III. Transparency in Health Care

Too Much Transparency Is Not Always a Good Thing?

Paul Nolan

A. Introduction

The roots of Western medicine stem from Ancient Greece where Hippocrates introduced numerous medical terms universally used by physicians, including symptom, diagnosis, therapy, trauma and sepsis.[1] There was the creation of empirical medicine grounded in ethical vows. Whilst most clinicians are familiar with the Hippocratic Oath, they are less likely to be familiar with the medical texts of that time. Many now view the Greek physician–patient relationship as paternalistic, in which the physician concealed diagnostic or prognostic information from the patient.[2]

With the requirement of informed consent so prevalent in recent decades, has that paternalistic concealment been largely obliterated or is there some remaining scope for a balancing exercise between the right to be informed and the risk that too much information may, inadvertently or otherwise, lead to harm. The Hippocratic Oath amongst other things adopts the Latin maxim '*primum non nocere*' which translates to 'First, do no harm'. It is followed later by, 'Then try to prevent it'. Is a clinician who prevents harm due to information overload necessarily acting in the patient's best interests? The question is open.

It is sometimes said: "A little knowledge is a dangerous thing. So is a lot".[3] It is also said that: "Knowledge without practice is useless and practice without knowledge is dangerous".[4] Regardless of which is right or wrong, there are certain situations where too much information (and in the situation about to unfold involving Artificial Intelligence in healthcare) too much transparency is not always a good thing. Human beings are

---

1 *C. F Kleisiaris/C. Sfakianakis/I. V. Papathanasiou*, Health care practices in ancient Greece: The Hippocratic Ideal, Journal of medical ethics and history of medicine 2014, available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4263393/ (last access: 30.09.2022).

2 *S. H. Miles*, Hippocrates and informed consent, The Lancet 2009, 1322–1323.

3 Albert Einstein, German-born physicist and founder of the theory of relativity 1879–1955; a variation on a quotation of Alexander Pope, English poet, 1688–1744.

4 Confucius, Chinese philosopher, 551–479BC.

individuals whose capacity to understand varies greatly. What may seem straightforward and intelligible to one person, is beyond comprehension to the next. Within those extremities there is a wide cognitive spectrum. Within the complexity of AI in medicine, the need for informed consent, and the requirements for AI transparency, it may be that therapeutic privilege will play an increasing role in assisting clinicians and, ultimately, patients to achieve optimal healthcare outcomes.

B. Current AI Developments in Healthcare

Artificial Intelligence (AI) technologies, such as machine learning (ML), are gaining importance in healthcare.[5] Healthcare is a priority area of AI development, with both governments and private sector pouring significant investments into the field. AI-enabled medical applications have been developed that promise to improve diagnosis;[6] assist in the treatment and prediction of diseases;[7] improve clinical workflow; enable high quality direct-to-consumer services,[8] such as wearable monitoring devices[9]; aid

5 For current paragraph, see generally: *R. Matulionyte/P. Nolan/F. Magrabi/A. Beheshti*, Should AI-enabled medical devices be explainable, International Journal of Law and Information Technology 2022, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4140234 (last access: 20.10.2022).

6 *G. Litjens/C. I. Sánchez/N. Timofeeva/M. Hermsen/I. Nagtegaal/I. Kovac/C. Hulsberge -van de Kaa/P. Bult/B. van Ginneken/J. van der Laak*, Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis, Scientific reports 2016, 1; *N. Zhang/G. Yang/Z. Gao/C. Xu/Y. Zhang/R. Shi/J. Keegan/L. Xu/H. Zhang/Z. Fan/D. Firmin.*, Deep learning for diagnosis of chronic myocardial infarction on nonenhanced cardiac cine MRI, Radiology 2019, 606.

7 *A. Cheerla/O. Gevaert*, Deep learning with multimodal representation for pan-cancer prognosis prediction, Bioinformatics 2019, i446 -i454; *M. Roberts/D. Driggs/M. Thorpe/J. Gilbey/M. Yeung/S. Ursprung/A. I. Aviles-Rivero/C. Etmann/C. McCague/L. Beer/J. R. Weir-McCall/Z. Teng/J. H. F. Rudd/E. Sala/C. Schönlieb*, Machine learning for COVID-19 detection and prognostication using chest radiographs and CT scans: a systematic methodological review, arXiv 2020, available at https://europepmc.org/article/ppr/ppr347321 (last access: 20.10.2022)

8 *T. P.Quinn/S. Jacobs/M. Senadeera/V. Le/S. Coghlan.*, The three ghosts of medical AI: Can the black box present deliver?, (2022) Artificial Intelligence in Medicine 2022, 124; *A. Rajkomar/J. Dean/I. Kohane*, Machine learning in medicine, New England Journal of Medicine 2019, 1347.

9 *N. D. Lane/S. Bhattacharya/P. Georgiev/C. Forlivesi/F. Kawsar*, An early resource characterization of deep learning on wearables, smartphones, and internet-of-things devices, in: Proceedings of the 2015 international workshop on internet of things towards applications, New York 2015, p. 7.

in genome interpretation[10] and biomarker discovery[11]; and in automated robotic surgery[12]. Deep learning, a sub-set of machine learning, has already achieved near-human performance in medical image classification, such as the diagnosis of diabetic retinopathy.[13] Gradually, AI-enabled medical devices are gaining regulatory approval and being released to the market. The US Food and Drug Administration (FDA) has approved hundreds of AI-enabled medical devices and this number will continue to increase sharply.[14]

C. AI and The Black Box

The feature of AI that poses specific challenges to the evaluation of liability and regulation is its 'black box' nature.[15] Unless specifically programmed

10 *J. Zou/M. Huss/A. Abid/P. Mohammadi/A. Torkamani/A. Telent* A primer on deep learning in genomics, Nature genetics 2019, 12.

11 *S. M. Waldstein/P. Seeböck/R. Donner/A. Sadeghipour/H. Bogunović/A. Osborne/U. Schmidt-Erfurt*, 'Unbiased identification of novel subclinical imaging biomarkers using unsupervised deep learning, Scientific reports 2020, 1; *L. Li/F. Wu/G. Yang/L. Xu/T. Wong/R. Mohiaddin/D. Firmin/J. Keegan/X. Zhuang*, Atrial scar quantification via multi-scale CNN in the graph-cuts framework, Medical image analysis 2020, available at https://www.sciencedirect.com/science/article/pii/S1361841519301355 (last access: 30.09.2022).

12 *A. I. Chen/M. L. Balter/T. J. Maguire/M. L. Yarmush*, Deep learning robotic guidance for autonomous vascular access, Nature Machine Intelligence 2020, 104.

13 *Quinn/Jacobs/Senadeera/Lea/Coghlan*, The Three Ghosts of Medical AI (n. 8); *V. Gulshan/L. Peng/M. Coram/M. C. Stumpe/D. Wu/A. Narayanaswamy/S. Venugopalan/K. Widner/T. Madams/J. Cuadros/R. Kim/R. Raman/P. C. Nelson/J. L. Mega/D. R. Webster*, Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs, Jama 2016, 2402; *R. Sayres/A. Taly/E. Rahimy/K. Blumer/D. Coz/N. Hammel/J. Krause/A. Narayanaswamy/Z. Rastegar/D. Wu/S. Xu/S. Barb/A. Joseph/M. Shumski/J. Smith/A. B. Sood/G. S. Corrado/L. Peng/D. R. Webster*, Using a deep learning algorithm and integrated gradients explanation to assist grading for diabetic retinopathy, Ophthalmology 2019, 552.

14 https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices (last access: 14.06.2022) (The list is not meant to be an exhaustive or comprehensive resource of AI/ML-enabled medical devices. Rather, it is a list of AI/ML-enabled devices across medical disciplines, based on publicly available information).

15 *W. Nicholson Price II.*, Black Box Medicine, Harvard Journal of Law and Technology 2015, 420; *W. Nicholson Price II.*, Medical Malpractice and Black-Box Medicine, Big Data, Health Law, and Bioethics 2017, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2910417 (last access: 01.10.2022).

to do so by AI designers and engineers, an algorithm does not provide a rationale or explanation for its output and remains opaque (thus, 'black box'). The black box nature is amplified by an evolutionary nature of AI which is especially characteristic of ML based techniques such as deep learning or neural networks. Its very purpose is to learn, that is, change conclusions based on new data with which the algorithms are furnished. As will be seen below, opacity and evolution pose challenges when analysing the tort of negligence.

The 'black box' phenomenon is something that is troubling the medical profession — one which has the notion of trust as its bedrock.[16] ML systems specifically, and AI systems in general, are so structurally complex and can process such vast amounts of data that 'there is no straightforward way to map out the decision-making processes of these complex networks of artificial neurons.'[17] Black box AI in healthcare is problematic from several perspectives.

The medical profession is built on trust. The functioning of contemporary medicine relies fundamentally on trust.[18] Doctors interact with their patients in a personal and physical way, and they are privy to sensitive information about their patients. Intrinsically, trust is what imbues the clinician-patient relationship with its singularity and importance. For medical AI to be successful, it must be trusted by governments, health professionals, and the public. The quality of medical advice depends on medical reasoning being open to clarification, scrutiny, and evaluation.

Optimal healthcare delivery also depends on a satisfactory dialogue between experts, which can afford the clinician added information about a patient or a condition. This in turn could usefully feed back into the initial clinician–patient dialogue about prognosis and treatment options. AI systems that lack sufficient understanding, arguably do not allow for

16 See: *Breen v Williams* (1996) 186 CLR 71 where Brennan J at para 81 held that '[T]he relationship of doctor and patient is one where the doctor acquires an ascendancy over the patient and the patient is in a position of reposing trust in the doctor'.

17 *Y. Bathaee*, The Artificial Intelligence Black Box and the Failure of Intent and Causation, Harvard Journal of Law and Technology 2018, 891.

18 *R. Rhodes*, The Professional Responsibilities of Medicine, in: R. Rhodes/L. P. Francis/A. Silvers (eds.), The Blackwell Guide to Medical Ethics, Hoboken 2007; *S. Nundy/T. Montgomery/ R. M. Wachter*, Promoting Trust Between Patients and Physicians In The Era Of Artificial Intelligence, Journal of the American Medical Association 2019, 497; *J. Hatherley*, Limits Of Trust In Medical AI, Journal of Medical Ethics 2020, 478.

the satisfactory development of this knowledge and discovery process.[19] Further, an ability (or inability) to explain decisions in a clear manner will impact on how clinicians will utilise the information gleaned from AI systems when treatment plans are put in place.[20] In other words, an inability to adequately explain the AI system will inhibit trust from two perspectives: 1) the clinician not understanding the AI and being reserved about its capabilities and ultimate decisions; and 2) it follows that if the clinician does not have adequate understanding and trust in the AI system, this will filter down to the patient. If a clinician is unsure about how the AI system functions and, earnestly, discloses that to a patient, then the patient's confidence in the system will hardly be galvanised.

Some scholars warn that black box algorithms can hamper patient autonomy in clinical decision making.[21] A patient must be able to make their own autonomous decision about treatment options. That can only be achieved when they have both the decision and the reasons for it explained to them. That is a fundamental requirement of medical care, with or without AI systems. An AI system, by reason of its opacity or unexplainability, precludes a patient from being fully informed of how a particular recommendation or decision was arrived at. That, in turn, compromises their ability to decide whether to accept or reject the AI recommendation. This may provide an ethical reason to oppose the introduction of black box AI systems in that it would violate the right to informed consent.[22]

Being mindful of the problems discussed above, AI-enabled medical devices are expected to comply with a number of ethical principles and policy recommendations. AI ethical guidelines,[23] including healthcare-spe-

---

19 *C. Rudin*, Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead, Nature Machine Intelligence 2019, 206; For over-reliance on AI, see: *D. Lyell/E. Coiera*, Automation Bias and Verification Complexity: A Systematic Review, Journal of the American Medical Informatics Association 2017, 423.

20 *D. Lyell/E. Coiera/J. Chen/P. Shah/F. Magrabi*, How Machine Learning Is Embedded to Support Clinician Decision Making: An Analysis of FDA-Approved Medical Devices, BMJ Health and Care Informatics 2021 available at https://pubmed.ncbi.nlm.nih.gov/33853863/ (last access: 20.10.2022).

21 *T. Grote/P. Berens*, On the Ethics of Algorithmic Decision-Making in Healthcare, Journal of Medical Ethics 2020, 205, *R. J. McDougall*, Computer Knows Best? The Need for Value-Flexibility in Medical AI, Journal of Medical Ethics, 156 (158).

22 See generally, *H. R. Sullivan/S. J. Schweikart*, Are Current Tort Liability Doctrines Adequate for Addressing Injury Caused By AI?, American Medical Association Journal of Ethics 2020, 160; *McDougall*, Computer Knows Best? (n. 21).

23 See e.g. OECD, Principles on AI (2019) https://www.oecd.org/going-digital/ai/principles/; European Commission, Ethics Guidelines for Trustworthy AI (2019)

cific AI ethical guidelines,[24] require AI-enabled medical devices to respect such principles as benevolence, privacy and protection of data, safety, fairness, accountability and responsibility, avoidance of bias, governance, and others.

A sought-after principle is that of transparency and/or explainability, which is found in most ethical AI guidelines.[25] It mandates that AI in healthcare should be transparent and/or explainable.

While there is no consensus on what an AI explainability principle means, it generally requires that outputs generated by AI should be explainable and interpretable by different stakeholders, such as healthcare providers. The term 'transparency' has been used interchangeably or as

---

https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai; G20, AI Principles (2019) https://www.g20-insights.org/wp-content/uploads/2019/07/G20-Japan-AI-Principles.pdf; Australia's AI Ethics Framework (2022) https://www.industry.gov.au/data-and-publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles (last access: 01.10.2022).

24 The Royal Australian and New Zealand College of Radiologists: Standards of Practice for Artificial Intelligence (2020), https://www.ranzcr.com/college/document-library/standards-of-practice-for-artificial-intelligence; Royal College of Physicians and Surgeons (Canada) (2020), https://www.royalcollege.ca/rcsite/health-policy/initiatives/ai-task-force-e; International Medical Device Regulators Forum (SaMD) (Sept 2017), https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-170921-samd-n41-clinical-evaluation_1.pdf; World Health Organisation (WHO), Ethics and Governance of Artificial Intelligence for Health, https://www.who.int/publications/i/item/9789240029200 (last access: 01.10.2022).

25 See: Australia's AI Ethics Framework (2022), 'There should be transparency and responsible disclosure so people can understand when they are being significantly impacted by AI and can find out when an AI system is engaging with them.' https://www.industry.gov.au/data-and-publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles; OECD – Recommendation of the Council on AI (2022) para 1.3 https://oecd.ai/en/dashboards/ai-principles/P7; European Commission – Ethics Guidelines for Trustworthy AI (2019) 14–19 https://www.aepd.es/sites/default/files/2019-12/ai-ethics-guidelines.pdf (last access: 01.10.2022); *J. Morley/C. C. V. Machado/C. Burr/J. Cowls/I. Joshi/M. Taddeo/L. Florid*, The Ethics Of AI In Health Care: A Mapping Review, Social Science & Medicine 2020 available at https://www.sciencedirect.com/science/article/abs/pii/S0277953620303919 (last access: 20.10.2022); *A. Jobin/M. Ienca/E. Vayena*, 'The Global Landscape Of AI Ethics Guidelines, Nature Machine Intelligence 2019, 389.

a synonym of 'explainability',[26] while in other cases these concepts are clearly delineated.[27]

There is debate as to the level of transparency that should surround medical AI. That is, exactly how much of the AI system do clinicians and patients really need to understand before they can comfortably make an informed decision as to its use. Enabling patients to understand how AI determined diagnosis and treatment options are arrived at is crucial but also complicated. Clinicians also must be able to provide clear and cogent explanations of diagnoses and treatment options. This is something that bears upon the principle of informed consent, namely, before a patient can fully consent to something, they should at the very least have knowledge of all material matters.[28] Similarly, clinicians should be able to provide the explanations sought. Current ML systems lack explainability. In healthcare, if AI makes a decision that will impact on a patient, then all material risks need to be recognised, explained, and understood. Obviously, that will include knowing how the AI arrived at a given decision, which needs to be explained in layperson's terms and not replete with technical jargon.

D. Drugs and Other Medical Black Boxes

Strictly speaking, AI is not the only black box in medicine. In an article researching the ability of AI to improve the prediction and treatment of sepsis in hospital patients, Sendak noted that "[T]he human body is in many ways "a black box," in which the causes and mechanisms of illnesses often elude explanation".[29]

---

26 *Quinn/Jacobs/Senadeera/Lea/Coghlan*, The Three Ghosts of Medical AI: Can the Black Box Present Deliver? (n. 8); *A. Poon/J. Sung*, Opening the Black Box of AI-Medicine, Journal of Gastroenterology and Hepatology 2021, 581.

27 *G. Yang/Q. Ye/J. Xia*, Unbox the Black Box for the Medical Explainable AI via Multi-modal and Multi-centre Data Fusion: A Mini-Review, Information Fusion 2022, 29 (31).

28 *G. I. Cohen*, Informed Consent and Medical Artificial Intelligence: What to Tell the Patient?, Georgetown Law Journal 2019, 1425; *J. Amann/A. Blasimme/E. Vayena/D. Frey/V. I. Madai*, Explainability for Artificial Intelligence in Healthcare: A Multidisciplinary Perspective', 2020 available at https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-020-01332-6 (last access: 20.10.2022)

29 *M. Sendak/M. C. Elish/M. Gao/J. Futoma/W. Ratliff/M. Nichols/C. O'Brien*, "The human body is a black box" supporting clinical decision-making with deep learning, in: Proceedings of the 2020 conference on fairness, accountability, and transparency, 2020, pp. 99–109.

Electroconvulsive therapy (ECT) is a safe and effective treatment for certain psychiatric disorders. ECT is most commonly used to treat severe depression (major depression). It is often the fastest and best treatment available for this illness. ECT is also sometimes used to treat other psychiatric disorders, such as mania and psychosis. During ECT, a small amount of electrical current is passed through the brain while the patient is under general anaesthesia. This current causes a seizure that affects the entire brain, including the parts that control mood, appetite, and sleep. It causes chemical and cellular changes in the brain that relieve severe depression. Since the introduction of ECT in 1938, the mechanism of action of this highly effective treatment has intrigued psychiatrists and neuroscientists who do not yet fully understand exactly how it works.[30]

Certain drugs also remain to be fully explained, for example, lithium. Doctors don't know exactly how lithium works to stabilise the mood of a patient, but it is thought to help strengthen nerve cell connections in brain regions that are involved in regulating mood, thinking and behaviour. Another is acetaminophen (paracetamol). Despite competing explanations for how acetaminophen works, we know that it is a safe and effective pain medication because it has been extensively validated in numerous randomised controlled trials (RCTs).[31]

The point to be made is that despite being largely unknown or a black box, these drugs and treatments are regularly used in the healthcare system. That is because they have undergone RCTs which have historically been the gold-standard way to evaluate medical interventions. It should be no different for AI systems.

E. Informed Consent

The law of negligence is premised upon the general rule that those whose acts or omissions might injure another should exercise reasonable care to avoid such an occurrence. The elements that are required to be made out in an action for negligence can be stated as follows: the existence of a duty of care; a breach of that duty by a negligent act or omission; and damage suffered as a consequence. Fundamental to these, but often considered

30 *T. G. Bolwig*, How does electroconvulsive therapy work? Theories on its mechanism, The Canadian Journal of Psychiatry 2011, pp. 13–18.

31 *K. Toussaint/X. C. Yang/M. A. Zielinski/K. L. Reigle/S. Nagar/R. B.Raffa*, What do we (not) know about how paracetamol (acetaminophen) works?, Journal of clinical pharmacy and therapeutics 2010, 617–638.

separately, is the requirement of a causal connection between breach and damage.

In the context of healthcare, one area that has gained increasing attention over the past decades when looking at the tort of negligence is the principle of informed consent. That is, an analysis must proceed on the legal acceptance that people have the right to decide for themselves whether or not they will undergo medical treatment. This includes being warned of material risks associated with said treatment. It is apposite to look at several judicial decisions.

I. Rogers v Whitaker

In the Australian High Court case of *Rogers v Whitaker*,[32] the salient facts were as follows: the patient had injured her right eye in a childhood accident and an ophthalmic surgeon advised her that an operation on the eye would not only improve its appearance but would likely also substantially restore sight to it. The operation was not successful, but it was performed with the requisite care and skill. Unfortunately, the patient suffered sympathetic ophthalmia post-operatively and, as a result of inflammation arising from that, lost all sight in the left eye. The patient was rendered almost totally blind.

In Australia, it had been accepted that the standard of care to be observed by a professional person is that of the ordinary skilled person exercising and professing to have that special skill. The question in *Rogers* was whether the observance of that standard of care required information regarding the risk associated with the aftermath of surgery to be given to the patient. The eye surgeon gave evidence that it had not occurred to him to mention sympathetic ophthalmia to the patient.

There was a body of evidence from other medical practitioners to similar effect. However, there was also evidence from other specialists that they would have given a warning to the patient. The state of the evidence may have signalled to the Court that the old rule was unsustainable. In England, the approach to the resolution of similar problems had been determined by a case which lends its name to the *Bolam* rule. The case of *Bolam v Friern Hospital Management Committee*[33] involved a patient who

32 *Rogers v. Whitaker* (1992) 175 CLR 479 (High Court of Australia delivered judgment on 19 November 1992).

33 *Bolam v Friern Hospital Management Committee* (1957) 1 WLR 582.

was injured whilst receiving electroconvulsive therapy ('ECT') without the prior administration of a relaxant drug. Evidence as to the accepted practice varied as between doctors, leading the Court to formulate a rule that has since been stated as follows:

> A doctor is not negligent if he acts in accordance with a practice accepted at the time as proper by a responsible body of medical opinion even though other doctors adopt a different practice. In short, the law imposes the duty of care; but the standard of care is a matter of medical judgment.[34]

It followed from this rule that so long as an acceptable number of medical practitioners adopted the practice in question, that would avail the practitioner a complete defence. It can be observed that the *Bolam* rule is directed to accepted practice in the actual provision of treatment, whereas *Rogers v Whitaker* was concerned with medical advice addressing the risks involved in treatment. In cases decided after *Bolam*, some judges held the view that the rule should only apply in matters involving negligent treatment or surgery, but not where the issue was the sufficiency or adequacy of the advice or information given. In *Rogers v Whitaker* the High Court decided that the rule should be restricted in that way.[35]

In relation to diagnosis and treatment, the Court accepted that the *Bolam* rule would continue to be influential, the reason being that whether a diagnosis or a method of treatment was negligent would depend largely upon medical standards, something known best by doctors. The issue of whether a risk is relevant to a patient, and about which they should be warned, is different. The High Court held that this was a question for the courts themselves. The High Court held:

> *"The law should recognize that a doctor has a duty to warn a patient of a material risk inherent in the proposed treatment; a risk is material if, in the circumstances of the particular case, a reasonable person in the patient's position, if warned of the risk, would be likely to attach significance to it or if the medical practitioner is or should reasonably be aware that the particular patient, if warned of the risk, would be likely to attach significance to it. This duty is subject to the therapeutic privilege."*[36]

---

34 *Sidaway v Governors of Bethlem Royal Hospital* (1985) AC 871, 881 (per Lord Scarman).

35 *Rogers v Whitaker* (1992) 175 CLR 479, 489–490.

36 *Ibid.*

Important to this ruling was the notion that an individual has autonomy and is entitled to make informed decisions about their life. Therefore, a patient must be informed of material risks. The High Court held that it would be reasonable for a person with one functioning eye to be concerned about the possibility of injury to it, particularly in the context of an elective procedure.

In *Rogers*, the Court did not entirely rule out the exercise of therapeutic judgment on the part of a doctor as to what information should be given to certain patients and how it is to be conveyed. The qualification the Court made to the duty owed to patients to give information about risks, was where there was a danger that the provision of all information would harm an unusually nervous, disturbed, or volatile patient.

II. Montgomery v Lanarkshire Health Board

In *Montgomery*,[37] the United Kingdom Supreme Court considered liability in negligence for failure to disclose material risks to patients as part of the process of informed consent. Nadine Montgomery was awarded £5.2 million compensation following birth complications. She was of small stature and had gestational diabetes and had expressed anxieties about vaginal delivery. Her obstetrician failed to warn of shoulder dystocia and her son was born with cerebral palsy. The Court found that had her son been born by elective caesarean section, it is more probable than not that he would have been born uninjured.[38] In a joint judgement, Lords Kerr and Reed (with whom the other justices agreed[39]) distinguished between cases concerning errors in treatment and diagnosis where the test set down in *Bolam* will continue to apply, and cases concerning the disclosure of risk and treatment alternatives which, it was held, are not purely a matter of professional judgement and cannot be decided by reference to a responsible body of medical opinion.[40] *Montgomery* galvanises a position that has been

37 *Montgomery v Lanarkshire Health Board* (2015) SC11 [2015] 1 AC 1430 (Judgment delivered by UK Supreme Court on 11 March 2015).

38 *Ibid*. at para 22.

39 Lady Hale's judgement makes additional observations about the context of childbirth. Unless stated otherwise, references in this article to Montgomery are to Lords Kerr and Reed's joint judgement.

40 Montgomery (2015) UKSC 11, at paragraph 86.

adopted in practice.[41] The Supreme Court declared that Lord Scarman in *Sidaway* had represented substantially the correct position, subject to the *Rogers v Whitaker* 'refinement'.[42]

Setting out a revised test, Lords Kerr and Reed stated:

> *"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, or the doctor is or should reasonably be aware that the particular patient would be likely to attach significance to it."*[43]

What can be distilled from these decisions of superior Courts is that a patient must be warned of material risks, and this is a subjective test. That is, what is significant to the individual patient in a particular case.[44] As noted though, this provision of information must be balanced: "On the one hand, physicians must provide all the information a patient needs to make an informed decision. On the other hand, complex medical information, including all aspects that are somehow relevant to the treatment, would rather prevent informed consent than promoting it".[45]

---

41 *Ibid*. at para 70 citing General Medical Council, Good Medical Practice (2013). See also para 78 citing GMC, Consent: Patients and Doctors Making Decisions Together (2008), para 5.

42 *Ibid*. at para 86–87.

43 *Ibid*. at para 87.

44 The issue of informed consent was looked at by OLG Hamm, judgment of June 18, 2013 – 26 U 85/12. It was held that, "*…the patient should recognize the "type and severity" of the intervention through the informational discussion. To do this, the risks do not have to be presented to him in all conceivable forms, but a "general picture of the severity and direction of the specific risk" ("broadly") is sufficient. However, jurisprudence recognizes that the patient must also be made aware of rare and even extremely rare risks, where these risks, if they materialize, are a heavy burden on the lifestyle and, despite their rarity, are specific to the procedure but surprising to the layperson are"*, available at https://openjur.de/u/645241.html (last access: 20.10.2022).

45 *B. Buchner/M. Freye*, Informed Consent in German Medical Law: Finding the right path between patient autonomy and information overload, 2022 available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4088631 (last access: 20.10.2022).

F. Therapeutic Privilege

There is an 'assumption that the physician cares not only for the patient's physiological health but for his psychological and moral wellbeing.'[46] The duty to have regard to the best interests of the patient and her welfare taken as a whole may clash with the patient's right to choose treatment based on adequate information. To consider provision of information solely in terms of the rights of the patient (and therefore the correlative duty of the health care professional) discounts the 'ethical and social dimension of medical treatments' and may potentially harm the relationship.[47]

Multiple studies that have attempted to determine and quantify the anxiety-generating effect of informed consent provide mixed results about whether a more detailed consent process is physiologically or psychologically harmful to a patient.[48]

Clearly, there are medical situations in which the information involved in planning is of such a nature that the decision-making capacity of a patient is overwhelmed by the sheer complexity or volume of the information they are confronted with. In such cases a patient cannot attain the understanding necessary for informed decision making, and informed consent is therefore not possible.[49]

When faced with these complex clinical contexts, physicians may wonder about the most appropriate ethical conduct. Is it the right time to tell a depressed patient about their cancer? Should they talk about a possible side effect or a risk when it could potentially lead the patient to refuse a medically necessary treatment? Should they discuss a prognosis when they know

---

46 *B. Barber*, Informed Consent in Medical Therapy and Research, New Brunswick 1980.

47 *N. J. Hanna*, Challenging medical decision-making: professional dominance, patient rights or collaborative autonomy?, Oxford Journal of Legal Studies 1998; 143.

48 *D. D.Kerrigan/R. S.Thevasagayam/T. O. Woods/I. Mc Welch/W. E. Thomas/A. J. Shorthouse/A. R. Dennison.*, Who's afraid of informed consent? BMJ 1993, 298; *J. J. Goldberger/J. Kruse/M. A. Parker/A. H. Kadish*, Effect of informed consent on anxiety in patients undergoing diagnostic electrophysiology studies, American Heart Journal 1997, 119; *N. Casap/M. Alterman/G. Sharon/Y. Samuni*, The effect of informed consent on stress levels associated with extraction of impacted mandibular third molars, Journal of Oral and Maxillofacial Surgery 2008, 878; *Z. N. Kain/S. M. Wang/L. A. Caramico/M. Hofstadter/L. C. Mayes*, Parental desire for perioperative information and informed consent: a two-phase study, Anesthesia & Analgesia 1997, 299.

49 *J. Bester/C. M. Cole/E. Kodish*, The limits of informed consent for an overwhelmed patient: clinicians' role in protecting patients and preventing overwhelm, AMA Journal of Ethics 2016, 869.

that it might precipitate an anxiety reaction? Should they withhold certain facts, or present them in a more favourable light? In other words, can they lie to their patients? The physician's judgement is based on clinical context as well as personal and professional values.[50]

In a clinical setting, reporting information without regard for the special conditions of each particular doctor – patient encounter reflects neither the spirit nor the letter of this conception of the truth. An honest medical relationship requires that the doctor consider the limits inherent in the therapeutic process and the circumstances and limitations of each patient. A relationship based on meaningful dialogue permits the doctor to determine these limits and adjust the conversation to the specific capacities and needs of each patient. This process may allow the doctor to consider responsibility and duty, demonstrating the complexity of the role.

Consider the following example: Patient A has been diagnosed with a form of cancer. A sophisticated AI system has recommended a course of drug treatment that is quite particular and unique to Patient A, given all the circumstances and data that is relied upon. It might be aptly described as 'personalised' treatment.

Patient A's prognosis will be much better with the treatment, but there are contraindications as with all chemotherapy. Clinician X knows that this is a treatment regime that will assist but is unable to understand how the AI system arrived at its decision due to the black box effect, let alone explain it in meaningful laypersons terms to the Patient.

What do they do? Save the patient from themselves at risk of compromising their autonomy on the basis of therapeutic privilege? Or admit that they cannot explain the decision adequately and risk not receiving consent for what will be a necessary life-preserving treatment? Time may be of the essence.

In consultations, where everything transpires in a single interview, given the avalanche of information, an informed decision is probably highly improbable for a significant number of patients. Unable to make a personal decision, the patient is often reduced to blindly trusting the doctor, which happens to varying degrees with any expert. Last, to assume that the patient's decision is based firmly on a complete understanding of the issues and without outside influence is neither realistic nor achievable and ultimately significantly underestimates the finitude of human beings.

50 *C. Richard/Y. Lajeunesse/M. T. Lussier*, Therapeutic privilege: between the ethics of lying and the practice of truth, Journal of Medical Ethics 2010, 353.

G. Variables Influencing the Sliding Scale of Capacity to Decide and Consent

As clinical medicine and the involvement of AI is evolving, medical ethics will face the challenge of keeping pace with the development and clinical application of new technologies and therapies. The doctrine of informed consent in specific clinical contexts may need to be looked at through this new prism. Informed consent has become so central and important to the way clinicians practice that there may be situations in which patients' ability to provide informed consent may be compromised or overlooked, particularly where complex information is involved.

I. Patient-related factors

One set of variables is patient related. An obvious case is a patient who is unconscious and lacks capacity to make even the most basic decisions. Others include patients under the influence of alcohol or drugs, young children, and patients with decreased cognitive function. There are also subtle cases where capacity is unclear. Some patients don't have the educational or intellectual ability to understand the choices before them, particularly if the choices are scientifically complex like with AI. Similarly, language and cultural barriers may also impose limits on capacity.

II. Information-related factors

If we are looking at a sliding scale, the more complex, scientifically advanced, and intellectually demanding information becomes, the greater the difficulty for patients to provide consent. Put simply, on one end of the scale is comprehensible, straightforward information about the AI involved in the process and its risks and benefits that is clear and easily understandable. As we move up the sliding scale, the information becomes more voluminous and more complex. If we keep going up the scale, we get to a point where people who ordinarily have capacity to make their own decisions find it all impossible to fully understand. There may be exceptions but for most patients, a full understanding—and truly informed consent—will be impossible.[51]

51 *Bester/Cole/Kodish*, The limits of informed consent (n. 49), 869.

III. Communication-related factors

Dumping an indigestible barrage of complex information about AI on a patient would challenge their understanding. The clinician's capacity to communicate complex information is therefore an important variable that impinges on the capacity to provide consent.

IV. Emotional overwhelm

It is reasonable that a patient may be emotionally overwhelmed by the illness experience and by the implications and complexity of decisions they are faced with. Being emotionally overwhelmed may make informed consent very difficult to achieve. Informed consent may still be possible in this case but is more difficult to attain as the patient's ability to make decisions is compromised.

V. Informational overload

A patient's ability to provide informed consent can easily be overwhelmed by the complexity, uncertainty, or sheer volume of information involved in the decision, as may occur with newer technologies such AI. In short, the information required to provide informed consent is of such complexity, volume, or uncertainty that it makes it impossible for a patient to make an informed choice because the patient is overwhelmed. They're in effect incapacitated for the decision in question.

H. Conclusion

There can be no escaping that AI in healthcare has well-arrived and will continue to exponentially increase. It is trite to say that it is a complex integration. Understanding it poses difficulties to those that design and construct it, let alone the hospitals and clinics that deploy it, the clinicians and others who use it, and the patients who trust and rely upon it.

Where does this leave consent in terms of AI healthcare? The clinical dilemma arises because of the tension between legal constraints and ethical practice. There must be an analysis on a case-by-case basis. Patients will still need to be given a broad overview but not to an intricate technical level.

Clinical validation develops a heightened importance and trust is essential – both in the AI system and in the clinician.

With the development and introduction of AI, arguably there should not be a hardline insistence on obtaining informed consent from patients who are clearly overwhelmed with information due to transparency requirements. Steps should be taken to provide the assistance that patients in each specific situation require. If this necessitates a tailored multi-factorial approach beyond the scope of this paper, then so be it. Too much information may be too much for the wrong patient.