Machine Learning for Diagnosis and Treatment:

Gymnastics for the GDPR

*Robin Pierce**

Machine Learning (ML), a form of artificial intelligence (AI) that produces iterative refinement of outputs without human intervention, is gaining traction in healthcare as a promising way of streamlining diagnosis and treatment and is even being explored as a more efficient alternative to clinical trials. ML is increasingly being identified as an essential tool in the arsenal of Big Data for medicine. ML can process and analyse the data resulting in outputs that can inform treatment and diagnosis. Consequently, ML is likely to occupy a central role in precision medicine, an approach that tailors treatment based on characteristics of individual patients instead of traditional 'average' or one-size-fits-all medicine, potentially optimising outcomes as well as resource allocation. ML falls into a category of data-reliant technologies that have the potential to enhance healthcare in significant ways. However, as such, concerns about data protection and the GDPR may arise as ML assumes a growing role in healthcare, prompting questions about the extent to which the GDPR and related legislation will be able to provide adequate data protection for data subjects. Focusing on issues of transparency, fairness, storage limitation, purpose limitation and data minimisation as well as specific provisions supporting these principles, this article examines the interaction between ML and data protection law.

Keywords: Machine Learning, GDPR, Data Protection, Artificial Intelligence in Medicine, Health Data, Automated Processing, Data Minimisation

I. Introduction

Machine Learning (ML), a form of artificial intelligence (AI) that produces iterative refinement of outputs based on ongoing inputs without human intervention holds significant promise to enhance health. ML is gaining traction as a way of streamlining diagnosis and treatment and is even being explored as a more efficient alternative to clinical trials. With a likely ability to improve precision and efficiency, ML is increasingly viewed as an essential tool in the arsenal of Big Data in medicine. Along with other data-driven technologies, ML can draw on a variety of sources, such as electronic health records, mobile health tracking devices, and other forms of eHealth technologies, and can process massive amounts of data to generate information that can ultimately serve to enhance treatment and diagnosis. As a result, ML can be expected to occupy a central location in the shift toward precision medicine. Harnessed almost inextricably to Big Data, ML falls into a category of data-reliant technologies that, by nature, invoke scrutiny regarding data protection law and other related norms.

This article explores various types of ML in likely or projected healthcare applications in relation to the GDPR. This is followed by brief considerations the implications for data protection when the technology is deployed as part of an integrated system. The inherent nature of ML presents natural challenges for data protection. In the context of health, the processing of sensitive data, how this tension is resolved is of particular significance given the pivotal role of

DOI: 10.21552/edpl/2018/3/11

* Robin Pierce, JD, PhD, Tilburg Institute for Law, Technology and Society, Tilburg University, The Netherlands. For correspondence: <r.l.pierce@tilburguniversity.edu>.

privacy in the successful functioning of healthcare systems.

On 25 May 2018 the GDPR went into effect throughout the European Union (EU).[1] Intended to harmonise privacy protections throughout the EU, give greater control to data subjects, and promote research and innovation while not compromising privacy interests, the GDPR introduces provisions strengthening data subject rights, eg data portability and the right to erasure, yet at the same time, the GDPR relaxes restrictions on research by allowing the re-use of health data for research purposes without explicit consent, provided adequate safeguards are in place, and subject to relevant national laws. This balancing between data protection and the promotion of medical research and innovation and efficient use of data is complex and challenging. This complexity increases as the web of national laws governing healthcare, health data, and medical research intersect with the GDPR in varying ways. Additionally, it is not clear whether the careful balancing act that the GDPR aims to achieve can withstand the vast potential mechanations of machine learning for diagnosis, treatment, and research. The drive to collect ever-increasing amounts of data to feed the promise of more effective treatment and personalised medicine, greater accuracy and efficiency in diagnosis, and faster research results, will need to be reconciled with fundamental rights, including data protection and non-discrimination. ML appears positioned to occupy a central role in delivering this promise and, as such, merits close scrutiny both at the design and program application level and as part of integrated systems. This article examines machine learning programs for health through a data protection lens. ML and Big Data must be viewed as tools not in unqualified pursuit of improvement in healthcare, but ones that must co-exist soundly and predictably with fundamental rights and other important values that are foundational to successful healthcare systems. Because privacy occupies such a critical role in healthcare, it is essential that any uptake of ML be integrated in ways that do not undermine this essential dimension of healthcare.

II. Machine Learning in Medicine

AI in medicine is rapidly gaining traction as a promising means by which to deliver substantial health benefits with greater efficiency and precision. Already showing promise in diagnosis involving pattern recognition of images, ML is also being explored as a way of refining and tailoring therapeutic dosage, informing treatment strategies, refining outcome predictions, enhancing risk assessments, and streamlining research. ML has been defined as 'an artificial intelligence discipline geared toward the technological development of human knowledge that allows computers to handle new situations via analysis, self-training, observation and experience.'[2] ML also refers to a type of data analysis that uses algorithms that learn from data. It is a type of artificial intelligence that provides systems with the ability to learn' without human intervention or being explicitly programmed[3] or without human intervention.

The use of ML in diagnostics is already a reality. One of the earliest applications was for the diagnosis of diabetic retinopathy,[4] a disease associated with diabetes that results in blindness. The standard method has been the laborious examination of MRI images to ascertain scam abnormalities that are associated with the onset of the condition. This is a particularly well-suited clinical target for improved diagnosis in that early detection is associated with significantly better outcomes and, in this case, early intervention could prevent the onset of irreversible blindness. With each patient, it incorporates new images, the ability of the machine to accurately identify patterns pertinent to diagnosis is, in principle, increased. Given the ability to process thousands of scanned images and identify even minute aberrations, the potential of this technology to deliver substantial health benefits is undeniable. However, the health context of ML presents several challenges for data protection not least because health data is regarded as sensitive data, exposing data subjects to

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

2 Prashant Natarajan, John Frenzel and Detlev Smaltz, *Demystifying Big Data and Machine Learning for Healthcare* (CRC Press 2017) 6.

3 ibid 7.

4 Varun Gulshan et al, 'Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs' (2016) 316 Journal of the American Medical Association 22.

greater vulnerability than non-sensitive personal data. This justifies the heightened protections for health data, a terrain that also requires navigation of relevant Member State laws.

III. Health Data

The GDPR defines health data as 'personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status.'[5] Under the GDPR health data continues to be recognized as a 'Special Category' of sensitive data (also including 'racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation').[6] Recognizing that these types of personal data render a person particularly vulnerable in multiple dimensions of life, the GDPR states that the processing of data falling into these special categories 'shall be prohibited'. Among the derogations to this general prohibition of the processing of health data are explicit consent,[7] vital interests of the data subject,[8] and scientific research[9]. Given that the processing of health information is necessary to the provision of healthcare, the GDPR allows for the processing of this otherwise off-limits data 'for the purposes of preventive or occupational medicine, for the assessment of the working capacity of the employee, medical diagnosis, the provision of health or social care or treatment of the management of health or social care systems...'[10] Indeed, the unhindered processing of health data for these purposes seems both reasonable and necessary to an efficient and well-functioning healthcare system. But, how wide this derogation opens the gate for the collection and processing of health-related data depends in no small part on the configuration, application, and integration of specific ML programs. Nevertheless, Member States retain considerable domain over the regulatory aspects of healthcare and, indeed, may provide further restrictions on processing.

This analysis of the data protection implications presented by ML will track principles of the GDPR as well as specific provisions to examine the extent to which certain ML applications can be reconciled with these data protection norms.

Below, several types of ML programs are analysed with regard to their data protection implications.

1. Transparency

Transparency is one of the six principles set forth in Article 5 of the GDPR. Under this provision personal data must be processed in a manner that is 'lawful, fair, and transparent'. Particular applications of ML can invoke scrutiny of all three dimensions of this principle. Moreover, new provisions in the GDPR regarding profiling and automated decision-making have clear relevance for several applications of ML in healthcare. These provisions provide critical support to the promotion of transparency and other principles set forth in Article 5.

ML has received considerable critique regarding its apparent lack of transparency, leading to the moniker 'black box' decision-making. The specific ways that ML raises concerns about transparency is useful to assess the nature and extent of any opaqueness. Concerns about transparency tend to be lumped into general critique of ML. This obscures the nature of what may not be transparent, procedurally or substantively, and whether a specific type of opaqueness should be of concern. Furthermore, it diminishes opportunities for exploration into what options might be possible in order to enhance transparency, and whether mechanisms or practices can be developed that address issues of transparency. Below, I briefly consider specific types of ML algorithms potentially used in health and examine the types of issues they raise.

2. Profiling and Automated Decision-Making

Among the most important provisions supporting the principle of transparency are Article 4(4) on pro-

5 See, ch 2 GDPR.

6 ibid.

7 ibid.

8 See chs 2 and 9, GDPR.

9 See ch 2 GDPR.

10 ibid.

filing and Article 22 of the GDPR pertaining to automated processing. Like several ML applications in healthcare, MBL can be seen as a form of 'profiling'. A goal of MBL is to identify which patients a new patient is most like, essentially aiming to 'profile' the patient. This, in turn, informs treatment and related care. Profiling is defined 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, economic situation, health, personal preferences, interest, reliability, behavior, location or movements.[11]

Working Party 29 (WP 29) discusses the elements of profiling – 1) automated form on processing 2) carried out on personal data where the 3) objective of the profiling is to evaluate personal aspects about a natural person.[12] Furthermore, WP 29 points out that the objective is central in understanding what constitutes profiling and cites the GDPR's reference to profiling as the automated processing of personal data for evaluating personal aspects, in particular to analyse or make predictions about individuals.[13]

Under Article 22(1) of the GDPR, the data subject has the right not to be 'subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her'. The prohibition of sole reliance on automated decision-making is one way to ensure a degree of transparency.

ML for informing treatment dosage and administration in which the algorithm generates determinations or evaluation about a patient could be fairly transparent by virtue of the biological criteria being assessed. Yet, the opportunity to challenge an automated treatment decision cannot practically be challenged and presumably is authorised by explicit consent to the use of automated processing if it is not necessary to the contract for care.

In the case of ML for health, while there is generally not a legal effect, there could be significant effects on the data subject-patient if he is denied treatment based solely on an automated decision, for example pertaining to risk prediction or treatment response that effectively indicates that this patient should not receive Treatment B, for example. This could lead to a substantial worsening of the condition or even death. There are three exceptions to this right. Particularly relevant to the health context are Article 22(2)(b) authorisation of such by an EU or Member State law that sets forth suitable measures to safeguard the rights, freedoms, and legitimate interests of the data subject. This appears to allow Member States to permit decision-making based solely on automated processing if suitable safeguards are in place. These suitable safeguards include the right to be informed about the logic involved and the significance and nature of the likely consequences (Article 13) and a way for the data subject to obtain human intervention, respond, and contest the decision.[14]

It is important to emphasize three points about the implications of these exceptions to the prohibition of decisions based solely on automated processing. First, if national laws permit automated decision-making, the safeguards may be extremely difficult to operationalise meaningfully in the context of ML for health. Physicians may have little understanding of the underlying logic of algorithmic processes, and therefore, be unable to provide much beyond vague and generic explanations. The right to human intervention as a meaningful safeguard is dependent on the nature and extent of that intervention, both of which may be limited by the capabilities, time, and expertise of the physician or other medical personnel. Second, explicit consent as a basis for the exception to the prohibition on sole reliance on automatic processing suffers from two critical considerations in healthcare. First, the inherent vulnerabilities of patients, particularly those suffering from debilitating, serious, or potentially fatal diseases, call into question the voluntary nature of explicit consent. This is above the well-recognized power imbalances inherent in the doctor-patient relationship.[15]

---

11 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Article 4(4)(c).

12 Article 29 Working Party (WP 29), 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' (2017) 6.

13 ibid 7.

14 WP 29, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' (2017) 9.

15 See Jay Katz, The Silent World of Doctor and Patient (Johns Hopkins Press, 2002)

The first exception set forth in Article 22(2)(a), 'necessary to the contract between the data subject and data processor', may appear to facilitate the use of automated decision-making in this situation, but also raise a concern about whether the patient actually has meaningful choice regarding whether to enter into that contract, again raising questions about meaningful alternatives to being subject to decisions based solely on automated processing. This points to the possibility that if there are no reasonable alternatives to automated decisions in a given health context, then the prohibited automated processing is 'necessary to the contract' for the provision of medical treatment. The extent to which this may undermine the protective intent of the provision and the required safeguards is worthy of consideration given that such decisions may ultimately have very significant impacts on one's life.

The possible resolutions seem rather unsatisfying. A patient and her clinical team may settle for generic or vague explanations that permit decisions based on automated processing, or the clinical team can provide nominal input into the decision for the sake of access to useful ML outputs, or the patient and medical team can simply forego the potential benefit of unexplainable outputs generated by automated processing. Again, Member State laws may address these issues and thus establish meaningful parameters, but reliance on this may prove troubling if it leads to uneven protection of data subjects or operates to provide uneven access that could lead to medical tourism. Third, the practicality of the safeguard ensuring the data subject 'the opportunity to respond' or context the decision is debatable. If transparency has been foregone by way of one of the exceptions permitting an automated decision, the ability of the data subject-patient to respond to or contest the decision is likely to be substantially diminished given that the basis for and merits of the decision are not known to the patient. The patient is merely able to express the desire for a different decision. This may be a profoundly inadequate redress for someone who has been denied a specific treatment on a basis that he or she does not understand and which the clinical team cannot explain.

16 Natarajan et al (n 2) 91.

17 HIC.

18 Natarajan et al (n 2) 95.

a. Memory-Based Learning (MBL)

Memory-Based Learning (MBL), one form of ML algorithm, basically compares newly acquired data with data that has been collected previously, with the goal of identifying what the new data is most like based on a subset of attributes. These input attributes can be developed by the program (unsupervised) or use labelled inputs provided by a human (supervised). MBL, an unsupervised type of ML is particularly useful to create cohorts of patients. Subsequent patients are not only compared to relevant cohorts identified by the algorithm but their data is incorporated and used to inform the constitution of those cohorts. These outputs can be used to inform and refine treatment strategies and risk modelling, for example[16].

Concerns about transparency can arise in connection with this type of ML in that it may not be apparent to the data subject-patient what the basis may be for either being placed in a particular cohort or being compared to a particular cohort as the relevant characteristics providing the basis for the grouping may not or cannot be known or explained. While the GDPR does not define transparency, the language used points to the need for the data subject to be able to understand the basis for output that affects them, and decisions made based on that output.

This lack of transparency can affect the use of ML in healthcare in a few different ways – 1) the lack of transparency is largely overlooked for the sake of deriving the benefit either by providing nominal interpretation and perfunctory explanations by medical personnel or 2) healthcare is required to forego the benefits of MBL because of the difficulties of complying with principles of transparency. The extent to which other options can be developed technologically, as a matter of regulation, or in practice (eg some sort of Health Information Counselors[17]) can contribute to addressing these challenges.

b. Recommender Systems

Recommender Systems are also a 'Patients Like Mine'[18] program, a related type of ML. These systems essentially learn what happened with other patients who are like the current patient. This could include, for example, disease profile, co-morbidities, age, and medical history. A clear use for this type of comparative system is to identify which therapeutic inter-

ventions are most likely to be effective for a particular patient given his or her similarities to other people like him or her. An example of this can be found in the Network of Enigmatic Exceptional Responders (NEER) study.[19] The NEER study seeks to understand the behaviour and attributes of exceptional responders to cancer treatments and ultimately, to prescribe and recommend treatment strategies based on the characteristics of patients with exceptionally positive outcomes. This type of recommender system, like MBL, is based on grouping patients based on similarities of attributes and characteristics of persons in the database and making inferences about the current patient. However, unlike MBL, Recommender Systems are supervised systems with inputs labelled by humans. Nevertheless, to the extent that it generates outputs based on similarities to other patients, it constitutes a form of profiling given that it employs automated processing of personal data for the purpose of analysing or making predictions about an individual. Here, the simple prediction that a particular patient will respond to Treatment D in a similar way as patients identified in the database could be a form of automated decision-making to the extent that this 'grouping' is deterministic of which treatment strategy is pursued.

The degree to which the Recommender Systems lack transparency depends, in part, on the degree to which the relevant behaviours and attributes, which serve as the basis for the comparisons are identified and explainable. Moreover, these systems may also involve the ability to weight characteristics as the algorithm learns from both training and new input data. Unless it is possible to ascertain both the relevant characteristics as well as the weighting of attributes involved in an output for a particular patient, it will be very difficult to explain the basis for the recommendation generated for any individual patient beyond the most general probable descriptors regarding disease, stage, or age, for example. In the NEER study, collection of sensitive data is fairly comprehensive, ranging from biometric data to social media networks and physical movement and more. Based on the idea that more information contributes to the robustness of the recommendation produced by the algorithm, this system exemplifies the scenario where data about virtually every aspect of a patient's world can be construed to be relevant to medical care, diagnosis, or treatment. As a result, the possibility of identifying precisely which data contributed to the production of a particular recommendation may even be further obscured by the sheer magnitude of the scope and amount of the data processed by ML.

c. Clustering

Clustering refers to the finding of natural groupings of patients based on an expansive array of data about patients.[20] Such clustering can be very insightful for healthcare personnel. Like the NEER study, finding similarities between patients can inform disease management strategies in addition to risk assessments. Transparency in this kind of clustering, as with previously discussed ML algorithms, can be critical to patients because of the implications that may attend being placed in one group versus another group without understanding or being able to challenge the basis for that grouping or clustering. This lack of transparency could even obscure inadvertent discrimination. For example, an individual who falls into a 'natural grouping' based on a 'prohibited' attribute under discrimination law, eg race, sexual orientation, political or religious affiliation, and these attributes are weighted by the ML algorithm as being highly informative of a relevant outcome, there is the possibility that the ML clustering inadvertently leads to 'unfair' or disparate disease management and possibly produces disparate outcomes on a prohibited basis. Discrimination of this sort does not necessarily lie outside of the scope of the GDPR, as provisions such as those regarding profiling and automated decision-making, while based on protection of personal data in these technological contexts, are intended to also serve as important protective blocks in support of anti-discrimination and other forms of unfair treatment. Both the requirement of transparency and the provisions regarding fully automated decision-making (Article 22) are examples of such provisions. Interestingly, while such characteristics may be noted by a physician and inform a treatment strategy, the attribute may merely be a proxy for other factors that can be verified or rejected by human observation. This also exemplifies why and how human intervention in the decision-making process can be critical to optimal outcomes.

19 People-Powered Medicine, 'NEER Study' <https://peoplepoweredmedicine.org/neer> accessed 31 August 2018 ('NEER Study').

20 Naharajat et al (n 2) 96.

d. Forecasting

Forecasting refers to the ability to forecast what is likely to happen with a particular patient (eg in the next 6-12 months). Also referring to predictive modelling, forecasting can be very beneficial to medical personnel, healthcare systems, as well as patients. This type of ML program might use neural networks that input past values and, based on those past values, predict next values, continuously receiving inputs and generating outputs through multiple layers rather than through a direct input-output correlation. This kind of predictive modelling can inform disease management and is likely to be of great interest for healthcare actors interested in risk prediction[21], eg insurers, providers, hospitals, etc.

The transparency problems stemming from predictive modelling or forecasting are in part related to the fact that this application falls squarely into the category of profiling and, depending on the manner of integration, may raise concerns about automated decision-making. Transparency also depends on the degree to which the inputs and processing of those inputs are known and understood by the clinical personnel. The layering nature of neural networks that underlie the algorithm can obscure the workings of ML prediction modelling, thus making it virtually impossible to track the exact basis of predictions generated by the program. This leaves the patient unable to understand or challenge a prediction that she may see as unfounded, unsubstantiated, or undesirable procedurally or substantively. This lack of transparency may have considerable downstream impacts in the healthcare setting in that the consequences of predictions based on forecasting outputs that has not accommodated a critical aspect of a particular patient could have dire consequences. Indeed, like the previous case, this could happen with a human physician, but the physician is likely to be able to articulate the basis for her decision, which can then be critiqued and contested, if necessary.

Even with the presence of human intervention somewhere in a decision-making process that is primarily based on ML algorithms, lack of transparency persists as a concern in view of the fact that the 'learning' takes place through interacting layers of neural networks and not on a direct input-output basis. While this may generate highly useful outputs that can assist in diagnosis, development of treatment strategies, early detection, and predictions about health outcomes or trajectories, the opaqueness of the process itself challenges the principle of transparency. The prohibition of automated decision-making by the GDPR has been noted by agencies charged with the regulating medical devices and support systems,[22] resulting in tentative guidelines prohibiting solely automated medical decisions. While this furthers the objectives of the GDPR, the lack of clarity about the degree of human intervention necessary can present a problem. Furthermore, any requirement that the data subject be able to receive an intelligible explanation of the basis for a ML output that affects them is likely to present challenges to many medical personnel who are not trained in ML or have a solid grasp of the nature of the processing of patient data that generates these outputs.

3. Purpose Limitation

Purpose Limitation is listed in Article 5 of the GDPR and requires that personal data must be 'collected and specified for explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes', further adding that scientific research shall not be considered incompatible (Article 89(1)). The challenges for AI in medicine regarding purpose limitation have been richly discussed in the literature. Personal data collected in ML processing may be put to almost infinite uses that cannot be explicitly articulated to the data subject at the time of the collection.

MBL, forecasting and clustering all involve some form of assessment based on comparisons with others whether to develop groups or clusters, predict treatment responses or health outcomes. While the personal data taken from a particular patient is done under the auspices of providing medical care to the data subject, when processed by ML this data enters a database that is continually updated to inform future comparisons. While this could conceivably be defensible under a 'pool' or 'collective' care concept, just as a doctor learns continuously from each patient that she sees, with ML the personal data collected from patients become a part of the database.

21 See ibid 92.

22 See eg Zachary Brennan, 'European Commission Offers Guidance on Standalone Software as a Medical Device or IVD' (17 August 2016).

An obvious challenge to purpose limitation is that the input data may be relevant to other uses to which the database may be put beyond that of the initial purpose of care for a particular patient. In this instance, the data may remain in use in ML systems for the purposes of refining treatment strategies for other patients, including those who may suffer from a different disease. Nevertheless, this kind of use is conceivably not incompatible with the initial purpose and may fall within permissible internal use for medical or administrative use. National legislation governing healthcare and health data may ultimately establish the permissible parameters of extended use of sensitive health data, although internal institutional use is generally regarded as permissible if safeguards are in place. Thus, purpose limitation issues for uses of narrow applications of ML algorithms may be relatively limited if the data is not transferred out of the immediate healthcare context. However, when these applications become part of integrated systems, particularly those involving links 'external' to the immediate healthcare setting, the potential for concern may be expected to increase.

ML for analysis of various types of text and documents is showing beneficial possibilities for healthcare, as well. For example, ML programs for Natural Language Processing (NPL) present an interesting challenge to purpose limitation. NLP utilizes different text processing tools to understand the implications of a text. It can, for example, recognize difference in diagnosis given to a single patient. Both Deep Learning (DL) and Text Mining also offer the ability to learn to recognize key features and can be used in concert to understand both text and context leading to, for example, more accurate understanding of conflicting language pertaining to diagnosis.[23] NLP involves databases could be used not only for unspecified purposes but also for uses that are incompatible with the initial use. This is a double dilemma in that while the GDPR seeks to support and facilitate public interest research and Open Source, in doing so may encourage violation of the purpose limitation requirement. NLP can present another type of challenge to purpose limitation in that, for reasons of efficiency, these databases are increasing open source, thus facilitating research. Depending on the governance of the database, and whether the open source is only for software or includes the database may inform purpose limitation regard. However, de-identified aggregated data is likely to be viewed as largely unproblematic.

4. Storage Limitation

The principle of storage limitation requires that data should not be kept in an identifiable form longer than is necessary for the purposes for which the personal data is processed (Article 5(e)). An important exception to this is archiving in the public interest, scientific or historical research (Article 89(1)). WP 29 has specifically addressed this consideration in the Report on Automated Decision-Making, stating that the storing of 'collected personal data for lengthy periods of time for the purpose of building up correlations means that organizations will be able to build up very comprehensive, intimate profiles of individuals'[24].

ML in the various forms explored here may well retain data beyond the time necessary to the purpose for which it was originally collected, but not necessarily in an identifiable form. In those cases where this is true, eg NLP or text mining not involving identification or non-identifiable image recognition, storage limitation is less problematic. It is less problematic rather than unproblematic because, as is widely recognized, the risk of re-identification still exists and increases in integrated systems and in context of Big Data.

For ML programs that function most optimally with maximum possible personal data, like MBL, Recommender Systems, or imaging to ascertain changes and trajectory, the value of data in this identifiable form is precisely what makes it valuable. Both Recommender Systems and MBL strive to identify 'patients like' the patient in front of them. The more information they have about each patient, the more accurate the comparisons. However, preserving this personal data beyond the time period for which it is useful in treating a data subject-patient may technically violate storage limitation.

Forecasting or predictive modelling presents an interesting challenge for storage limitation. Predictive modelling, which takes past values and yields an-

23 Nahajarat et al (n 2) 97.

24 WP 29, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' (2017) 20.

ticipated next values, relies on massive amounts of data over time for optimal accuracy. If data in this database is removed as soon as the purposes of the initial patient intake have been achieved, the ML program continuously loses power and the robustness of the outputs diminish accordingly. Storage limitation is directly at odds with the beneficial application of this technology. Moreover, WP29 points out that this inherent feature of ML may conflict with the proportionality consideration.[25]

IV. Collapsing of Clinical Care and Research

A fundamental challenge of several applications of ML in healthcare is that it collapses clinical care and research activities into one as the collection and retention of that beyond what is necessary for the care of patient M can no longer be said to be under the auspices of medical care for that patient, but rather is more in the nature of research, an activity with its own sets of norms, guidelines, and regulation. The nature of ML is that it uses the data from each new patient to inform future outputs, which can be viewed as a form of health research. Aside from the verifiability problem, to the extent that this is a 'research activity', it bypasses longstanding norms and governing frameworks specifically devised to protect the rights and interests of research participants. This becomes particularly important regarding the re-use of data under the scientific research exemption. Health research must observe EU rules as well as relevant Member State laws. Thus, whatever requirements Member States place on categories of medical research may also pertain to ML in the healthcare setting.[26]

Such provisions may pertain to consent (to research) or the research use of patient data collected in the course of clinical care. An argument could be made that the scientific research exemption should apply to extended storage for the development of medical research database, but this collapsing of clinical and research activity has several implications, not least of which that these activities are governed by a web of parameters based on national laws that could sculpt an uneven landscape for the use of ML in research.

V. ML in Context: Integrated Systems

With a potential substantial link to Big Data and Internet of Things (IoT) in healthcare, data protection issues for ML must also be considered when it operates as part of an integrated system. This is a critical step as it addresses the joint power and potential both for benefit as well as for risk of harm that neither independently may produce. Potential applications of ML for care and diagnostic aid can come in many forms, including.[27]

AI robotics are being specifically designed to identify patterns associated with the onset of various types of mental illness[28] Equipped with a variety of information-gathering mechanisms that detect and relay behavioural changes, these robotics monitor movement and activity, perform geospatial monitoring, aggression indicators (eg cortisol levels), ascertain aspects of speech patterns, pauses, articulation, as well as collect biometric data – weight, consumption, blood pressure, heart rate, etc.

1. Predictive and Behavioural Modelling

Conditions like schizophrenia, dementia, and depression are familiar examples of outbreaks that can result in harm to both the patient and others. The standard health mantra that 'earlier interventions lead to better outcomes' is a guiding beacon in the quest to identify impending health problems at the earliest possible stage. Through ML technology, patterns can be identified that could, in principle, signal incipient mental illness. Changes in patterns of speech, behaviour, gait, movement, and so on have long been known to precede outbreaks of schizophrenia, and other mental health conditions. With the emergence and integration with an ever-growing array of tracking devices, sensors, robotics and other eHealth technologies, the ability to detect pat-

25 ibid 20.

26 See, eg AEGLE Project for survey of country regulations pertaining to health data research <http://www.aegle-uhealth.eu/en/> accessed 2 September 2018.

27 Amy Liang et al, 'A Pilot Randomized Trial of a Companion Robot for People With Dementia Living in the Community' (2017) 18(10) J Am Med Dir Assoc 871-878.

28 Laurel Riek, 'Robotics Technology in Mental Healthcare' in D Luxton (ed), *Artificial Intelligence in Behavioral Health and Mental Health Care* (Elsevier 2015).

terns prior to frank presents a very attractive intervention for many mental health conditions. 'Digital phenotyping', a method of quantifying individual characteristics by analysing data generated from an individual's use of personal digital devices, eg analysing swiping and scrolling movements is being used to correlate with onset of depression.[29] Therapeutic robotics for mental health patients aim to detect and alert subtle symptoms, phenomena, and biological or behavioural changes based on algorithmically derived predictive indicators.[30] This could include, for example, information on speech patterns, social interactions, gaze characteristics[31] in addition to biological phenomenon such as heart rate and blood pressure. In fact, in a data-driven approach to disease detection and treatment, little stands outside of the rather vague parameters of 'personal data related to the physical or mental health of a natural person' that is informative of health status.

ML is now being engaged to collect virtually every type of personal data to enhance the accuracy and efficiency of early detection of Alzheimer's disease[32], a disease notorious for the breadth and scope of personal information potentially relevant to a diagnosis.

2. Data Minimisation

Machine learning for diagnosis and treatment has widely been acknowledged as presenting an inherent tension with data protection.[33] In the context of ever-increasing types of data being collected via IoT, mobile health, various forms of eHealth, and so forth, the ability of ML to process and analyse this data identifying patterns and correlations with health presents substantial challenges to the force and operationalisation of the principle of data minimisation the collection and processing of virtually any type of data can be justified as being 'relevant to health status'. In this way, such an expanded scope of data may bypass the protections aiming to reduce vulnerability that the principle of data minimization aims to provide. Recital 39, Article 5(c) of the GDPR states that

> The personal data should be adequate, relevant and limited to what is necessary for the purposes for which they are processed. This requires, in particular, ensuring that the period for which the personal data are stored is limited to a strict minimum. Personal data should be processed only if the purpose of the processing could not reasonably be fulfilled by other means.[34]

3. Scientific Research Exemption and Machine Learning

The second potential challenge to the ability of the GDPR to provide adequate protection of the rights and interests of data subjects in the integrated context of ML for health is the relaxed restrictions on the re-use of data for scientific research.[35] The pertinent recitals in the GDPR state that further processing of lawfully collected personal data is allowed for scientific research in the public interest without explicit consent if certain conditions are met. Critically, this requires that adequate safeguards be in place and that this derogation is subject to national laws. That is, where a Member State requires explicit consent for the use of sensitive data for re-use in research, regardless of the presence of safeguards, an organisation cannot rely on the relaxed restrictions of the GDPR to conduct that research without consent. The AEGLE project has already yielded publications making clear the extent of the variation across Member States in the degree and manner in which the scientific research exemption can be deployed. While this provides heightened protections in some cases, these protections are uneven across the EU.

The data protection implications of ML in health extending solely within the context of the health sphere is challenging enough, but ML can and does

29 Thomas Insel, 'Digital Phenotyping: Technology for a New Science of Behavior' (2017) 318 Journal of American Medical Association 13.

30 Marcello Ienca et al, 'Intelligent Assistive Technology for Alzheimer's Disease and Other Dementias: A Systematic Review' (2017) 56 Journal of Alzheimer's Disease 1301–1340.

31 ibid.

32 Even without a cure, there are good arguments to pursue early detection in that the pathology associated with AD is understood to begin 10-20 years before symptoms manifest and some theorise that the reason that no cure has been found is because the interventions are introduced too late in the disease progression. Consequently, the quest for early detection has consumed much of AD research in the past 15 years and has witnessed an increase in the number of biomarkers and means to detect AD.

33 Menno Mostert et al, 'Big Data In Medical Research And EU Data Protection Law: Challenges To The Consent Or Anonymise Approach' (2015) European Journal of Human Genetics 24.

34 GDPR, art 5(1)(c), recital 39.

35 GDPR, art 6(4).

involve external processors, and other IT services, particularly as part of an integrated system. However, an important phenomenon has emerged in the partnering of healthcare institutions and commercial enterprises for the provision of infrastructural services. In what became a high-profile arrangement, Royal Free London NHS Foundation Trust entered into a deal with Google DeepMind between July 2015 and October 2016. This arrangement involved the transfer of identifiable patient records from the NHS without explicit consent for the purpose of developing an app to provide clinical alerts for kidney injury.[36] It turned out that the data of all patients within the NHS had been received by Google DeepMind and not just for patients with kidney problems. The argument was made that this would enable research into preventive measures. In the research context, valuable research is being conducted using ML to identify patterns and correlations that could ultimately contribute to enhanced healthcare delivery in treatment, care, and drug development. For example, the NEER study, based at the Department of Bioinformatics at Harvard Medical School in the US, is partnered with Amazon and research participants receive confirmation emails from Amazon, an organisation with whom the researchers acknowledge prior partnership.[37] The NEER study aims to better understand those cancer patients who far exceed initial prognoses and expectations, with the goal of using what is learned about these exceptional responders to improve treatment, care, and counsel to cancer patients to affect better outcomes for them. Nevertheless, a partnership with Amazon, or any other commercial enterprise, raises serious questions in the context of the data protection rights and interests of data subjects, particularly in the collection and processing of sensitive data being processed for re-use without explicit consent. While Cambridge Analytica[38] appears to be a worst case scenario, the partnering of health institutions legitimately processing data for the permissible purposes with commercial institutions complicates the risk scenario and, possibly, the nature of the risk.

VI. Conclusion

As ML increasingly slows the ability to enhance healthcare through the greater efficiency and precision in diagnosis and development of treatment strategies, this data-driven technology will and should come under scrutiny for novel ways of processing personal data. The very nature of ML is such that it will result in a lack of transparency, which can have downstream effects that negatively affect other rights, eg non-discrimination. Additionally, purpose limitation, storage limitation, and provisions regarding automated processing and profiling, if applied strictly, could significantly hinder the ability of ML to deliver on its enormous promise in medicine. The extent to which the GDPR potentially stands as a gatekeeper to medical innovation is partly contingent on how these principles and provisions are operationalised. This points to a key role for data controllers – in the assessment of risk, the provision of adequate safeguards, and the responsible and judicious use of health data. This places the data controller in a critical position of not only ensuring compliance with data protection laws, but ensuring that the trust in medical privacy, which is pivotal to healthcare, is not put in jeopardy.[39] Without the trust that assurances of privacy inspire, a healthcare system can fail both its patients and the society it serves. The extent to which the ML is able to navigate relevant principles and provisions and still deliver on its potential of making major contributions to treatment, diagnosis, and research may well require that ML engage in the gymnastics necessary to ensure not only data protection compliance, the maintenance of medical privacy such that trust in the system is not eroded.

36 Julia Powles and Hal Hodson, 'Google DeepMind and healthcare in an age of algorithms' (2017) 7(4) Health Technology (Berl) 351-367.

37 NEER Study (n 19).

38 ibid.

39 Robin Pierce, 'Medical Privacy: Where Deontology and Consequentialism Meet' in Bart van der Sloot et al (eds), *The Handbook of Privacy Studies* (APC 2018).