

Lifecycle Regulation and Evaluation of Artificial Intelligence and Machine Learning-Based Medical Devices

Kerstin N. Vokinger, Thomas J. Hwang, and Aaron S. Kesselheim

1.1 INTRODUCTION

Artificial intelligence- and machine learning (AI/ML)-based technologies aim to improve patient care by uncovering new insights from the vast amount of data generated by an individual patient, and by the collective experience of many patients.¹

Though there is no unified definition of AI,² a good working definition is that it is a branch of computer science devoted to the performance of tasks that normally require human intelligence.³ A major subbranch of this field is ML, in which, based on the US Food and Drug Administration's (FDA) definition, techniques are applied to design and train software algorithms to learn from and act on data.⁴ When intended to diagnose, treat, or prevent a disease or other conditions, AI/ML-based software is a medical device under the Food, Drug, and Cosmetic Act in the United States as well as the Council Directive 93/42/EEC and Therapeutic Products Act in the European Union and Switzerland, respectively.⁵ Examples of AI/ML-based medical devices include an imaging system that uses algorithms to give diagnostic information for skin cancer or a smart electrocardiogram device that estimates the probability of a heart attack.⁶

¹ T.J. Hwang et al., Lifecycle Regulation of Artificial Intelligence and Machine Learning-Based Software Devices in Medicine, 322 JAMA 2285 (2019); M.E. Matheny et al., Artificial Intelligence in Health Care: A Report from the National Academy of Medicine, 323 JAMA 507 (2020).

² M. Hutson, AI Glossary: Artificial Intelligence, in *So Many Words*, 357 Science 19 (2017).

³ A.S. Adamson & H.G. Welch, Machine Learning and the Cancer-Diagnosis Problem – No Gold Standard, 381 N. Engl. J. Med. 2285, 2285–7 (2019).

⁴ *Id.*; see US Food & Drug Admin., Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) (Apr. 2, 2019), www.fda.gov/media/122535/download; G. Hinton, Deep Learning – A Technology with the Potential to Transform Health Care, 320 JAMA 1101 (2018).

⁵ US Food & Drug Admin., Proposed Regulatory Framework, *supra* note 4.

⁶ US Food & Drug Admin., Artificial Intelligence and Machine Learning in Software as a Medical Device, www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device.

Medical devices that are AI/ML-based exist on a spectrum from locked to continuously learning. “Locked” algorithms provide the same result each time the same input is provided.⁷ Such algorithms need manual processes for updates and validation. By contrast, adaptive or continuously learning algorithms change their behavior using defined learning processes. These changes are typically implemented and validated through a well-defined and possibly fully automated process that aims at improving performance based on analysis of new or additional data.⁸

While AI/ML-based technologies hold promise, they also raise questions about how to ensure their safety and effectiveness.⁹ In April 2019, the FDA published a discussion paper and announced that it was reviewing its regulation of AI/ML-based medical devices.¹⁰ The distinctive characteristics of AI/ML-based software require a regulatory approach that spans the lifecycle of AI/ML-based technologies, allowing necessary steps to improve treatment while assuring safety outcomes.

In this chapter, we analyze the regulation of the clearance and certification of AI/ML-based software products in the United States and Europe. Due to the distinctive characteristics of AI/ML-based software, we believe that a regulatory approach is required that spans the lifecycle of these technologies, allowing indicated steps to improve treatment and ensure safety.¹¹ We conclude by reviewing the regulatory implications of this approach.

1.2 CLEARANCE OF AI/ML-BASED MEDICAL DEVICES IN THE UNITED STATES

There is no separate regulatory pathway for AI/ML-based medical devices. Rather, in the United States, the FDA reviews medical devices based on the risks of the devices primarily through the 1) premarket approval pathway (most stringent review for high-risk devices), 2) the 510(k) pathway, or 3) de novo premarket review (for low- and moderate-risk devices).¹² Additionally, the humanitarian device exemption can apply to medical devices intended to benefit patients in the treatment or diagnosis of diseases or conditions that affect fewer than 8,000 individuals in the United States per year.¹³

Premarket approval (PMA) is the most likely FDA pathway for new Class III medical devices. Class III devices are those that support or sustain human life, are of

⁷ US Food & Drug Admin., Proposed Regulatory Framework, *supra* note 4.

⁸ US Food & Drug Admin., Proposed Regulatory Framework, *supra* note 4; Hwang et al., *supra* note 1.

⁹ W.N. Price, Regulating Black-Box Medicine, 116 Mich. L. Rev. 421 (2017).

¹⁰ US Food & Drug Admin., Proposed Regulatory Framework, *supra* note 4.

¹¹ Hwang et al., *supra* note 1.

¹² Hwang et al., *supra* note 1; US Food & Drug Admin., Premarket Notification 510(k), www.fda.gov/medical-devices/premarket-submissions/premarket-notification-510k; US Food & Drug Admin., Premarket Approval (PMA), www.fda.gov/medical-devices/premarket-submissions/premarket-approval-pma.

¹³ US Food & Drug Admin., Humanitarian Device Exception, www.fda.gov/medical-devices/premarket-submissions/humanitarian-device-exemption.

substantial importance in preventing impairment of human health, or which present a potential unreasonable risk of illness or injury. The FDA determined that general and special controls alone are insufficient to guarantee safety and effectiveness of such devices. Thus, such devices require a PMA application to obtain marketing approval. Premarket approval requires the demonstration of “reasonable assurance” that the medical device is safe and effective and generally includes at least one prospective trial.¹⁴ Clearance through the 510(k) pathway is intended for devices for which a PMA is not required (Class I, II, and III devices). In contrast to the PMA, the 510(k) pathway only requires “substantial equivalence” to an already marketed device.¹⁵ The de novo pathway is an alternate pathway to classify novel medical devices that had automatically been placed in Class III after receiving a “not substantially equivalent” (NSE) determination in response to a 510(k) submission. There are two options for de novo classification for novel devices of low to moderate risk. In the first option, any sponsor that receives an NSE determination may submit a de novo request to make a risk-based evaluation for classification of the device into Class I or II. In option 2, any sponsor that determines that there is no legally marketed device upon which to base a determination of substantial equivalence may submit a de novo request for the FDA to make a risk-based classification of the device into Class I or II, without first submitting a 510(k) and receiving an NSE determination.¹⁶ The de novo pathway allows new devices to serve as references or predicates for future 510(k) submissions.¹⁷

A majority of AI/ML-based medical devices are cleared through the 510(k) pathway.¹⁸ However, the 510(k) pathway has been criticized for not sufficiently guaranteeing safety and effectiveness. The 510(k) clearance can lead to chains of medical devices that claim substantial equivalence to each other, but over years or even decades, may diverge substantially from the original device.¹⁹ For example, certain metal-on-metal hip implants were cleared without clinical studies and based on predicate medical devices that did not demonstrate safety and effectiveness or were discontinued.²⁰ Indeed, past clearance of AI/ML-based medical devices can be traced back to other devices that do not have an AI/ML component. For example, the AI/ML-based medical device, Arterys Oncology DL, cleared in 2018, which is

¹⁴ US Food & Drug Admin., Premarket Approval (PMA), *supra* note 12.

¹⁵ US Food & Drug Admin., The 510(k) Program: Evaluating Substantial Equivalence in Premarket Notifications [510(k)] (Feb. 5, 2018), www.fda.gov/regulatory-information/search-fda-guidance-documents/510k-program-evaluating-substantial-equivalence-premarket-notifications-510k.

¹⁶ US Food & Drug Admin., Evaluation of Automatic Class III Designation (De Novo) Summaries (Oct. 27, 2020), www.fda.gov/about-fda/cdrh-transparency/evaluation-automatic-class-iii-designation-de-novo-summaries.

¹⁷ Hwang et al., *supra* note 1.

¹⁸ Id.; U.J. Muehlemaier et al., Artificial Intelligence and Machine Learning Based Medical Devices in the US and Europe (2015–2020) – A Comparative Analysis (accepted at The Lancet Digital Health).

¹⁹ Hwang et al., *supra* note 1.

²⁰ B.M. Ardaugh et al., The 510(k) Ancestry of a Metal-on-Metal Hip Implant, 368 N. Engl. J. Med 97, 97–100 (2013).

indicated to assist with liver and lung cancer diagnosis, can be traced back to cardiac imaging software cleared in 1998, which was considered as substantially equivalent to devices marketed prior to 1976.²¹ The clearance decision does not provide any information regarding clinical validation, and such testing may not have been done.²²

Changes or modifications after marketing of a device requires additional FDA notification and possibly review, either as a supplement to the premarket approval or as a new 510(k) submission.²³ Of course, this is a further challenge for AI/ML devices, since adaptive algorithms that enable continuous learning from clinical application and experience may result in outputs that differ from what has initially been reviewed prior to regulatory approval.²⁴

The FDA publishes summaries of the cleared medical devices' safety and effectiveness as well as statements. However, only rarely does the device description state whether the medical device contains an AI/ML component.²⁵ One example in which this was indicated was BriefCase, a radiological computer-aided triage and notification software that was 510(k) cleared in 2018 and indicated for use in the analysis of nonenhanced head CT images. According to the FDA's summary, BriefCase uses an artificial intelligence algorithm to analyze images and highlight cases with detected intracranial hemorrhage on a standalone desktop application in parallel to the ongoing standard of care image interpretation. The user is presented with notifications for cases with suspected intracranial hemorrhage findings.²⁶ Another example is AiCE (Advanced Intelligent Clear-IQ Engine), an AI/ML-based medical device that was 510(k) cleared in 2020. AiCE is a noise-reduction algorithm that improves image quality and reduces image noise by employing deep convolutional neural network methods for abdomen, pelvis, lung, cardiac, extremities, head, and inner ear applications.²⁷ However, the FDA's summaries and statements do not reveal whether a cleared AI/ML-based medical device contains locked or adaptive algorithms.²⁸ For example, Illumeo System, an image management system software used with general purpose computing hardware to acquire, store, distribute, process, and display images and associated data throughout the

²¹ Hwang et al., supra note 1; Letter from Robert Ochs, Director, US Food & Drug Admin., to John Axerio-Cilies, Chief Operating Officer, Arterys, Inc. (Jan. 25, 2018), www.accessdata.fda.gov/cdrh_docs/pdf17/K173542.pdf.

²² See Ochs, supra note 21.

²³ Hwang et al., supra note 1.

²⁴ *Id.*

²⁵ Muehlematter et al., supra note 18.

²⁶ Letter from Robert Ochs, Director, US Food & Drug Admin., to John J. Smith, Partner, Hogan Lovells US LLP (Aug. 1, 2018), www.accessdata.fda.gov/cdrh_docs/pdf18/K180647.pdf.

²⁷ Letter from Robert Ochs, Director, US Food & Drug Admin., to Orlando Tadeo Jr., Senior Manager, Canon Medical Systems USA (Feb. 21, 2020), www.accessdata.fda.gov/cdrh_docs/pdf19/K192832.pdf.

²⁸ Muehlematter et al., supra note 18.

clinical environment, is promoted as “adaptive” on the manufacturer’s website, but this is not explicitly mentioned in the FDA’s summary.²⁹

1.3 CE MARKING OF AI/ML-BASED MEDICAL DEVICES IN EUROPE

In Europe, there is also no specific regulatory pathway for AI/ML-based medical devices.³⁰ In contrast to the United States, medical products are not approved by a centralized agency. Apart from the lowest-risk medical devices (Class I) that can be carried out under the sole responsibility of the manufacturer, initial review of medical devices of higher-risk Classes (IIa, IIb, and III) are handled by private so-called notified bodies.³¹ In Vitro Medical Devices (IVD) are, based on their risks, either marketed on the basis of the sole responsibility of the manufacturer or handled by notified bodies.³² The EU Member States, EFTA States (Liechtenstein, Iceland, Norway, and Switzerland), and Turkey concluded treaties with regard to the mutual recognition of conformity assessments for medical devices.³³ For simplicity, we use “Europe” to refer to these countries, unless otherwise denoted. Each of these European countries recognize certificates (“Conformité Européenne” [CE] marks) issued by accredited private notified bodies in the other European countries, meaning that after a manufacturer obtains a CE mark in one European country, direct distribution is possible across Europe. Country-specific requirements remain valid, such as mandatory notification for new medical devices, requirements regarding the languages in which the product information must be provided, provisions regarding the prescription and professional use, advertising, reimbursement by social insurances, surveillance.³⁴

Studies show that medical devices are often certified in Europe prior to approval in the United States.³⁵ However, faster access in Europe brings with it important risks that have been well documented. Recent changes to the current European device regulatory system are intended to better safeguard

²⁹ Compare Royal Philips, Philips Illumeo with adaptive intelligence has been selected by University of Utah Health radiologists, Philips News Center (Nov. 26, 2018), www.philips.com/a-w/about/news/archive/standard/news/press/2018/20181126-philips-illumeo-with-adaptive-intelligence-has-been-selected-by-university-of-utah-health-radiologists.html, with Letter from Robert Ochs, Director, US Food & Drug Admin., to Yoram Levy, QA/RA Consultant, Philips Medical Systems Technologies Ltd. (Jan. 12, 2018), www.accessdata.fda.gov/cdrh_docs/pdf17/K173588.pdf.

³⁰ K.N. Vokinger et al., Artificial Intelligence und Machine Learning in der Medizin, Jusletter (Aug. 28, 2017), www.zora.uzh.ch/id/eprint/142601/.

³¹ *Id.*

³² Swissmedic, Guide to the Regulation of Medical Devices, www.swissmedic.ch/swissmedic/en/home/medical-devices/regulation-of-medical-devices/medical-device-regulation_online-guide.html.

³³ *Id.*

³⁴ *Id.*

³⁵ Muehlematter et al., *supra* note 18; T.J. Hwang et al., Comparison of Rates of Safety Issues and Reporting of Trial Outcomes for Medical Devices Approved in the European Union and United States: Cohort Study, 353 *BMJ* 3323 (2016).

patient safety.³⁶ For example, the revised laws (Regulation 2017/745 on Medical Devices [MDR] and Regulation 2017/46 on in vitro diagnostic medical devices [IVDR]) raised the certification threshold for medical products. However, these new laws still do not address AI/ML-based medical devices specifically. Due to the COVID-19 pandemic, the date of implementation of these laws by Member States has been postponed by one year to May 2021 for the MDR and May 2022 for the IVDR.³⁷

In contrast to the United States, Europe does not have a publicly accessible, comprehensive database for certified medical devices and summaries of the regulatory decisions. The EC database on medical devices (Eudamed) is a repository for information on market surveillance exchanged between national competent authorities and the Commission. However, its use is restricted to national competent authorities, the country-specific device regulatory authorities for medical devices, such as Swissmedic in Switzerland.³⁸ In some European countries, for example, Germany, the United Kingdom, or France,³⁹ such authorities have publicly accessible databases for registered medical devices in their country. However, such databases only reflect a fraction of the medical devices CE marked in Europe.

1.4 IMPLICATIONS FOR LIFECYCLE REGULATION OF AI/ML-BASED MEDICAL DEVICES

The traditional paradigm of medical device regulation in both the United States and Europe was not designed for (adaptive) AI/ML technologies, which have the potential to adapt and optimize device performance in real time. The iterative and autonomous nature of such AI/ML-based medical devices require a new lifecycle-based framework with the goal of facilitating a rapid cycle of product improvement and to allow such devices to continuously improve while providing patients' safety.⁴⁰

First, we believe it is important to address the currently limited evidence for safety and effectiveness available at the time of market entry for such products. Both in the

³⁶ *Id.*; A.G. Fraser et al., Commentary: International Collaboration Needed on Device Clinical Standards, 342 *BMJ* 2952 (2011); N. Williams, The Scandal of Device Regulation in the UK, 379 *Lancet* 1789–90 (2012); D. Cohen, Patient Groups Accuse European Parliament of Putting Economic Interests Ahead of Safety on Medical Devices, 347 *BMJ* 6446 (2013); D.B. Kramer et al., Regulation of Medical Devices in the United States and European Union, 366 *N. Engl. J. Med.* 848–55 (2012).

³⁷ European Comm'n, Medical Devices – EUDAMED, https://ec.europa.eu/growth/sectors/medical-devices/new-regulations/eudamed_en.

³⁸ *Id.*

³⁹ BAM, Recherche in öffentlichen Medizinprodukte Datenbanken, www.dimdi.de/dynamic/de/medizinprodukte/datenbankrecherche/; MHRA, Medical Device Manufacturers by Name, <http://aic.mhra.gov.uk/era/pdr.nsf/name?openpage&start=2001&count=1000>; ANSM, Mise sur le marché des dispositifs médicaux et dispositifs médicaux de diagnostic in vitro (DM/DMIA/DMDIV), [www.ansm.sante.fr/Activites/Mise-sur-le-marche-des-dispositifs-medicaux-et-dispositifs-medicaux-de-diagnostic-in-vitro-DM-DMIA-DMDIV/DM-classe-I-DM-sur-mesure-assemblage-Declaration/\(offset\)/5](http://www.ansm.sante.fr/Activites/Mise-sur-le-marche-des-dispositifs-medicaux-et-dispositifs-medicaux-de-diagnostic-in-vitro-DM-DMIA-DMDIV/DM-classe-I-DM-sur-mesure-assemblage-Declaration/(offset)/5).

⁴⁰ US Food & Drug Admin., Proposed Regulatory Framework, *supra* note 4.

United States and in Europe, a majority of the cleared and CE-marked AI/ML-based medical devices have not required new clinical testing.⁴¹ This can deprive patients and clinicians of important information needed to make informed diagnostic and therapeutic decisions. Ideally, AI/ML-based medical devices that aim to predict, diagnose, or treat, should be evaluated in prospective clinical trials using meaningful patient-centered endpoints.⁴² More rigorous premarket assessment of the performance of AI/ML-based medical devices could also facilitate trustworthiness and thus broader and faster access to these new technologies.⁴³ Implementation of AI/ML-based medical devices in clinical care will need to meet particularly high standards to satisfy clinicians and patients. Mistakes based on the reliance of an AI/ML-based medical device will drive negative perceptions that could reduce overall enthusiasm for the field and slow innovation. This can be seen with another AI-fueled innovation, autonomous and semi-autonomous vehicles. Even though such vehicles may be, on average, safer than human drivers, a pedestrian death due to such a vehicle error caused great alarm.⁴⁴ As pointed out in a prior study, it is also crucial to ensure that new regulations help contribute to an environment in which innovation in the development of new AI/ML-based medical devices can flourish.⁴⁵ Thus, the prerequisites for clinical testing must be aligned with the risks of AI/ML-based medical devices.

Second, to address the postapproval period (“surveillance”), manufacturers and the agencies (FDA in the United States, national authorities in Europe) should work together to generate a list of allowable changes and modifications that AI/ML-based medical devices can use to adapt in real time to new data that would be subject to “safe harbors” and thus not necessarily require premarket review. This is especially crucial for devices with adaptive algorithms. Such a “safe harbor” could, for example, apply to modifications in performance, with no change to the intended use or new input type, provided that the manufacturer agrees that such changes would not cause safety risks to patients.⁴⁶ These modifications should be documented in the manufacturer’s change history and other appropriate records. However, modifications to the AI/ML-based medical device’s intended use (e.g., from an “aid in diagnosis” to a “definitive diagnosis”) could be deemed to fall out of the “safe harbor” scope and require submission of a new review.⁴⁷ Depending on the modification, it may be reasonable that a focus of the review lies on the underlying algorithm changes for a particular AI/ML-based medical device.

⁴¹ Hwang et al., *supra* note 1; Muehlemitter et al., *supra* note 18.

⁴² T.M. Maddox et al., *Questions for Artificial Intelligence in Health Care*, 321 *JAMA* 31, 31 (2019); W. Stead, *Clinical Implications of Artificial Intelligence and Deep Learning*, 320 *JAMA* 1107, 1107 (2018).

⁴³ Hwang et al., *supra* note 1.

⁴⁴ Maddox et al., *supra* note 42.

⁴⁵ Price, *supra* note 9.

⁴⁶ Hwang et al., *supra* note 1.

⁴⁷ US Food & Drug Admin., *Proposed Regulatory Framework*, *supra* note 4.

Since even anticipated changes may accumulate over time to generate an unanticipated divergence in the AI/ML-based software's eventual performance, there should be appropriate guardrails as software evolves after its initial regulatory approval. One possibility would be to develop built-in audits for regular intervals using data from ongoing implementation and assessing outcomes prespecified at the time of approval.⁴⁸ Another example would be to implement an automatic sunset after a specific amount of years, such as five years.⁴⁹ This would allow the regulatory agencies to periodically review accumulated modifications and postapproval performance to ensure that the risk-benefit profile for the device remains acceptable.⁵⁰ A stronger focus on the postapproval period is also in line with the FDA's discussion paper that proposes, among other things, that manufacturers provide periodic reporting to the FDA on updates to their software.⁵¹

Lastly, transparency has the potential to improve the usefulness, safety, and quality of clinical research by allowing agencies, regulators, researchers, and companies to learn from successes and failures of products.⁵² It also fosters trust.⁵³ Function and modifications of AI/ML-based medical devices are key aspects of their safety, especially for adaptive software, and should therefore be made publicly accessible. Since modifications to AI/ML-based medical devices may be supported by the collection and monitoring of real-world data, manufacturers should also provide information about the data being collected in an annual report. A further approach to enhance transparency and trustworthiness could be that manufacturers actively update the FDA and European agencies, as well as the public (clinicians, patients, general users) with regard to modifications in algorithms, change in inputs, or the updated performance of the AI/ML-based medical devices.⁵⁴

A stronger focus on transparency should also be pursued by the FDA and European agencies. For example, medical devices that contain an AI/ML component should be indicated as such in the FDA's summaries. The FDA should also clarify in the summaries whether such AI/ML-based medical devices include locked or adaptive algorithms. In Europe, the public does not have access to reviews or summaries of notified bodies or national authorities. National authorities in Europe should adopt the FDA's approach.

⁴⁸ Hwang et al., *supra* note 1.

⁴⁹ *Id.*; R.B. Barikh et al., Regulation of Predictive Analytics in Medicine, 363 *Science* 810, 810–12 (2019).

⁵⁰ Hwang et al., *supra* note 1.

⁵¹ US Food & Drug Admin., Proposed Regulatory Framework, *supra* note 4; T. Minssen et al., Regulatory Responses to Medical Machine Learning, 7 *J. Law & Biosciences* 1 (2020).

⁵² T.J. Hwang et al., Evaluating New Rules on Transparency in Cancer Research and Drug Development, 5 *JAMA Oncol.* 461 (2019).

⁵³ *Id.*

⁵⁴ US Food & Drug Admin., Proposed Regulatory Framework, *supra* note 4.

Medical devices that are AI/ML-based pose new chances and challenges. Current regulations in the United States and in Europe are not designed specifically for AI/ML-based medical devices, and do not fit well with adaptive technologies. We recommend a regulatory approach that spans the lifecycle of these technologies.