Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/CLSR

Computer Law & Security Review

Transparency of machine-learning in healthcare: The GDPR & European health law

Miranda Mourby[a,*], *Katharina Ó Cathaoir*[b], *Catherine Bjerre Collin*[c]

[a] *Centre for Health, Law and Emerging Technologies ('HeLEX'), Faculty of Law, University of Oxford, Ewert House, Oxford OX2 7DD, UK*

[b] *Faculty of Law, University of Copenhagen, Karen Blixens Plads 16, 2300 Copenhagen, Denmark*

[c] *Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Blegdamsvej 3B, 2200 Copenhagen, Denmark*

ARTICLE INFO

Keywords:
Clinical decision support software
Explanations
Healthcare
Machine-Learning
Patients' rights
Transparency

ABSTRACT

Machine-learning ('ML') models are powerful tools which can support personalised clinical judgments, as well as patients' choices about their healthcare. Concern has been raised, however, as to their 'black box' nature, in which calculations are so complex they are difficult to understand and independently verify. In considering the use of ML in healthcare, we divide the question of transparency into three different scenarios:

1) Solely automated decisions. We suggest these will be unusual in healthcare, as Article 22(4) of the General Data Protection Regulation presents a high bar. However, if solely automatic decisions are made (e.g. for inpatient triage), data subjects will have a right to 'meaningful information' about the logic involved.
2) Clinical decisions. These are decisions made ultimately by clinicians—such as diagnosis—and the standard of transparency under the GDPR is lower due to this human mediation.
3) Patient decisions. Decisions about treatment are ultimately taken by the patient or their representative, albeit in dialogue with clinicians. Here, the patient will require a personalised level of medical information, depending on the severity of the risk, and how much they wish to know.

In the final category of decisions made by patients, we suggest European healthcare law sets a more personalised standard of information requirement than the GDPR. Clinical information must be tailored to the individual patient according to their needs and priorities; there is no monolithic 'explanation' of risk under healthcare law. When giving advice based (even partly) on a ML model, clinicians must have a sufficient grasp of the medically-relevant factors involved in the model output to offer patients this personalised level of medical information. We use the UK, Ireland, Denmark, Norway and Sweden as examples of European health law jurisdictions which require this personalised transparency to support patients' rights to make informed choices. This adds to the argument for post-hoc, rationale explanations of ML to support healthcare decisions in all three scenarios.

© 2021 Miranda Mourby, Katharina Ó Cathaoir, Catherine Bjerre Collin. Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

* Corresponding author.
E-mail address: miranda.mourby@law.ox.ac.uk (M. Mourby).

https://doi.org/10.1016/j.clsr.2021.105611

0267-3649/© 2021 Miranda Mourby, Katharina Ó Cathaoir, Catherine Bjerre Collin. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

1. Introduction

Machine-learning ('ML') is a type of artificial intelligence ('AI') in which algorithms are trained to infer patterns from a large amount of data.[1] It has recently emerged as a dominant form of AI.[2] ML is associated with a lower level of 'interpretability,'[3] meaning it is difficult for a human being to understand its workings without a second model (i.e. another piece of software) providing an explanation of what features in the data influenced its conclusions. This is referred to as a 'post-hoc' explanation.

There has been some controversy as to whether such post-hoc explanations are sufficient in high-stakes contexts,[4] or if ML can be built in a more 'interpretable' way for use in healthcare.[5] Nevertheless, in silico modelling, such as, ML is rapidly changing healthcare. From imaging software to mortality prediction,[6] ML models can process larger volumes of heterogeneous information than the conscious human mind, and can in some cases yield more accurate classifications and predictions than skilled clinicians working alone. It has been suggested that deep-learning models in particular, with their capacity to process multiple parameters simultaneously, offer the opportunity to 'personalise' medicine to the individual patient, taking into account their social and biological profile.[7]

Academics and policy-makers alike have expressed optimism that ML can be used to support clinical decision-making to the broader benefit of healthcare systems[8][9]—these types of models fall within the category of software described as 'Clinical Decision Support Software.'[10] At the same time, reservations have been expressed about the 'black box' opacity of such models,[11] as well as their potential for bias.[12] The locus of much of these discussions in a European context has been the General Data Protection Regulation ('GDPR')[13] and the extent to which it does, or does not, adequately regulate algorithmic data processing. While this paper touches on this debate in considering the use of ML in healthcare, we will also expand the frame of reference by including healthcare law.

The alleged 'right to an explanation' of automated decisions under the GDPR has been a hotly debated issue,[14][15] with some suggesting the focus on whether the GDPR's information requirements amount to an 'explanation' distracts from the substance of the legislation.[16][17][18] We therefore use the terms 'information' and 'transparency' as these reflect the language of the GDPR, and of healthcare law. Where the word 'explanation' is used, it refers to a post-hoc method of using an additional model to explain ML to end-users; we do not suggest these explanations are necessarily the type of information which should be given to patients, as the number of clinically relevant features involved may become too complex to constitute meaningful information for any individual who is not a computer scientist or a doctor. Instead, we focus on rights to information under the GDPR and healthcare law. The 'meaningful information' patients require may not be the kinds of explanations devised by computer scientists,[19] but rather a medical framing of their condition as constructed (collaboratively with their input) by a clinician.

[1] European Commission, *White Paper on Artificial Intelligence - A European approach to excellence and trust* (Brussels, 19 February 2020), COM(2020) 65 final, available from: <https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf>

[2] Information Commissioner's Office & The Alan Turing Institute, 'Explaining decisions made with artificial intelligence', Part I 'What is AI?' available from <https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/explaining-decisions-made-with-artificial-intelligence/part-1-the-basics-of-explaining-ai/definitions/> accessed 24 March 2021. Hereafter referenced as 'ICO and ATI.'

[3] Ibid, at 5

[4] Cynthia Rudin, 'Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead' (2019) 1 Nature Machine Intelligence 206

[5] Richard Li,; Ashwin Shinde, An Liu et al, 'Machine Learning–Based Interpretation and Visualization of Nonlinear Interactions in Prostate Cancer Survival' (2020) 4 JCO Clinical Cancer Informatics 637

[6] Annelaura B Nielsen et al, 'Survival prediction in intensive-care units based on aggregation of long-term disease history and acute physiology: a retrospective study of the Danish National Patient Registry and electronic patient records' (2020) 1 Lancet Digital Health 2 https://doi.org/10.1016/S2589-7500(19)30024-X

[7] Georgios Z. Papadakis et al 'Deep learning opens new horizons in personalized medicine' (2019) 10 Biomedical Reports 4

[8] I. Glenn Cohen et al, 'The legal and ethical concerns that arise from using complex predictive analytics in health care' (2014) 33 Health Aff. 1139

[9] European Commission, White Paper (note 1)

[10] Tamra Lysaght et al, 'AI-Assisted Decision-making in Healthcare: The Application of an Ethics Framework for Big Data in Health and Research' (2019) 11 Asian Bioethics Review 299

[11] Agata Ferretti et al, 'Machine Learning in Medicine: Opening the New Data Protection Black Box' (2018) 3 European Data Protection Law 320, <https://edpl.lexxion.eu/data/article/13107/pdf/edpl_2018_03-011.pdf>

[12] Maja Brkan, 'Do algorithms rule the world? Algorithmic decision-making and data protection in the framework of the GDPR and beyond' (2019) 27 International Journal of Law and Information 2

[13] Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/ (General Data Protection Regulation) [2016] OJ L119/1, which will be cited as 'the GDPR'

[14] Bryce Goodman and Seth Flaxman, 'European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation"' (2016) ICML Workshop on Human Interpretability in Machine Learning, arXiv:1606.08813 (v3); (2017) 38 AI Magazine 50.

[15] Sandra Wachter, Brent Mittelstadt, and Luciano Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 76.

[16] Andrew D Selbst & Julia Powles 'Meaningful information and the right to explanation' (2017) 7 *International Data Privacy Law* 4

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

[18] Margot E Kaminski, 'The Right to Explanation, Explained' (2019) 34 Berkley Technology Law Journal 189. See also Brkan, note 12.

[19] Edwards and Veale, note 17

Explanations of ML are thus a means of achieving transparency, not transparency itself,[20] and often require human mediation to become contextually useful information. The delineation between these terms, as used in this paper, can be summarised as follows:

Information	Details an individual should receive by law. These may be the details which a data subject is entitled to know under data protection law, or which a patient should receive under healthcare law. This is sometimes referred to as 'meaningful information' or 'accessible information' in the GDPR, and 'material information' in UK and Irish healthcare law.
Explanation	An account of how a model reaches (or 'reached', in the case of post-hoc explanation) its output. For a model used to support decision-making in healthcare, we suggest such explanations should be targeted towards the knowledge and priorities of a healthcare professional, who will need to translate the model's recommendations into 'information' for patients.
Transparency	The ultimate goal of information and explanations alike; a principle supporting individuals in awareness of—and active involvement in– any interference with their fundamental rights.[21]

There exists a broad range of explanation types, and tools by which they can be generated in a post-hoc fashion (i.e. *after* a specific output has been reached by the model). The UK's Information Commissioner's Office ('ICO') has, along with the Alan Turing Institute, identified six types of explanation,[22] as well as provided a very comprehensive schedule of supplementary explanation models which can be used to explain AI (including ML).[23] This schedule goes into the respective benefits and limitations of these methods of explaining artificial intelligence—a full review of which is beyond the scope of this paper. However, we broadly concur with the ICO's guidance that some additional explanatory tool should be used to shed light on how ML reached its output, even if this explanation cannot convey the full inner workings of a more complex model. We agree with the ICO that, in a medical context, an accurate and reliable 'rationale' explanation will support the evidence-based judgement of the professionals involved and is a priority for patients to understand (at least to an extent) *why* an ML model arrived at a particular recommendation relevant to their healthcare.[24]

We therefore advocate for rationale explanations of ML in healthcare, without championing any particular explanatory tool. Selection of the most appropriate supplementary model is a highly technical and contextual decision, which is (again) outside the scope of this paper. We instead consider the legal standards of transparency which these explanations must ultimately satisfy. For example, local explanation (in which the main factors involved in a model's output are identified) and counterfactual explanation (in which the respective influence of each of these factors is evaluated) have been identified as both technically feasible forms of explanation, and as having a high degree of correspondence with the GDPR's requirements.[25] Some form of post-hoc explanation—which is possible even for complex neural network models that process large volumes of pixel-data within medical images[26] —is necessary not only for GDPR compliant automated decisions and profiling,[27] but also under healthcare law, where patient decisions are based at least partly on the recommendations of an ML model.

Patient decisions represent a particularly important case-study. Such decisions, in which patients make choices impacting their bodily integrity, engage fundamental rights which supplement GDPR information rights. Outside the specific context of health, Bart van der Sloot has argued generally that the case law of the European Court of Human Rights ('ECtHR') imposes a higher standard of regulation than the GDPR where profiling is concerned, as it promotes 'decisional privacy' (essentially, autonomy of choice).[28] Similarly, Hildebrandt has framed the ability to reflect on the profiles that are applied to us as central to our freedom of action, and sees the GDPR as potentially revolutionary in this regard.[29] Mantelero has suggested that data protection assessment should be augmented

[20] Anastasiya Kiseleva, 'AI as a Medical Device: Is It Enough to Ensure Performance Transparency and Accountability?' (2020) 4 European Pharmaceutical Law Review 1

[21] See Kaminski (note 18): the nature of transparency required ultimately depends on the nature of the right it is designed to support.

[22] ICO & ATI (see note 2), 'What goes into an explanation?' available from < https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/explaining-decisions-made-with-artificial-intelligence/part-1-the-basics-of-explaining-ai/what-goes-into-an-explanation/#explanation_3> accessed 24 March 2021

[23] Ibid, 'Annexe 3: Supplementary models', available from: <https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/explaining-decisions-made-with-artificial-intelligence/annexe-3-supplementary-models/> accessed 24 March 2021

[24] Ibid, 'Part 2 Explaining AI in practice'.

[25] Maja Brkan and Grégory Bonnet, 'Legal and Technical Feasibility of the GDPR's Quest for Explanation of Algorithmic Decisions: of Black Boxes, White Boxes and Fata Morganas' (2020) 11 European Journal of Risk Regulation 1

[26] For example, the neural networks developed by the Moorfield Eye Hospital in the UK can provide eye care professionals with information that explains their output in a clinically meaningful way (e.g. visuals of features of eye disease identified in the image, and level of confidence in the conclusion). See <https://www.moorfields.nhs.uk/content/breakthrough-ai-technology-improve-care-patients> accessed 3 March 2021.

[27] As also argued by Amitojdeep Singh, Sourya Sengupta, and Vasudevan Lakshminarayanan, 'Explainable deep learning models in medical image analysis' (2020) 6 Journal of Imaging 6

[28] Bart van der Sloot, 'Decisional privacy 2.0: the procedural requirements implicit in Article 8 ECHR and its potential impact on profiling' (2017) 7 International Data Privacy Law 3

[29] Mireille Hildebrandt, 'The Dawn of a Critical Transparency Right for the Profiling Era' in J. Bus et al (eds) 'Digital Enlightenment Yearbook 2012', available from: < doi:10.3233/978-1-61499-057-4-4141> accessed 4 March 2021

by human rights considerations.[30] We agree with these authors: engagement of human rights, and particularly Article 8 ECHR,[31] brings deeper considerations of personal autonomy and dignity, and not just the general 'fairness' of data processing under data protection law.[32] We focus specifically on decisions in healthcare which are informed by ML predictions. As we will show, healthcare law bears some relationship with ECtHR jurisprudence, but is drawn from heterogeneous sources of law across Europe; these in turn have a complex intersection with the GDPR when in silico predictive modelling is used in healthcare.

In terms of the additional contribution of healthcare law, we focus on common law jurisdictions such as the UK and Ireland, and civil law such as Denmark, Sweden and Norway. Although there are variations between these national levels of law, we suggest that all these jurisdictions require information to be tailored to a patient's preferences and circumstances, and that this amounts to an overarching requirement of 'personalised transparency' under European healthcare law.

As a final caveat: the recently proposed EU Regulation on AI[33] is not explored in this paper, as at the time of writing it was very newly published and required further consideration. It is worth noting, however, that its proposed text would designate all AI which is (or forms part of) a medical device a 'high risk' AI system.[34] As such, it must satisfy additional transparency requirements aimed at the 'user' (in this instance, the clinician) and be designed to accommodate human interface and oversight.[35] Although further research into this new Regulation is needed, it appears to provide a compelling additional reason for clinical users of ML to be permitted a reasonable level of insight into its outputs.

Despite the variation of transparency standards in these heterogenous sources of law, the outcome is the same: decision-makers need to know why, in medical terms, a particular outcome is recommended by a model, so they can make a sufficiently informed decision for the purposes of data protection and healthcare law. We therefore argue that post-hoc, 'rationale' model explanations (if not complete interpretability) are needed to support healthcare decision-making, even if such explanations do not provide a full picture of an ML's inner workings. Transparency is not the same thing as certainty, and too much information about ML is likely to obfuscate rather than clarify, but patients still need the option of some rationale as to why a particular treatment is recommended for them to provide informed consent.[36]

We will begin with consideration of healthcare decisions made in sole reliance on ML, and the relevant GDPR information requirements.

2. Solely automated decisions: 'meaningful information'

It would be, we suggest, unusual for a decision to be made in the course of healthcare *solely* on the basis of an ML output, with no human mediation. This is, in part, because of the limited exceptions the GDPR allows for the processing of health-related data under Article 22(2). However, where such decisions *are* made without human input, data subjects would be entitled to meaningful information about the logic involved in the automated decision-making, so the ML could not be entirely opaque.

As an illustration, we will briefly consider an example of healthcare automation proposed in a US context. I. Glenn Cohen and colleagues expressed optimism in a 2014 paper that predictive analytics could help make healthcare systems stronger and more dynamic. In particular, they hypothesise that a predictive model could be used to advise doctors (or, later in the article, hospital administrators) as to which patients should be admitted to the Intensive Care Unit, suggesting such a model could take into account:

'*the risk of all patients in a hospital, their individual therapeutic goals and preferences, hospital staffing (including staff members' experience and performance), resource constraints, and external conditions such as whether other hospitals are diverting patients in the emergency department in the case of a disaster.*'[37]

The authors emphasise that such automated triage would have to be subject to clinical (or, potentially, administrative) scrutiny. However, if there were (again, hypothetically) instances where the automated triage was not overseen or overruled by a hospital staff member, this would present difficulties under the GDPR.

2.1. GDPR & automated decision-making

Article 22(1) contains the GDPR's most direct and specific regulation of automated 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.*'

This means that, by default, no 'significant' decisions should be made through automated processing, including by machine-learning models and other forms of algorithmic processing.

To apply the GDPR to this ICU admission example, therefore, it seems uncontroversial that the decision whether to admit a patient to the ICU is one which would significantly affect them. The Article 29 Working Party guidance on automated processing, for example, lists decisions that affect someone's access to health services as producing significant effects,[38]

[30] Alessandro Mantelero, 'AI and Big Data: A blueprint for a human rights, social and ethical impact assessment' (2018) 34 Computer Law & Security Review 754

[31] European Convention on Human Rights, as arbitrated by the ECtHR

[32] Edward S Dove, 'The EU General Data Protection Regulation: Implications for International Scientific Research in the Digital Era' (2019) 46 Journal of Law, Medicine & Ethics 4

[33] European Commission, 'Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts Brussels, 21.4.2021 COM (2021) 206 final

[34] Ibid, Article 6(1)(a) & Annex II

[35] Ibid, Articles 13 & 14

[36] Kiseleva, note 20

[37] Cohen et al, note 8

[38] Article 29 Working Party, 'Guidelines on Automated individual decision-making and Profilingfor the purposes of Regulation

and this (we suggest) should include assessments which could determine the level of care a patient receives within a service. In extremis, the decision could even engage a patient's right to life.

Where admission decisions are made 'solely' on the basis of automated processing, Article 22 GDPR will be engaged. This Article establishes a general right for individuals not to be subject to significant decisions based solely on automated processing. Article 22(2) specifies some exceptions to this right—where authorised by law, necessary for contract or the explicit consent of the data subject has been obtained. However, this is subject to a further qualification in 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.

This is important in a medical context, as data relating to health (and genetic data) are special categories of personal data, and are covered by Article 9. They should therefore not be used for solely automated decisions unless:

a) The data subject's explicit consent has been obtained (Article 9(a)), or
b) The processing is necessary for reasons of substantial public interest, on the basis of EU or Member State law which is proportionate to the aim pursued (Article 9(g)).

2.2. *EU or member state law*

Article 9(g) appears to require a specific piece of legislation to legitimate the automated decision-making. It could be inferred that this is not likely to include general legislation for the provision of healthcare. This is because Article 9(h) is the condition on which special category data can be processed for medical treatment or the management of healthcare services, and it is not among the conditions included in Article 22(4). Neither is the scientific research condition—Article 9(j). This strongly suggests that healthcare and related research alone are not sufficient for Article 9(g), and it would require specialist legislation, well-evaluated for proportionality, to legitimate automated decision-making under this condition.

Some EU jurisdictions have made provision for automated decision-making in their national law.[39] Even, then, however, it is not always clear-cut. The UK, for example, has provided a basis in s.14 Data Protection Act 2018, but the Explanatory Notes to the Act for observe that:

'*Article 22(2)(b) of the GDPR does not require the law to expressly provide that a decision can be made based solely on automated processing before that decision can be taken on the basis of automated processing. It is enough that automated processing is a reasonable way of complying with a requirement, such as a regulatory obligation or licence condition. Such obligation may be provided in general terms, such as a requirement to maintain fraud and financial crime detection systems.*'[40]

This suggests some flexibility in the way Article 22(b) is relied upon in the UK, but not all jurisdictions will necessarily take this broad approach with reference to a vague principle of 'reasonable' compliance. A diversity of approaches to automated decision-making have been taken across the EU.[41] For example, Denmark has not made provision for automated decision making in its national Data Protection Act, but the Data Protection Authority and preparatory works suggest that automated decision making (including in healthcare) could be permitted, provided there is a legal basis and the individual affected is afforded adequate guarantees, for example, can appeal the decision to a body that does not make decisions based on automated decision-making.[42] It could, however, be difficult to prove the adequacy of this appeal as a safeguard if decisions are being made in urgent situations such as ICU admission.

2.3. *Explicit consent*

Explicit consent may initially appear a more promising option. Ferretti and colleagues see 'informed consent' as the obvious way of legitimating automatic decisions, although they suggest that innovative consent models may be needed to impart information about machine learning.[43] The difficulty with this, however, is that guidance subsequently issued has emphasised the distinction between informed consent and consent under the GDPR, and has suggested that consent may not be an appropriate condition for processing in a clinical context.

The Article 29 Working Party,[44] the European Commission,[45] and the European Data Protection Supervisor,[46] have all emphasised the difference between informed consent and consent as a basis for processing under the GDPR. While the former is still essential for lawful medical treatment, the latter might be difficult, as consent to processing cannot be 'freely'

2016/679' (WP251 rev.01, as revised and adopted on 6 February 2018)

[39] Brkan, Note 12

[40] Explanatory Notes to the Data Protection Act 2018, para 115, available from: <http://www.legislation.gov.uk/ukpga/2018/12/pdfs/ukpgaen_20180012_en.pdf>

[41] Gianclaudio Malgieri, 'Automated decision-making in the EU Member States: The right to explanation and other "suitable safeguards" in the national legislations' (2019) 35 Computer Law & Security Review 5

[42] Datatilsynet, Vejledning om de registreredes rettigheder (July 2019) 49; Justitsministeriets, Betænkning om Databeskyttelsesforordningen (2016/679) – og de retlige rammer for dansk lovgivning Del I, bind 1 nr. 1565, 379-383. See also, Det Etiske Råd, Redegørelse om sundhedswearables og big data (2019), p. 95-96.

[43] Ferretti et al, note 11

[44] Article 29 Working Party, 'Guidelines on Consent under Regulation 2016/ 679' WP 259 rev.1, as last revised 10 April 2018

[45] European Commission Directorate-General for Health and Food Safety, 'Question and Answers on the interplaybetween the Clinical Trials Regulation and the General Data Protection Regulation', page 7, available at:<https://ec.europa.eu/health/sites/health/files/files/documents/qa_clinicaltrials_gdpr_en.pdf>,

[46] European Data Protection Supervisor, 'A Preliminary Opinion on data protection and scientific research' (Brussels, 6 January 2020), available at: <https://edps.europa.eu/sites/edp/files/publication/20-01-06_opinion_research_en.pdf>

given if withholding it might cause detriment. For example, a doctor cannot treat a patient without noting this in their records, and so refusing consent for the record could result in the denial of treatment.

Even if it were possible, therefore, in the Cohen et al. example for a patient to decline to have their data processed within the hospital's triage modelling, it seems highly unlikely they could do so without the risk of exclusion from (or at least a delay in entering) the ICU even when their in-patient data might warrant admission.

In the (we suggest) unusual circumstances in which explicit consent to healthcare data processing is deemed an appropriate GDPR condition, and the fear of detriment can be overcome, this will nonetheless require a high level of transparency about the nature and consequences of the processing to ensure such consent is informed and unambiguous.[47] Although the scope of the 'meaningful information' required under Articles 13–15 have been debated,[48][49] it is worth considering the higher level of information about the processing that would have to be disclosed as part of an adequate solicitation of consent under Articles 6,7 and 9 GDPR.

In summary, while it is not impossible that consent to the GDPR standard could be obtained in a healthcare context, in circumstances where e.g. admission to the ICU may be warranted, the potential detriment of opting out of processing appears too significant to justify any claim that the consent was freely given. If it were possible to overcome the detriment objection, a high level of transparency would be required to support an unambiguous consent, in addition to the overarching transparency requirements of Articles 5, 13–15 and 22.

2.4. *Meaningful information*

In the exceptional cases where healthcare decisions are made automatically,[50] the data subject is entitled to 'meaningful information about the logic involved,'[51] provided in 'accessible' form.[52] Sufficiency and extent of 'meaningful information', and whether it amounts to an 'explanation' has been discussed at length elsewhere.[53] For the purposes of this paper, however, it is enough to note that post-hoc explanations such as counterfactuals or deconvolution have been advocated as adequate sources of meaningful information about a model's logic for the purposes of Article 22 GDPR.[54][55][56][57] We therefore suggest that models used for any solely automated decisions in healthcare should be supplemented with rationale, post-hoc explanation; at least to the standard of highlighting the factors and considerations which were taken into account in the decision.[58] Even for models used to process large volumes of image data to identify urgent cases, such explanations have been shown to be possible.[59]

In the case of automated triage, which would conventionally be decided upon manually but without necessarily informing patients of the criteria for prioritisation, this raises the intriguing question as to whether such 'meaningful information' about triage logic would make the process more transparent. The counterpoint would be whether a patient ill enough to warrant admission to the ICU would really be in an adequate position to exercise their data protection rights and challenge the automated decision by querying the logic of their triage. This casts doubt on the adequacy of transparency as a safeguard in such instances. This is, however, mostly speculative and beyond the scope of this paper. Furthermore, Recital 71 GDPR (which has recently been recognised by a Dutch District Court as an important means of interpreting Article 22[60]) also states that measures based on profiling should not concern a child. This is another reason why solely automated decisions may not be routinely appropriate in healthcare, as it could be difficult to ensure that no children were affected by such measures.

We therefore conclude this section by suggesting that purely automated decisions in healthcare should be the exception to the rule under the GDPR.

3. Clinical decisions: GDPR & profiling

For the reasons explored above, we suggest the majority of decisions taken in healthcare should be mediated by human intervention. Where such decisions are made by clinical staff (medical or otherwise) we refer to them as 'clinical decisions.' For clarity, we are referring only to decisions made in the context of healthcare delivery (such as ICU admission, or medical diagnosis). Decisions made in the course of medical or scientific research are out of scope where they do not impact directly upon patients.

Article 22 GDPR only governs decisions made 'solely' on the basis of automated processing. This means that human-mediated decisions are still subject to transparency requirements, but not the default prohibition of Article 22. This means it is easier for healthcare decisions to be made with partial reliance on ML.

Although Wachter et al. have suggested that the GDPR requires very little explanation of automated processing when decisions are (even minimally) human-mediated,[61] a num-

[47] Article 29 Working Party, note 38 p.13
[48] Selbst and Powles, note 16
[49] Brkan, note 12
[50] Rather than, for example, the clinical criteria for inclusion in a screening programme, or for ICU admission being set at a policy/hospital level, and merely *implemented* using an automated process.
[51] GDPR, Articles 13-15
[52] GDPR, Article 12
[53] Wachter et al note 15; Selbst & Powles note 16
[54] Sandra Wachter, Brent Mittelstadt, & Chris Russell, 'Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR' (2018) 31 Harvard Journal of Law & Technology 2
[55] Ferretti et al, note 11
[56] Brkan and Bonnet, note 25
[57] Gianclaudio Malgieri and Giovanni Comandé, 'Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 4
[58] Brkan, note 12
[59] See notes 26-22
[60] C / 13/689705 / HA RK 20-258, *Ola drivers v. Ola Cabs (transparency requests)*, unofficial English translation available from < https://ekker.legal/2021/03/13/dutch-Court-rules-on-data-transparency-for-uber-and-ola-drivers/>
[61] Note 15

ber of authors have subsequently favoured a more robust interpretation of the GDPR transparency obligations, which is now supported by guidance of the Article 29 Working Party ('A29WP') on automated decision-making and profiling.[62] The A29WP's guidance is authoritative, coming as it does from representatives of the data protection authorities across the EU, and is thus likely to influence how courts and supervisory authorities will interpret the GDPR transparency obligations.

The A29WP soften the distinction between profiling and automated decision-making, and make clear that both forms of data processing are subject to the GDPR's overarching transparency obligations under Article 5. Profiling is defined in the GDPR as:

'*any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation,* ***health****, personal preferences, interests, reliability, behaviour, location or movements*' (emphasis added).[63]

The A29WP make it clear that human intervention does not remove automated processing from the scope of profiling, as long as ML is used to evaluate or predict aspects concerning a person's health:

'*In particular, where the processing involves profiling-based decision making* ***(irrespective of whether it is caught by Article 22 provisions)****, then the fact that the processing is for the purposes of both (a) profiling and (b) making a decision based on the profile generated, must be made clear to the data subject.*' (emphasis added)[64]

Even when doctors make the final clinical decision, therefore, predictive modelling in particular is likely to fall into the GDPR definition of profiling and attract its enhanced transparency requirements. This is highlighted by Recital 60 GDPR, which sets out:

'*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. Furthermore, the data subject should be informed of the existence of profiling and the consequences of such profiling.*'

Recital 60 informs the Article 5 Transparency principle where profiling is concerned. It does fall short of the 'meaningful information' about the logic involved which patients would be guaranteed under Article 22 where solely automated decisions are made. However, full explanation of 'consequences' may mean that data subjects (patients) should be told what factors will affect their risk scores, which may in turn affect the decisions which are made about their healthcare.

This requires data controllers using ML for health-related profiling to have a sufficient understanding of the clinical factors at work in a model's predictions, so they can at least frame its outputs in the context of a medical rationale, and in that way provide meaningful insight into the 'consequences' of profiling. Admittedly, this is not as high a standard as the requirement to provide 'meaningful information' about the logic of a model, but data subjects may need to know of any characteristics picked up in their health data which could impact on their healthcare, as this would form part of the 'consequences' of the profiling. This aligns well with the information doctors often wish to provide to their patients, allowing them to better understand their situation and change those risk factors which can be changed (e.g. smoking; alcohol consumption etc.).

In short, 'clinical' decisions made with the support of ML require only a more basic level of GDPR transparency, but Recital 60 does make it clear that patients should be aware of the 'consequences' of the profiling. This means at least being aware of its existence, and how it will impact upon their healthcare (which, arguably, would mean how they will be individually impacted, and what aspects of their medical records will affect their profile). This implies some knowledge of feature-relevance[65] (i.e. an exposition of which medical factor(s) were important to the model's output) would be necessary for this level of 'consequence' orientated transparency.

For patients' decisions, however, a higher level of transparency is needed to support their right to make informed choices. While this does not necessarily mean the ML model itself needs to be more 'explainable' or interpretable, this higher legal standard places greater weight on the importance of clinicians understanding the most important clinical features in a ML prediction, if they rely upon it when discussing treatment with a patient. This is explored further in the next section.

4. Patient decisions: 'material information' & personalised transparency

When a patient has to choose between treatment options, clinicians must provide sufficient information to support this individual, autonomous choice.[66] This role relies on a transparent, collaborative dialogue between the healthcare professional and the patient. While no current treatment guideline or standard decision-tree aligns with every patient's subjective thought process, models can be adapted to allow personalised weighting of varying outcomes or risk factors, and could assist clinicians and patients in visualising priorities and structuring decisions. The decision-making process varies according to each particular patient's concerns, goals and values, and the decision the patient makes will not necessarily be motivated by clinical factors alone.[67]

We shall first outline the concept of patients' rights to information in European healthcare law, followed by the nature of the clinical advisory duty to support patient choice, and the information about risk this requires. While previous sections have been generically pan-EU by focusing on the GDPR, here

[62] Selbst and Powles, note 16; Brkan note12; Kaminski note 18; Malgieri and Comandé note 57. See also Bryan Casey, Ashkan Farhangi and Roland Vogl, "Rethinking Explainable Machines: The GDPR's "Right to Explanation" Debate and the Rise of Algorithmic Audits in Enterprise' (2019) 34 Berkeley Technology Law Journal 143

[63] Article 29 Working Party, note 38

[64] Ibid

[65] Erik Štrumbelj and Igor Kononenko 'Explaining prediction models and individual predictions with feature contributions' (2014) 41 Knowledge and Information Systems 3

[66] *Montgomery v Lanarkshire Health Board* [2015] UKSC 11, [2015] A.C. 1430 in the UK—other sources of law considered later on in the section.

[67] Ibid, para 49

we narrow our focus to particular jurisdictions to permit consideration of the national detail. We use several European jurisdictions as case studies in illustrating the interplay between patient rights and physician duties, and how this has evolved to a patient centred standard. We will then turn to the current state of the art in explainable machine learning models in healthcare. We compare this with the standard of information currently required in the jurisdictions examined.

Although the sources and precise nature of healthcare law vary between civil and common law jurisdictions across Europe, we focus on three civil law jurisdictions (Norway, Denmark and Sweden) and two common law (the UK and Ireland) to illustrate the commonalities in the information which must be available to patients, and therefore the standard to which machine-learning models would be held when their predictions are intended to inform patients' choice of treatment.

4.1. *Patients' rights to information*

During the 1990s, two important statements on patients' rights to information were adopted at a European intergovernmental level, in light of a growing international consensus on patient autonomy. Firstly, the 1994 WHO Declaration on the Promotion of Patient's Rights outlines a detailed right to information. This includes information on medical facts about one's condition; about the proposed medical procedures, together with the potential risks and benefits of each procedure; about alternatives to the proposed procedures, including the effect of non-treatment; and about the diagnosis, prognosis and progress of treatment. Although non-binding, the declaration has been influential, as is apparent from the Danish health law (see below).

A more limited right to information is recognised in the 1997 Council of Europe Convention on Human Rights and Biomedicine ('the Biomedicine Convention'), whereby prior to consent, patients must be given "appropriate information" as to the purpose and nature of the intervention as well as on its consequences and risks.[68] While the Convention is legally binding, twelve Council of Europe Member States have still not ratified, including the United Kingdom and Ireland. Still, the Convention is an important source of obligations, which the European Court of Human Rights cites in its judgements, including *Glass v. UK*, discussed in the next section. Despite the divergences in healthcare law at a national level, through Council of Europe conventions, European jurisdictions share common threads in the information which must be offered to patients.

4.2. *Common law rights*

In terms of the UK and Ireland, the information requirements are articulated as an 'advisory' duty of disclosure. This duty will exist in some form in any jurisdiction governed by the European Convention on Human Rights[69] ('ECHR') as the importance of patient decision-making has been emphasised by the European Court of Human Rights (ECtHR).[70] Similarly, states that have ratified the Biomedicine Convention are obligated to ensure that clinicians provide patients with "objective information" that is "sufficiently clear and suitably worded".[71] For the purposes of the UK example, we will focus on the articulation of the advisory duty by the UK Supreme Court in *Montgomery v Lanarkshire Health Board*,[72] as this judgment distilled the ECtHR case law on Article 8 ECHR into an obligation under negligence law to provide adequate information to patients about material risks of treatment. Some form of this duty should thus apply in any jurisdiction seeking to comply with the ECtHR. Versions of the *Montgomery* duty of disclosure have also been developed in a number of Commonwealth jurisdictions,[73] such as Singapore.[74] This was preceded in Ireland by the case of *Geoghegan v Harris*,[75] however, which established a similar, patient-centred, duty of disclosure in elective surgery.

The evolution of the distinct advisory duty of care has been gradual, and has been particularly informed in Europe by the jurisprudence on Article 8 of the ECHR. Originally and ostensibly a right to respect for one's private and family life, it has evolved a depth of nuance whereby a 'private life' is not just a sphere of domestic activity, but also captures a person's physical and psychological integrity, and the autonomy to control intervention in this physical and emotional bodily space.[76] Influenced by developments in pan-European law, the duty to advise patients on treatment options casts the clinician in a supportive and responsive role, with the patient (or their legal proxy in cases of incapacity) at the heart of decision-making.

Box 1. Glass v UK

> The judgment of the ECtHR in *Glass v United Kingdom*[77] illustrates the fundamental importance of patient/representative control over life and death decisions. David Glass had severe disabilities, with his mother making decisions on his behalf. He became very unwell following surgery and was ventilated in the intensive care unit. The prognosis was poor, and Ms Glass was informed her son was dying. Against her wishes, he was given diamorphine, which appeared to cause deterioration in his condition. In the midst of a physical altercation with hospital staff, Ms Glass resuscitated her son. His condition improved and he was able to return home the same day. The ECtHR found that the administration of the morphine

[68] Article 5, Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine, Oviedo, 4.IV.1997 ETS 164. Hereafter cited as 'the Biomedicine Convention.'

[69] Convention for the Protection of Human Rights and Fundamental Freedoms, Rome, 1950 Council of Europe European Treaty Series 5.

[70] *Glass v UK* (2004) EHRR 341; *Tyslac v Poland* (2007) 45 EHRR 42

[71] Explanatory Report – ETS 164 – Human Rights and Biomedicine Convention, para 35, 36.

[72] Note 66

[73] Louise V Austin, '*Hii Chii Kok v (1) Ooi Peng Jin London Lucien; (2) National Cancer Centre*: Modifying Montgomery' (2019) 27 Medical Law Review 2

[74] Lysaght et al, note 10

[75] [2000] 3 IR 536

[76] Bart van der Sloot, 'Where is the Harm in a Privacy Violation? Calculating the Damages Afforded inPrivacy Cases by the European Court of Human Rights' (2017) 8 JIPITEC 322

against Ms Glass' wishes, and without authorisation by a court, was a breach of Article 8 and a violation of the patient's physical integrity.

It is clear based on Box1, that, even where a model might yield a gloomy prediction of mortality, potentially life-shortening palliative care could not be administered without the consent of the patient or their proxy (a relative or a court where necessary). The more difficult question is what information the patient/proxy will need about the predictions or recommendations of the model. Lysaght and colleagues note the advisory duty of care under Singaporean negligence law, requiring the patient to be provided with material information. However, they argue what the level of information about an ML model would be considered 'material' is unclear.[78]

Obermeyer and Emmanuel predict that better estimates of survival:

'*could transform advance care planning for patients with serious illnesses, who face many agonizing decisions that depend on duration of survival.*'[79]

The judgment in *Glass* illustrates not only the importance of better estimates chances of survival, but also of transparent estimates of survival. It was not enough for the clinicians to estimate survival and alter the treatment plan accordingly; the patient's mother (or, in the face of her opposition, a court) needed information about the basis for the prediction, the degree of confidence, and the ensuing options for treatment. If a clinician were relying on an automated prediction of mortality through an ML model, they would still need to provide this same level of information to the patient, meaning they would need to understand (at least from a medical, if not a statistical, point of view) why the ML had made its prediction.

To get a better idea of what this detail could look like, it is worth turning to the UK Supreme Court's decision in *Montgomery*, described in Box 2.

Box 2. Montgomery

Like *Glass v UK*, the *Montgomery* case was brought by a mother—maternity being a key battleground in which the lines of patient autonomy have been drawn.[80] In this case, Ms Montgomery was pregnant, diabetic and of below average height. As a diabetic, she was at risk of having a larger than average baby. Despite the concerns Ms Montgomery voiced about delivering a larger baby, her doctor chose not to advise her of the risk of shoulder dystocia, and the potential complications which could stem from it.

In the event, the risk of shoulder dystocia did materialise, and the ensuing complications meant the claimant's son developed cerebral palsy. The Court accepted that if informed of the risk Ms Montgomery would have elected to have a caesarean section and her child might have been born uninjured. Ms Montgomery should have been told about risks the reasonable patient in her position would deem material (or which a doctor should reasonably know was material to her, having spoken to her about the birth).

As such, the advisory duty of care was solidified into UK law, reflecting the jurisprudence of the ECtHR. The test as articulated in para 87 of the judgment can be summarised as follows:

The doctor is therefore under a duty to take reasonable care to ensure that the patient is aware of any material risks involved in any recommended treatment, and of any reasonable alternative or variant treatments. The test of materiality is whether, in the circumstances of the particular case:

- a reasonable person in the patient's position would be likely to attach significance to the risk (i.e. the 'objective test'), or
- the doctor is or should reasonably be aware that the particular patient would be likely to attach significance to it (i.e. the 'subjective' strand of the test).

For the purposes of this paper, it is the 'subjective' strand of the Montgomery test which is particularly important. It means that medical information cannot be generic, but must be tailored to patients, at least to the extent that the doctor should be reasonably aware of their particular concerns.

Similarly, the Irish courts have recognised a clinical 'duty to warn' in advance of elective surgery. In 2000, the Irish High Court held that:

the 'reasonable patient' test, which requires full disclosure of all material risks incident to proposed treatment, is the preferable test to adopt, so that the patient, thus informed, rather than the doctor, makes the real choice as to whether treatment is to be carried out. It is the view of this Court that assessment of the duty of disclosure on this basis is more logical than the professional standard test, whereby the Court adopts the standard of the medical profession, yet reserves the right to override the views of the medical experts as and when it sees fit.[81]

In this case, although the risk (of chronic neuropathic pain from a dental procedure) was remote and experts testified that they too would not have warned the patient, the Court held that the physician was under an obligation to warn.[82] It restated the general principle that the patient has a right to know and the physician has a duty to advise of all material risks. Ultimately, it is for the court to decide what is material, which should include consideration of (a) the severity of the consequences and (b) statistical frequency of the risk.[83] In 2007, the Irish Supreme Court confirmed that a '*patient centred test is preferable*'(to a doctor-centred approach).[84]

Thus, the amount of information provided has to be tailored to the patient under UK and Irish law, although there are basic, minimum standards of disclosure where 'material risk' is concerned. It therefore follows that, if an ML calculation is relied upon in discussing treatment options, a doctor should be capable of elaborating on the medical reasons behind the prediction, so the patient can understand whether they are static (e.g. height, genetic profile) or potentially dynamic (e.g. age, BMI, timing of procedure).

[78] Note 10

[79] Ziad Obermeyer, and Ezekiel J. Emanuel, 'Predicting the Future — Big Data, Machine Learning, and Clinical Medicine' (2016) 357 New England Journal of Medicine 13

[81] *Geoghegan v. Harris* [2000] IEHC 129; [2000] 3 IR 536 (21st June 2000), para 165

[82] Ibid para 86

[83] Ibid para 98

[84] *Fitzpatrick -v- White* [2007] IESC 51

The next subsection considers Denmark. Sweden and Norway and reveals similar themes in their healthcare transparency requirements.

4.3. *Civil law rights*

In civil law jurisdictions, physicians' duties and patients' rights have been similarly influenced by international developments on patient autonomy, such as the Biomedicine Convention. In contrast, rights are primarily defined in legislation and administrative regulations.[85] These statutory rights to information have substantive overlap with the common law duty of disclosure.

In the Nordic context, legislation – not case law - enshrines the patient's right to information and participation in healthcare. Compared to the common law context, legislation provides greater detail as to the information to which patients are entitled. For example, the Danish health law resembles the WHO Declaration, stating that:

'*Patients have a right to receive information on the state of their health and on treatment options, including on risks for complications and side effects…*

The information must be provided on an ongoing basis and provide an understandable explanation of the disease, the examination and the intended treatment. The information must be given in a considerate manner and be adapted to the patient's individual requirements, including age, maturity, and experience.

The information shall include information on relevant prevention, treatment and care options, including information on other, clinically sound treatment options, as well as information on the consequences of no treatment being initiated. The information must also include details about potential consequences for treatment options, …. The information must be more comprehensive when the treatment entails an obvious risk of serious complications and side effects…'[86]

This provision was introduced in 1998, whereby for the first time legislation framed information as a right – as opposed to a clinical duty.[87] The purpose of the 1998 law was to strengthen the legal position of, and legal certainty for, patients in light of international and national developments related to self-determination, notably to bring Danish law in line with the Biomedicine Convention, which was adopted in 1996 and the WHO Declaration on Patients' Rights in Europe.

The preparatory works of the legislation outline the two 'scales':

a) 'serious complications' to 'trivial complications';
b) 'frequent complications' to 'rare complications'.

Four combinations are possible:
1) severe + frequent;
2) severe + rare;
3) trivial + frequent;
4) trivial / rare.

Following the preparatory works, cases 1) and 2), must always be thoroughly informed. Case 3) should often be informed. In case 4) information is usually not required. Furthermore, the healthcare professional should take the particulars of the patient into account, for example, for an athlete, surgery poses risks that are not present for an office worker. However, further complicating matters, guidance from the Board of Health differs from the above, finding that case 2 should often be informed (but not necessarily always).[88] The legislation thereby does not specify when and which risks must be transmitted, only that more comprehensive information is required where there is an obvious risk of serious complications and side effects.

As an aside, it can be noted that a slightly different approach is found in the Norwegian Patient Rights Act, which recognises a right to information that is necessary to gain an insight into the state of one's health and treatment.[89] The Swedish law states that patients must be informed of significant risks of complications and side effects.[90] All Nordic countries recognise that information must be adapted to the patient's situation, such as age, maturity, experience and cultural and language background.[91]

Applying the facts of *Montgomery* yields a similar result under Danish law, highlighting the commonalities between the two jurisdictions. The risk of shoulder dystocia in this case was calculated at 9–10% and the risk of severe injury was very small, Danish law would require that the patient be informed of the risk of shoulder dystocia and would usually require the patient to be informed of the danger of injury to the baby. In the particular case, given that the woman had expressed concern about the size of the foetus, it would be likely that this risk should also be conveyed, given the focus on personalising information to the patient's needs.

The specificity of the right to information is illustrated in a decision of the Danish Agency for Patient Complaints, where it criticised a hospital for inadequately informing a patient prior to induction. The patient was a first time mother, who was four weeks overdue (41 + 4). She was prescribed misoprostol to induce childbirth. The complainant argued that she had not been informed of the risks associated with this drug, although the hospital disputed this. The Agency held that patients should be informed of the specific side effects and complications associated with medication used to induce. It was insufficient to simply state that serious side effects were possible.[92]

Nordic health law provides that patients have a right to information on treatment methods that could be considered

[85] Mette Hartlev, 'Informed consent in the Nordic countries' in *Informed Consent and Health: A Global Analysis* (Global Perspectives on Medical Law) Eds Thierry Vansweevelt, Nicola Glover-Thomas (2020, Elgar)

[86] Sundhedsloven, LBK nr 903 af 26/08/2019 (hereafter the Danish Health Act), § 16

[87] The legislation draws on a circular from the Ministry for Health from 1992.

[88] Vejledning om information og samtykke og om videregivelse af helbredsoplysninger mv. Sundhedsstyrelsens vejledning nr. 161 af 16/9 1998, at 3.3. This approach is often followed by the Healthcare workers Disciplinary Board (see further, Mette Hartlev et al, Sundhed og Jura (Jurist-og-.Økonomforbundets Forlag, 2017), p. 190)

[89] Patient and User Rights Act (Patient and User Rights Act) (consolidated, 2018) §3.2. See further, Anne Kjersti Befring, *Persontilpasset medicin: Rettslige Perspektiver* (2019, Gyldendal).

[90] Patient's Act (2014:821), chapter 3.2

[91] Patient and User Rights Act (Patient and User Rights Act) (consolidated, 2018) § 3.5

[92] Case nr. 14POB085 (24. November 2014) https://stpk.dk/da/afgoerelser/afgoerelser-fra-styrelsen-for-patientklager/behandlingssager/2014/14pob085/

akin to the right to an explanation. Notably, the law does not necessarily limit patients' entitlements to information on 'material risks'. Instead, patients are entitled to understandable information, inter alia, regarding treatment and, importantly, potential alternative treatments. Here again, health law underlines that patients are entitled to understandable information.

From the above overview of the information requirements within healthcare law, we can summarise the information which should be provided to patients in different jurisdictions as follows:

Country	Source of Law	Information Required
UK	Common law duty of disclosure (*Montgomery*), albeit broadly underpinned by Article 8 ECHR and Human Rights Act 1998.	• Material risks (per reasonable patient, or per individual patient if doctor reasonably should be aware that patient would regard the risk as material) • Treatment options
Ireland	Common law duty of disclosure (*Geoghegan*).	• Material risk (per reasonable patient)
Denmark	Statutory right, s.16 Health Act 1998	• Clinically sound treatment options • Understandable explanation of disease and treatment • Right to refuse information[93] • Consequences of no treatment • More information required if there is an obvious risk of serious complications
Norway	Statutory right, Patient Rights Act	• Information necessary to gain an insight into health and treatment • Possible risks and side effects • Injuries or serious complications & right to apply for compensation
Sweden	Statutory right, Patient's Act 2014	• Patient's state of health, • The methods available for examination, care and treatment, • The aids available for people with disabilities, • At what time he or she can expect to receive care, • The expected course of care and treatment, • Significant risks of complications and side effects, • Aftercare • Methods of preventing disease or injury. • The possibility of choosing treatment alternatives, • The possibility of obtaining a new medical assessment • The care guarantee, and • The possibility to obtain information from the Swedish Social Insurance Agency about care in another country within the European Economic Area

While there are variations in the above informational requirements, the following can be consolidated as a list of information a patient decision-maker must be offered that cuts across these jurisdictions:

1) The treatment options reasonably available to them (including the consequences of no treatment);
   The 'material risks' of treatment, according to what most patients would deem material (or according to that particular patient if the doctor should be aware of their concerns from talking to them) including significant complications and side-effects;
   More detail is required for a more serious or more likely risk.

The level of information offered to enable informed decision-making should be calibrated based on factors intrinsic to the patient, and the information in which they express an interest as part of a consultative dialogue and not factors intrinsic to the technology used by the clinician (such as whether it constitutes automated processing.)

4.4. *Explaining ML models*

As suggested above, we agree with those who suggest that adequate explanation of ML can be achieved through post-hoc use of models which illustrate feature relevance. However, we do not suggest that such 'explanations' should be given directly to patients, as the result may be too complex to support their decision making. Wachter and colleagues[94] suggest a counter-factual style of explanation could be given to patients:

2) Person 1: If your 2-Hour serum insulin level was 154.3, you would have a score of 0.51.
   Person 2: If your 2-Hour serum insulin level was 169.5, you would have a score of 0.51.
   Person 3: If your Plasma glucose concentration was 158.3 and your 2-Hour serum insulin level was 160.5, you would have a score of 0.51.

This style of explanation does not lend itself to patient communication in all cases, however. Having used the judgement in *Montgomery* as a key UK benchmark for informed consent, we have considered a 'counterfactual' style explanation based on the facts of this case:

Box 3. The Maternity Counterfactual

If an expectant mother has her risk of birth complications calculated by a model, the type of information taken into account could include the following:
- Her height and stature
- Her medical history, including the diagnosis of diabetes and any history of previous dystocia, anaesthesia and c-sections
- The baby's apparent size at ultrasound
- The baby's due date
- The margin of error in calculating the baby's birth-weight
- The increasing risk of shoulder dystocia for each additional 0.5 kg of birth-weight

[93] Danish Health Act, §16(2).

[94] Note 54

- The separate percentage risks of shoulder dystocia leading to a significant negative outcomes such as post-partum haemorrhage, plexus injuries or permanent neurological disorders of varying severities, including cerebral palsy

Some factors are static, while others are dynamic, e.g. if more than one ultrasound is performed, or the baby is delivered sooner or later than predicted. The margin of error would be important if the patient wanted to consider the worst-case likely scenario. The extent to which an additional 0.5 kg of birth-weight would affect the risk is also worth knowing, as a 10% margin of error could mean a significant increase in risk compared to the working estimate. Conversely, if a small amount of additional weight would not make a difference, the patient may feel less troubled by the margin of error.

The proliferation of numbers and risk factors to be compared quickly becomes so extreme as to be unhelpful, even when clearly presented. For example, the above case, where the individual risk of each type of complication for multiple possible courses of action at multiple possible birth dates and weight, would provide many hundreds of thousands of risk assessments. As the UK ICO observe, counterfactuals can therefore have limitations that originate in the variety of possible features that may be included in representing alternative outcomes.[95] This illustrates why certain methods of explaining ML may work well in some contexts, but not others, and therefore why we have refrained from advocating any particular explanatory tool in this paper.

Bearing in mind that a patient should receive a minimum amount of information about material risks of different treatment options—with additional information only if desired—we do not see that the style of explanation outlined in Box 3 necessarily lends itself to a tailored, collaborative discussion between doctor and patient. While some patients may find this level of quantifiable detail assists their reasoning, many may also prefer to focus on the core risk factors, and narrow down their options accordingly.

For this reason, we advocate post-hoc, rationale explanations of ML, to assist clinicians in identifying why it has generated an output for their particular patient—i.e. what were the key factors in its conclusion. This can be combined with their own medical expertise, and independently verified in light of their training and experience. They can then translate this into appropriately personalised information given to a patient. This is why we have argued that 'explanations' of a model's outputs should be targeted at healthcare professionals, so they have sufficient clarity to translate the outcome into the personalised level of 'information' an individual patient requires. This level of information is ultimately open-ended under healthcare law, and so it is the mediating professional who must have sufficient comprehension of the model to deliver the contextually necessary details to obtain informed consent. While the explanation a clinician will require can be predicted and automated (e.g. visuals of the areas of a scan identified as features of eye disease[96]), this is not the case for patient information, which will vary according to the priorities and values of that individual,[97] and thus cannot be standardised.

To make an informed choice, it will suffice for a patient to be informed of key risks, which in *Montgomery* were timing, and the attendant risks of birth weight and dystocia—e.g. '*if you are induced at week X, your baby's weight is likely to be Y, which will mean the risk of shoulder dystocia is Z.*' We do not suggest that patients should, as a matter of course, be given multiple detailed counterfactuals. But one or two counterfactuals, at the level of: '*if you go into labour at week A, instead of being induced at week X, risk of dystocia might be as low as B.*' This level of detail may be helpful for a patient to make an informed decision about their care, but ultimately it is for the data controller to choose the most suitable method of explaining ML to clinicians (and, thus, indirectly to patients).

Whatever supplementary model is selected, the information offered to a particular patient must be tailored to that patient's priorities, concerns and interests; 'explainability' of models should mean that a clinician has access to information about the factors which play a role in a clinical prediction or recommendation. This does not mean simplifying an ML model to the extent that its utility and accuracy are sacrificed; an additional model can instead be used to highlight the model's feature weighting.[98]

For this reason, we do not advocate that explanations of ML operations should be routinely provided to patients as the 'information' they require to make decisions. As O'Hara has argued, such computed accounts do not amount to an 'explanation' in the social sense, which is not so much a *text* as a *process* designed to bring about understanding in the recipient.[99] While explainable ML may put clinicians in a clearer position to have this dialogue, the information the patient receives will ultimately be co-produced via their questions, clarifications and challenges in their attempts to achieve their desired level of understanding.[100] Some have argued that model outputs do not need to be explained to patients at all, and it is sufficient for patients to receive an automated list of prioritised treatment options, with the clinician merely providing information as to the medical implication for each option.[101] We do not consider this to be enough, however: if a patient is told that a particular option is recommended as a priority, the 'material information' requirement would suggest they need some sense as to *why* it has been prioritised. Without at least some rationale explanation of the recommendation, the patient will be in the dark as to why the ML recommends an

[96] See note 26

[97] T.T Arvind & Aisling McMahon, 'Responsiveness and the Role of Rights in Medical Law: Lessons from *Montgomery*' (2020) 28 Medical Law Review 3

[98] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin, '"Why should I trust you?" Explaining the predictions of any classifier'(2016) Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining., pp. 1135–1144. doi: 10.1145/2939672.2939778

[99] Kieron O'Hara, 'Explainable AI and the philosophy and practice of explanation' (2020) 39 Computer Law & Security Review 105474

[100] Ibid

[101] Juan Manuel Durán & Karin Rolanda Jongsma, 'Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI' (2021) Journal of Medical Ethics available from <doi: 10.1136/medethics-2020-106820>

option for them, and this mystery interacts awkwardly with the logic of an 'informed' decision. We therefore concur with Kiseleva that rationale explanations are necessary to support informed consent in medicine.[102]

The patient (or their representative) has the ultimate authority to decide their preferred choice of treatment, and this decision is acknowledged to involve objectives and values which go beyond a purely medical assessment.[103] This subjective thought process cannot be automated, even in 'personalised' medicine, but must be supported by information relating to material risks. The doctor's role here is socially responsive; they must help the patient relate the clinical information to their personal values.[104] If a model's predictions form part of this clinical information, the information should be offered to the patient not just for data protection reasons, but because it is their right under the ECHR to make their own informed and autonomous decisions about their healthcare.[105] While we have borrowed the term 'transparency' from the GDPR, patients' rights under national healthcare law have evolved from more fundamental (and indeed more intimate) considerations of dignity and self-determination, rather than the fair use of information within a common data market. This means associated rights to information are inevitably more person-focused.

Under the assorted sources of European healthcare law, it is the patient who must 'personalise' their healthcare in collaboration with their clinical team and those with whom they have close, caring relationships. This network of caring relationships, at the centre of which is the relationally autonomous patient, is what ultimately 'personalises' healthcare, not an automated model.[106] The 'autonomy' the patient exercises is thus of a more 'relational' form, which prioritises collaborative discussion, joint responsibility, and mutually inter-dependent decision-making.[107] Medical information provided in advice should therefore be capable of adaption to support this collaborative communication. Explainable ML can help clinicians provide personalised information, but no computed 'explanation' of a model's workings could be provided as a substitute.

5. Conclusion

We have, in this article, explored the use of ML models to support three broad categories of decision in healthcare:

1) System-wide, fully automated decisions (e.g. resource allocation & patient triage): these solely automated decisions will unambiguously be subject to the GDPR requirement to provide data subjects with 'meaningful information' about the logic involved in the processing, as well as its significance and envisaged consequences (Articles 13, 15 and 22). If a more complex form of ML is used—such as a neural network—this should at least be subject to post-hoc explanation so that 'meaningful information' is available for data subjects.
2) Clinical Decisions: decisions taken by healthcare professionals—with only partial reliance on ML and not requiring the informed consent of the patient (e.g. a diagnosis) represent a relative lacuna. The information required for these types of decisions under the GDPR falls short of the 'meaningful information about the logic involved,' although the A29WP guidance suggests patients should still be made aware of profiling & its consequences in these cases. We still suggest that models should be at least explainable to clinicians in these circumstances, however, for the following reasons:
   a Some comprehension of the ML output assists the clinician in their general duty to provide reasonable, non-negligent care in making the decision in partial reliance upon it;
   b Comprehension of the consequences of profiling may still require some meaningful information to patients, even under the lower GDPR standard;
   c The boundary between clinical and patient decisions is not airtight: patients may query a diagnosis—and its level of confidence—in making choices about their treatment.

   For these reasons, we do not support a lower standard of ML explanations in clinician-mediated decisions, even if the GDPR's requirements may be less strict in this regard.

1) Patient decisions: where the informed consent of a patient (or their representative) is required, and medical advice is even partly based on an ML output, there is a more personalised standard of material information required across various national healthcare laws. The requirement to provide context-specific, tailored information to patients in support of their decisional autonomy is an additional reason why healthcare users of ML should have access to explanations, to support them in their advisory obligations.

In all three cases, rationale explanations of ML models should be available to healthcare professionals, so that patients can be given the information to which they are entitled under these different standards of transparency.

It has already been argued that the GDPR's transparency obligations should be expanded across all medical machine-learning models for ethical reasons, regardless of whether Article 22 applies or not (i.e. whether the decision is 'solely' automated).[108] The duty of disclosure is another compelling reason for patients to have access to enough information about how 'material risks' are calculated. We argue that patients' existing rights to adequate information under health law in fact includes the right to adequate information, regardless of the type of medical science involved in creating the information.

[102] Note 20
[103] *Montgomery* (note 66) para 45
[104] T.T Arvind & Aisling McMahon, note 97
[105] This has also been termed 'decisional privacy,' see Bart van de Sloot note 28
[106] Jonathan Herring, 'Law and the Relational Self' in *Law and the Relational Self* (Cambridge, CUP 2019)
[107] Jennifer K Walter and Lainie Friedman Ross, 'Relational Autonomy: Moving Beyond the Limits of Isolated Individualism' (2014) 133 Paediatrics S16
[108] Ferretti et al, Note 11

As an aside, it is worth noting that the advisory duty discussed in this paper does not apply only in a healthcare context. Following *Montgomery* in the UK, a duty to disclose material risk has also been found in the financial[109] and legal[110] sectors. This additional source of transparency obligations will therefore have implications for models used in other contexts, and perhaps wherever a legally significant decision is made by the data subject with the assistance of a model.

We have suggested that only a minority of significant decisions in healthcare will be made on a solely automated basis. While some GDPR transparency requirements will still apply when clinicians oversee or make decisions based on automated profiling, this is potentially to the lower standard of general transparency under Article 5 GDPR. Patients should still be made aware of any profiling, and its consequences, following the general transparency requirements as interpreted via Recitals 60 and 71 GDPR, even when decisions are not fully automated.

When patients must make treatment decisions for themselves, however, the additional support of the ECHR and the ECtHR case law creates a distinct kind of disclosure obligation. The ability of adults with capacity to make healthcare decisions to be offered enough information to make an informed choice is a key cornerstone in healthcare law. Even where a decision is made on behalf of a child or an adult lacking capacity, the same level of information will be required for their proxy. The judgment in *Glass v UK* suggests that the principle of decisional autonomy could also apply to treatment based on survival prediction, meaning that palliative care should only be obtained on the basis of sufficiently transparent information to be capable of discussion between clinicians and patients.

We therefore recommend post-hoc, rationale explanations—aimed at healthcare professionals, rather than directly at patients— as aligning with minimum standards of accuracy and transparency for ML models. While human interpretable models may be preferable when the standard of prediction and recommendations they can give are of sufficient accuracy, in the cases where accurate predictions can be obtained only through opaque (non-human-interpretable) models because of the complexity of relevant input data, these should fulfil minimum criteria of transparency through post-hoc, rationale explanation, empowering clinicians in as far as possible to give tailored information to patients.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data for reference

No data was used for the research described in the article.

Acknowledgments

The authors would like to thank Eugenijus Gefenas, Jurate Lekstutiene, Mette Hartlev, Piotr Jaroslaw Chmura, Vilma Lukaseviciene and our collaborators in EU-STANDS4PM for their help developing this paper.

The authors of this article are part of the EU-STANDS4PM consortium (www.eustands4pm.eu) that is funded by the European Union Horizon2020 framework programme of the European Commission under Grant Agreement #825843.

Katharina Ó Cathaoir also acknowledges that this work was partly supported by NordForsk and Innovation Fund Denmark through funding to PM Heart, project number 90580. Catherine Bjerre Collin acknowledges support from Novo Nordisk Foundation (grant agreement NNF14CC0001).

The funders have played no role in the conception or writing of this article, or the decision to submit it for publication.

[109] *O'Hare v Coutts & Co* [2016] EWHC 2224 (QB))

[110] *Baird v Hastings (t/a Hastings and Co Solicitors)* [2015] NICA 22, [2015] 5 WLUK 107