as a question about what the 'baseline' for judging significance is. Might a suitable baseline be an individual's reasonable expectations? For fraudulent transactions, this seems appropriate: individuals do not expect to have their card stopped, but would be surprised and impeded when it is. (If they are indeed committing fraud, it could be argued that being stopped for fraud is insignificant given that they should expect to be treated more suspiciously than the average person.) Yet in other domains, such reasonable expectations are likely to be more subjective, and particularly problematically, tied to privilege. A financially secure person does not expect to be denied a loan. An individual from a marginalized community too often expects to be screened into questioning at a border. Treating significance as a function of expectations seems to go exactly against the purpose of Article 22 of securing fair processing where rights and freedoms are most at risk.

Foreclosing outcomes upstream

Let us assume for the sake of argument that having an account wrongfully frozen can constitute a significant effect—or, indeed, that the system in general produces significant effects. For the individual who has been filtered as an anomaly into a human review process for more specific consideration, and then has subsequently had their account frozen, has an automated decision been taken? On one hand, it could be argued that because a human has considered their case subsequent to the automated step, the decision is not 'based solely' on automated processing. However, this would be to ignore how without the filtering at the automated step, the individual would, with certainty, not have had their account frozen. 'But for the automation, the account would not have been frozen.'

In that case, the same cannot be said in reverse. An individual whose account was 'not' frozen could have experienced that outcome regardless of the automated system, as it is an end point of both branches. This is not always the case. Imagine a system where there are four outcomes (Figure 5). The first step is an automated assignment to one of two groups of human reviewers: A or B. Outcomes 1 and 2 are significantly different to outcomes 3 and 4, but human reviewers A can only allocate cases to outcome 1 or 2, and human reviewers B can only allocate cases to outcome 3 or 4. These systems are not fanciful, but have real analogues. This situation might occur in systems which are used to flag the best and worst performers in a workplace or education institution for human assessment leading to special measures or treatments such as promotion, remedial measures, or similar. Geography also presents a good example. Such a situation may occur where an automated system chooses the part of a country that an individual is eligible for social housing in, and a case-worker then chooses between available housing in that specific area.

In this case, the initial automated step, which assigns people to the A or B reviewers, is sufficient to determine that someone can or cannot obtain a certain outcome.

32 Commission of the European Communities, *Amended proposal for a Council Directive on the protection of individuals with regard to the processing of personal data and on the free movement of such data* (15 October 1992, COM(92) 422 final - SYN 287) 26.

33 Veale and Edwards (n 8).

34 Law Enforcement Directive (n 6), art 11 ('Member States shall provide for a decision based solely on automated processing, including profiling, which produces an adverse legal effect concerning the data subject or significantly affects him or her'). Note that some Member States have placed safeguards in relation to Article 22(2)(b) for automated decisions which trigger 'adverse' effects. Whether such derogations are permitted given the nature of the GDPR as a maximum harmonization instrument are outside of the scope of this article. See further Malgieri (n 9), 19.

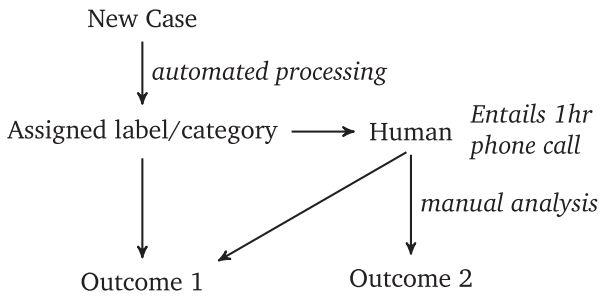

Figure 6. Anomaly detection with user cost

In all cases they might be eligible for a house: but will it be in Leeds or Swansea? This differs from the anomaly detection case, because in that case the human reviewers could assign cases to any of the possible outcomes. In this case, the human reviewers are effectively fettered in their decision-making power by the initial automated step. It is hard to say that an automated decision has not been taken. It could be argued that there are two separate decision outcomes in the system: the location in which an individual will receive a house, and the specific house itself. As the exact house relies on location, all subsequent decision-making that would be rendered non-applicable by a change in that upstream choice has the potential to be affected by a use of Article 22. However, aspects of the decision unaffected by that change (such as the logic of the choice between outcomes 3 and 4) would not be equally challengeable.

It is worth noting that with a small change, the anomaly detection triaging case falls under the same problem as well (Figure 6).

In this scenario, the two outcomes in the anomaly detection triaging case have *de facto* been increased to three. How so? Imagine the credit card fraud case: outcome 1 is that an individual can continue to spend as usual, while outcome 2 is that the card is frozen. Yet reaching outcome 1 through the human-intervention side incurs some delay or other cost: in this case, you must make a long phone call to plead your existence as a *bone fide* customer to a caseworker. The difference between automatic approval and manual human review (which might even involve providing more personal data) could be significant.

The consequences of a delay could have financial implications. An unequal application of human intervention to only those cases that could result in a negative outcome might lead to an unequal distribution of the lesser injustice (those who eventually get approved by human had to wait longer for their benefit). This unequal distribution could arguably in itself be a significant effect. Furthermore, automatic approvals in cases where resources are limited may directly affect the chances of others who have to wait for human assessment to get the positive outcome, because thresholds may be altered dynamically as a result of the number of e.g. loans given in a particular time period.

In other cases, being pulled aside for security screening at a border—even if the individual always makes it through in the end—is certainly possible to construe as a significant effect, if not from the impact on time, for the impact on an individual's dignity. Automated decision-making may have its costs, but so can the selective application of manual processes.

Is the final step decisive?

Yet further complications arise when we consider the order in which automated and human elements of the decision-making process are placed. As alluded to in the title of this article, it is common to think of the 'final' decision in a decision-making process as being final not only in the temporal and sequential sense, but also as the locus of decision-making power. It is therefore understandable that we might look to the final step in a decision-making process, and if that step is automated, judge the entire process to be automated. Conversely, a human making the final decision renders the process non-solely automated. However, as some of the previous cases suggest, neither inference will always hold true.

First, not all processes which end with an automated step are necessarily fully automated. Take the examples of summarization noted above. A system which applies OCR to handwritten scores and tallies them up to generate a decision (eg an employee assessment) may be the final step in a process, but the upstream human input could be substantial enough to render the process not solely automated. Summarization therefore provides an interesting counterpoint to the otherwise reasonable assumption that the final step in a process needs to be a human one in order for the decision to not be solely automated. It was for this reason that the decision has historically required a profile, although this is now only in the recital (the decision must be 'evaluating personal aspects relating to him or her').[35]

35 The European Commission initially defined one of the three conditions for the article's application as '[t]he processing must apply variables which determine a standard profile (considered good or bad) to the data concerning the data subject; this excludes all cases where the system does not define a personality profile: for example, the fact that a person is unable to obtain the sum of money he wants from an automatic cash dispenser because he has exceeded his credit limit would not fall inside this definition.' See Commission of the European Communities, *Amended proposal for a Council Directive on the protection of individuals with regard to the processing of personal data and on the free movement of such data* (15 October 1992, COM(92) 422 final - SYN 287) 26.

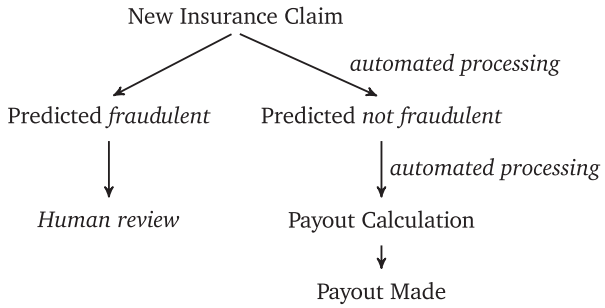

Figure 7. Chained upstream analysis

There will evidently be edge cases, in which the automated step involves something more than a consolidation or basic mathematical summary of underlying data. For instance, a system which mined user-generated text reviews using natural language processing to detect sentiment in order to evaluate a seller on a peer-to-peer e-commerce site, instead of simply aggregating their numerical scores, would arguably make the final automated processing step into a separate decision in its own right.

But generally, where those who generate the underlying scores do so on the basis of meaningful, considered judgments, and exercising the right level of authority and competence, the final step being automated should not necessarily render the whole process solely automated. Without them, the automated process has nothing to go on and no decision can be made.

Conversely, some processes which 'end' with a human step may 'still' constitute decisions which are solely automated. The examples raised above, of triaging processes which foreclose certain outcomes and thus fetter human decision-makers, demonstrate why this might be the case. Even if all outcomes involve a final human step, the outcomes available to the human to choose between have been constrained by an automated process at an earlier stage. That automated process therefore could be said to have produced a significant effect by constraining the options, even if a human chose between the subset of remaining options. So, whether the final step in a decision-making process is human or automated is not a steadfast guide to determining the overall status regarding article 22.

There is another other reason to be cautious about placing too much weight on the final step in a process as the locus of decision-making power. To outline why, let us consider a case in which Article 22 is clearly triggered, but only after multiple automated decision-making steps are chained together (Figure 7). For instance, an insurance provider might process claims first by applying a fraud detection algorithm, then depending on the result, it may apply a further algorithm to calculate an insurance payout or hand the case to a human to process the claim.

In cases like this (where payouts are made automatically), the first automated step does not by itself result in any solely automated decisions which produce legal or similarly significant effects; such decisions are only reached via a further step. Does this remove any upstream automated processing steps from the scope of Article 22?

The logic according to which substantial human input at the final step renders upstream automated processing out of scope (so long as it does not foreclose any outcomes), could be extrapolated in such a way as to apply here. Namely, it could be argued that in general Article 22 applies only to automated processes which are 'individually sufficient' to produce a decision with legal or similarly significant effects. As such, a prior step which provided a necessary, but insufficient condition for the downstream decision, would be outside the scope.

This could result in a partial loophole that is somewhat analogous to that presented by the perfunctory insertion of a human to rubber-stamp automated decisions, that the EDPB guidance sought to close. Instead of adding a human step on to the end of an automated process in order to render it non-solely automated, this loophole would be exploited by adding an additional 'automated' step, in order to render any prior automated steps outside of the scope of Article 22 altogether. Similarly, by switching around the order in which various automated steps take place, scrutiny over the more contentious steps that Article 22 might have provided could be avoided. For instance, the insurance provider in the above example could avoid scrutiny of the pricing algorithm by placing it *before* the fraud detection algorithm, or vice-versa. If Article 22 only applies to the final decision step, data controllers would have the discretion to hide contentious automated processing steps 'upstream'.

Closing this loophole would mean bringing upstream automated processing steps back into the scope of Article 22. Automated steps which 'indirectly' result in significant effects via other automated steps would be within scope. But relaxing the interpretation of 'producing a legal or similarly significant effect' to include 'indirect' production would be inconsistent with the consensus regarding previously discussed cases, where indirect production of such effects via a further human step is not enough to bring the decision within the scope of the Article. If 'automated' steps which indirectly produce qualifying effects are within scope when the further step is also automated, why would they not also be in scope when the further step is taken by a human?

We do not aim to resolve this tension here. We raise it to highlight the consequences of placing primary emphasis on the final step in the decision-making process to determine the status of the entire processing operation. This might sometimes make sense when the contention concerns whether the human input is sufficient (aforementioned summarization cases notwithstanding). But the examples above demonstrate that the contentious questions about the scope of Article 22 do not only concern the degree of human involvement or the significance of the effects of the decision. Even in cases where the final step is clearly solely automated, it may not be clear how far back up in a chain of automated steps the scope should stretch. Multi-stage automated decision-making scenarios therefore raise more fundamental questions about the scope of Article 22 which require further analysis in their own right.

Ways forward

Above, we have detailed a range of complications that come when considering Article 22 in common situations of multi-layered or selective automation. Unfortunately, none of these complications appear, at least to these authors, to have easy answers without drifting radically from the set-up of Article 22. Each of the answers we considered above presents its own difficulties, either casting the net too wide, too narrow, or both for different cases, as we have discussed in the relevant sections. If you the reader (potentially even a curious regulator) have come this far in search of a marvellous interpretative medicine for this tangle, we must let you down. If courts or regulators are presented with a wide range of such cases to investigate or adjudicate regarding, we anticipate they will find it difficult to distil a set of clear, fair and consistent interpretations within multi-level profiling systems.

Some might conclude that Article 22 is thus conceptually beyond saving. It is unusual as one of relatively few parts of the European data protection regime which seeks to lay out a bright line rule regime, as opposed to a balancing test or risk-based approach. Fall on the wrong side of it, and the legal gymnastics to secure a basis for decision-making can sink data controllers' dreams of automation entirely.

Perhaps the regime should move to a more explicitly subjective test, considering risks to rights and freedoms rather than the extent of automation? If the common narrative that this provision stems somewhat from an especially European unease with automation in light of human dignity is to be believed, blurring this bright line might prove controversial.[36] Controllers would likely seize the opportunity to claim that systems they operate are not risky, and given the opacity surrounding their use, and limited regulatory capacity, few claims would be likely to go challenged. Even under current automated decision rules, controllers operating hugely impactful automated systems argue against their classification under Article 22,[37] and similar battles around interpretation have long occurred with the definitions of legitimate interest and necessity for contract. With more flexibility, it seems unlikely there will be a queue to reclassify systems as 'within' scope of further obligations. Furthermore, the issues we have discussed in this article do not emerge from low or high risks to fundamental rights or freedoms, but from the complex features of certain practical set-ups. A risk-based framework may open up an escape route such that these complexities can avoid consideration in low-risk cases, but would do little for a situation where both high risks and such complexities were present.

Resolving these challenges might also follow a different route. Echoing the expansive approach to the definition of personal data in European law, courts in Europe and elsewhere could also move to 'expand' the notion of an Article 22–qualifying decision, while making the obligations less draconian. A similar cliff-edge area of the GDPR, the conditions needed to lift the prohibition on the processing of special category personal data in Article 9, has already seen the CJEU perform some gymnastics of its own. In *GC and Others*,[38] the Court was faced with the possibility it might need to require Google, whom it had previously declared a data controller,[39] to obtain a legal basis to process special category data before indexing it in its search engine. There are few, if any, feasible bases other than consent for this, barring the introduction of novel Member State or Union

36 See generally Margot E Kaminski, 'Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability' (2019) 92 Southern California Law Review 1529.

37 In England and Wales, the DPIA for the Bluetooth COVID-19 contact tracing app, which notifies people of exposure without human intervention with Government guidance which at the time of writing stated they should isolate from all other persons for 10 days, now states 'We consider it is arguable as to whether or not Article 22 is engaged by the contact tracing function of the app'. See Department of Health & Social Care, 'NHS COVID-19 App: Data Protection Impact Assessment' (*GOV.UK*, 22 February 2021) <https://www.gov.uk/government/publications/nhs-covid-19-app-privacy-information/nhs-covid-19-app-data-protection-impact-assessment> accessed 28 March 2021.

38 Case C–136/17 *GC and Others v Commission nationale de l'informatique et des libertés (CNIL)* ECLI:EU:C:2019:773.

39 Case C-131/12 *Google Spain* ECLI:EU:C:2014:317 [33].

law. Likely fearing it would 'break the Internet' (a fear it made explicit in *Lindqvist*[40]), the CJEU reframed the *ex ante* requirement of an Article 9 exemption into an *ex post* data subject right, arguing that this was justified by the specific 'responsibilities, powers and capabilities' of search engines. A seemingly unworkable opt-in became a guaranteed opt-out, running against the conditions for consent in both the GDPR and accompanying case law.[41] It seems feasible that this is a strategy the Court could use to develop its case-law to alleviate some controllers of *ex ante* obligations to lift the Article 22 prohibition, too—potentially, as in *GC and Others*, by substituting it with a strong *ex post* right to object or contest.

Where any such line between controllers or systems could be drawn would still be challenging given the concerns raised in this paper. Determining the nature of automation seems considerably more difficult a controller asking themselves whether or not they are a search engine. One can also ask how wise it is to continue the creation of sector-specific case-law for a horizontal instrument like the GDPR. In relation to search engines, the statutory right to erasure under Article 17 now sits oddly alongside, rather in place of, the right to delisting established in *Google Spain*.[42] Further developments of this kind may smooth some of the regime's bumpier aspects, but will only contribute to layering on more complexity and compliance costs. Yet just as the GDPR eventually codified several aspects of CJEU and A29 interpretation of the Data Protection Directive, perhaps that is the rollercoaster we end up boarding with Article 22.

Concluding considerations

Provisions addressing automated decision-making in international data protection law are, perhaps, the least-understood, most contentious, but perhaps amongst the most important parts in the years to come. In this article, we have not started with these provisions' philosophical basis in relation to the fundamental rights of privacy and data protection, or the adequacy and nature of the obligations, safeguards, and rights they entail. Instead, our points of departure have been some seemingly basic questions of their scope, in light of a range of common applications of (fully or semi) automated decision-making systems in the public and private sector, similarly applicable to the many international regimes with similar wording. Aside from the previously recognized ambiguities around 'solely automated' and 'legal or similarly significant' effects, these raise a set of difficult questions.

In this paper, we have raised five distinct (although in practice, likely interrelated) challenges and complications relating to automated decision provisions in European law and in many of the other instruments where similar or identical text appears. These challenges include: the potential for selective automation on subsets of data subjects despite generally adequate human input; the ambiguity around where to locate the decision itself; whether 'significance' should be interpreted in terms of any 'potential' effects or only 'realised' effects; the potential for upstream automation processes to foreclose downstream outcomes despite human input; and finally, that focusing on the final step may mislead as to the status of upstream processes.

We have argued that the CJEU or other courts overseeing data protection law with similar provisions will have to do more than carefully interpret these provisions' phrasing in order to rid this provisions of some of its thorniest tensions and trade-offs. Courts or legislators may seek a risk-based approach, although this would seem less a fix and more an attempt to lower the chance that such a complex case may end up needing firm analysis. They may mirror the CJEU's recent gymnastics in the widening and narrowing of aspects of provisions relating to personal data, erasure and delisting, and seek to, in certain situations, transform stubborn *ex ante* concepts like lawful bases into *ex post* oversight. This may patch over some of the most extreme tensions we have presented. The cost however would be legal uncertainty and complexity throughout the GDPR, and civil society groups may well ask how many other explicit provisions are up for judicial redefining in this manner.

However we, courts, or legislators, proceed, we can be sure of two things. First, that navigating the right way will require significant human judgment; and second, that it will be highly likely to have significant effects.

doi:10.1093/idpl/ipab020

40 Case C-101/01 *Lindqvist* EU:C:2003:596 [69] ('if the Commission found . . . that even one third country did not ensure adequate protection, the Member States would be obliged to prevent any personal data being placed on the internet').

41 See e.g. Case C-673/17 *Planet49 GmbH* ECLI:EU:C:2019:801; Case C61/19 *Orange Romania* ECLI:EU:C:2020:901.

42 Case C-131/12 *Google Spain* ECLI:EU:C:2014:317. See generally Jef Ausloos, *The Right to Erasure in EU Data Protection Law* (OUP 2020).