phg foundation
making science work for health

Black box medicine and transparency

Regulating transparency

A PHG Foundation report for the Wellcome Trust

UNIVERSITY OF CAMBRIDGE

Authors

Johan Ordish, Colin Mitchell, Alison Hall

Acknowledgements

The *Black Box Medicine and Transparency* project was funded by the Wellcome Trust as a part of their 2018 Seed Awards in Humanities and Social Sciences [Grant Number: 213623/Z/18/Z]. We thank the Wellcome Trust for their support.

The series of reports is informed and underpinned by a series of roundtables and interviews. These roundtables and interviews are detailed in the Report of Roundtables and Interviews. Further, highlights from both are seeded throughout all reports, being found in 'A Salient Feature' boxes.

Disclaimer

The following report is intended to provide general information and understanding of the law. The report should not be considered legal advice, nor used as a substitute for seeking qualified legal advice.

URLs in this report were correct as of February 2020

This report is available from www.phgfoundation.org

Published by PHG Foundation 2 Worts Causeway, Cambridge, CB1 8RN, UK

+44 (0)1223 761900

February 2020

Correspondence to: intelligence@phgfoundation.org

How to reference this report:

Ordish J, Mitchell C, Hall A. *Black Box Medicine and Transparency: Regulating Transparency*. PHG Foundation. 2020.

PHG Foundation is an exempt charity under the Charities Act 2011 and is regulated by HEFCE as a connected institution of the University of Cambridge. We are also a registered company No. 5823194, working to achieve better health through the responsible and evidence based application of biomedical science

Contents

1. Regulating transparency

This report considers the requirements of the General Data Protection Regulation (GDPR) on machine learning used for healthcare and medical research. In particular, it analyses the legal requirements for transparency and interpretability, and explores how these impact on the nature, timing and content of explanations.

Why make your machine learning model transparent? Why render it interpretable? Why explain its outputs? One reason is that lawful use of your machine learning system might require you to do just that: make your model transparent, interpretable, or explainable. These requirements may stem from the law in the form of legislation or common law, or may be underpinned by ethics or governance.[1] So law and regulation are only one practical reason why providing an explanation of a machine learning model might be required. For instance, if your model is to be trusted by clinicians and patients, some kind of interpretability may be necessary or highly desirable. Moreover, as noted by the previous Ethics of Transparency report, there may be persuasive ethical reasons to think that some kind of interpretability or explanation might be owed to patients or consumers. This report considers what the General Data Protection Regulation (GDPR) requires of machine learning for healthcare/research by way of transparency, interpretability, or explanation.

a. Transparency elsewhere

The tools to require transparency, interpretability, or explainability of machine learning are not unique to the GDPR. Other parts of the law may be leveraged to generate a duty of transparency, interpretability, or to explain. Briefly, other notable places in law that might generate such a duty include the following:

I. The Council of Europe have similar provisions to the GDPR regarding automated processing and explanation in Convention 108+ and have adopted recommendations containing further provisions for health-related data.[2][3]
II. The Medical Device Regulation (MDR), *In Vitro* Diagnostic Medical Device Regulation (IVDR), and associated harmonised standards may require devices that qualify as medical or *in vitro* diagnostic medical devices to provide some form of interpretability for proper risk assessment and mitigation. See our report *Algorithms as medical devices* for more details.[4]
III. In the public sector, judicial review (in some circumstances) may provide a duty to give reasons and so may be used in combination with data protection law to leverage some kind of explanation.[5]
IV. The 'patient centred' standard of care for communicating risk in *Montgomery v Lanarkshire* may require models used in a clinical context to be rendered (somewhat) interpretable, to avoid claims in professional negligence.[6]
V. Some harmonised standards require that products, services, or processes comply with relevant EU legislation.[7] While ISO/IEC 29100 on privacy frameworks and ISO/IEC 27701 on privacy information management remain unharmonised, these standards reference the GDPR, containing similar, often complementary provisions.[8]

Thus the GDPR is just one piece of the many, albeit the most prominent, potential sources of law that might generate a duty of transparency, interpretability, or explainability. The GDPR is no panacea for transparency - too much should not be asked of the GDPR and its rights and

duties - it is one piece of the puzzle, one means to the ends of transparency, interpretability, or explainability. Indeed, transparency is perhaps best achieved by combining data protection principles and data subject rights with rights and principles found in other parts of the law.[i]

b. Where the GDPR fits

The GDPR is a form of data protection with a defined (and restricted) purpose. This purpose colours the way in which its provisions should be interpreted. This section briefly introduces the body of regulation known as data protection, considers the protection of data in the UK, and outlines the pedigree and predecessor of the GDPR.

i. Data protection and privacy

The GDPR is a form of data protection. Privacy has a central place in data protection law. However, data protection is not synonymous with privacy.[9] Arguably, there are three reasons to think data protection is not solely concerned with privacy.[10] First, data protection protects a set of values apart from those protected by traditional concepts of privacy.[ii] [11] For example, the GDPR includes data subject rights that are related but not derived from privacy, for instance, the right to data portability.[12] Second and relatedly, swathes of data protection law concern the quality of the data processed. For example, the GDPR includes the principle of accuracy and right to rectification.[13] Third, data protection law often seeks to facilitate the processing of data for legitimate ends and does not seek to protect privacy alone. For example, while the GDPR includes principles like data minimisation, it also provides the means to lawfully process data for specified, explicit, legitimate purposes.[14] Accordingly, the lens with which the GDPR's provisions on transparency is interpreted is data protection. This is important as any kind of duty of transparency, interpretability, or explainability found in the GDPR will ultimately be a data protection solution that protects data protection interests and values.

ii. Protection of data in the UK

The GDPR is by no means the only legal mechanism that protects data. One practical way to categorise the different forms of regulation of data in the UK is to distinguish between law that is within the Information Commissioner's Office's (ICO) statutory authority and law that is outside the regulator's authority. In addition, regulation may apply to personal data generally or be sector specific, i.e. to machine learning/AI in general and to the healthcare and research data in particular.

---

[i] For example, the right of access in the German context has been paired with German labour law to reveal otherwise hidden information. See: Elteste U, Van Quathem K. *German court decides on the scope of GDPR right of access*. Available from:https://www.insideprivacy.com/international/european-union/german-court-decides-on-the-scope-of-gdpr-right-of-access/ [Accessed 9th February 2020].

[ii] See Westin's definition of privacy 'is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others.' Westin AF. *Privacy and Freedom*. New York: Atheneum; 1967: 7.

In the UK, the ICO has statutory authority in regards to the GDPR, the DPA 2018, and other legislation.[15] Notably, the ICO also covers the Privacy and Electronic Communications (EC Directive) Regulations 2003 (PECR). Amongst other things, PECR implements the ePrivacy Directive.[16] This Directive will likely be replaced by an EU Regulation shortly, but, for now, is the more specific (in legal terms, *lex specialis*) and so the law that governs in matters like cookies and certain electronic communications.[17] Consequently, while the GDPR governs personal data generally, other parts of EU law regulate specific types of processing that would otherwise be governed by the GDPR. In addition, the ICO also has within its statutory authority certain domestic legislation. One notable example of this is the Freedom of Information Act 2000 (FOIA 2000). It is clear from recent case law like *University of Bristol v John Peters* that data protection law (in this case the Data Protection Act 1998) can be difficult to interpret consistently with FOIA 2000 - the latter's purpose being the disclosure of data, the former's effect often resulting in non or restricted disclosure.[18] Regardless, the ICO must interpret both, while attempting not to make compliance with one noncompliance for the other. Regulation of data within the ICO's authority clearly includes more than the GDPR and may raise difficult conflict of laws questions.

Data is also regulated by law outside the statutory authority of the ICO. For instance, the common law of confidentiality and the tort of misuse of private information are the domain of the courts. ICO must therefore reconcile issuing guidance that is a) consistent with EU data protection authorities like the European Data Protection Board (EDPB) and Court of Justice of the European Union (CJEU), while also b) being cognizant that there is a parallel system of common law that data controllers must also comply with but not necessarily advise on.[iii] In this regard, not only is there an issue of consistency between the law within the ICO's discretion but also with the regulation that lies outside its discretion.

A Salient Feature | Roundtable 3

Select participants underlined the critical distinction between data protection under the GDPR and duties of confidentiality, with data protection being the jurisdiction of the Information Commissioner's Office, and common law duties of confidentiality being within the jurisdiction of the court.

Finally, there is also sector-specific regulation of data for healthcare and research as well as possible new regulation specifically for AI on the horizon. Regarding the former, there are many pieces of law that govern data in healthcare, for instance, if the data form part of a medical report then the Access to Medical Reports Act 1988 applies. There is a complex web of law, regulations, and orders that govern the flow of data through the NHS to other bodies such as NHS Digital and select disease registries: often Section 251 of the NHS Act 2006 underpins these transfers. Additional rules apply to very specific requests. For instance, where a third party requests health data in the context of disclosure in a civil case, Civil Procedure Rules 31.16 or 17 govern the release.[19] Regarding the former, the European Commission since 2018 have had a roadmap for 'Artificial Intelligence for Europe', this roadmap setting out the Commission's plans for supporting the market for AI in the EU and noting plans to ensure an 'appropriate ethical and legal framework.'[20] Recently, the European Parliament also passed a resolution on 'enabling the digital transformation of health and care in the Digital Single

[iii] N.B. The EDPB reconstitutes the Article 29 Working Party (WP29), adopting many of WP29's guidelines with respect to the GDPR. The EDPB offers authoritative interpretations of EU data protection law, the ultimate arbiter being the CJEU.

Market.'[21] This resolution acknowledges the importance of processing personal data according to the GDPR but notes certain points of ambiguity, urging clarification of how to use secondary data for research and coagulation of best practice for health data sharing.[22]

iii. The GDPR's genesis and pedigree

The GDPR is the product of earlier data protection measures and finds much of its inspiration in human rights instruments. Convention 108 contains the seeds of the GDPR and data protection principles, containing the core principles of lawful, fair, and purpose-limited processing.[23] In the UK, the first Data Protection Act was passed in 1984 and Convention 108 was ratified in 1985. Adopted in 1995, the Data Protection Directive (DPD), reflected concern that some Member States had not adopted Convention 108, and that protection of personal data therefore varied across the EU. In response to the DPD, the UK passed the Data Protection Act 1998, providing a national foothold for the Directive and its principles.

When interpreting the GDPR we must keep in mind its treaty basis and purpose. Notably, the GDPR differs from the DPD in its treaty basis. The DPD's treaty basis was Article 7a of the Treaty Establishing the European Union, that is, the most common treaty basis referencing the power to adopt measures to harmonise the single market.[iv] [24] The DPD nodded to the influence of human rights instruments such as Convention 108 but lacked a treaty footing to secure this purpose.[25] By contrast, the GDPR's treaty basis is Article 16 of the Treaty on the Functioning of the European Union (TFEU) as introduced by the Treaty of Lisbon. Article 16(2) confers the following power:

> 'The European Parliament and the Council, acting in accordance with the ordinary legislative procedure, shall lay down the rules relating to the protection of individuals with regard to the processing of personal data by Union institutions, bodies, offices and agencies, and by the Member States when carrying out activities which fall within the scope of Union law, and the rules relating to the free movement of such data.'

The human rights basis of the GDPR is further sealed by the first Recitals (1)-(4) offering a preamble that forcibly underlines the right of protection of personal data as enshrined in Article 16 TFEU and Article 8(1) of the Charter of Fundamental Rights of the European Union.

This human rights basis is also reflected in the preparatory materials (in legal terms - *travaux préparatoires*) of the GDPR. These materials emphasise the primary purpose of the GDPR to provide 'a comprehensive approach on personal data protection' in the EU to ensure the fundamental right to data protection is consistently applied across the EU.[26] Harmonisation of the single market and facilitation of data flows is mentioned but not given the same prominence as the fundamental right to protection of personal data. Further, the preparatory materials also note that this emphasis on fundamental rights, was driven by the widespread worry of Europeans that too much personal data was being requested of them online and the apparent lack of control over this data.[27] Consequently, the headline purpose of the GDPR to 'put individuals in control of their personal data' and the strong statement that:[28]

[iv] The contemporaneous treaty basis for this being: Consolidated Version of the Treaty on Functioning of the European Union [2016] OJ C202/95, art 115.

> '... individuals have the right to enjoy effective control over their personal information. Data protection is a fundamental right in Europe, enshrined in Article 8 of the Charter of Fundamental Rights of the European Union, as well as in Article 16(1) of the Treaty on the Functioning of the European Union (TFEU), and needs to be protected accordingly.'

This difference in treaty basis between the DPD and GDPR often underpins different interpretations of otherwise similar provisions between the two.[29] That is, the human rights basis and emphasis of the GDPR provides an opportunity to reinterpret similar provisions carried from the DPD to the GDPR. It is unclear how much we should invoke the GDPR's status as a human rights instrument to reinterpret provisions that were the same under the DPD. Certainly, the human rights basis does not give us wholesale license to reinterpret otherwise settled law under the DPD. However, where this change in basis does matter, it should colour our interpretation and give us pause for reinterpretation. Nevertheless, we note that where the text is the same under the DPD, the GDPR provision should only be reinterpreted on the basis of strong supporting evidence.

It should now be clear that if the GDPR does contain a duty of transparency, interpretability, or explainability then these duties will be coloured by the Regulation's status as an instrument of data protection law. Further, we should also note that interpretation of the GDPR or the DPA 2018 is by no means the last word on the law of data protection in the UK. A full description of any controller's position requires assessment of the other regulation in this space.

c. Why the GDPR?

This report focuses on the GDPR and the rights, duties, and principles it contains. It touches only incidentally on the UK Data Protection Act 2018 (DPA 2018) for three reasons:

I. The common belief that the DPA 2018 is the UK implementation of the GDPR is false. The GDPR is an EU regulation and so has been directly applicable (a part of UK law) since its publication, although only in force since the 25th of May 2018. It is contrary to EU law even to transcribe a regulation's requirements into domestic law.[30] Consequently, those looking for a domestic foothold for the GDPR should look no further than the GDPR itself.

II. The DPA 2018 supplements the GDPR but only where there is domestic competency to do so. There is no domestic competency over most of the relevant Articles of the GDPR from which we derive the duties of transparency and explanation. One notable exception to this is Section 14 DPA 2018 which gives Member States competence to legislate further exemptions to the prohibition/right against automated processing found in Article 22(1) GDPR. We discuss this Member State competence found in Article 22(2)(b) further in Section 5(h)(i) of this report.

III. The DPA 2018 sections commonly cited that relate to transparency and explanation actually only relate to processing of data in the context of law enforcement or intelligence services.[31] In other words, where the GDPR does not directly apply and where there is domestic competence to legislate. Given this, these sections of the DPA 2018 are mostly irrelevant for healthcare and research, we exclude them from our analysis.

d. The GDPR post-Brexit

A recurring question is whether, and to what extent the GDPR will still be relevant in the UK after Brexit. Pre-Brexit, the GDPR - as a regulation - was directly applicable and a part of UK law. After the end of the transition period, the GDPR will (all going to plan) become the 'applied GDPR' or 'UK GDPR,' being transferred under the authority of the European Union (Withdrawal) Act 2018 with the modifications listed in Schedule 1 of The Data Protection, Privacy and Electronic Communications (Amendments etc) (EU Exit) Regulations 2019.[32] At the time of writing, these modifications do not alter Articles to which this report relates, that is, the modifications mostly concern procedure and not substantive content on transparency, interpretability, and so on. Consequently, for our purposes, the GDPR and the UK GDPR are mostly the same as they relate to transparency and the supposed 'right to explanation.'

Section 1 key messages:

- **The GDPR is one source of regulation that might generate a duty of transparency, interpretability, or explainability but is no panacea.**
- **Any duty of transparency, interpretability, or explainability that the GDPR offers will be a data protection solution that seeks to protect data protection interests and values.**
- **Data is protected by a complex web of regulation in England and Wales. Notably, ICO has statutory authority over other law that often governs the same space. In parallel with data protection is a system of common law which includes duties of confidentiality as well as the tort of misuse of private information, and there may be other potential sector-specific regulation for AI in the future.**

2. GDPR basics

This section outlines the basics of the GDPR to orientate the reader. To analyse what tools the GDPR provides in regards to transparency, interpretability, or explainability, we must first understand some basics about the GDPR - its scope, the rights it confers, the duties it assigns, and the principles it contains.

a. Material scope

The GDPR has a material scope and territorial scope. The material scope - the subject matter to which the GDPR applies - is limited to the processing of personal data.[v]

> *Personal data* means 'any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly…'[33]

In short, to determine whether data is personal data or not, we consider whether an individual is identifiable using the means reasonably likely to be used.[34] This is a context-specific process and may be a difficult assessment to make when dealing with pseudonymised health research data.[35] We highlight both the importance of this assessment but also the vexed and context-specific nature of this judgment.

> *Processing* means 'any operation or set of operations which is performed on personal data or on sets of personal data.'

In short, it is difficult to envision an action that could be performed on personal data that would not count as 'processing.' Consequently, the collection and operations performed on personal data as a part of machine learning will likely count as 'processing' for the purposes of the GDPR.

One important consideration is how machine learning for healthcare and research might be caught within the material scope of the GDPR. In this way, there are a number of scenarios. First, with respect to training or test data, this data may have been anonymised and the data controller of this data may not be regarded as a processor or joint controller. In this scenario, the training and test data are likely beyond the scope of the GDPR and not governed by its provisions. Second, with respect to training and test data, often anonymisation to the standard required by the GDPR removes much of the richness of the data. Moreover, in some scenarios, it is important to retain the ability to return a diagnosis or finding back to a research participant. In both cases, the training and test data may remain personal data and so be within the scope of the GDPR. Third, machine learning models, depending on the application, require a set of inputs for the model to process for a particular instance. For example, a model

[v] There are a number of exceptions to the material scope. For example, the household activity exception found in Article 2(2)(c) GDPR.

to assess surgical risk might require inputs such as age, height, and BMI to provide the output of surgical risk for any given patient. In this way, data as an input for a particular instance of processing might also be caught by the GDPR. Broadly, it is important to distinguish between personal data processed as a part of training or test data and personal data as an input for an instance of processing. This is because the consequences may differ between the two. For instance, as we cover in Section 5(f)(i), it is not clear any 'decision' is necessarily being made where personal data are processed as training or test data. Consequently, the more stringent requirements found in Article 22(1) may not apply to training or test data where personal data is purely used to train or test a model. The analysis may be very different where personal data is input to be processed by the model to generate an output. In healthcare and where healthcare blends into research, often the entire point is to make what might count as a 'decision' with respect to a patient or consumer. In short, it is important not only to establish that data is within the material scope of the GDPR but what data is within scope and how.

b. Territorial scope

The territorial scope of the GDPR is governed by two sets of rules found in Article 3. It is important to consider the territorial scope of the GDPR as it demonstrates how those developing or offering machine learning for healthcare or research may be caught by the Regulation. Further, to establish the legal position of a data controller or processor, it is important not only to understand whether one is caught by the GDPR's territorial scope but also how the controller/processor is caught. Notably, the GDPR's territorial scope can extend beyond the EU (the Union) and European Economic Area (EEA) via Article 3(1) and (2).[vi]

Article 3 Territorial scope

1. This Regulation applies to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union, regardless of whether the processing takes place in the Union or not.

2. This Regulation applies to the processing of personal data of data subjects who are in the Union by a controller or processor not established in the Union, where the processing activities are related to:

    (a) the offering of goods or services, irrespective of whether a payment of the data subject is required, to such data subjects in the Union; or

    (b) the monitoring of their behaviour as far as their behaviour takes place within the Union.

[vi] N.B. 'Union' here means 'European Union' but also extends to the wider European Economic Area, see: EEA Joint Committee, 'Decision of the EEA Joint Committee amending Annex XI (Electronic communication, audiovisual services and information society) and Protocol 37 (containing the list provided for in Article 101) to the EEA Agreement [2018/1022]' (Decision) 154/2018.

We provide a brief summary of Article 3(1) and (2) and how each provision might apply to machine learning for healthcare or research below.

i. Article 3(1)

> **Article 3(1)** This Regulation applies to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union, regardless of whether the processing takes place in the Union or not.

In regards to Article 3(1), there are two main points of interpretative contention.

First, 'an establishment in the Union.' In this regard, under the DPD, the CJEU departed from a formalistic approach considering place of registration, ruling that 'establishment' means 'extends to any real and effective activity — even a minimal one — exercised through stable arrangements.'[36] Consequently, it is difficult to consider healthcare provision or a research activity that might escape the definition of 'establishment.'

Second, 'in the context of the activities of' has been interpreted by the EDPB so that the processing in question need not be carried out by the EU-established entity itself.[37] That is, an establishment inside the EU may not process data but its connection to a controller outside the EU that does process data may mean this processing is within the scope of the GDPR. The EDPB note that two factors might assist when questioning 'in the context of the activities of': the relationship between a data controller (or processor) outside the EU and its local establishment in the EU as well as revenue raising in the EU. Both of these elements may assist in considering whether the processing activities are 'inextricably linked' to the EU establishment.[38]

The upshot of Article 3(1) is that processing may be caught, even if it takes place outside the EU. Consider Example 5 of the operation of Article 3(1) given by the EDPB:

> 'Example 5: A pharmaceutical company with headquarters in Stockholm has located all its personal data processing activities with regards to its clinical trial data in its branch based in Singapore. In this case, while the processing activities are taking place in Singapore, that processing is carried out in the context of the activities of the pharmaceutical company in Stockholm i.e. of a data controller established in the Union. The provisions of the GDPR therefore apply to such processing, as per Article 3(1).'

This example should give those working in machine learning for healthcare or research pause for thought. Application of Article 3(1) may be problematic in these circumstances. For instance, consider two examples.

> Example A: a multinational developer headquartered in the EU trains their machine learning model outside the EU using datasets of personal data from citizens outside the EU. The machine learning system is sold in the EU through the EU-established entity.
>
> Example B: a research institution headquartered in the EU conducts research as a part of a consortium outside of the EU. The research consortium is funded by EU grants. The machine learning model is trained on datasets of personal data of citizens outside the EU. The model directly benefits research programmes and studies in the EU-established research institution.

Both of these examples are relatively close to EDPB's Example 5 - it is unclear whether they might be distinguished. Indeed, these examples all consider territorial scope via establishment and context of activities only, as Article 3(2) makes clear, territorial scope also extends beyond establishing a related EU establishment.

ii. Article 3(2)

Article 3(2) This Regulation applies to the processing of personal data of data subjects who are in the Union by a controller or processor not established in the Union, where the processing activities are related to:

(a) the offering of goods or services, irrespective of whether a payment of the data subject is required, to such data subjects in the Union; or

(b) the monitoring of their behaviour as far as their behaviour takes place within the Union.

In regards to Article 3(2), this Article makes clear that even if there is no establishment in the EU, the offering of goods or services or monitoring behaviour of subjects of the EU are within the territorial scope of the GDPR. We outline each requirement in turn. Both of these requirements are couched by the EDPB as 'targeting' of EU subjects.[39]

1. Offering of goods or services, irrespective of payment

Article 3(2)(a) stipulates that where processing activities relate to 'the offering of goods or services, irrespective of whether a payment of the data subject is required to such data subjects in the Union' the GDPR applies. For our purposes, the 'irrespective of whether payment is required' clause is critical since research institutions and studies that target Union subjects may well be caught. Recital 23 clarifies the extent of Article 3(2)(a) offering us two clarifications. First, there is clarification over what would be insufficient:

> 'whereas the mere accessibility of the controller's, processor's or an intermediary's website in the Union, of an email address or of other contact details, or the use of a language generally used in the third country where the controller is established, is insufficient to ascertain such intention [to offer goods]'

Second, ideas of what might be sufficient to count as 'offering of goods or services:'

> 'in order to determine whether such a controller or processor is offering goods or services to data subjects who are in the Union, it should be ascertained whether it is apparent that the controller or processor envisages offering services to data subjects in one or more Member States in the Union… factors such as the use of a language or a currency generally used in one or more Member States with the possibility of ordering goods and services in that other language, or the mentioning of customers or users who are in the Union, may make it apparent that the controller envisages offering goods or services to data subjects in the Union.'

In this way, if the processing of Union subject personal data is incidental and not envisioned, the processing may be beyond the territorial scope of the GDPR.

Translated into the healthcare context, developers offering machine learning as software, as an application, or as a service may well be caught, depending on how they position their good or service. Translated to health research, studies and consortia may be caught, depending on how the scope of the study is construed and communicated. For instance, if we replace 'customer' with 'research participant' in Recital 23, many of the same criteria may equally apply to the research context.

2. Monitoring of behaviour

Article (3)(2)(b) stipulates that where processing activities are related to the 'monitoring of their [data subjects] behaviour as far as their behaviour takes place within the Union,' the GDPR will apply. EDPB in *Guidelines 3/2018 on the territorial scope of the GDPR* break this judgment into two, cumulative parts.[40] First, the behaviour monitored must relate to a data subject in the EU. Second, the monitored behaviour must take place within the Union. Recital 24 offers some clarification over these two criteria:

> '... to determine whether a processing activity can be considered to monitor the behaviour of data subjects, it should be ascertained whether natural persons are tracked on the internet including potential subsequent use of personal data processing techniques which consist of profiling a natural person, particularly in order to take decisions concerning her or him or for analysing or predicting her or his personal preferences, behaviours and attitudes.'

The EDPB notes that this thinking, while directed toward tracking via the internet, might also be applied to other types of technology or networking.[41] Further, the EDPB also highlights that 'monitoring' implies that the controller or processor has a specific purpose in mind - the mere fact of online collection or analysis of EU subject personal data is insufficient.[42] Finally, the EDPB also provides a number of examples of monitoring that might trigger Article 3(2)(b). Of interest to the machine learning for healthcare or research context are the examples of behavioural advertisement, online personalised diet and health analytics services, and monitoring or regular reporting on an individual's health status, such as Example 20:[43]

> 'Example 20: A US company has developed a health and lifestyle app, allowing users to record with the US company their personal indicators (sleep time, weight, blood pressure, heartbeat, etc…). The app then provide users with daily advice on food and sport recommendations. The processing is carried out by the US data controller. The app is made available to, and is used by, individuals in the Union. For the purpose of data storage, the US company uses a processor established in the US (cloud service provider)
>
> To the extent that the US company is monitoring the behaviour of individuals in the EU, in operating the health and lifestyle app it will be 'targeting' individuals in the EU and its processing of the personal data of individuals in the EU will fall within the scope of the GDPR under Art 3(2).'

Given Recital 24, EDPB commentary and examples, it appears that machine learning for healthcare and research when delivered as an application or software as a service will be especially vulnerable to being within the territorial scope of the GDPR, despite a lack of established presence within the jurisdiction. Consequently, the rights the GDPR provides to data subjects and the correlated duties it requires of controllers may extend beyond the EU and EEA.

c. Rights, duties, principles

The GDPR confers rights upon data subjects, assigning correlated duties and principles to data controllers or processors. What do these three terms mean?

> A *data subject* is a natural (*not* a legal or corporate) person who is identified or identifiable by information.[44] That is, a person to whom the personal data relates.
>
> A *data controller* (hereafter 'controller') is a person or body that alone or jointly 'determines the purposes and means of the processing of personal data.'[45] In this regard, data controllership is context specific and considers control over, authority over, and responsibility for processing of personal data - there is no requirement that the entity itself processes or handles the data in question.
>
> A *data processor* is a person or body that 'processes personal data on behalf of the [data] controller.'[46] In other words, a data processor processes personal data but does not determine the purposes of processing that data. For the sake of simplicity, this report primarily considers the obligations of controllers. However, much of the analysis may also apply to processors.[vii]

[vii] N.B. The broad responsibilities for controllers and joint controllers are found in Article 24 and 26 GDPR respectively. To compare the broad responsibilities of processors, see Article 28 GDPR.

The GDPR contains a number of rights, however none of these rights are conferred upon all data subjects without qualification. Rather, these rights only trigger in certain circumstances or are subject to heavy exceptions (often found in Member State law). Broadly, these data subject rights include: the rights to information, of access, rectify, move, or erase data, as well as to restrict, object to data processing, and other restrictions such as those relating to automated individual decision-making.[47] Much of this report considers under what circumstances these rights arise, what they require, and how they might be marshalled to provide a right to transparency, interpretability, or explanation.

It is the controller's responsibility - their duty - to comply with these rights.[48] Further, it is also their responsibility to comply with a number of principles when processing personal data.[49] These principles both assist when interpreting associated rights but also constitute their own grounds by which enforcement actions may be directed.[50] That is, controllers may be fined and be otherwise liable for breaches of principle not just breaches of specific data subject rights.[51] This is an important interpretative point: an action may breach a specific right or fall foul of a specific prohibition but ultimately the more amorphous principle of transparent processing might itself also ground a claim. We discuss this possibility further in Section 4(a) below.

Data processing principles dictate that personal data shall be:

a) Processed lawfully, fairly and in a transparent manner in relation to the data subject;[52]
b) Collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes;[53]
c) Adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed;[54]
d) Accurate, and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay;[55]
e) Kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed;[56]
f) Processed in a manner that ensures appropriate security of the personal data.[57]

One key principle is found in a) - the principle of lawful processing. Lawful processing includes a number of considerations but always starts with identifying a lawful basis for processing. Lawful bases are found in Article 6 of the GDPR. The most commonly relied upon lawful bases for health care or research in the UK include public interest,[58] legitimate interests,[59] and vital interests.[60]

The GDPR also contains further provisions for processing of 'special category data.' Relevant to machine learning for healthcare, special category data includes the processing of genetic data, biometric data (for the purpose of uniquely identifying a natural person), and data concerning health, as well as data revealing racial or ethnic origin or data concerning a person's sex life or sexual orientation.[61] The definition of each can be found below.

> *Data concerning health* means 'personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status.'[62]
>
> *Genetic data* means 'personal data relating to the inherited or acquired genetic characteristics of a natural person which give unique information about the physiology or the health of that natural person and which result, in particular, from an analysis of a biological sample from the natural person in question.'[63]
>
> *Biometric data* means 'personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images…'[64]

Lawful processing of special category data is subject to two key restrictions. First, processing of special category data is only lawful if a derogation (an exception) found in Article 9(2) applies.[viii] Commonly relied upon derogations for healthcare or research include: preventative or occupational medicine,[65] public health,[66] or research purposes.[67] Second, many derogations allow controllers to process special category data but only when such processing is supported by appropriate safeguards and supported by Member State law.[68] Given these two restrictions, special category data is subject to further restrictions and safeguards to establish lawful processing.

Section 2 key messages:

- **The GDPR is limited by its material scope: 'personal data.' It is important to consider how machine learning might be caught as personal data, distinguishing between training/test data as personal data and data used as an input to a model as personal data.**
- **The GDPR is limited by its territorial scope. First, what counts as 'processing of personal data in the context of the activities of an establishment' is likely broad, potentially including training/test datasets outside the Union if the models are eventually sold in the EEA. Second, where not established in the Union, provisions relating to 'offering of goods or services' and 'monitoring' are expansive. In this way, no money need be exchanged to count as 'offering of goods or services' and EDPB examples of 'monitoring' including applications that use personal predictors to provide personal recommendations.**
- **Broadly, the GDPR furnishes data subjects with data subject rights, assigning correlated duties to controllers and processors. It is the controller's**

[viii] Note that some commentators argue that Article 9 derogations exclude the need for Article 6 legal bases, see: Dove ES. The EU General Data Protection Regulation: implications for international scientific research in the digital era. *The Journal of Law, Medicine & Ethics*. 2018; 46(4): 1020.

Molnár-Gábor F. Germany: A fair balance between scientific freedom and data subjects' rights?. *Human genetics*. 2018; 137(8): 620.

Article 29 Data Protection Working Party. *Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC*. 2014: 14-15.

responsibility to ensure these rights are complied with and data protection principles are upheld. Further, in the context of healthcare and research, 'biometric data', 'genetic data', and 'data concerning health' all count as special category data and are subject to special restrictions and safeguards.

This report considers how the GDPR's rights and principles may be marshalled to provide transparency, interpretability, or explainability to machine learning in healthcare and research. This analysis makes the following assumptions:

a) **The data processed is within the material scope of the GDPR**
b) **The data controller or processor is within the territorial scope of the GDPR**
c) **The data processed may constitute special category data**

3. General principles, particular rights

The GDPR provides a number of tools that might generate a duty to make transparent, interpretable, or explainable machine learning models. Broadly, we should distinguish between:

a) The general principle of transparent processing;
b) How this principle supports other data subject rights; and
c) How the general principle and associated rights interact with the automated individual decision-making requirements.

It is important to emphasise that these three elements are intimately related. The general principle of transparent processing helps us interpret specific data subject rights, just as specific data subject rights add flesh to the bones of an otherwise vague principle. Further, automated processing requirements do not sit alone but interact with specific data subject rights and in turn ought to be interpreted in light of the general principle of transparent processing. Accordingly, principles, specific rights, and specific requirements are a part of tiered structure. One may claim a breach of principle, breach of a specific data subject right, or breach of automated processing requirements - each relates to the other but constitutes a distinct way to ground a claim. We examine each claim in the context of machine learning for healthcare and research asking:

a) Does the general principle of transparency require interpretability or explainability?
b) Do the specific data subject rights individually or collectively together with the principle of transparency constitute a right to interpretability or explainability?
c) Do the automated individual decision-making requirements generate a duty to render interpretable or explainable machine learning models?

Further, we consider the collective effect of these claims; whether the sum of these claims is a duty of interpretability or explainability for machine learning in the context of healthcare and research. This allows us to provide a comprehensive reply to the question of what tools the GDPR provides in regards to transparency, interpretability, or explainability.

A Salient Feature | Roundtable 3

Some participants at Roundtable 3 thought it important to distinguish between provisions that lend themselves to transparency and those provisions directed toward explainability. A duty of explainability may require something very different to the general principle of transparency.

Section 3 key messages: there are three interrelated claims that may be marshalled to generate a duty of transparency, interpretability, or explainability:

I. The general principle of transparency; and
II. How this principle interacts with specific data subject rights; and
III. Automated individual decision-making requirements.

4. The general principle of transparency and associated rights

The GDPR contains a general principle of transparent processing. This principle works through and is instantiated by particular data subject rights and specific requirements regarding automated processing. For our purposes, we ask whether this general principle and its interaction with other data subject rights might reasonably be interpreted to constitute a duty of interpretability or explainability?

a. The general principle of transparent processing

The GDPR contains a general principle of transparency. As Working Party 29 (WP29) Guidelines note,[ix] [69] this is a common feature of EU law, the foundational Treaty on European Union (TEU) and the Treaty on the Functioning of the European Union (TFEU) making multiple references to transparency, openness, transparent dialogue, and rights of access to documents of EU institutions.[70] Article 5(1)(a) GDPR contains the core of the principle of transparent processing ('the principle'):

> Article 5(1) Personal data shall be:
>
> (a) processed lawfully, fairly and in a transparent manner in relation to the data subject

Recitals 39, 58, and 60 all add colour to what this principle requires, both in the kind of information that the principle of transparent processing might require a controller to disclose and the manner in which this information should be communicated. We address these clarifications below.

Transparency under the GDPR is generally interpreted as being a 'user-centric rather than legalistic' concept.[71] WP29 in their *Guidelines on Transparency under Regulation 2016/679* highlight the many dimensions of transparency and its importance, noting that transparency is about engendering trust,[72] is an expression of fairness,[73] and is intrinsically linked to the principle of accountability.[74] Further, it is also noted that transparency requirements exist throughout the life cycle of data processing and irrespective of the legal basis relied upon. WP29 emphasise that the quality, accessibility, and comprehensibility of the transparency information is just as important as the content of information communicated to data subjects.[75] At the same time, WP29 also note the great practical importance of complying with the notification duties of controllers found in Article 12-14.[76] In this way, the provision of basic information like the identity of the controller and the legal basis for processing are not to be overlooked. As a result, two general questions arise when addressing the general principle of transparent processing:

[ix] These WP29 Guidelines have now been adopted by the EDPB on the 29th of November 2017.

I. Has the controller communicated the correct kinds of and sufficient information to comply with the principle of transparent processing? A question of substance; and
II. Has the controller communicated this information in the right manner? A question of form.

Broadly, on the manner in which this information should be communicated, Recital 39 notes that all information and communications should be easily accessible, easy to understand, and use clear, plain language. Recital 58 adds that information addressed to the public or data subjects should also be concise, and, where appropriate, use visualisation. Additionally, Recital 60 notes that information may be provided using standardised icons to give a meaningful overview of intended processing in an 'easily visible, intelligible and clearly legible manner'. Finally, Article 12(1) places special emphasis on communication of information when complying with data subject rights under Articles 15-22, and 34 GDPR, echoing many of clarifications offered in Recitals 39, 58, and 60.

On the content to be provided, many of the data subject rights, especially the rights to information (Article 13 and 14) and access (Article 15), have specific requirements as to what information must be provided to data subjects. In this regard, compliance with data subject rights tells us much of the information that should be provided to comply with the general principle of transparent processing. This is the topic of the next section. However, there are two points of caution.

First, it is open to national supervisory authorities, national courts, and the Court of Justice of the European Union (CJEU) to take a broad purposive approach when considering the content of information to be provided.[77] That is, they may simply ask what information is required to vindicate data subject rights, enforce accountability, and encourage transparency in the round and act accordingly. In this way, data protection authorities (especially the CJEU) will be particularly concerned to uphold the principles and integrity of the GDPR and not be overly concerned whether the information concerned is listed in Articles 13-15. Indeed, the GDPR over and above the DPD strengthens accountability, noting that controllers are responsible for compliance with data protection principles and must demonstrate such compliance.[78]

Second, supervisory authorities often target their enforcement actions against failures to both comply with specific data subject rights but also violation of data protection principles generally. Indeed, many of the fines issued by supervisory authorities so far have been based on a breach of Article 5 general principles.[79] [80] Given this, while controllers should communicate specific information required by Articles 13-15, and facilitate other data subject rights found in Articles 16-22, it is important to stand back and consider compliance with the spirit of the Article 5 principles.

A Salient Feature | Interviews

Interviewees and multiple roundtable participants highlighted the importance of interpreting transparency requirements according to their purpose and not imposing an overly restrictive, mechanistic reading of the GDPR's principle of transparent processing.

Finally, it should also be emphasised that the principle of transparent processing is not an assessment to be made in isolation from other principles. For instance, transparency is stated alongside the principles of lawfulness and fairness of processing. This is fitting as judgments of what information to disclose and how to disclose this information should be coloured by judgments of what is lawful and fair. For instance, Recital 60 emphasizes this point:

> 'The controller should provide the data subject with any further information necessary to ensure fair and transparent processing taking into account the specific circumstances and context in which the personal data are processed.'

Indeed, this is reflected in WP29, EDPB, and ICO's position that transparency is not formulaic but a context-sensitive judgment.[81] [x] WP29 also emphasise a similar point in their Guidelines, extending such thinking to the consequences of processing:

> 'A central consideration of the principle of transparency outlined in these provisions is that the data subject should be able to determine in advance what the scope and consequences of the processing entails and that they should not be taken by surprise at a later point about the ways in which their personal data has been used.'[82]

The interpretation of data subject rights is coloured by these ideas of being user-centric, context-sensitive, and that there are inherent links between the different principles, such as transparency and fairness.

b. The role of transparent processing in upholding associated rights

In relation to some data subject rights, transparency places a 'triple obligation' upon controllers, they must:[83]

I. Comply with the principle of transparency when communicating with data subjects (as mentioned above), elaboration of this being found in Article 12(1); and
II. Provide information to data subject on their rights following Article 13(2)(b) and 14(2)(c) (provisions of the rights to information); and
III. Facilitate the exercise of data subject rights following Articles 15-22.

We have covered the first obligation above - this obligation concerns the manner and form by which information should be communicated. We now consider the remaining two obligations below.

Arguably, transparency's role in vindicating data subject rights is the principle's *principal* function. Indeed, under the DPD, the Opinion of Advocate General Villalón in *Smaranda Bara v Președintele Casei Naționale de Asigurări de Sănătate* (C-201/14) noted:

[x] N.B. The UK ICO's Project Explain is instructive, providing further details on this approach to transparency. See: Information Commissioner's Office. *Project ExplAIn: Explaining Decision Made with AI*. 2020, 33-37

> 'the requirement to inform the data subjects about the processing of their personal data, which guarantees transparency of all processing, is all the more important since it affects the exercise by the data subjects of their right of access to the data being processed, referred to in Article 12 of Directive 95/46, and their right to object to the processing of those data, set out in Article 14 of that directive.'[84]

In other words, transparent processing ensures that data subjects are in a position to uphold their data subject rights. If data subjects do not know their personal data is processed, they are in no position to enforce any right. If data subjects do not know what information is processed, they do not know whether the information requires rectification or not. Critically for our purposes, if data subjects do not know how the decision was arrived at, it is difficult to see how they might successfully contest the decision. As a consequence, Article 12(1) and (2) make plain that controllers must take 'appropriate measures' to provide information outlined in Articles 13 and 14 and facilitate the exercise of data subject rights found in Articles 15-22.

The following sections outline the relevant data subject rights, when these rights are available to data subjects, how the principle of transparent processing interacts with each right, and whether the right might generate a duty of interpretability or explainability. Broadly, we ask both what these rights require but whether the principle of transparency when combined with certain data subject rights require some kind of interpretability or explainability.

We provide a typology for how data subject rights might be marshalled to generate a duty of interpretability or explainability, distinguishing between data subject rights that:

I. Require disclosure of information that might constitute interpretability or an explanation. Candidates for this type include:
    A. The rights to information; and
    B. The right of access; and
    C. The right to data portability.

II. While not being directed toward the provision of information itself, to be vindicated may require some interpretability or explanation. Candidates for this type include:
    A. The right to rectification; and
    B. The right to object; and
    C. The right to erasure; and
    D. The right to restriction of processing.

With respect to II. the rights to erasure and restriction of processing seem less pertinent to the following discussion, our preliminary analysis indicating that these rights offer little more by way of interpretability or explanation. However controllers should note that apart from interpretability, these rights may be of critical operational importance.

i. Restricting data subject rights

All data subject rights are subject to some kind of qualification, restriction, or derogation. Rather than speaking in general terms about data subjects possessing various rights, it is better to talk about these rights in context: when these rights are triggered and under which circumstances the right is available to data subjects, exempted, or restricted. That is, asking what a data subject right requires in general is very different to asking what the right requires

in the context of healthcare and research. Understanding the contextual limitations of these data subject rights allows us to better grasp the character of these rights and their purpose. Further, considering these rights in their context allows us to ask how useful these rights are in healthcare and research. The derogations, exemptions, or restrictions that apply differ according to the right in question. Nevertheless, there are three broad scenarios in healthcare and research that limit many of the relevant data subject rights and in turn limit whatever duty of interpretability or explainability they might contain.

1. Processing that does not require identification

Article 11 and Article 12(2) both contain caveats regarding the identifiability of data subjects.

Article 11(1) contains the general rule that if a controller's purpose for processing data no longer requires identification of data subjects, they shall *not* be obliged to 'maintain, acquire or process' additional information to identify the data subject 'for the sole purpose of complying' with the GDPR.

Article 11(2) specifies that where a controller is:

I. Able to demonstrate that they are 'no longer in a position to identify that data subject';[xi] and
II. The data subject does not provide additional information enabling their identification; then

Certain data subject rights do not apply.

Indeed, Article 12(2) clarifies that while controllers should facilitate the exercise of data subject rights, where the requirements of Article 11(2) are met, data subject rights found in Article 15-22 do not apply. What rights are found in these Articles? Data subject rights found in these Articles include:

I. Right of access
II. Right to rectification
III. Right to erasure
IV. Right to restriction of processing
V. Notification duties regarding rectification, erasure, and restriction; and
VI. Right to data portability
VII. Right to object
VIII. Restrictions over automated individual decision-making

[xi] N.B. The controller must also notify the data subject that they are no longer in a position to identify the data subject.

For our purposes, notably absent from this list are the rights to information (Articles 13-14). Otherwise, the caveat regarding identifiability applies to all data subject rights analysed for the remainder of this report.

The assessment of what counts as 'no longer in a position to identify that data subject' is likely to be a vexed technical and legal question, especially in the health research context where data are often pseudonymised.[85] Moreover, it is not uncommon in health research contexts - especially in the field of genetics - for multiple entities to control the means and processing of personal data, that is, to have joint controllers.[86] Accordingly, the identifiability assessment with respect to data subject rights becomes a knottier problem. Complexity aside, those that rely on research purposes are specifically told to respect the principle of data minimisation and use measures like pseudonymisation where possible.[87] The research context is therefore one particular context that might lend itself to invoking Article 11 and 12(2) as a shield to data subject rights. However, depending on context, *health* research in particular may not lend itself to successful reliance on Article 11 and 12(2). In health research, especially genomics, there are often countervailing research ethics reasons to keep the data subject identifiable - for example, duties to recontact, to update individuals with new, clinically significant information relating to diagnosis or testing.[88] Where research blends into healthcare, the more likely identifiability will be necessary and the less likely successful reliance on Articles 11 and 12(2) will become.

To summarise, whether a controller can successfully invoke Articles 11 or 12(2) involves a vexed assessment of whether they are in a position to identify the data subject. The possibility of successfully relying on Article 11 and 12(2) is a live option in the research context, although it becomes less likely where research becomes enmeshed with treatment of research participants. If Article 11 and 12(2) are successfully relied upon, the controller may successfully block the exercise of all data subject rights apart from the rights to information.

2. Article 23 Member State restrictions

Some restrictions to data subject rights attach to processing in specific contexts. Article 23(1) GDPR provides scope for Union and Member State law to restrict by way of legislation data subject rights and processing principles in certain circumstances.[xii] Restrictions are permissible insofar as they respect 'the essence of the fundamental rights and freedoms' and constitute a necessary and proportionate measure in a democratic society to safeguard.[89] Following Article 23(1), the DPA 2018 lays down certain restrictions in Schedule 3, Part 2 for data concerning health.

Schedule 3, Part 2 restricts the application of specified data subject rights in relation to health data. All but one data subject right is restricted in some way by this Schedule, the missing right being the restrictions on automated individual decision-making in Article 22.[90]

The restrictions found in Schedule 3, Part 2 seek to preserve the framework that governed health data under the previous DPD and DPA 1998, this framework primarily being found in

[xii] N.B. It is unclear which ground in Article 23(1) GDPR Schedule 3, Part 2 DPA 2018 purports to rely on - perhaps the most plausible candidates are Article 23(1)(e), (h), and at points (i) GDPR.

The Data Protection (Subject Access Modification)(Health) Order 2000.[91] We examine how this framework operates with respect to individual data subjects below. Schedule 3, Part 2 restricts the exercise of data subject rights in three main circumstances:

I. Where the serious harm test is met (only in relation to the right of access), namely that the application of Article 15 of the GDPR to the data 'would be likely to cause serious harm to the physical or mental health of the data subject or another individual' (DPA Schedule 3, Part 2, Section 2(2))[92]
II. Where data is processed by courts[93]
III. Where a request is made by a person with responsibility over the data subject and this request contravenes the expectations or wishes of the data subject[94]

With regard to II. health data processed by a court or supplied in evidence to a court according to the rules listed in Schedule 3, Part 2, Section 3(2) are exempt from the listed data subject rights. The effect of this with respect to data concerning health is that most court-related processing is shielded from most data subject rights in this context. Where III. applies, there are special exemptions where a third party has parental responsibility over a person under 18 or responsibility over a person who is incapable of managing their own affairs. Section 4(2) restricts data subject rights insofar as they would disclose information that was given in the expectation that the information would not be further disclosed either to the person making the request, generally through the consent process, or where the data subject expressly indicated they did not want the data disclosed.[95] In short, Section 4(2) allows vulnerable parties to disclose data concerning health and have this data shielded from their guardian where they make this wish clear. We address I, that is, the serious harm test with respect to the right of access in Section 4(b)(iii)(2) below.

3. Article 89 research purposes derogations

In the context of health research, controllers will commonly rely on the Article 9(2)(j) derogation to process special category data for research purposes.[96] This is for two reasons.

First, as mentioned earlier, processing of special category data requires an Article 9 derogation for lawful processing. Health-related data, genetic data, and biometric data all count as 'special category data.'[97] The research purposes derogation is sometimes a natural fit for non-commercial health-related research projects, as many other derogations carry with them significant disadvantages. In short, sometimes selecting a derogation is less selecting the ideal option but picking what is available given restrictions and what derogation's rigours can be feasibly complied with.

Second, reliance on the Article 9(2)(j) derogation requires the research to be in accordance with the requirements of Article 89(1). However, where this is the case, Article 89(1) provides for further flexibility. The requirements of Article 89(1)-(2) and the flexibility it provides for research are considered below.

Article 89(1) GDPR provides scope for Union and Member State law to restrict by way of legislation certain data subject rights. Derogations may be provided for via legislation:

I. So long as appropriate safeguards are in place, in particular, procedures to ensure the principle of data minimisation is respected, for example, pseudonymisation;[98] and
II. Insofar as the data subject right would 'render it impossible or seriously impair the achievement of the specific purposes' of processing.[99]

The GDPR treats differently data processed for a) 'scientific or historical research purposes or statistical purposes'[100] and b) archiving in the public interest.[101] We are primarily interested in the former but it is important to note that reliance on archiving allows Member States to derogate from more data subject rights than research purposes. Research purposes provide the opportunity for Member States to derogate from the following data subject rights:[102]

I. Right of access
II. Right to rectification
III. Right to restriction of processing
IV. Right to object.

In regards to Article 89, research purposes allow derogation from the listed rights subject to the caveat that the right would otherwise likely render impossible/severely impair the purposes of processing and that the derogation is accompanied by appropriate safeguards.[103] The DPA 2018 provides for such derogations from the rights in Schedule 2, Part 6, Section 27, stating that the rights do not apply to processing for research purposes insofar as:

I. The right would 'prevent or seriously impair the achievement of the purposes in question;' and
II. The data is processed in accordance with Article 89(1) and its elaboration in Section 19 DPA 2018.

In regards to point II, Section 19 notes that to meet the Article 89(1) requirements, the research must:

A. *Not* be likely to cause substantial damage or distress to a data subject; and
B. If the processing is carried out for the 'purposes of measures or decisions with respect to a particular data subject' the processing will not meet the requirements of Article 89(1), *unless* the purposes are 'approved medical research.'

Let us take each in turn.

The upshot of A is that reliance on the Article 89 research exemption is only possible where the processing is not likely to lead to substantial damage or distress to the data subject. In the medical or health research context, the risk of substantial damage or distress is likely to be heightened. However, this provision regarding damage and distress was the same under the Data Protection Act 1998 (DPA 1998) that implemented the previous DPD.[104] Given this, research that was compliant under the old regime in relation to distress and damage should be compliant under the new regime of GDPR and DPA 2018.

To examine B, let us break the provision into its constituent parts.

First, the reference to 'purposes of measures or decisions with respect to a particular data subject.' The DPA 1998 also included reference to data not being processed 'to support measures or decisions with respect to particular individuals.'[105] The explanatory notes to the DPA 2018 note that Section 19 replicates the safeguards under Section 33 of the 1998 Act. In this way, the interpretation under the 2018 Act may be the same. Despite interpretive continuity, it is inherently unclear what will count as processing for 'purposes of measures or decisions with respect to a particular data subject.' Nevertheless, it does seem clear from our analysis of where machine learning is being used in healthcare and research (see the Machine Learning Landscape report) that machine learning will increasingly be used to support measures or decisions with respect to particular individuals. Indeed, if machine learning often counts as 'personalised medicine', frequently the technology will tend to be directed toward particular data subjects.[106] However, contrary to this is the rise in 'operational uses' for machine learning. For instance, machine learning for scheduling, rotas and so on. These kinds of uses, depending on their context, may avoid being classified as processing that counts as 'purposes of measures or decisions with respect to a particular data subject.'

Second, the addition of 'unless the purposes for which the processing is necessary include the purposes of approved medical research.' The explanatory notes of the DPA 2018 explain that the DPA 2018 replicates this 1998 provision.[107] However, this is not strictly true, as the 2018 Act differs from the 1998 Act and its supporting statutory instrument the Data Protection (Processing of Sensitive Personal Data) Order 2000 by including the clause '*unless* [my emphasis] the purposes for which the processing is necessary include the purposes of approved medical research.'[108] This aside, 'approved medical research' in Section 19 is given the following definition under the DPA 2018:[109]

> '"approved medical research" means medical research carried out by a person who has approval to carry out that research from—
>
> (a) a research ethics committee recognised or established by the Health Research Authority under Chapter 2 of Part 3 of the Care Act 2014, or
>
> (b) a body appointed by any of the following for the purpose of assessing the ethics of research involving individuals—
>
> (i) the Secretary of State, the Scottish Ministers, the Welsh Ministers, or a Northern Ireland department;
>
> (ii) a relevant NHS body;
>
> (iii) United Kingdom Research and Innovation or a body that is a Research Council for the purposes of the Science and Technology Act 1965;
>
> (iv) an institution that is a research institution for the purposes of Chapter 4A of Part 7 of the Income Tax (Earnings and Pensions) Act 2003 (see section 457 of that Act);
>
> "relevant NHS body" means—

> (a) an NHS trust or NHS foundation trust in England, (b) an NHS trust or Local Health Board in Wales,
>
> (c) a Health Board or Special Health Board constituted under section 2 of the National Health Service (Scotland) Act 1978,
>
> (d) the Common Services Agency for the Scottish Health Service, or
>
> (e) any of the health and social care bodies in Northern Ireland falling within paragraphs (a) to (e) of section 1(5) of the Health and Social Care (Reform) Act (Northern Ireland) 2009 (c. 1 (N.I.)).'

The consequence of the above, is that there is often a rigorous process to be declared 'approved medical research.' Research emerging from universities will often be familiar with such processes. It is less clear how familiar commercial bodies from other sectors are with these procedures. For instance, large technology bodies entering the healthcare space may be less accustomed to research ethics committees and their requirements.

Schedule 2, Part 6 of the DPA 2018 provides exceptions to the listed data subject rights along the lines of Article 89(2) GDPR, adding in Section 27(3) that the exemptions to the right of access only apply so long as the results of research do not identify data subjects. In short, the DPA 2018 states Article 89(2), merely adding one caveat related to identifiability. We consider how Article 89 impacts upon data subject rights in the context of research with respect to each right below.

Data subject rights might therefore be restricted in a number of ways. As we have seen, if the controller is no longer in a position to identify the data subject and the data subject does not provide additional information to enable their identification, then all data subject rights apart from the rights to information do not arise. However, additional restrictions may apply. As discussed in Section 4(b)(i)(2), health data is subject to specific restrictions in UK law: specific restrictions attach to data subject rights in the context of health data if disclosure of data meets the serious harm test, is processed by a court, or where data is requested by proxy but is contrary to the data subject's expectations and wishes. It is important to keep in mind the totality of restrictions and derogations when considering whether data subject rights generate a duty of interpretability or explainability as the data subject right in the context of healthcare or research may be blocked, derogated from, or limited.

4. Trade secrets and intellectual property

Another notable restriction on data subject rights relates to trade secrets and intellectual property. This restriction found in Recital 63 primarily relates to the right of access, the Recital noting:

> 'That right [the right of access] should not adversely affect the rights or freedoms of others, including trade secrets or intellectual property and in particular the copyright protecting the software.'

Recital 63 gives particular emphasis to copyright protections of software. Accordingly, this restriction is especially important in relation to duties of transparency or interpretability as they apply to machine learning.

There are two main points to consider in relation to the restriction.

First, the restriction primarily arises in relation to the right of access. However, the restriction likely applies more generally to other rights, for instance, the rights to information, especially the requirement to provide 'meaningful logic' under Articles 13(2)(f), 14(2)(g) and 15(1)(h). Indeed, there is likely scope for Member States to introduce further provisions to secure intellectual property and trade secrets under Article 23(1)(i), these considerations likely being included as 'rights and freedoms of others.' Currently, the DPA 2018 does not provide specific provisions to protect intellectual property or trade secrets against the operation of data subject rights. In short, controllers should keep in mind that trade secrets and intellectual property may act to restrict the operation of some data subject rights. Controllers should be especially wary of disclosing trade secrets or intellectual property of third parties when complying with data subject rights.

Second, the restriction does not act to entirely block the operation of data subject rights. Recital 63 clarifies:

> 'However, the result of those considerations should not be a refusal to provide all information to the data subject. Where the controller processes a large quantity of information concerning the data subject, the controller should be able to request that, before the information is delivered, the data subject specify the information or processing activities to which the request relates.'

In short, trade secrets or intellectual property likely act to temper disclosure, not block disclosure. If a request of access would necessarily reveal such information, the likely response is to ask for clarification and specification of the request, allowing disclosure that does not divulge trade secrets and other intellectual property.

We now turn to data subject rights that influence any duty of transparency that emerges or that can be used to formulate a duty of interpretability or explainability for machine learning in healthcare or research.

ii. Rights to information

The rights to information might ground a duty of transparency, interpretability, or explainability. The following provides a description of the rights, when in time they trigger, what processing triggers the rights, what the rights require, and how these requirements might apply to machine learning for healthcare and research.

1. Timing of the rights to information

Articles 13 and 14 contain the rights to information. The rights to information apply:

I. 'At the time when personal data are obtained' (Article 13);[110] or
II. Where the personal data has *not* been obtained from the data subject (Article 14):
    A. within 'a reasonable period' after obtaining the personal data (not exceeding one month), or
    B. when the personal data is used to communicate with the data subject, or
    C. where disclosure to another recipient is envisaged; [111] or
III. Where the information already communicated changes and this change is a fundamental change to the nature of processing, the controller should notify the data subject of these changes.[112]

In this way, the rights to information are triggered at the outset of processing when the controller obtains the personal data and also when this information changes.

2. What processing triggers the rights to information?

Unlike many of the other data subject rights, there are comparatively few exceptions to the rights of information. The only exceptions to Article 13 being if the data subject already has the information or if there is a special Member State exception legislated for under Article 23(1).[113] However, Article 14 (where information has *not* been obtained from a data subject) contains three further exceptions. We address Article 23 - exceptions and restrictions that apply to both Articles 13 and 14 - and then restrictions and exemptions that apply solely to Article 14 in turn.

Article 23 exceptions and restrictions

As noted above, Article 23 allows Member States certain restrictions or exceptions to data subject rights. Schedule 3, Part 2 of the DPA 2018 lays down such restrictions for 'health data.' The rights to information under Article 13(1)-(3) and Article 14(1)-(4) are explicitly included in the GDPR provisions that may be restricted by Schedule 3.[114] For our purposes, Part 2 does not contain many relevant restrictions to the rights to information that concern machine learning. To summarise the restrictions, the dual rights may be restricted where the data is processed by courts, or where disclosure is contrary to the wishes of children or those incapable of managing their own affairs. The effect of Schedule 3, Part 2 is felt most sharply with respect to Article 15 and the right of access, we consider its effect at Section 4(b)(iii)(2) below.

Article 14 exceptions

Where information has *not* been obtained from a data subject, three further exceptions may apply:

I. Where the provision of such information proves impossible or would require disproportionate effort, in particular for research purposes (must also safeguard data subject rights, including making the information publicly available);[115] and
II. Obtaining or disclosure is expressly laid down by Union or Member State law (must also safeguard data subject rights);[116] and

III. Where personal data must remain confidential subject to an obligation of professional secrecy regulated by Union or Member State law.[117]

All three of these exceptions are potentially relevant to the healthcare and research sector. For example, in regards to I. many research uses of machine learning will not collect data from subjects directly and will use the Article 6(1)(e) public interest legal basis and research purposes Article 9(2)(j) derogation.[118] That is, many research uses of machine learning may fall under this exception. This exception likely requires exceptional circumstances to trigger. It will be incumbent upon the controller to produce convincing reasons as to why the provision of information generally available to data subjects by right should be so destructive to be disapplied in their situation - the 'disproportionate effect' element. Further, even if the exception applies, some disclosure of information to the data subject or public may be required. In regards to II. disclosure expressly laid down by Union or Member State law, much of the data collected by the NHS and funnelled to bodies like NHS Digital is expressly laid down according to UK law.[119] While much of this data will be anonymised by the time it reaches secondary users including commercial partners, any machine learning conducted using routinely collected identifiable clinical data likely falls within this exception.[120] Finally, in regards to III. information disclosed within the bounds of a clinician-patient relationship is generally subject to duties of confidentiality.[121] In this regard, those using machine learning as part of healthcare caught by this exception may refuse to disclose information about processing that may reveal sensitive data regarding other patients. The combined effect of the three exceptions is that, where data is not collected from data subjects, the right to information may not straightforwardly apply in the healthcare and research context. We examine the effect of this in the next section.

3. What the rights to information require

Both rights to information detail categories of information that must be communicated to data subjects.[122] In terms of content, the rights to information appear to be identical - the status and importance of the information required being the same.[123] However, as noted above, there are more exceptions to the right to information when the personal data is *not* collected from the data subject. Given this, we separate out the analysis, considering what the rights to information require once triggered and what might be required if an exception applies.

What the rights to information require generally

What information must be communicated according to the dual rights where no exception applies? Articles 13(1)-(2) and Articles 14(1)-(2) all note specific categories of information that must be communicated to the data subject. For our purposes, we can distinguish between two possible kinds of information:

I. Information that is administrative in nature, information that constitutes a 'notification duty' only. For example, the identity and contact details of the controller,[124] the fact that the data subject has a right to lodge a complaint with a supervisory authority,[125] and so on
II. Information that might be used to construct or assist with a right to interpretability or explainability

The latter kind of information is of most interest to us. While the information that could be used to construct such rights is a matter of judgment, the following two provisions seem especially relevant:

> Article 13(2)(f)/14(2)(g) 'the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject'; and
>
> Article 14(2)(f) 'from which source the personal data originate, and if applicable, whether it came from publicly accessible sources.'

Articles 13(2)(f) and 14(2)(g) have a special relationship with the automated individual decision-making requirements found in Article 22. Consequently, we examine what these provisions require and whether these requirements constitute a right to interpretability or explanation in Section 5(d).

Article 14(2)(f) is potentially interesting as it relates to data not collected from the data subject and requires the disclosure of the source of data and if this source was public. Along with the general provision of Article 13, this means that data subjects should typically be in a position to understand where their data came from. That is, in most cases, data subjects by right should be told their personal data is processed and the source of this data. In the context of machine learning this might include notification that personal data is processed in training the model and disclosure of the source of this dataset. This by no means constitutes a right to interpretability or explainability. However, disclosure of the source of data might be an important foundational building block to combine with other rights to constitute such a right. For instance, this right may be particularly powerful if the dataset is publicly available. Further, disclosure of the source of data may also allow the intrepid data subject to lodge a request for data portability or access from the source to understand the range of possible inputs used to construct the model. We consider the right of access in Section 4(b)(iii) and data portability in Section 4(b)(iv)

While we consider Articles 13(2)(f) and 14(2)(g) in Section 5(d) as they reference the Article 22 requirements regarding automated processing, there is little else in these rights to information that might constitute a duty of interpretability or explainability. The remaining tools of the right to information mostly amount to straightforward notification duties and requirements to communicate basic, administrative information.

Where an Article 14(5) exception applies

It is less clear what information must be provided where an exception to the rights to information applies. Despite this lack of clarity, we can be reasonably confident about two aspects where an exception applies.

First, save where the data subject already has the information, an exception applying does not typically exempt controllers from providing any information whatsoever.[126] For example, the

Article 14(5)(b) research purposes exception is restrictively worded to begin with only applying where 'the provision of such information proves impossible or would involve a disproportionate effort.' Further, Article 14(5)(b) also stipulates that controllers must take 'appropriate measures to protect the data subject's rights and freedoms and legitimate interests, including making the information publicly available.' Similarly, the Article 14(5)(c) exemption on Union and Member State law also requires 'appropriate measures to protect the data subject's legitimate interests.' In both cases, these provisions are likely not licence to dispense with disclosure completely or not install some alternative to disclosure.

Second, the effect of an exception applying will likely depend upon the exception relied upon and the context in which the exception triggers. It is especially important to consider who is restricted from receiving such information. For example, suppose a controller relies upon the Article 14(5)(d) exception on confidentiality and professional obligations of secrecy. Suppose further that the obligation of secrecy referenced here is the duty of confidentiality that exists between patient and clinician.[127] In this case, the duty of confidentiality typically prevents the communication of confidentiality information from the clinician to a third party, not from the clinician to the patient. In this way, disclosure of information that reveals other data subjects might be restricted on this basis but not information that solely relates to the patient disclosed between clinician and patient. This kind of situation is very different to where a controller relies on the Article 14(5)(b) research exemption to protect the integrity of a blinded clinical trial or where a controller argues that contacting data subjects would be too onerous. In this case, the exemption restricts information communication between the controller and data subject in question.

It is difficult to indicate what information must be provided if one of the exceptions listed in Article 14(5) apply. However, it is likely that these exceptions will not typically operate to block all information required under the rights to information. Moreover, the way the exception will apply will be contingent upon the exception relied upon and the context in which the exception has been triggered. Even given this contextualisation, it seems clear that the rights to information are not strengthened by the application of an exemption. Accordingly, there is little apart from the Article 13(2)(f) and 14(2)(g) (provisions that relate to automated individual decision-making) that might constitute a duty of interpretability or explainability.

Finally, returning to what the rights to information require generally, the rights to information require specific information to be communicated to the data subject, much of this information falls under the umbrella of 'notification duties' and is administrative in nature. For instance, identity and contact details of the controller. Some information like Article 13(2)(f) and 14(2)(g) (covered at Section 5(d) below) may be more substantive. However, it is worth repeating: this list ought not be read as exhaustive - express mention of certain information does not exclude further information from being required by the general principle of transparent processing or other data subject rights. Moreover, much of the guidance released by WP29 and the ICO place heavy emphasis on the proper communication of information rather than just its content - it often addresses the 'how' you communicate not just 'what' you communicate.

4. Application to machine learning for healthcare and research

To summarise, some machine learning for healthcare and research will either use personal data to train the model or, as a part of the function of the model itself, process personal data. At various stages the purposes for processing might change. For instance, a university

researcher might 'process' personal data to develop a model for research purposes. Subsequently, when the trained model is deployed, it might use the personal data of patients to provide predictions for healthcare purposes.

The rights to information may apply to processing for both of these purposes, requiring the disclosure of information to data subjects. It is important to note not just that the rights to information apply but *how* the rights apply. There are a number of scenarios where the rights to information might apply:

I. The right to information under Article 13(1) straightforwardly applies at the point of collection by the controller; or
II. The right to information under Article 13(1) is restricted by Article 23(1) and Schedule 3 of the DPA 2018; or
III. The right to information under Article 14(1) straightforwardly applies where the controller obtains the data from a third party and within a reasonable period of time/when the personal data is used to communicate with the data subject/where further disclosure is envisaged; or
IV. The right to information under Article 14(1) has an exception applied under Article 14(3); or
V. The right to information under Article 14(1) is restricted by Article 23(1) and Schedule 3 of the DPA 2018.

The rights to information will be softened if a restriction or exception applies. Nevertheless, some information will likely still have to be disclosed to data subjects. Article 13(2)(f), 14(2)(g), and 15(1)(h) aside, it is unlikely that the requirements of the rights to information, even where the rights apply in their fullest form, require interpretability or explainability. Nevertheless, the rights to information, even if they mostly constitute 'notification duties', are still important as they facilitate the function of other data subject rights.

iii. Right of access

In concert with other data subject rights, the right of access supports the principle of transparency and is another candidate with which to formulate a duty of interpretability or explainability. The right of access gets relatively little attention in WP29's *Guidelines on Transparency under the GDPR*, the Guidelines mostly focusing on the general principle of transparency, Article 12's general application to data subject rights, and specific application to the rights to information.[128] Nevertheless, the right of access is potentially the most pivotal right that might establish a duty of interpretability or explainability. Indeed, following Ausloos et al, the right of access is 'a *sine qua non* for meaningfully exercising other data subject rights in Chapter III of the GDPR.'[129] The following provides a description of the right, when in time it triggers, what processing triggers the right, what the right requires, and how these requirements might apply to machine learning for healthcare and research.

1. Timing of the right of access

Article 15(1) notes:

> 'The data subject shall have the right to obtain from the controller confirmation as to whether or not personal data concerning him or her *are* [my emphasis] being processed…'

The usage of 'are being processed' along with the clarifications present in Recital 63 mean that the right of access triggers after the collection of personal data and remains available, allowing the data subject to request information at reasonable intervals. It is likely that the meaning of 'reasonable intervals' references the ability of the controller to reject vexatious uses of the right meant to bombard the controller. It is unlikely that 'reasonable intervals' allow a controller to structure rounds of disclosure, as the right of access is always triggered on request from the data subject. In this regard, the right of access is a right that remains available to the data subject throughout the lifecycle of processing.

In terms of timing, the right of access differs from the rights to information - the latter triggering at the outset of collection, the former being available at intervals throughout processing. This difference in timing is potentially significant in the argument for whether a right to explanation exists or not. We address this timing difference and its importance to the right to explanation later at Section 5(g)(2).

Another notable difference is that the right of access - and many of the following rights - typically requires a data subject to request action to be taken.[130] In this way, information provided under the right of access requires a request to trigger disclosure. The distinction between the rights to information that require communication without any request by the data subject and other rights that require action on the part of data subjects is important in practice. Indeed, only a small minority of data subjects will likely take any positive action in relation to their personal data. Consequently, even if the right of access and other rights requiring positive action from data subjects contain rich, illuminating information, this information will likely only reach a minority of data subjects.

2. What processing triggers the right of access?

The right of access is restricted or unavailable to some data subjects. In the context of healthcare and research, there are three elements of interest that restrict its use for data subjects. As we have already seen, this right is restricted where Articles 11 and 12(2) on identifiability apply. In regards to healthcare, additional restrictions are based on Article 23(1). In regards to research, the restrictions are based on Article 89(1).

Articles 11 and 12(2)

The right of access is often restricted or entirely set aside where the purpose for processing no longer requires identification of data subjects. As noted earlier in Section 4(i)(1), where the controller demonstrates that they are 'no longer in a position to identify that data subject' and

the data subject does not provide additional information enabling identification, then the right of access no longer applies.[131] Notably, these provisions did not apply to the rights to information. How will Article 11 and 12(2) restrict the right of access as applied to machine learning for healthcare and research?

In the healthcare context, it is important to distinguish between data processed as a part of training data and data processed input into the model as a variable.

With regards to training data, developers are often not in a position to identify data subjects in the prepared training dataset or the produced model, even if data subjects were to provide further data in an effort to allow identification. Still, it is important to note that well-meaning efforts to facilitate data subject rights, for instance, the right of access, have led to data remaining identifiable. In this way, there is a tension between the data protection by design/default requirement and vindicating the rights that require identification.[132] Conservative approaches to GDPR compliance favour treating data as identifiable (even if it may not be) to ensure key data subject rights are given effect.[133]

With respect to personal data as an input for a machine learning model, identification of a data subject is often necessary and in fact the entire point of the processing activity. For instance, if values based on attributes of a patient are input into a model in order to produce a diagnosis, the very purpose of the processing is to return a diagnosis to that patient. One potential complication is that healthcare institutions often purchase technology from vendors. In this way, the vendor alone may not be able to identify data subjects but processes data through their model, through their server, or simply licence the software for use in that institution. In these situations, it is important to note that the vendor may still be subject to the GDPR in two main ways. First, as a processor processing personal data on behalf of the controller healthcare institution.[134] Second, as a joint controller if they jointly with the healthcare institution 'determine the purposes and means of the processing of personal data.'[135] The possibility of being a joint controller has been made probable with recent CJEU case law applying an expansive view of the concept.[136] In short, even if the data is not identifiable in the hands of the vendor, the vendor may still be subject to data subject rights if the healthcare institution does identify the data subject - Articles 11 and 12(2) may be of little assistance.

In the research context, often research uses of machine learning blend into healthcare uses or necessitate the need to retain the capacity to recontact data subjects. In this way, when investigating diagnostics and therapeutics, it is often necessary as a part of the research or as an ethical imperative to keep the data subject identifiable. Pseudonymization of datasets, while an important method to preserve privacy and maintain security, do not necessarily put the data beyond the scope of the GDPR.[137] Consequently, many important research uses require identifiability as a part of their research purposes - Articles 11 and 12(2) are unlikely to apply.

Article 23 restrictions

As noted, Article 23(1) GDPR provides scope for Union and Member State law to restrict by way of legislation data subject rights and processing principles. Schedule 3, Part 2 provides for such restrictions in the context of health data. The right of access, along with the other rights discussed at Section 4(b)(i)(2), is subject to a) exemption in regards to data processed by courts and b) special rules relating to a data subject's expectations and wishes where a person manages a data subject's affairs.[138] However, the right of access is subject to a special

restriction and exemption over and above other rights in Sections 5 to 6 of Schedule 3, Part 2. This restriction and exemption:

I. Restricts who can disclose data concerning health under the right of access to those who are either a:
    A. Health professional; or
    B. Controller who has sought an opinion from 'the appropriate health professional' noting that the serious harm test has not been met; and
II. Exempts information under the right of access from disclosure where the 'serious harm test' is met.[xiii]

The 'appropriate health professional' generally means the 'health professional who is currently or was mostly recently responsible for the diagnosis, care or treatment of the data subject in connection with the matters to which the data relates.'[139] However, rules can differ, depending on the circumstance.[140] The upshot of this being that typically a health professional must provide an opinion noting that the serious harm test is not met when a request under the right of access is requested.[141] The serious harm test has the following definition:

> 'The "serious harm test" is met with respect to data concerning health if the application of Article 15 of the GDPR to the data would be likely to cause serious harm to physical or mental health of the data subject or another individual.'[142]

The general effect of both sections is that where the serious harm test with respect to health data is met, the right of access is blocked. We will explore the effect of this in the following sections.

Article 89(1) restrictions

In regards to research purposes, Article 89(1) GDPR provides scope for Union and Member State law to restrict by way of legislation certain data subject rights. Article 89(2) includes the right of access in this list. Derogations may be provided for via legislation:

I. So long as appropriate safeguards are in place, in particular, procedures to ensure the principle of data minimisation is respected;[143] and
II. Insofar as the data subject right would 'render it impossible or seriously impair the achievement of the specific purposes' of processing.[144]

The DPA 2018 in Schedule 2, Part 6 provides for exceptions along the lines of Article 89(2), adding in Section 27(3) that the exemptions to the right of access only apply so long as the results of research do not identify data subjects. Given this, the converse is true: if research

[xiii] N.B. See The Data Protection (Subject Access Modification) (Health) Order 2000, Section 2 and Schedule 3, Part 2, Section 2(1) DPA 2018 for a definition of 'appropriate health professional.'

purposes are relied upon and the research identifies data subjects, the right of access applies under UK law.

The effect of the above is that the right of access may well be blocked in circumstances where the controller can make a convincing case that access would seriously impair the research purposes pursued and where this research does not identify data subjects. Notably, this case may be more difficult to make where special category data (for our purposes, data concerning health, genetic data, or biometric data) are processed.

Recital 63 trade secrets and intellectual property

As described in Section 4(b)(i)(4), Recital 63 notes that the right of access 'should not adversely affect the rights or freedoms of others, including trade secrets or intellectual property and in particular the copyright protecting the software.' Notably, this restriction does not operate to entirely block the right but tempers disclosure. Recital 63 also clarifies:

> 'However, the result of those considerations should not be a refusal to provide all information to the data subject. Where the controller processes a large quantity of information concerning the data subject, the controller should be able to request that, before the information is delivered, the data subject specify the information or processing activities to which the request relates.'

Controllers should be wary of disclosing their or third parties' trade secrets or intellectual property but the response should be clarification and specification of what the data subject seeks access to, not outright refusal. The hope being that this specification allows both trade secrets/intellectual property to be preserved and the right of access to be accommodated.

3. What the right of access requires

The right of access requires disclosure of some of the same information as under the rights to information. However, there are at least five notable differences for the right of access.

First, while some of the information required may be the same, the difference in timing may produce radically different consequences. For instance, consider the Article 15(1)(h) requirement to provide 'meaningful information about the logic involved as well as the significance and the envisaged consequences of such processing' in relation to automated individual decision-making. There may be relatively little information a controller could provide prior to processing to satisfy such disclosure. Disclosure will be limited to generalities. However, after processing, the controller will generate further data, data also potentially subject to the right of access and thereby disclosure. Given this, the same requirement triggered at different times can require different disclosures. This difference should not be underestimated, especially considering the possible breadth of scope of the right of access. We consider this difference later at Section 5(g)(ii).

Second, Article 15(3) requires controllers 'shall provide a copy of the personal data undergoing processing.' The provision further states that where the request is made by electronic means

(and the data subject does not request otherwise), the information should be provided in a commonly used electronic form. In this way, Article 15(3) acts as adjunct right similar to the right to data portability we examine in Section 4(b)(iv) below. The strength of the right to data portability over and above the right found in Article 15(3) being that the data subject can receive back *structured machine-readable format* data.[xiv] However, the strength of the right of access is in its possible breadth. The right to data portability is limited to data provided by the data subject. The right of access has no such restriction. The scope of 'personal data undergoing processing' is potentially large. For instance, Ausloos et al (2019) note that with judgments like *Nowak* (C-434/16) in mind, the right of access might extend to opinion and inferences about the data subject, so long as data is sufficiently linked to a person 'by reason of its content, purpose or effect.'[145]

Third, some authors argue that the right of access under Article 15(3) requires more than just a mass disclosure of data concerning the data subject. For example, Ausloos et al argue that 'accommodating the right of access should – where needed – include the tools rendering the entire data-set understandable.'[146] Basing their argument on the Article 12(2) transparency requirements to facilitate data subject rights, Ausloos et al argue that the layered approach to privacy notices advocated for in WP29 Guidelines, equally applies to facilitation of data subject rights.[147] As the argument goes, disclosure should include simple, understandable summaries but also comprehensive, technical information. Arguably then, the Article 15(3) right of access at its barest interpretation might encourage the disclosure of explanation of an otherwise uninterpretable machine learning model.

Fourth, the right of access is restricted and excluded in different ways from the rights to information. Specifically, under Article 11 and 12(2) in regards to the inability to identify the data subject, Article 23 and Schedule 3, Part 2 of the DPA 2018 in regards to the serious harm test, and Article 89(2) in regards to research purposes that do not identify data subjects.

Fifth, the right of access includes some flexibility around disclosure that does not necessarily apply to the rights to information. For instance, Recital 63 clarifies:

> 'Where the controller processes a large quantity of information concerning the data subject, the controller should be able to request that, before the information is delivered, the data subject specify the information or processing activities to which the request relates.'

Accordingly, because the right of access is in response to a request, there is some scope to ask for specification of the request, especially if the request would reveal personal data of others or infringe trade secrets/intellectual property (as noted in Section 4(b)(i)(4) above).

[xiv] N.B. Some jurisdictions, for instance, the Supervisory Authority of Hesse (a German state) interpret the requirements of Article 15(3) as requiring only a summary of the data subject's data. See: Elteste U, Van Quathem K. *German court decides on the scope of GDPR right of access*. Available from: https://www.insideprivacy.com/international/european-union/german-court-decides-on-the-scope-of-gdpr-right-of-access/ [Accessed 9th February 2020].

Following the above five differences, we note that the right of access is a distinct right from the rights to information in its timing and context.

4. Application to machine learning for healthcare and research

What does the right of access and its relation to the principle of transparency require of machine learning for healthcare and research in practice? Notably, the right of access may be restricted or wholly put aside as it applies to health data following Schedule 3, Part 2 of the DPA 2018. Moreover, the right may also be restricted in the context of research according to Schedule 2, Part 6 of the DPA 2018. Where available, the right of access requires the provision of information, much of this information being similar to that required under the rights to information. However, because the right of access remains available post-processing and applies to data (so long as it is 'personal') generated as a part of processing, the disclosure required is likely more extensive and might constitute a possible foothold for a duty to render interpretable or explain. Moreover, the right of access requires under Article 15(3) that a copy be provided of personal data undergoing processing. This Article 15(3) right likely extends to opinions and inferences made about the data subject. Further, this right set in the context of the general principle of transparency and WP28 and EDPB Guidelines arguably may require more than just mass disclosure. Following Ausloos, this disclosure should be contextualised, possibly with some kind of explanation. It is unclear exactly what form this explanation might take. However, it is important to note that the right of access might serve as a foothold to require explanation. To clarify, this is apart from any provision related to automated processing like Article 13(2)(f), Article 14(2)(g), and Article 15(1)(h). We consider what these provisions require later at Section 5(d).

iv. Data portability

Unlike the DPD, the GDPR contains a right to data portability. This section provides a brief analysis of the right, the upshot of this analysis being that the right to data portability is not a good candidate on which to build a right to interpretability or explainability.

The nub of the right to data portability is found in Article 20(1):

> 'The data subject shall have the right to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format and have the right to transmit those data to another controller without hindrance from the controller to which the personal data have been provided, where
>
> a) the processing is based on consent pursuant to point (a) of Article 6(1) or point (a) of Article 9(2) or on a contract pursuant to point (b) of Article 6(1); and
> b) the processing is carried out by automated means.'

The nature of the right to data portability becomes clearer when we consider the numerous restrictions to which it is subject.

1. The restrictive nature of right to data portability

There are five key restrictions that shape the right to data portability and limit its usefulness for generating a duty of interpretability or explainability.

First, the right to data portability is restricted to the 'personal data concerning him or her, *which he or she has provided to a controller* [our emphasis].'[148] When processing data, especially in a machine learning context, new data will be generated. This data, so long as it is personal data, is also within the material scope of the GDPR. However, the right to data portability does not require this data to be disclosed. Accordingly, we should distinguish between:

I. The personal data the data subject has provided and the structuring of this data - the data which might be subject to the right to data portability; and
II. The personal data generated by the controller, data that is not conceivably part of the structure of the provided data - data not subject to the right to data portability.

Following the above, the personal data to which the right attaches is limited.

Second and relatedly, the exercise of the right to data portability may conflict with the rights of other data subjects to not have their data disclosed.[149] Article 20(1) makes clear that the right to data portability is restricted to data that 'concerns him or her [the data subject].' Further, Article 20(4) clarifies that the right to data portability 'shall not adversely affect the rights and freedoms of others.' Accordingly, the data within the scope of this right is also restricted to data that does not concern other data subjects.

Third, processing is limited to 'processing carried out by automated means.'[150] It is unclear how 'automated' the processing must be to restrict the application of the right. Indeed, the usage of 'automated' remains unqualified unlike Article 22(1) that refers to 'a decision based solely on automated processing.' Perhaps the only meaning we can read into this difference is that the bar for 'automated' will be lower - perhaps significantly lower - than 'based solely on automated processing.' Nevertheless, processing that is not automated is beyond the right to portability.

Fourth, the right only triggers when certain legal bases or derogations are relied upon. Namely, if consent as a legal basis (Article 6(1)(a)) or derogation (Article 9(2)(a)) is relied upon or if contract as a legal basis (Article 6(1)(b)) is relied upon, the right to data portability might arise. In the context of UK healthcare and research, neither legal basis or derogation is an attractive option. In regards to consent, consent has a prominent place in healthcare and research but is often an onerous legal basis or derogation to rely upon for data processing. Accordingly, the Health Research Authority and the Information Governance Alliance discourage public bodies conducting health research to rely on consent as their legal basis for processing or Article 9 derogation.[151] In regards to contract, recent EDPB guidelines on the interpretation of contract as a legal basis also mean that contract is a potentially impractical legal basis with which to conduct healthcare or research.[152] The effect of this is that, in healthcare and research, especially if this research is undertaken by the public body, the right to data portability will not arise.

Fifth, the right to data portability is specifically blocked or its operation restricted in a number of circumstances, namely:

I. If the processing is necessary for a task in the public interest, following Article 20(3); or[xv]
II. Where an Article 23 restriction according to Member State law applies, in the case of the UK Schedule 3, Part 2 DPA 2018; or
III. Where the Article 89(3) derogation on archiving according to Member State law applies, in the case of the UK, Schedule 2, Part 6, Section 28 DPA 2018.

In regards to Article 23, Schedule 3, Part 2 of DPA 2018 mentions the right to data portability as being among the rights that Part 2 derogates from.[153] Accordingly, data concerning health that is processed by the courts is exempt from the right to data portability.[154] Further, special provisions are made for data concerning health and data subject expectations and wishes.[155] Of note is that data portability is *not* listed alongside the right of access in relation to the serious harm test.[156] That is, contrary to the right of access, the serious harm test does *not* exempt information from being returned under the right to data portability. This makes sense as, unlike the right of access, the right to data portability only requires the returning of data already provided by the data subject. As an example, suppose a healthcare professional notes that the sharing of a psychiatric diagnosis would meet the serious harm test. If a data subject access request were made, there are two possible claims at play here: the claim relating to data access and the claim relating to data portability. Under the right of access claim, the access claim may include the opinions of healthcare professionals but can also be blocked if the serious harm test is met. Under the right to data portability, this right only returns data provided already by the data subject just in a machine readable format.

In regards to Article 89, only Article 89(3) - archiving in the public interest - allows for derogation from the right to data portability.[157] As a consequence, those relying on the research interest derogation found in Article 89(2) may have to meet the right to data portability. That is, controllers relying upon research interests may well trigger the right to data portability if they rely upon consent or contract as a legal basis (and in the case of consent as an Article 9 derogation).[158]

To summarise, this section establishes that the right to data portability is a right to a subset of data processed by automated means, is a right constrained by the rights of others, that only triggers when certain legal bases or derogations are relied upon, and is subject to myriad restrictions.

[xv] N.B. Article 20(3) GDPR includes rather strange drafting. Our interpretation is that 20(3) is belt and braces, noting that the right does not apply where public interest is relied upon. In most cases, this should be obvious as the public interest legal basis is not included in the list of legal bases that trigger the right to data portability. However, it is unclear what happens if a controller relies upon legitimate interests as their legal basis and then public interest as their Article 9 derogation. In this case, public interest is relied upon, just not as the controller's legal basis. While it is unclear what happens here, it is equally unclear why a controller would rely on public interest as their Article 9 derogation but not for their legal basis.

2. What the right to data portability requires

If the right to data portability does apply, what does the right require and might this right support a duty of interpretability or explanation? As we have seen, this right only arises in narrow circumstances and attaches to a subset of information. Even if it does apply - the right cannot conceivably be stretched to formulate a duty of interpretability or explainability. The main reason for this is that the right requires little disclosure of information beyond what the data subject has already provided.

In terms of content, the right to data portability requires that the personal data, which the data subject has provided, be returned in a structured, commonly used and machine-readable format.[159] Exactly what the three terms 'structured', 'commonly used' and 'machine-readable' means is subject to debate.[160] However, what does seem clear is that no reasonable interpretation of any of these adds much in terms of interpretability or explanation. Indeed, much of the debate surrounds what file format is most appropriate in what circumstances.[161] While the right to portability certainly represents a power shift in favour of data subjects in the contexts of social media and insurance, the right provides slim foundation for a duty of interpretability or explainability. Perhaps the only contribution to interpretability of explainability might be information gleaned from the structure in which the data is returned and the ability to reinterpret the data more easily.

v. Other data subject rights

We have considered so far the rights to information and the right of access. Generally, these are the data subject rights used to argue that a right to interpretability or explainability exists within the GDPR. However, a number of additional data subject rights, rather than themselves constituting a right to interpretability or explanation, may require some form of interpretability or explanation if they are to be vindicated. In this regard, while some rights may not be obviously geared toward interpretability or explainability, they may, as a consequence of their implementation, require some interpretability or explainability. The following analysis considers whether the rights to rectification and the right to object might require some kind of interpretability or explanation to be vindicated.

1. Right to rectification

Like the DPD, the GDPR contains a principle of accuracy, Article 5(1)(c) notes that personal data shall be:[162]

> 'accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay.'

Relatedly, Article 16 GDPR contains a specific right to rectification:[xvi]

> 'The data subject shall have the right to obtain from the controller without undue delay the rectification of inaccurate personal data concerning him or her. Taking into account the purposes of the processing, the data subject shall have the right to have incomplete personal data completed, including by means of providing a supplementary statement.'

On the face of it, the right to rectification has little to do with interpretability or explanation. However, the argument for the right's relevance might go as follows: how can I as a data subject verify the accuracy of data if the system that processes the data is a black box? First we examine when the right to rectification is available to data subjects.

When is the right to rectification available?

The right to rectification is not available to all data subjects at all points. Similar to previous data subject rights discussed, there are restrictions and derogations from the right found in Article 23 and Article 89. We take each in turn.

In regards to Article 23, there is scope for Member State or Union law to restrict the right to rectification with the caveat that these restrictions respect the essence of fundamental rights and freedom and constitute a necessary and proportionate measure in a democratic society.[163] Schedule 3, Part 2 of the DPA 2018 contains few restrictions or derogations in relation to data concerning health - the restrictions only relating to data processed by courts[164] and requests of data contrary to data subject expectations and wishes.[165] Consequently, there are few restrictions to the right to rectification for data concerning health in particular.

In regards to Article 89, research purposes allow derogation from the right to rectification subject to the caveat that the right would otherwise likely render impossible/severely impair the purposes of processing and that the derogation is accompanied by appropriate safeguards.[166] As outlined in Section 4(b)(i)(2), under Section 19 DPA 2018, Article 89 restrictions will not apply where the processing is likely to cause substantial damage or distress to the data subject.[167] Moreover, where processing is carried out for the 'purposes of measures or decisions with respect to particular data subject,' the processing will have to count as 'approved medical research' for any Article 89 restriction to apply.[168] Consequently, the right to rectification may more often be restricted in the context of approved medical research. For instance, where research is carried out under a recognised research ethics committee or within a relevant NHS body.[169]

As outlined in Section 4(b)(i)(3), under Section 19 DPA 2018, Article 89 restrictions will apply so long as the processing is not likely to cause substantial damage or distress to the data subject and, where the processing relates to 'measures or decision with respect to a particular data subject', the processing is carried out as 'approved medical research.'

[xvi] N.B. the DPD also contained a provision on rectification but this provision was found under the right of access and rolled together with other provisions that might belong to the group of provisions related to 'data quality.' See: DPD, art 12(b).

What the right to rectification requires

When considering this right, we distinguish between two kinds of data that might be inaccurate and require rectification:

I. Personal data provided by the data subject to the controller that is inaccurate. That is, the processing containing the error involves the collection or storage of the data - perhaps the data was improperly recorded, contains errors, was attributed to the wrong data subject and so on; and
II. Personal data that was generated about the data subject by the controller. That is, the processing containing the error involves the algorithm that processes personal data- perhaps the algorithm relied upon the wrong data, relied on improper features to reach a conclusion, or simply 'misclassifies' the data subject.

Inaccurate data that fits within I. is data that may be rectified under the right but not conceivably used to leverage interpretability or explainability. In this situation, the data subject does not need to know anything about the algorithm that processes their data to know whether their data requires rectification or not. However, some inaccurate data falling within 2. may be conceivably used to leverage some kind of interpretability or explainability. For example, diagnostic machine learning models often process data concerning health to produce some kind of output to aid or provide a diagnosis. If the data subject only has the data they provided and the output from the model, but the processing itself is human uninterpretable and goes unexplained, arguably the data subject is not a position to know whether the output is accurate or inaccurate. In such a case, it is arguable that the data subject might be entitled to some type of interpretation or explanation in order to identify potential errors.

2. Right to object

Another potential argument in support of a duty of interpretability or explainability is that it is required in order for data subjects to properly exercise their right to object if the processing involved uses a black box model. The following outlines the right to object and why it does not generate a duty of interpretability or explainability.

When and what triggers the right to object

The right to object is a specialised tool, not a general right to object to processing of one's personal data.[170] Accordingly, the right only triggers under narrow circumstances and only requires disclosure of information in line with its limited purpose.[171] The right to object triggers in time to challenge the weighing that must take place when relying on certain legal bases. Specifically, the right to object triggers:[172]

I. When the legal basis relied upon is public interest (Article 6(1)(e); or
II. When the legal basis relied upon is legitimate interests (Article 6(1)(f).

In the context of healthcare and research, both of these legal bases will frequently be relied upon. Where public bodies conduct research, the public interest legal basis is a common choice.[173] Where commercial bodies are concerned, legitimate interests is one of the few

feasible legal bases that allow commercial health-related research under the GDPR. The right to object is therefore a live issue in regards to healthcare and research processing.

If successfully invoked, the right to object results in the controller no longer processing the personal data of the data subject.[174] This result is achieved by challenging the balancing exercise that must be undertaken if a controller relies on either of these legal bases. For instance, in regards to legitimate interests, controllers may only rely on this legal basis if these legitimate interests are not overridden by the interests or fundamental rights and freedoms of the data subject.[175] This balancing act is a vexed exercise and reliance on this legal basis in effect means that a controller is saying: 'our legitimate interests override the interests and fundamental rights at play.' A number of rules dictate when such challenges will be successful. The general rule is that if the controller relies upon public interests or legitimate interests and the data subject objects, the personal data should no longer be processed *unless* the controller demonstrates 'compelling legitimate grounds for the processing which override the interests, rights and freedoms of the data subject…'[176] There are also subordinate rules. Relevant to us is the subordinate rule that relates to Article 89(1) research purposes. This states that where research purposes are invoked, the right to object should operate *unless* the processing is necessary for the performance of a task carried out for reasons of public interest, that is, the public interest legal basis is relied upon.[177] If the public interest legal basis is relied upon, then the general rule, the balancing act applies. The consequence for the right to object is that two groups will emerge. First, the right to object will not necessarily trigger where public interest is claimed and relied upon. Second, the right to object will apply where research purposes are relied upon but no public interest is claimed or relied upon. In this way, the right to object will be more likely to apply where the legitimate interest legal basis is relied upon as opposed to public interest legal basis.

What the right to object requires

Might the right to object require some kind of interpretability or explainability to be vindicated? The purpose of the right to object is to challenge the balancing exercises found in the legitimate interests and public interest legal bases. The question is what information would a challenge to legitimate interest or public interest balancing require, and, would this information require some kind of interpretability or explainability? Let us take the balancing test for legitimate interests as our litmus test to answer these two related questions.

Lawful reliance on the Article 6(1)(f) legitimate interest legal basis likely requires the controller to apply a tripartite test.[178] While this tripartite test was developed under the DPD, GDPR guidance from national authorities still references the test as persuasive law, the test requiring:[179]

I. The identification of a *legitimate* interest; and
II. The personal data processed must be necessary to fulfil that purpose; and
III. The fundamental rights and freedoms must not override these legitimate interests.

All three limbs of this test are relevant when thinking of a right to object. That is, a data subject invoking their right to object may potentially challenge the legitimacy of the interest pursued, the necessity of the processing to achieve that purpose, and the balance between these interests and their rights and freedoms.

There is little in the three parts of this test that might require interpretability or explainability. Of course, any information provided to assist the data subject in vindicating their right to object needs to be interpretable. However, the test as noted under the Directive by WP29 and the CJEU mainly involves a broad-based assessment of the interests and rights at play.[180] It seems likely then that the right to object does not require much information beyond the standard notification duties found under the rights to information (see Section 4(b)(ii) for notification duties). Perhaps one area that might provide for more information than the standard notification duties would be information relating to the envisaged consequences of the processing - information akin to that mentioned through the provisions that relate to automated processing found in Article 13(2)(f), 14(2)(g), and 15(1)(h). That is, envisaged consequences of processing seem highly relevant when balancing legitimate interest with the right and freedoms of the data subject. Regardless, it still seems unlikely that the right to object could be wielded to require disclosure of 'logic' or the kind of interpretability/explainability discussed in previous reports.

We now consider the totality of what the principle of transparency and data subject rights (excluding Article 22 related provisions) require with respect to interpretability and explainability.

c. What the principle of transparency and data subject rights require

We have considered what the general principle of transparent processing requires and how this principle interacts with other data subject rights. We now consider the cumulative effect of both together: do these amount to a duty of interpretability or explainability?

The principle of transparency, as interpreted by Recital 60 requires controllers to

> 'provide the data subject with any further information necessary to ensure fair and transparent processing taking into account the specific circumstances and context in which the personal data are processed.'

Accordingly, the guiding element for assessing whether interpretability or explainability is required is to question whether they are, in that context, needed to facilitate fair and transparent processing.

Data subject rights instantiate and flesh out the principle of transparency and its requirements. In healthcare and research, many of these rights are restricted or unavailable to data subjects. Regardless, even where these rights apply in their fullest extent (provisions relating to automated individual decision-making aside), there are no direct requirements to provide interpretability or explainability. There are provisions that may require some kind of interpretability or explainability to fully vindicate but nothing concrete. This is not to say that data subject rights (excluding Articles 13(2)(f), 14(2)(g), and 15(1)(h)) provide no tools relevant to interpretability. On the contrary, they often supply critical information or access to data which may be useful to construct interpretability or explanation.

The principle of transparency and associated data subject rights are more than the sum of their parts with respect to interpretability and explanation. Namely, the wise data subject will use the totality of rights and principles available to them to construct a right to interpretability or explanation. One of the potentially most powerful tools at the data subject's disposal are the automated individual decision-making restrictions and associated provisions. We now turn to consider these restrictions and provisions below.

Section 4 key messages:

- **The general principle of transparent processing is context-specific and user-centric. It requires controllers to consider the form in which they communicate (accessibility, simplicity, and intelligibility) as well as the content. In regards to the content, Recital 60 clarifies that controllers should: 'provide the data subject with any further information necessary to ensure fair and transparent processing taking into account the specific circumstances and context in which the personal data are processed.'**
- **The general principle places a triple obligation on controllers, requiring that they comply with the principle when communicating with data subjects, disclose information required under the rights to information, and facilitate other data subject rights found in Articles 15-22.**
- **Depending on the context, data subject rights may be qualified, restricted, or derogated from.**
- **In the context of healthcare and research, four restrictions to data subject rights and data protection principles are particularly relevant. First, where the controller is no longer in a position to identify the data subject (Articles 11 and 12(2)). Second, the Article 23(1) restrictions that apply to health data according to the DPA 2018's Schedule 3, Part 2. Third, the flexibility for research purposes found in Article 89 and in the DPA 2018's Section 19 and Schedule 3, Part 6. Fourth, the restrictions relating to disclosure of trade secrets and intellectual property in Recital 63 and Article 23(1)(i).**
- **Considering rights that might directly require some kind of interpretability or explainability, the rights to information (Articles 13(2)(f), 14(2)(g) aside), of access (Article 15(1)(h) aside), and portability require little interpretability or explanation. However, these rights may provide useful tools to construct interpretability or explanation.**
- **Considering rights that might indirectly require some kind of interpretability or explainability, the rights to rectification and to object may arguably require some interpretability or explainability to be vindicated. However, such arguments are context-specific, turning on their facts, rather than constituting a tangible right for data subjects to invoke.**
- **The general principle of transparency combined with the data subject rights outlined is more than the sum of its parts. The wise data subject will use the totality of the rights available to them to leverage interpretability or an explanation.**

5. Automated individual decision-making

The GDPR contains specific provisions on automated individual decision-making. It is these provisions and the Articles throughout the GDPR that make reference to them that are the focal point for the debate over whether there is a 'right to explanation.'[181] Much of the ink spilled over this question uses the same materials - the same Recitals and Articles of the GDPR - but comes to radically different conclusions about when and what the GDPR requires. Our method to answering such questions is to consider the automated individual decision-making requirements and how these might be combined with related data subject rights and principles. We then reflect upon whether they are sufficient to constitute a 'right to explanation.'

This section on the automated individual decision-making considers two main questions:

A. When the automated individual decision-making conditions are triggered; and
B. What the automated individual decision-making conditions require once triggered.

This section clarifies what the automated individual decision-making conditions are, when data subjects will be able to avail themselves of any right that emerges, what the conditions require of controllers, and how these conditions might apply to machine learning for health. Finally, we consider to what extent these conditions constitute a 'right to explanation.'

a. The structure of automated individual decision-making conditions

If a 'right to explanation' exists, it is a composite right, a right to be read across multiple Articles of the GDPR. Accordingly, if a right to explanation exists, it is a result of considering the combined effect of different data protection principles, data subject rights, and other requirements. This begs the question, where do we find this 'composite right'?

The most important parts of this composite right are found by reading across:

I. Article 22 on automated individual decision-making; and
II. The rights to information and access, specifically Articles 13(2)(f), 14(2)(g), and 15(1)(h) that reference parts of Article 22 and mention 'meaningful information about the logic involved;' and
III. Recital 71 that provides an interpretative aid to interpret the above provisions.

If the right to explanation exists, it exists by reading these Articles and Recital in conjunction with the general principle of transparent processing. What do each of these Articles and their Recitals say?

Article 22(1) includes a general prohibition or right against decisions based on a certain kind of processing, stating:

> '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.'

The rest of Article 22, does two things:

A. Article 22(2) lays down exceptions to this general prohibition/right
B. Article 22(3) requires, where these exceptions apply, that certain safeguards must be in place

Article 22(1) contains no reference to transparency or any provision that by itself might be construed as a 'right to explanation.' If the right to explanation exists, the 'explanatory' content is found elsewhere, namely in either:

1. Recital 71 that assists in the interpretation of Article 22; and/or
2. The safeguards and reference made to the right to 'contest the decision' mentioned in Article 22(3); and/or
3. Provisions that reference Article 22(1) in the rights to information and access, specifically Articles 13(2)(f), 14(2)(g), and 15(1)(h).

The next sections analyse Article 22, Recital 71, and the relationship between Article 22(1) and Articles 13(2)(f), 14(2)(g), 15(1)(h).

b. The spirit of Article 22

Where does this concern in regards to automated decision-making come from? How should we interpret the Article?

Article 22 was the result of negotiation and bargaining when settling the final text of the GDPR, and the influence of other legislative measures is apparent. There are four points to note when interpreting Article 22.

First, Article 15 DPD (the equivalent of Article 22 GDPR) saw wide variation in Member State implementation. For instance, Italy's implementation only prohibited judicial or administrative decisions 'based solely on the automated processing of personal data,' subjecting private sector equivalents to a qualified right to object.[182] As a consequence, Article 15 was an example of the fragmentation of data protection protections across the EU that the GDPR sought to address.

Second, Article 22 has its roots in the GDPR's predecessor the DPD and France's 1978 Act on data processing files and individual liberties.[183] The preparatory material for the DPD (*travaux préparatoires*) clarifies the purpose of Article 15 DPD:[184]

> 'The danger of the misuse of data processing in decision-making may become a major problem in future: the result produced by the machine, using more and more sophisticated software, and even expert systems, has an apparently objective and incontrovertible character to which a human decision-maker may attach too much weight, thus abdicating his own responsibilities. Article 16(1) therefore lays down the principle that a person is not obliged to accept a decision of a public administration or of a private party which adversely affects him if it is based solely on automatic processing defining a personality profile.'

To summarise, the DPD Article reflects the broad concern that human decision makers will abdicate their responsibilities to automated processing. In this way, Article 15 DPD sought to mitigate against the predicted decline of having a human in the loop.

Third, Article 22 differs from Article 15 under the DPD in a number of ways. One significant change is the inclusion and emphasis on 'profiling.' Explicitly, the emphasis on profiling was inspired by the Council of Europe Recommendation CM/Rec(2010)13 on *The protection of individuals with regard to automatic processing of personal data in the context of profiling*.[185] This Recommendation warned of a number of risks the rise of profiling posed to data subjects and society.[186] Its influence is seen in the development of the GDPR, especially in draft proposals that feature profiling much more prominently in the title of the Article and in the general prohibition/right.[187]

Fourth, the spirit of Article 22 may also be useful when considering how to interpret Article 22(1). Notably, there are two broad interpretations of Article 22(1).[188] First, Article 22(1) as a qualified prohibition. This interpretation reads Article 22(1) as applying by default prohibiting 'decisions based solely on automated processing' that have 'legal effect' or 'similarly significantly affect' the data subject. Second, in the alternative, some commentators read into the clause 'the data subject shall have the right not to be subject to a decision,' noting that the kind of decision-making is not prohibited by default but merely provides another right for the data subject to invoke. Indeed, this latter interpretation seems most consistent with the preparatory materials of Article 15 DPD, these materials emphasising the choice of the data subject: 'a person is not obliged to accept a decision of a public administration or of a private party which adversely affects him if it is based solely on automatic processing.'[189]

To summarise, Article 22 seeks to harmonise approaches to automated processing across the EU. Article 22's spirit is not singular but a mishmash of concerns over the rise of processing without a human in the loop and profiling. We should interpret Article 22 as an example of how the GDPR seeks to give data subjects control over their data and furnish them with 'efficient and operational means to make sure they are fully informed about what happens to their personal data and to enable them to exercise their rights more effectively.'[190]

Finally, it is important to note that Article 22 derived arguments for a duty of interpretability or explainability are not the only arguments to be made for such duties. Commentary on the supposed 'right to interpretation' has often focused on Article 22 and its connection to other GDPR provisions as the sole basis upon which to leverage such a right. However, as noted, the primary principle is the principle of transparency in context: 'The controller should provide the data subject with any further information necessary to ensure fair and transparent processing taking into account the specific circumstances and context in which the personal data are processed.'[191] This principle in conjunction with data subject rights is the primary obligation

directed toward controllers. Article 22 derived obligations are an instantiation of the general principle of transparency but they do not exhaust the principle, nor are they the ceiling for the principle's requirements.

c. Recital 71

Article 22 does not use the words 'explanation', 'interpretability', or any synonym. As we shall see, Articles 13(2)(f), 14(2)(g), and 15(1)(h) use the phrase 'meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing.' However, even this phrase does not necessarily invoke explainability. Where does the supposed 'right to explanation' derive its name?

Goodman and Flaxman (2017) coined the term in their paper *EU regulations on algorithmic decision-making and a 'right to explanation'*, the paper leaning on what is now Recital 71 of the GDPR.[192] Recital 71 primarily aids in the interpretation of Article 22 and related provisions. The relevant part of Recital 71 reads:

> 'Such [automated] processing should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an *explanation* [my emphasis] of the decision reached after such assessment and to challenge the decision.'

In their original paper, Goodman and Flaxman lean heavily on this Recital to bolster their argument that a right to explanation is a required safeguard where processing counts as a) 'a decision based solely on automated processing, including profiling' and b) 'produces legal effect or similarly significantly affects [the data subject].'[193]

This reliance on Recital 71 and the very existence of a right to explanation as construed by Goodman and Flaxman is challenged by Wachter et al (2017).[194] Wachter et al, note that Recital 71 is a recital, not an operative provision but an interpretative aid.[195] Further, Wachter et al also find significance in the apparent intentional non-inclusion of 'a right to explanation of specific decision' in transition from draft versions of the GDPR to the final text.[196] As a consequence, say Wachter et al, Recital 71 is a poor foundation on which to build a right to explanation.

Recital 71 is indeed non-binding, its non-inclusion in the Articles of the GDPR is significant. Alone, Recital 71 is a poor foundation on which to build a right to explanation, and few people argue that Recital 71 alone establishes a right to explanation. Rather, the argument is to interpret operative provisions of the GDPR - the Articles - in light of Recital 71.

There are multiple tools in the operative provisions of the GDPR that might be used to construct a right to explanation. Goodman and Flaxman in their first paper primarily used Recital 71 combined with what is now Article 22.[197] Later, Goodman and Flaxman relied upon the rights to information.[198] Selbst and Powles rely heavily on the rights to information and

access.[199] Other authors emphasise Article 22(3) safeguards.[200] We consider these various possibilities below.

d. Article 22 and Articles 13(2)(f),14(2)(g), and 15(1)(h)

How do the Article 22(1) provisions on automated processing that produce legal effect/similarly significant affect the data subject relate to the 13(2)(f), 14(2)(g), and 15(1)(h) Articles that reference them?

Articles 13(2)(f), 14(2)(g), and 15(1)(h) are identical in the text they contain, requiring the following information be disclosed to the data subject. For example:

> Article 13(2)(f) 'the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.'

There are three notable elements when interpreting these Articles and their relationship to Article 22:

I. The interpretation of 'existence of automated decision-making, including profiling referred to in Article 22(1) and (4)'
II. The interpretation of 'at least in those cases'
III. The interpretation of what is then required: 'meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject'

The first two issues go to when the supposed right triggers, the last goes to what the right requires. We consider the two issues below and the question of what the right requires later at Section 5(g).

i. 'At least in those cases'

Article 13(2)(f) 'the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.'

Wachter et al argue that a right to explanation derived from Article 22 'would only apply to a narrow range of decisions 'solely based on automated processing' and with 'legal' or 'similarly significant' effects.'[201] Does this accurately characterise the relationship between Articles 13(2)(f), 14(2)(g), and 15(1)(h) to Article 22?

Articles 13(2)(f), 14(2)(g), and 15(1)(h) all contain at least two ambiguities regarding how they relate to Article 22.

First, Articles 13(2)(f), 14(2)(g), and 15(1)(h) all contain the phrase 'the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4).' It is unclear exactly what this means. For instance, does this clause refer to the cumulative conditions of a decision that is a) 'based solely on automated processing, including profiling' and b) 'produces legal effects' or 'similarly significantly affects the data subject and so is a very narrow requirement? Or, does this clause merely refer to decisions 'based solely on automated processing' but does not have legal or similarly significant effect and so is a potentially wide requirement? We examine what these conditions mean later but this ambiguity puts the very question of how Articles 13(2)(f), 14(2)(g), and 15(1)(h) relate to Article 22.

Second, notable in Articles 13(2)(f), 14(2)(g), and 15(1)(h) is the use of 'at least in those cases.' This clause references our first point of ambiguity. At first read, this clause might be read as a minimum legal requirement.[202] That is, where processing counts as a) 'a decision based solely on automated processing, including profiling' and b) 'produces legal effect or similarly significantly affects [the data subject],' it is required to provide the information specified in Articles 13(2)(f), 14(2)(g), and 15(1)(h) - situations beyond this are matters of best practice.

To provide some clarity to this ambiguity, we can glean some answers from recitals, WP29 Guidelines, and emerging national supervisory authority guidance. In regards to recitals, Recital 63 provides some support for the position that Articles 13(2)(f), 14(2)(g), and 15(1)(h) are a wide requirement applying to automated processing, including profiling that does not necessarily have legal or similarly significant effect. Recital 63 relates to the interpretation of the right of access. The relevant part of Recital 63 states:

> 'Every data subject should therefore have the right to know and obtain communication in particular with regard to the purposes for which the personal data are processed, where possible the period for which the personal data are processed, the recipients of the personal data, *the logic involved in any automatic personal data processing and, at least when based on profiling, the consequences of such processing* [my emphasis]'

The italicized portion of the Recital indicates that the requirement to disclose 'logic' applies to 'any automatic personal data processing.' Perhaps this constitutes support for a wide application of the requirements to provide meaningful logic and information about envisaged consequences.

However, WP29's *Guidelines on automated individual decision-making and profiling for the purposes for the GDPR* in some places favour a narrow application of requirements. When interpreting Articles 13(2)(f) and 14(2)(g), WP29 Guidelines state:

> '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.'[203]

This wording seems to assume that the requirement to provide meaningful logic is narrow and cumulative.

Ultimately, WP29 Guidelines and tentative guidance from UK's ICO favour a more principled approach rather than defining minimum legal requirements. For instance, emphasis is placed on the connection between transparent processing and fairness as well as the contextual nature of the judgment of what information must be disclosed.[204] In this respect, Recital 60 is key:

> 'The controller should provide the data subject with any further information necessary to ensure fair and transparent processing taking into account the specific circumstances and context in which the personal data are processed.'

Accordingly, perhaps the best way to think about the relationship between Article 22 and any requirements that link to it, is to consider less whether a machine learning system counts as a decision based solely on automated processing that has legal or similarly significant effect. Instead, controllers need to think more about what information would have to be disclosed to facilitate fair and transparent processing in that particular context.

e. Article 22(3) safeguards

Another notable element of the automated individual decision-making requirements is Article 22(3). A brief reminder of how Article 22(3) fits with Article 22 as a whole:

I. Article 22(1) contains a general prohibition/right against decisions that a) are based solely on automated processing' and b) have legal effect or similarly significantly affect the data subject
II. Article 22(2) notes the exceptions to this general prohibition/right, the prohibition/right being disapplied where the processing is
    A. Necessary for the performance of a contract
    B. Authorised by EU or Member State law; or
    C. Is based on the data subject's explicit consent.

Article 22(3) stipulates the conditions that apply where the Article 22(2) exceptions on contract or consent apply. Article 22(3) notes the following:

> 'In the cases referred to in points (a) and (c) of paragraph 2 [contract and consent], the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human

> intervention on the part of the controller, to express his or her point of view and to contest the decision.'

In regards to the Article 22(2)(b) exception on EU and Member State law, this exception includes within it provision for 'suitable measures to safeguard the data subject's rights and freedoms and legitimate interests.' In short, where an exception applies, safeguards must be in place.

The interpretation of these Article 22(3) safeguards is another major tool used to construct a 'right to explanation.' Indeed, the key interpretative aid to Article 22, Recital 71 primarily mentions 'explanation' as a safeguard:

> 'such processing [Article 22(1) processing] should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision.'

Arguably then, explanation is a safeguard that may apply if a controller triggers Article 22(1) and relies on an Article 22(2) exception. We consider what Article 22(3) safeguards might require by way of transparency, interpretability, and explainability in Section 5(g)(3).

f. What processing triggers Art 22(1)?

Article 22(1) '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.'

Section 5(g)(1) recommends that a decision to provide 'meaningful information about the logic involved' and 'envisaged consequences' is best interpreted as a context-specific judgment connected to assessments of what information facilitates fair and transparent processing. Alternatively, what if the requirements to provide an explanation or meaningful logic require the triggering of Article 22(1)? What kind of processing triggers this 'right to explanation'? How much of the processing that takes place in machine learning for healthcare and research gives rise to such a right?

The following analysis considers five points of interpretation of Article 22(1), the interpretations of:

I. 'A decision based'
II. 'Solely on automated processing'
III. 'Including profiling'
IV. 'Legal effect'
V. 'Similarly significant affects'

The two most important elements are 'a decision based solely on automated processing, including profiling' and the clause 'produces legal effects concerning him or her or similarly significantly affects him or her.' These two elements combined limit the application of Article 22(1) to machine learning for healthcare and research (see Table 1 below).

Table 1: Triggering Article 22(1)	**Is based solely on automated processing**	**Is NOT based solely on automated processing**
Produces legal effect and/or similarly significant effect	A Article 22(1) triggered	B Not triggered
Does NOT produce legal effect and/or similarly significant effect	C Not triggered	D Not triggered

i. A decision

> **Article 22(1)** '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.'

Article 22(1) refers to 'a decision based solely on automated processing, including profiling.' The usage of 'decision' here is curious, given that the rest of the GDPR (and its predecessor the DPD) talk in terms of processing rather than decisions. Following Mendoza et al, the usage of 'decision' is a reflection of Article 22's roots in traditional administrative law that typically regulates government decision-making.[205] The roots of Article 22 aside, it is unclear how we should interpret 'decision' here. There is some evidence that the EDPB interprets 'decision' as different from 'processing' defined in Article 4(2). As we cover later in Section 5(f)(i), the EDPB Guidelines on automated processing at one point distinguish between: profiling generally, decision-making based on profiling, and profiling that counts as 'a decision based solely on automated processing.' The EDPB's conceptual schema at least holds out the logical possibility of there being 'solely automated processing' but with no view to a decision being made. It is unclear how solely automated processing that does not make a decision fits with Article 22(1). We consider the interpretation of 'profiling' further at Section 5(f)(iii).

Even if we have a clear understanding of what 'decision' means and where it fits with Article 22(1), what counts as the reference 'decision' at stake here? In the context of healthcare, patient pathways are complex branching strings of judgments and decisions. While we might think of diagnosis as a single activity or decision, in reality, diagnosis is the outcome of a string of decisions and findings. If we are to appraise whether 'a decision' is based solely on automated processing and has legal effect/similarly significantly affects a data subject, we need to know what decision is in question. This is important because a different reference

decision may radically change our assessment of whether the decision is captured by the right to explanation. Consider the following example with different interpretations of the referent 'decision.'

> Example: consider a histopathology machine learning system that interprets biopsies, classifies the sample as cancerous or benign, stratifies those classified as cancerous according to prognosis, and triages patients to different patient pathways (including no treatment) following the former two tasks. There are multiple interpretations of what counts as 'the' decision at stake here:
>
> Interpretation A: there are at least three decisions at play here - classification, stratification, and triaging.
>
> Interpretation B: there are only two decisions at play here, the task of classification being distinguished as lacking decisional elements, perhaps constituting more of a result.
>
> Interpretation C: there is only one 'ultimate decision' being made here, the triaging outcome. In this way, the relevant 'decision loop' concerns itself with overall outcomes, the input of humans into processes generally. In short, the usage of 'a decision' may not allow data subjects to challenge each decision made in a chain but the overall process.

In short, it is not obvious what counts as the 'decision' in this circumstance. In healthcare there are strings of 'decisions' that result in a diagnosis or treatment. Consequently, it is unclear how best to interpret this provision, especially where complex systems are concerned.

A Salient Feature | Roundtable 3

Roundtable 3 participants highlighted the importance of what 'decision' is the reference decision when considering the questions of 'based solely on automated processing' and 'legal effect/similarly significant.' Indeed, participants indicated that focusing on one 'decision' in the context of healthcare may be difficult and somewhat artificial.

In summary, it is unclear what counts as a 'decision', nor is it certain how the term 'decision' relates to the triggering of Article 22(1) Moreover, even if these ambiguities were made clear, it is especially uncertain what the referent 'decision' will be in healthcare or research.

ii. Automation

Article 22(1) '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.'

When does a decision count as being 'based solely on automated processing'? How might this apply to machine learning for health care and research?

1. 'Based solely on automated processing'

A tempting analysis is to take Article 22(1) at face value: if there is 0.001 human in the loop, the decision is not based solely on automated processing. However, there are three reasons to think this simple but tempting analysis is false.

First, Article 22(1) talks about a '*decision based* [my emphasis] solely on automated processing, including profiling.' Consequently, the decision itself need not be solely automated, only the processing upon which the decision is 'solely based.' In this way, processing mentioned in Article 22(1) need not be like clockwork - a mechanistic process from beginning to end. However, it is notable that proposals to include wider wording for the text of the GDPR were rejected. For instance, the European Parliament Committee on Civil Liberties, Justice and Home Affairs preferred wording 'solely or *predominantly* [my emphasis] on automated processing' did not make it into the final text of the Regulation.[206]

Second, WP29 in their *Guidelines on automated individual decision-making and profiling for the purposes of GDPR* ('the Guidelines') elaborate on how best to interpret 'based solely on automated processing.' The WP29 Guidelines start by noting that 'based solely on automated processing' means that there is 'no human involvement in the decision process.'[207] However, the Guidelines row back from this blanket statement, noting that human involvement cannot be 'fabricated', meaning that if a human merely rubber stamped any automatically generated profile to an individual, this decision would still be based solely on automated processing.[208] Further, the Guidelines go on to offer three general comments to clarify what might qualify as human intervention:[209]

I. Oversight of the decision must be 'meaningful' and more than just a 'token gesture;' and
II. Intervention should be from someone who has the 'authority' and 'competence' to change the decision; and
III. This person should consider 'all' the available input and output data.

The WP29 Guidelines should give us confidence that the assessment of what counts as 'based solely on automated processing' implies a more complicated assessment of the meaningfulness of this intervention.

Third, the idea that even token or trivial human influence in the decision would be enough to avoid the automated processing requirements was likely false under the predecessor DPD. For instance, consider the clarification of the 'solely by automatic processing' clause under the preparatory materials for Article 15 DPD:[210]

> '... what is prohibited is the strict application by the user of the results produced by the system. Data processing may provide an aid to decision-making, but it cannot be the end of the matter; human judgment must have its place.'

Article 15 DPD was subject to relatively little challenge through its application. Consequently, we do not have a rich case law to aid our interpretation. However, there was at least one case in the German Federal Court of Justice that sought to interpret the relevant provision of German's Federal Data Protection Act 1990. As reported by Mendoza and Bygrave, the *SCHUFA* case considered automated credit-scoring systems, the court holding on appeal that the system fell outside the German version of the automated processing requirements. In this judgment the court noted that 'the automated elements of the decisional process pertained to the preparation of evidence; the actual decision to provide credit was made by a person.'[211]

Article 22(3) as an interpretative aid

Article 22(3) 'In the cases referred to in points (a) and (c) of paragraph 2, the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.'

While helpful (and authoritative), the WP29 Guidance has a notable gap in its analysis. Namely, consideration of how the Article 22(3) right to obtain human intervention fits with the interpretation of 'decision based solely on automated processing' in Article 22(1). Our view is that proper interpretation of this safeguard assists with the interpretation of 'based solely on automated processing.' However, the introduction of this safeguard introduces a potential paradox into the assessment of whether a system is based solely on automated processing or not. Consider Article 22(3) above.

Article 22(3) only applies to processing that has already been declared solely automated. Specifically, the Article lays down safeguards where the general prohibition/right against this kind of processing does not apply (Article 22(1)) and where certain lawful bases and derogations are relied upon (Article 22(2)).

As a part of these safeguards, Article 22(3) specifically mentions three safeguards:

I. The right for the data subject to obtain human intervention on the part of the controller;

II. The right for the data subject to express their point of view;
III. The right for the data subject to contest the decision.

That is, the decision that has already been declared solely automated - declared to be without meaningful human intervention - must have these three rights. This seems paradoxical. The human intervention mentioned by Article 22(3) seems to check many of the boxes WP29 note in their elaboration of the concept of 'based solely on automated processing' - the intervention appears meaningful and the person appears to have the authority to change the decision. At first glance, the right to contest safeguard appears to make any 'solely automated processing' that it applies to no longer 'solely automated.'

There is a general rule when interpreting legislation: we should interpret provisions to be consistent with one another wherever we can.[212] Given this, is there a way to square the right to human intervention with the elaborations we have on 'based solely automated processing'? One way to square the two is as follows: a decision that *could* include human intervention, but, as a matter of process does not, is still solely automated. To clarify, consider the two hypothetical examples:

> Example A: a machine learning model generates a risk score that is assigned to patients, this score is interpreted in the round with other medically relevant information by a human to determine whether a patient will be offered treatment or not.

> Example B: a machine learning model generates a risk score that is assigned to patients, this risk score automatically approves or denies treatment. Patients may request a review of and contest the decision.

In Example A, we have a decision process that always includes human intervention - this intervention is likely meaningful, more than token, and so on. Moreover, the human intervention is, by default, a part of the decision process. Consequently, the system likely does not count as a decision based solely on automated processing. Example B includes a decision process that *might* include meaningful human intervention. However, this parallel process of human intervention is only available as a parallel process upon request from the patient. Accordingly, the system may count as a decision based solely on automated processing.

2. Application to machine learning for health

We now have the tools to consider what uses of machine learning for health might be caught or form part of a decision process that counts as a 'decision based solely on automated processing.' Drawing on our assessment of the uses for machine learning in healthcare in the Machine Learning Landscape, it is clear that for the short term, most machine learning applications will be assistive only. In this way, machine learning acts as a decision support tool, as a second reader, or as an interpretative aid for a healthcare professional. Accordingly, more often than not, there will be a human who, by default, has the authority and information to provide meaningful oversight of the machine learning system. In the research context, much the same analysis likely applies, with the qualification that systems for investigational use, clinical trials, or research may be more ambitious in their automation than those systems in use for the healthcare system. Of course, the above analysis depends upon how machine

learning systems are implemented, the check and balances that are put in place, and how any of these systems are labelled for use.

iii. Profiling and automated processing

Article 22(1) '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.'

Article 22(1) refers to 'a decision based solely on automated processing, *including profiling* [my emphasis].' What is 'profiling'? How does profiling fit with Article 22(1) and the 'automated processing' condition? Article 4(4) defines profiling:

> *Profiling* means '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'

WP29 clarify that the profiling according to Article 4(4) has three elements:[213]

I. The processing must be *automated*
II. It must be carried out on *personal data*
III. The objective of the processing must be to *evaluate personal aspects* about that person

Further, WP29 also clarify that profiling that does not necessarily have a predictive purpose but simply classifies individuals according to characteristics may count as profiling. Indeed, classification of persons that merely classifies individuals according to personal aspects may still count as 'profiling', even if this assessment does not seek to predict anything.

WP29 Guidelines distinguish between profiling of three types:[214]

I. General profiling
II. Decision-making based on profiling
III. Solely automated decision-making, including profiling

It is unclear what WP29 has in mind with category I., the Guidelines providing no example, nor commentary. Presumably, WP29 has in mind profiling, that is classification of persons on the basis of their personal data with no decision in mind. For instance, profiling for the purpose of research might constitute 'general profiling' that does not directly result in a decision for the data subject. In regards to II. and III. WP29 distinguish between the rule that applies to profiling that is automated yet does not satisfy Article 22(1) and profiling that counts as a 'decision based solely on automated processing.' Consequently, we distinguish between the

rules that apply to profiling that does not trigger Article 22(1) and the rules that apply to profiling as a subset of automated processing that triggers Article 22(1).

1. Application to machine learning for health

How does the interpretation of 'profiling' and its relationship to Article 22(1) relate to machine learning for healthcare and research? Profiling is given broad definition in Article 4(4). Plausibly, many activities in healthcare and health-related research might indeed count as 'profiling' under Article 4(4). For example, many screening programmes seem a natural fit for the definition. However, some of this profiling will not be directed toward a decision or be involved in decision-making relating to that data subject. For instance, in a research context with no duties to report findings or treat patients, the goal of the profiling is scientific research, the profiling has no view to make decisions about the data subject. Moreover, only a subset of this profiling will count as 'solely automated' (we considered this requirement at Section 5(f)(ii)(1)). Finally, a subset of this subset will also be processing that has legal effect or similarly significantly affect the data subject (we consider this requirement at Section 5(f)(iv) below). In short, because the requirements of Article 22(1) are cumulative, profiling caught by the Article is a subset of a subset.

iv. Legal effect/similarly significant means

A machine learning application caught by the 'based solely on automated processing' element may still yet fall outside of Article 22(1). Indeed, the machine learning application must also produce legal effect or similarly significantly affect the data subject to trigger Article 22(1). This begs the question: what does 'legal effect' mean, what does 'similarly significant affects' mean? Moreover, how do these concepts apply to machine learning for healthcare and research?

1. 'Legal effect'

> **Article 22(1)** '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 [my emphasis] or her or similarly significantly affects him or her.'

'Legal effect' is not defined within the GDPR. Nevertheless, WP29 provide a relatively straightforward definition:

> 'Legal effect suggests a processing activity that has an impact on someone's legal rights… A legal effect may also be something that affects a person's legal status or right under a contract.'[215]

WP29 Guidance goes further noting a number of examples of 'legal effect,' including the entitlement or denial of a social benefit.[216] While the definition of 'legal effect' necessarily includes some ambiguity, assessments such as these are not uncommon in law. Nevertheless, the issue with 'legal effect' is finding a reasonable limit on its application. For example, Mendoza and Bygrave note that the combined effect of 'legal effect or similarly significant

effect' is that the decision must be more than trivial for a person's welfare, the more adverse the consequence, the more likely it will be caught by Article 22(1).[217] While this seems to capture the tenor of 'legal effect' and 'similarly significant' together, it perhaps does not capture the breadth of 'legal effect' capturing rights under contract. It seems that the interpretation of 'legal effect' may not be so straightforward, the idea of 'legal effect' also requiring that this effect be significant. To clarify, a nuanced interpretation of 'legal effect' here might exclude some decisions that have legal effect if that effect is trivial, for example, a variation in rights under a contract that has little impact upon the data subject

2. Application to machine learning for health and research

Only a subset of machine learning applications for healthcare and research will likely have 'legal effect.' Consider the following examples of machine learning for healthcare and research that may or may not have legal effect:

> Example A: consider the related example of the failure to deliver a disability facilities grant. Following *R v Birmingham City Council ex parte Mohammed*, the failure to provide a disability facility grant counted as a decision as to whether a person has a particular right or legal entitlement, being subject to judicial review.[218] Consequently, the decision almost certainly has 'legal effect.'

> Example B: consider an analogous example in healthcare: the denial of a hip replacement due to obesity. The policy that underpins any machine learning operationalisation of the decision will likely have legal effect. That is, it concerns the distribution of public authority resources.

> Example C: consider a machine learning system that schedules hospital appointments within a short defined period. In this example, there is likely no legal effect, even though such scheduling may have a small detrimental effect on patient health. No social benefit is denied, no legal status is changed, no rights under contract are altered.

To make a broad assessment, there are some machine learning systems for healthcare and research that will likely have legal effect. As noted in Example A, the field of judicial review may provide some analogous case law useful in providing some foundation for the 'legal effect' test. Specifically, the scope of interests subjected to judicial review has expanded in recent years to include an exercise of power that 'manifests itself in a decision that has a discernible effect on an individual.'[219] While judicial review is a distinct line of case law not to be directly applied to 'legal effect', the kind of judgment being made here seems similar. However, perhaps the majority of machine learning uses for healthcare and research are not an easy fit for 'legal effect.' For instance, systems directed toward diagnosis or treatment, although they may have grave consequences for their data subject, do not directly have legal effect. This suggests that although the concept of 'legal effect' has some resolution to it - the edges of the concept are fuzzy - for many machine learning applications it may be unclear whether they have legal effect or not.

Another complication when considering the legal effect of machine learning in healthcare and research is the use of contract in each context. Consenting is not the same as a contract, although sometimes both are necessary in public sector provision of healthcare and research. Nevertheless, the comparative lack of contracts in the public sector may make this sector less likely to produce legal effect than its private counterpart. Consequently, many machine learning systems in these contexts may be better candidates for 'similarly significant affect.'

3. 'Similarly significantly affects'

> **Article 22(1)** '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* [my emphasis] him or her.'

While assessment of 'legal effect' is not uncommon in law, the inclusion of 'similarly significantly affects' the data subject complicates the matter. The previous Data Protection Directive (DPD) differed, the test being just 'significantly affects,' rather than '*similarly* [my emphasis] significantly affects' under the GDPR.[220] WP29 reads into this difference, noting that it suggests that the threshold for significance must be similar to a decision that has legal effect.[221] That is, the question of significance is inherently tied to the significance of legal effect. Mendoza et al note that this 'may signal an intention that such consequences must have a non-trivial impact on the status of a person relative to other persons – just as legal effects typically do.'[222] However, the tying of significance to legal effect is potentially unhelpful. Legal effect does not guarantee that the effect is significant. A right modified under contract may be insignificant and unimportant yet, undeniably, has legal effect - WP29 emphasises as such. In this regard, the drafters appear to have made the assumption that all legal effects are significant - an assumption that is patently false. It is more accurate to think of legal effect as legal consequences, legal consequences tend to be significant, although they need not be. Consequently, the test for 'similarly significant effect' is a vexing one - the test potentially boiling down to the significance of the interest at stake. As noted by Mendoza and Bygrave (2017), it is likely that the effect will have to have a 'more than trivial impact' upon the data subject, the more adverse the consequence, the higher the probability the processing will count as having 'legal effect of similarly significant effect.'[223] Notably, under the Directive, Church and Millard (2010) argued that 'similarly significant' need not be 'pecuniary,' that is, involve monetary loss.[224] As the argument goes, a significant consequence might be '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.'[225] However, Bygrave (2020) note that this interpretation has been made less probable with the inclusion of 'similarly', the inclusion of this term perhaps ruling out mere emotional distress like that described above.[226] Given the lack of clarity, it is difficult to say what machine learning systems in healthcare or research will be caught by this provision.

4. Application to machine learning for health

As described above, it is unclear how we should interpret 'similarly significantly affects.' Moreover, there is little case law under the Directive and sparse guidance from the EDPB to direct us. Perhaps the most conservative definition that provides the concept its widest breadth, is that the decision must have 'more than trivial impact' on the data subject. While not directly applicable, Joel Feinberg's account of 'harm' and 'injury' might provide some assistance. For instance, 'affect' might be interpreted as 'injury' defined as a 'setback to an

interest.'[227] Notably, there is also a threshold at which these setbacks might become candidates for harm, notably, mere inconvenience, irritation, or annoyance fall below such a threshold.[xvii] In this way, a setback to an interest that constitutes more than mere inconvenience might provide some framework to structure assessment of 'more than trivial.' In addition to finding a threshold for 'more than trivial', we might also identify instances that almost certainly have similarly significant effect. That is, there are likely core interests whose setting back counts as more than trivial. For instance, Sen and Nussbaum's capabilities approach outlines the goods necessary for people to achieve a minimally flourishing life.[228] Nussbaum provides a list of 10 central capabilities: life, bodily health, bodily integrity, senses/imagination/thought, emotions, practical reason, affiliation, other species, play, and control over one's environment.[229] Notably, healthcare and health research can significantly impact upon many of these capabilities, which is to say that decisions in healthcare and research often impact upon one's ability to live a minimally flourishing life. Consequently, machine learning in these sectors compared to other sectors tend to have similarly significant effect. To summarise the above, the more core the interest at stake in the decision, the more likely the decision will have 'similarly significant effect.'

v. Consideration of the two elements together

Let us revisit what machine learning for healthcare and research might trigger Article 22(1) (see Table 1 below).

Table 1: Triggering Article 22(1)	**Is a decision based solely on automated processing**	**Is NOT a decision based solely on automated processing**
Does produce legal effect and/or similarly significant effect	A Article 22(1) triggered	B Not triggered
Does NOT produce legal effect and/or similarly significant effect	C Not triggered	D Not triggered

Machine learning in category A will necessarily be a subset of a subset. That is, only a subset of machine learning applications will count as 'a decision based solely on automated processing.' Further, only a subset of this subset will have legal effect or similarly significantly affect the data subject.

[xvii] N.B. The setback to the interest must also be 'wrongful' to constitute a harm on Feinberg's account. Feinberg J. *The Moral Limits of the Criminal Law Volume 1: Harm to Others*. Oxford: Oxford University Press; 1987, 31-64.

g. What the right to explanation requires

Once triggered, what do the automated individual decision-making requirements require of controllers, what information does it provide to data subjects?

Critically, what the right requires depends upon how we formulate the right. As outlined earlier, there are four main sources to formulate such a right, namely:

I. The general principle of transparency contextualised
II. Recital 71 and Article 22
III. Article 22 and Articles 13(2)(f), 14(2)(g), and 15(1)(g)
IV. Article 22(1) and Article 22(3)

We analyse what the 'right to explanation' requires according to these constructions below.

i. Logic and consequences

If the right to explanation is derived from Article 22(1) and Articles 13(2)(f), 14(2)(g), and 15(1)(g), much hangs on the requirement to provide 'meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.' There are two elements to this provision:

I. The requirement to provide 'meaningful information about the logic involved'
II. The requirement to provide information on the 'significance and envisaged consequences' of such processing for the data subject

We consider each in turn.

1. Meaningful information about the logic involved

How ought we interpret the clause 'meaningful information about the logic involved'? We may further breakdown the elements of this phrase into its constituent components:

I. What constitutes 'logic involved'
II. To whom must the information be meaningful?

We consider each in turn.

'Meaningful information about the logic involved'

Article 13(2)(f) 'the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, *meaningful information about the logic involved* [my emphasis], as well as the significance and the envisaged consequences of such processing for the data subject'

WP29 in their *Guidelines on automated individual decision-making* give the following interpretation of 'logic involved':

> 'The controller should find simple ways to tell the data subject about the rationale behind, or the criteria relied on in reaching the decision without necessarily always attempting a complex explanation of the algorithms used or disclosure of the full algorithm.'[230]

In relation to the requirement under the DPD, Bygrave elaborates on what 'logic involved' might require, noting:

> 'decision makers themselves must be able to comprehend the logic of the automated steps involved. This further means, in effect, that the logic be documented and that the documentation be kept readily available for consultation and communication (both inside and outside the decision maker's organization). The documentation must set out, at the very least, the data categories which are applied, together with information about the role these categories play in the decision(s) concerned.'

This interpretation seems sensible. However, it is unclear how it might be apply to machine learning. For instance, Kamarinou et al highlight that 'logic involved' could refer to:[231]

I. The data used to train the algorithm
II. The way in which the algorithm itself works in general
III. The specific policies and criteria that fed into the algorithm
IV. The variables and weights attributed to these variables

It is likely that 'logic involved' might mean all or none of these possibilities. To clarify, the kind of information 'logic involved' refers to is likely to be context-sensitive. In this way, the question of what counts as 'logic involved' ultimately reflects what information is 'necessary to ensure fair and transparent processing taking into account the specific circumstances and context.'[232] Further, while the Guidelines note that controllers should not always attempt 'a complex explanation of the algorithms, the advice given in relation to the rights to information and access likely also applies in this context. That is, a layered explanation may be best practice - a simplified, accessible explanation, and, if appropriate, a technical explanation for data subjects that want a more thorough, albeit complex disclosure. The context-sensitive approach is also mirrored in the approach of national authorities, authorities like ICO providing a list of methods that might assist with transparency of machine learning but not settling on

one method to cover the breadth of applications of machine learning.[233] In short, the proper interpretation of 'logic involved' does not give us one method, one kind of information that should be disclosed - it is a context-sensitive judgment, requiring different information in different circumstances, and perhaps a layered approach to explanation for best practice.[xviii]

To whom must the 'meaningful information about logic involved' be meaningful? Kamarinou et al note that this judgment should be assessed from the data subject's perspective.[234] Indeed, WP29 Guidelines agree: 'the information provided should, however, be meaningful to the data subject.[235] While the test is unlikely to be subjective, that is, what any given data subject themselves thinks is meaningful, it may be helpful to think about the following elements to approximate an answer to what counts as 'meaningful':

- What interests of the data subject are at stake with respect to the decision?
- Does the information disclosed provide the data subject with a good idea of how the decision was arrived at?
- Does the information disclosed allow the data subject to interrogate the system for fairness?
- Does the disclosure allow the data subject to challenge the decision in an informed way?

The methods discussed in the Interpretable Machine Learning report may, to varying extents, satisfy different questions. Apart from these questions the WP29 Guidelines emphasise the manner of communication when disclosing 'meaningful information.'[236] Specifically in regards to automated processing, WP29 emphasises clarity and simplicity, highlighting that the controller should:

> 'find simple ways to tell the data subject about the rationale behind, or the criteria relied on in reaching the decision without necessarily always attempting a complex explanation of the algorithms used or disclosure of the full algorithm.'[237]

Further, best practice may dictate that disclosure of 'meaningful information' should be layered - allowing data subjects of access different information, depending on their needs and wants.[238] In this way, seeking *a* or *the* solution to meeting the requirements of 'meaningful information about the logic involved' is likely a flawed approach. Methods to render machine learning interpretable may represent an important and useful way to provide such information. However, this information must be contextualised to be 'meaningful.' For example, it may be extremely helpful for data subjects to be told what a machine learning model found significant for their instance of processing (local interpretability). But, these findings must also be supported with an idea of how the model's finding contributed to the decision in question. That is, local interpretability is useful but only meaningful if we understand how it contributed to the decision in question.[xix] In short, technical disclosure using interpretable machine learning must

[xviii] N.B. The tasks suggested by ICO's *Project ExplAIn* are also likely helpful to ensure 'meaningful information' is supplied. See: Information Commissioner's Office, The Alan Turing Institute. *Project ExplAIn: Explaining Decisions Made with AI.* 2020.

[xix] N.B. Controllers may find the list of explanation strategies and tools listed in ICO's Project Explain Guidance instructive. See: Information Commissioner's Office, The Alan Turing Institute. *Project ExplAIn: Explaining Decisions Made with AI.* 2020, 120-122.

be supported by the basics of communication - how the machine learning system is used and how it contributes to the decision.

Finally, controllers should also note the restrictions on disclosure of trade secrets and intellectual property discussed in Section 4(b)(i)(4) above. As noted earlier, these restrictions, while primarily relating to the right of access, are also relevant more generally, especially with respect to Article 13(2)(f) and 14(2)(g). Accordingly, trade secrets and intellectual property, particularly the trade secrets/intellectual property of a third party, may restrict what is disclosed under 'meaningful logic.'

2. 'Significance and the envisaged consequences'

Article 13(2)(f) 'the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the *significance and the envisaged consequences* [my emphasis] of such processing for the data subject'

Connected to the requirement to provide 'meaningful information about the logic involved' is the requirement to provide information on 'the significance and the envisaged consequences of such processing.' What does this requirement mean in the context of healthcare or research? WP29 in their *Guidelines on automated individual decision-making* provides some idea on what 'significance' and 'envisaged consequences' mean. Specifically, the Guidelines note:

> 'information must be provided about intended or future processing, and how the automated decision-making might affect the data subject. In order to make this information meaningful and understandable, real, tangible examples of the type of possible effects should be given.'[239]

Further, the Guidelines also provide an example in the context of insurance:

> 'An insurance company uses an automated decision making process to set motor insurance premiums based on monitoring customers' driving behaviour. To illustrate the significance and envisaged consequences of the processing it explains that dangerous driving may result in higher insurance payments and provides an app comparing fictional drivers, including one with dangerous driving habits such as fast acceleration and last-minute braking.
>
> It uses graphics to give tips on how to improve these habits and consequently how to lower insurance premiums.'

Notably, this example highlights the importance of communicating to data subjects the inputs and their relationship to the outputs, as well as the decision in question. This kind of

information is relatively simple to provide where there is a linear relationship between inputs and outputs of a simple model. However, where machine learning is concerned, as we noted in the Interpretable Machine Learning report, the relationship between inputs and outputs is less clear and more complex. For example, the features of a machine learning model may be filtered through nodes, each node with different weights, some weights being dependent on the weighting of other nodes. Consequently, the simple, linear relationship between input and output that WP29 envision in their example may not be a good fit for some machine learning models. In this way, a rule-based post hoc method for interpretation may better suit disclosure of 'significance' and 'envisaged consequences' of processing.

3. The right to contest and Article 22(3)

Article 22(3) In the cases referred to in points (a) and (c) of paragraph 2, the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.

As noted in Section 5(e) earlier, some construe any supposed 'right to explanation' as a safeguard as mentioned in Article 22(3). Indeed, as we noted, this interpretation is implied by Article 22(3) and a number of recitals, notably Recital 71:

> 'In any case, such processing [in relation to Article 22(1)] should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision.'

Moreover, the Council of Europe when considering the modification of Convention 108 to reflect the GDPR, notes the following about explanation:[240]

> 'Data subjects should be entitled to know the reasoning underlying the processing of their data, including the consequences of such a reasoning, which led to any resulting conclusions, in particular in cases involving the use of algorithms for automated-decision making including profiling. For instance in the case of credit scoring, they should be entitled to know the logic underpinning the processing of their data and resulting in a 'yes' or 'no' decision, and not simply information on the decision itself. Without an understanding of these elements there could be no effective exercise of other essential safeguards such as the right to object and the right to complain to a competent authority.'

This statement is significant as it situates explanation as the lynchpin that underpins other safeguards, principally the right to contest. Notably, the interpretation of explanation as a safeguard also means that where the rights to information or access are blocked or restricted, the safeguards may still apply. That is, explanation as a safeguard may act as a possible

complementary but also independent source of any duty to explain found in Articles 13(2)(f), 14(2)(g), and 15(1)(h). What would explanation as a safeguard require of controllers? How might this explanation differ from any explanatory elements found under the rights to information and access?

As noted in Section 5(e), Article 22(3) specifically mentions the 'right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.' It is the latter part of Article 22(3) referencing the right to 'contest the decision' that Recital 71 appears to link to explanation. Accordingly, we can read 'explanation as a safeguard' as primarily being concerned with data subjects possessing sufficient information to contest the decision. Consequently, this interpretative slant on explanation potentially provides something novel, something over and above the kind of explanation required by Articles 13(2)f), 14(2)(g), and 15(1)(h).

Interpreting explanation under Article 22(3) emphasises the ability for data subjects to contest the decision. This is potentially significant for the machine learning context. The controller in this interpretation must provide sufficient information to allow the data subject to contest the decision. As noted in the Interpretable Machine Learning report, some methods of interpretability explain what the model generally finds significant (global interpretability), some explain what the model found significant for that instance of processing (local interpretability), and some do both. Undoubtedly, both forms of information might assist the data subject in contesting the decision. However, arguably more is required. For instance, especially where the decision relates to the division of resources, the ability to interrogate the model for fairness is likely relevant. Accordingly, the ability to contrast a data subject's decision with that of a similarly situated data subject may be an important feature. Indeed, in this regard, elegant solutions like counterfactual explanations may not provide sufficient information to contest the decision.

ii. Ex ante and ex post explanation

One major question in regards to the supposed 'right to explanation' is whether any such requirement is restricted to provide explanation before processing (*ex ante*) or extends to after processing (*ex post*). Indeed, the title of Wachter et al's (2017) paper is '*Why a Right to Explanation of Automated Decision-Making Does Not Exist in the GDPR*', yet the conclusion of this paper is more complicated than the title suggests. That is, Wachter et al's conclusion is that there is no 'right to explanation' but a more limited 'right to be informed.'[241] The difference between the two rights, says Wachter et al, is that the right to be informed only provides information that is possible to disclose prior to processing.[242] Consequently, information to be disclosed under the right to explanation is limited to 'system functionality', information on the general functionality of the model, not 'specific decisions', the reasons given for a specific decision.[243]

Why does the difference matter? As addressed in the Interpretable Machine Learning report, global interpretability of a machine learning model is not necessarily the same as local interpretability of a model. That is, global interpretability considers the 'whole logic of the model', local interpretability considers the reasons for a specific decision.[244] Global interpretability is no mean feat in relation to some machine learning algorithms. However, at a more general level, global interpretability of certain machine learning algorithms may be

satisfied more easily than local interpretability. For instance, the release of a decision tree or breakdown of what a model generally finds significant may be easier than analysing each instance of processing for local interpretability. Moreover, while the tools to provide some global interpretability may sometimes also provide some local interpretability (for example, LIME), they can also require different tools, different methods. Consequently, limiting any right to explanation to *ex ante* explanation usefully limits the problem of interpretability to a manageable, digestible level for controllers.

If there is a right to explanation, is it only restricted to *ex ante* explanation? There is good reason to doubt the claim that any explanation will be restricted to *ex ante* information. As discussed in Section 4(b)(iii)(3), we noted a number of differences between the rights to information and access. For instance, that the rights differ in their timing and that this timing may require very different information to be disclosed. Wachter downplays these differences, restricting the obligation to provide 'meaning logic' under the right of access to having the same effect as under the right to information. Mendoza and Bygrave (2017) also provide a number of reasons to doubt the conclusions of Wachter et al (2017).[245] The tenor of Mendoza and Bygrave's critique being that Wachter et al ignore evidence that *ex post* explanation is certainly a possibility. For instance, the wording of Article 15(1)(h) seems to indicate that a decision has already taken place in its wording of 'existence of' automated processing and requirement to provide the 'significance and envisaged consequences' of the decision. Further, taking into account Article 22(3) and explanation being tied to a right to contest, Mendoza and Bygrave also highlight that explanation here acts as an *ex post* means akin to a right to appeal. In short, the requirement to provide *ex post* explanation should not be ruled out. Again, a layered approach, providing an explanation *ex ante* and *ex post* may best satisfy the myriad requirements of the GDPR.

h. Complications of legal bases, derogations, and automated individual decision-making conditions

Suppose a controller triggers Article 22(1), that is, the machine learning system counts as a) a decision based solely on automated processing, including profiling and b) produces legal or similarly significantly affects the data subject. Suppose further that the controller has a good idea about their obligations in regards to transparency and explainability, implementing appropriate measures along these lines. Still, there are barriers to lawful processing. There are two further barriers to note. First, the exceptions to be met in Article 22(2) and second, the restrictions that apply to special category data under Article 22(4).

i. Article 22(2) restrictions

As outlined in Section 5(a), to lawfully process data that triggers Article 22(1), an Article 22(2) exception must apply. These exceptions are where the processing is:

A. Necessary for entering into or performance of a contract
B. Authorised by EU or Member State law
C. Based on the data subject's explicit consent

As noted earlier, these conditions do not fit well with the favoured Article 6 legal bases and Article 9 derogations for healthcare and research controllers. For instance, consent is given a high bar and it may be difficult to comply with its rigours - the Information Governance Alliance and the HRA noting as much in their guidance.[246] Further, contract is also given a challenging interpretation by recent EDPB Guidelines (albeit in the context of online services).[247] For example, these Guidelines place emphasis on the clause 'necessary' meaning very few clauses count as 'necessary' for entering into or performance of a contract. In this way, controllers either have to rely on different legal bases or derogations or must build in some consent or contract into their procedures to fall within Article 22(2)(a) or (c) exceptions.

In regards to Article 22(2)(b) on Member State authorisation, Section 14 of the DPA 2018 legislates this exception into the law for the UK. Section 14 has two elements of note.

First, Section 14(3) stipulates what counts as a 'qualifying significant decision' to render the automated decision exempt under Article 22(2)(b). Section 14(3) has three cumulative conditions that add little to Article 22(2)(b), a decision is a 'qualifying significant decision' if:

A. It is a significant decision in relation to a data subject. However, Article 22(1) already requires that the decision have legal or similarly significantly affect the data subject
B. It is required or authorised by law. Notably, this is stated in the text of Article 22(2)(b)
C. It does not fall within Article 22(2)(a)-(c). This requirement merely stipulates that you cannot rely on Article 22(2)(b) - authorisation by Member State law - if consent or contract apply instead

As demonstrated, the national implementation of Article 22(2)(b) is permissive, adding little to the text of the GDPR. Consequently, those controllers whose processing is caught by Article 22(1) and is authorised by law, may find the Article 22(2)(b) a useful exception. Notably, the permissive nature of Section 14(3) does not help many private sector commercial companies – much of the processing conducted by these bodies failing to be 'required or authorised by law.' Accordingly, Section 14 is of little assistance to commercial entities caught by Article 22(1).

Second, Section 14(4) lays down further conditions where the Article 22(2)(b)/Section 14(3) exception applies. Namely, the controller must notify the data subject of the decision based solely on automated processing, allow the data subject to request reconsideration of the decision or take a new decision not based solely on automated processing.[248]

ii. Special category data

As outlined in Section 2(c), special category data are also subject to further restrictions and safeguards. For our purposes, genetic data, biometric data, and data concerning health are explicitly included in Article 9(1) as special category data. If a controller processes any special category data (which, given the context of healthcare and health research, this is probable), an Article 9(2) derogation must be found. In addition, there are also specific restrictions that attach to special category data where Article 22(1) and (2) apply. Namely, Article 22(4):

Article 22(4) Decisions referred to in paragraph 2 shall not be based on special categories of personal data referred to in Article 9(1), unless point (a) or (g) of Article 9(2) applies and suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place.

To explain, Article 9(2)(a) and (g) are two possible derogations when processing special category data. These derogations are:

(a) Explicit consent

(g) Substantial public interest

It is notable that Article 9(2)(a) and (g) does not include the derogations most favoured by controllers in healthcare and research, namely:[249]

(h) Preventative or occupational medicine

(i) Public health

(j) Research purposes

The effect of Article 22(4) being that those controllers relying on Article 9(2)(b)-(f) or (h)-(j) derogations will now have to also comply with another Article 9(2) derogation if their system counts as a decision based solely on automated processing that has legal effect or similarly significant effect. This may be difficult task, as the eligible Article 9(2)(a) and (g) derogations sometimes simply do not apply in the context of healthcare or research processing, or are an odd fit for the context, are a high bar to successfully rely upon, or may leave processing vulnerable to data subject rights.

The combined effect of Article 22(1), (2), and (4) is as follows. If a controller processes personal data making a decision based solely on automated processing that has legal effect or similarly significant effect, Article 22(1) will apply. To lawfully process using this automated system, the controller must fall under an Article 22(2) exception. Namely, the processing must be necessary for the performance of a contract, authorised by EU or Member State law, or gain the data subject's explicit consent. If the processing involves special category data, the controller will also have to comply with one of Article 9(2)(a) and (g) derogations. As noted, this provides further inflexibility and conditions for controllers to meet in order to lawfully process data. In short, if machine learning for healthcare or research does trigger Article 22(1), the legal requirements placed upon controllers intensifies, the legal options available to lawfully process narrow.

Section 5 key messages:

- **Provisions relating to automated individual decision-making are often the most prominent tools used to construct a 'right to explanation.' There are two broad questions to consider when exploring whether such a right exists. First, when the right is triggered and second, what the right requires once triggered.**
- **In regards to triggering, provisions used to evidence a right to explanation are spread across Article 22, Recital 71, Articles 13(2)(f), 14(2)(g), and 15(1)(h). However, it is unclear how we should read such provisions – commentators use the same Recitals and Articles to either affirm or deny that there is such a right.**

- **Article 22 lays down the conditions for automated individual decision-making. However, it is one manifestation of the general principle of transparency – not the beginning nor the end of any requirement to make interpretable or to explain. In short, we should not fixate on Article 22(1) as being the only source for interpretability or transparency.**
- **Article 22(1) captures only a narrow range of processing. Namely, 'a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.'**
- **When interpreting 'a decision', it is manifestly unclear what counts as 'a decision' and how to frame such 'decisions' in the context of healthcare and research. Indeed, in these contexts, it is common to have strings of decisions rather than just one decision.**
- **'Based solely on automated processing', is interpreted by WP29 to mean that any human in the loop to count must have authority and have meaningful input. In the context of machine learning for healthcare and research, in the near-term, most machine learning is assistive, requiring healthcare professionals to contextualise and interpret its results. Typically, healthcare professionals will have the skills and authority to meaningfully intervene.**
- **'Legal effect' relates to a change in legal rights, status, or rights under contract for a data subject. The term is inherently fuzzy but in the context of machine learning for healthcare and research, those systems that approve or deny a social benefit (including healthcare) may count as having 'legal effect.'**

- **'Similarly significant effect,' is difficult to interpret with the addition of 'similarly.' Nevertheless, the more core the interest at stake, the more likely the decision will have 'similarly significant effect.' With respect to what the right requires if triggered, there are three elements to interpret: 'meaningful information about the logic involved', 'significance and the envisaged consequences', the 'right to contest under Article 22(3).'**
  - **'Meaningful information about the logic involved' as applied to machine learning may require the disclosure of different kinds of information utilising different methods discussed in the Interpretable Machine Learning report. However, the usage of 'meaningful' likely requires a user-centric, layered approach. Accordingly, there may not be a one-size-fits-all approach to render machine learning interpretable.**
  - **'Significance and the envisaged consequences' as interpreted by WP29 appears to require some idea of how inputs into the model influence its outputs and the eventual decision. In the context of machine learning, this may be difficult as there is often not a linear relationship between an input and a particular output.**
  - **Emphasising the right to contest under Article 22(3) may add extra interpretative depth to any 'right to explanation', perhaps requiring disclosure of information to allow data subjects to interrogate the model for fairness.**
- **In regards to timing, we note that any right to explanation may require both explanation before processing and explanation after processing. Consequently, the GDPR may require both global interpretability of the model overall but also local interpretability of particular instances of processing.**
- **If a controller is caught by Article 22(1), the requirements of Article 22(2) narrow and complicate the legal position of the controller, especially if the controller processes special category data.**

6. The GDPR and tools for transparency

This report has sought to provide a comprehensive analysis of what the GDPR requires by way of transparency, interpretability, and explainability as it applies to machine learning for healthcare or research. We summarise our findings as follows:

Machine learning for healthcare and research may be subject to many different spheres of regulation that might require transparency, interpretability or explainability. The GDPR is but one of these pieces of regulation, albeit one of the most prominent.

The general principle of transparency underpins and informs any duty of transparency, interpretability, or explanation. Notably, these duties will be context-sensitive, Recital 60 clarifies the application of the principle, noting: 'The controller should provide the data subject with any further information necessary to ensure fair and transparent processing taking into account the specific circumstances and context in which the personal data are processed.'

The general principle of transparency is instantiated by associated data subject rights. However, data subject rights are often blocked, restricted, or derogated from, depending on the legal position of the controller. Consequently, the protections the rights offer can vary.

Data subject rights, specifically the rights to information, the right of access (Articles 13(2)(f), 14(2)(g), and 15(1)(h) aside), and right to data portability do not directly require interpretability or explainability of machine learning systems. However, these rights do offer useful tools for those wishing to leverage some form of interpretability or explanation.

Data subject rights, specifically the rights to rectification, object, erasure, and restriction of processing do not themselves require the disclosure of information that might constitute 'interpretability' or an 'explanation.' Nevertheless, some interpretability or explanation may be necessary to vindicate these rights. That is, to know whether personal data is correct or to properly test the balancing exercise at issue in the right to object, some explanation may be necessary. This aside, any such requirement is highly context-dependent and so subject to the facts of that particular case. In other words, interpretability or explanation will have to be necessary in that data subject's particular case to vindicate their rights.

The provisions on automated individual decision-making are one manifestation of the general principle of transparency. These provisions do not exhaust, nor are they the only source for potential duties to render interpretable, or provide an explanation. These provisions represent a particular concern regarding the lack of human involvement in decision-making, restricting automated processing and providing the tools to challenge such processing in some circumstances.

There is widespread disagreement over how to interpret the automated individual decision-making requirements, namely: Article 22(1), Recital 71, Articles 13(2)(f), 14(2)(g), and

15(1)(h), and Article 22(3). This disagreement is a particular flashpoint when considering whether there is a right to explanation or not.

A major interpretative issue is how to interpret Article 22(1)'s main subject matter: 'a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.' The interpretation of 'a decision', 'solely on automated processing, including profiling', and 'legal effect or similarly significantly affects' are all major sources of interpretative uncertainty. These points of confusion are yet to be clarified by national courts or the CJEU. Nevertheless, reasonable interpretations lead us to believe that only a subset of a subset of machine learning for healthcare and research will trigger Article 22(1). Namely, many machine learning systems will constitute a 'decision' or have 'similarly significant effect' but any one system is likely to count as a) a decision that b) is based solely on automated processing, including profiling, and c) has legal effect or similarly significantly affects the data subject. Indeed, for the short term, clinicians remain the ultimate arbiter for medical interventions, using machine learning to merely assist but not make most decisions.

It is also unclear what Article 22(1) requires if it does apply. Broadly, if we consider Articles 13(2)(f), 14(2)(g), and 15(1)(h), controllers must provide both 'meaningful information about the logic involved' as well as 'the significance and the envisaged consequences of such processing for the data subject.' It is unclear how each requirement applies to machine learning for healthcare but any answer likely requires special attention be paid to the context in which the processing takes place. Further Article 22(3) also emphasises that data subjects should be given sufficient information to contest the decision being made. This interpretative take on interpretability is potentially demanding, perhaps straying into the ability for the data subject to test the model for fairness. Moreover, it is also possible that controllers will have to explain the outputs of their systems after the fact (*ex post*). Accordingly, explanation of particular instances of processing for particular data subjects is potentially required.

Finally, we also note the unenviable position that some controllers may be in if they trigger Article 22(1). Broadly, controllers, especially if they do not fit within Section 14 DPA 2018, may be left with a complex web of legal bases, derogations, and special requirements under Article 22(2) to satisfy. In short, their legal position narrows and increases in complexity.

Overall, we recommend that controllers consider interpretability or explainability of their machine learning system throughout the development and lifecycle of their system.[xx] Specifically, transparency and associated requirements should be interpreted, being sensitive to the interests of their data subjects, keeping in mind the general precept to provide the 'data subject with any further information necessary to ensure fair and transparent processing, whilst also taking into account the specific circumstances and context in which the personal data are processed.'[250]

[xx] N.B. The Information Commissioner's Office *Project ExplAIn* also provides a list of principles to keep in mind when explaining machine learning. See: Information Commissioner's Office, The Alan Turing Institute. *Project ExplAIn: Explaining Decisions made with AI*. 2020, 38-44.

References

[1] Floridi L. Soft ethics, the governance of the digital and the General Data Protection Regulation. Philosophical Transactions of the Royal Society: Mathematical, Physical and Engineering Sciences. 2018; 376(2133):20180081.

[2] Protocol amending the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data [2018] CETS No.223.

Convention for the Protection of Individuals with regards to Automatic Processing of Personal Data [1981] ETS No.108.

[3] Committee of Ministers Recommendation CM/Rec (2019)2 of the Committee of Ministers to member States on the protection of health-related data [2019]

[4] Ordish J, Murfet H, Hall A. *Algorithms as medical devices*. PHG Foundation. 2019.

[5] Cobbe J. Administrative law and the machines of government: judicial review of automated public-sector decision-making. *Legal Studies*. 2019; 39(4): 21-22.

[6] *Montgomery v Lanarkshire Health Board* [2015] UKSC 11, [2015] A.C. 1430

Ordish J. AI for health: Is there a regulatory gap? *Digital Health Legal*. 2018; 5(6): 3-5.

[7] European Commission. Harmonised Standards. Available from: https://ec.europa.eu/growth/single-market/european-standards/harmonised-standards_en [Accessed 9th February 2020]

[8] British Standards Institution. BS ISO/IEC 29100:2011+A1:2018 Information technology - Security techniques - Privacy framework. London: BSI; 2018.

British Standards Institution. BS ISO/IEC 27701:2019 Security techniques - Extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy information management - Requirements and guidelines. London: BSI; 2019.

[9] Lynskey O. Deconstructing data protection: the 'added-value' of a right to data protection in the EU legal order. *International & Comparative Law Quarterly*. 2014; 63(3): 569-597.

[10] Bygrave LA. The place of privacy in data protection law. *University of New South Wales Law Journal*. 2001; 24(1): 14-16.

[11] Bygrave LA. *Data Protection Law: Approaching its Rationale, Logic, and Limits*. New York: Kluwer Law International; 2002: 144-159.

[12] Council Regulation (EC) 2016/679 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 (General Data Protection Regulation) (GDPR) [2016] OJ L119/1, art 20.

[13] GDPR, art 16, art 5(1)(d).

[14] GDPR, art 5(1)(c), art 5(1)(b), art 6, art 9.

[15] Data Protection Act 2018, s 114-141.

GDPR, art 51-67.

The Information Commissioner's Office. *Legislation We Cover*. Available from: https://ico.org.uk/about-the-ico/what-we-do/legislation-we-cover/ [Accessed 5 February 2020]

[16] Council Directive (EC) 2002/58 concerning the processing of personal data and the protection of privacy in the electronic communications sector [2002] OJ L201/37 (as amended by Council Directive (EC) 2009/136 amending Directive 2002/58/EC).

[17] European Commission, 'Proposal for a Regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC' (Communication) COM (2017) 10 final.

[18] *University of Bristol v John Peters* EA/2018/0142.

[19] Civil Procedure Rules 31.16-17.

Durham County Council v Dunn [2012] EWCA Civ 1654, [2013] 1 W.L.R. 2305.

[20] European Commission, 'Artificial Intelligence for Europe' (Communication) COM(2018) 237 final.

[21] European Parliament, 'Motion for a Resolution on enabling the digital transformation of health and care in the Digital Single Market; empowering citizens and building a healthier society' (Resolution) B9-0239/2019.

[22] European Parliament, 'Motion for a Resolution on enabling the digital transformation of health and care in the Digital Single Market; empowering citizens and building a healthier society' (Resolution) B9-0239/2019, para 21.

[23] Convention for the Protection of Individuals with regards to Automatic Processing of Personal Data [1981] ETS No.108.

[24] Codified Version of the Treaty Establishing the European Community [1992] OJ C224, art 7a.

[25] Directive (EC) 95/46 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (DPD) [1995] OJ L281, recital 11.

[26] European Commission, 'Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (GDPR)' (Communication) COM (2012) 11 final, 1-2.

[27] European Commission, 'Safeguarding Privacy in a Connected World A European Data Protection Framework for the 21st Century' (Communication) COM(2012) 9 final, 4-5.

[28] European Commission, 'Safeguarding Privacy in a Connected World A European Data Protection Framework for the 21st Century' (Communication) COM(2012) 9 final, 2-5.

[29] Selbst A, Powles J. Meaningful Information and the right to explanation. *International Data Privacy Law*. 2017; 7(4): 236.

[30] Case 39/72 *Commission v Italy* [1973] ECR 00101, para 17.

[31] Department of Health & Social Care, *Guidance: Code of conduct for data-driven health and care technology*. Available from: https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology. N.B. Principle 7 cites Chapter 3 of Part 4 of the Data Protection Act 2018, Part 4 only relating to intelligence services processing.

Fisher M. *The Data Protection Act v Machine Learning Algorithms*. Available from: https://ukhumanrightsblog.com/2019/05/10/the-data-protection-act-v-machine-learning-algorithms/ [Accessed 13th May 2019]. N.B. The blog has since been corrected, although there is no note acknowledging the corrections.

[32] The Data Protection, Privacy and Electronic Communications (Amendments etc) (EU Exit) Regulations 2019, SI 2019/419, schedule 1.

[33] GDPR, art 4(1).

[34] GDPR, recital 26.

[35] Mitchell C, Ordish J, Hall A. *Data protection and genomic data*. PHG Foundation. [Preprint] 2020.

Mourby M, Mackey E, Elliot M, et al. Are 'pseudonymised' data always personal data? Implications of the GDPR for administrative data research in the UK. *Computer Law & Security Review*. 2018; 34(2): 222-233.

[36] C-230/14 *Weltimmo s.r.o. v Nemzeti Adatvédelmi és Információszabadság Hatóság* [2015] ECR I-639, para 31.

European Data Protection Board. *Guidelines 3/2018 on the territorial scope of the GDPR (Article 3)*. 2018: 5-6.

[37] Ibid, 7.

[38] Ibid, 8.

[39] Ibid, 13-21.

[40] Ibid, 19.

[41] Ibid, 19-20.

[42] Ibid, 20.

[43] Ibid, 20-21.

[44] GDPR, art 4(1).

[45] GDPR, art 4(7).

[46] GDPR, art 4(8).

[47] GDPR, arts 13-18, 20-22.

[48] GDPR, art 12(2).

[49] GDPR, art 5(2).

[50] GDPR, art 83(5)(a).

[51] GDPR, art 82(1).

[52] GDPR, art 5(1)(a).

[53] GDPR, art 5(1)(b).

[54] GDPR, art 5(1)(c).

[55] GDPR, art 5(1)(d).

[56] GDPR, art 5(1)(e).

[57] GDPR, art 5(1)(f).

[58] GDPR, art 6(1)(e).

[59] GDPR, art 6(1)(f).

[60] GDPR, art 6(1)(d).

[61] GDPR, art 9(1).

[62] GDPR, art 4(15).

[63] GDPR, art 4(13).

[64] GDPR, art 4(14).

[65] GDPR, art 9(2)(h).

[66] GDPR, art 9(2)(i).

[67] GDPR, art 9(2)(j).

[68] GDPR, arts 9(2)(b), (g), (h), (i), (j), 9(3).

[69] Article 29 Working Party. *Guidelines on transparency under Regulation 2016/679*. 2018: 4.

[70] Consolidated Version of the Treaty on European Union [2012] OJ C326-0001, art 11(2).

Consolidated Version of the Treaty on Functioning of the European Union [2016] OJ C202/95, art 15.

[71] Article 29 Working Party. *Guidelines on transparency under Regulation 2016/679*. 2018: 5.

[72] Ibid, 4.

[73] Ibid, 5.

[74] Ibid.

[75] Ibid.

[76] Ibid.

[77] Llorens AA. The European Court of Justice: more than a teleological court. *Cambridge Yearbook of European Legal Studies*. 1999; 2: 381-383.

[78] GDPR, art 5(2).

Information Commissioner's Office. *Big data, artificial intelligence, machine learning and data protection*. 2017: 51.

[79] Mazars. Mazars analysis shows that the finance sector has received the most GDPR fines to date. Available from: https://www.mazars.ie/Home/News-and-Insights/Latest-News/Finance-sector-receiving-most-GDPR-fines [Accessed 9th February 2020].

[80] Grentzenberg V, Spittka J. Germany: Berlin Data Protection Authority Imposes Eur 14.5 Million Fine for 'Data Cemetery'. Available from: https://blogs.dlapiper.com/privacymatters/germany-berlin-data-protection-authority-imposes-eur-14-5-million-fine-for-data-cemetery/ [Accessed 9th February 2020].

[81] Article 29 Working Party. *Guidelines on transparency under Regulation 2016/679*. 2018: 5.

Information Commissioner's Office. *Project ExplAIn Interim Report*. 2019: 1-31.

[82] Article 29 Working Party. *Guidelines on transparency under Regulation 2016/679*. 2018: 7.

[83] Ibid, 26.

[84] Case C-201/14 *Bara v Președintele Casei Naționale de Asigurări de Sănătate* [2015] ECR I-638, para 74.

[85] Mitchell C, Ordish J, Hall A. *Data protection and genomic data*. PHG Foundation. [Preprint] 2020.

[86] Ibid.

[87] GDPR, art 89(1).

[88] Carrieri D, Howard HC, Benjamin C, et al. Recontacting patients in clinical genetics services: recommendations of the European Society of Human Genetics. *European Journal of Human Genetics*. 2019; 27: 169-182.

[89] GDPR, art 23(1).

[90] DPA 2018, schedule 3, part 2, section 1.

[91] The Data Protection (Subject Access Modification) (Health) Order 2000, SI 2000/413

Data Protection Act 2018 Explanatory Notes, para 697.

[92] DPA 2018, schedule 3, part 2, section 5.

[93] Ibid, section 3.

[94] Ibid, section 4.

[95] Ibid, section 4(2).

[96] NHS Health Research Authority. *Legal basis for processing data*. Available from: https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/data-protection-and-information-governance/gdpr-detailed-guidance/legal-basis-processing-data/ [Accessed 9 February 2020].

Information Governance Alliance. *The General Data Protection Regulation: Guidance on Lawful Processing*. 2018.

[97] GDPR, art 9.

[98] GDPR, art 89(1).

[99] GDPR, art 89(2)-(3).

[100] GDPR, art 89(2).

[101] GDPR, art 89(3).

[102] GDPR, art 89(2).

[103] GDPR, art 89(1)-(2).

[104] Data Protection Act 2018 Explanatory Notes, para 142.

Data Protection Act 1998, section 33(1)(b).

[105] DPA 1998, section 33(1)(a).

[106] Raza S, Blackburn L, Moorthie S, et al. *The personalised medicine technology landscape*. PHG Foundation. 2018: 96-99.

[107] Data Protection Act 2018 Explanatory Notes, para 142.

[108] DPA 2018, section 19(3).

DPA 1998, section 33.

Section 19(3) DPA 2018

Section 33, DPA 1998

Data Protection (Processing of Sensitive Personal Data) Order 2000, Section 9.

[109] DPA 2018, section 19(4).

[110] GDPR, art 13(1).

[111] GDPR, art 14(3).

[112] GDPR, art 13(3).

GDPR, art 14(4).

Article 29 Working Party. *Guidelines on transparency under Regulation 2016/679*. 2018: 17.

[113] GDPR, art 13(4).

[114] DPA 2018, schedule 3, part 1.

[115] GDPR, art 14(5)(b).

[116] GDPR, art 14(5)(c).

[117] GDPR, art 14(5)(d).

[118] NHS Health Research Authority. *Legal basis for processing data*. Available from: https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/data-protection-and-information-governance/gdpr-detailed-guidance/legal-basis-processing-data/ [Accessed 9 February 2020].

[119] National Health Service Act 2006, section 251.

[120] NHS Digital. *Appendix 1: Section 251 of the National Health Service Act 2006*. Available from: https://digital.nhs.uk/services/data-access-request-service-dars/how-the-national-data-opt-out-affects-data-released-by-nhs-digital/national-data-opt-out-guidance-for-researchers/appendix-1-section-251-of-the-national-health-service-act-2006 [Accessed 10th February 2020).

[121] *W v Egdell* [1990] Ch. 359, [1990] 2 W.L.R. 471.

[122] GDPR, art 13(1), 14(1).

[123] Article 29 Working Party. *Guidelines on transparency under Regulation 2016/679*. 2018: 14.

[124] GDPR, arts 13(1)(a), 14(1)(a).

[125] GDPR, arts 13(2)(d), 14(2)(e).

[126] GDPR, art 14(5)(a).

[127] *W v Egdell* [1990] Ch. 359, [1990] 2 W.L.R. 471.

[128] Article 29 Working Party. *Guidelines on transparency under Regulation 2016/679*. 2018.

[129] Ausloos J, Mahieu R, Veale M. Getting Data Subject Rights Right: A Submission to the European Data Protection Board from Data Protection Academics. 2019. para 10.

[130] GDPR, art 12(3).

[131] GDPR, arts 11(1)-(2), 12(2).

[132] GDPR, art 25.

[133] Ordish J, Cook S, Hall A, Burton H. My healthy future: Discussion notes on privacy and autonomy. 2019.

[134] GDPR, art 4(8).

[135] GDPR, art 4(7).

[136] Case C-210/16 *Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein v Wirtschaftsakademie Schleswig-Holstein GmbH* [2018] ECR II-388.

Case C-25/17 *Tietosuojavaltuutettu* [2018] ECR II-551.

[137] Article 29 Data Protection Working Party. *Opinion 05/2014 on Anonymisation Techniques*. 2014: 20-25.

[138] DPA 2018, schedule 3, part 2, sections 3-4.

[139] DPA 2018, schedule 3, part 2, section 2(1).

[140] DPA 2018, schedule 3, part 2, section 2(1)(b)-(c).

[141] DPA 2018, schedule 3, part 2, section 6.

[142] DPA 2018, schedule 3, part 2, section 2(2).

[143] GDPR, art 89(1).

[144] GDPR, art 89(2)-(3).

[145] Ausloos J, Mahieu R, Veale M. Getting Data Subject Rights Right: A Submission to the European Data Protection Board from Data Protection Academics. 2019. paras 31-35.

Case C-434/16 *Peter Nowak v Data Protection Commissioner* [2017] ECR I-994, para 35.

[146] Ausloos J, Mahieu R, Veale M. Getting Data Subject Rights Right: A Submission to the European Data Protection Board from Data Protection Academics. 2019. para 29.

[147] Ibid.

Article 29 Working Party. *Guidelines on transparency under Regulation 2016/679*. 2018. 19-20.

[148] GDPR, art 20(1).

[149] Li W. *A tale of two rights: exploring the potential conflict between right to data portability and right to be forgotten under the General Data Protection Regulation*. 2018; 8(4): 309-317.

[150] GDPR, art 20(1)(b).

[151] NHS Health Research Authority. *Legal basis for processing data*. Available from: https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/data-protection-and-information-governance/gdpr-detailed-guidance/legal-basis-processing-data/ [Accessed 9 February 2020].

Information Governance Alliance. *The General Data Protection Regulation: Guidance on Lawful Processing*. 2018.

[152] European Data Protection Board. *Guidelines 2/2019 on the processing of personal data under Article 6(1)(b) GDPR in the context of the provision of online services to data subjects*. 2019. 7-10.

[153] DPA 2018, schedule 3, part 2, section 1(g).

[154] DPA 2018, schedule 3, part 2, section 3.

[155] DPA 2018, schedule 3, part 2, section 4.

[156] DPA 2018, schedule 3, part 2, sections 5-6.

[157] GDPR, art 89(3).

DPA 2018, schedule 2, part 6, section 28.

[158] DPA 2018, schedule 2, part 6, section 27.

[159] GDPR, art 20(1).

[160] Wong J, Henderson T. The right to data portability in practice: exploring the implications of the technologically neutral GDPR. *International Data Privacy Law*. 2019; 9(3): 173-191.

[161] Ibid.

[162] GDPR, art 5(1)(d).

DPD, art 6(1)(d).

[163] GDPR, art 23(1).

[164] DPA 2018, schedule 3, part 2, section 3.

[165] DPA 2018, schedule 3, part 2, section 4.

[166] GDPR, art 89(1)-(2).

[167] DPA 2018, section 19(2).

[168] DPA 2018, section 19(3).

[169] DPA 2018, section 19(4).

[170] Ausloos J. *Balancing in the GDPR: legitimate interests v right to object.* Available at: https://www.law.kuleuven.be/citip/blog/balancing-in-the-gdpr-legitimate-interests-v-right-to-object/ [Accessed 9th February].

[171] GDPR, arts 6(1)(e)-(f), 6(3), recital 69.

[172] GDPR, art 21.

[173] NHS Health Research Authority. *Legal basis for processing data*. Available from: https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/data-protection-and-information-governance/gdpr-detailed-guidance/legal-basis-processing-data/ [Accessed 9 February 2020].

[174] GDPR, art 21.

[175] GDPR, art 6(e).

[176] GDPR, art 21.

[177] GDPR, art 21(6).

[178] Case C-13/16 *Valsts policijas Rīgas reģiona pārvaldes Kārtības policijas pārvalde v Rīgas pašvaldības SIA 'Rīgas satiksme'* [2017] ECR I-336. para 28.

[179] Information Commissioner's Office. *What is the 'legitimate interests' basis?* Available from: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/legitimate-interests/what-is-the-legitimate-interests-basis/ [Accessed 9th February 2020].

[180] Article 29 Data Protection Working Party. *Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC*. 2014.

[181] Goodman B, Flaxman S. EU regulations on algorithmic decision-making and a 'right to explanation.' *ICML Workshop on Human Interpretability in Machine Learning*. 2016; 26-30.

Wachter S, Mittelstadt B, Floridi L. Why a Right to Explanation of Automated Decision-Making Does Not Exist in the GDPR. *International Data Privacy Law*. 2017; 7(2): 76-99.

Selbst AD, Powles J. Meaningful Information and the Right to Explanation. *International Data Privacy Law*. 2017; 7(4): 233-253.

[182] Mendoza I, Bygrave LA. The Right not to be Subject to Automated Decisions based on Profiling. In: Synodinous T, Jougleux P, Markou C, et al (eds.) *EU Internet Law: Regulation and Enforcement*. Springer, London; 2017: 3.

Personal Data Protection Code 2003, art 14.

[183] Mendoza I, Bygrave LA. The Right not to be Subject to Automated Decisions based on Profiling. In: Synodinous T, Jougleux P, Markou C, et al (eds.) *EU Internet Law: Regulation and Enforcement*. Springer, London; 2017: 3.

[184] 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' COM (92) 422 final, 26.

[185] European Commission, 'Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)' COM (2012) 11 final, 9.

Council of Europe Committee of Ministers, 'The protection of individuals with regard to automatic processing of personal data in the context of profiling' CM/Rec (2010) 13 and explanatory memorandum.

[186] Council of Europe Committee of Ministers, 'The protection of individuals with regard to automatic processing of personal data in the context of profiling' CM/Rec (2010) 13 and explanatory memorandum: 28-32.

[187] European Commission, 'Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)' COM (2012) 11 final,

Proposal for a Regulation COM 2012/0011 (COD), 54.

[188] Bygrave LA. Article 22 Automated individual decision-making, including profiling. In: Kuner C, Bygrave LA, Docksey C. (eds.) *The EU General Data Protection Regulation (GDPR): A Commentary*. Oxford University Press, Oxford; 2020. 529-532.

[189] 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' COM (92) 422 final, 26.

[190] European Commission, 'Safeguarding Privacy in a Connected World a European Data Protection Framework for the 21st Century' (Communication) COM (2012) 9 final, 5.

[191] GDPR, recital 60.

[192] Goodman B, Flaxman S. EU regulations on algorithmic decision-making and a 'right to explanation.' *ICML Workshop on Human Interpretability in Machine Learning*. 2016; 28.

[193] GDPR, art 22.

Goodman B, Flaxman S. EU regulations on algorithmic decision-making and a 'right to explanation.' *ICML Workshop on Human Interpretability in Machine Learning*. 2016; 28.

Selbst AD, Powles J. Meaningful Information and the Right to Explanation. *International Data Privacy Law*. 2017; 7(4): 244-445.

[194] Wachter S, Mittelstadt B, Floridi L. Why a Right to Explanation of Automated Decision-Making Does Not Exist in the GDPR. *International Data Privacy Law*. 2017; 7(2): 78.

[195] Ibid.

[196] Ibid.

[197] Goodman B, Flaxman S. EU regulations on algorithmic decision-making and a 'right to explanation.' *ICML Workshop on Human Interpretability in Machine Learning*. 2016; 28.

[198] Goodman B, Flaxman S. EU regulations on algorithmic decision-making and a 'right to explanation.' *AI Magazine*. 2017; 50.

[199] Selbst AD, Powles J. Meaningful Information and the Right to Explanation. *International Data Privacy Law*. 2017; 7(4): 244-445.

[200] Mendoza I, Bygrave LA. The Right not to be Subject to Automated Decisions based on Profiling. In: Synodinous T, Jougleux P, Markou C, et al (eds.) *EU Internet Law: Regulation and Enforcement*. Springer, London; 2017.

[201] Wachter S, Mittelstadt B, Floridi L. Why a Right to Explanation of Automated Decision-Making Does Not Exist in the GDPR. *International Data Privacy Law*. 2017; 7(2): 78.

[202] Ibid.

[203] Article 29 Data Protection Working Party. Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679. 2016, 13.

[204] Ibid.

[205] Mendoza I, Bygrave LA. The Right not to be Subject to Automated Decisions based on Profiling. In: Synodinous T, Jougleux P, Markou C, et al (eds.) *EU Internet Law: Regulation and Enforcement*. Springer, London; 2017, 3.

[206] Committee on Civil Liberties, Justice and Home Affairs, 'On the proposal for a regulation of the European Parliament and of Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)' A7-0402/2013 (2013), amendment 115.

Mendoza I, Bygrave LA. The Right not to be Subject to Automated Decisions based on Profiling. In: Synodinous T, Jougleux P, Markou C, et al (eds.) *EU Internet Law: Regulation and Enforcement*. Springer, London; 2017, 11.

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

[208] Ibid, 10.

[209] Ibid.

[210] 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' COM (92) 422 final, 26.

[211] Mendoza I, Bygrave LA. The Right not to be Subject to Automated Decisions based on Profiling. In: Synodinous T, Jougleux P, Markou C, et al (eds.) *EU Internet Law: Regulation and Enforcement*. Springer, London; 2017, 11.

[212] Dworkin R. *Law's Empire*. Oxford: Hart Publishing; 1998, 242-244.

[213] Article 29 Data Protection Working Party. Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679. 2016, 6.

[214] Ibid, 8.

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

[216] Ibid.

[217] Mendoza I, Bygrave LA. The Right not to be Subject to Automated Decisions based on Profiling. In: Synodinous T, Jougleux P, Markou C, et al (eds.) *EU Internet Law: Regulation and Enforcement*. Springer, London; 2017, 12.

[218] *R v Birmingham City Council, ex parte Mohammed* [1999] 1 W.L.R. 33.

[219] Lewis C. *Judicial Remedies in Public Law 5th Edition*. London: Sweet & Maxwell; 2014: 4-005-4-007.

[220] DPD, art 15.

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

[222] Mendoza I, Bygrave LA. The Right not to be Subject to Automated Decisions based on Profiling. In: Synodinous T, Jougleux P, Markou C, et al (eds.) *EU Internet Law: Regulation and Enforcement*. Springer, London; 2017, 9.

[223] Ibid.

[224] Church S, Millard C. Commentary on Article 15 of the Data Protection Directive. In: Bullesbach A, Bijbrath S, et al. (eds.) *Concise European IT Law*. Alphen aan den Rijn: Kluwer Law International; 2010. 84.

[225] Bygrave LA. Automated Profiling: Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling. *Computer Law & Security Review*. 2001; 17(1): 19.

[226] Bygrave LA. Automated Profiling: Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling. *Computer Law & Security Review*. 2001; 17(1): 19.

[227] Feinberg J. *The Moral Limits of the Criminal Law Volume 1: Harm to Others*. Oxford: Oxford University Press; 1987, 31-64.

[228] Nussbaum M. *Creating Capabilities: The Human Development Approach*. Harvard: Harvard University Press; 2011, 32-33.

[229] Ibid, 33-34.

[230] Article 29 Data Protection Working Party. Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679. 2016, 14.

[231] Kamarinou D, Millard C, Singh J. Machine Learning with Personal Data. *Queen Mary University of London, School of Law Legal Studies Research Paper*. 2016; 247: 20.

[232] GDPR, recital 60.

[233] Information Commissioner's Office, The Alan Turing Institute. *Explaining decision made with AI: Draft guidance for consultation: Part 2: Explaining AI in practice*. 2019.

[234] Kamarinou D, Millard C, Singh J. Machine Learning with Personal Data. *Queen Mary University of London, School of Law Legal Studies Research Paper*. 2016; 247: 20.

[235] Article 29 Data Protection Working Party. Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679. 2016, 14.

[236] Article 29 Working Party. *Guidelines on transparency under Regulation 2016/679*. 2018: 8-9.

[237] Article 29 Data Protection Working Party. Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679. 2016, 14.

[238] Information Commissioner's Office, The Alan Turing Institute. *Explaining decision made with AI: Draft guidance for consultation: Part 2: Explaining AI in practice*. 2019, 68.

[239] Article 29 Data Protection Working Party. Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679. 2016, 14-15.

[240] Council of Europe, 'Explanatory Report to the Protocol amending the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data' Council of Europe Treaty Series (2018) 223, para 77.

[241] Wachter S, Mittelstadt B, Floridi L. Why a Right to Explanation of Automated Decision-Making Does Not Exist in the GDPR. *International Data Privacy Law*. 2017; 7(2): 78.

[242] Ibid.

[243] Ibid, 78.

[244] Guidotti R, Monreale A, Ruggieri S, et al. A survey of methods for explaining black box models. *ACM Computing Surveys*. 2018; 51(5): 1-42.

[245] Mendoza I, Bygrave LA. The Right not to be Subject to Automated Decisions based on Profiling. In: Synodinous T, Jougleux P, Markou C, et al (eds.) *EU Internet Law: Regulation and Enforcement*. Springer, London; 2017, 15-17.

[246] NHS Health Research Authority. *Legal basis for processing data*. Available from: https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/data-protection-and-information-governance/gdpr-detailed-guidance/legal-basis-processing-data/ [Accessed 9 February 2020].

Information Governance Alliance. *The General Data Protection Regulation: Guidance on Lawful Processing*. 2018.

[247] European Data Protection Board. *Guidelines 2/2019 on the processing of personal data under Article 6(1)(b) GDPR in the context of the provision of online service to data subjects*. 2019, 7-10.

[248] DPA 2018, section 14(4)(a)-(b).

[249] NHS Health Research Authority. *Legal basis for processing data*. Available from: https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/data-protection-and-information-governance/gdpr-detailed-guidance/legal-basis-processing-data/ [Accessed 9 February 2020].

Information Governance Alliance. *The General Data Protection Regulation: Guidance on Lawful Processing*. 2018.

[250] GDPR, recital 60.

The Black box medicine and transparency report was funded by the Wellcome Trust as part of the 2018 Seed Awards in Humanities and Social Sciences [Grant Number: 213623/Z/18/Z].

We thank the Wellcome Trust for their support.

The PHG Foundation is a non-profit think tank with a special focus on how genomics and other emerging health technologies can provide more effective, personalised healthcare and deliver improvements in health for patients and citizens.

For more information contact:

intelligence@phgfoundation.org

UNIVERSITY OF CAMBRIDGE

phg foundation

making science work for health