The EU Proposal for Regulating AI: Foreseeable Impact on Medical Robotics

Maria-Camilla Fiazza[1]

***Abstract*— In April 2021, in the wake of a number of preparatory documents, the European Commission published a proposed regulatory framework for Artificial Intelligence (AI). This comprehensive proposal is the first institutional effort that goes beyond first principles to lay down detailed requirements. The impact is expected to be worldwide, with an ensuing potential to cause widespread process changes.**

This paper examines key aspects of the Proposal and discusses the framework's foreseeable impact, in particular on research in medical robotics, classified among the high-risk applications and thus subject to a heavy regulatory burden.

On the technical side, the focus rests in particular on the shift toward system-level requirements and an implied shift toward systems thinking. After discussing some open issues on the nature of decision making in AI-based systems, the paper explores issues connected with developing medical robots with progressively higher levels of autonomy.

On the organizational side, attention is on the tension between two conflicting drives: the differentiation of independent regulatory ecosystems and the universal adoption of the most restrictive standard. Finally, regulatory burden differentials are examined in the light of a new division of research labor between academia and commercial spin-offs.

I. Introduction: The AI Act

On April 21, 2021 the European Commission released official documents detailing its proposed regulatory framework for AI, the *Artificial Intelligence Act* (AI Act) [1]. The draft comes in the wake of a number of preparatory documents addressing a variety of related topics.

The document is potentially a game-changer. An institution of the Commission's caliber has put its weight behind a specific proposal, ending the phase in which researchers and politicians discuss the principles and the approach, and starting the debate on specifics. This Proposal has the power to force change at many levels and steer innovation in research labs and technology-based companies, not just in Europe but around the world.

The scope of the regulation is set in Article 3.1, where AI is defined by enumeration as any software built with the techniques listed in the (easily amended) Annex I. AI here is intended in the most general sense possible (past, present and future), thus including traditional symbolic AI, expert systems, all flavors of machine learning, inductive logic programming, hybrid systems and statistical methods– and leaving room for techniques yet to come.

*This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 742671 "ARS").

[1]The author is with the Department of Computer Science, University of Verona, Strada le Grazie, 15, 37134, Verona, Italy. Write to: mariacamilla.fiazza@univr.it

The design and use of AI-based components, and the critical step of integrationwill all have to be re-examined under a new lens: a special attention to the issue of *safety*. Safety is understood here in a broader societal sense. The working notion in use in robotics (freedom from bodily harm caused by interaction with a robot) is functionally expanded to include harm from reduction of economic opportunity, discrimination, and loss of human rights.

The focus is on the technological acceptance that comes from safety, captured under the banner of *trustworthy AI*. Input safety, output trust: the desired result is the acceptance and widespread adoption of AI-based technologies, with the declared goal of enabling sustained economic expansion and establishing the strategic relevance of European technology.

One of the document's objectives is to provide a framework that is sufficiently general to account for future developments and sufficiently specific to actively constrain them—a form of forward-compatibility in innovation engineering. The resulting constraints need to be taken into account by researchers writing grant proposals in the EU, by spin-offs and R&D departments in orienting technical development and also by financial management, in anticipating and providing for the economic burden of regulations.

With this Proposal, the Commission aims at building the much-needed infrastructure surrounding AI as a technology. It provides an *environment* for the core technology–a way to contextualize it, to integrate it within the other over-structures that give form to our society. The document, in addition to describing (guidelines for) regulations, establishes that the legal framework for AI is that of *product* certification. What is being regulated is not the technology per se, in the abstract, but AI *systems* in specific use cases. EU values and ethical guidelines are embedded at both regulatory and legal levels. One of the clear benefits is leading technology entrepreneurs out of the uncertainty of regulatory limbo, providing certainty of the law and clarity on applicable rules, thus allowing for realistic assessment of liability profiles and costs.

What was released on April 21 is a proposal. It has not been signed into law yet, and it may not be approved by the European Parliament and all Member States in that exact form. However, the Proposal contains clear policy and delineates the regulatory strategy chosen by the third largest market for robotics. With high likelihood, standards for certification will be developed within this framework.

A. Scope and Organization of This Paper

The AI Act can be approached from a multiplicity of perspectives, including law, human rights, economics, policy

and ethics [2], with analysis in mind or with the intent to offer amendments to the text, as done in [3]. This paper focuses on the implications of the text as presented. It is written from the viewpoint of a practitioner in medical robotics, a field that is maximally impacted by the Proposal due to its intrinsic risk to health and to the sensitive nature of patient data. Practitioners of robotics in less safety-critical applications may recognize in the exposition key themes at play in their own domains. The paper is intended as a way to start the discussion, especially in the medical robotics community, about the practical ways in which the advent of the AI Act and its guidelines will change the research and development landscape—and about how to respond to the challenge.

The rest of the paper is organized as follows: Section II discusses select aspects of the Proposal, focusing on technical requirements; Section III broadens the view to the international effects of the regulation; Section IV discusses possible impacts of the regulatory burden on how research is organized. Section V concludes with an invitation to proactively embrace the challenge to design better systems.

II. Risk and Requirements

The framework's backbone is contextual risk assessment, organized by use cases and followed by continuous risk management. Although the Proposal covers all techniques and flavors of AI, it tries not to over-regulate the market, making a distinction between applications that *need* to be regulated and applications that do not. AI systems are divided into 4 categories of risk (unacceptable, high, limited, minimal). Obligations are imposed depending on the level of risk.

No regulatory burden is imposed for minimal-risk applications, where there is no detectable threat to safety, access to livelihood or protected rights. Limited-risk applications must provide enough transparency and information for the user's consent to be meaningful. On the opposite end of the risk spectrum, for the first time the EU outright bans some computer systems: those that manipulate users into actions that could cause harm, or that could be used for the suppression of human rights in the presence of power differentials (e.g., between government and citizen). In particular, social scoring technology is explicitly banned.

Unsurprisingly, AI systems in the medical application domain, such as AI-guided surgical robots or AI-assisted diagnosis, fall into the high-risk category, which comes with very strict requirements and a heavy regulatory burden. Regulation is especially impactful in light of the recent efforts to raise the level of autonomy in surgical robots.

Some systems in medical robotics are accessory and do not interact with the patient, not even through patient data. Examples are training systems for robotic surgery with automatic assessment of surgical skills (e.g., dexterity) and medical simulators with case-generation systems, often used during a surgeon's medical training and at times included in physician certification and licensing. On first inspection, these systems may not be considered high-risk. The AI Act requires broadening the scope of how risks are conceptualized. Here the danger is to exclude unfairly: Any automatic assessment with real-world consequences (licensing for use of a robotics system) is subject to in-depth scrutiny and is considered high-risk. Training may end up cognitively biasing the surgeon, even if it is aimed at acquiring motor skills. The AI Act asks of AI developers to think about the larger systems in which AI is immersed.

The Proposal's solutions to high risk are risk reduction through constant risk management, a stringent ex-ante conformity assessment, and ex-post surveillance plans with explicit indication of metrics for monitoring. Hefty documentation, described in Annex IV, is required throughout the entire life cycle of the AI "product."

Annex IV places the emphasis on *motivating* design choices, explicitly discussing trade-offs, listing underlying assumptions, formalizing what the system is optimized for and even discussing the trade-offs induced by the need to comply with the regulation. A perhaps-desired side effect of being subjected to this regulatory burden is that product development has been made into a fully deliberate, conscious process, in which developers are constantly asking themselves the reasons and the consequences of their choices, thinking in an integrated (systems) manner. As Quinn et al. point out in [4] while advocating for the establishment of expert groups for medical AI, it takes a remarkable level of interdisciplinary expertise to perform at this level. The AI Act could result in a very significant mindset shift into treating design requirements by default as *system-level* requirements, as opposed to module-level requirements.

A. The Age of Decision Making

The prescriptions in Annex IV in fact mandate that every step of design, development, validation and testing leave an auditable trail. This auditing pertains to the development process of systems whose defining characteristic is, in turn, partial decisional autonomy, and whose decisional performance is the subject of interest of the regulation.

To touch briefly on a topic worthy of a longer discussion, the Proposal mandates in Article 10 that all datasets used "shall be relevant, representative, free of errors and complete," "exhibiting the appropriate statistical properties," and that they be evaluated for possible biases. In so doing, the Proposal represents in the wider context of AI a problem already visible in the General Data Protection Regulation [5]: it is not presently known how to fulfill the bill. The problem is especially thorny in medical robotics in connection to accounting for rare events, anatomical variants and rare pathologies in the datasets.

A version of the problem restricted to anti-discrimination is well discussed in [6] and still unresolved: "under the minimal interpretation the non-discrimination requirement is ineffective, under the maximal interpretation it is infeasible." Although the legislative intent is clear, its technical translation is not. Using solely the criterion of correlation with protected features (e.g., race and gender) leads to either disregarding the intent of the regulation or to disregarding the

purpose of the system. Advances in understanding the mechanisms of decision-making at the heart of discrimination and bias are needed to overcome the impasse.

When what is automated is how data is being put together and readied for a decision, researchers as a community are called to clarify what it is that drives our decisions: what principles, what trade-offs, and what context-capturing *structures*. Comprehensive AI regulation in a sense starts the age of systematic decision research.

B. On Autonomy and Supervision

In the Proposal, three related elements describe the power relationship between citizens and AI systems: the obligation to provide for human supervision, the right for a human to override an automated decision, and the right to obtain human intervention. Intervention may differ from overriding by affecting or changing elements on which the machine decision is based. AI systems are intended to remain fully under human control.

Article 22 of the GDPR prohibits any decision "based solely on automated processing, including profiling" which "significantly affects" a data subject. In forbidding automated processing from resulting, by itself, in any action that could significantly affect humans, EU law creates the space for the right to human intervention and forbids *full autonomy* (i.e., systems at level of autonomy 5 in [7]).

This prohibition would also entail that human supervision is now an intrinsic element of AI system design. Researchers will have to successfully tackle both the issue of how to effectively design to facilitate human intervention (including override and supervision) and the much more difficult question of how to effectively supervise a super-intelligence. Most AI systems may currently perform very far from superhuman intelligence, but the day is not far when they will exceed human performance even in tasks of higher cognition.

Article 13 requires that the AI's "operation is sufficiently transparent to enable users to interpret the system's output and use it appropriately." The system must in fact be designed specifically to enable effective human oversight, not just effective use. The two may coincide in practice at times, but should be kept conceptually distinct.

Human intervention needs to be consciously *designed* into the system at different levels, already in the earliest phases. Good design of the user interface is necessary but certainly not sufficient to provide transparency. The architectural provisions needed to integrate the human element have far-reaching consequences. At what level of abstraction can the human affect the system's data, choices, intermediate results? With what consequences? The Proposal clearly intends to enable human review at least at the last stage, as a vetting of the final output. However, decisions not based "solely on automatic processing" occur also when humans can intervene at lower levels. This is an intriguing research question in architectures for shared control.

A cornerstone objective is to research methods to facilitate the interpretability of machine decisions, making their meaning in context accessible to users and to those with oversight. Users must be "fully informed of the capabilities and limitations" of the system, in itself a research question.

Depending on the nature of the problem addressed by the machine and on the nature of the computation, some features of AI (e.g., opacity and complexity) can make human oversight ineffective in spite of thoughtful design. Human supervision can also introduce novel sources of bias. Under the hypothesis that a specific machine output was overridden because it was in fact incorrect, if the AI does not learn from the new, human-generated data point, it will retain its propensity to repeat the mistake; if it does learn, it may introduce new bias. Neither naive course of action is safe.

Some systems crunch vast amounts of data to generate a recommendation, making it very hard to evaluate their output both for developers (to discern if the AI is learning "correctly") and for users (to discern if the output is correct in that particular circumstance). Often, due to the volume of data, truly assessing the performance of machines requires other machines.

III. Regulatory Ecosystems

The AI Act thwarts the onset of national solutions to the problem of AI. It prevents the fragmentation of the EU market and the advent of a regulatory babel—also averting a loss of attractiveness in the eyes of strategic innovators. Fragmentation may be more than geographical. So far, United States government agencies have independently established their internal guidelines for the use of AI. For example, H.R. 4468, a 2021 bill recently introduced for discussion in the House of Representatives, entrusts the head of *each* agency to "establish an AI Strategy, Objectives, and Metrics Plan [8]", defining "values, ethics and principles" for the agency's use of AI, and pursuing a now-familiar list of desiderata—spanning from high-quality, reliable, and representative training data to data protection to embedding mechanisms for human supervision.

Conversely, with the notable exception of Defense agencies (military applications are outside of the scope of the AI Act), all administrative bodies in European Member States stand to implement the same shared vision of the acceptable boundaries of AI technology, thus magnifying the regulation's impact in the experience of the citizen.

The AI Act does more than prevent a problem; it builds an organic web of coherent elements, animated by the same logic, the same approach and the same priorities found in other EU laws. The Proposal incorporates the EU approach to privacy seen in the GDPR, to product safety in the CE certification, and to human rights in the Charter of Fundamental Rights [9]. Predictably, after adapting them to the case of high-risk AI-based applications, the Proposal adopts the same strategies and solutions that are found in the GDPR: risk assessment, proportionate impact evaluation, reliance on the certification system, and risk-dependent limitations.

The AI Act stands as one element in a tight network of interwoven regulations that reinforce and explain each other. The remarkable degree of integration between AI Act and

other EU guiding documents creates a genuine regulatory *ecosystem.*

Although the AI Act is European law, its impact extends beyond the borders of Member States. Europe will stand on the international scene as an appealing integrated market with well-defined entry conditions. Meanwhile, across the Atlantic the United States have launched their National AI Initiative in 2020. As part of its implementation, in June 2021 members of the new National Artificial Intelligence Research Resource Task Force were appointed and tasked with establishing a research infrastructure to sustain technological leadership in AI, also addressing security and civil rights. Soon, the US will stand with a second, competing, value proposition. Other countries are also looking with interest, seeking to establish their own guidelines and define their stance.

AI providers are witnessing the formation of different markets. There is tension between two conflicting drives: the specialization of independent ecosystems into separate markets with different regulatory logic and the collapse of less restrictive regulations, when the most restrictive standard is adopted *by providers* everywhere. This second case occurred with the GDPR, which became the *de facto* standard for transnational corporations. China's response is the Personal Information Protection Law (PIPL) [10], whose second draft was released in April 29th, 2021.

Unified global regulation has clear benefits in terms of cost reduction. Whereas costs considerations always favor unification under one standard, considerations of *ability* can favor specialization. Banning certain types of research judged ethically risky (especially in biotechnology and genetic engineering) resulted in the research being physically moved to less regulated countries, where hard laws are replaced by soft guidelines. China made itself especially attractive to researchers who wished to pursue technical progress beyond the legal bounds in effect in their states, but legal elsewhere.

The Proposal prohibits some applications that are judged intrinsically too risky under a human rights lens. Social-scoring technology, which has never been introduced in the EU, would be permanently banned in Europe upon the approval of the AI Act; social scoring AI is deployable—and currently deployed ubiquitously—in China.

Although the European and US legal systems have maintained their unique logic and individuality, their regulatory standards historically have tended to converge, building the backbone of international consensus later formalized in standards. The main approval bodies recognize each other's certifications as substantial.

When large states or unions such as the US, Europe and China are vying for strategic dominance in the key market of AI, even small regulatory advantages can result in an important competitive edge. First-to-market and early market penetration often lock in a position of advantage against competitors. Regulatory advantage can come in the form of a head start on what will later become the global consensus. Compelling evidence of its critical importance can be found in a recent legislative effort [11]. If the bill is signed into law, US technology companies would enjoy financial incentives to direct their experienced and highly qualified personnel to participate in the meetings of international standards organizations.

IV. The Regulatory Burden

The AI Act contains an explicit acknowledgment that the regulatory burden may negatively impact innovation, in particular by barring less-mature (and therefore not-yet-certifiable) technology from access to deployment. To counter this risk, there are provisions to help small enterprises and developers of high-risk applications. The primary innovation-support measure is the creation of regulatory sandboxes (Title V, Art. 53-55) [1], "controlled environment[s] to test innovative technologies for a limited time on the basis of a testing plan agreed with the competent authorities" (explanatory memorandum sec. 5.2.5). Even though the Proposal determines the legal basis for regulatory sandboxes (§72) it falls short of actually establishing them. The text merely *encourages* national authorities to provide them and to establish process and parameters for access, as well as rules for governance, supervision and liability.

Sandboxes are intended to accompany product development, serving as safe places for experimentation and testing up to the pre-market phase. The AI provider is, however, still fully liable for adverse events that may occur in the sandbox as a result of the technology. A second concern is the absence of language acknowledging that scale is a critical factor in the thorough testing of AI systems; sandbox platforms, when made available by national governments, may not be able to support testing *at scale*.

In medicine—and consequently also in medical robotics—almost all decision making is patient-specific and relies on medical profiling so as to enable leveraging results from medical literature (evidence-based medicine). There is no question that medical decisions have the potential of affecting the data subject very significantly; error can have catastrophic consequences. Article 54 clarifies that the handling of sensitive data processed in sandboxes is also restricted and highly regulated. Testing must not result in decisions or measures affecting the data subject, so the type of testing that can be done in this context is virtual and self-contained. Data must be deleted after testing; processing logs can be kept for a year only for accountability purposes. Results, at this point detached from the data, become part of the product testing documentation described in Annex IV and required by Article 11(1).

A. Requiring What is Not Yet Available

Litigation and the court system will ultimately establish to what degree the regulatory burden of high-risk AI applications needs to be met in practice while technology matures. It is not a simple matter of fulfilling or not fulfilling a requirement, in the sense that requirements for fairness, absence of bias, robustness, etc. rest on open issues in computer science. There is still no established way to certify machine learning [12], [13], let alone systematic techniques to design machine-learning systems and data collection efforts that guarantee the

desired outcome in terms of emergent properties. In addition to explainability and dataset bias, other open issues are protection of privacy and robustness to adversarial machine learning. These issues are open at the technical (and perhaps philosophical) level only; the legislative intent is clear.

The AI Act can be expected to fuel innovation in specific directions in much the same way as the GDPR did. European law recognizes a *right to an explanation* whenever algorithms make decisions on user-level data that affect the data subject "significantly" [5, art. 13, 14, 22]. Although explainable AI is still in its infancy, it has experienced an astounding acceleration in growth starting in 2017. The right to an explanation was already present in the preceding Data Protection Directive dated 1995 [14]. However authoritative, directives are guidance for EU Member State lawmakers and are not laws in themselves. The GDPR was signed into law in 2016, with the understanding that it would go into force in 2018. The effect on the research environment was widespread, sustained and nearly immediate.

The GDPR estblished a requirement for "meaningful information about the logic" driving a decision [5, Art. 13-14] well before reliable technology was in place to fulfill the requirement; similarly, the AI Act introduces pressure to develop methods suitable to certify decision making and systemic properties.

Of necessity, the law will have to be interpreted in the direction of AI providers making a reasonable effort to comply. Providers can be expected to incorporate the latest techniques up to the time of product release. Although keeping abreast of research developments is highly desirable, outside of academia the need to do so introduces additional costs and cost uncertainty.

AI providers of medical devices may be held to an unreasonably high standard in the courts–perhaps due to political and social factors and the need for the technology to be perceived as more mature than it currently is. This could be a major disincentive to engage in high-risk applications and even more so in safety-critical fields.

B. On the Effect of Burden Differentials

For all technologies that benefit from being deployed in order to mature, being brought to market is an enabling factor not just in terms of the developer's economic viability, but also technically. Disincentives are apparent in the enormous difference in regulatory burden between medical applications and the vast ocean of minimal-risk processes that could be enabled by AI.

Some needed advances in AI-asssisted surgical robotics and diagnostics are critically dependent on breakthrough general-purpose solutions for thorny problems such as contextual reasoning with real-world variability, anatomical navigation, and safe handling of rare, adverse or unforeseen events. It is possible that these key technological solutions, although motivated by the needs of medical robotics, may be best developed in other (non-safety-critical) contexts first.

Working on a related problem in a non-safety-critical application domain could be a way to generate funds and revenue from a breakthrough technology and sustain it to maturation. Only later, when the technology has already proven reliable and trustworthy, would it be ported to medical robotics. On the other hand, there may be aspects of a general technology that cannot be successfully worked out in other application fields—because of dependencies that are not apparent in other, less complex, domains, or due to additional constraints. For example, in some use cases contextual reasoning must occur in near real time to be compatible with the intra-operative needs of surgical robotics. Other complexities requiring that a general-purpose innovation be developed as a *medical* technology can stem from the systemic nature of patient safety and the integration with other legal requirements specific to healthcare data, such as restricted access to sensitive data.

Differentials in regulatory burdens can be expected to drive a novel decomposition of a research area into aspects that have to be addressed in the medical application field and aspects that can be investigated elsewhere. Implicitly, those that can be tested elsewhere *should* be. Medical robotics will likely find itself specializing further, primarily as the place where available robotic technology is embedded into a network of system safety, patient safety and system integration considerations.

As a consequence of the AI Act, it is reasonable to anticipate that safety and system integration concerns will extend into progressively earlier phases of the design process and will receive a larger proportion of the energy and resources than they have been afforded so far. Techniques will tend to be imported into the medical application domain from elsewhere rather than be developed natively.

An example of technology developed natively is the automatic extraction of procedural knowledge from surgical books, intended to enable robots to learn surgical workflows and make surgical plans [15]. There are many non-medical versions of this problem. Some, for which a few studies exist in literature, are learning to cook specific dishes from recipe books, or providing automated technical support after processing product documentation and maintenance manuals. Progress in these domains would not immediately generalize to good performance on surgical texts because of very significant differences in the degree of exploitable structure and in the semantic complexity of the context. A non-safety-critical version that could generalize is learning to mix, colorize, shape and bake artistic ceramics from books.

Wherever technique innovation can be fastest and return viable products, there technical progress will become the core concern. The AI Act, on the other hand, places an incentive for safety and integration, not novel technique, to become the primary focus of safety-critical fields of research.

C. Organization of Research

The selective pressure induced by differentials in the regulatory burden could result in changes at the organizational level, mirroring changes in problem decomposition and formulation. In response to liability pressure, research entities active in medical robotics may turn to a "dual track" strategy

for their (native) research questions, especially concerning technologies whose accurate evaluation is strongly dependent on large numbers and which may thus require market deployment of prototype technology. The first step is to actively seek out related minimal-risk (parallel) problems that could benefit from the target innovative technology. In a first phase, the safety-critical (core) components of the native problem would be kept in-house or addressed in collaboration with academia as pure research, whereas the parallel problem would be entrusted to a spin-off created specifically for the purpose. There is no regulatory burden on pure research, because regulations pertain to AI-based *products*; in the parallel domains, on the other hand, the certification burden is minimal. An example of native problem is modeling drug interactions and their effects in the body; a parallel problem could be modeling ingredient interaction in recipe design and their effects on food properties.

Market deployment of the solutions in the parallel domain can enable the critical technical maturation needed to justify the expense of developing regulated products as solutions for the native problem. In addition, the revenue generated in parallel can contribute to funding this second phase of research.

Spin-offs have typically come *after* the research results (solutions), as part of a project's exploitation plan; once the solution was transferred from the research project to the spin-off, the latter moved forward as a new, independent entity. In the scenario under examination, on the other hand, spin-offs would be contributing to the maturation of the target technology while it is under development—the very purpose of the spin-offs' existence—and insight would feed directly back into the originating research laboratory. It is possible that research entities in medical robotics will start operating as a company with two departments: medical (safety and integration) and non-medical (technique and revenue generation).

V. Conclusion

The main practical take-home point from the AI Act is that work toward certification of a future product starts at prototype design time, instead of after final testing, and continues throughout the development process. Product safety has functionally been expanded well beyond the current scope, mostly concerned with impacts, contact forces and mechanical harm.

Design that is not informed by the requirements laid out in the Proposal is a liability, in the sense that the resulting product may not be certifiable. Taking the time to examine the Proposal's implications is a worthy investment, especially in a field like medical robotics where major capital is needed to bring products to market. Being informed and taking the guidelines into account can help research entities anticipate change and be found ready when, 24 months after the AI Act is adopted as law, the policies come into effect with the strength of regulation and the reach of European law.

Regulatory certainty is recognized as a key enabling factor for the flourishing of targeted investments in AI-based systems and for Europe's economic development. The AI Act provides clarity on the legal framework and the legislative intent, establishing the boundaries of permissible use and calling for the institution of regulated support structures (e.g., sandboxes). However, the Proposal cannot dispel the uncertainty entirely because full compliance with the regulation is at present beyond the state of the art. Perhaps as a result of the approval of the framework, market deployment in non-regulated fields of application will spearhead the availability of innovative solutions for the medical and surgical settings.

Once the AI Act is approved, compliance can be expected to pose significant challenges to developers, at least initially. It is also a unique opportunity to embrace higher standards of quality in the design process, mindful that technology ought to be a tool to empower humanity and is not an end in itself.

References

[1] European Parliament and Council of the European Union. *Artificial Intelligence Act: Regulation Laying Down Harmonised Rules on Artificial Intelligence and Amending Certain Union Legislative Acts*, Proposal for Regulation COM/2021/206 final. Brussels, Belgium, April 21, 2021.

[2] "Ethics, Safety and human centricity: Intelligent Machines under the scope of the European AI Regulation Act" Workshop 7, 3rd Conference of the Italian Institute of Robotics and Intelligent Machines. Oct 9, 2021. [Online] Available: `https://www.youtube.com/channel/UCd1GRFZu1SOuwj_uyykrCuw/videos`

[3] N. Smuha et al, "How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence Act." August 5, 2021. SSRN. DOI: 10.2139/ssrn.3899991

[4] T. P. Quinn, M. Senadeera, S. Jacobs, S. Coghlan, and V. Le, "Trust and medical AI: the challenges we face and the expertise needed to overcome them," *Journal of the American Medical Informatics Association*, Volume 28, Issue 4, April 2021, Pages 890–894, DOI:10.1093/jamia/ocaa268.

[5] European Parliament and Council of the European Union. *General Data Protection Regulation*, Regulation EU 2016/679. Brussels, Belgium, April 27, 2016.

[6] B. Goodman, S. Flaxman. "European Union regulations on algorithmic decision-making and a 'right to explanation'." *AI magazine*, 38(3):50–57, 2017.

[7] G.-Z. Yang et al., "Medical robotics—Regulatory, ethical, and legal considerations for increasing levels of autonomy," *Science Robotics*, vol. 2, no. 4, March 2017.

[8] United States House of Representatives, 117th Congress. *AI for Agency Impact Act*. Bill H.R.4468, July 16, 2021.

[9] European Union, *Charter of Fundamental Rights* (consolidated version of the Lisbon Treaty) in Official Journal of the European Union. Brussels, Belgium, October 26, 2012.

[10] R. Creemers, M. Shi, L. Dudley, and G. Webster, "Translation: Personal Information Protection Law of the People's Republic of China (Second Review Draft)." [Online] Available: `https://digichina.stanford.edu/news/translation-personal-information-protection-law-peoples-republic-china-draft-second-review`

[11] United States Senate, 117th Congress. *Leadership in Global Tech Standards Act of 2021*. Bill S.1849, May XX, 2021.

[12] F. Tambon et al, "How to Certify Machine Learning Based Safety-critical Systems? A Systematic Literature Review," 2021. [Online]. Available: arXiv:cs.LG/2107.12045v2.

[13] R. Sheh, "Evaluating Machine Learning Performance for Safe, Intelligent Robots," 2021 IEEE International Conference on Intelligence and Safety for Robotics (ISR), 2021, pp. 388-393, doi: 10.1109/ISR50024.2021.9419381.

[14] European Parliament and Council of the European Union. *Data Protection Directive*, Directive 95/46/EC, October 24, 1995.

[15] M. Bombieri, M. Rospocher, D. Dall'Alba and P. Fiorini, "Automatic detection of procedural knowledge in robotic-assisted surgical texts", in International Journal of Computer-Assisted Radiology and Surgery, pp.1-9, 2021.