Dr. AI, Where did you get your degree?

Edward Raff[1], Shannon Lantzy[1], and Ezekiel Maier[1]

Booz Allen Hamilton
{raff_edward, lantzy_shannon, maier_ezekiel}@bah.com

Abstract. Federal health agencies are currently developing regulatory strategies for Artificial Intelligence based medical products. Regulatory regimes need to account for the new risks and benefits that come with modern AI, along with safety concerns and potential for continual autonomous learning that makes AI non-static and dramatically different than the drugs and products that agencies are used to regulating. Currently, the U.S. Food and Drug Administration (FDA) and other regulatory agencies treat AI-enabled products as *medical devices*. Alternatively, we propose that AI regulation in the medical domain can analogously adopt aspects of the models used to regulate *medical providers*.

Keywords: Regulation · Continuous Learning · Clinical Applications

1 Introduction

Governmental agencies like the FDA are currently drafting regulatory guidance for software as a medical device (SaMD). The FDA's new Digital Health Program is running an eight-company pilot program to pre-certify organizations developing static SaMD for streamlined premarket review. In this paper, we extend the SaMD discussion to a regulatory framework for Artificial Intelligence (AI) as a medical "device." For the purpose of this paper, we define AI-enabled medical device as a software product that actively learns after it is released to the market, and that is intended to inform or make decisions on behalf of a doctor or patient.

There have been instances of promising innovation in the AI-enabled medical device field. For example, in late 2017, Arterys Inc. received FDA clearance for its web-based medical imaging software. However, broader advancement of these medical devices has been stymied by the lack of clear regulatory guidance and FDA approval pathways. The development of clear AI regulation in the medical domain will provide market stability and encourage innovation due to: (1) improved consumer confidence in the safety and efficacy of products; (2) clearer pathways for industry to develop and attain approval for products; (3) availability of standards for use by academics and Institutional Review Boards to design and review studies. One of the main challenges for agencies developing regulation is the recognition that AI is not static. Thus, a novel regulatory schema must account for AI-based devices that actively learn over time. As AI researchers, it is critical we have a voice in how this regulation forms to ground expectations and ensure that innovation is not unduly stifled.

Medical products, such as drugs, biologics, and non-AI devices, have historically been static products that are reviewed using an evidence based evaluation of safety and efficacy. AI-enabled devices upend this traditional regulatory paradigm as these novel devices are dynamic, via continuous learning and updating. The often "black-box" nature of AI has spurred considerable demand for interpretability and explainability in any AI-based medical device [4], and a "right to explanation" has already been enshrined in the European Union's laws [2]. We contend interpretability is excessively burdensome for AI-enabled devices. Regulatory review of medical products traditionally focuses on evidence of *safety* and *effectiveness*, not on interpretability or mechanism of action. To focus on interpretability for AI would stifle progress.

Rather than focusing regulation on algorithms and models, we propose a framework analogous to the standards used to license medical providers. Similar to accredited *medical schools* which train medical doctors, AI-enabled devices should be trained utilizing accredited *data collection and validation methods*. AI devices trained using these accredited data collection and validation methods should be evaluated based on outcomes. Like medical providers, we need an infrastructure to quantify the outcomes for patients, ensuring up-to-date treatment, and sanction of failure cases.

2 Regulation Points

To ensure that the immense potential of AI is not hampered, researchers must actively engage in the development of the regulatory framework. Further, researchers' participation in this discussion will help to avoid the hype and fear that has led to previous AI winters. Therefore, we argue that the methodological accreditation and outcomes-focus framework, outlined below, will enable regulatory agencies to accomplish their mandate of protecting public health, while allowing for innovation by AI researchers.

Accrediting Our Methods While much of the discussion around AI focuses on the algorithms used, data collection and quality is critically important to the success of any model. AI is in no way immune to the "garbage-in garbage-out" problem, and so ensuring that high-quality algorithms are developed means we must ensure data is of an equally high quality. Accrediting the process by which data is acquired and prepared provides the foundation needed for any level of trust in the results. Accreditation should include any intended tasks and applications for which the data will be used, and should minimally consider: an appropriate diversity of patient backgrounds (e.g., age, BMI, etc); a diversity of feature sources (e.g., MRI images used for training must come from multiple MRI machines of differing versions and differing vendors); the consistency of feature sources between the training and clinical contexts; the completeness of data meta-information; defined measurable and clinically-relevant outcomes (e.g., real-time insulin levels), rather than measures that may be available (e.g., unqualified claims records).

As part of Booz Allen Hamilton's organization of the 2016 and 2017 Data Science Bowl competitions [1], which focused on detecting heart function and lung cancer respectively, organizers used these methods to ensure that the competition data was high-quality and resulted in useful algorithms. For example, meta-information describing the hospital that labeled the cardiac MRI images proved to be strongly predictive of a specific heart measurement, despite having no clinical diagnostic power. If this meta-information was not recorded, organizers would not have discovered the correlated, but not actionable feature, and could have led to model overfitting to the training data. This exemplifies why data should be acquired from a diversity of locations, and why trained medical providers must be part of the data preparation process. As one step toward ensuring the safety, AI-enabled devices must be robust to a diversity of input sources. The best way to achieve this robustness is to utilize a diverse high-quality data set for training.

Focus on the Outcomes By their definition, learning algorithms learn from well-defined outcomes which are measured while the device is in use. Therefore, post-market surveillance (i.e., the challenge of monitoring the safety of a medical product after it has been released on the market) is built directly into an AI product. Regulators should focus on the process by which an AI device developer defines, collects, and uses post-market outcomes to refine and improve the model. Similar to a doctor who is subject to review and possible sanctions by her state medical board, regulators should sanction and/or withdraw an AI device from the market for egregious errors.

We propose that, like medical review boards for medical providers, regulators should institute AI review boards consisting of a multidisciplinary group of internal and external experts. The AI boards would include continuing education-like requirements to update AI models using new standards and ground truth data, sanctioning AI producers for errors or AI misconduct or bias, and removal of an AI product when it does harm. Trials and studies will remain necessary to ensure that the device is both safe (does no harm), and effective (provides meaningful and quantifiable improvement in outcomes).

3 So We Treat AI Like A Doctor?

Framing the regulation of AI in the same manner as medical doctors provides a basis for constructing regulation for non-static products. This approach allows regulators, the AI community, and the general public to coherently reason about the opportunities and obstacles of AI-enabled medical devices.

A primary psychological benefit of this approach is to avoid the problem of "moving goal posts." The public is often unwilling to trust a machine to perform a task if the outcome is *only* as good as a human can produce. This thought process ignores the intrinsic benefits of availability and faster decision making. For example, AI-enabled medical devices can provide both routine care in rural and poor communities that would have no access otherwise, and faster

diagnosis, leading to improved patient outcomes. With regulation focused on data accreditation and clinical outcomes, regulators avoid unnecessarily delaying adoption of AI technology for medicine.

This regulatory framework also provides guidance on ensuring AI devices remain safe over time. Physicians are not simply told to do no harm. Rather, physicians progress from interns to specialist over their careers, and as they progress their responsibilities and autonomy increases. AI devices should follow a similar (task-dependent) progression. However, AI devices need not progress completely to autonomous continually learning agents (i.e., a specialist). Instead AI devices can ultimately be tools, which have utility to physicians irrespective of their autonomous continually learning capability.

With this regulatory approach we must collectively recognize that errors and mistakes will be made. Doctors currently, and eventually AI will, unintentionally kill patients. Deaths caused by software bugs have already occurred [3] and were incidents that the FDA studied in order to remediate and prevent future incidents. While the use of AI devices promises to reduce the frequency of such unfortunate incidents, the same lessons will apply to the AI space. Researchers who acquire and prepare the data, and develop models to analyze and act on it must understand this risk. Given the potential greater autonomy of AI-enabled devices, AI developers may require a form of "malpractice" insurance. This insurance would provide fiscal and regulatory incentives to encourage safety and provide financial recompense when incidents occur.

4 Conclusion

Regulatory agencies are currently developing policy and guidance for static SaMD, and will soon codify rules to govern dynamic SaMD (i.e., AI medical devices). Rather than developing new regulations based on our existing rules for static medical products, we have proposed a novel regulatory framework for AI-enabled devices that is analogous to the approach used to accredit medical doctors. We argue this regulatory framework provides a natural paradigm to address the public's concerns about the use of AI in healthcare. Though the accreditation process for medical doctors is not perfect, the approach has served society for decades and can serve as the foundation for regulating AI-enabled medical devices.

References

1. Data Science Bowl (2018), `https://datasciencebowl.com/`
2. Goodman, B., Flaxman, S.: European Union Regulations on Algorithmic Decision-Making and a Right to Explanation. AI Magazine **38**(3), 50 (10 2017). https://doi.org/10.1609/aimag.v38i3.2741
3. Leveson, N., Turner, C.: An investigation of the Therac-25 accidents. Computer **26**(7), 18–41 (7 1993). https://doi.org/10.1109/MC.1993.274940
4. Lipton, Z.C.: The Doctor Just Won't Accept That! In: Interpretable ML Symposium at NIPS (2017), `http://arxiv.org/abs/1711.08037`