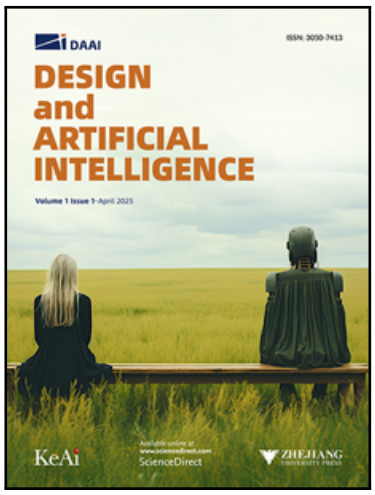
Journal Pre-proof



Design and Expert Validation of an AI Change Management Process (AI-CMP) Assessment for AI-Enabled Medical Devices: Towards Agentic AI Change Management with Human Oversight

Niamh St John Lynch , Róisín Loughran , Martin McHugh , Fergal McCaffrey

PII: S3050-7413(26)00040-6
DOI: https://doi.org/10.1016/j.daai.2026.100101
Reference: DAAI 100101

To appear in: *Design and Artificial Intelligence*

Received date: 11 February 2026
Revised date: 29 July 2026
Accepted date: 19 August 2026

Please cite this article as: Niamh St John Lynch , Róisín Loughran , Martin McHugh , Fergal McCaffrey , Design and Expert Validation of an AI Change Management Process (AI-CMP) Assessment for AI-Enabled Medical Devices: Towards Agentic AI Change Management with Human Oversight, *Design and Artificial Intelligence* (2026), doi: https://doi.org/10.1016/j.daai.2026.100101

Highlights

- Introduces a validated AI Change Management Process (AI-CMP) framework to govern iterative AI lifecycle changes in safety-critical medical devices.
- Demonstrates how design research can translate complex multi-standard AI regulatory requirements into a usable, auditable change assessment artefact.
- Provides empirical evidence that automation and agentic AI can significantly reduce regulatory change management burden while preserving human accountability.
- Proposes a human-governed agentic AI architecture for repeatable, regulator-aligned change assessments across the Total Product Lifecycle.
- Advances design-oriented AI governance by reframing AI change control as a socio-technical, human-in-the-loop design problem rather than a purely technical task.

Design and Expert Validation of an AI Change Management Process (AI-CMP) Assessment for AI-Enabled Medical Devices: Towards Agentic AI Change Management with Human Oversight

Niamh St John Lynch[1[0009-0009-4150-4970]], Róisín Loughran[2[0000-0002-0974-7106]], Martin McHugh[3[0000-0003-4275-3302]], Fergal McCaffrey[4[0000-0002-0839-8362]]

[1,2,3,4] Regulated Software Research Centre, DkIT, Dundalk, Co. Louth, Ireland
niamh.stjohnlynch@dkit.ie

Abstract

Agentic Artificial Intelligence presents a significant opportunity to address one of the most resource-intensive obligations facing manufacturers of AI-enabled Medical Devices (AIeMDs): the assessment and documentation of iterative lifecycle changes. Recall evidence indicates that inadequately controlled modifications are a leading contributor to safety issues in AI/ML-enabled devices, yet change assessment remains largely manual. A prerequisite for safe agentic automation is a validated, structured instrument for the agent to execute. This paper reports the design and expert validation of the AI Change Management Process (AI-CMP) Assessment, a multi-jurisdictional instrument consolidating the change management expectations of EU MDR 2017/745, UK MDR 2002, FDA guidance, and the EU AI Act 2024/1689, together with standards including ISO 13485, ISO 14971, IEC 62304, IEC 81001-5-1 and IEC PAS 63621. In an exploratory, design-science study, the 35-section instrument was evaluated by a purposive panel of ten domain experts (eleven records, including one re-scored response) drawn from legal manufacturers and specialist consultancies across EU, UK and US regulatory contexts. Ratings indicated strong content validity for regulatory coverage and completeness, while lower ratings for length and resource burden strengthened the case for automation. Two implementations are reported: a proof-of-concept web application, and a deployed agentic prototype (AIa-CMP) operating under tiered, mandatory human oversight, applied in a real-world case study within a medical device organisation. These findings constitute purposive expert evaluation of content validity rather than population-level evidence of effectiveness; formal evaluation of the agentic implementation is identified as future work.

Introduction

This section introduces the regulatory framework for traditional active medical devices such as *Software as a Medical Device* (SaMD) as a standalone software device and *Software in a Medical Device* (SiMD). A SiMD is defined as such when its software is a component or accessory to a hardware medical device. This research responds to the growth of Artificial Intelligence-enabled Medical Devices (AIeMD) in recent years and acknowledges the reciprocal absence of harmonised standards and guidance. The introduction provides an overview of regulatory expectations arising within a changing technological landscape and offers a solution for change management performed by Medical Device Manufacturers that must adapt to these changing technologies and regulatory requirements. Agentic AI presents a significant opportunity to reduce the burden of these repeat assessments; however, safe agentic automation in a regulated context presupposes a validated, structured instrument that the agent can execute under documented human oversight. This paper makes three contributions: (1) a validated multi-standard AI change management assessment framework (AI-CMP), (2) proof-of-concept digitisation web application, and (3) a deployed agentic AI implementation evaluated through a real-world change management use case within a medical device organisation that has been approved for use by QA/RA and multifunctional team for use in the QMS.

The scope of the AI-CMP is deliberately multi-jurisdictional. The instrument consolidates the change management expectations of EU MDR 2017/745, UK MDR 2002, FDA guidance and the EU AI Act 2024/1689 within a single assessment, such that a manufacturer placing an AIeMD on more than one market may evaluate a proposed change against the requirements of each relevant territory in one structured exercise, rather than maintaining separate, parallel assessments. Whilst certain individual questions necessarily reflect the wording of a particular framework, and may therefore appear specific to one jurisdiction, the instrument is intended to be applied across all three regulatory territories, with sections that are not applicable to a given device, change or market marked as such with a documented rationale.

1.1. Background and Related Work

Regulatory expectations for Software in Medical Devices (SiMD) are well established under the IEC 60601 series of standards [1]. IEC 62304 [2] and IEC 82304-1 [3] define the software development lifecycle and validation respectively, that are under update to include AI requirements. PD IEC TR 80002-1 [4] guides software risk engineering, whereas ISO 14971 [5] governs product risk management for medical devices generally and has been extended to include ISO 24971-2 [6] specifically for AI risk management, whereas IEC 81001-5-1 [7] addresses cybersecurity across the full lifecycle. These and other product and process standards are required for medical device development within a Quality Management System (QMS) framework as defined by ISO 13485. Additional

emerging AI-specific standards include IEC PAS 63621 [8] for data management, IEC 63450 [9] for AI test methods and more amongst the range of standards under development by IEC's Technical Committee TC 62 for Medical Equipment, Systems and Software.

1.2. Rise of Artificial Intelligence-enabled Medical Devices

Artificial Intelligence (AI) has rapidly become a foundational component of advanced medical technologies, accelerating diagnostic accuracy, enabling personalised treatment pathways, and supporting real-time clinical decision-making [10]. AIeMDs and AI-enabled Health Software (AIeHS) are now deployed across radiology, cardiology, pathology, ophthalmology, and digital therapeutics to name a few [11,12]. Recent analyses of the healthcare sector show that transformation aims to reshape operational and clinical processes, requiring organisations to rapidly adapt to emerging technological and organisational change [13,14].

As AI models increasingly incorporate machine learning, deep learning, and adaptive algorithms, medical devices now behave less like static engineered products and are more dynamic and data-driven [15]. This creates unprecedented opportunities for safety improvement and innovation whilst also introducing new risks relating to data quality, model drift, explainability, bias, and cybersecurity. These unique characteristics demand modernised regulatory and operational frameworks for lifecycle development, oversight, risk management and change control [16–19].

1.3. Regulatory expectations and motivations for this work

Regulators worldwide have recognised the need for dedicated oversight of AIeMDs [20,21]. While health-sector literature shows that organisational change practices (e.g., Kotter, Lewin, ADKAR, PDCA) are widely applied to clinical and technical transformations, these frameworks must be paired with rigorous regulatory processes to ensure patient safety [22–24].

Compliance expectations are intensifying and uncertain given the EU Medical Device Regulation (MDR 2017/745) and IVDR (IVDR 2017/746) and the EU AI Act 2024/1689, which classifies AIeMDs as "high-risk systems" that are subject to strict governance, transparency, and lifecycle monitoring requirements [25]. The industry has called for a simplification of the regulations whilst retaining the necessary controls for safety and performance of medical devices [26]. The FDA's recent change to acceptance of ISO 13485 whilst retaining QMSR additional requirements [27] relies on a Total Product Lifecycle (TPLC) *process* approach. These requirements raise the bar for medical device manufacturers requiring greater levels of compliance and obligation to reduce defects whilst assuring continuous improvement in QMSs. A significant difference from these newer regulations is that medical device manufacturers will for the first time have to disclose management reviews, internal audit reports and external/supplier audit data to the FDA. This change is substantial and is compounded by the mandatory use of eSTAR for FDA 510(k) submissions required before a medical device can be placed on the US market.

This taken together with the use of AI (Elsa) being used by FDA and other regulatory bodies piloting similar AI applications, makes scrutiny that much quicker and potentially more in-depth [28]. The industry needs to adapt and ensure their processes are equally capable of timely, accurate, and effective processes that yield comprehensive analysis of changes and maintenance of documentation to support them.

With the need for heightened scrutiny to reduce defects in the market and the introduction of AI into medical devices, regulators are re-emphasising the need for traceable, evidence-based change assessments and in addition, robust data governance, bias mitigation, model performance monitoring, and documented human oversight. AI introduces a high rate of iterative updates, far exceeding traditional device updates. This amplifies the burden on manufacturers to repeatedly assess the safety, regulatory impact, and design-control implications of each update. The PCCP may go some way to reduce the regulatory burden on regulatory bodies. However, PCCPs reduce the burden on the regulatory agency, whilst placing more burden on the medical device industry that needs to continually assess AIeMDs for planned changes including drift and must document all change assessments. Hence, increased number and type of change assessments are necessary and must be fit for purpose to adequately document all changes [29].

The medical device sector was selected as the context for this research because it presents a uniquely demanding environment for the management of change in AI systems. The legal manufacturer bears singular, non-delegable accountability for the safety and performance of a device across its entire lifecycle, including all post-market modifications, under EU MDR 2017/745, FDA QMSR and UK MDR 2002. The consequences of inadequate change management are directly patient-facing, as evidenced by the disproportionately high recall rates reported for AI/ML-enabled devices [30,32]. Furthermore, the convergence of medical device regulations with the EU AI Act within a single product lifecycle creates a level of multi-jurisdictional complexity that is arguably unmatched in other AI application domains, whilst the TPLC approach now embedded in FDA and EU expectations requires change management to operate continuously from concept through to decommission. These characteristics make the medical device industry a critical and instructive setting in which to develop and stress-test a change management framework for AIeMDs, with lessons that are likely transferable to other safety-critical AI domains.

1.4. Safety and Performance of SaMD and AIeMD

To date medical device safety and performance have not yielded entirely satisfactory results. Recent regulatory analyses have shown that a significant proportion of safety issues and recalls involving SaMDs and AIeMDs arise, not from initial design flaws predominantly, but from insufficiently controlled design modifications, software updates, or lifecycle changes [30,31]. A 2025 systematic review of 27 years of US AI/ML-enabled device recalls found that AI/ML devices experienced a disproportionately high number of recalls compared to all other software-related devices, with recall root-causes frequently tied to software defects, inadequate testing of updates, or failures in managing post-market

modifications [32]. Similarly, case studies of software-related device recalls show recurring patterns of poor version control, inadequate verification and validation, and failures to identify the regulatory impact of software changes, many of which directly contributed to field safety issues and device recalls [33–35]. Evidence from real-world QMS inspection findings reinforce this, documenting instances where SaMD and AIeMDs were released with insufficiently analysed design changes, weak traceability, and missing documentation on risk impacts, signalling breakdowns in change management and design control processes [36]. The FDA's 2025 draft guidance for AI-enabled devices highlights these failures explicitly, emphasising that post-market performance monitoring and update governance must be integrated into the TPLC to prevent unsafe or unvalidated algorithm changes from reaching patients [37]. Collectively, this evidence demonstrates that breakdowns in change management, such as untested AI retraining updates, inadequate dataset modifications, and failure to reassess risks after design iterations remain a major cause of device malfunction, regulatory non-compliance, and patient safety risks in the current generation of SaMD and AIeMD products. Hence, the motivation for this work on change management for AIeMD is clear; manufacturers need scalable, standardised, and regulator-aligned methods to evaluate the impact of changes to AI models, datasets, software components, and clinical behaviours across the device lifecycle from concept through to decommission and disposal.

1.5. Change Management within Medical Devices

Change management has always been a core requirement of medical device manufacturing, given regulatory demands for design control, verification/validation, risk management, and documentation traceability under ISO 13485 [38] and FDA 21 CFR 820/QMSR [27]. However, research in healthcare demonstrates that change initiatives often fail without structured frameworks, and that models such as Kotter's 8-Step Process, Lewin's Three-Stage Model, ADKAR, the McKinsey 7-S Framework, and the Plan Do Check Act (PDCA)/ Plan Do Study Act (PDSA) cycle are widely applied to successfully manage complex organisational and technical transitions. However, Change Management performed by legal manufacturers in medical device industry is typically a hybrid of these models with the primary focus being on regulatory compliance to requirements such as ISO 13485 [38].

1.6. Change Management Requirements

ISO 13485 and the updated FDA QMSR as of February 2026, require manufacturers to implement a closed-loop, documented risk-managed change control process for changes affecting the Quality Management System and medical device product design, process or documentation. Key requirements include: 1) Evaluation of all changes for their impact on the Quality Management System (QMS) and medical devices, including regulatory, safety, and performance impacts; 2) Controlling changes according to documented procedures that meet both ISO 13485 and regulatory authority requirements; 3) Maintaining detailed documentation and traceability for every change, including version history, approvals, verification, and rationale; 4) Management of QMS process and

ongoing regulatory compliance (Section 4.1.4); 5) Design and development changes (7.3.9) must be formally reviewed, verified, validated (as appropriate), and approved before implementation; 6) Risk management integration must be applied throughout product realisation and explicitly linked to changes and updated risk assessments; and finally 7) Document control (4.2.4) mandates that changes to documents are reviewed, approved, and clearly traceable. The burden can be significant, and unintended short-cuts are inevitable when resources and competency are in demand or when multinational organisations do not have fundamental controls locked-in, easy access to a complete design history file or full understanding of the technology.

1.7. Challenges to Traditional Change Management

Whilst complying with these clearly defined requirements the burden cannot be underestimated. Significant resources are required to ensure the necessary competency and understanding of changes is available by organisations to ensure these requirements are effectively demonstrated. The data reported on Field Safety Corrective Actions and Recalls suggests that these systems are not always effective for traditional SaMD/SiMD products never mind AIeMDs. The Change Management Process (CMP) approach is itself challenged by the technological advancements made possible by AI. As device design increasingly incorporates AI, the nature of change itself shifts.

Software updates will require model retraining of released AIeMDs; data updates can expect to alter clinical performance; hardware and algorithmic changes can modify risk controls; new evidence may require updated clinical claims; continuous learning introduces ongoing or periodic change assessment. The State of the Art is quickly evolving, and with it, new regulatory requirements and standards and guidance continue to emerge. It is challenging to keep on top of it and is resource intensive.

These challenges create a mismatch between existing change control processes and the rate of complexity and opacity of AIeMD complexity and related modifications. Traditional change management models provide structure, but AI-specific risk factors such as drift, fairness, explainability, emergent behaviours, and dependency on configured data streams, require new tooling, documentation templates, and multidisciplinary analysis methods. This paper addresses this gap by providing a comprehensive, industry-validated AI Change Management Process (AI-CMP) Assessment tool and proposes an AI-powered assessment agent that applies this validated AI-CMP in a regulatory context such as medical device industry. Gilbert et al. highlight that standards alone are not a recipe for success and call for specific guidelines for adaptive ML-based SaMD for algorithm change protocols and real-world use of AIeMDs [39]. This research also states that improving safety and performance requires more than guidelines. It requires validated tools that meet regulatory requirements, capable of preventing poor performance in the field and maintenance of device documentation that reflects the product without undue burden being placed on professionals, not always equipped to adapt in time. A validated change management tool is necessary so that it can be integrated or overlaid onto existing eQMSs preferably with software performance monitoring capabilities.

1.8. Improving Safety, Regulatory Compliance, and Reduced Burden

Safe deployment of AIeMDs depends on the manufacturer's ability to detect, evaluate, and justify changes across the AI lifecycle in line with existing and emerging regulatory requirements [39]. Without clear frameworks, organisations risk inconsistent assessments, overlooked hazards, regulatory non-compliance, changing performance and ultimately loss of clinical trust. The literature on change management models underscores that successful change management in organisations requires structured models, leadership engagement, and systematic processes, particularly when introducing complex technologies like AI and this is essential in medical device organisations, who are also subject to regulatory requirements and scrutiny. The focus can sometimes shift to regulatory compliance instead of effectiveness without intention. This work seeks to enhance device safety by providing a structured assessment of model performance, bias, drift, data integrity, and clinical impact of changes to name a few [40]. The CMP artefact proposed here aligns with AI-specific regulatory obligations, improving early detection of risks associated with changes in AIeMDs including retraining or algorithmic updates. This paper demonstrates how change assessments performed by AI agents with human oversight can strengthen regulatory compliance and reduce resource burden. The validated CMP assessment aligns with the EU MDR[41], FDA[27], UK MDR from MHRA[42], EU AI Act [43] with compliance to ISO 13485 [38], ISO 14971 [44], IEC 62304[2], and IEC PAS 63621 [8] Data Management specifications. Using this CMP ensures traceability of changes to risk controls, verification, validation, and technical documentation. It also supports compliance with Predetermined Change Control Plans (PCCPs) [29].

By integrating evidence-based change management models with regulatory change-control expectations, this work provides a scalable, repeatable, and industry-aligned approach to managing AI/ML lifecycle updates in AIeMDs. It also reduces the already constrained resources for manufacturers by automating time-consuming documentation and regulatory assessments, standardising multidisciplinary input from Regulatory Affairs, Quality Assurance, Clinical Affairs, Data Science, Software and Cybersecurity teams. The reduction of documentation workload supports consistent, reproducible output and improves internal alignment using a structured and validated change framework for Trustworthy AI in healthcare.

While existing regulatory mechanisms such as internal change control procedures, Predetermined Change Control Plans (PCCPs), and ad hoc regulatory impact assessments are commonly used by medical device manufacturers, these approaches remain fragmented and weakly standardised across AI-specific lifecycle domains. In contrast, the AI-CMP provides a unified, validated, and regulator-aligned change assessment framework that operationalises requirements spanning software lifecycle management, AI governance, data management, risk management, cybersecurity, and post-market oversight within a single structured process. This work uniquely demonstrates the evolution of this validated framework from a static artefact into a digital platform and an agentic AI system capable

of supporting repeatable, auditable, and human-governed regulatory change assessments for AIeMD.

1.9. AI-CMP Assessment Tool (Manual)

The AI-CMP Assessment v4.0 is an MS Word file with 35 interrelated sections as a checklist with guidance designed to act as a holistic change assessment for AIeMDs. It covers device identification, change description, intended purpose, regulatory classification, software classification and compliance, IEC 62304 and IEC 82304-1 lifecycle activities, Medical Electrical Equipment under IEC 60601 (where applicable), AI governance (ISO 42001 or EN 18286 healthcare standard [45]), data management (IEC PAS 63621), training/re-training (BS ISO 30440), algorithm efficiency, third-party PTM/SOUP assessment, model cards, AI model development and testing, product and software risk management (ISO 14971; PD IEC TR 80002-1), AI risk management (ISO/DTS 24971-2), clinical risk management (e.g., NHS DCB0129/0160), security lifecycle (IEC 81001-5-1), SBOM/AIBOM, usability (IEC 62366-1), safety and performance claims, PMS and monitoring, PCCP, EU AI Act compliance, sustainability, technical documentation, deliverables, regulatory approvals, change approval, release gating and implementation effectiveness planning. The change assessment form allows skip logic with editable rationale for "*Not Applicable*" selection in line with Good Documentation Practices required by medical device organisations. It provides a manual editable document file that can be uploaded to existing e-QMS workflows. It serves either as a stand-alone artefact or a supplemental checklist for use to demonstrate compliance to regulatory requirements for AIeMD undergoing change assessment.

The content presented in this research was validated by several leading medical device organisations as described in the methods section. The next phase of the development project was to create a web-based application utilising the validated content from the AI-CMP Assessment tool**.** Note: The term AI-CMP refers to the validated AI Change Management Process assessment framework described in this study and subject to design research by case study. AI-CMP Web Application denotes the digitised, non-agentic implementation of this framework. The term AI*a*-CMP refers to the Agentic AI system that operationalises the AI-CMP through autonomous agents operating under defined human oversight and approval controls. These terms are used consistently throughout this paper.

2. Methods

2.1. Validation Method of AI-CMP Tool

This study employed a design science and mixed-methods expert validation approach to evaluate the AI-CMP artefact with respect to usability, regulatory completeness, and adoption readiness within regulated medical device development environments against US, EU and UK territory requirements. Structured Likert-scale survey is combined with qualitative free-text responses, targeting professionals involved in medical device

development, AI lifecycle management, software engineering, quality systems, and regulatory affairs. The objective was to assess whether the AI-CMP instrument adequately supports change management decisions for AIeMDs and SaMDs, prior to digitisation into a web application. Validation was performed using a 35-item questionnaire instrument hosted online, capturing both quantitative scoring and qualitative user experience feedback. Raw response data were obtained from AI-Change Management Process Assessment Validation Survey results, while summary results and visualisation outputs were extracted from MS forms.

Participants and Recruitment: A total of ten individual and anonymised expert respondents (n=10) completed the validation survey. One respondent (ID#4 repeated validation under ID#5), following greater understanding following post-survey interview deemed necessary due to 'poor' responses across almost all sections. For this reason, the dataset presented shows eleven respondents (n=11) though since one is a repeat, both are captured in the analysis for a true and representative dataset. This participant rated all sections as 'poor' due to length and following interview and updates related to security section removal; the participant re-scored with favourable ratings. All participants represented large multinational manufacturers, SMEs, micro-enterprises and specialist consultants, ensuring exposure to varying levels of QMS maturity and AI governance capability. Most importantly, the selected experts were primarily chosen from IEC international cohort of experts in the field, who are supporting emerging regulatory requirements for medical devices and are therefore elite experts. While the cohort size does not support statistical generalisation, it was intentionally designed to provide expert validation of regulatory adequacy, usability and practical integration rather than population-level inference. This provides content validity required by this study.

Validation of AI-CMP Tool using a Survey Instrument: Participants reviewed or utilised this template by applying it to real or hypothetical AI/ML change scenarios from their organisations. All participants were asked to complete a survey following review of the AI-CMP assessment tool. The survey captured ratings across four domains: 1) Usability & Structure including clarity, navigation, usefulness in change management, and support for competency development. 2) Regulatory Coverage with consideration for MDR 2017/745, UK MDR 2002, FDA guidance, IEC 62304, and the EU AI Act. 3) Domain Completeness across intended purpose, data management, risk management, AI QMS, software development, AI training, model testing, PCCP, security, sustainability, and model cards. 4) Practicality of Use including feasibility, suitability, length, resource burden and likelihood of adoption. A final question asked for readiness to transfer to an Agentic AI system using a 5-star rating scale. Validation feedback informed iterative refinement of the AI-CMP artefact in line with design science research principles. Specifically, participant feedback regarding assessment length, duplication, and applicability resulted in targeted modifications, including the removal of a standalone security section to ISO 27001, not strictly mandated by medical device regulations.

Separately, consultation with an academic domain expert prompted the development of a companion Standard Operating Procedure (SOP) to provide instruction on a proportionate and scalable approach to assessment. The SOP clarifies that not every change requires the comprehensive assessment in full: it guides the assessor in determining the depth of assessment proportionate to the change, distinguishing organisational quality management system changes from new product introductions and from post-market product or model changes, whilst the not-applicable mechanism allows sections irrelevant to a given change to be excluded with a documented rationale. A formal mapping of change type to a prioritised subset of sections, such that minor or non-AI-impacting changes need not trigger the full assessment, is identified as a development priority for the modular AI*a*-CMP. These refinements improved usability while maintaining regulatory completeness and alignment with relevant standards.

Survey Validation Analysis: Participants completed the online validation survey and rated the AI-CMP using Likert scales (1–5 or 1–6 depending on item), providing qualitative feedback where relevant. Quantitative results were extracted from MS Forms as an MS excel sheet and statistical analysis using Minitab v. 22.4.0. In summary, most respondents rated usability and structure as 'Very Good' or higher, with lower scores concentrated among respondents expressing concern regarding assessment length, resource burden or applicability across all change types. Quantitative data were aggregated across all participants to generate mean scores, bar charts, and domain-specific completeness summaries using the exported dataset. Qualitative responses were analysed thematically using inductive coding across feedback statements, including comments such as: "This is an excellent resource… will be further enhanced by a digital version." "It's great, it has everything." (Medtronic). Several respondents noted that the full assessment may be more appropriate for significant changes, suggesting a need for clearer guidance on scope and applicability. However, these respondents assumed use of the manual system. Instead, their validation was to prove that the content was appropriate for use within an agentic system prior to building and assessment as part of case study assessment.

Quantitative Analysis and Results: Descriptive statistics were generated for each Likert-scale domain. Usability & Structure: most sub-items received top ratings i.e., 'Very Good' or higher. Regulatory Coverage: most respondents scored MDR, FDA, UK MDR, IEC 62304, and EU AI Act coverage as '*Very Good'*-'*Exceptional'*. Domain Completeness: most responses fell in the '*Very Good'*-*'Excellent'* range across 13 domains. Practicality: feasibility, suitability, and acceptance most received '*Very Good'*-'*Exceptional'* ratings. Overall Usefulness & Adoption Likelihood: many respondents selected the top two categories. These outputs provide content validity by elite domain experts that the AI-CMP tool meets or exceeds regulatory requirements and industry expectations in the measured domains, apart from resource burden/length which was the primary concern stated.

The clustering of responses at the upper end of the Likert scale suggests a potential ceiling effect. This is not unexpected given the composition of the validation cohort, which consists of highly specialised experts with deep involvement in AIeMD development and exposure to evolving regulatory frameworks and standards. As the AI-CMP framework reflects current and emerging regulatory expectations within this domain, the alignment between evaluator expertise and artefact content may reduce score variability, resulting in compressed ratings at the upper scale. This effect should therefore be interpreted as a function of expert–artefact alignment rather than an absence of critical evaluation. That the cohort exercised critical and discriminating judgement, rather than offering uniform approval, is evident in the ratings themselves, whereby scores were markedly lower and more dispersed for length, ease of navigation, resource burden and likelihood of adoption, with several respondents, for example ID#4, ID#10 and ID#11, returning 'poor' or 'very poor' ratings on these dimensions, in contrast to the consistently high ratings awarded for regulatory coverage and domain completeness. Several respondents further provided substantive constructive feedback, including recommendations to separate organisational quality management system assessment from product- and change-specific assessment, to differentiate the depth of assessment by change type, to sharpen the focus on AI-specific considerations, and to incorporate ISO 13485, which informed both the refinements described above and the modular, change-type-specific development identified as a priority for future work.

Qualitative Results Summary: Qualitative data, although limited, were useful for an understanding of strengths and weaknesses across different company sizes and quality approach. The themes were mapped to usability, regulatory completeness, and workflow integration. The responses were coded into three categories: 1) Strengths included comprehensiveness, clarity and regulatory completeness. 2) Improvement Areas saw a request for reduction of resource burden and clarity on when to use this tool (i.e. some suggested the tool was not practical for all changes due to the length and heavy resource burden), more visual examples were requested for junior personnel with additional clarification for application across different territories. 3) Digitalisation Benefits were demonstrated with strong interest in an interactive system, reinforcing the progression toward a web-based version of the AI-CMP tool with resounding agreement that this tool is suitable for Agentic AI. Figure 1 demonstrates the results of all 10 participants who scored between 3 and 5 on a 5-star rating scale on the suitability of use as an Agentic AI system, with the majority scoring 5 as the highest rating. On further qualitative analysis, the lowest-scoring respondent (ID#4), who rated several domains 'low' or 'poor', attributed these ratings in follow-up interview primarily to the *length* of the assessment and *resource burden* of applying such an extensive instrument, rather than to its regulatory content. These ratings may also reflect the respondent's organisational context as a contract/consultancy organisation, for which the effort of a comprehensive assessment is experienced differently than by a legal manufacturer. Legal manufacturers, who bear full

responsibility and accountability for the safety and performance of their devices on the market and costs associated with field safety corrective actions, returned high ratings, indicating acceptance of the tool and agreement on its suitability as the basis for an Agentic AI system.

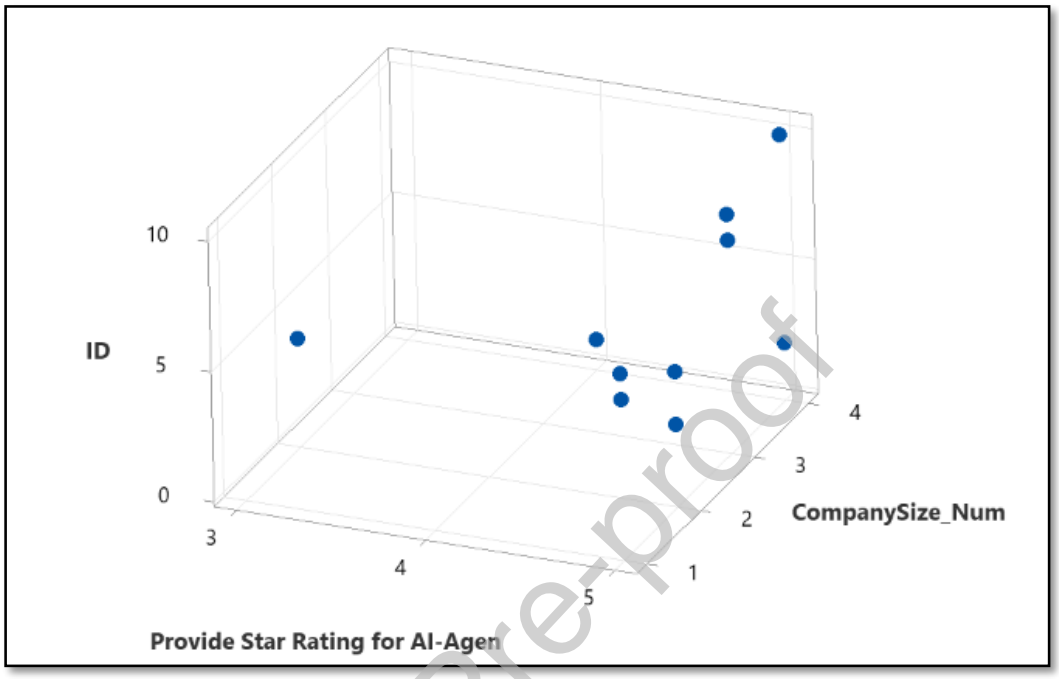


Figure 1 3D Scatterplot of Result (ID) - Company Size vs 5-Star Rating

Use of AI-CMP Across Change Types: validation feedback also highlighted the need for clearer guidance on when a comprehensive AI-CMP assessment is required versus when a reduced or targeted assessment would be appropriate. Following initial feedback during design research a Standard Operating Procedure was developed to support the AI-CMP assessment tool and provide such guidance. Respondents (ID#8, #10) suggested explicit differentiation between organisational QMS changes and new product introductions and post-market product or model changes. Several participants noted that applying the full assessment to minor or non-AI impacting changes may introduce unnecessary burden and reduce adoption.

The AI-CMP tool allows for N/A of relevant sections with a rationale, thereby reducing this burden. However, since the sections remain visible in an MS Word version of the tool, it can be challenging to see beyond these questions which remain open to auditors and quality assurance checks. This is overcome by the AI-CMP Web-based application, which is marked as 'not applicable' (n/a) and disappears from reader's view. This challenge is overcome still further through an Agentic AI system that needs to be aware of all questions necessary to ask but is allowed to remove with human review, sections not applicable. Therefore, removing sections or separating assessments at this stage takes away from the holistic nature of the tool and opportunities for automating it as

an Agentic AI tool. Indeed, respondents who rated the manual AI-CMP lower due to length or feasibility simultaneously rated the concept of an automated or agent-assisted implementation significantly higher. This suggests that resistance was not to regulatory depth, but to manual execution, providing empirical justification for the agentic architecture.

Quantitative Results Summary: The validation demonstrates a strong multi-stakeholder agreement on the AI-CMP tool's usability, completeness, and regulatory adequacy. Resource burden/length was of concern to several respondents, ref. Figure 2. This reflects a need for automation since it is also accepted that the tool, AI-CMP MS Word method, is comprehensive and suitable for use for the assessment of changes of AIeMDs.

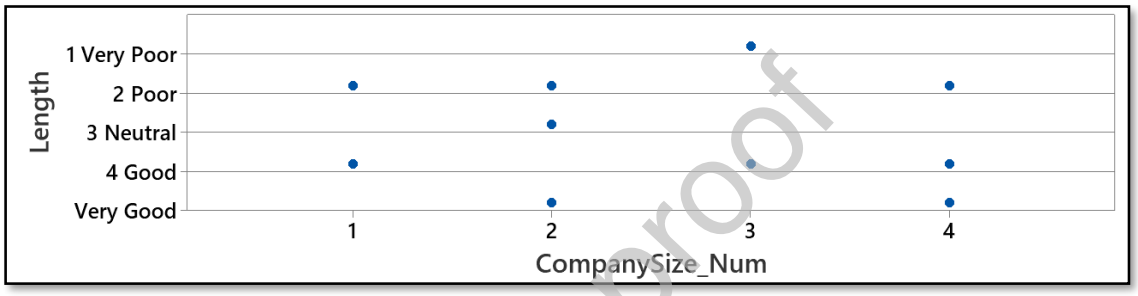


Figure 2 Dotplot Likert Scale Ratings of Length by Company Size (1=Micro,4=Large)

The data across domain areas reflects the strengths and weaknesses seen in the qualitative analysis, such as comprehensiveness and regulatory completeness and clear relationships between the 13 sections in meeting regulatory requirements for AIeMDs. We can see from Figure 3 that respondents viewed MDR (EU) 2017/745 as highly rated across all organisations. This data compares similarly with other territories.

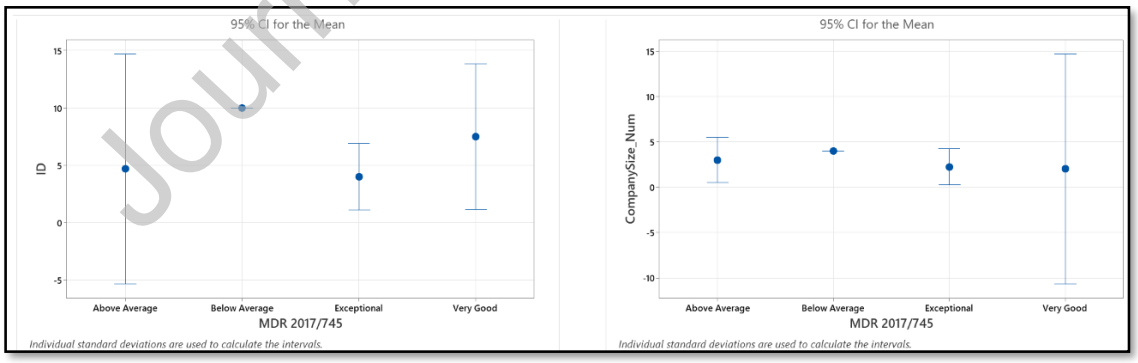


Figure 3 Interval Plot by ID (participant) and Company Size Likert Ratings of MDR (EU)

Pearson Correlation demonstrates a clear relationship between the elements assessed under the regulatory territories (EU, UK, US) included in the study, ref. Figure 4. Additional data is presented in Figures 5–10.

Correlations

	AI Act 2024/1689	ISO 62304	FDA Guidance	UK MDR 2002
ISO 62304	0.778			
FDA Guidance	0.870	0.671		
UK MDR 2002	0.945	0.623	0.928	
MDR 2017/745	0.943	0.688	0.898	0.968

Figure 4 Correlation Analysis of Various Regulatory Requirements (Minitab 22.4.0)

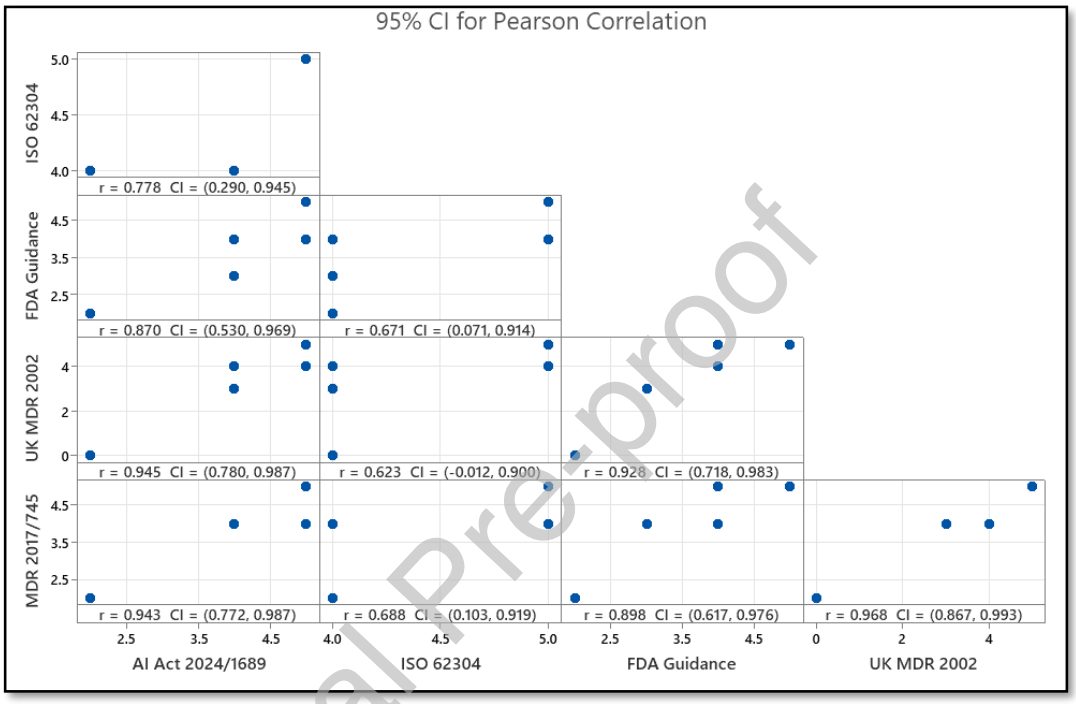


Figure 5 Pearson r Correlation Analysis of Regulatory Requirements

The results taken together are evidence that the AI-CMP framework functions effectively as a pre-market and post-market change assessment instrument. There is clear justification to proceed with development of an AI-CMP web application and AI*a*-CMP solution to reduce resource burden and improve workflow integration. The validation process establishes that the AI-CMP assessment framework is fit-for-purpose, forming the evidence base for subsequent digital transformation into the AI-CMP web-based application and an AI*a*-CMP tool. Ratings improve when the AI-CMP's manual Word tool is proposed as the basis for an Agentic AI system, reporting an almost perfect high score; the majority giving a 5-star rating '5' (highest rating) as shown in Figure 10; total possible score is 50 across all domains.

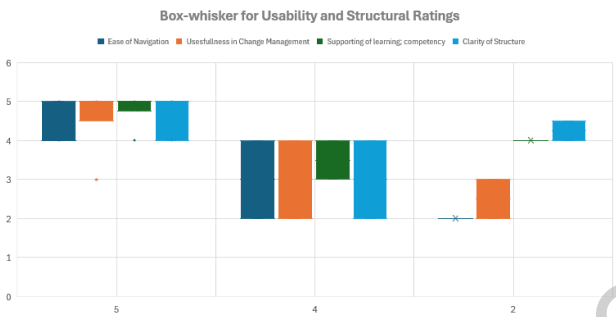


Figure 6 Box-Whisker for Usability & Structural Ratings

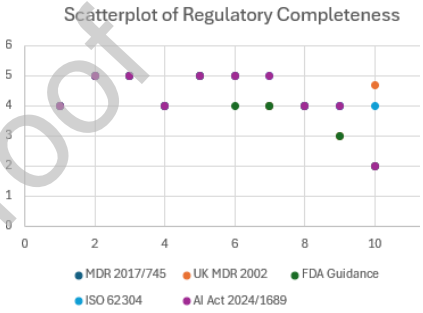


Figure 7 Scatterplot of Regulatory Completeness

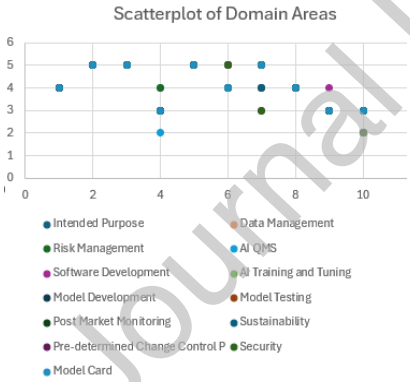


Figure 8 Scatterplot of Domain Areas

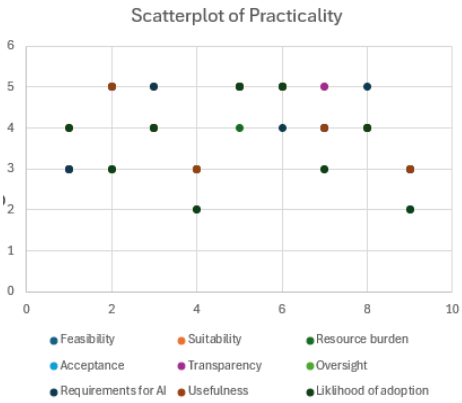


Figure 9 Scatterplot of Practicality

Figure 10 Radar of Likert Scale Assessments by Domain Areas

3. AI-CMP Web Application (Non-Agentic)

The AI-CMP web-application is an interactive, browser-based platform designed to operationalise the AI Change Management Process Assessment (AI-CMP) framework. It provides a structured, end-to-end digital environment for evaluating changes to AIeMDs, SaMD and AIeHS in alignment with global regulatory expectations. The platform implements the full 35-section AI-CMP model covering device identification, software lifecycle, AI governance, risk management, regulatory compliance, and post-market monitoring as listed above and transforms it into a usable, traceable, and auditable application workflow.

The web application is a single-page, client-side system built using standard web technologies (HTML, CSS, JavaScript), loaded with a structured JSON definition of all 35 sections, their regulatory standards, and their associated questions and guidance. Each section maps directly to compliance obligations captured in the AI-CMP DOCX/PDF, including MDR, FDA QMSR, IEC 62304, ISO 14971, ISO 42001, IEC PAS 63621, BS 30440, IEC 81001-5-1, and the EU AI Act. The application automatically renders these requirements dynamically in the user interface. A key benefit of the application is its role as a digital regulatory assessment tool, enabling manufacturers to perform structured, repeatable, and standardised evaluations of AI/ML lifecycle changes. It provides a centralised interface for gathering evidence, documenting decisions, assessing risk, and generating documentation required for regulatory submissions or QMS approval. The

regulatory requirements can be updated in the HTML code to ensure all users and sites, can avail of timely implementation of the State-of-the-Art regulatory requirements and harmonised standards and guidance instead of relying on individuals to be trained to the next version of a form and ensure correct download and use. However, agentic AI allows us to take this a step-further and thereby yield considerable advances to efficiency, consistency and effectiveness of change assessments. Available for review at https://github.com/niamh888/CMPWebApplication.git.

3.1. AI*a*-CMP System Architecture (Agentic)

The AI*a*-CMP prototype used as part of real-world use case is available at https://github.com/niamh888/AIa-CMP and deployed at https://a-ia-cmp.vercel.app. A case study assessment report is provided as supporting evidence from a real-use case change that demonstrates the system's performance and integration within an operational QMS environment. It ensures that MemoryTell, as a new-start medical device organisation, understands their requirements prior to implementation of a change. AI*a*-CMP is a browser-based, single-page web application designed to automate the drafting of regulatory change management assessments for AIeMDs. The application follows a three-tier serverless architecture: a vanilla JavaScript/HTML5 client provides the user interface and manages a seven-step guided workflow; a Cloudflare Worker acts as a secure API gateway, proxying requests to the Anthropic API without exposing credentials to the browser; and Claude Sonnet (claude-sonnet-4.5 or above) serves as the AI drafting engine, invoked via the tool-use API to produce structured JSON output. Upon receiving a structured change description, the system sequentially generates a comprehensive 35-section Change Management Process assessment, already validated for content, covering areas including risk management, software lifecycle impacts, cybersecurity, usability, and regulatory notifications. Each section is annotated with confidence ratings (High/Medium/Low), clause-level regulatory citations, and categorical flags such as Safety Critical, PCCP impact and regulatory uncertainty (e.g. as is the case for the AI Act that has yet to be finalised for medical devices). The regulatory scope spans EU MDR 2017/745, FDA QMSR/ISO 13485:2016, IEC 62304, ISO 14971, ISO 42001, and transitional provisions of the EU AI Act 2024/1689. The application is explicitly designed as a human-in-the-loop, co-creation tool: reviewers may accept, edit, reject, or mark each AI-drafted section as not applicable, with rejected sections triggering targeted AI regeneration using the reviewer's feedback as an additional instruction. Human review is tiered by risk: any assessment involving a new or modified intended purpose, a change to a safety-critical function, training data outside the previously validated distribution, or a potential PCCP boundary condition requires mandatory, documented human review and explicit approval before any QMS action is taken, whereas lower-risk outputs, such as documentation updates with no design impact, may be queued for batch review. Human intervention is also mandatory where the system flags regulatory uncertainty or conflict between territories, or where no precedent exists for the change type within the organisation's assessment history. Reviewers validate agent outputs against the

clause-level regulatory citations and confidence ratings generated for each section, ensuring that drafted content is grounded in current requirements rather than unsupported generation. This design operationalises the human oversight obligations of Article 14 of the EU AI Act [43] for high-risk AI systems and preserves the accountability of the legal manufacturer required under ISO 13485 [38] and FDA QMSR [27]. All workflow actions, including reviewer identity, timestamps, section decisions, and rationale are captured in an ISO 13485 compliant audit log, and final records are exportable as PDF with 21 CFR Part 11 style signature blocks and as structured JSON for integration with eQMSs. Draft persistence is maintained via browser localStorage, with no server-side data storage, positioning the prototype as a lightweight decision-support tool rather than a validated medical device software system during prototype implementation.

Quantitative Validation Results of AI*a*-CMP Changes. Two assessors (n=2) reviewed three changes (n=3) presented as preliminary quantitative data in support of AI*a*-CMP benefits. Additional assessors were not possible within MemoryTell due to the lack of skilled resources capable of performing such a review. Nevertheless, this data provides early case study data supporting validation of AI*a*-CMP. Table 1 provides a **s**tage-by-stage comparison of the organisation's established manual turnaround against AI*a*-CMP task time recorded automatically in the system audit trail, for two senior assessors (a QA/RA Director and CEO) across three real-world changes for MemoryTell. AI*a*-CMP durations are presented for Change 1: AIa-CMP-20260728-DGCBPS; Change 2: AIa-CMP-20260728-TI3794; Change 3: AIa-CMP-20260728-9301M0. Manual change assessment durations represent typical documented turnaround of similar changes, rather than a stopwatch-timed re-enactment of the individual changes. All change reports and audits trails are provided.

Table 1 AIa-CMP Change Comparison to Manual Change Management

	Measure	Manual	AIa-CMP
Change 1: application transfer AI Model to Ireland /EU AWS **Reviewer 1**	Time to complete change assessment	1–3+ days	18m 4s
	Time to review	1–3+ days	10m 2s
	Updates / corrections	2+ days	Minimal
	Notification of actions	5 mins + (× persons identified by task)	< 1 min
	End-to-end cycle time	**1 week +**	**32m 2s**
Change 2: Change from ML to	Time to complete change assessment	1–3+ days	16m 46s
	Time to review	1–3+ days	7m 44s
	Updates / corrections	2+ days	Minimal

Deterministic Logic Model **Reviewer 1**	Notification of actions	5 mins + (× persons identified by task)	<1m
	End-to-end cycle time	**1 week +**	**26m 36s**
Change 3: UI/Software Control Change (Toggle tasks) **Reviewer 2**	Time to complete change assessment	1–3+ days	13m 22s
	Time to review	1–3+ days	6m 12s
	Updates / corrections	2+ days	Minimal
	Notification of actions	5 mins + (× persons identified by task)	<1m
	End-to-end cycle time	**1 week +**	**21m 24s**

Note. Change AIa-CMP intervals are derived from the system audit trails.

4. Discussion

The AI-CMP framework provides full regulatory coverage across the AIeMD lifecycle. In practice, not all sections are relevant to every change scenario. Rather than requiring uniform completion, the AI*a*-CMP system dynamically evaluates the context of the change and identifies sections that are not applicable (N/A), supported by a clear, audit-ready rationale. This capability was reflected in validation feedback, where concerns regarding the perceived burden of a full 35-section assessment were primarily associated with the manual (DOCX) format and in response, a companion Standard Operating Procedure (SOP) was developed to support identification of a proportional and scalable approach to assessment. This would enable users to determine appropriate scope based on the nature and significance of the change. However, in contrast, the agentic implementation applies contextual understanding of the change, combined with embedded SOP logic, to selectively present relevant sections and justify exclusions. This reduces unnecessary burden while maintaining completeness, traceability, and regulatory alignment. Furthermore, when the assessment is initially scoped by the AI*a*-CMP system, it can evaluate changes consistently and transparently based on its embedded knowledge base, while allowing human reviewers to modify outputs as required. This reduces the initial burden on individual team members, who would otherwise rely heavily on personal competency and judgement, thereby supporting greater standardisation and reducing variability in change assessment across projects, teams, and organisations.

Interpretation of expert reviews during AI-CMP validation activities demonstrate participant readiness for a comprehensive AI*a*-CMP assessment tool with a complete understanding of regulatory requirements associated with an AIeMD. The AI-CMP is grounded in established medical device change control principles and extends beyond traditional change processes by operationalising AI-specific risks including model drift, retraining effects, data distribution shifts, algorithmic bias, and AI Act obligations. The development and validation of the AI-CMP assessment tool, both in its original Word

(DOCX) format and its evolution into an interactive web application and as an Agentic AI application offers several significant strengths and contributions to the field of AIeMD design, regulatory science and digital health change management. This is illustrated in the new start organisation (MemoryTell Ltd) case study, where the AI*a*-CMP not only identifies the impact of the change but also highlights potential gaps in the Quality Management System processes for review by the team, with prompts for consideration of required deliverables in alignment with standards including ISO 13485 expectations. This demonstrates the system's ability to adapt assessment scope while maintaining regulatory rigour across different organisational contexts.

Notably, the AI*a*-CMP tool demonstrates clarity, usability, and structural robustness, as suggested by survey respondents who rated its usability, clarity of structure, and ease of navigation as "Very Good" or above. These ratings were observed across all usability dimensions in the aggregated results. Also, as seen from the survey of the AI-CMP tool, the agentic solution provides comprehensive regulatory coverage across global frameworks, with all participants rating coverage of MDR 2017/745, UK MDR 2002, FDA guidance, IEC 62304, and the EU AI Act in the top scoring tiers ("Very Good" to "Exceptional"). This content was used to develop the agentic application for real-world implementation. The AI-CMP data confirms the framework's relevance for diverse international markets and demonstrates its value as a harmonised, standards-aligned change assessment instrument that unifies regulatory expectations for AI-enabled devices into a single structured process. The tool in all formats delivers domain completeness, encompassing intended purpose, risk management, data management, AI QMS, software development lifecycle, AI training, model evaluation, post-market surveillance, PCCP readiness, and security lifecycle. Respondents rated all 13 domain completeness areas as "Very Good" to "Excellent." This breadth of domain coverage is a key contribution, given that AIeMDs require intersectional oversight spanning technical, regulatory, clinical, cybersecurity, and organisational domains. The tool demonstrates practicality and potential to adopt particularly when automated. Feasibility, suitability, acceptance, and overall helpfulness for QMS integration were not all favourable, predominantly due to the burden of a long list of questions spanning 57 pages in total when presented with guidance text across 35 sections. Qualitative comments reinforced this, with reviewers from both SMEs and large multinationals describing the tool as "*excellent*," "*comprehensive*," and "*particularly helpful for SMEs*," but "too long".

These results provide strong industry validation that the AI-CMP instrument can be readily incorporated into real-world quality systems and regulatory decision-making workflows which has now been achieved as a prototype for further longitudinal monitoring and independent review by competent teams and consultants. The AI-CMP assessment framework serves as a bridge between traditional medical device change control processes and emerging AI-specific regulatory expectations. By integrating requirements from ISO 13485, ISO 14971, IEC 62304, ISO 42001, IEC PAS 63621, BS ISO 30440, IEC

81001-5-1, and the EU AI Act, the AI-CMP and the agentic tool uniquely operationalises regulatory AI governance into a structured, repeatable assessment process. No equivalent tool currently synthesises these multi-standard obligations into a unified lifecycle evaluation template. The closest existing alternatives available to manufacturers are internal change control procedures, which are organisation-specific, unstandardised and not designed to address AI-specific risks such as drift or retraining effects; PCCPs [29], which reduce the regulatory burden of pre-specified, bounded changes but do not constitute a holistic change assessment across the lifecycle domains covered by the AI-CMP; and ad hoc regulatory impact assessments, which are by definition unstructured and assessor-dependent. The most directly comparable published work is the algorithm change protocol proposed by Gilbert et al. [39], which the AI-CMP extends by operationalising multi-standard requirements into a single, validated and digitisable assessment instrument, ref. Table 1. The digitisation of the AI-CMP into an Agentic AI application constitutes a major contribution to design and AI practice. It transforms a static regulatory template into an interactive, error-preventing, audit-ready system with built-in progress tracking, data persistence, structured and justifiable workflows with machine-readable export formats. This digital transformation directly addresses user feedback from the validation study, where participants expressed that a digital version would "further enhance" the tool's usability and impact.

Finally, the AI-CMP and agentic tool aim to contribute to patient safety and regulatory readiness by enabling manufacturers to systematically assess the impact of all AI-related changes. This can only be monitored through longitudinal studies that are not part of this research. The framework explicitly supports evaluation of data quality shifts, model drift, algorithmic updates, AI-specific risks such as bias, and post-market monitoring obligations, including such areas identified as having common root causes of historical SaMD and AI device failures. By providing a structured mechanism for assessing such risks, the AI*a*-CMP tool helps ensure that AIeMDs are assessed consistently to State-of-the-Art regulatory requirements. The opportunities and benefits from using an Agentic AI application are compelling. Preliminary quantitative data demonstrates significant reduction of time in performing an assessment of the change based on change description and validated AI-CMP framework. The three changes assessed by the AI*a*-CMP tool demonstrate that inconsistencies and inaccuracies are highlighted for correction, State-of-the-Art and medical device regulatory requirements are well defined, and actions reflect the change under consideration. To perform AI-CMP assessments comprehensively, which is expected by regulatory authorities, Agentic AI should improve the performance of devices on the market with more rigorous review of market data and internal parsing of documentation in real-time for more comprehensive assessments and accurate results. With an initial prototype study, the most we can claim is efficiency of review and speed for parsing existing documentation and auto-generation of proposals for updates to QMS and technical documentation. With this efficiency, it is anticipated that greater effectiveness

will follow, by improving consistency of change assessments, reducing assessor-to-assessor variability, and strengthening inspection readiness through standardised, traceable outputs. By enforcing structured evaluation logic and preserving complete audit trails, the AI*a*-CMP supports regulator expectations for explainability and evidence-based decision-making across repeated AI lifecycle updates. While at the same time, the approval process provides the human oversight and action necessary to mitigate against AI bias and other potential failures. This reinforces the PDCA cycle for Agentic AI with humans being in the loop and performing the final 'check' before acting. Compared to traditional QMS and SDLC-based change control approaches, which rely heavily on manual interpretation and fragmented documentation. The AI-CMP provides a unified, structured framework that integrates regulatory, technical, and AI-specific considerations. Its agentic implementation further enables dynamic adaptation of assessment scope and improved consistency, addressing key limitations of existing approaches in managing AI-driven lifecycle changes, ref. Table 2.

A dimension not addressed by the current study is the perspective of patients regarding AI involvement in the governance of devices that affect their care. The AI-CMP, as designed, focuses on manufacturer-side governance, that is, the processes by which a legal manufacturer evaluates, documents and justifies changes to an AIeMD prior to implementation. Patient-facing explainability of such governance decisions, and the extent to which patients would accept or expect transparency regarding AI involvement in change assessment, represents a distinct but complementary research question. This is particularly relevant given the transparency obligations of the EU AI Act [43] for high-risk AI systems, which extend beyond technical documentation to meaningful communication with affected persons. Future work should consider how AI*a*-CMP outputs, such as model cards, change summaries and performance monitoring reports, could be adapted to support patient-accessible transparency, engaging patient representatives in the design and evaluation of such outputs.

Table 2 Comparison of Benefits/Weakness of each mode of application

Dimension	Traditional Change Control	Manual AI-CMP Form	AI-CMP Web App	Agentic AI + Human Approval
Regulatory State of the Art Requirements	Low–Variable	High	High	Very high
Efficiency	Variable	Low	Medium	Very high
Consistency	Low	Variable	High	Very high
Depth of insight & competency	Human Dependent	Human Dependent	Human Dependent	Hybrid (AI + expert)

Scalability	Poor	Poor	Good	Excellent
Validation burden	Low	Low	Medium	High
Risk of error	High	High	Medium	Low (with review)
Inspector explainability	Low–Variable	High	High	High if well governed

4.1. Future Work

Following pilot implementation and continuous feedback via longitudinal studies the AI*a*-CMP application can be extended to provide an enterprise level AI*a*-CMP architecture with cross-rater analysis. The proposed architecture for enterprise level processing under Figure 11 can improve a) complete automated draft answers for high-burden narrative responses b) regulatory impact prediction using AI and machine learning models c) detection of missing evidence or incomplete sections d) context-aware recommendations linked to IEC/ISO clauses or AI Act provisions and emerging State of the Art guidance (MDCG, FDA, IMDRF, etc.). This would further reduce resource burden on manufacturers and strengthen consistency across assessors whilst improving compliance, safety and performance of AIeMDs. The automated system could cater for modular assessment by type and cater for territory-specific toggling. Also, an opportunity to re-introduce requirements for certification to IEC 27001 [46] or other device specific standards, feasible without being burdensome, once modular assessments are possible.

4.2. Limitations

The current evaluation focuses primarily on expert validation of usability, regulatory coverage, and domain completeness, rather than measuring outcomes such as patient safety, inter-assessor consistency, or quantified burden reduction. Accordingly, claims relating to these outcomes are positioned as potential benefits rather than empirically demonstrated results.

Content validation of the MS Word (DOCX) version of the AI-CMP assessment tool presents several methodological and contextual limitations that must be considered when interpreting the findings. While participants represented a mix of multinational organisations, SMEs, and micro-enterprises, the limited number restricts statistical generalisability. However, statistical generalisation was not the intention of this study. Expertise in AIeMD change management remains concentrated among a small number of specialists with access to international technical committees and awareness of standards still in development; the focus was therefore on gathering expert validation for content validity from this purposively selected cohort. Also, the respondent pool was not stratified across regulatory roles or AI development maturity levels. This is necessary for

representative generalisation, which is not required for this study. Since content validity is the primary concern, stratification is not required.

The validation focused on user experience, regulatory completeness perception, and structural clarity, but did not include empirical testing of the AI-CMP tool's ability to detect regulatory gaps, identify significant vs. non-significant changes, or improve decision consistency across evaluators. Such outcomes require scenario-based or comparative experimental evaluation. Evaluation is therefore positioned within the context of longitudinal real-world deployment, where the AI*a*-CMP system is currently in active use within a medical device organisation. Three (n=3) changes are presented as part of preliminary quantitative data demonstrating benefits from completion time. The assessors included a skilled QA/RA Director and CEO. Additional independent assessors were not feasible in the present company's case study due to nature of a small start-up team. This enables exposure to more complex, ambiguous, and edge-case change scenarios, providing naturalistic stress-testing under real regulatory and operational conditions. Implementation of the AI*a*-CMP has gone through QA/RA review and approval, and initial changes have been assessed with approval from a new start organisation (MemoryTell). This deployment is further strengthened by ongoing oversight from multidisciplinary teams, including Quality Assurance and Regulatory Affairs professionals, an independent expert consultant and clinical expert as CEO within the Quality Management System. Initial deployment feedback indicates that the AI*a*-CMP outputs are aligned with regulatory expectations and support efficient assessment processes and preliminary quantitative results indicate significant improvement in change assessments times. However, formal empirical evaluation of accuracy and efficiency remains outside the scope of this study.

5. Conclusion

The integration of AI into medical devices represents one of the most significant transformations in contemporary healthcare technology, introducing capabilities that far exceed those of traditional software. However, these advances come with new forms of risk, uncertainty, and regulatory complexity. This research indicates that the AI-CMP Assessment provides a comprehensive, expert-validated, and regulator-aligned framework to address the unique lifecycle challenges posed by AIeMD and SaMDs. Through design research and mixed methods exploratory validation study, the manual Word (DOCX) AI-CMP tool was assessed by domain experts as usable, complete in its regulatory coverage, with limited practicality for integration within medical device QMS. Feedback from multinational and SME stakeholders indicated that the AI-CMP is both methodologically sound and operationally relevant.

The development of the AI-CMP web application extends these capabilities further by transforming a static assessment document into a dynamic, structured, and auditable digital system. This digitalisation supports multidisciplinary collaboration, traceability, and documentation completeness.

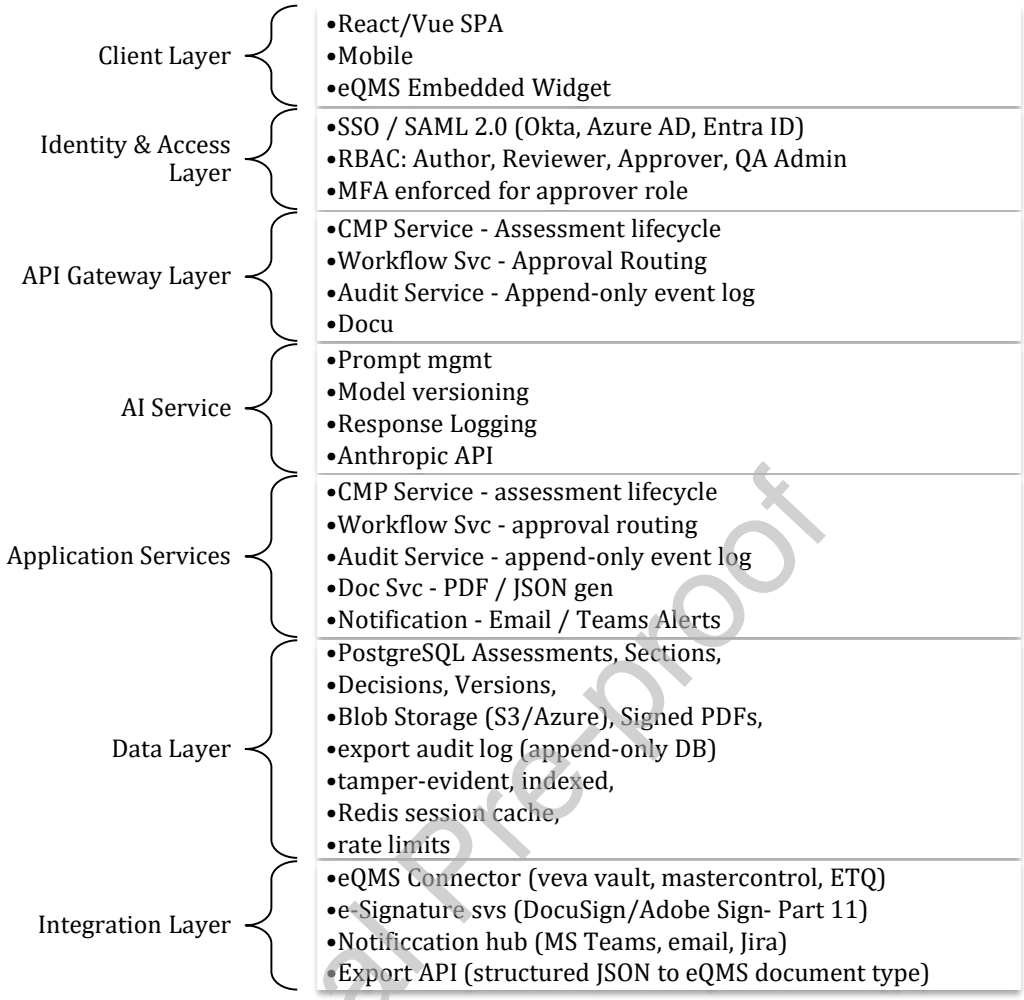


Figure 11 Future Recommendations for Enterprise Architecture

These are critical elements in ensuring the safe deployment of AI technologies in regulated environments. The remaining resource concerns may be mitigated by the agentic solution. The proposed agentic AI*a*-CMP architecture advances this work by demonstrating how autonomous systems, operating under human oversight, can streamline regulatory assessments, maintain continuous traceability, and reduce the risk of documentation errors or omissions.

Together, these contributions address a critical gap in AI lifecycle management: the absence of standardised, scalable tools capable of supporting manufacturers in evaluating the safety, regulatory impact, and performance implications of iterative AI updates. By unifying regulatory requirements such as ISO 13485, ISO 14971, IEC 62304, IEC PAS 63621, BS ISO 30440, IEC 81001-5-1, and the EU AI Act into one coherent methodology, the AI-CMP framework supports consistent, evidence-based decision-making across the TPLC.

As the regulatory landscape continues to evolve and AI systems become increasingly adaptive, tools such as the AI-CMP and their agentic extensions, will be essential in supporting safe innovation. This research provides a robust foundation for future advances in automated regulatory compliance, AI governance, and trustworthy AI for healthcare. Through continued refinement, integration with e-QMS and MLOps infrastructures, and further longitudinal studies from the agentic AI*a*-CMP case study, AI-CMP has the potential to become a cornerstone of AI-enabled medical device oversight, with potential for improving patient safety, ensuring regulatory compliance, and reducing resource burden for manufacturers worldwide. These findings should therefore be interpreted as initial expert validation of artefact content and usability, with subsequent real-world deployment providing a complementary phase of evaluation through practical application and continuous refinement in real-world use.

Acknowledgements. The author acknowledges the AI-CMP Assessment v4.0 and associated guidance developed with contributions of software, quality, clinical and regulatory industry colleagues who provided peer-review and feedback as well as a case study implementation by MemoryTell Ltd. Reviews by cross-functional teams are graciously thanked with survey responses from Medtronic, MemoryTell Ltd., Device Advice Ltd., Blue Bridge Technologies Ltd., and Siemens Healthineers. The author gratefully acknowledges the significant contribution of MemoryTell Ltd, who reviewed and approved the AI-CMP for use within their quality management system, adopted it as a prototype, and reviewed each change for accuracy, thereby contributing to the validation of the tool itself. The author also thanks Dr. Gilbert Regan of Dundalk Institute of Technology for his constructive feedback and his suggestion to develop a Standard Operating Procedure to accompany the AI-CMP. This research is supported by the Research Ireland (RI), Frontiers for the Future Programme under the Regulatory Compliance Framework for Trustworthy AI Medical Device Software (Reg-Fr-AIMs) project, ID 21/FFP-A/9255; a collaborative project undertaken by DkIT and UCD, Ireland**.**

Competing interests. The author provides independent regulatory affairs and quality management consultancy services within the medical device sector and from time to time advises MemoryTell Ltd. in QA/RA activities. MemoryTell Ltd. had no role in the design or development of the validation study, the analysis of survey data, or the decision to publish.

6. References

[1] IEC 60601-1, Medical electrical equipment-Part 1: General requirements for basic safety and essential performance (IEC 60601-1:2005 (EQV)), 2006.

[2] EN ISO/IEC-62304, Medical device software: software life-cycle processes, (2015) 84.

[3] IEC 82304-1, Health Software Part 1: General requirements for product safety, BSI Standards (2017).

[4] IEC-TR-80002-1, Medical device software. Part 1: Guidance on the application of ISO 14971 to medical device software, 1.0, British Standards Institution, 2009.

[5] EN/ISO-14971, Medical Devices - Application of risk management to medical devices EN ISO 14971:2019+A11:2021, 2021. www.gov.uk.

[6] ISO/DTS-24971-2, Medical Devices - Guidance on the application of ISO 14971-Part 2: Machine learning in artificial intelligence (draft) Technical Specification ISO/DTS 24971-2:2025, 2025. www.iso.org.

[7] IEC-81001-5-1, Health software and health IT systems safety, effectiveness and security. Part 5-1: Security - Activities in the product life cycle, British Standards Institution, 2022.

[8] IEC-PAS-63621, Artificial Intelligence-enabled Medical Devices - Data Management, IEC Technical Specification (2025). https://osd.iec.ch/#/iec/archive/22c98f06-0898-ba79-e063-6d13000a8b96/en/PNW/1 (accessed January 11, 2025).

[9] IEC-63450, Testing of Artificial Intelligence / Machine Learning-enabled Medical Devices, (2026).

[10] M.Y. Shaheen, AI in Healthcare: medical and socio-economic benefits and challenges, ScienceOpen (2021). https://doi.org/10.14293/S2199-1006.1.SOR-.PPRQNI1.v1.

[11] T. Riaz, M. Akram, U. Laila, M. Talha Khalil, R. Zainab, M. Iftikhar, F. Ahmet Ozdemir, G. Sołowski, E. Alinia-Ahandani, M. Altable, C. Egbuna, A. Sfera, M. Adnan, P. Parmar, Transforming Healthcare with AI, Transforming Healthcare with Artificial Intelligence (AI). IAIM 10 (2023) 41–50. http://iaimjournal.com/.

[12] M. Matheny, S.T. Israni, M. Ahmed, D. Whicher, AI in Health Care, Washington, DC, 2022.

[13] FDA, Artificial Intelligence & Medical Products: How CBER, CDER, CDRH and OCP are Working Together, (2024).

[14] S. Grossberg, A Path Toward Explainable AI and Autonomous Adaptive Intelligence: Deep Learning, Adaptive Resonance, and Models of Perception, Emotion, and Action, Front. Neurorobot. 14 (2020). https://doi.org/10.3389/fnbot.2020.00036.

[15] C.L. Curchoe, Unlock the algorithms: regulation of adaptive algorithms in reproduction, ScienceDirect, Fertility & Sterility 120 (2023) 38–43. https://doi.org/./j.fertnstert...

[16] J. Allgaier, R. Pryss, Practical approaches in evaluating validation and biases of machine learning applied to mobile health studies, Communications Medicine 4 (2024) 76. https://doi.org/10.1038/s43856-024-00468-0.

[17] F. Hasanzadeh, C.B. Josephson, G. Waters, D. Adedinsewo, Z. Azizi, J.A. White, Bias recognition and mitigation strategies in artificial intelligence healthcare applications, NPJ Digit. Med. 8 (2025). https://doi.org/10.1038/s41746-025-01503-7.

[18] EU, FRA, Data quality and artificial intelligence-mitigating bias and error to protect fundamental rights, Vienna, Austria, 2019. https://doi.org/10.2811/615718.

[19] S. Zhu, M. Gilbert, I. Chetty, F. Siddiqui. The 2021 landscape of FDA-approved artificial intelligence/machine learning-enabled medical devices: An analysis of the characteristics and intended use, Int. J. Med. Inform. 165 (2022). https://doi.org/10.1016/j.ijmedinf.2022.104828.

[20] B. Meskó, E.J. Topol. The imperative for regulatory oversight of large language models (or generative AI) in healthcare, NPJ Digit. Med. 6 (2023). https://doi.org/10.1038/s41746-023-00873-0.

[21] K. Skov Hilding, AI Trustworthiness Framework Part 1: Logging, Transparency and Human Oversight, 2025.

[22] R. Harrison, S. Fischer, R.L. Walpola, A. Chauhan, T. Babalola, S. Mears, H. Le-Dao. Where do models for change management, improvement and implementation meet? A systematic review of the applications of change management models in healthcare, J. Healthc. Leadersh. 13 (2021) 85–108. https://doi.org/10.2147/JHL.S289176.

[23] T.M. Welbourne, Change Management Needs a Change, Employ. Relat. Today 41 (2014) 17–23. https://doi.org/10.1002/ert.21449.

[24] S. Cummings, T. Bridgman, K.G. Brown, Unfreezing change as three steps: Rethinking Kurt Lewin's legacy for change management, Human Relations 69 (2016) 33–60. https://doi.org/10.1177/0018726715577707.

[25] MDCG-2025-6, Medical Devices Interplay between the MDR & IVDR and the Artificial Intelligence Act (AIA), Brussels, 2025.

https://health.ec.europa.eu/document/download/b78a17d7-e3cd-4943-851d-e02a2f22bbb4_en?filename=mdcg_2025-6_en.pdf (accessed June 20, 2025).

[26] EC, Proposal for a Regulation of the EP and Council amending MDR 2017/745 and IVDR 2017/746, European Commission, Brussels, Europe, 2025. https://www.who.int/health-topics/medical-devices#tab=tab_1.

[27] FDA, Quality Management System Regulation (QMSR) Rules and Regulations, Fed. Regist. 89 (2026) 7496–7525. https://www.fda.gov/medical-devices/postmarket-requirements-devices/quality-management-system-regulation-qmsr (accessed March 25, 2026).

[28] J. Kasic, AI Enters The 510(k) Submission World What Device Developers Need to Know, MedDevice Online (2025). https://www.meddeviceonline.com/doc/ai-enters-the-k-submission-world-what-device-developers-need-to-know-0001#:~:text=Beyond%20the%20surge%20in%20development%20of%20AI-enabled%20devices,Resource%20%28eSTAR%29%20mandatory%20for%20most%20510%20%28k%29%20submissions. (accessed March 24, 2026).

[29] F. Al-Faruque, FDA official discusses use cases, limits for device PCCPs _ RAPS, Regulatory News, A RAPS Publication (2024) 1–6. https://www.raps.org/news-and-articles/news-articles/2024/2/fda-official-discusses-use-cases,-limits-for-devic?mkt_tok=MjU5LVdMVS04MDkAAAGRhppSgHBLhGlcDnSeGSIOIqLE3qlV7t58ZBxaJT7kqJiLeTeKEmeMqS3iu9Ege7c6Na7dVfOeZrkuJuXLOZDhwa5u9NgB_1YK90jIWwxK (accessed February 27, 2024).

[30] D.R. Wallace, D.R. Kuhn, Failure Modes in Medical Device Software: An analysis of 15 years of recall data, 2001. https://doi.org/https://doi.org/10.1142/S021853930100058X.

[31] Z. Fu, C. Guo, S. Ren, Y. Jiang, L. Sha, Study of Software-Related Causes in the FDA Medical Device Recalls, in: Proceedings of the IEEE International Conference on Engineering of Complex Computer Systems, ICECCS, Institute of Electrical and Electronics Engineers Inc., 2017: pp. 60–69. https://doi.org/10.1109/ICECCS.2017.20.

[32] W.P. Chen, W.G. Teng, C.B. Kuo, Y.J. Yen, J.Y. Lian, M. Sing, P.T. Chen, Regulatory Insights From 27 Years of Artificial Intelligence/ Machine Learning–Enabled Medical Device Recalls in the United States: Implications for Future Governance, JMIR Med. Inform. 13 (2025). https://doi.org/10.2196/67552.

[33] H.K. Rajaram, J. Loane, S.T. MacMahon, F. McCaffrey, Taxonomy Based Testing and Validation of a new Defect Classification for Health Software, Journal of Software: Evolution and Process (2018). http://mc.manuscriptcentral.com/jsme.

[34] Sedgwick, Recall Index - European Union, 2023.

[35] Sedgwick, Recall Index - United States of America, 2023.

[36] M. McHugh. Integrating agile practices within a medical device plan driven software development lifecycle to overcome the challenge of changing requirements, Ph.D., Dundalk Institute of Technology, 2014.

[37] FDA, Artificial Intelligence-Enabled Device Software Functions: Lifecycle Management and Marketing Submission (draft) Recommendations, Rockville MD, 2025. https://www.fda.gov/vaccines-blood-biologics/guidance-compliance-regulatory-information-.

[38] ISO 13485, Medical Devices - Quality management systems - Requirements for regulatory purposes, 2016. www.ili-info.com.

[39] S. Gilbert, M. Fenech, M. Hirsch, S. Upadhyay, A. Biasiucci, J. Starlinger, Algorithm change protocols in the regulation of adaptive machine learning-based medical devices, J. Med. Internet Res. 23 (2021). https://doi.org/10.2196/30545.

[40] M. Nagendran, Y. Chen, C.A. Lovejoy, A.C. Gordon, M. Komorowski, H. Harvey, E.J. Topol, J.P.A. Ioannidis, G.S. Collins, M. Maruthappu, Artificial intelligence versus clinicians: Systematic review of design, reporting standards, and claims of deep learning studies in medical imaging, The BMJ 368 (2020). https://doi.org/10.1136/bmj.m689.

[41] EU, MDR 2017/745, 2017. https://eur-lex.europa.eu/eli/reg/2017/745/oj/eng (accessed June 19, 2025).

[42] UK Gov., The Medical Devices Regulations 2002 No. 618 (as amended), London, 2002. https://www.legislation.gov.uk/uksi/2002/618/contents (accessed March 25, 2026).

[43] EU, The AI Act 2024/1689, Brussels, 2024. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401689 (accessed April 8, 2024).

[44] EN/ISO-14971, Medical Devices - Application of risk management to medical devices, 2021. www.gov.uk.

[45] CEN/CLC/JTC 21, Artificial Intelligence. Quality management system for EU AI Act regulatory purposes., (2026).

[46] IEC-27001, Information security, cybersecurity and privacy protection-Information security management systems-Requirements, British Standards Organisation (2024).

Credit Author Statement

Niamh St John Lynch, MSc (ORCID: 0009-0009-4150-4970) – Dundalk Institute of Technology (DkIT), Ireland
Roles: Conceptualization, Methodology, Validation, Formal Analysis, Investigation, Resources, Data Curation, Writing – Original Draft, Writing – Review & Editing, Project Administration, Visualization.

Róisín Loughran, PhD (ORCID: 0000-0002-0974-7106) – Dundalk Institute of Technology (DkIT), Ireland
Roles: Supervision, Writing – Review & Editing.

Martin McHugh, PhD – Dundalk Institute of Technology (DkIT), (ORCID: 0000-0003-4275-3302) Ireland
Roles: Supervision, Writing – Review & Editing.

Fergal McCaffery, PhD (ORCID: 0000-0002-0839-8362) – Dundalk Institute of Technology (DkIT), Ireland
Roles: Supervision, Methodology, Writing – Review & Editing.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: The author Niamh St John Lynch is a member of several Technical Committees for Healthcare International Standards, and Chair of TC10 Medical Equipment, Systems and Software. Technical Committee expert membership is held under IEC TC 62 and CEN-CENELEC, including Software Network Artificial Intelligence Group (SNAIG) and SNAG Task 2 (Artificial Intelligence) members with insight into medical device and healthcare standards development. This does not negatively impact the quality of research or conflict with policies and procedures as a researcher. Also, from time to time, MemoryTell has sought expert advice from the author Niamh St John Lynch without remuneration, who is not employed and is not contracted to provide consultancy services to MemoryTell.