

Regulatory-Driven Software Engineering Methodology for AI/ML-Based Medical Software Development for SMEs

Stephanie Batista^{1,†}

¹University of Coimbra, CISUC/LASI – Centre for Informatics and Systems of the University of Coimbra, Department of Informatics Engineering, Coimbra, Portugal

Abstract

The increasing integration of artificial intelligence into software as a medical device is transforming healthcare, but it is also amplifying the regulatory and engineering challenges faced by small and medium-sized enterprises. Artificial intelligence-based medical software is governed by a fast-evolving and heterogeneous regulatory landscape, particularly across the European Union and the United States. Existing development methodologies do not adequately reconcile agile, iterative practices with the rigorous documentation, traceability, risk management, and post-market obligations imposed by the applicable regulations and standards. This doctoral project aims to design, develop, and empirically validate a regulatory-oriented engineering framework tailored to the structural constraints of small and medium-sized enterprises developing artificial intelligence-based medical software for the European Union and the United States. Results to date include two systematic literature reviews and an empirical industry study involving international professionals, which prompt the hybrid development lifecycle framework currently under construction and validation. This paper presents the research motivation, questions, methodology, results already achieved, and next steps.

Keywords

Software as a Medical Device, Artificial Intelligence, Regulatory Compliance, Small and Medium-Sized Enterprises

1. Introduction

Medical software incorporating Artificial Intelligence (AI) and Machine Learning (ML) is rapidly expanding as a class of high-impact digital health technology, offering substantial benefits in diagnosis, prognosis, treatment planning, and clinical workflow optimization [1, 2]. From a software engineering perspective, this growth materializes in mandatory lifecycle, risk-management and validation processes prescribed by regulatory frameworks and harmonized standards such as IEC 62304, ISO 14971 and ISO 13485, that AI/ML capabilities further extend with concerns around data governance, model versioning, adaptive change control and post-market monitoring.

Reconciling these requirements with contemporary practices such as agile development and continuous integration remains a recognized open challenge, as the iterative, data-driven nature of AI/ML systems is fundamentally at odds with the plan-driven, evidence-heavy assumptions of medical device regulation [3, 4, 5].

Small and Medium-Sized Enterprises (SMEs) represent more than 90% of medical technology manufacturers in the EU and a comparable share globally [6], typically operating with lean engineering teams, restricted access to clinical data, limited internal regulatory expertise, and strong dependence on external regulatory consultants [7, 8]. While these structural features support innovation, they become a serious limitation under the highly prescriptive, evidence-driven regulatory frameworks that govern medical devices, particularly when AI/ML components introduce additional concerns such as data quality, generalization, bias, transparency, cybersecurity, and post-market change management [2, 8, 9].

IDOESE'26: International Doctoral Symposium on Empirical Software Engineering, October 07, 2026, München, Germany

✉ stephanie@dei.uc.pt (S. Batista)



© 2026 Copyright for this paper by its author. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2. Motivation

This doctoral project is motivated by a clear gap between regulatory expectations for AI/ML-SaMD (Software as a Medical Device) and the engineering reality of SMEs. Current frameworks are predominantly oriented toward regulatory and quality specialists in large enterprises, while the practical conditions of technical teams in resource-constrained organizations are underexplored. As a result, SMEs often produce regulatory evidence retrospectively, struggle to reconcile agile methodologies with regulatory requirements, and rarely operationalize AI-specific requirements such as a Predetermined Change Control Plan (PCCP) and subgroup performance evaluation.

This research is also driven by a strong professional motivation: across a career in software quality assurance, verification and validation, and medical devices regulatory affairs, the author has contributed to the development and certification of more than 30 medical devices at consulting firms, manufacturers, and a notified body (EU) and approved body (UK). This first-hand experience has made the difficulties faced by AI/ML-SaMD manufacturers particularly visible, including documentation produced under pressure at the end of the lifecycle, limited internal regulatory expertise in technical teams, and substantial duplication of effort when targeting both the EU and the US markets.

The central problem addressed by this doctoral project is therefore the absence of regulatory-oriented engineering frameworks that are simultaneously tailored to the structural conditions of SMEs, explicitly aligned with the dual-market regulatory pathway of the EU and the US, and capable of supporting the AI/ML-specific challenges of medical software.

3. Research Questions

This research is guided by four Research Questions (RQs):

- **RQ1:** What are the main technical, organizational, and regulatory challenges faced by SMEs in the development and validation of AI/ML-SaMD targeting the EU and the US?
- **RQ2:** How do current development frameworks for AI/ML-SaMD address regulatory compliance throughout the lifecycle, and what are their main limitations for SMEs?
- **RQ3:** Which engineering practices, regulatory artifacts, and automation mechanisms can be integrated into a hybrid lifecycle framework that reconciles agile software engineering with regulatory rigor in AI/ML-SaMD development for SMEs?
- **RQ4:** To what extent does the proposed regulatory-oriented framework improve compliance readiness, traceability, and development efficiency in SMEs developing AI/ML-SaMD for the EU and the US?

4. Related Work

Proposed frameworks in the literature serve distinct but complementary purposes across four lifecycle domains: software development, AI implementation, cybersecurity and data protection, and Post-Market Surveillance (PMS). In the software development domain, early contributions such as MDevSPICE [3] established a structured process assessment approach grounded in IEC 62304, ISO 14971, and ISO 13485. Subsequent work sought to harmonize requirements across safety-critical domains [10], embed regulatory constraints into multidisciplinary design optimization [11], and operationalize IEC 62304 compliance through automated DevOps workflows [5]. None of these frameworks incorporates AI/ML-specific mechanisms, gaps that AI implementation frameworks have progressively addressed. Granlund et al. [4] adapted MLOps practices to regulated SaMD contexts. Assadi et al. [12] combined systems engineering, human factors, and iterative development to support safe ML deployment. More recently, Shrivastava et al. [13] proposed a regulatory governance framework drawing on FDA's Total Product Lifecycle and Good Machine Learning Practice principles, while Janev et al. [14] introduced a trustworthiness evaluation framework referencing both the EU AI Act and MDR.

Cybersecurity represents a comparatively underexplored dimension. Early contributions identified risk classification approaches and systemic vulnerabilities in networked medical devices [15, 16]. More recently, Chattoraj and Joshi [17] introduced a knowledge graph automating regulatory compliance interpretation, while Jayaneththi et al. [18] concluded that current frameworks provide only high-level principles without AI-specific threat modeling or dynamic risk quantification.

In the PMS domain, Zinchenko et al. [19] proposed one of the first structured PMS methodologies tailored to AI/ML-SaMD. Uddin and Ge [20] introduced an automated adverse event classification framework leveraging Large Language Models (LLMs). Pechmann et al. [21] proposed CRISP-PCCP, supporting the FDA's PCCP for continual learning systems.

These limitations are particularly significant for SMEs, for which limited resources, lack of regulatory expertise, and increasing complexity under medical devices and emerging AI regulations lead to certification delays, unsuccessful submissions, and barriers to market entry [22]. The absence of integrated, SME-oriented frameworks thus represents not only a technical gap but a barrier to innovation and economic growth in the medical device ecosystem [7, 8]. This doctoral project positions itself at this intersection, combining practitioner evidence with a systematic review of current frameworks.

5. Research Methodology

The research follows a Design Science Research (DSR) methodology [23], organized into four phases: (i) problem identification and characterization, combining a systematic literature review of regulatory and technical challenges for AI/ML-SaMD with an empirical industry study based on an anonymous online questionnaire targeting technical and managerial professionals in SMEs (RQ1); (ii) state of the art analysis of current development frameworks through a second systematic literature review, organized into a taxonomy of four categories and assessing lifecycle coverage, regulatory alignment, and suitability for SMEs (RQ2); (iii) framework design, articulating a hybrid lifecycle that reconciles the agility of contemporary software engineering with regulatory rigor and AI-specific obligations for the EU and the US, the phase currently in progress (RQ3); and (iv) framework validation through experts evaluation and case study application with SMEs, assessing compliance readiness, traceability, and development efficiency (RQ4).

6. Progress to Date

6.1. Systematic Literature Review on AI/ML-SaMD Challenges

The first systematic literature review identified the main technical, ethical, and regulatory challenges of AI/ML-SaMD development across 4 major jurisdictions (Canada, the EU, the UK, and the US). The review covers 87 studies published from 2015 to the first quarter of 2026, proposes a taxonomy of 4 categories (AI, software engineering, regulatory affairs, and stakeholders) and 20 subcategories, with in-depth analysis of the AI and software engineering categories, and is currently under re-submission to a Q1 journal. Key findings include the dominance of cybersecurity and data privacy concerns, the persistent tension between agile practices and plan-driven regulatory requirements, and the underrepresentation of software validation studies.

6.2. Systematic Literature Review on Current AI/ML-SaMD Development Frameworks

The second systematic literature review identifies and analyzes existing AI/ML-SaMD development frameworks proposed over the last decade, concluding that no existing framework fully reconciles regulatory rigor with the iterative, data-driven nature of AI systems. Most identified frameworks address requirements from outdated regulatory references, provide limited AI implementation and cybersecurity guidance, adopt rigid plan-driven processes, or offer insufficient support for post-market adaptation and scalability.

6.3. Industry Challenges Empirical Study

Complementing the literature analysis, the empirical industry study collected 36 valid responses from professionals in organizations active in AI/ML-SaMD development across 17 countries, with approximately 70% of the participating organizations qualifying as SMEs. The study was approved by the Ethics Committee for Research of the University of Coimbra (reference CEIUC_R-23-02/2025). Key results include: 94.3% of respondents perceive the documentation burden as a moderate-to-high challenge (51.4% at the highest level); 86.1% identify the reconciliation of agile methodologies with regulatory requirements as a challenge for their technical team; 91.7% consider the regulatory differences between the EU and the US to be a significant strategic obstacle; 80.6% of organizations report no form of regulatory automation, and 88.9% have no formally defined PCCP or Algorithm Change Protocol (ACP); 63.9% of respondents consider that their technical (non-regulatory) team does not have sufficient dedicated regulatory expertise; and 73.5% of organizations evaluate device performance in specific subgroups only rarely or never. These results are organized into an empirical paper under submission to the ESEM Technical Track and directly motivate the framework design phase.

6.4. Framework Design Progress

Building on this empirical evidence and the identified gaps, the framework design is currently in progress. The proposed framework follows a hybrid lifecycle structure with 3 mutually reinforcing layers: (i) a set of practical artifacts (guides and checklists) targeted at the areas of highest regulatory pressure, namely PCCP/ACP, risk management, technical documentation, and software validation; (ii) capability-building components addressing the structural shortage of internal regulatory expertise; and (iii) methodological adjustments that embed regulatory activities into the engineering lifecycle from the earliest phases, including early regulatory planning, automation of documentation and traceability, and mapping of common EU and US regulatory artifacts. Preliminary versions of the lifecycle structure and traceability schema have been drafted, and the next iterations will focus on detailed activity definition and tooling integration.

7. Expected Contributions

The expected contributions of this doctoral project are:

- A consolidated body of empirical evidence on the technical, organizational, and regulatory challenges faced by SMEs developing AI/ML-SaMD for the EU and the US.
- A taxonomy of AI/ML-SaMD frameworks and challenges, supporting comparative analysis and gap identification.
- A regulatory-oriented engineering framework tailored to SMEs, articulating a hybrid lifecycle that reconciles agile engineering practices with EU and US regulatory expectations and AI-specific obligations.
- A set of practical artifacts (guides, checklists and traceability structures) directly usable by SMEs in their day-to-day regulatory and engineering practice, along with empirical validation evidence.

8. Next Steps

The remaining work of this doctoral project is structured along 3 main directions: (i) completion of the framework design, with detailed specification of the lifecycle phases, regulatory artifacts, traceability mechanisms, and AI-specific safeguards, (ii) the definition of supporting checklists, tooling and regulatory automation mechanisms (e.g., eQMS integration, document generation, and LLMs) compatible with the constraints of SMEs, and (iii) empirical validation through experts evaluation and case studies with SMEs.

Participation in IDoESE 2026 is a valuable opportunity to debate this research with experts from the software engineering community. Feedback is particularly sought on the scope, granularity, and modularity of the framework and its applicability to SMEs with different maturity levels, on the methods and metrics for validating it in real industrial settings, on the methodological rigor of the empirical components, and on strategies to maximize its practical impact and industry adoption.

Declaration on Generative AI

The author has not employed any Generative AI tools.

References

- [1] K. Zhou, G. Gattinger, The evolving regulatory paradigm of AI in MedTech: A review of perspectives and where we are today, *Therapeutic Innovation & Regulatory Science* 58 (2024) 456–464. doi:10.1007/s43441-024-00628-3.
- [2] U. J. Muehlematter, P. Daniore, K. N. Vokinger, Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis, *Lancet Digital Health* 3 (2021) e195–e203. doi:10.1016/S2589-7500(20)30292-2.
- [3] M. Lepmets, F. McCaffery, P. Clarke, Piloting MDevSPICE: the medical device software process assessment framework, in: *Proceedings of the 2015 International Conference on Software and System Process (ICSSP '15)*, ACM, New York, NY, USA, 2015, pp. 9–16. doi:10.1145/2785592.2785598.
- [4] T. Granlund, V. Stirbu, T. Mikkonen, Towards regulatory-compliant MLOps: Oravizio's journey from a machine learning experiment to a deployed certified medical product, *SN Computer Science* 2 (2021) 342. doi:10.1007/s42979-021-00726-1.
- [5] M. R. Martina, E. Bianchini, S. Sinceri, M. Francesconi, V. Gemignani, Software medical device maintenance: DevOps based approach for problem and modification management, *Journal of Software: Evolution and Process* 36 (2024). doi:10.1002/smr.2570.
- [6] MedTech Europe, MedTech Europe's facts & figures 2025, <https://www.medtecheurope.org/resource-library/medtech-europes-facts-figures-2025>, 2025.
- [7] S. Reddy, Global harmonization of artificial intelligence-enabled software as a medical device regulation: Addressing challenges and unifying standards, *Mayo Clinic Proceedings: Digital Health* 3 (2025). doi:10.1016/j.mcpdig.2024.100191.
- [8] S. Santra, P. Kukreja, K. Saxena, S. Gandhi, O. V. Singh, Navigating regulatory and policy challenges for AI enabled combination devices, *Frontiers in Medical Technology* 6 (2024) 1473350. doi:10.3389/fmedt.2024.1473350.
- [9] F. Faheem, M. Haq, M. Derhab, R. Saeed, U. Ahmad, J. S. Kalia, Integrating ethical principles into the regulation of AI-driven medical software, *Cureus* 17 (2025). doi:10.7759/cureus.79506.
- [10] A. B. Bujok, S. T. MacMahon, P. F. Grant, D. Whelan, W. J. Rickard, F. McCaffery, Approach to the development of a unified framework for safety critical software development, *Computer Standards & Interfaces* 54 (2017) 152–161.
- [11] S. R. Mishra, K. Behdian, Multidisciplinary design analysis and optimization framework for regulatory driven medical device development, *Proceedings of the Design Society* 3 (2023) 2735–2744. doi:10.1017/pds.2023.274.
- [12] A. Assadi, P. C. Laussen, A. J. Goodwin, S. Goodfellow, W. Dixon, R. W. Greer, A. Jegatheeswaran, D. Singh, M. McCradden, S. N. Gallant, A. Goldenberg, D. Eytan, M. L. Mazwi, An integration engineering framework for machine learning in healthcare, *Frontiers in Digital Health* 4 (2022) 932411. doi:10.3389/fdgth.2022.932411.
- [13] A. Shrivastava, U. Jain, M. Al-Farouni, Y. Kumar, R. J. Anandhi, M. Bhavana, Proposed framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical

- device (SaMD), in: 7th International Conference on Contemporary Computing and Informatics (IC3I), 2024, pp. 1755–1763. doi:10.1109/IC3I61595.2024.10829073.
- [14] V. Janev, M. Nenadovic, D. Paunovic, S. Vahdati, J. Li, M. H. Yousuf, J. Montanya, R. Theilen, J. Wittenstein, S. Tsurkan, R. Huhle, IntelliLung AI-DSS trustworthiness evaluation framework, in: 2024 32nd Telecommunications Forum (TELFOR), 2024, pp. 1–4.
 - [15] T. L. Lewis, J. C. Wyatt, mHealth and mobile medical apps: a framework to assess risk and promote safer use, *Journal of Medical Internet Research* 16 (2014) e210. doi:10.2196/jmir.3133.
 - [16] P. A. H. Williams, A. J. Woodward, Cybersecurity vulnerabilities in medical devices: a complex environment and multifaceted problem, *Medical Devices: Evidence and Research* (2015) 305–316. doi:10.2147/MDER.S50048.
 - [17] S. Chattoraj, K. P. Joshi, MedReg-KG: Knowledge graph for streamlining medical device regulatory compliance, in: 2024 IEEE International Conference on Big Data (BigData), 2024, pp. 3382–3390. doi:10.1109/BigData62323.2024.10825686.
 - [18] B. Jayaneththi, F. McCaffery, G. Regan, An evaluation of risk management standards and frameworks for assuring data security of medical device software AI models, in: *Proceedings of the 19th International Conference on Software Technologies (ICSOFT 2024)*, 2024, pp. 91–99.
 - [19] V. V. Zinchenko, K. M. Arzamasov, S. F. Chetverikov, A. V. Maltsev, V. P. Novik, E. S. Akhmad, D. E. Sharova, A. E. Andreychenko, A. V. Vladzymyrskyy, S. P. Morozov, Methodology for conducting post-marketing surveillance of software as a medical device based on artificial intelligence technologies, *Sovremennye Tekhnologii v Meditsine* 14 (2022) 15–23. doi:10.17691/stm2022.14.5.02.
 - [20] M. M. Uddin, M. Ge, Data analytics framework for identifying relevant adverse events in medical software, in: *Proceedings of the 9th International Conference on Information and Communication Technologies for Ageing Well and eHealth (ICT4AWE 2023)*, 2023, pp. 81–90.
 - [21] L. Pechmann, Y. Potdevin, K. Brehmer, D. Nowotka, M. Leucker, CRISP-PCCP – a development methodology supporting FDA approval for machine learning enabled medical devices, in: *Bridging the Gap Between AI and Reality*, 2025. doi:10.1007/978-3-031-73741-1_14.
 - [22] M. Draghi, The future of european competitiveness: A strategic vision for the EU in a changing global economy, https://ec.europa.eu/info/publications/future-european-competitiveness_en, 2024.
 - [23] J. Brocke, A. Hevner, A. Maedche, Introduction to design science research, in: *Design Science Research. Cases*, 2020, pp. 1–13. doi:10.1007/978-3-030-46781-4_1.