Institutional Review Boards as Soft Governance Mechanisms of R&D: Governing the R&D of AI-based Medical Products

Toni Lorente, Kings College London and the Future Society

Abstract

Risk-based approaches to governance bear an ambiguous stance regarding the R&D stages of AI, for they the possibility of explicit risks before they are posed by a given finalized product. In this context, Institutional Review Boards (IRBs) stand as unique governance mechanisms, capable of addressing the step from general research to concrete product development. However, IRBs face several challenges in governing AI-based medical products, including: (a) achieving consistency, (b) being exhaustive, (c) ensuring process transparency, and (d) reducing the existing capacity and knowledge asymmetry between different stakeholders.

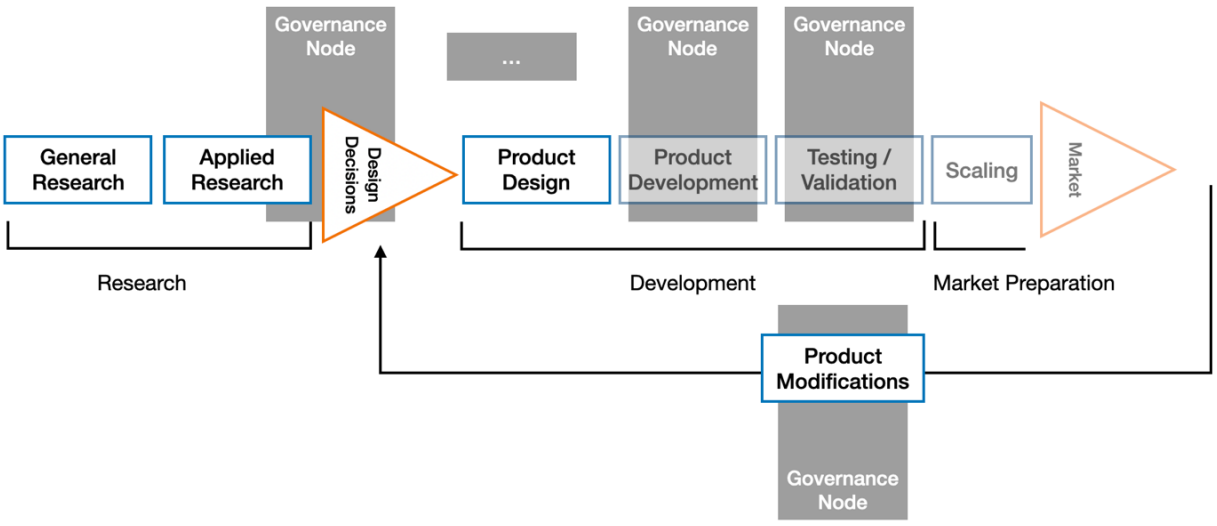
This article explores four governance levers that can be used to effect change, four governance entry-points throughout a product's lifecycle, and five different behaviors that IRBs should try to advance to ensure the effective governance of the R&D stages of AI-based medical projects. In doing so, IRBs can seize the unique opportunity they present to bring principles into practice, increase research quality, reduce governance costs, and bridge the knowledge gap between stakeholders.

1. Introduction

Institutional Review Boards (IRBs) have played a critical role in providing a channel to translate research ethics standards (such as The Belmont Report [1978] or the Declaration of Helsinki [1964]) into actual practices, overseeing research proposals in healthcare for over 6 decades (Friesen et al., 2021:36). However, the efforts to include Artificial Intelligence (AI) across all sectors is pushing the limits of competence of IRBs when it comes to evaluating AI-based medical products, presenting unique challenges and opportunities for these boards to effectively govern the development of AI.

By considering the lifecycle of AI research projects and the current regulatory context (which is heavily leaning towards risk-based approaches), this article discusses IRBs as unique governance mechanisms to tackle AI at the research and development (R&D) stages – that is, the process of developing algorithms and architectures, training models, and developing them into concrete products or services.

Figure 1: High-level representation of the product development pipeline to illustrate key stages of R&D, including the design decisions that divide research from development stages and different governance nodes.



Over the last years, and with AI becoming a promising avenue in healthcare research, IRBs' mandate has been expanded to incorporate AI-enabled medical products and research proposals as part of their scope. But the rapid development within the field of AI is imposing a heavy burden on IRBs that manifests via different challenges these boards face – of which this article considers four: (a) achieving consistency, (b) being exhaustive, (c) ensuring process transparency, and (d) addressing the existing capacity and knowledge asymmetry between certain private actors leading the development of AI and some independent evaluators. Failing to address these challenges may not only result in IRBs playing a role for which they have not been given the tools and resources to fulfill but could also mean the loss of a unique governance opportunity for AI research in healthcare (given that, in most jurisdictions, the risks and dangers stemming from AI systems tend to be defined after systems are deployed).

To address these challenges, this article develops in 8 stages. Section 1 provides a brief introduction. Section 2 presents IRBs and their broad regulatory context in the EU and the US. Section 3 presents the four main challenges that IRBs face to remain effective in governing the safe development of AI-enabled medical products. Section 4 discusses IRBs in the AI context as critical governance mechanisms due to their pre-deployment scope. Section 5 presents four different governance levers or paths and mechanisms to ensure effective governance (i.e., monitoring, harmonizing, observing norms and standards and training) and their sources of authority. Section 6 presents four governance nodes or entry-points throughout a product's lifecycle (i.e., the steps from general research to proposal drafting, from project proposal to product development, results publication, and modifications or updates). Section 7 identifies five key behaviors that different actors must include to empower and equip IRBs. Finally, Section 8 compiles a set of opportunities for key stakeholders, as well as related risks and recommendations.

2. Problem & Context

During the 1980s and 1990s, some of the world's leading economic powers went through a de-regulatory process that was, in part, triggered by a political move seeking to reduce the burden and overall costs of regulation industry. Across the UK, Europe, and US, different political waves tipped the scales towards a leaner, evidence-based, and industry-inspired regulatory approach. In this context, the private sector became the benchmark and gold standard for new regulatory proposals, percolating cost-benefit analyses, "objective and transparent" and, ultimately, risk-based approaches into the public sector (Hutter, 2005:1-3).

Since then, the power of the "central government" has been fragmented and crystallized into industry-specific or independent regulatory agencies, and alternative forms of self-regulation have been encouraged (Hutter, 2005:3). In parallel, succeeding industry-led technological booms throughout the last decades have reinforced this multi-channel approach to governance, where different mechanisms and stakeholders control different risks entailed by a given technology.[1]

When the concept of "risk" was borrowed and industry adopted it as the backbone of some new approaches to governance, the lack of a neatly defined notion of risk in the legal context gave rise to several questions — most of which remain unanswered. In this sense, and in the European context, for example, regulatory crutches like tort law or product liability directives may have to cover the gap left by how the criticality of the risks related to AI are divided (Chamberlain, 2023:1). This is particularly important considering how the efforts by the industry to govern technological progress have prioritized operational or those hazards faced by the industry over those faced by other stakeholders.

When it comes to EU's AI-healthcare applications, efforts have been mostly directed towards developing an infrastructure capable of supporting forthcoming requirements to evaluate AI-enabled

[1] There are, nonetheless, historical anomalies such as the GDPR, where the regulator took a rights-based approach to data management and data protection. Even though there are guidelines and requirements to perform risk assessments, the underlying idea is to clearly state and operationalize the citizens' rights in relation to their data.

medical products. In this sense, both the European Medicines Agency (EMA) and the European Commission (EC) have published different reports[2] laying out the priorities and objectives for Europe. However, all medical devices – including those enabled by AI – are subject to and must, comply with existing EU regulation (Aggarwal et al., 2022:10). In practice, IRBs have been stretched to include and evaluate AI-based research projects without necessarily having the skillset to do so, or without having a clear mandate and strict protocols to address these types of projects.

Moreover, the EU has notable variation in terms of regulation, applicable laws, and other relevant practices among different countries, which can sometimes lead to a lack of clarity for international projects (illustrating the coherence and consistency challenges discussed in Section 3) and, in some cases, to inefficiencies, highlighting the need for greater procedural harmonization not only to reduce costs but also to increase the quality of the evaluation and prevent resource waste (Timmers et al., 2020:2).

In the US, the FDA is responsible for regulating IRB's and how they address AI research projects in healthcare.[3] In 2017, the FDA adopted the principles governing software as medical device (SaMD) developed by the International Medical Device Regulators Forum IMDRF, a voluntary group of medical device regulators. These were later developed into several Food and Drug Administration (FDA) texts which covered, in part, the safe and effective development of AI-based software. [4] As a result, AI-based software developed with the aim to treat, diagnose, cure, mitigate, or prevent disease or other conditions is considered as a medical device.

In this context, soft approaches to governance have been central and, in most cases, have spearheaded best practices in responsible research and innovation for cutting-edge technology. An important reason is that these governance approaches tend to be more agile than hard law while simultaneously requiring the direct involvement of the actors leading the path, partially addressing the pacing problem by closing the gap between regulatory efforts and technological progress (Marchant, 2011:19). This high degree of industry involvement is one of the main advantages of soft governance approaches as it streamlines processes and narrows the scope of specific measures or bodies. However, it can also water down the types of risk addressed in voluntary or self-regulating protocols as they usually gravitate around operational risks that are covered by existing standards and norms that validate, in turn, their own processes.

AI, nonetheless, is a type of general-purpose technology that has the potential to be revolutionary (or a type of technology that supports a fundamental transformation in the nature of economic production; Garfinkel, 2022:2). Before this possibility, self-governance and industry-wide soft governance efforts will play a crucial role in operationalizing the guidelines and objectives defined collectively, bridging the distance between the government or regulator and the organizations developing AI. But because of AI's transformative potential, soft governance mechanisms must be coordinated and thoroughly scrutinized to ensure that they tackle the risks and threats relevant to the general public (beyond operational or corporate risks) and, ultimately, contribute to operationalize the underlying rights, values, and principles that constitute each society.

A rapid response mechanism to address the rising trend of AI-based research in healthcare has been to expand the scope of IRBs. However, IRBs were not originally formed to review AI-based projects. Considering this, and in the current regulatory context where AI is governed on a risk-based approach (with a pervasive focus on the risks that deployed systems entail), the legacy of IRBs' structures and processes, some of the concept definitions used (e.g., regarding human subjects or what constitutes

[2] See https://www.ema.europa.eu/en/documents/minutes/hma/ema-joint-task-force-big-data-summary-report_en.pdf and https://commission.europa.eu/document/d2ec4039-c5be-423a-81ef-b9e44e79825b_en

[3] See Section 201(h) of the "Federal Food, Drug, and Cosmetic Act" (FD&C Act) https://www.fda.gov/regulatory-information/laws-enforced-fda/federal-food-drug-and-cosmetic-act-fdc-act_

[4] See https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device.

publicly available data), and a lack of resources and protocols to enforce their mandate could result in the challenges that IRBs face surpassing the governance opportunities they present.

3. Main Challenges

According to the FDA's definition, IRBs are appropriately constituted groups that are:

> "*[…] formally designated to review and monitor biomedical research involving human subjects. In accordance with FDA regulations, an IRB has the authority to approve, require modifications in (to secure approval), or disapprove research. This group review serves an important role in the protection of the rights and welfare of human research subjects.*
>
> *The purpose of IRB review is to assure, both in advance and by periodic review, that appropriate steps are taken to protect the rights and welfare of humans participating as subjects in the research. To accomplish this purpose, IRBs use a group process to review research protocols and related materials (e.g., informed consent documents and investigator brochures) to ensure protection of the rights and welfare of human subjects of research.*"

As such, IRBs present different opportunities to govern the use and development of AI-enabled products and services in healthcare: on the one hand, the *raison d'être* of IRBs is to ensure that a set of different requirements, especially ethical guidelines, are observed in research projects, posing a distinctive opportunity to operationalize and embed ethically relevant discussions into the lifecycle of AI research projects. On the other hand, and closely related to this point, IRBs act upon different moments of the project's lifecycle or governance nodes, offering a unique longitudinal perspective over research projects. But, at the same time, and because of their almost exceptional status, IRBs face several challenges when assessing the risks entailed by AI-enabled research projects.

In December 2022, the Ada Lovelace Institute published a report exploring the role of IRBs in evaluating AI-enabled research from an ethical standpoint to identify the main challenges these boards face. To do so, they conducted a literature review on the challenges faced by IRBs as well as several workshops with members of IRBs and researchers working on AI ethics (Ada Lovelace Institute, 2022:6). From their findings, and in the context of this article, three challenges stood out:

- **Consistency**: An increasing interest in AI intersects with the nature of research projects in the domain of healthcare which are, in many cases, the product of the efforts from different actors and, sometimes, public-private partnerships. This is likely to trigger multiple evaluations of the same research project, highlighting in turn some of the challenges of governance and consistency between the standards observed among different IRBs. (Ada Lovelace Institute, 2022:7) This is not only problematic but could also lead to the "gamification" of ethical reviews, exploiting the discrepancies among IRBs to the benefit of the promoter by selecting the better-suiting committee according to the project. Furthermore, private organizations are more and more well-positioned to conduct research exclusively overseen by their own "advisory IRB", circumventing traditional review processes.

- **Exhaustiveness**: Many of the harms entailed by AI-enabled systems are only observable after the system has been deployed into the real world. However, IRBs are de facto tasked with assessing AI-enabled research projects *before* they are finalized, making it virtually impossible to predict some of the effects of the resulting product. Moreover, certain types of harms (spanning from bias to misuses of AI, but also in terms of the broader impacts of AI research in society) are hard to predict by a single committee (Ada Lovelace Institute, 2022:7). The lack of tools to predict some of these harms and protocols to address wider societal concerns increases the chance of involuntary ethics washing, stressing the need to develop and coordinate thorough protocols.

- **Process transparency**: Private IRBs are closer to the ground truth, for they usually have privileged information about the research. The privilege is protected by trade secrets and NDAs, but results in a lack of transparency with respect to their processes that raises concerns from the research community and other stakeholders, ultimately putting at risk the efficacy of public-private research partnerships (Ada Lovelace Institute, 2022:7-8). However, they serve on an advisory role, leaving the final decision to researchers or project managers. In this sense, a lack of transparency perpetuates the knowledge gap between research promoters and evaluators and fosters an information asymmetry between independent evaluators and research promoters.

But other than these, IRBs must also face:

- **Capacity Asymmetry**: Some private companies have the means to privately develop systems or services relying on powerful AI models that would qualify as medical products. These are often developed using publicly available data or data that is not necessarily private in a strict medical sense. The impact that these can have on society is tangible but given the limited mandate of IRBs and the lack of process transparency by private organizations, their R&D efforts are rarely overseen beyond a company-owned ethics review committee (Friensen et al., 2021: 37-8). Ultimately, this poses a direct threat to the credibility and effectiveness of IRBs as a mechanism to govern R&D.

Addressing these challenges is crucial to ensure IRBs remain effective in governing the safe development of AI-based medical products and ability to provide a unique governance entry point at the R&D stages of AI.

4. Harnessing R&D: A Deeper Look into the Context

AI R&D has boomed over the last decade, attracting talent and capital from all over the world. This boom has led to a substantial growth of the AI market and, therefore, the potential impact of AI over everyone's lives. To address this increased potential impact, several AI-related Bills have been passed, going from 1 in 2016 to 37 in 2022 across 127 countries (Maslej et al., 2023:268).

However, AI R&D remains an under-regulated field, largely from the rapid progress the field has experienced in the last years, which clashes with longer timelines for hard regulation and the knowledge asymmetry between the regulators and AI developers. AI's transformative potential is sure but still indeterminate, and the risks derived from leaving research and development of cutting-edge systems be governed by market interests are unknown.

The interest on the benefits and risks of GPT technology in chatbots applied to medicine, (Lee et al., 2023) or more ambitiously on the paradigm shift that foundation models can imbue into medicine and healthcare (Moor et al., 2023) is prescient. Yet the opacity throughout the research process and the corporate interests affecting the decision on which product to develop are not transparent, especially given the private nature of the actors leading the technological progress.

Google, Microsoft, Anthropic, and OpenAI announced the constitution of the "Frontier Model Forum" to "*promote the safe and responsible development of frontier AI systems: advancing AI safety research, identifying best practices and standards, and facilitating information sharing among policymakers and industry.*"[5] And while this announcement shows signs of some responsible research practices, it also strengthens the case for more blunt action upon the R&D stages of AI systems.

More concretely, IRBs are well positioned to affect one of the most crucial decision-points within the early stages of AI's lifecycle: the step from research to development. For even though it is common to concatenate research and development almost like a mantra, sometimes even blurring the difference, the truth is that the step from a strictly scientific or technological research to specific product development with a real-world application not only entails several decisions by different actors that are strongly political in nature (in the sense that they ultimately determine how the public

[5] Verbatim from the press release by OpenAI, available at https://openai.com/blog/frontier-model-forum

is affected as well as the nature of the benefit they afford) but is, perhaps, one of the key dividing moments in an AI-product lifecycle.

This division is significant, for even though there are some risks tied to general research[6] (e.g., in developing new algorithms or expanding capabilities, particularly for larger models), specific risks to society are ultimately determined by the decisions made to develop the technology into a product and, ultimately, risk-based approaches to regulation and governance apply, by definition, to existing products that entail real risks. Yet there is a major difference between merely explorative research and embedding its findings into concrete products or technologies. Considering certain aspects (or failing to do so) during the step from research to development and project definition determine some of the risks that the end-of-the-line product shall pose to society, being a critical decision point or governance node in the project lifecycle of an AI-based product.

Hence, IRBs occupy a privileged position within the project lifecycle. Namely, IRBs oversee research projects regardless of the promoter – that is, they are the universal and independent gatekeepers – and do so from an early stage of the project's lifecycle. This confers IRBs' unique governance role in the AI in healthcare landscape. However, in the current context where the mechanisms to embody principles into practices are still under construction, IRBs are exposed and in need for protocols and material means to address the knowledge gap between the promoters and the evaluators.

The universality of IRBs, however, hints at the potentially disparate impact that longer review processes can have on smaller and medium-sized organizations doing research. Given the nature, composition, and workload of IRBs, review processes can take weeks or months, whereas research at the forefront of AI and, even more so, in the start-up context usually takes weeks, if not days (Molnar et al., 2023). Thus, the greater impact that compulsory review processes could have on start-ups and smaller and medium enterprises (in comparison with less money-constrained organizations) should be considered when drafting best practices and procedures for IRBs. This would not only ensure excellence in innovation for all parties involved but would also contribute to reducing power concentration by ensuring competitiveness among all types of organization.

5. Governance Levers and Sources of Authority

Governance levers include the different paths or mechanisms by virtue of which effective governance is induced – which are, in turn, closely tied to the source of authority on which they stand. For AI-enabled medical products, we can distinguish at least four governance levers and their respective sources of authority:

a) **Monitoring**: This lever does not refer to the activity conducted by IRBs on specific projects, but to the effort by the central authority to oversee that the practices followed by IRBs are compliant with their statutes and underlying laws, and that such practices address the risks that AI-enabled medical products pose. In the American context, monitoring is ultimately conducted by the FDA, the authority that has the power to declare and cease IRBs.

b) **Harmonizing**: In a similar line, the FDA has the role to harmonize the efforts put forward by different IRBs, making sure that all review boards observe standards equally, reducing the probability of promoters gamifying the review process.[7] On the other hand, if State Laws generate asymmetries, harmonizing efforts are intended to address them.

[6] Especially for Frontier Models, (Bommasani et al., 2022) the emergent capabilities of which cannot be predicted before training and could lead to a change of paradigm, but also related to broader notions of safety, control, or privacy.
[7] Organisms like the International Conference on Harmonisation (ICH) produce guidelines for global pharmaceutical development (which apply to Europe, USA, Canada, Japan, and Switzerland). These guidelines are intended to reduce duplication of clinical trials and improve the efficiency of assessment processes. Creating a discussion group within ICH or finding a way to display a similar approach to AI-based medical products could facilitate the safe circulation of AI-based medical products by strengthening the harmonization governance lever.

c) **Observing Norms and Standards**: Even though this lever consists of IRBs being familiar with relevant norms and standards and making sure researchers observe them, the underlying authority of this lever comes from the international bodies that develop best practices and desirable behaviors into norms and standards.

d) **Providing Training**: By means of disseminating the desired outcomes and standards across researchers and stakeholders, both the central authority (i.e., the FDA) and the different organisms or IRBs set the bar and ethical guidelines, which ultimately percolates into the researchers' best practices.

6. Governance Nodes

Governance nodes are the different decision points throughout a project's lifecycle, in which a stakeholder can effect change and/or steer the stream of work, altering the behavior of other actors and, ultimately, the end-result. IRBs, via different governance levers, and by virtue of their processes and protocols, have access to different moments of a project's lifecycle and can therefore shape the end-result with their decisions.

For AI-enabled projects, one of the most critical governance nodes is the moment in which general research is crystallized into a project proposal to develop a specific product. This draws the line between explorative research and concrete, market-facing development efforts, bringing in turn a unique opportunity to land principles into concrete red lines and practices to be observed during the following research.

In the report developed by The Ada Lovelace institute, a multi-stage ethics review process is proposed not only to break down the complexity and burden of the review process, but also to make justice to the governance nodes that appear throughout a given project's lifecycle, from the first draft designing the research protocol to post-publication reviews to understand how the evaluations made by the IRB have altered the results (Ada Lovelace Institute, 2022:77-80). On the other hand, the transition from "locked" models (or those models that, once developed, are not modified) to continuous learning approaches to AI and periodic modifications justifies this break down to streamline updates.

In this sense, the recommendation is to clearly determine the governance nodes in which an AI-enabled product or service is to be evaluated by the relevant IRB, including:

a) **Background**: In which the need for an AI-enabled solution affecting or relying on the use of real health data is justified and linked to theoretical results deriving from fundamental research.

b) **Product Development**: In which not only the scope of the product, but also the performance conditions or ethical considerations under which its release will be halted (e.g., unequal performance across protected groups, or unacceptable failure modes).

c) **Results Publication**: In which relevant ethical considerations (both in terms of safety and possible harms to society) are systematically evaluated and included in the final report and/or product description.

d) **Relevant Modifications**: In which the processes to re-evaluate and allow future modifications, fine-tuning, or continuous learning is addressed (for example, as intended by the Predetermined Change Control Plan for AI/ML-enabled device software functions developed by the FDA which is currently under development).

These nodes apply to any AI-based medical product, yet some private organizations are capable of developing services and systems that would in practice qualify as medical products while circumventing traditional ethical review processes. These, even when overseen by company-owned ethical review committees, pose a structural threat to the trustworthiness of IRBs to effectively govern the R&D of AI in healthcare. As such, regulatory agencies should revisit foundational texts like the Common Rule, where critical terms used to demarcate the projects that need an independent ethical review are defined, with the aim of incorporating some of these private efforts. Doing so is

particularly important considering the weight that "publicly available" data is playing in the development of privately owned AI-based services, which can ultimately have a direct impact on the population and that, in many cases, make use of medical or medical-adjacent data – drawing again the line between general wellness products and medical devices.

7. **Target Behaviors and Actors**

One of the challenges of IRBs as governance mechanisms for R&D is that we do not speak of a single actor, but of a network of independent boards. In this sense, governance comprises both public or private and individual or collective actors, and includes hard governance or regulation, industry-wide self-governance, and company self-governance.[8]

The FDA grants the status, but the power is independently wielded by each board. Due to their proximity with R&D, boards and members have a direct impact on the final products delivered into the market from early stages. But to effectively activate the different governance levers, IRBs must be equipped to:

a) **Coordinate**: Mainly to avoid research promoters gamifying the review process. Even though IRBs are, by definition, independent, individual boards should be equipped and with universal basic means to ensure the preparedness of any board to address AI-based products. This is the instrumental to prevent IRBs becoming the enabler of ethics-washing efforts.

b) **Seek Coherence**: Through the coordination efforts mentioned above, the need for coherence with a set of foundational principles arises. In this sense, overarching bodies (like the FDA) should clearly provide the baseline principles on which specific protocols operate, and require the inclusion of ethics experts, product engineers, or tech lawyers among others in IRBs to ensure that protocols are implemented observing such principles. These need not to be original or new but establish the foundation of any activity conducted by IRBs.[9]

c) **Seek Consistency**: On the one hand, pursuing internal consistency contributes to effective evaluations throughout the different governance nodes (e.g., between the project definition stage and a subsequent review). Decisions should be traceable and expandable throughout the different governance nodes, and the nature of these should respond to each stage of the project (i.e., listing potential risks in the background or first draft of the project, versus voicing concrete ethical concerns in the later stages). On the other hand, protocols and principles should be derived from a well-curated list of internationally relevant standards and best practices, both to ensure interoperability, international compatibility, and prevent an unnecessary *ad hoc* and heterogeneous maze of resources used by different boards across a given jurisdiction.

d) **Limit the Scope**: On the one hand, to ensure that the approval of a given product does not lead to the *de facto* approval of any subsequent modification of such product. On the other hand, to ensure that the IRB model is not stretched, as some point it already is (Friesen et al., 2021:36). By delimiting the mandate and the capabilities required to execute it for IRBs, upcoming needs will be more easily identified and potentially addressed.

   While IRBs have the capacity to withhold approval based on ethical considerations, it would be advisable to explicitly state their capacity to request changes in the fundamental research proposal (for example, asking for a less-invasive approach or technique to

[8] Cf. Auld et al., 2022 for a discussion of motivations and actions driving AI Governance

[9] They could be inspired, for example, by the US Whitehouse's Blueprint for an AI Bill of Rights, which includes algorithmic discrimination protections, data privacy, notice and explanation, and human alternatives, consideration and fallback. https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf

develop a ML model), which would be consistent with the mandate of IRBs by the FDA to "*require modifications (to secure approval).*"

e) **Seek Transparency**:[10] Even though transparency is mainly instrumental (that is, it does not stand as a value but as a means to an end), ensuring individual boards seek transparency is a means to first, reduce coordination costs among different boards and second, close the knowledge gap between project promoters and evaluators. In this case, transparency does not aim to facilitate algorithm explainability *per se* (that is, a better technical comprehension by a lay or non-involved audience) but rather capacity building and effective evaluation throughout IRBs.

8. Conclusion

IRBs occupy a unique position as governance mechanisms for AI-based medical products, for even though they exist in a regulatory context based on the risks posed by AI, they are conceived to voice ethical concerns during or even before research projects are executed. And while this could well mean that IRBs' mandate is being stretched beyond what these boards were initially designed to achieve, it can also be understood as an anomaly that presents four unique opportunities, with their respective risks and recommendations.

First, IRBs offer an opportunity to embed ethical principles into material governance protocols from the earliest stages of research. Nevertheless, and given the work required in terms of coherent guidelines and processes, review processes could lead to ethics washing if members of each one of the individual boards are not adequately equipped or surrounding norms and regulation lack teeth. To address this, regulatory bodies (like the FDA) should implement and enforce a modular approach to reviewing AI-based projects by experts, limiting the scope of the Predetermined Change Control Plan or PCCP[11]. Moreover, this proposal could also encourage the gamification of the review process, especially in light of weak coordination among the nodes constituting the IRB network. To prevent such gamification, the FDA should ensure common baselines and provide support to different boards, fostering harmonization.

Second, IRBs offer a chance to prove the increase in quality derived from overseeing R&D in AI and, potentially, to propagate the relevance of IRBs for AI R&D beyond the domain of healthcare. However, and given the nature of the review process, key stakeholders (mainly larger private organizations) could potentially find ways of circumventing or deferring IRBs' oversight, creating greater asymmetries among different types of organizations and reducing the impact of this governance proposal. To address this, the signatories of the Common Rule should redefine critical concepts to ensure that target stakeholders are also reviewed by IRBs (for example, by instead of focusing on the nature of data exclusively, considering the scale or aggregation of databases as critical elements to trigger independent reviews).

Third, IRBs can be a way of reducing overall governance costs by legitimizing themselves as standposts of R&D best practices. However, this could end up stretching even further the scope and mandate of IRBs, rendering them either as mere tick-boxes in an exercise of ethics washing or, at the other side of the spectrum, as the wardens of R&D excellence. To prevent this, the organizations signing the Common Rule should limit and update the scope of IRBs, as well as provide them with tools and ensure a suitable composition, in accordance with the state of the art of AI.

Fourth and last, IRBs can tackle the pacing problem by bridging the gap between private organizations developing cutting-edge technology and public and/or independent boards reviewing their research efforts. Moreover, the lessons learned could be aggregated and used as input to develop codes of conduct and governance best practices to foster interoperability and global coordination. This, however, faces two risks.

[10] See Larsson & Heintz, 2020 for a more extensive discussion of transparency's role in AI.
[11] See https://www.regulations.gov/docket/FDA-2022-D-2628

On the one hand, the inherent asymmetry of power and interests between private actors and IRBs leaves most of the agency in the hands of the first, with limited incentives and unclear benefits. To address this, regulators should foster process transparency by establishing collaboration frameworks where key companies developing cutting-edge AI technology and IRBs could interact. This would practically act as a sandbox, where AI companies would benefit from the input of external and independent experts and IRBs would ensure their relevance as technology advances.

On the other hand, smaller and medium sized companies subject to the longer review processes of IRBs could see their capacity to innovate and to compete with larger organizations hampered. This would cause an asymmetry, given that larger organizations can usually absorb the cost of extended timelines more easily, disincentivizing innovation among the smaller companies and increasing power concentration among the bigger ones. To avoid this, and to ensure competitiveness, best practices and procedures for IRBs should account for the potential disparate impact that extended timelines could have, establishing clear and defined timelines to include them in research planning and agile response mechanisms to address time-sensitive reviews.

Establishing a clear framework to govern research, development, and other key nodes in the life cycle of AI-based medical products through IRB oversight will require seizing the opportunities discussed above while addressing the risks discussed. Doing so will not only contribute to the governance map with a unique pathway to address issues related to the research and development phases of AI-based medical products but will also contribute towards developing trust and excellence in research, taking the quality of the next generation of AI products one step further.

References

Ada Lovelace Institute. (2022). Looking before we leap: Ethical review processes for AI and data science research. Available at: https://www.adalovelaceinstitute.org/report/looking-before-we-leap/

Aggarwal, N., Matheny, M. E., Shachar, C., Wang, S. X. Y., & Thadaney-Israni, S. (2022). Artificial Intelligence in Healthcare. In The Oxford Handbook of AI Governance. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197579329.013.50

Auld, G., Casovan, A., Clarke, A., & Faveri, B. (2022). Governing AI through ethical standards: learning from the experiences of other private governance initiatives. *Journal of European Public Policy*, *29*(11), 1822-1844. https://doi.org/10.1080/13501763.2022.2099449

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*. https://doi.org/10.48550/arXiv.2108.07258

Chamberlain, J. (2023). The risk-based approach of the European Union's proposed artificial intelligence regulation: Some comments from a tort law perspective. *European Journal of Risk Regulation*, *14*(1), 1-13. https://doi.org/10.1017/err.2022.38

Friesen, P., Douglas-Jones, R., Marks, M., Pierce, R., Fletcher, K., Mishra, A., ... & Sallamuddin, T. (2021). Governing AI-Driven Health Research: Are IRBs Up to the Task? *Ethics & Human Research*, *43*(2), 35-42. https://doi.org/10.1002/eahr.500085

Garfinkel, B. (2022). The Impact of Artificial Intelligence. In The Oxford Handbook of AI Governance (p. C5.S1-C5.N23). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197579329.013.5

Hutter, B. M. (2005). *The attractions of risk-based regulation: accounting for the emergence of risk ideas in regulation* (Vol. 33). London: CARR.

Larsson, S., & Heintz, F. (2020). Transparency in artificial intelligence. *Internet Policy Review*, *9*(2). https://doi.org/10.14763/2020.2.1469

Lee, P., Bubeck, S., & Petro, J. (2023). Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. *New England Journal of Medicine*, *388*(13), 1233-1239. https://doi.org/10.1056/NEJMsr2214184

Marchant, G. E. (2011). *The growing gap between emerging technologies and the law* (pp. 19-33). Springer Netherlands. https://doi.org/10.1007/978-94-007-1356-7

Maslej, N., Fattorini, L., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Ngo, H., Niebles, J.C., Parli, V., Shoham, Y., Wald, R., Clark, J., & Perrault, R., *The AI Index 2023 Annual Report*, AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2023. Available at: https://aiindex.stanford.edu/report/

Molnar, A., Stanley, D., & Valeriani, D. (2023). Neurotechnology, Stakeholders, and Neuroethics: Real Decisions and Trade-Offs from an Insider's Perspective. In *Policy, Identity, and Neurotechnology: The Neuroethics of Brain-Computer Interfaces* (pp. 271-283). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-26801-4_15

Moor, M., Banerjee, O., Abad, Z. S. H., Krumholz, H. M., Leskovec, J., Topol, E. J., & Rajpurkar, P. (2023). Foundation models for generalist medical artificial intelligence. *Nature*, *616*(7956), 259-265. https://doi.org/10.1038/s41586-023-05881-4

Timmers, M., Van Dijck, J. T., Van Wijk, R. P., Legrand, V., Van Veen, E., Maas, A. I., ... & Kompanje, E. J. (2020). How do 66 European institutional review boards approve one protocol for an international prospective observational study on traumatic brain injury?

Experiences from the CENTER-TBI study. *BMC medical ethics*, *21*, 1-14. https://doi.org/10.1186/s12910-020-00480-8