phg foundation
making science work for health

Algorithms as medical devices

UNIVERSITY OF CAMBRIDGE

Authors

Johan Ordish, Hannah Murfet and Alison Hall

Acknowledgements

This report was informed by the workshop and report *Regulating algorithms in healthcare: the GDPR and IVDR in practice*. The PHG Foundation is extremely grateful to the expert contributions of all participants. We further acknowledge the expertise of Julian Hitchcock (Bristows LLP) and Sara Payne (PHG Foundation Associate) for reviewing drafts of this report.

Hannah Murfet (PHG Foundation Fellow) contributed to this report by way of in-kind support from Microsoft Research Ltd. Any opinions expressed are the author's own, and may not represent the view of Microsoft.

URLs in this report were correct as of October 2019

This report can be downloaded from:
www.phgfoundation.org

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

+44 (0)1223 761900

Sep 2019

Correspondence to:
intelligence@phgfoundation.org

How to reference this report:
Algorithms as medical devices
PHG Foundation (2019)

ISBN 978-1-907198-33-5

Contents

Executive summary

Algorithms as medical devices considers the position of digital health under medical device regulation and the challenges the technology poses for the regulatory framework.

This report is divided into three sections, each addressing a distinct set of issues. Section 1 describes the challenges that the digital health sector might pose for regulators and developers, Section 2 how digital health devices might be regulated as medical devices under EU and US law, and Section 3 addresses the specific challenges that machine learning might pose for medical device regulation due to the black box nature of many machine learning medical devices, especially those that constantly retrain. Overall, this report offers an accessible resource for developers, policy makers and researchers.

Our conclusions are that healthcare is changing; medical device regulation must change with it. This report considers three interrelated challenges this evolution poses.

Challenge 1: Digital health and medical device regulation

- The medical device market is shifting. The dramatic expansion of digital health means that the quantity and variety of digital health devices grows daily. This market shift poses a threat in terms of sheer numbers: regulators may soon be swamped by a tidal wave of digital health devices. Further, because digital health devices differ from traditional medical devices the skillset to test and properly analyse these devices requires different skills and expertise.

- The growth in algorithms and software medical devices means that the nature of the medical devices sector is changing, with more developers being exposed to medical device regulation, often for the first time, without the institutional support and benefits of scale that manufacturers of more traditional medical devices might typically have.

Given these broad conclusions, we recommend the following:

1. Bodies such as the Medicines and Healthcare products Regulatory Agency (MHRA) and notified bodies should ensure that they possess sufficient expertise to assess this new generation of digital health devices, including machine learning.

2. Regulators like the MHRA should ensure that market actors that may be caught by medical device regulation for the first time are aware of their obligations under the Medical Devices Regulation (MDR) and In Vitro Diagnostic Medical Devices Regulation (IVDR).

Challenge 2: Digital health and device qualification

- The line between what qualifies as a medical device and what constitutes a life-style or wellbeing device is increasingly blurred. The availability of highly accurate sensors outside the clinic means that 'consumer devices' and 'medical devices' are no longer mutually exclusive. In this context, the law is also changing. What qualifies as a medical device must be sufficiently flexible for regulators to regulate risky devices but also rigid enough for manufacturers to be given some degree of certainty over whether their device will qualify as a medical device or not.

Given these broad conclusions, we recommend the following:

1. That guidance akin to MEDDEV be issued that clarifies the issue of device qualification under the MDR and IVDR.
2. That the European Commission should continue to update and expand upon its *Manual on Borderline and Classification in the Community Regulatory Framework for Medical Devices* under the Regulations – this practical resource is useful for manufacturers to understand how their device might be regulated.

Challenge 3: Machine learning as a medical device

- Machine learning is set to change healthcare as the number of uses increase across the sector. However, machine learning is not necessarily qualitatively different from its manually programmed counterparts.

- Not all machine learning models are black boxes (human uninterpretable). Some models are relatively transparent, their decisions being able to be represented with visualisations. Other models, while opaque, may be explained by constructing model-agnostic explainers or by testing particular instances with example-based explanations.

- Not all machine learning models will constitute moving targets for regulators. Not all machine models retrain, nor do all machine learning models incorporate streaming data. However, those models that do constantly retrain may pose safety concerns as some machine models can be fragile, that is, small changes in data or the model may cause dramatically different outputs. In this regard, National Institute for Health and Care Excellence (NICE) *Evidence standards framework for digital health technologies* distinguishes between 'fixed algorithms' on one hand and 'adaptive algorithms' on the other.

- While human interpretability and explanation are key issues, no current standard directly addresses them – this is despite the fact that human interpretability may be relevant when assessing the safety, effectiveness and risk of some devices.

- Machine learning models that constantly retrain do not fit well with current medical device regulation. The MDR/IVDR and harmonised standards, while envisioning change in devices, do not envision the kind of dynamic change that some machine learning models represent.

- There are other regulatory strategies that might assist in dealing with the specific problems that machine learning devices might pose. For instance, the proposed US Food and Drug Administration (FDA) AI Framework, FDA Pre-Certification Programme, concepts from the NICE Evidence standards framework for digital health technologies, and regulatory sandboxes might provide lessons for medical device regulation.

Given these broad conclusions, we recommend the following issues should be considered:

1. The extent to which the black box problem should be addressed by harmonised standards at an EU and national level
2. Whether it is helpful to distinguish in harmonised standards and supplementary guidance like MEDDEVs between those machine learning medical devices that retrain and those that are static, at an EU and national level
3. Whether harmonising standards like EN 82304 for Health Software might capture risk factors (i.e. security concerns) not currently captured by other standards, at an EU and national level
4. Whether a programme akin to the FDA's Pre-Certification Programme and AI Framework might work within the EU context
5. How the NHS Digital and MHRA joint project on synthetic devices might address the black box and dynamic devices problem: this project seems to be an ideal opportunity for testing what a balanced regulatory strategy might look like.

Overall, we urge regulators to first utilise the regulatory tools that already exist in medical device regulation before imposing new systems of regulation on machine learning. Further, we stress that machine learning is a diverse set of tools and does not always represent a novel challenge to our current regulatory framework. Consequently, we urge caution when regulating machine learning, it would be unwise to regulate the entire field according to an exceptional subset of machine learning tools.

1. Digital health and medical device regulation

The practice of medicine is changing as healthcare is increasingly planned and delivered through digital technologies. This section explores how this new era of health will be regulated and the challenges this presents for medical device law.

Digital medicine is 'the transformation of health care through the use of computer technology in the creation and application of medical knowledge.'[1] This transformation is accelerating as the fruits of digital health including 'mobile health, health information technology, wearable devices, telehealth, telemedicine, and personalised medicine'[2] come to bear with many of these devices increasing in both number and variety (Table 1).[3] In short, the practice of medicine is changing with the diagnosis and treatment of patients being increasingly tempered by digital health technology. With these market shifts, concerns regarding the quality, reliability, and safety of this new generation of devices have been raised. [4–6]

Table 1: Examples of digital health devices

Manufacturer/Developer	Device/App	Function of device
Natural Cycles	Natural Cycles App	Contraception app using body temperature readings and menstrual cycle information to prevent pregnancy[7]
Viz.ai	Viz LVO	AI to automatically identify suspected large vessel occlusion strokes and notify specialists directly[8]
AliveCor	KardiaMobile, KardiaBand	Medical-grade EKG to detect normal heart rhythm or AFib[9]
Tricella	Pillbox	Smart pillbox that monitors medication compliance[10]
VitalConnect	VitalPatch	Monitoring of heart rate, respiratory rate, skin temperature, posture, fall detection, and activity[11]
Babylon	GP at Hand	Virtual GP appointments and other related services[12]

Often these devices are regulated end to end: from the data that trains the model, to requirements for CE marking of the device itself, to statute and case law that determines who is liable if something goes wrong. Some of this regulation is extremely broad in scope. For instance, the GDPR is broad in terms of its territorial application but also the subject matter that it regulates. The definition of 'processing'[13] includes most operations that one could want to perform on data, and 'personal data'[14] includes most data that identifies individuals. Other types of regulation are tighter in scope, applying to a subset of digital health devices but also imposing comparatively burdensome requirements on developers. Arguably, medical device regulation fits this latter description.

1.1 What is medical device regulation?

EU medical device regulation is in a state of transition and comprises of two sets of EU law: an outgoing set of Directives and an incoming set of Regulations. First, the outgoing set of Directives: Directive 93/42 on medical devices,[15] Directive 98/79 on in vitro diagnostic medical devices[16] and Directive 90/385 on active implantable medical devices.[17] Second, the incoming set of Regulations collapses the previous three Directives into two Regulations: Regulation 2017/745 on medical devices (MDR)[18] and Regulation 2017/746 on in vitro diagnostic medical devices (IVDR).[19]

Together the Regulations represent the future of EU regulation of medical devices. The MDR and IVDR are both regulations, meaning that they have been directly applicable law in Member State law since their adoption and in force since 25 May 2017. Despite this, both Regulations have a transition period, the MDR fully replacing its predecessor Directives on 26 May 2020 and the IVDR doing the same in 2022.[20 i] The focus of this document is on the incoming MDR and IVDR.

The primary function of the MDR/IVDR is to ensure the harmonization of the single market by providing uniform standards for the quality and safety of medical devices.[21, ii] The regulation of medical devices should be sharply distinguished from the regulation of medicines and services, medicines being regulated by the European Medicines Agency and services being subject to oversight by the Care Quality Commission. Uniform standards are achieved by imposing CE marking requirements on manufacturers (or in this case, developers) of medical devices to a) bring their devices to market and b) keep their devices on the market. Broadly, the structure of this system can be broken down into three parts.

1. The Regulations set out what qualifies as a medical and in vitro diagnostic medical device, that is, what counts as a medical or in vitro diagnostic device – what is regulated?[22]
2. The Regulations classify these devices according to their risk profiles – how are they regulated?[23]
3. According to risk classification, 'conformity assessment procedures' may be required: self-certification, notified body review, demonstration of clinical evidence, post-market performance, and so on – what is required?

Apart from these purposes, the MDR/IVDR also outline the responsibilities of competent authorities (the MHRA in the UK) and the notified bodies under their supervision.[24] In brief, 'notified bodies' carry out the conformity assessment checks, while 'competent authorities' provide supplementary advice while also ensuring notified bodies remain objective and perform their delegated duties to a high standard. Finally, while the MDR/IVDR capture and regulate devices that qualify as a medical or in vitro diagnostic device, there are broad exemptions from many requirements. For instance, the health institution exemption, exempts certain laboratory developed tests from many MDR/IVDR requirements.[25–26] Hence, qualification as a medical device (whether a device is regulated as a medical or in vitro diagnostic device) does not necessarily mean that the device must comply with all the rigours of conformity assessment procedures.

Medical device regulation also includes a multitude of harmonised standards. These are a means for manufacturers to demonstrate conformity with the general safety and performance requirements as well as the quality assurance and risk management requirements.[27] Following *James Elliott Construction v Irish Asphalt*, harmonised standards are a part of EU law so long as they are adopted by the appropriate procedure – in this case harmonisation.[28] Generally, harmonised standards set out processes that may assist in meeting many of the broader requirements of the MDR and IVDR.[29]

1.2 Where does software fit into the MDR/IVDR?

The Regulations mention both 'algorithms'[30] and 'mobile computing platforms'[31] by name but the primary term used to regulate digital devices is 'software.' Software is not defined in the MDR or IVDR. However, past guidance has defined 'software' as 'a set of instructions that processes input data and creates output data.'[32] Software, whether as a component in a wider medical device or in its own right, can qualify as a medical device.[32] For example, software used to control blood glucose monitors would have software as a component in the wider blood glucose device, whereas risk prediction application might constitute a medical device in and of itself. Both the MDR and IVDR explicitly reference software in their definition of medical device[34] and in vitro diagnostic medical device.[35] This represents a subtle shift from the definition under the previous set of Directives. Under the amended Directives, software was included in the definition of medical device but included extra clarification: 'software intended by its manufacturer to be used specifically for diagnostic and/or therapeutic purposes and necessary for its proper application.'[36]

Supplementary guidance to the Directives MEDDEV 2.1/6 provided specific guidance on the qualification and classification of standalone software. Some of this guidance has been included in the new MDR/IVDR. A notable difference is that the MDR/IVDR no longer speaks in terms of 'standalone software.'[37] If software has a medical purpose, then it is within the scope of the MDR/IVDR, if not, then it falls outside the scope of the MDR/IVDR. However, the classification rules in Annex VIII do distinguish between 'software that drives or influence the use of a device' and software 'independent of any other device.'[38]

1.3 Why does medical device regulation matter?

The MDR/IVDR represents a barrier to entry for devices entering the medical device market. The proliferation of digital health poses a number of challenges for the MDR/IVDR, its regulators, and manufacturers/developers. Our analysis suggests a number of possible challenges:

- The line between what is and is not a medical device is blurred, making it difficult to know the scope of the regulatory burden. As more digital health devices come onto the market, the number of borderline devices (devices where it is unclear whether the device is or is not a regulated device) may increase
- Even with a clear and restricted regulatory scope, a modest proportion of the swelling number of digital health devices threatens to create a regulatory tidal wave[39]

- Compliance with medical device regulation is a specialised area of knowledge with a limited pool of qualified professionals. Proper assessment of the safety and effectiveness of digital health devices may also require different tools, methods, and standards. As a consequence, skills to properly assess digital health devices as a medical device may be even more specialised

- Digital health has also brought new market participants into the medical devices sphere.[40] Many digital health developers may encounter medical device regulation for the first time and so be unaware or ill-prepared to comply with regulatory requirements. In this way, some developers may find it difficult to compete with established medical device manufacturers

- The medical device market was once primarily a business-to-business market, devices being sold for use by clinicians.[41] However, digital health may begin to turn the market on its head, most apps being marketed for direct to consumer use. This development is of special concern as devices for self-testing are often subject (rightly) to further requirements in relation to labelling and testing[42]

The significance of many of these challenges will be contingent upon what qualifies as a medical device. The next section seeks to explore the US and EU methods to determine device qualification, keeping in mind the above digital health challenges.

2. Device qualification and digital health

This section considers the position of software as a medical and in vitro diagnostic medical device in both the EU and the US. The boundary issue between regulated medical devices and lifestyle or general purposes is evaluated, focusing on two specific claims. First, that EU and US methods of determining device qualification differ (whether a device is regulated as a medical or in vitro diagnostic device); second, that the US method is more flexible, being better able to capture risky digital health devices. Although both claims have their supporters, we argue that the EU and US systems offer similar scope for regulators to capture risky digital health devices,[43] thereby offering similar levels of protection and reassurance to users.

2.1 EU position on qualification

The scope of the MDR and IVDR is primarily determined by whether a device qualifies as a medical or in vitro diagnostic device. If a device qualifies as either, so long as the device is not health institution exempt, the device will likely be subject to pre-market certification and post-market surveillance procedures.

The primary way in which a device might qualify as a medical or in vitro diagnostic device is via Article 2(1) MDR or Article 2(2) IVDR (see Table 2 below). That is, the manufacturer intends their device to be used for one of the medical purposes specified in either definition. A full list of how the MDR/IVDR captures devices, parts of devices, and accessories can be found in Appendix 1. In this report, the focus is on the intended purpose route. There are two elements to this route: what counts as 'intended purpose' and what counts as 'specific medical purpose.' We explain each in turn.

Intended purpose

The definitions of medical and in vitro diagnostic devices both include the concept of 'intended purpose'. This concept makes plain that it is not enough that a device possesses the features listed in either definition; rather the manufacturer must also intend their device be used for one of these medical purposes. This concept of intended purpose is one shared by both US and EU jurisdictions.[43] Similarly, the interpretation of the concept is contentious on both sides of the Atlantic, the FDA seeking (but failing) to elucidate the meaning of 'intended purpose'[44-45] and the European Court of Justice (ECJ) attempting clarification on at least two occasions.[46–47]

Article 2(12) of the MDR/IVDR provides the key clarification of what 'intended purpose' means, the MDR version stating:

> '"intended purpose" means the use for which a device is intended according to the data supplied by the manufacturer on the label, in the instructions for use or in promotional or sales materials or statements or as specified by the manufacturer in the clinical evaluation;'

The MDR also recognises that a device may have multiple intended purposes, some of which will be non-medical. In this case, so long as there is one intended medical purpose, the device will qualify as a medical device and so be within the scope of the Regulation.[48] The next question then is what counts as a 'specific medical purpose'?

'Specific medical purpose'

If a device is to qualify as either a medical device or IVD, the device must have a 'specific medical purpose.' There are three points to note in regards to this specific medical purpose provision.

First, while the definition of medical device and IVD differ, the definition of IVD is predicated on the definition of medical device. That is, the Article 2(2) IVDR definition starts by noting that IVD 'means any medical device,' this definition being the Article 2(1) MDR definition of medical device (see Table 2 below).[49] This is important to note because what counts as a 'specific medical purpose' is then primarily determined by reference to the purposes listed under the definition of medical device in Article 2(1) MDR. In short, when looking for what a specific medical purpose is, we should look to the definition of medical device.

Second is consideration of how 'medical purpose' was interpreted under the Directives. Disputes over the medical purpose of products have arisen under the Directives. One notable dispute is *Brain Products v BioSemi*.[50] *Brain Products* – a preliminary ruling by the ECJ – concerned the question of whether a system that enabled the recording of human brain activity qualified as a medical device. The ECJ's reply was forcible, following closely the MEDDEV 2. 1/1 guidance on the definition of 'medical devices': manufacturer's intended purpose must be specifically medical - use in a medical context is not enough for a device to qualify as a medical device.[51] As the ECJ reasoned, many 'sports goods' measure the function of certain organs but for performance purposes not medical purposes, so these products should not qualify as medical devices. In short, *Brain Products* stands for the proposition that context is insufficient but indispensable when construing medical purpose.

Third, the definition of 'medical device' is subtly different between the Medical Device Directive and the MDR.[iii] This was changed with the amendment to the Directive with Directive 2007/47 (see Table 2). This amendment brought the Directive in line with the definition found in MEDDEV 2.1/1. The amended definition now reading: 'to be used specifically for diagnostic and/or therapeutic purposes and necessary for its proper application...' It is unclear whether the streamlined definition in the MDR that now includes explicit mention of 'specific medical purposes' change the legal meaning. We suggest that the European Commission update their MEDDEV 2.1/1 guidance for the MDR as soon as possible.

Table 2: EU definitions of medical and *in vitro* diagnostic devices

Medical Device Directive 93/42 (as amended by Directive 2007/47/EC, Article 2(1)(a)(i)) Article 1(2)(a)	"medical device" means any instrument, apparatus, appliance, software, material or other article, whether used alone or in combination, including the software intended by its manufacturer to be used specifically for diagnostic and/or therapeutic purposes and necessary for its proper application, intended by the manufacturer to be used for human beings for the purpose of: ■ diagnosis, prevention, monitoring, treatment or alleviation of disease ■ diagnosis, monitoring, treatment, alleviation of or compensation for an injury or handicap ■ investigation, replacement or modification of the anatomy or of a physiological process ■ control of conception and which does not achieve its principal intended action in or on the human body by pharmacological, immunological or metabolic means, but which may be assisted in its function by such means;'
Medical Device Regulation (EU) 2017/745 Article 2(1)	' "medical device" means any instrument, apparatus, appliance, software, implant, reagent, material or other article intended by the manufacturer to be used, alone or in combination, for human beings for one or more of the following specific medical purposes: ■ diagnosis, prevention, monitoring, prediction, prognosis, treatment or alleviation of disease ■ diagnosis, monitoring, treatment, alleviation of, or compensation for, an injury or disability ■ investigation, replacement or modification of the anatomy or of a physiological or pathological process or state ■ providing information by means of in vitro examination of specimens derived from the human body, including organ, blood and tissue donations and which does not achieve its principal intended action by pharmacological, immunological or metabolic means, in or on the human body, but which may be assisted in its function by such means…'
In Vitro Diagnostic Device Regulation (EU) 2017/746 Article 2(2)	' "*in vitro* diagnostic medical device" means any medical device which is a reagent, reagent product, calibrator, control material, kit, instrument, apparatus, piece of equipment, software or system, whether used alone or in combination, intended by the manufacturer to be used in vitro for the examination of specimens, including blood and tissue donations, derived from the human body, solely or principally for the purpose of providing information on one or more of the following: (a) concerning a physiological or pathological process or state (b) concerning congenital physical or mental impairments (c) concerning the predisposition to a medical condition or a disease (d) to determine the safety and compatibility with potential recipients (e) to predict treatment response or reactions (f) to define or monitoring therapeutic measures…'

Wellbeing and general purpose devices

Both the MDR and IVDR contain specific clarification of whether software qualifies as a medical or in vitro diagnostic device. Moreover, the European Commission's *Manual on Borderline and Classification in the Community Regulatory Framework for Medical Devices* provides specific examples of software and mobile applications that would and would not be likely to qualify as a medical device.[52] In regards to the Regulations, Recital 19 MDR and Recital 17 IVDR note:

'...software in its own right, when specifically intended by the manufacturer to be used for one or more of the medical purposes set out in the definition of a medical device [an *in vitro* medical device], qualifies as a medical device [an *in vitro* medical device], while software for general purposes, even when used in a healthcare setting, or software intended for life-style and well-being purposes is not a medical device [an *in vitro* medical device]...'

These recitals therefore distinguish between three kinds of software:

1. Software that qualifies as a medical device or in vitro diagnostic device
2. Software for general purposes (and so not a medical or in vitro diagnostic device)
3. Software for life-style or well-being purposes (and so not a medical or in vitro diagnostic device)

While the recitals treat these distinctions as categorical – a device is either a medical device or a well-being device – in reality, there is no sharp line between what counts as a medical device and what counts as a well-being device (the same being true of devices for general purposes).

The MEDDEV 2.1/6 guidance further clarifies how to draw the line between software as a medical device and software for general purposes. While this MEDDEV guidance relates to the previous set of directives, and so its authority under the regulations is in doubt, it may be persuasive. MEDDEV 2.1/6 clarifies:

'if the software does not perform an action on data, or performs an action limited to storage, archival, communication, 'simple search' or lossless compression… it is not a medical device.'

The guidance goes on to state that software intended to create, modify, embellish, or alter the representation of data or medical information for a medical purpose could be a medical device.[53] Additionally, MEDDEV 2.1/6 also gives a number of examples of software not considered to be for the benefit of individual patients: aggregations of population data, software for 'generic diagnostic or treatment pathways', models and templates for epidemiologic studies or registers.[54]

To summarise, the general method of determining intended purpose and medical purpose in combination with MEDDEV 2.1/6 and Recitals 19 and 17 helps us determine what software might qualify as a medical device/IVD.

SNITEM

MEDDEV guidance is just that: guidance. Despite this, the guidance is widely cited and, some argue, has been partially codified by the recent C-329/16 *SNITEM* (*Syndicate National de l'Industrie des Technologies Médicales*) ruling.[55] The *SNITEM* preliminary ruling concerned the qualification of drug prescription assistance software as a medical device.[56] The judgment cites multiple parts of the MEDDEV 2.1/6 guidance to reach its verdict. In particular, the guidance references the software qualification flowchart as well as other sections unrelated to qualification.[57, iv] Accordingly, *SNITEM* has been seen as endorsement of the MEDDEV 2.1/6 process to determine standalone software qualification as a medical device.

For many, *SNITEM* was mostly a restating of the orthodox method to determine device qualification.[58] The orthodox position being that the manufacturer's intended purpose is the intended purpose relevant when considering medical device qualification. To determine such purposes, it is necessary to look to promotional and instructional material listed in Article 2(12) MDR/IVDR. In other words, national authorities may not substitute their own understanding of intended purpose to ensure device qualification but must look to the evidence listed in Article 2(12) to determine intended purpose.

Although the *SNITEM* ruling came to the expected conclusion citing predictable law, the method and emphasis of the Court's approach was surprising. While the Court relied upon previously cited law, the ruling demonstrates a lack of deference toward manufacturers' intended purpose. Arguably, the judgment constitutes a sea change toward increasing reliance on the function of a device to determine device qualification. Evidence for this is found in the nub of the judgment:

'I[t] follows that software, of which at least one of the functions makes it possible to use patient-specific data for the purposes, inter alia, of detecting contraindications, drug interactions and excessive doses, is, in respect of that function, a medical device…'[59]

From the usage of 'follows' and the failure to defer to the evidence referred to in the definition of 'intended purpose' it seems clear that function is a key part of determining device qualification.[iv] *SNITEM* seems to indicate that the ECJ will deal with intended purpose rather briskly, using the function of the device to establish medical purpose and so device qualification. Rather than bolstering the manufacturers' role in determining intended purpose, *SNITEM* suggests that function, not intended purpose is king when determining device qualification.

2.2 US position on qualification

Like the EU, the US method to determine device qualification utilises statutory definition and the rider of 'intended use' (see Table 3). The US method has two main similarities to the EU method:

- First, the definitions of medical device and in vitro diagnostic product are similar to equivalents found in the EU MDR and IVDR, as well as the International Medical Device Regulators Forum's guidance *Software as a Medical Device* (SaMD): Key Definitions (see Table 3)[60]
- Second, as in the EU, intended use is not the only route by which a device might qualify as a medical device[vi]

Table 3: US and IMDRF definition of the terms 'device', 'in vitro diagnostic product' and 'medical device'

Federal Food, Drug, and Cosmetic Act 2018 Section 201(h)	The term "device" (except when used in paragraph (n) of this section and in sections 301(i), 403(f), 502(c), and 602(c)) means an instrument, apparatus, implement, machine, contrivance, implant, in vitro reagent, or other similar or related article, including any component, part, or accessory, which is - […] (2) intended for use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment, or prevention of disease, in man or other animals, or (3) intended to affect the structure or any function of the body of man or other animals, and which does not achieve its primary intended purposes through chemical action within or on the body of man or other animals and which is not dependent upon being metabolized for the achievement of its primary intended purposes. The term "device" does not include software functions excluded pursuant to section 520(o).
Title 21 Code of Federal Regulations 809.3 (a)	"In vitro diagnostic products are those reagents, instruments, and systems intended for use in the diagnosis of disease or other conditions, including a determination of the state of health, in order to cure, mitigate, treat, or prevent disease or its sequelae. Such products are intended for use in the collection, preparation, and examination of specimens taken from the human body […]."
IMDRF, Software as a Medical Device (SaMD): Key Definitions Section 5.2.1	'Medical device' means any instrument, apparatus, implement, machine, appliance, implant, reagent for in vitro use, software, material or other similar or related article, intended by the manufacturer to be used, alone or in combination, for human beings, for one or more of the specific medical purpose(s) of: ■ diagnosis, prevention, monitoring, treatment or alleviation of disease ■ diagnosis, monitoring, treatment, alleviation of or compensation for an injury ■ investigation, replacement, modification, or support of the anatomy or of a physiological process ■ supporting or sustaining life ■ control of conception ■ disinfection of medical devices ■ providing information by means of in vitro examination of specimens derived from the human body

The US method differs from the EU in two ways. In the US context, 'intended use' is defined as 'objective intent.' When determining objective intent,[61, vii] the test is not what the actual person believed but what a reasonable observer would determine their intention to be from their representations.[viii] In this way, intention is constructed from the evidence submitted. In this instance, the instructions, advertising materials, promotional materials and so on. Objective intent is not a concept found in the EU MDR/IVDR.

While it seems the ECJ and the various national supervisory bodies engage in a similar process, the MDR/IVDR appears to question subjective – not objective – intent. Despite this disparity, the evidence used to determine intended use or purpose remains much the same across the two jurisdictions, this intent being determined by reference to the manufacturer's statements and the context that surrounds them.[62] As in the EU, labelling, advertising claims, and other statements are all eligible evidence to determine intended use. The US method is arguably 'risk-based' rather than being a question of qualification as laid down in the MDR/IVDR.[63] In this way, while a particular device might meet the definition of 'device', the FDA exercises its discretion in refusing to regulate 'low-risk' devices. This differs from the EU position where the scope of the MDR/IVDR is determined by whether devices qualify as devices.

Some commentators such as Quinn have claimed that this second feature is critical in distinguishing between US and EU approaches. In the remainder of this section, we question the significance of this distinction.

Low-risk digital health devices

The claim that US and EU jurisdictions differ in their approach to regulating digital health devices is often premised on the contrast between the FDA drawing the boundary of medical device law using its risk-based discretion and the EU with device qualification. In other words, what devices qualify as medical or in vitro diagnostic devices may differ in these two jurisdictions. In short, two elements to the FDA's strategy supposedly differ from the EU. First, the FDA have broad powers to regulate devices but choose not to regulate a subset of these devices. Second, the FDA use risk to draw the boundary between those devices they choose and exclude from regulation.

When establishing this distinction, Quinn cites the FDA's *Mobile Medical Applications: Guidance for Industry and Food and Drug Administration Staff*.[64] The guidance states that the FDA intends to apply its regulatory oversight to 'only those mobile apps that are medical devices and whose functionality could pose a risk to a patient's safety if the mobile app were to not function as intended.'[65] Further to this, Section V-B and Appendix B of the same document provide examples of devices that may meet the definition of device, but which the FDA intends to exercise its enforcement discretion, that is, not regulate as medical devices.[66] The implication of this is that the FDA has jurisdiction but chooses not to enforce its powers against low-risk devices.

Released in 2015, the *Mobile Medical Applications* guidance is now out of date. Specifically, Section 3060 of the 21st Century Cures Act 2016 determines what software is within FDA oversight:

'Certain software is exempted from requirements for medical devices, including software that provides medical recommendations and the basis for those recommendations to health care professionals. Software remains subject to regulation as a medical device if: (1) the software acquires, processes, analyzes, or interprets medical information; or (2) the FDA identifies use of the software as reasonably likely to have serious adverse health consequences.' [67]

While the FDA's remit has been recast under the 21st Century Cures Act, its discretion to regulate software remains largely unchanged. Specifically, the provision that includes software that 'acquires processes, analyzes, or interprets medical information' is extremely broad. Further, those devices that do not fall neatly within this broad provision may then be caught by the provision that captures risky devices. In other words, while the FDA has less discretion, the difference is slight; the only devices outside their remit being low risk devices that do not acquire, process, analyse, or interpret medical information.[68]

The position of the FDA method for determining device qualification then remains much the same as Quinn described it: for the most part, the FDA uses their risk-based discretion to determine whether a digital health device will be subject to medical device law. The EU lacks this jurisdictional discretion, the scope of the MDR/IVDR being determined by the question of device qualification. However, there is reason to believe that this risk-based strategy may not be so different from the EU.

2.3 Comparing EU and US positions

Given the differences between the EU and US methods of determining device qualification, does the US method retain more flexibility to capture risky digital health devices? Certainly, the EU's national bodies lack the discretion the FDA has to refuse to regulate certain devices. However, does the EU approach lack the sensitivity to risk that the US approach possesses?

The argument that the EU does have some flexibility and sensitivity to risk is supported by the *SNITEM* judgment. The judgment emphasises the importance of device function to establish device qualification. However, function is not necessarily helpful to distinguish the regulated from the unregulated, that is, medical devices from well-being devices. Indeed, well-being devices perform similar functions to many regulated devices – they often monitor patients for conditions in a similar way that a medical device might, but their intention is quite different. Given this, it is likely that the competent authorities and the ECJ may look at the kind of risk the devices pose to assess whether the device really does have a specific medical purpose. If this is true, then risk will play an increasing role in determining device qualification.

Both jurisdictions are likely to emphasise the risk a device poses to determine whether to regulate or not regulate a device. For the FDA, risk may put the device within the FDA's remit or may determine whether the FDA will exercise its regulatory discretion. For the EU, risk may assist when distinguishing between medical devices and well-being devices. Given this, despite any apparent differences, the two jurisdictions may come to similar conclusions using slightly different means.

3. Machine learning as a medical device

This section considers a subset of algorithms in healthcare: the set of methods known as machine learning and whether these techniques pose a novel problem for medical device regulation. Determining whether machine learning constitutes a novel problem for medical device law depends upon a number of considerations, for example, what counts as machine learning, the application of this technology in healthcare, and what medical device regulation requires. This section considers these details in relation to two supposed novel problems that machine learning for healthcare might pose for medical device regulation. First, the problem of assessing the safety and effectiveness of 'black box' models. Second, the problem of assessing the safety and effectiveness of machine learning models if these models constantly retrain and change.

3.1 What is machine learning?

In order to consider whether machine learning poses a novel problem for medical device regulation, a robust definition of machine learning is required. Machine learning is the more specific term compared to the wider and more amorphous concept artificial intelligence.The term 'artificial intelligence'[ix] often conjures ideas of algorithms with broad powers of cognition.[69] By contrast, machine learning describes an approach to programming that produces algorithms with bounded, task-specific intelligence.[70] In a phrase, machine learning algorithms are narrowly intelligent, broadly unintelligent. As a concept, machine learning has many definitions, but the starting point is often Tom Mitchell's classic definition:

'A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E.'[71]

To give an applied example, if we develop a machine learning algorithm to play checkers:

'E = the experience of playing games of checkers

T = the task of playing checkers, and

P = the probability that the program will win the next game'[72]

Consequently, the core of any definition is that machine learning systems learn and so improve from experience.[73–74] With some simplification, the difference between traditional programming and machine learning is as follows.[75] Traditional programming combines data and the program to produce the desired output. Machine learning combines the data and output to create the program.

Much of the practice of machine learning focuses on creating a 'well-posed learning problem,'[76] in other words, using 'the right features to build the right models that achieve the right tasks.' Indeed, machine learning consists of these three ingredients:[77] features that 'define a language in which we describe the relevant objects, be they emails or complex organic molecules'; tasks being the abstract representation of problems we wish to solve with these objects; and models being the product of applying the machine learning algorithm to the training data.

The term machine learning describes a diverse set of methods to detect and predict patterns - there is no one machine learning technique.[78] It is common to divide the field into three paradigms of machine learning:

- 'Supervised' (or 'predictive') learning use training data consisting of labelled sets of input-output pairs.[79] Following these pairs, the model will then learn the features of the input data associated with the labelled outputs. For example, to construct an email spam filter, sample emails (inputs) known to be or not be spam will be labelled as such (output) to constitute a model

- 'Unsupervised' (or 'descriptive') learning approaches attempt to find patterns of interest in the data.[80] Unlike supervised learning, the relationship between the inputs and outputs is unknown. Many unsupervised machine learning models are directed toward finding structure in a dataset, often a necessary step to solve a supervised machine learning problem[81]

- 'Reinforcement learning' tells us 'how to act or behave when given occasional reward or punishment signals.'[82] In this way, an 'agent' receives information about its environment and learns to pick actions that maximise some reward.[83] Reinforcement learning has applications across a diverse set of fields, for instance, self-driving, robotics, resource management, and education

Apart from these distinctions, machine learning as a field is one characterised by deep theoretical differences. These approaches underpin how developers frame and solve their respective problems.[84] In this way, while machine learning algorithms may use similar features to describe their objects and seek to solve similarly defined tasks, the models produced differ widely[85]

3.2 Machine learning for medicine

There are three near-implementation applications for machine learning in medicine:[86]

- Automation/semi-automation of tasks currently performed by humans, e.g. segmentation of medical images

- Mining of large datasets to uncover novel patterns and insights for discovery, e.g. novel disease biomarkers of drug targets

- Prediction of health and disease by complex patterns recognition, e.g. disease detection and diagnostics, clinical decision support

Applications of machine learning extend from the research that underpins medical practice to changing the practice of medicine by assisting clinicians to diagnose and treat patients.

Machine learning for healthcare has a number of current or near use applications (see Table 4).

Table 4: Machine learning for healthcare tools in current or near use applications

Challenge the tool addresses	Example of a tool	Solution the tool provides	Status of the tool
The task of compiling systematic reviews. An estimated 2.5 million English-language scientific articles are published each year and rising.[87]	Project Transform (with EPPI-Centre at University College London), Cochrane.[88]	Machine learning to assist with searching, study eligibility assessment, data extraction, and evidence synthesis.[89]	Present to near future implementation (although non-machine learning text mining is already in use).[90]
It is time consuming and expensive to manually delineate radiological images	Microsoft Research's Inner Eye.[91]	Machine learning for automatic delineation of healthy anatomy and tumours.	Near use (in development).[92]
The difficult task of diagnosing wrist fractures in a timely manner.	OsteoDetect.	AI analysis of wrist radiographs to highlight regions of distal radius fractures.	FDA market authorisation granted.[93]
Detection of normal heart rhythm and atrial fibrillation often requires medical intervention by skilled medical professionals.	AliveCor with SmartRhythm monitoring.[94]	Portable medical-grade EKG and analysis for consumer use.	FDA market authorisation granted.[95] CE marked
Developing effective treatments for rare diseases can be difficult as getting novel drugs to market is an expensive exercise and rare disease populations are, by definition, small.	Healx's HealNet.[96]	Machine learning to draw on a number of datasets: clinical trials, disease symptoms, drugs targets, multi-omic data, and chemical structures to identify new uses for existing drugs. That is, old drugs, new tricks.	In use.[97]
In the UK, an estimated 1.5 to 3 million people who attended emergency departments 'could have had their needs addressed in other parts of the urgent care system.'[98]	Babylon Health's Babylon Check.	Automated triage system to route patients to the appropriate service.[99]	Chatbot triage service CE marked.[100]
Differentiating between malignant and benign tissue in breast cancer surgery.	iKnife.[101]	A knife that heats tissue as it cuts, analysing the smoke with Rapid Evaporative Ionisation Mass Spectrometry to differentiate cancer from normal breast tissue.[102,103]	Near use (in trials).[104]

Machine learning promises to change healthcare from the research that underpins the field, to the discovery of new treatments, all the way to the operating theatre.

These applications have a number of features that impact on how they are used and regulated:

First, machine learning comprises a set of tools to solve a problem. Different machine learning techniques underpin the technologies listed in Table 1. For example, Inner Eye utilises Deep Neural Decision Forests – essentially a long series of branching decisions – to analyse radiological images.[105] By comparison, a fully-fledged machine learning systematic review aggregator would use several different machine learning methods to compile systematic reviews.[106]

Second, machine learning has a number of different applications. Machine learning is a set of techniques; its use in text mining presents very different risks than its use in surgery. Given this, while there may be common features across machine learning as a set of techniques, nuanced regulation ought to look at the particular use the technique purports to solve.

Third, different machine learning techniques require varying amounts of data and amounts of human intervention.[107] Depending on the technique utilised and problem in question, the resulting machine learning use will require small datasets and little human decision-making.

To summarise, given the breadth of machine learning uses across healthcare, the breadth of techniques in use, and the breadth of data and human involvement, the requirements for any given machine learning algorithm depends heavily on context.

3.3 Two challenges posed by machine learning for healthcare

The range and scale of potential uses of machine learning for healthcare raises many potential issues for medical device regulation. For example, quality machine learning requires quality data; biased data can lead to a biased model and inevitably an algorithm that functions differently in different populations. While data bias is undoubtedly an issue for machine learning and medical device regulation, it is not a problem unique to machine learning.

The following sections focus on two potential novel problems that machine learning for healthcare poses for medical device law, namely:

1. **The opacity of black box algorithms**. A concern often levelled at the increasing use of machine learning is that these systems are black boxes. That is, models 'whose internal workings are either unknown to the observer or known but uninterpretable to humans.'[108] This concern is widely discussed in relation to the General Data Protection Regulation and its supposed right to explanation.[109–110] Yet, opacity may also constitute an issue for machine learning as a medical device. Specifically, might opacity in machine learning also impair our ability to assess a medical device's safety and effectiveness?
2. **The dynamic nature of machine learning**. Machine learning may be more 'dynamic' than its traditionally programmed counterparts. Namely, machine learning might constitute a moving target for regulators as some machine learning models constantly retrain using incremental learning from streaming data. The question arises how can we assess the safety and effectiveness of these models under the MDR/IVDR?

3.4 Issue 1: Black boxes

This section first considers whether machine learning algorithms are necessarily black boxes and what 'black box' might mean, and then goes on to consider what the MDR/IVDR and their harmonised standards require. Ultimately, the question is whether medical device law requires developers and manufacturers to open their black boxes?

Are all machine learning algorithms black boxes?

The term 'black box' has multiple dimensions emphasised across multiple disciplines.[111] For our purposes, 'black box' may refer to a) the human interpretability of the model, 'the degree to which a human can understand the cause of a decision'[112] or b) the testability of software whose internals are hidden from the tester. In either usage of the term, something about the algorithm or model is hidden, opaque, or unavailable to an observer or tester.

Machine learning models vary in the extent to which they are interpretable by humans. One way to have interpretable machine learning is to use only interpretable machine learning models.[113] Some forms of machine learning like neural networks are often difficult to interpret and explain.[114] Other models, while being premised on reams of data and trained rather than explicitly programmed, may still nevertheless lend themselves to human interpretation.[115] For instance, 'decision trees' lend themselves to interpretation since they can be graphically represented, focus on relevant attributes only, and operate according to a hierarchical structure.[116] Other examples of 'interpretable' models may include 'additive models',[117] 'attention-based networks',[118] and 'sparse linear models'.[119] In short, it is a fallacy to think of machine learning as a magician's hat where data goes into the hat and a model is pulled out as if by magic.[120]

It is clear that there are machine learning methods where it is possible to meaningfully interpret the model's conclusions directly.[121] However, not all machine learning models are readily interpretable to humans. Moreover, developers may also have persuasive reasons to choose opaque models over interpretable models, as the cost of interpretability may sometimes be accuracy.[122] In this way, the computational goal of interpretable outputs may use different tools and constitute a significantly different objective from only achieving accurate outputs.[x]

Another way to achieve interpretable machine learning is to interrogate an otherwise opaque machine learning model. Broadly, this could be done in two ways. First, a 'model-agnostic method' could be used to extract the necessary information to interpret the model.[123] Model-agnostic methods 'extract post-hoc explanations by treating the original model as a black box', deriving an interpretable model using the predictions of the black box model.[124] In other words, another model is built to interpret an otherwise opaque model. Second, an 'example-based explanation method' could be used to explain the behaviour of the model.[125] Example-based models select instances of the dataset and seek to explain that decision: these instances then might tell us something about the model itself. For example, 'counterfactual explanations' note the closest possible world where the desired outcome would have been reached.[126] Counterfactual explanations take the form: 'if you earned X amount more, you would have been offered the loan' and so provide some explanation as to why a particular decision was reached.[127] Whichever technique used, there are various ways to interpret otherwise uninterpretable models and the decisions they make.

Apart from human interpretability, there are also methods to test software whose internal structure is hidden from the tester.[128] Broadly, software testing techniques can be classified as 'black box' testing or 'white box' testing.[129] Black box testing (also known as specification-based testing) understands the function of software in terms of mapping from inputs to outputs.[130] With this kind of testing, the function of software may be tested without access to the source code, without opening the black box. White box testing (also known as code-based testing) is based on information derived from the source code.[131] If desired, a white box tester may test individual lines of code, not just how inputs map to outputs. In this regard, white box testing assures the tester that each line of code is correct. While both forms of testing have their own adherents, it is generally agreed that each strategy has its strengths and a full test of software would typically include both white and black box testing.[132] However, does the MDR/IVDR require such a test?

What does the MDR/IVDR require?

Manufacturers have a general obligation to ensure that their devices have been designed and manufactured in accordance with the MDR/IVDR when placing their devices on the market.[133] This begs the question, what does the MDR and IVDR require? In what senses might the MDR/IVDR require manufacturers to open their black boxes? In this respect, examination of the clinical evidence requirements of the MDR/IVDR and requirements for software in particular, is relevant.

The MDR/IVDR (fully in force in 2020 and 2022) emphasises a life-cycle approach to regulating medical devices.[135] In short, the MDR/IVDR, as well emphasising the path to gaining a CE mark, increasingly places emphasis on a rigorous post-market strategy to ensure device safety and effectiveness.

In summary, in order to place a device on the market (i.e. to gain rather than retain CE marking), devices must conform with applicable requirements in the Annexes to the MDR/IVDR. For example, MDR/IVDR conformity assessment procedures ensure that devices meet the General Safety requirements found in Annex I MDR/IVDR.

General safety and performance

Annex I anchors what the conformity assessment procedure requires of each device by providing the broad principles and structure that support more specific requirements found elsewhere in the Regulations or harmonised standards. These general safety and performance requirements advise on aspects for consideration, to ensure that the device is designed and manufactured in ways that are safe and effective.[136] In particular, these sections require manufacturers to mitigate risk and ensure the positive balance of benefit over risk (harmonised standards are explored later in this chapter). When meeting these general requirements, the manufacturer must establish one of two types of evaluation. If the software qualifies as a medical device, manufacturers must provide a sufficient level of 'clinical evaluation': if the software qualifies as an IVD, manufacturers must provide sufficient 'performance evaluation'. We examine each in turn.

Clinical evaluation – medical devices

What constitutes clinical evaluation? Manufacturers are expected to produce a clinical evaluation report.[137] The clinical evidence that informs this report comprises critical evaluation of the relevant scientific literature, critical evaluations of the results of all available clinical investigations,and consideration of currently available alternative treatment options.[138]

The level of 'sufficient clinical evidence' required will depend on three main factors.[139] First, the intended use of the device; second, the evaluation of the side-effects of the device; third, the acceptability of the risk-benefit ratio that emerges. Broadly, manufacturers must justify the level of clinical evidence they provide as being sufficient to meet conformity standards. Specifically, where the manufacturer is unable to demonstrate sufficient depth of clinical evidence, clinical investigations may be required.

How does this apply to machine learning medical devices? It is foreseeable that a manufacturer (in our case, a developer) could give a critical evaluation of the relevant literature. For example, if a developer seeks to establish that a chatbot for mental health has clinical evidence, they may review any available literature on that subject. However, in regards to evidence used for clinical investigations, this may prove more difficult as often this literature does not exist for machine learning devices. Indeed, comparisons to other machine learning devices are unlikely to meet the equivalence requirements to be used as clinical investigations.[140] Given this, manufacturers may have to perform their own clinical investigation study. Finally, evidence related to alterative treatment options may be problematic in the machine learning context. Most developers will simply not have access to other developers' models and the data used to train them, making it difficult to properly compare these 'alternative treatment options.' In the context of software, this information is highly sensitive and often constitutes a trade secret. Given this, it is hard to see how machine learning device manufacturers could rely on this branch of evidence.

Performance evaluation - IVDs

What constitutes performance evaluation? Manufacturers are expected to produce a performance evaluation report.[141] The clinical evidence that informs this report comprises the following.[142] First, scientific validity refers to the association of an analyte to a clinical condition or physiological state. Second, analytical performance refers to the ability of a medical device to correctly detect and measure a particular analyte. Third, clinical performance refers to the ability of the device to yield results that relate to a particular clinical condition or physiological state for the intended use and in accordance with the target population, and applicable to the intended user.

The level of 'sufficient clinical evidence' required will depend on three main factors.[143] First, the intended use of the device; second the evaluation of interferences and cross-reactions; third, the acceptability of the risk benefit-ratio. In summary, determining a sufficient level of evidence broadly follows the MDR, with one substitution.

Counterintuitively, performance evaluation under the IVDR may be a better fit for diagnostic machine learning devices than the clinical evaluation system under the MDR. This is interesting as most diagnostic machine learning devices will likely qualify as medical devices rather than IVDs. Table 5 below details how the IVDR performance evaluation system might be interpreted for a machine learning device that attempts to provide its users with a risk score. Specifically, the table compares how a standard IVD test and a machine learning device for surgical risk prediction might satisfy the performance evaluation criteria.

Table 5: How performance evaluation might fit with diagnostic machine learning

Clinical evidence for performance evaluation	Interpretation for a standard IVD, e.g. a diagnostic laboratory test (C-reactive protein for rheumatoid arthritis)	Possible interpretation for diagnostic machine learning e.g. machine learning for surgical risk prediction
Scientific evidence	Peer reviewed literature establishes that a particular protein is linked to a clinical condition e.g. C-reactive protein has been associated with rheumatoid arthritis.	Peer reviewed literature establishes that a particular feature is linked to a clinical condition e.g. diabetes might constitute a risk factor for developing pneumonia.
Analytical performance	Laboratory-based testing to confirm that given protein can be adequately detected and that there are no interfering substances e.g. in the development of a commercial C-Reactive protein test.	Using test data to determine the accuracy of the features of your model e.g. through training the model learns that diabetes is a risk factor for pneumonia. Using test data that the model has not been trained on, testing can show the accuracy of the model as a predictor of pneumonia.
Clinical performance	Clinical-based testing to confirm that the commercial C-Reactive protein test performs as expected across its intended use.	Using external test data to confirm the causality between the feature and the risk factor and ensuring the model functions across its intended use e.g. confirmation that diabetes is a relevant risk factor across a range of populations.

Following the above table, using the performance evaluation methods contained in the IVDR may mitigate against assigning causality to what are in fact mere correlations.

For example, the machine learning model to predict pneumonia risk in Caruana *et al* (2015) found a correlation between having asthma and being less likely to develop pneumonia.[144] However, subsequent analysis found that asthma was only protective because this population received a different treatment, that is, the treatment that asthmatics received for their asthma was protective, not the condition itself. In this regard, the model's predictions had to be understood in their context. Arguably then, the IVD performance evaluation method may require manufacturers to unpack the findings of their black box model.

The upshot is this: while there is no general MDR/IVDR requirement for a model to be human interpretable, the IVD performance evaluation method may encourage manufacturers to render their model at least somewhat intelligible – to unpick causation from mere correlation. Given this, at least for diagnostic algorithms, a pathway that follows closely the IVD performance evaluation may be more appropriate than the clinical evaluation pathway found in the MDR.

Specific provisions for software

The MDR and IVDR also contain specific requirements for software in particular. Appendix 2 outlines these additional requirements. Do these specific requirements necessitate software to be either human interpretable or white box testable? There are few clues in these more specific requirements. For instance, 17.1, Annex I, MDR speaks to designing to ensure 'repeatability, reliability and performance in line with their intended use.' However, a device can be a black box yet still meet such requirements, being highly predictable and reliable yet also opaque – a lot hangs on the device's intended use. Further, 17.2, Annex I, MDR notes that 'software shall be developed and manufactured in accordance with the state of the art taking into account the principles of development life cycle, risk management...' Interpreting the meaning of 'state of the art' requires looking at the harmonised standards, the subject of the next section.

Harmonised standards – verification, validation, and risk management

Harmonised standards provide clarity to many of the requirements found in the MDR/IVDR, providing a framework to assist with compliance. In this regard, it makes sense to examine the relevant harmonised standards to consider whether medical device regulation might require human interpretability. Appendix 3 provides a brief description of some of the standards most relevant to machine learning as a medical device. In the following sections we consider whether the relevant harmonised standards might require manufacturers to open an otherwise black box model. Key to understanding some of the main principles of the harmonised standards is the differentiation of verification, validation, and risk management.

Verification and validation

'Verification' is 'confirmation, through the provision of objective evidence, that specified requirements have been fulfilled.'[145] In other words, verification questions whether the device meets certain standards or requirements, and can recast as the question 'are we building it right?' 'Validation' is 'confirmation, through the provision of objective evidence, that the requirements for a specified use or application have been fulfilled.'[146] In other words, validation questions whether the device meets the intended use of the product, and can be recast as the question: 'are we building the right thing?'

Proper verification and validation of machine learning models might require manufacturers to pay special attention to a number of factors. These factors may include implementing controls over the planning, segregation, and maintenance of training data (data used to train the model) versus test data (data kept aside to test that model) such that bias in the dataset is avoided and assurances can be made regarding the accuracy of the model's output.

In addition to these factors, it may be necessary to consider what real world validation is necessary to ensure that the algorithm fulfils its intended purpose and is sufficiently generalisable.[147] For example, some machine learning models are highly accurate but highly localised and may not function outside a given location or setting. For example, if one develops a risk tool to predict surgical complications in a specific hospital setting, real world testing across a number of different hospitals and scenarios may be necessary.

Finally, the principles of human factors engineering may be relevant in relation to transparency and explainability, especially where explainability is used as a means of reducing risk. For example, consider again a surgical risk tool designed to predict complications that might arise in surgery. If this model were to have clinical use, the end user, in this case, the surgeon, must understand the risk score generated as an output and be able to put this score in its proper context (risk management principles are further explored in the next section). Given this, it may be best practice to perform user studies to validate to what extent the model is human interpretable.

To summarise, developers of machine learning models will probably have to test their model with test data, may have to understand how their model functions across a range of situations, and, in some circumstances, may have to consider the explainability of their model. While no general requirement to open a black box exists, the cumulative requirements of verification and validation may require some machine learning models to be made somewhat intelligible. However, this requirement is heavily dependent on the intended use and risk the device poses. In this way, explainability and transparency may be a means to comply with the verification and validation process but are not core to those processes.

Risk management

A key element of the lifecycle process of medical devices is the implementation of a risk management strategy. Risk management strategies or plans form one of the main principles for medical devices regulation. There are two areas where the standards' risk management approaches might constitute a poor fit for machine learning systems.

First, it is unclear how might training and test data fit reduction of risk requirements. ISO 14971 tells us that manufacturers must reduce risk as-low-as-reasonably-practicable.[148] However, both the Directives and the Regulations differ saying manufacturers must reduce risks 'as far as possible'.[149] Where there is a discrepancy, manufacturers must take the approach in Directives/Regulations. However, this might be problematic, especially in regards to machine learning. For example, where a manufacturer takes their training data from a high quality dataset but identifies a potential for data bias, how far must they go to reduce this risk given that under the Directive/Regulations they cannot use economic considerations (i.e. time and cost) as a justification that a risk has been reduced as far as possible? Taken literally these requirements in context of machine learning would require a training dataset to be completely representative.

Second, machine learning models can often be fragile, often being sensitive to small changes in their data. In this regard, there may be challenges in profiling the risks associated with a small change to a machine learning dataset. For example, certain types of machine learning may be particularly sensitive to small changes in their dataset, i.e. deep neural networks resulting in catastrophic forgetting.

Following this assessment, it is unclear how machine learning might fit with the risk management elements of the harmonised standards. That is, the opacity of machine learning over its training data may require some creative interpretation of the applicable harmonised standards. This difficulty is further compounded on examination of the issue of change control under harmonised standards (see the next section).

Are black box models consistent with medical device regulation?

This assessment of black boxes, machine learning, and medical device regulation suggests the following conclusions:

- Not all machine learning constitutes a black box – either in terms of testing or in terms of human interpretability
- If a particular model is a black box, there are often methods to render its decision-making process at least somewhat human interpretable
- There is no general requirement from the MDR/IVDR or harmonised standards that machine learning be human interpretable
- There is no general requirement from the MDR/IVDR that machine learning be white box testable
- The IVDR performance evaluation method may be better equipped than the MDR clinical evaluation method to assessing the safety of diagnostic machine learning devices
- Depending on intended purpose, labelling, and the risk that the model presents, harmonised standards may require some probing of an otherwise black box model
- Neither the verification, validation, nor the risk management elements of the harmonised standards appear to be a good fit for some machine learning models

The upshot of this analysis is this: black box machine learning algorithms are not necessarily incompatible with the demands of medical device regulation. The MDR/IVDR's focus on safety and effectiveness means that the Regulations and their associated harmonised standards do not always require explanation for all device types. As no harmonised standard directly addresses this issue and the current set of standards seem insufficient, this may be a good opportunity to develop common specifications as outlined in Article 9 MDR/IVDR.

3.5 Issue 2: Dynamic machine learning devices

Machine learning learns by combining data and labelled outputs to create a model. Consequently, these models may retrain and so change over time. This ability to retrain may prove problematic for medical device regulation as the MDR/IVDR and their associated harmonised standards assess the safety and effectiveness of devices in a structured, systematic way. This section considers the 'dynamic nature' of some machine learning models and whether retraining is compatible with medical device regulation.

Consider Trigger's Broom, a take on the Ship of Theseus thought experiment:

'Trigger: This old broom has had 17 new heads and 14 new handles in its time.

Sid: How can it be the same broom, then?

Trigger: Here's a picture of it, what more proof do you need?'[150]

Trigger's Broom illustrates the problem of static identity over continuous change. Medical device regulation also faces a similar conundrum. If a medical device undergoes constant iterative change, when does this change necessitate re-execution of verification and validation, or additional clinical evidence, indeed, when does this iterative change constitute a different device entirely?

Medical devices, especially those containing software, are increasingly iterative, being continuously tweaked over their lifecycle.[xi] While the medical device industry is inherently more conservative than the software sector in general, it is clear the future points toward a more dynamic approach to medical device development. This dynamic nature calls for consideration of how the MDR/IVDR handles change in medical devices, and also whether machine learning is uniquely affected by this problem of iterative change. Perhaps the first question we should ask, though is how iterative or 'dynamic' is machine learning?

How 'dynamic' is machine learning?

Machine learning algorithms are made of data. As explained earlier, machine learning differs from traditional programming in that machine learning combines data and output to create a model. Some machine learning models require large quantities of data to function adequately, others require comparatively modest quantities.[151] Regardless, a quality model requires quality data, or, in a phrase: garbage in, garbage out. Consequently, a model that incorporates more data, more often might be thought a better model. The future of machine learning may be one informed by 'big data', trained continuously with 'streaming data', and set within a 'learning health system' that efficiently learns from experience. However, is this the case? Does machine learning constitute a moving target for regulators?

'Machine learning from streaming data' may mean one of two things.[152] The term might refer to models where predictions are made taking into account recent history. For example, weather prediction models will consider the weather today to predict the weather tomorrow. While the data input changes, the model remains the same. This is distinct from situations where the model itself evolves. With this second usage, the training data arrives over time and the model re-trains continuously. These algorithms might retrain 'incrementally', whenever it sees a new training instance, or as a 'batch', that is, 'every so often.'[153] The idea of a machine learning algorithm that learns incrementally perhaps represents the zenith of a learning health system that is hyper responsive and constantly learning from its experience.[154] Despite this, constantly evolving models are not necessarily the future of all machine learning for medicine.

Incremental learning is not a strategy appropriate for all machine learning problems – in fact, it may sometimes create more problems than it solves. Machine learning models can provide highly accurate predictions, yet they can also be highly fragile. For example, neural networks can fall victim to problems like 'catastrophic forgetting', the tendency of a network to abruptly forget previously learned information upon learning new information.[155] While these issues are surmountable, models that constantly retrain can often fail spectacularly. On the other hand, models to solve dynamic problems might die a slow death of irrelevance if the data that underpins them is no longer relevant. In other words, the choice of how and when to incorporate new data and how radically to retrain the model depends on the problem at hand. A stock trading algorithm requires streaming data and constant retraining, an algorithm that automatically delineates tumours may not require such time-sensitive data and retraining.[xii]

To summarise, streaming data may be incorporated as an input into the machine learning model or be used to constantly retrain the model. Second, the choice to constantly retrain a model is a design choice that stems from the kind of problem the model has been created to solve. In short, machine learning models are not necessarily hyper dynamic, they need not incorporate streaming data, and, if they do, this data need not retrain the model itself.

Dynamic devices under the MDR/IVDR

Machine learning models may or may not utilise streaming data. Regardless, supposing a model does utilise streaming data by retraining continuously, how would the MDR/IVDR deal with such iterative change?

Manufacturer's responsibilities do not end with the granting of a CE mark. On the contrary, the MDR/IVDR may require post-market surveillance. Moreover, manufacturers are required to be responsive to change. In this regard, as a part of the manufacturer's post-market surveillance plan, periodical safety reviews may be required.[156] More concretely, manufacturers are obliged to inform the notified body of 'substantial changes' to the quality management system or to parts or components of the device which might significantly change the performance or safety characteristics or the intended purpose of the device.[157] In this way, as the characteristics of the device change, manufacturers may have to update their performance assessments as necessary.

Appendix 4 outlines the broad responsibilities of manufacturers to keep pace with such change. Major categories of responsibility include: quality management systems, parts and components, performance evaluations, clinical investigations, and unique device identifiers. The provisions listed in Appendix 4 often speak in terms of 'substantial' and 'significant' change – requiring manufacturers to notify their notified body of any changes that reach this threshold. This generates two questions. First, what change counts as 'substantial' or 'significant'? Second, while a single change might not count as 'substantial' or 'significant', what if multiple changes do amount to a 'substantial' or 'significant' change? It seems that at least a partial answer to these questions may be found in the harmonised standards, the topic of the next section.

The MDR/IVDR are somewhat flexible in that they envision a device whose characteristics change over time. Nevertheless, methods such as incremental learning paired with streaming data seem difficult to reconcile with the Regulations, as the model constantly retrains specifically to improve performance. Indeed, the MDR/IVDR still seems anchored in a static conception of medical devices, many of its requirements appearing to be impractical for a device as dynamic as one that incorporates incremental learning. For instance, the requirements relating to clinical investigation (required only for a subset of devices) listed in Appendix 4 seem to envision that most changes to devices will be superficial, directed toward usability or quality of life updates rather than changes directly to affect device performance. Given this, examination of the more specific harmonised standards and how these standards view change in devices is necessary.

Harmonised standards – change control

As a general requirement, the harmonised standard ISO 13485 requires manufacturers (developers in our case) to control design and development changes.[158] Organisations must understand the significance of changes as they impact the function, performance, usability, safety, and applicable regulatory requirements for the medical device and its intended use.[159] Before implementation, changes must be reviewed, verified, validated (as appropriate), and approved.[160] These general requirements are supplemented by more specific guidance in the harmonised standard BS EN 62304 on medical device software and software life-cycle processes.[161]

BS EN 62304 includes the concept of 'software maintenance.'[162] This concept is given wide definition and constitutes a core consideration that requires manufacturers to monitor, evaluate, and resolve feedback. To illustrate, consider a medical imaging algorithm that fails to properly analyse atypical anatomy, a problem that is then reported by the end user to the manufacturer. The manufacturer would then have to consider whether this represented an adverse event and respond accordingly. One possible appropriate response might be to initiate a change request.

BS EN 62304 requires a controlled approach to change, requiring a process starting with a change request, implementation of change, and verifying change. One such source of change is feedback, that is, software maintenance. A response to this change might include retesting.

In regards to specific requirements for testing software, the manufacturer must retest after changes have been made. In particular, when making changes during software system testing the manufacturer must repeat or perform additional tests to verify the effectiveness of the change in correcting the problem, conduct tests to demonstrate that unintended side effects have not been introduced, and perform relevant risk management activities.[163]

Another consideration is the need for revalidation. As previously mentioned, validation is necessary to show the device fulfils its intended purpose. If the manufacturer makes a significant change, one might expect that this process will have to be repeated. However, the concept of re-validation is not considered by BS EN 62304. Although unharmonised, BS EN 82304 does address re-validation.[164] Consequently, there may be an argument to harmonise BS EN 82304 or to amend BS EN 62304 to ensure that major software maintenance has not impaired the device's ability to meet its intended purpose.

To summarise, it is unclear how machine learning, especially machine learning that utilises incremental learning and streaming data, might be compatible with the above change control requirements. The very idea of a dynamic algorithm that retrains constantly seems a challenging concept when compared to the structured rigour of the change control processes outlined above.

Are dynamic devices consistent with medical device regulation?

This assessment of machine learning, dynamic devices, and medical device regulation suggests the following conclusions:

- Not all machine learning is necessarily dynamic – some models may remain static and not retrain at all, some may retrain in batches, others may retrain constantly

- The MDR/IVDR requires manufacturers to be responsive to change in their devices and continue to ensure their device is both safe and effective once placed on the market
- However, the MDR/IVDR process, while acknowledging that change may occur, does seem geared toward a relatively static conception of devices – devices that change slowly and methodically if at all
- The relevant harmonised standards are a poor fit for dynamic devices – current standards envision a highly structured way to initiate change in devices. It is difficult to see how incremental learning might be consistent with such processes

Overall, it is clear that retraining machine learning models do not fit well with the MDR/IVDR and its associated harmonised standards. In fact, such devices require a strained interpretation of current medical device regulation. While it is true that many machine learning devices will not retrain in such a dynamic way, if these devices are to come to market as medical devices, regulators may be met with a decision: either update how harmonised standards deal with change to incorporate such dynamic devices as medical devices or deny all of these dynamic devices CE marking. Alternatively, as no harmonised standard directly addresses this issue and the current set of standards seem insufficient, this may be a good opportunity to develop common specifications as outlined in Article 9 MDR/IVDR.

3.6 Does machine learning pose novel problems?

This section of the report considers two supposed 'novel' problems that machine learning might pose for medical device regulation. First, how it may be difficult to assess the safety and effectiveness of any black box machine learning models. Second, how machine learning might constitute a moving target for medical device regulation.

Analysis suggests that machine learning does not necessarily pose such problems. Rather than assuming that machine learning is inherently exceptional and using the technology as a cipher for the future of AI, it would be more accurate to highlight the unexceptional nature of many examples of machine learning that are neither opaque nor dynamic, and do not pose such regulatory challenges. Where machine learning is exceptional, the technology may not easily fit existing regulatory frameworks. Where this is the case, further regulatory guidance, strategies, or harmonised standards may need to be developed.

3.7 New strategies to regulate machine learning

Regulatory bodies are well aware that machine learning (and AI in general) might upset existing regulatory frameworks.[165] In this section, three regulatory developments that might be of assistance when regulating machine learning as a medical device are considered. First, the recent developments in the FDA strategy to regulate machine learning as a medical device. Second, how other standards like BS EN 82304 might be of use. Third, recent developments in the regulation of machine learning in other related areas, specifically NICE evidence standards and regulatory sandboxes.

FDA strategy

On April 2nd of 2019, the FDA released a discussion paper *Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning Based Software as a Medical Device* (FDA AI Framework). The FDA AI Framework elaborates on the previously released three-pronged approach to a) ensure that currently existing regulatory tools such as software validation tools are flexible and responsive to the changing field, b) utilise new programmes such as the FDA Pre-Certification Programme, and c) develop new regulatory tools to examine the validity and strength of AI-derived evidence.[166] This section describes the Pre-Certification Programme and the new FDA AI Framework and offers comments on points of praise, points of concern or limitations, and questions of how a similar strategy might fit in the EU context.

FDA Pre-Certification Programme

The FDA Pre-Certification Programme focuses on the quality of the manufacturer and their processes rather than the device itself. The FDA Pre-Certification Programme, while promising, is a pilot programme and is therefore limited in scope in two respects. First, at least initially, the programme operates only within the De Novo classification pathway and will be limited to software that is not a part of a hardware medical device.[167] Second, the programme is envisioned to be voluntary and only represents one regulatory pathway open to developers. If the device is eligible and a developer elects to use this regulatory pathway, the Pre-Certification Programme may represent a better fit for machine learning. Explicitly, the programme acknowledges that the traditional approach to regulating hardware-based medical devices is not well-suited to the iterative design and development important for software-based devices.[168] In an effort to remedy this, the new programme seeks to establish trust in an organisation, ensuring they have a 'culture of quality and organizational excellence.'[169] By emphasising transparency of process and demonstrating capacity to manufacture and monitor devices, the FDA hopes to shift the regulatory emphasis to organisations rather than products and to post-market surveillance over pre-market review.

FDA AI Framework

Released after the Pre-Certification Programme, the FDA AI Framework starts to conceptualise how the FDA views AI as a medical device.[170] The Framework includes elements of concern and elements of interest.

The FDA AI Framework has many laudable elements that jurisdictions like the EU Commission or national authorities might consider adopting. Namely, the Framework recognises that machine learning models may be dynamic and constitute a moving target for regulators. However, the Framework helpfully distinguishes between different kinds of modifications: modifications related to performance with no change to intended use or new input type, modifications related to inputs with no change to the intended use, and modifications related to intended use. Following these categories, the Framework then regulates each accordingly. This process of assessing whether machine learning is dynamic or not and then categorising what kind of dynamic change might be involved is a sound way to divide AI medical devices and regulate each according to the risk it poses.

Like the Pre-Certification Programme, the FDA AI Framework is limited in scope, only relating to 'software as a medical device.' Given this, the Framework, at least initially, does not cover AI that comprises a part of a wider hardware device. Further, while the Framework does show some flexibility toward accepting change, the documentation that manufacturers might have to provide for each device still seems onerous and perhaps a shift away from the process-based approach seen in the Pre-Certification Programme.

Will this work in the EU?

It remains to be seen how successful this new strategy will be and how a similar strategy might fit into the EU context. For example, there does appear to be some regulatory convergence between the EU and FDA, both jurisdictions emphasising the adoption of quality processes by manufacturers rather than requiring specific assessments. Moreover, it is also unclear how much the FDA Pre-Certification programme and AI Framework will differ from the EU MDR/IVDR and harmonised standards such as ISO 13485 and BS EN 62304. Indeed, these EU instruments already emphasise transparency of process and capacity to monitor medical devices. In short, any recommendation to adopt a strategy similar to that of the FDA should fully consider the similarities and differences between this programme and existing EU methods.[xiii]

New and existing standards for harmonisation

The safety and effectiveness of machine learning as a medical device is increasingly premised on these systems being transparent, ensuring the datasets remain free of bias, that any predictions the device makes are accurate, and that it operates securely. However, medical devices standards are not designed with many of these challenges in mind. Indeed, at best, these standards may be cobbled together to face these challenges, at worst they may be wholly inadequate.

Standards such as BS EN 82304 go some way to address these concerns, rolling in concerns about validation and verification with concerns regarding the security of health software products. However, no standard that might relate to medical devices seems to contemplate the issues of human interpretability and the dynamic nature of some machine learning devices. Consequently, new standards or harmonisation of other existing standards may go some way to addressing the two novel problems machine learning poses for medical device regulation. Nevertheless, no standard, nor the harmonised standards taken as a whole, fully equip developers with the processes to build and monitor safe machine learning devices for health.

NICE evidence standards and regulatory sandboxes

Machine learning is set to shakeup healthcare in general as well as other sectors. Consequently, medical device regulation may be able to learn from how other actors respond to the challenge. The NICE *Evidence standards framework for digital health technologies* is considered as well as the use of regulatory sandboxes in other sectors.

Medical device regulation is not a body of regulation that imposes standards on digital health devices. Specifically, NICE evidence standards assist in selecting technologies that are being considered for commission in the UK health and care system. Apart from this, NHS Digital has its own digital clinical safety regulations to help NHS organisations ensure the clinical safety of their IT software.[171]

Of further note is the Department of Health & Social Care's *Code of conduct for data-driven health and care technology*, which may go some way to ensuring transparency for machine learning. However, of particular note is the NICE *Evidence standards framework for digital health technologies* that assists in selecting technologies being considered for commissioning in the UK health and social care system. These standards distinguish between machine learning as a 'fixed algorithm' and machine learning as an 'adaptive algorithm' (those that constantly change).[172] This distinction seems to be a sensible one to make when assessing machine learning and perhaps a distinction that should be made when assessing machine learning as a medical device.

The FDA Pre-Certification programme is one way to give developers some latitude to develop innovative products but in a supervised environment. However, this kind of programme is not unique to the FDA. Indeed, in the realm of data protection, the Information Commissioner's Office (ICO) is in the beta phase of their 'Regulatory Sandbox.' The purpose of the sandbox is to assist developing innovative products in developing a shared understanding of what compliance might mean in these new areas. Similar sandboxes have been mentioned in relation to medical devices in the *Life Sciences Industrial Strategy*.[173] Further, NHS Digital have announced a joint project with the MHRA on 'synthetic devices.' This project aims at 'increasing the capability to accurately measure the effectiveness of new algorithms, including artificial intelligence and machine learning in medical devices in order to validate them.'[174] At the time of writing, little more is known about this project. However, if properly directed, it will hopefully address many of challenges outlined in this report.

In summary, there are other regulatory strategies that might be transplanted to regulate otherwise troublesome machine learning medical devices. The NHS Digital/MHRA joint project on synthetic devices hopefully represents a new and innovative regulatory strategy, that will address many of the challenges that some machine learning devices might pose.

Summary and recommendations

Healthcare is changing; medical device regulation must change with it. This report considered three interrelated challenges this evolution poses.

Challenge 1: Digital health and medical device regulation

- The medical device market is shifting. The dramatic expansion of digital health means that the quantity and variety of digital health devices grows daily. This market shift poses a threat in terms of sheer numbers: regulators may soon be swamped by a tidal wave of digital health devices. Further, because digital health devices differ from traditional medical devices – the skillset to test and properly analyse these devices requires different skills and expertise

- The growth in algorithms and software medical devices means that the nature of the medical devices sector is changing, with more developers being exposed to medical device regulation, often for the first time, without the institutional support and benefits of scale that manufacturers of more traditional medical devices might typically have

Given these broad conclusions, we recommend the following:

1. Bodies such as the MHRA and notified bodies should ensure that they possess sufficient expertise to assess this new generation of digital health devices, including machine learning

2. Regulators like the MHRA should ensure that market actors that may be caught by medical device regulation for the first time are aware of their obligations under the MDR and IVDR

Challenge 2: Digital health and device qualification

- The line between what qualifies as a medical device and what constitutes a life-style or wellbeing device is increasingly blurred. The availability of highly accurate sensors outside the clinic means that 'consumer devices' and 'medical devices' are no longer mutually exclusive. In this context, the law is also changing with cases like *SNITEM*. What qualifies as a medical device must be sufficiently flexible for regulators to regulate risky devices but also rigid enough for manufacturers to be given some degree of certainty over whether their device will qualify as medical device or not

Given these broad conclusions, we recommend the following:

1. That guidance akin to MEDDEV be issued that clarifies the issues of device qualification under the MDR and IVDR
2. That the European Commission should continue to update and expand upon its *Manual on Borderline and Classification in the Community Regulatory Framework for Medical Devices* under the Regulations – this practical resource is useful for manufacturers to understand how their device might be regulated

Challenge 3: Machine learning as a medical device

- Machine learning is set to change healthcare as the number of uses increase across the sector. However, machine learning is not necessarily qualitatively different from its manually programed counterparts
- Not all machine learning models are black boxes (human uninterpretable), some models are relatively transparent, their decisions being able to be represented with visualisations. Other models, while opaque, may be explained by constructing model-agnostic explainers or by testing particular instances with example-based explanations
- Not all machine learning models will constitute moving targets for regulators. Not all machine models retrain, nor do all machine learning models incorporate streaming data. However, those models that do constantly retrain may pose safety concerns as some machine models can be fragile, that is, small changes in data or the model may cause dramatically different outputs. In this regard, NICE *Evidence standards framework for digital health technologies* distinguishes between 'fixed algorithms' on one hand and 'adaptive algorithms' on the other
- While human interpretability and explanation are key issues, no current standard directly addresses them – this is despite the fact that human interpretability may be relevant when assessing the safety, effectiveness and risk of some devices
- Machine learning models that constantly retrain do not fit well with current medical device regulation. The MDR/IVDR and harmonised standards, while envisioning change in devices, do not envision the kind of dynamic change that some machine learning models represent
- There are other regulatory strategies that might assist in dealing with the specific problems that machine learning devices might pose. For instance, the proposed FDA AI Framework, FDA Pre-Certification programme, concepts from the NICE Evidence standards framework for digital health technologies, and regulatory sandboxes might provide lessons for medical device regulation

Given these broad conclusions, we recommend the following issues should be considered:

1. The extent to which the black box problem should be addressed by harmonised standards or common specifications at an EU and national level
2. Whether it is helpful to distinguish in harmonised standards and supplementary guidance like MEDDEVs between those machine learning medical devices that retrain and those that are static, at an EU and national level
3. Whether harmonising standards like EN 82304 for Health Software might capture risk factors (i.e. security concerns) not currently captured by other standards, at an EU and national level
4. Whether a programme akin to the FDA's Pre-Certification Programme and AI Framework might work within the EU and national context
5. How the NHS Digital and MHRA joint project on synthetic devices might address the black box and dynamic devices problem: this project seems to be an ideal opportunity for testing what a balanced regulatory strategy might look like.

Overall, we urge regulators to first utilise the regulatory tools that already exist in medical device regulation before imposing new systems of regulation on machine learning. Machine learning is a diverse set of tools and does not always represent a novel challenge to our current regulatory framework. Consequently, we urge caution when regulating machine learning; it would be unwise to regulate the entire field according to an exceptional subset of machine learning tools.

Appendix 1: Routes to device qualification under the MDR/IVDR

Broadly, a device might qualify as a medical or in vitro diagnostic device via three routes:[xiv]

- Via the intended purpose and specific medical purpose route found in Article 2(1) MDR and Article 2(2) IVDR
- Automatic qualification including via Annex XVI MDR or by being a device for the control or support of conception mentioned in Article 2(1) MDR
- Hybrid qualification by being a product specifically intended for the cleaning, disinfection or sterilisation of devices referred to in Article 1(4) MDR or a device specifically intended to contain or preserve specimens, mentioned in Article 2(2) and 2(4) IVDR

Further, a device may not qualify as a medical or in vitro diagnostic device but be regulated by the MDR/IVDR via three routes:

- While not being a medical or in vitro diagnostic device in and of itself, software (or another kind of module) may be a component or part of a device that is a medical or in vitro diagnostic device under Article 23 MDR or Article 20 IVDR
- While not being a medical or in vitro diagnostic device in and of itself, software may be combined with other devices to be placed on the market as a system or procedure pack under Article 22 MDR
- While not being a medical or in vitro diagnostic device itself, the manufacturer intends the device to be used as an accessory to a device that is a medical or in vitro diagnostic device

Appendix 2: MDR/IVDR additional requirements for software

Requirement	Applies to	Citation
6.1 (b) software verification and validation (describing the software design and development process and evidence of the validation of the software, as used in the finished device. This information shall typically include the summary results of all verification, validation and testing performed both in-house and in a simulated or actual user environment prior to final release. It shall also address all of the different hardware configurations and, where applicable, operating systems identified in the information supplied by the manufacturer)	Technical documentation, specifically pre-clinical and clinical data requirements	MDR, Annex II, 6.1
6.4 Software verification and validation The documentation shall contain evidence of the validation of the software, as it is used in the finished device. Such information shall typically include the summary results of all verification, validation and testing performed in-house and applicable in an actual user environment prior to final release. It shall also address all of the different hardware configurations and, where applicable, operating systems identified in the labelling.	Technical documentation, specifically pre-clinical and clinical data requirements	IVDR, Annex II, 6.4
17.1 Devices that incorporate electronic programmable systems, including software, or software that are devices in themselves, shall be designed to ensure repeatability, reliability and performance in line with their intended use. In the event of a single fault condition, appropriate means shall be adopted to eliminate or reduce as far as possible consequent risks or impairment of performance	General Safety and Performance requirements, specifically Chapter II on design and manufacture	MDR, Annex I, Chapter II, 17.1 IVDR, Annex I, Chapter II, 16.1
17.2 For devices that incorporate software or for software that are devices in themselves, the software shall be developed and manufactured in accordance with the state of the art taking into account the principles of development life cycle, risk management, including information security, verification and validation	General Safety and Performance requirements, specifically Chapter II on design and manufacture	MDR, Annex I, Chapter II, 17.2 IVDR, Annex I, Chapter II, 16.2

Appendix 3: Harmonised standards relevant to medical device software

Standard	Name	Scope
EN ISO 13485:2016	Medical devices - Quality management systems	Recommends a process approach to quality management.
EN ISO 14971:2012	Medical devices - Application of risk management to medical devices	Recommends a framework for managing risks, primarily to the patient, but also to the operator, other persons, other equipment and the environment.
BS EN 62304:2006	Medical device software - Software life cycle processes	Recommends a framework of life cycle processes for the safe design and maintenance of medical device software.
EN 82304-1:2017	Health software	Recommends requirements for the safety and security of health software products.

Appendix 4: Dynamic responsibilities of manufacturers

Requirement	Applies to	Citation
9. Manufacturers shall ensure that procedures are in place to keep series production in conformity with the requirements of this Regulation. Changes in device design or characteristics and changes in the harmonised standards or CS [common specifications] by reference to which the conformity of a device is declared shall be adequately taken into account in a timely manner. Manufacturers of devices, other than devices for performance study, shall establish, document, implement, maintain, keep up to date and continually improve a quality management system that shall ensure compliance with this Regulation in the most effective manner and in a manner that is proportionate to the risk class and the type of device	Manufacturer's general responsibilities	MDR, Article 10(9) IVDR, Article 10(8)
2.4 The manufacturer in question shall inform the notified body which approved the quality management system of any plan for substantial changes to the quality management system, or the device-range covered. The notified body shall assess the changes proposed, determine the need for additional audits and verify whether after those changes the quality management system still meets the requirements referred to in Section 2.2.	Quality management system	MDR/IVDR, Annex IX, Chapter I, 2.4
1. The manufacturer shall establish, document and implement a quality management system as described in Article 10(9) and maintain its effectiveness throughout the life cycle of the devices concerned	Quality management system	MDR/IVDR, Annex IX, Chapter 1, 1.
2. An item that is intended specifically to replace a part or component of a device and that significantly changes the performance or safety characteristics or the intended purpose of the device shall be considered to be a device and shall meet the requirements laid down in this Regulation.	Parts and components	MDR, Article 23(2) IVDR, Article 20(2)
6. The performance evaluation and its documentation shall be updated throughout the life cycle of the device concerned with data obtained from implementation of the manufacturer's PMPF [post-market performance follow-up] plan in accordance with Part B of Annex XIII and the post-market surveillance plan referred to in Article 79.	Performance evaluations	IVDR, Chapter VI, Article 56, 6.
1.3.3 The clinical evidence and its assessment in the performance evaluation report shall be updated throughout the life cycle of the device concerned with data obtained from the implementation of the manufacturer's PMPF plan in accordance with Part B of this Annex, as part of the performance evaluation and the post-market surveillance system referred to in Article 10(9)	Performance evaluations reports	IVDR, Annex XIII, Part A, 1.3.3
1. If a sponsor intends to introduce modifications to a performance study that are likely to have a substantial impact on the safety, health or rights of the subjects or on the robustness or reliability of the data generated by the study, it shall notify, within one week, by means of the electronic system referred to in Article 69, the Member State(s) in which the performance study is being or is to be conducted of the reasons for and the nature of those modifications… 2. The Member State shall assess any substantial modification to the performance study in accordance with the procedure laid down in Article 67.	Performance studies	IVDR, Article 71(1)-(2)

1. If a sponsor intends to introduce modifications to a clinical investigation that are likely to have a substantial impact on the safety, health or rights of the subjects or on the robustness or reliability of the clinical data generated by the investigation, it shall notify, within one week, by means of the electronic system referred to in Article 73 the Member State(s) in which the clinical investigation is being or is to be conducted of the reasons for and the nature of those modifications 2. The Member State shall assess any substantial modification to the clinical investigation in accordance with the procedure laid down in Article 71	Clinical investigations	MDR, Article 75(1)-(2)
3.9 A new UDI-DI shall be required whenever there is a change that could lead to misidentification of the device and/or ambiguity in its traceability; in particular, any change of one of the following UDI database data elements shall require a new UDI-DI: (b) device version or model, [see 6.5.1-6.5.3 for software specific rules] 5.4 Manufacturers shall periodically verify the correctness of all of the data relevant to devices they have placed on the market, except for devices that are no longer available on the market. 5.8 Manufacturers shall update the relevant UDI database record within 30 days of a change being made to an element, which does not require a new UDI-DI.	Unique Device Identifiers	MDR/IVDR, Annex VI, Part C, 3.9, 5.4, 5.8.
6.5.1. The UDI shall be assigned at the system level of the software. Only software which is commercially available on its own and software which constitutes a device in itself shall be subject to that requirement. The software identification shall be considered to be the manufacturing control mechanism and shall be displayed in the UDI-PI. 6.5.2. A new UDI-DI shall be required whenever there is a modification that changes: (a) the original performance; (b) the safety or the intended use of the software; (c) interpretation of data. Such modifications include new or modified algorithms, database structures, operating platform, architecture or new user interfaces or new channels for interoperability 6.5.3. Minor software revisions shall require a new UDI-PI and not a new UDI-DI. Minor software revisions are generally associated with bug fixes, usability enhancements that are not for safety purposes, security patches or operating efficiency. Minor software revisions shall be identified by a manufacturer-specific form of identification	Unique Device Identifiers – Device Software	MDR, Annex VI, Part C, 6.5.1, 6.5.2, 6.5.3. IVDR, Annex VI, Part C, 6.2.1, 6.2.2, 6.2.3.

Project summary

This report is part of the PHG Foundation project *Regulating algorithms in healthcare* that included a series of workshops and reports that sought to clarify:

- How algorithms in healthcare are regulated
- How algorithms in healthcare should be regulated

Regulating algorithms in healthcare considers how algorithms in healthcare are regulated, from the data that is used to train an algorithm to the question of who is liable if something goes wrong. The project considered the following general spheres of regulation:

- Algorithms as data (the General Data Protection Regulation and the Data Protection Act 2018)
- Algorithms as medical devices (the Medical Devices Regulation and In Vitro Diagnostic Medical Devices Regulation)
- Algorithms as intellectual property (including patent, copyright, and trade secret protections)
- Algorithms as a source of liability (clinical negligence, product liability, statutory compensation schemes)

Working with the Centre for Law, Medicine and Life Sciences at the University of Cambridge, the project convened two workshops bringing together academics, legal practitioners, regulators, developers, and clinicians.

Workshop 1 – *Regulating algorithms in healthcare – the GDPR and MDR/IVDR in practice* considered the following issues.

In regards to algorithms as data:

- The particular ethical issues algorithms might pose
- Whether the GDPR contains a right to explanation?
- Whether counterfactual explanations might satisfy such a right?

In regards to algorithms as medical devices:

- The position of software under the MDR and IVDR
- How software qualifies as a medical device in EU and US law
- The particular problems machine learning might pose for medical device regulation

Workshop 2 – *Regulating algorithms in healthcare – intellectual property and liability* was convened with the support of the Centre for Advanced Studies in Biomedical Innovation Law, University of Copenhagen. This workshop considered the following issues.

In regards to intellectual property:

- The patentability of 'computer implemented inventions' post *Alice*
- The viability of using open source software in the healthcare sector

In regards to liability:

- What scheme of liability might be most appropriate for artificial intelligence?
- The place of predictive analytics in medical malpractice

This report considers one part of the wider *Regulating algorithms in healthcare* project: algorithms as medical devices.

References

1. Shafer DW, Kirgin CM, Kaput JJ, *et al.* What is digital medicine? Studies in health technology. 2002; 80: 195–204.
2. United States Food and Drug Administration. Digital Health Innovation Action Plan. 2018.
3. Aitken M, Clancy B, Nass D. The Growing Value of Digital Health: Evidence and Impact of Human Health and the Healthcare System. IQVIA Institute for Human Data Science. 2017; p1.
4. Huckvale K, Adomaviciute S, Prieto JT, *et al.* Smartphone apps for calculating insulin dose: a systematic assessment. 2015; 13: 106–116.
5. Ferrero NA, Morrell DS, Burkhart CN. Skin scan: a demonstration of the need for FDA regulation of medical apps on iPhone. Journal of the American Academy of Dermatology. 2013; 68(3): 515–516.
6. Huckvale K, Prieto JT, Tilney M, *et al.* Unaddressed privacy risks in accredited health and wellness apps: a cross-sectional systematic assessment. BMC Medicine. 2015; 13: 214–227.
7. United States Food and Drug Administration. FDA allows marketing of first direct-to-consumer app for contraceptive use to prevent pregnancy. FDA. 2018
8. Viz.ai. Home https://www.viz.ai
9. AliveCor. How it works. https://www.alivecor.com/how-it-works
10. Tricella. Home. https://www.tricella.com
11. VitalConnect. Solutions https://vitalconnect.com/solutions
12. Babylon GP at Hand. About us. https://www.gpathand.nhs.uk
13. Regulation (EU) 2016/679 of the European Parliament and of the Council 27 April 2016 on the protection of natural persons with regard to the processing of personal data. Article 4(2). https://publications.europa.eu/en/publication-detail/-/publication/3e485e15-11bd-11e6-ba9a-01aa75ed71a1/language-en
14. Ibid. Article 4(1).
15. Council Directive 93/42/EEC of 14 June 1993 concerning medical devices. https://eur-lex.europa.eu/eli/dir/1993/42/oj
16. Council Directive 98/79/EC of 27 October 1998 on in vitro diagnostic medical devices. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A31998L0079
17. Council Directive 90/385/EEC of 20 June 1990 on the approximation of the laws of the Member States relating to active implantable medical devices.. https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX%3A31990L0385
18. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32017R0745
19. Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32017R0746
20. Article 120, Regulation (EU) 2017/745.
    Article 110, Regulation (EU) 2017/746.
21. Recital 2, Regulation (EU) 2017/745.
    Recital 2, Regulation (EU) 2017/746.

22. Article 2(1), Regulation (EU) 2017/745.
Article 2(2), Regulation (EU) 2017/746.
23. Annex VIII, Regulation (EU) 2017/745.
Annex VIII, Regulation (EU) 2017/746.
24. Article 35, Regulation (EU) 2017/745.
Article 31, Regulation (EU) 2017/746.
25. Recital 30, Article 5(5), Regulation (EU) 2017/745.
Recital 29, Article 5, Regulation (EU) 2017/746.
26. Draft guidance on the health institution exemption (HIE) – IVDR and MDR. MHRA; 2018.
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/675419/Health_institution_exemption_draft_for_public_consultation.pdf
27. Recital 22, Regulation (EU) 2017/745.
Recital 20, Regulation (EU) 2017/746.
28. Judgment of 27 of October 2016, *James Elliott Construction*, C-613/14, EU:C:2016:821, paragraph 40.
http://curia.europa.eu/juris/document/document.jsf?text=&docid=184891&doclang=en
29. Harmonised standards – Medical devices. About us. European Commission. 2019.
https://ec.europa.eu/growth/single-market/european-standards/harmonised-standards_en
30. Annex VI, Part C, 6.5.2, Annex XIV, Part A, 3, Regulation (EU) 2017/745.
Annex VI, Part C, 6.2.2, Annex XII, Part A, 2.3.2., Regulation (EU) 2017/746.
31. Annex I, Chapter II, 17.3, Regulation (EU) 2017/745.
Annex I, Chapter II, 16.3, Regulation (EU) 2017/746.
32. Meddev 2.1/6. Qualification and classification of standalone software. p1.
https://ec.europa.eu/docsroom/documents/17921/attachments/1/translations
33. Recital 19, Regulation (EU) 2017/745.
Recital 17, Regulation (EU) 2017/746.
34. Article 2(1), Regulation (EU) 2017/745.
35. Article 2(2), Regulation (EU) 2017/746.
36. Article 2(1), Directive 2007/47/EC of the European Parliament and of Council of 5 September 2007 amending Directive 90/385/EEC, Directive 93/42/EEC, and Directive 98/8/EC.
https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32007L0047
37. Loh E, Boumans R. Understanding Europe's New Medical Devices Regulation (MDR 2017/745). Emergo. 2017; p4.
38. Annex VIII, Chapter II, 3.3, Regulation (EU) 2017/745.
Annex VIII, 1.4, Regulation (EU) 2017/746.
39. Aitkin M. The Growing Value of Digital Health in the United Kingdom. IQVIA Institute for Human Data Science; 2017.
40. United States Food and Drug Administration. Digital Health Innovation Action Plan.2018
41. Grennan M. Price Discrimination and Bargaining: Empirical evidence from medical devices. American Economic Review. 2013; 103(1): 145-177.
42. Guidance for notified bodies on the regulation of IVDs for self-testing. MHRA; 2012.
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/377885/Guidance_for_notified_bodies_on_the_regulation_of_IVDs_for_self-testing.pdf

43. Quinn P. The EU Commission's risky choice for a non-risk based strategy on assessment of medical devices. Computer Law & Security Review. 2017; 33(3): 361–370.
44. Title 21 Food and Drugs, Section 201.128, Section 801.4, U.S. Electronic Code of Federal Regulations. https://www.ecfr.gov/cgi-bin/text-idx?SID=38fbeb12b19089e6dc7f734f2b38b2f0&mc=true&tpl=/ecfrbrowse/Title21/21tab_02.tpl
45. Clarification of When Products Made or Derived From Tobacco Are Regulated as Drugs, Devices, or Combination Products; Amendments to Regulations Regarding "Intended Uses"; Partial Delay of Effective Date. Federal Register. 2018; 83(52): 11639-11642. https://www.govinfo.gov/content/pkg/FR-2018-03-16/pdf/2018-05347.pdf
46. Judgment of the Court (Third Chamber) 22 November 2012, *Brain Products GmbH v BioSemi VOF and Others*, C-219/11, EU:C:2012:742.
47. Judgment of the Court (Fourth Chamber) 7 December 2017, *Syndicat national de l'industrie des technologies médicales v Premier ministre, Ministre des Affaires sociales et de la Santé (SNITEM)*, C-329/16, EU:C:2017:947.
48. Recital 19, Regulation (EU) 2017/745.
49. Article 2(1), Regulation (EU) 2017/746.
50. *Brain Products,* C-219/11.
51. Ibid. paras 17–18.
European Commission, MEDDEV 2. 1/1 Guidelines relating to the application of: Directive 90/385/EEC on active implantable medical devices and Directive 93/42/EEC on medical devices, Article 1.1(b). https://ec.europa.eu/docsroom/documents/10278/attachments/1/translations
52. Manual on Borderline and Classification in the Community Regulatory Framework for Medical Devices Version 1.22. European Commission. 2018; pp 77–83. https://ec.europa.eu/docsroom/documents/35582/attachments/1/translations/en/renditions/native
53. Meddev 2.1/6. Qualification and classification of standalone software. p1. https://ec.europa.eu/docsroom/documents/17921/attachments/1/translations
54. Ibid. p12.
55. Vollebreght E. An unsurprising case of software qualification with an interesting twist. Medicaldeviceslegal; 2018. https://medicaldeviceslegal.com/2018/01/12/an-unsurprising-case-of-software-qualification-with-an-interesting-twist/
56. *SNITEM*, C-329/16, paras. 17-20.
57. Ibid. para 33.
58. Vollebregt E. An unsurprising case of software qualification with an interesting twist. 2018.
59. *SNITEM*, C-329/16 para 34.
60. 5.2.1 International Medical Device Regulators Forum, Software as a Medical Device (SaMD): Key Definitions. 2013. http://www.imdrf.org/docs/imdrf/final/technical/imdrf-tech-131209-samd-key-definitions-140901.pdf

61. Title 21 Food and Drugs, Section 801.4, U.S. Electronic Code of Federal Regulations.
62. Title 21 Food and Drugs, Section 801.4, U.S. Electronic Code of Federal Regulations.
US v. William Facteau [2016] 15–10076.
https://casetext.com/case/united-states-v-facteau-2
US v. Vascular Solutions [2016] 181.
https://casetext.com/case/united-states-v-vascular-solutions-inc-1
63. Quinn P. The EU commission's risky choice for a non-risk based strategy on assessment of medical devices. 2017; pp 361-370.
64. Mobile Medical Applications: Guidance for Industry and Food and Drug Administration Staff. United States Food and Drug Administration. 2015; p4.
65. Ibid.
66. Ibid. pp 15-19.
67. Section 3060. 21st Century Cures Act 2016.
https://www.congress.gov/114/plaws/publ255/PLAW-114publ255.pdf
68. Evans O, Ossorio P. The Challenge of Regulating Clinical Decision Support Software After 21st Century Cures. American Journal of Law & Medicine. 2018; 44(2–3): p 389.
69. Lawrence DR, Brazier M. Legally Human? 'Novel Beings' and English Law. 2018; 26(2): pp 312–313.
70. Preparing for the Future of Artificial Intelligence. Executive Office of the President National Science and Technology Council Committee on Technology. 2016; p7.
71. Mitchell TM. Machine Learning. McGraw-Hill; 1997. p3.
72. Ng, A. Introduction to Machine Learning. Coursera. https://www.coursera.org/learn/machine-learning-duke?
73. Flach P. Machine Learning: The Art and Science of Algorithms that Make Sense of Data. Cambridge University Press. 2012; p3.
74. Craig C. McCarthy N, Montgomery J, *et al.* Machine learning: The power and promise of computers that learn by example. Royal Society. 2017; pp15–29.
75. Brownlee J. Basic Concepts in Machine Learning. Machine Learning Mastery. 2015. https://machinelearningmastery.com/basic-concepts-in-machine-learning/
76. Mitchell TM. Machine Learning.1997; pp 1-17.
77. Flach P. Machine Learning: The Art and Science of Algorithms that Make Sense of Data. 2012; pp 13-48.
78. Murphy KP. Machine Learning: A Probabilistic Perspective. MIT Press. 2012; p1.
79. Ibid.
80. Ibid.
81. Chollet F. Deep Learning with Python. Manning Shelter Island. 2018; p86.
82. Murphy KP. Machine Learning: A Probabilistic Perspective. 2012; p1
83. Chollet F. Deep Learning with Python. 2018; p87.
84. Domingos P. The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World. Allen Lane. 2015.
85. Flach P. Machine Learning. 2012. p13.

86. Raza S, Blackburn L, Moorthie S, *et al.* The personalised medicine technology landscape. PHG Foundation. 2018; pp96–99.
87. Ware M, Mabe M. The STM Report: An overview of scientific and scholarly publishing. STM Association. 2015; p6.
88. Takeda K. Text Mining to Improve the Health of Millions of Citizens . Microsoft Machine Learning Blog. 2017. https://blogs.technet.microsoft.com/machinelearning/2017/04/20/text-mining-to-improve-the-health-of-millions-of-citizens
89. Thomas J, Noel-Storr A, Marshall I, *et al.* Living systematic reviews: 2. Combining human and machine effort. Journal of Clinical Epidemiology. 2017; 91: pp31–37.
90. O'Mara-Eves A, Thomas J, McNaught J, *et al.* Using text mining for study identification in systematic reviews: a systematic review of current approaches. Systematic Reviews. 2015; 4(5): pp1–22.
91. Project InnerEye – Medical Imaging AI to Empower Clinicians. Microsoft Research. https://www.microsoft.com/en-us/research/project/medical-image-analysis/
92. Ibid.
93. United States Food and Drug Administration. FDA permits marketing of artificial intelligence algorithm for aiding providers in detecting wrist fractures . United States Food and Drug Administration. 2018. https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-algorithm-aiding-providers-detecting-wrist-fractures
94. AliveCor. How It Works. https://www.alivecor.com/press/press_release/alivecor-launches-ability-to-detect-a-serious-heart-condition-in-an-ecg-on-a-mobile-device-receives-ce-mark-on-atrial-fibrillation-algorithm-for-uk-and-ireland/
95. AliveCor. AliveCor Launches Ability to Detect a Serious Heart Condition in an ECG on Mobile Devices. AliveCor. 2015. https://www.alivecor.com/press/press_release/alivecor-launches-ability-to-detect-a-serious-heart-condition-in-an-ecg-on-a-mobile-device-receives-ce-mark-on-atrial-fibrillation-algorithm-for-uk-and-ireland/
96. Healx. Fragile X Syndrome. Drug Repurposing Summary Report. Healx; 2017. https://www.fraxa.org/wp-content/uploads/2017/06/Healx-Summary-Report_FRAXA_June-2017.pdf
97. Ibid.
98. Next steps on the NHS Five Year Forward View. NHS England. 2017. p 4.
99. Middleton K, Butt M, Hammerla N, *et al.* Sorting our symptoms: design and evaluation of the 'babylon check' automated triage system. 2016. https://arxiv.org/abs/1606.02041
100. Babylon. Babylon Health Regulatory. https://www.babylonhealth.com/regulatory
101. Francis D, Takats Z. i-Knife for breast cancer surgery. NIHR Imperial Biomedical Research Centre. 2016. https://imperialbrc.nihr.ac.uk/2016/06/26/i-knife-for-breast-cancer-surgery/
102. St John E, Balog J, McKenzie JS, *et al.* Rapid evaporative ionisation mass spectrometry of electrosurgical vapours for the identification of breast pathology: towards an intelligent knife for breast cancer surgery. Breast Cancer Research. 2017; 19(1): pp 59–73.
103. Balog J, Sasi-Szabo L, Kinross J, *et al.* Intraoperative Tissue Identification Using Rapid Evaporative Ionization Mass Spectrometry. Science Translational Medicine. 2013; 5(194): pp 1–12.

104. Wong S. "Intelligent knife" tells surgeon if tissue is cancerous. Imperial College London. 2013. https://www.imperial.ac.uk/news/126106/intelligent-knife-tells-surgeon-tissue-cancerous/
105. Kontschieder P, Fiterau M, Criminisi A, *et al.* Deep Neural Decision Forests. The IEEE International Conference on Computer Vision. 2015. p1470.
106. Elliot JH, Synnot A, Turner T, *et al*. Living systematic review: 1. Introduction-the why, what, when, and how. J Clin Epidemiol. 2017; 91: pp 23-30.
107. Beam AL, Kohane IS. Big Data and Machine Learning in Healthcare. JAMA. 2018; 319(13): pp1317–1318.
108. Guidotti R, Monreale A, Ruggieri S, *et al.* A Survey of Methods for Explaining Black Box Models. ACM Computer Surveys. 2019; 51(5): p 5.
109. Wachter S, Mittelstadt B, Floridi F. Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. International Data Privacy Law. 2017; 7(2): pp 76–99.
110. Selbst AD, Powles J. Meaningful information and the right to explanation. International Data Privacy Law. 2017; 7(4): pp 233-242.
111. Guidotti *et al*. A Survey of Methods for Explaining Black Box Models. 2019.
112. Biran O, Cotton C. Explanation and Justification in Machine Learning: A Survey. IJAI-17 Workshop on Explainable AI. 2017; 8: pp 8–13.
113. Molnar C. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. 2018. https://christophm.github.io/interpretable-ml-book/cite.html
114. Weld DS, Bansal G. The Challenge of Crafting Intelligible Intelligence. Communications of the ACM. 2018. p1.
115. Freitas AA. Comprehensible classification models: A position paper. ACM SIGKDD. 2014; 15(1): pp1–10.
116. Ibid. p 2.
117. Caruana R, Lou Y, Gehrke J, *et al.* Intelligible Models for Healthcare: Predicting Pneumonia Risk and Hospital 30-day Readmission. KDD 2015. 2015: pp 1721–1730.
118. Xu K, Ba JL, Kiros R, *et al.* Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. International conference on machine learning. 2015: pp 2048–2057.
119. Usten B, Cynthia R. Superspace linear integer models for optimised medical scoring systems. Machine Learning. 2016; 102(3): pp 349–391.
120. Domingos P. Ten Myths About Machine Learning. Medium. 2016. https://medium.com/@pedromdd/ten-myths-about-machine-learning-d888b48334a3
121. Ribeiro MT, Singh S, Guestrin C. Model-Agnostic Interpretability of Machine Learning. ICML Workshop on Human Interpretability in Machine Learning. 2016. p91.
122. Ibid.
123. Molnar C. Interpretable Machine Learning. 2018.
124. Riberio MT *et al.* Model-Agnostic Interpretability of Machine Learning. 2016. p91.
125. Molnar P. Interpretable Machine Learning. 2018.
126. Wachter S, Mittelstadt B, Russell C. Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR. Harvard Journal of Law & Technology. 2018; 31(2): pp 1–52.

127. Ibid.
128. Nidhra S, Dondeti J. Black Box and White Box Testing Techniques – A Literature Review. International Journal of Embedded Systems and Applications. 2012; 2(2): pp 29–50.
129. Jorgensen PC. Software testing: A craftsman's approach. CRC Press; 2014. pp 5–9.
130. Ibid. p7.
131. Liu H, Tan HBK. Covering code behaviour on input validation in functional testing. Information and Software Technology. 2009; 51(2): p 546.
132. Jorgensen PC. Software Testing: A craftsman's Approach. 2014. pp 8–9
133. Article 10, Regulation (EU) 2017/745
134. Article 10, Regulation (EU) 2017/746
135. Loh E *et al*. Understanding Europe's New Medical Devices Regulation MDR 2017/745. p 2.
136. Annex I, Regulation (EU) 2017/745.
Annex I, Regulation (EU) 2017/746.
137. Article 61, Regulation (EU) 2017/745.
138. Article 61(3)(a), Regulation (EU) 2017/745.
139. Article 61(3)(b), Regulation (EU) 2017/745.
140. Annex XIV, Part A, 3. Regulation (EU) 2017/745.
141. Article 56(3), Regulation (EU) 2017/746.
142. Article 56(3), Regulation (EU) 2017/746.
143. Article 56(1), Regulation (EU) 2017/746.
144. Caruana et al. Intelligible Models for Healthcare: Predicting Pneumonia Risk and Hospital 30-day Readmission. 2015.
145. 3.8.12. ISO 9000:2015 Quality management systems: Fundamentals and vocabulary. International Organization for Standardization; 2015.
146. 3.8.13. ISO 9000:2015 Fundamental. Quality management systems: Fundamentals and vocabulary; 2015.
147. United States Food and Drug Administration. Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML) – Based Software as a Medical Device (SaMD). United States Food and Drug Administration; 2019.
https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device
148. D.8. BS EN ISO 14971:2012 Medical devices – Application of risk management to medical devices. British Standards Institution; 2012.
149. Annex ZA.3. BS EN ISO 14971:2012.
Annex I, 2. Regulation (EU) 2017/745,
Annex 1, 2. Regulation (EU) 2017/746.
150. Only Fools and Horses. Episode 59: Heroes and Villains. 1996.
151. Murphy KP. Machine Learning. 2012 p2.

152. Parker C. Machine Learning From Streaming Data: Two Problems, Two Solutions, Two Concerns, and Two Lessons. BigML. 2013. https://blog.bigml.com/2013/03/12/machine-learning-from-streaming-data-two-problems-two-solutions-two-concerns-and-two-lessons/
153. Gepperth A, Hammer B. Incremental learning algorithms and applications. European Symposium on Artificial Neural Networks. 2016. pp 1–12.
154. Krumholz HM. Big Data and New Knowledge in Medicine: The Thinking, Training, and Tools Needed for a Learning Health System. Health Affairs. 2014. 33(7): pp 1163-1170.
155. Kirkpatrick J, Pascanu R, Rabinowitz N, *et al.* Overcoming catastrophic forgetting in neural networks. PNAS. 2017; 114(13): p 3521.
156. Articles 84-86, Regulation (EU) 2017/745.
Articles 79-81, Regulation (EU) 2017/746.
157. Annex IX, Chapter 1 2.4 Regulation (EU) 2017/745.
Annex IX, Chapter I, 2.4, Regulation (EU) 2017/746.
158. 7.3.9. BS EN ISO 13485:2016 Medical devices – Quality management systems – Requirements for regulatory purposes. British Standards Institution. 2016.
159. Ibid.
160. Ibid.
161. BS EN 62304:2016 + A1:2015 Medical device software – Software life-cycle processes. British Standards Institution; 2015.
162. 5.7.3. BS EN 62304:2016 + A1:2015.
163. Ibid.
164. BS EN 82304-1:2017. Health Software - Part 1: General requirements for product safety. British Standards Institution; 2017.
165. Gottlieb S. Transforming FDA's Approach to Digital Health. United States Food and Drug Administration; 2018. https://www.fda.gov/news-events/speeches-fda-officials/transforming-fdas-approach-digital-health-04262018
166. Ibid.
167. Software Precertification Program: Regulatory Framework for Conducting the Pilot Program within Current Authorities. United States Food and Drug Administration; 2019. https://www.fda.gov/media/119724/download
168. Developing a Software Precertification Program: A Working Model. United States Food and Drug Administration; 2019. p6. https://www.fda.gov/media/113802/download
169. Ibid.
170. Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning. United States Food and Drug Administration.

171. DCB0160: Clinical Risk Management: its Application in the Deployment and Use of Health IT Systems. NHS Digital; 2018. https://digital.nhs.uk/data-and-information/information-standards/information-standards-and-data-collections-including-extractions/publications-and-notifications/standards-and-collections/dcb0160-clinical-risk-management-its-application-in-the-deployment-and-use-of-health-it-systems
DCB0129: Clinical Risk Management: its Application in the Manufacture of Health IT Systems. NHS Digital; 2018. https://digital.nhs.uk/data-and-information/information-standards/information-standards-and-data-collections-including-extractions/publications-and-notifications/standards-and-collections/dcb0129-clinical-risk-management-its-application-in-the-manufacture-of-health-it-systems
172. Evidence Standards Framework for Digital Health Technologies for Digital Technology. National Institute for Health and Clinical Excellence. 2019. p33. https://www.nice.org.uk/Media/Default/About/what-we-do/our-programmes/evidence-standards-framework/digital-evidence-standards-framework.pdf
173. Bell J *et al*. Life Sciences Industrial Strategy – A report to the Government from the life sciences sector. Life Sciences Industrial Strategy Board; 2017. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/650447/LifeSciencesIndustrialStrategy_acc2.pdf

174. NHS Digital. NHS Digital welcomes funding for joint project with MHRA on creating synthetic devices. NHS Digital; 2018. https://digital.nhs.uk/news-and-events/latest-news/nhs-digital-welcomes-funding-for-joint-project-with-mhra-on-creating-synthetic-devices

Notes

i. This difference in timing is important to determine the position of the MDR/IVDR under the European Union (Withdrawal) Act 2018.

ii. Article 114 and 168(4)(c) TFEU form the treaty basis for both the MDR and IVDR. Article 114 refers to the EU's exclusive competence to harmonise the single market. Article 168(4)(c) refers to the derogation allowing EU competence to set standards of 'quality and safety for medicinal products and devices for medical use.'

iii. We thank Julian Hitchcock for alerting us to this point.

iv. The definition of 'intended purpose' remains largely the same in Regulations as the Directives. The Regulations only add an extra sentence to the Article 1(2)(g) Medical Device Directive definition, specifically: 'or sales materials or statements or as specified by the manufacturer in the performance evaluation.'

v. The other sections of the Meddev 2.1/6 guidance cited refer to the modules chapter of the guidance.

vi. For instance, following Section 201(h)(1) of the Federal Food, Drug and Cosmetic Act, if a device is listed as such under the official National Formulary or United States Pharmacopeia, the device automatically qualifies as a medical device.

vii. As previously mentioned, the FDA had recently tried (and failed) to elaborate on the meaning of 'intended use.' This elaboration would have revised the interpretation of 'intended use' for both drugs and devices. As of March 2019, this elaboration has been indefinitely delayed after controversy over the added phrase 'totality of evidence,' see Statement from FDA Commissioner Scott Gottlieb, M.D., on the FDA decision to seek additional time to reassess rule that would have changed longstanding practices for how the agency determined the 'intended use' of medical products. FDA; 2018.

viii. For an example drawn from English contract law, see Lord Clarke's dictum at RTS Flexible Systems v Molkerei Alois [2010] UKSC 14. 45.

ix. For various definitional difficulties with the definition of 'artificial intelligence', see Russell SJ, Norvig S. Artificial Intelligence: A Modern Approach. Prentice-Hall; 1995: 31-52.

x. There may be a parallel to literature on differential privacy here: '…by rethinking the computational goal, one can often obtain far better results than would be achieved by methodically replacing each step of a non-private computation with a differentially private implementation,' see Dwork C, Roth A. The Algorithmic Foundations of Differential Privacy. Foundations and Trends in Theoretical Computer Science. 2014; 9(3): 211-407.

xi. The reader should note that the current FDA 510(k) process has been heavily criticised, see: Heneghan CJ, Goldacre B, Onakpoya I, *et al.* Trials of transvaginal mesh devices for pelvic organ prolapse: A systematic database review of the US FDA approval process. BMJ Open. 2017; 7(7): 1-8.

xii. A relevant concept to assess the need for agility in decision-making is the Observe-Orient-Decide-Act (OODA) loop. This concept describes the process of decision-making and the strategic advantages that swift decision-making and progress though OODA loops might provide, see Osinga PB. Science Strategy and War: The Strategic Theory of John Boyd. Routledge; 2007.

xiii. At the time of writing, the authors are unaware of the publication of any such examination.

xiv. There are also provisions for 'companion diagnostics' and 'genetic tests.' Arguably, these provisions do not constitute a separate route to qualification. Rather, Recital (10) IVDR should be considered an elaboration on how to interpret the standard intended purpose/specific medical purpose route to qualification.

phg foundation
making science work for health

About the PHG Foundation

The PHG Foundation is a pioneering independent think-tank with a special focus on genomics and other emerging health technologies that can provide more accurate and effective personalised medicine. Our mission is to make science work for health. Established in 1997 as the founding UK centre for public health genomics, we are now an acknowledged world leader in the effective and responsible translation and application of genomic technologies for health. In April 2018 we became part of the University of Cambridge.

We create robust policy solutions to problems and barriers relating to implementation of science in health services, and provide knowledge, evidence and ideas to stimulate and direct well-informed discussion and debate on the potential and pitfalls of key biomedical developments, and to inform and educate stakeholders. We also provide expert research, analysis, health services planning and consultancy services for governments, health systems, and other non-profit organisations.

phg foundation
making science work for health

PHG Foundation
2 Worts Causeway
Cambridge
CB1 8RN

+44 (0) 1223 761900

@phgfoundation

www.phgfoundation.org