Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

From Trustworthy AI Principles to Verifiable System Requirements: A Multi-Stakeholder Engineering Framework for Clinical Decision Support Systems

Antonino Angi

antonino.angi@ub.edu

Universitat de Barcelona

Joan Perramon-Llussà

Universitat de Barcelona

Ilia Stepin

Universidad Autónoma de Madrid

Miriam Cabrita

Shine 2Europe (Portugal)

Laura Arbelaez Ossa

Universitat de Barcelona

A. Anil Sinaci

SRDC Software Research and Development Consulting

Senan Postaci

SRDC Software Research and Development Consulting

Carina Dantas

Shine 2Europe (Portugal)

Machteld J. Boonstra

Amsterdam University Medical Centers

Saskia Haitjema

University Medical Center Utrecht

Andreas Triantafyllidis

Centre for Research and Technology Hellas

Shishir Rao

University of Oxford

Stefan Smorenburg

Amsterdam Cardiovascular Sciences

Folkert Wouter Asselbergs

Amsterdam University Medical Centers

Karim Lekadir

Universitat de Barcelona

Research Article

Keywords: FUTURE-AI, Multi-stakeholder elicitation, Requirements engineering, System requirements specification, Trustworthy AI

Posted Date: August 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10802694/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: Competing interest reported. SR is an editor for BMJ Heart and reports consultancy fees from Lucem Health and grants from Oxford University Hospital Trust, the Medical Research Council, and F. Hoffmann-La Roche AG (Roche). The remaining authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

From Trustworthy AI Principles to Verifiable System Requirements: A Multi-Stakeholder Engineering Framework for Clinical Decision Support Systems

Antonino Angi[1*], Joan Perramon-Llussà[1], Ilia Stepin[1,7], Miriam Cabrita[3], Laura Arbelaez Ossa[1], A. Anil Sinaci[2], Senan Postaci[2], Carina Dantas[3,15], Machteld J. Boonstra[4,9,10], Saskia Haitjema[11], Andreas Triantafyllidis[12], Shishir Rao[13], Stefan Smorenburg[10,14], Folkert Wouter Asselbergs[4,5,6], Karim Lekadir[1,8]

[1*]Artificial Intelligence in Medicine Lab (BCN-AIM), Departament de Matemàtiques i Informàtica, Universitat de Barcelona, Barcelona, 08007, Spain.
[2]SRDC Software Research Development and Consultancy Corporation, Ankara, Turkey.
[3]SHINE 2Europe Lda, Coimbra, Portugal.
[4]Department of Cardiology, Amsterdam Cardiovascular Sciences, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, The Netherlands.
[5]Institute of Health Informatics, University College London, London, United Kingdom.
[6]National Institute for Health Research, University College London Hospitals, Biomedical Research Centre, London, United Kingdom.
[7]Departamento de Ingeniería Informática, Escuela Politécnica Superior, Universidad Autónoma de Madrid, Madrid, Spain.
[8]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.

[9]Department of Medical Informatics, Amsterdam University Medical Center, University of Amsterdam, Amsterdam, The Netherlands.
[10]Amsterdam Cardiovascular Sciences, Heart Failure and Arrhythmias, Amsterdam, The Netherlands.
[11]Central Diagnostic Laboratory, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands.
[12]Information Technologies Institute, Centre for Research and Technology Hellas, Greece.
[13]Deep Medicine, Nuffield Department of Women's and Reproductive Health, University of Oxford, Oxford, UK.
[14]Department of Surgery, Amsterdam University Medical Center, University of Amsterdam, Amsterdam, The Netherlands.
[15]School of Medicine and Biomedical Sciences, University of Porto, Porto, Portugal.

*Corresponding author(s). E-mail(s): antonino.angi@ub.edu;

Abstract

Background: A persistent gap separates the trustworthy AI principles articulated by consensus guidelines and regulatory instruments from the deterministic, verifiable specifications required to develop clinical decision-support systems. While frameworks such as FUTURE-AI operationalize trustworthiness through detailed sub-principles and lifecycle recommendations, the translation of these principles into formally structured engineering requirements remains under-specified. This paper presents a requirements engineering framework that systematically translates stakeholder needs, preferences, concerns, and regulatory constraints into formal, verifiable system specifications for clinical AI.

Methods: The framework follows a three-phase methodology: *(i)* hybrid multi-stakeholder elicitation combining bottom-up clinical, patient, and ethical inputs with top-down regulatory and contractual constraints, followed by scoping to define the operational system boundary; *(ii)* formalization of scoped inputs into structured, verifiable requirements via the ISO/IEC/IEEE 29148 standard, applying its syntactic template, obligation levels (*shall*/*should*/*may*), functional and non-functional typing, and per-requirement means of verification (MoV); and *(iii)* systematic mapping of the formalized requirements to the FUTURE-AI taxonomy as a normative coverage criterion rather than a descriptive classification scheme.

Results: Applied to the AI4HF project, a Horizon Europe initiative for personalized risk prediction in chronic heart failure, the framework yielded a fully specified, verifiable requirement corpus, demonstrating that it produces standards-compliant specifications with exhaustive FUTURE-AI coverage and per-requirement verification. The instantiation generated 173 requirements, each paired with a defined MoV; inter-rater validation of the FUTURE-AI mapping on

a stratified subset yielded Cohen's $\boldsymbol{\kappa} = 0.72$ (substantial agreement). The resulting corpus is publicly available via Zenodo [1], constituting, to our knowledge, the first openly available FUTURE-AI-aligned, ISO/IEC/IEEE 29148-compliant requirements dataset for a high-risk clinical AI system.
Conclusions: The methodology is designed to be reproducible across clinical AI projects beyond the AI4HF instantiation and is directly applicable to multi-stakeholder development efforts subject to high-risk regulatory classification (EU AI Act, MDR/IVDR), supporting trustworthy-by-design specification of healthcare AI systems.

Keywords: FUTURE-AI, Multi-stakeholder elicitation, Requirements engineering, System requirements specification, Trustworthy AI

1 Background

The deployment of machine learning (ML) and artificial intelligence (AI) in healthcare has precipitated a fundamental shift in diagnostic and prognostic practice. Clinical decision support systems (CDSS) that leverage multimodal patient data to generate personalized risk predictions are increasingly positioned as tools capable of augmenting clinical judgment and improving patient outcomes [2, 3]. However, translating the theoretical promise of such systems into safe, trustworthy, and deployable systems that comply with regulatory requirements remains a profound engineering challenge, one that is inadequately addressed by the current state of both AI ethics literature and clinical AI development practice [4–7]. Many policies have responded by articulating normative principles for trustworthy clinical AI, but these remain high-level and difficult to implement. Even binding regulation such as the European Union AI Act, which classifies AI systems used in medical diagnosis and treatment as high-risk and imposes obligations around transparency, human oversight, data governance, and conformity assessment, leaves open the question of how to translate these requirements into deployable terms and concrete engineering features [8]. The FUTURE-AI guidelines [9] derived from expert consensus across 117 interdisciplinary experts into six trustworthiness principles (*i.e.,* Fairness, Universality, Traceability, Usability, Robustness, and

Explainability), each decomposed into sub-principles, alongside seven *General* recommendations. These guidelines represent an important step toward operationalization. However, it goes beyond principles, setting out best practices that address technical, clinical, socio-ethical, and legal dimensions and span the entire AI lifecycle from design, development, and validation to regulation, deployment, and monitoring. By moving from abstract values to structured, lifecycle-spanning recommendations, they offer concrete guidance for the design and deployment of clinical AI. While a good foundation, an open challenge remains in translating such recommendations into specific engineering features and system-level design choices, particularly as they must be adapted to the requirements of individual clinical use cases. Without a formalized requirements engineering pipeline, the translation of trustworthy AI is relegated to ad-hoc interpretation by individual developers and deployers, introducing risks of non-compliance, regulatory misalignment, algorithmic bias, and breakdown of clinical trust [10].

This paper addresses this gap by proposing a systematic requirements engineering framework that translates heterogeneous multi-stakeholder inputs, ethical principles, and regulatory obligations into formal, verifiable specifications for clinical AI decision support systems. The framework formalizes trustworthy-AI principles into engineering syntax through three sequential phases. In the first, diverse stakeholder inputs are consolidated into a unified elicitation corpus, combining bottom-up empirical feedback (*e.g.,* patient narratives, clinician preferences, ethical analyses) with top-down requirements derived from health technology assessments, regulatory mandates and contractual obligations. In the second, each scoped need is translated into a formal system requirement using the syntactic templates of the ISO/IEC/IEEE 29148 standard [11], including an explicit obligation level (*shall*, *should*, *may*), end-to-end traceability to the originating stakeholder, and a defined Means of Verification (MoV). In the third, every requirement is mapped to the FUTURE-AI taxonomy, used as

a normative coverage criterion to ensure trustworthiness completeness. Because both the formalization rules and the coverage criteria are drawn from established international standards and open consensus guidelines rather than project-specific tooling, the pipeline is reproducible: any clinical AI development team can apply the same three phases with their own stakeholder constituencies and clinical domain to produce a comparably structured and auditable specification corpus. We demonstrate the framework on the AI4HF project [12], a Horizon Europe initiative developing a federated, multi-modal clinical decision-support system for personalized risk prediction in chronic heart failure (HF), deployed across European, South American, and African clinical sites. The instantiation produced 173 formally specified requirements covering the FUTURE-AI taxonomy, analyzed across four dimensions: stakeholder source, obligation level, trustworthy-AI principle coverage, and verification strategy.

1.1 Related Work

The use of AI in healthcare raises concerns about reliability, transparency, fairness, and clinical trust, which should be built into system design from the start [13, 14]. Many frameworks propose guidance to address these issues, such as FUTURE-AI, the EU Ethics Guidelines for Trustworthy AI [15], and the World Health Organization (WHO) guidance on AI for health [16]. However, there is still a gap between these high-level principles and the clear, testable outputs needed by software teams [17]. A recent review confirms that, despite numerous ethics frameworks, real-world use in healthcare remains limited and is rarely linked to patient safety [18]. Bridging this gap requires looking at the research areas that connect ethical ideas to system design. We group the literature into five dimensions: *(i)* stakeholder elicitation, which collects needs from clinicians, patients, developers, and regulators; *(ii)* governance translation, which turns ethical principles into design guidance; *(iii)* formal specification of requirements for AI, which encodes that guidance into syntactically well-formed system requirements

grounded in an engineering standard; *(iv)* verification and assurance, which establishes the means by which conformance to each individual requirement can be demonstrated; and *(v)* regulatory operationalization, which translates obligations from medical and AI regulations into formally specified, verifiable requirements through the creation of standards and implementing acts. These five steps are connected: elicitation feeds translation, translation shapes specification, specification enables verification, and the entire chain must align with regulation. While progress has been made in each area, they remain weakly integrated, which is the gap this paper addresses.

Within the first perspective, Labkoff et al. [19] recruited 200 participants to a consensus process that produced recommendations for trustworthy AI-enabled clinical decision support. Bergquist et al. [20] and Goisauf et al. [21] investigated trust dimensions and stakeholder perspectives in clinical radiology and medical AI, respectively, and emphasized that trustworthiness is a relational process that requires diverse engagement throughout the AI lifecycle. Lopez-Ramos et al. [22] employed co-creation workshops and Planguage [23] to gather low-level ethical requirements for a clinical AI project, demonstrating the feasibility of operationalization through stakeholder engagement while also revealing the limitations of Planguage as a formal specification tool and the absence of patient representation. Recently, Dantas et al. [24] proposed a structured methodology integrating participatory social innovation with requirements engineering for AI healthcare solutions. Their approach combines multi-stakeholder elicitation, MoSCoW prioritization, and traceable requirements management, producing a catalog of business, stakeholder, regulatory, and solution requirements aligned with the FUTURE-AI framework. However, the requirements neither follow an internationally recognized specification standard nor attach explicit MoV to each requirement. Muda and Sulaiman [25] adapted Quality Function Deployment methodology to map patient feedback from hospital reviews onto prioritized technical requirements for healthcare AI, demonstrating renewed interest

in systematic translation pipelines while remaining limited to a single stakeholder group and producing no formally verifiable specifications. In a recent contribution, Koutra et al. [26] adapted the MapStakes framework [27] into a five-step co-creation roadmap (MapStakes-PH) for structuring stakeholder engagement in public health; despite procedural rigor, its outputs remain at the level of participatory artifacts rather than formally expressed, verifiable system requirements aligned with an engineering standard.

A second perspective shifts focus from elicitation to governance, translating trustworthy AI principles into governance criteria and evaluation structures. Economou-Zavlanos et al. [28] developed lifecycle checkpoint evaluation criteria for algorithmic technologies at an academic medical center. Moreno-Sanchez et al. [29] enumerated informal requirements per the EU High-Level Expert Group (EU HLEG) trustworthiness principle using MoSCoW prioritization, an agile-based technique used in project management which classifies requirements as *Must have*, *Should have*, *Could have*, or *Won't have* [30]. Diaz-Rodriguez et al. [31] connected Trustworthy AI (TAI) requirements to regulation and auditing, proposing the concept of a responsible AI system. Solanki et al. [32] and Amugongo et al. [33] provided solution-based and lifecycle-integrated operationalization frameworks. None of these contributions applies a formal requirements engineering standard, and regulatory obligations are discussed as contextual background rather than formally expressed as verifiable non-functional requirements (NFR).

Within the third strand, work on requirements engineering for AI systems (RE4AI) addresses the elicitation, structuring, and formal expression of requirements. Ahmad et al. [34] systematically mapped 43 RE4AI approaches and identified challenges in adapting traditional RE practices to AI systems, particularly for human-centered and trust-related qualities. Habibullah and Horkoff [35] examined the specification of NFR for ML systems and reported difficulties in formulating qualities such as fairness,

explainability, and robustness with engineering decisions. While these studies show the difficulty of specifying AI requirements, they do not bind the resulting requirements to an internationally recognized engineering standard.

Within the fourth strand, attention shifts from how requirements are specified to how their satisfaction is demonstrated. Festor et al. [36] formalized safety constraints for a sepsis AI systems, illustrating the value of constraint-level verification while remaining confined to a single trustworthiness dimension. Jia et al. [37] explicitly identified the absence of pre-defined specifications against which clinical ML systems can be systematically assessed, framing this gap as a primary obstacle to reproducible evaluation. Taken together, these works highlight a gap between high-level evaluation of medical AI systems and explicit requirement-level verification.

Within the fifth strand, regulatory compliance in medical AI has been examined by Fraser et al. [38] and Muehlematter et al. [39], both highlight important gaps in translating regulatory obligations into development specifications. More recently, Niemiec et al. [40] specifically examined whether the EU AI Act's dataset quality and bias mitigation provisions are sufficient to address algorithmic bias in medical AI, concluding that while helpful, they stop short of providing the specification-level guidance that developers require. Shojaei et al. [41] extended this analysis by mapping the overlapping post-market obligations of the EU AI Act and Medical Device Regulation (MDR) for high-risk AI Software, identifying convergences and critical gaps, particularly around human oversight and system modification, that current regulatory guidance leaves unresolved at the specification level.

The remainder of this section consolidates the comparison in Table 1. The Table presents only those studies that satisfy at least one of the five evaluation dimensions defined in its caption; additional contributions discussed in the surrounding prose are excluded from the table when they offer no point of alignment along any of the five axes. As the table shows, no prior work simultaneously satisfies all five dimensions.

Reference	Formal RE Standard	Multi-Stakeholder Elicitation	TAI Coverage Criterion	Regulatory Operationalization	Per-Requirement Verification
Moreno-Sánchez et al. [29]	✘	✘	✔	✘	✘
López-Ramos et al. [22]	✘	✔	✔	✘	✘
Economou-Zavlanos et al. [28]	✘	✘	✘	✘†	✘
Labkoff et al. [19]	✘	✔	✘	✘	✘
Díaz-Rodríguez et al. [31]	✘	✘	✔	✘	✘
Ahmad et al. [34]	✘	✘	✘	✘	✘
Solanki et al. [32]	✘	✘	✘†	✘	✘
Amugongo et al. [33]	✘	✔	✘†	✘	✘
Bergquist et al. [20]	✘	✔	✘	✘	✘
Goisauf et al. [21]	✘	✔	✘	✘	✘
Festor et al. [36]	✘	✘	✘	✘	✘†
Jia et al. [37]	✘	✘	✘	✘	✘
Koutra et al. [26]	✘	✔	✘	✘	✘
Dantas et al. [24]	✘	✔	✘	✔	✘†
This work	✔	✔	✔	✔	✔

Table 1 Comparative analysis of related work against five dimensions. ✔ indicates the dimension is satisfied; ✘ indicates it is not. † indicates partial alignment that was conservatively classified as unmet under the strict binary evaluation protocol. The evaluated dimensions are: *(1)*: explicit citation and application of ISO/IEC/IEEE 29148 or equivalent international standard; only informal checklists, MoSCoW prioritization, and Planguage do not qualify. *(2)*: multi-stakeholder elicitation. *(3)*: use of a trustworthy-AI taxonomy as a normative coverage criterion; *(4)*: obligations from instruments such as the EU AI Act, GDPR, or MDR are translated into concrete and actionable requirements. *(5)*: a specific MoV is defined for each individual requirement produced. No prior work satisfies all five dimensions simultaneously.

2 Methods

In this section, we present how we built the proposed framework that translates heterogeneous multi-stakeholder inputs into formally structured, verifiable system requirements through three sequential phases: stakeholder elicitation and scope definition; formalization of requirements following ISO/IEC/IEEE 29148; and systematic mapping to the FUTURE-AI sub-principle taxonomy. Each phase produces an output that serves as input to the next, ensuring traceability from initial needs to final software specifications. Figure 1 gives an overview of the process and also follows one example requirement through all phases, showing how a stakeholder concern becomes a formal, verifiable and trustworthiness-linked requirement.

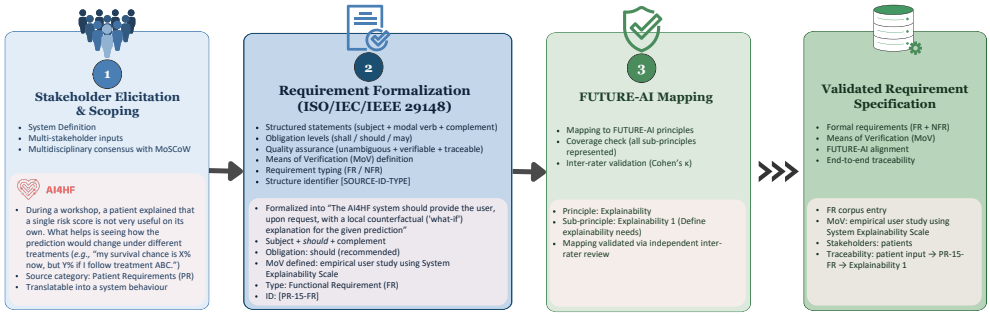

Fig. 1 Overview of the proposed three-phase requirements engineering framework (top row), which transforms multi-stakeholder inputs into a validated requirement specification, illustrated with a running example from the AI4HF project (bottom row). Multi-stakeholder inputs are elicited and scoped, formalized into ISO/IEC/IEEE 29148-compliant requirements with associated means of verification (MoV), and mapped to the FUTURE-AI taxonomy to ensure trustworthiness coverage. The framework supports end-to-end traceability from stakeholder concerns to validated software requirements. The bottom row follows requirement *PR-15-FR* through the framework stages.

2.1 Stakeholder Elicitation and Scope Definition

The elicitation phase follows a human-centered social innovation approach, where requirements emerge through continuous feedback between technical developers and the constituents that shape the socio-technical lifecycle of clinical AI [42]. We describe here the requirements engineering paradigm adopted in the project as a whole, building on the work of [43–46]. This paradigm is implemented through a two-layer structure. The first layer consists of multi-stakeholder sessions where different groups identify system concerns from their own perspective, describing general perspectives and barriers and facilitators. Each session is thematically scoped (*e.g.,* clinical, ethical, and trustworthiness aspects), and repeated at increasing levels of detail as the system evolves. The second layer consists of local working groups, which provide context-specific input throughout the project lifecycle. This structure separates *breadth* (cross-site workshops) from *depth* (local engagement), ensuring both general and clinically grounded requirements. A parallel desk-research stream complements this process by combining literature on ethics, privacy risks, and accountability with stakeholder input to align experience with existing guidance (Figure 2). Some sessions divided participants into

homogeneous discussion groups before reconvening in plenary, mitigating hierarchical dynamics (*e.g.,* between patients and clinicians) by letting each group voice concerns from its own expertise before negotiating shared perspectives; others kept mixed groups throughout, promoting early cross-stakeholder interaction and collaborative understanding from the outset.

Since requirements originate from both predefined project specifications and emergent stakeholder concerns, the framework adopts a hybrid elicitation approach. Top-down inputs derive from project-level obligations and alignment with established technical standards (*e.g.,* provenance modeling and lifecycle documentation frameworks), and are translated into functional (FR) and non-functional (NFR) requirement candidates. Bottom-up inputs arise from stakeholder dialogue and are subsequently validated against and, where needed, refined by the top-down specifications. This hybrid approach ensures that the resulting specification is project-compliant, standards-aligned, and grounded in stakeholder needs, reducing single-source bias. To structure this diversity into actionable engineering inputs, the project adopted a five-stage upstream pipeline [42]: (1) extract clinical cases from stakeholder dialogues during co-creation workshops; (2) construct storyboards that capture patient journeys and clinical workflows; (3) identify key challenges; (4) translate these challenges into candidate requirements; and (5) define use cases with actors and interactions.

Prior to formalization, each candidate stakeholder requirement underwent a consensus-based validation and prioritization step, conducted jointly by the project's participatory-research, clinical, and technical leads (with an ethics advisor for ELSI-sourced concerns) using the MoSCoW framework. Requirements that did not reach consensus were rejected or retained only as *rationale* annotations; accepted requirements were carried forward into the formalization phase described in Section 2.2.

2.2 Formalization via ISO/IEC/IEEE 29148

Each MoSCoW-accepted stakeholder requirement is formalized as a system requirement following ISO/IEC/IEEE 29148, the international standard for requirements engineering.

According to the standard, well-formed requirements must satisfy several properties: necessity (the requirement is essential for system capability or compliance), unambiguity (it has only one interpretation), verifiability (it can be tested against measurable criteria), feasibility (it can be implemented within known constraints), and traceability (it can be linked to a documented stakeholder source). To meet these criteria, each requirement is expressed using a standard structure with three components: a subject (the system or component responsible), a modal verb (indicating obligation), and a complement (describing the required behavior or property in implementation-free terms). This implementation-free framing preserves design flexibility while maintaining clarity. The modal verbs follow ISO/IEC/IEEE 29148 conventions: *shall* indicates a mandatory requirement; *should* indicates a recommended behavior that may only be omitted with justification; and *may* indicates an optional capability. When stakeholder inputs are expressed as negative constraints (*i.e.,* what the system should not do), they are reformulated as positive, testable statements, since negative requirements are difficult to verify directly, whereas their positive counterparts can be validated.

Each requirement is assigned a structured identifier of the form `[SOURCE]-[NN]-[TYPE]`, where *SOURCE* identifies the originating stakeholder group, *NN* is a sequential index, and *TYPE* denotes either *FR* or *NFR*. FRs describe system behaviors and data operations, while NFRs specify qualities and constraints such as security, performance, or regulatory compliance. To ensure feasibility, all requirements are reviewed by the technical partner responsible for deployment to confirm that they can be implemented within the target environment. To ensure verifiability, each

requirement is paired with a defined MoV, linking specification to testing. MoVs are grouped into three categories: (1) Automated System Audits (*e.g.,* CI/CD pipeline checks for bias and regulatory compliance); (2) Functional Testing (*e.g.,* validation of access controls and secure data transmission); and (3) Empirical User Validation (*e.g.,* structured usability assessments during user testing). NFRs related to usability, performance, and security typically require broader validation than unit testing, including user studies, performance benchmarking and security validation.

2.3 Systematic Mapping to the FUTURE-AI Sub-Principle Taxonomy

To ensure that the resulting requirements architecture captures the socio-technical complexity of clinical AI deployment, the FUTURE-AI framework [9] is used as the normative coverage criterion against which the requirement corpus is evaluated [42]. While widely recognized guidelines, such as those from the EU HLEG and the WHO, provide important general principles for trustworthy AI, they often lack the domain specificity required to formalize requirements in healthcare systems. In contrast, the FUTURE-AI framework provides a specialized, lifecycle-oriented taxonomy designed specifically for clinical AI and medical imaging. Rather than being applied solely as a post-hoc classification scheme, the taxonomy functions here as a normative coverage instrument: after the initial mapping of formalized requirements, an explicit gap analysis identifies under-represented principles, and targeted requirements are engineered through follow-up consultation with the relevant stakeholder groups to close those gaps. This iterative completeness loop preserves the integrity of FUTURE-AI as a generative criterion without requiring it to have driven the upstream elicitation instruments. The taxonomy comprises 30 sub-principles organized into two layers: 23 sub-principles grouped under six guiding principles, and 7 general recommendations addressing cross-cutting concerns such as stakeholder engagement, data privacy, risk

management, and regulatory compliance [9]. Each requirement is mapped to the most relevant FUTURE-AI sub-principle through an analytical review process. This involves comparing the requirement statement and its underlying rationale with the definitions of candidate sub-principles. To avoid ambiguity, each requirement is assigned to a single primary dimension reflecting its main motivation. For convenience, we report the FUTURE-AI principles and sub-principles in Additional file 1.Since mapping technical requirements to high-level ethical principles involves a degree of interpretation, an additional validation step is performed. A stratified subset of requirements was independently mapped by a second researcher, and inter-rater agreement is measured using Cohen's κ to assess consistency and reproducibility.

The final output of this process is a fully specified requirement corpus that integrates formal requirement statements, stakeholder rationales, FUTURE-AI mapping, and associated MoV.

3 Results

To evaluate the framework, we instantiate it on the AI4HF clinical decision-support system as a representative high-risk, multi-stakeholder validation case. The results below characterize the framework's outputs and properties through this instantiation, rather than reporting an AI4HF-specific deliverable. Applying the three-phase pipeline to this system produced a corpus of 173 formally specified requirements; these requirements, along with their stakeholder origins, FUTURE-AI mappings, and associated verification methods, are publicly available via the project's Zenodo repository [1].

3.1 Demonstration on the AI4HF System

AI4HF (*Trustworthy Artificial Intelligence for Personalised Risk Assessment in Chronic Heart Failure*) is a Horizon Europe project that develops a federated, multimodal clinical decision-support system for personalized risk prediction in chronic

HF [12]. The system integrates imaging, electrocardiography, cardiac biomarkers, and structured clinical reports across geographically distributed clinical sites in Europe, South America, and Africa. It provides a representative use case for the proposed framework due to its high-risk regulatory classification, cross-jurisdictional deployment, and socio-technical complexity. Following the hybrid elicitation design of Section 2.1, six stakeholder constituencies were engaged: Healthcare Professionals (HPR), Patients (PR), Ethical, Legal, and Societal Issues experts (ELSI), Health Technology Assessment specialists (HTA), technical/IT contributors and AI engineers (AIPP[1]), and the Grant Agreement (GA), the latter representing top-down contractual requirements derived from the AI4HF Description of Action (DoA). Top-down contractual and standards-aligned inputs (AIPP, GA) were combined with bottom-up inputs gathered across three thematically-scoped multi-stakeholder rounds, spanning clinical conceptualization, ELSI-focused concerns (including autonomy, confidentiality, re-identification, and accountability), trustworthy-AI co-design, and five ongoing local clinical/patient working groups. The full engagement structure, participant counts, and session-level detail are reported in Figure 2 and in the source participatory studies [43–46].

The AIPP constituency followed the same hybrid pathway: top-down inputs were aligned with established technical standards (W3C PROV [47] and PROV-ML [48] for provenance modeling, AI FactSheets [49] for governance documentation, MLOps [50] and ModelOps [51] for lifecycle management, and Google Model Cards [52] for structured reporting), while bottom-up refinements were gathered through the third multi-stakeholder round. The resulting AIPP requirements span five lifecycle phases: study definition; dataset and feature extraction; model generation and evaluation; deployment and monitoring; and passport generation.

[1]The AIPP label is retained for consistency with the requirement identifiers (*e.g.,* AIPP-04-FR) and encompasses the project's broader technical/IT/AI contributions, not solely AI passport-specific input.

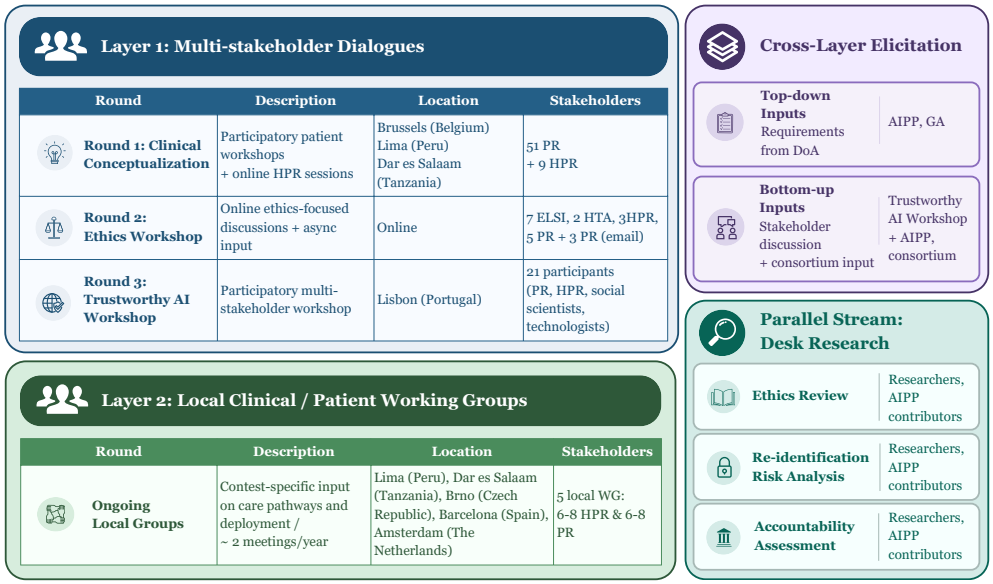

Fig. 2 Multi-stakeholder elicitation and scope definition pipeline applied to the AI4HF project. Layer 1 consists of three rounds of multi-stakeholder dialogue (clinical conceptualization, ethics, and trustworthy AI), generating co-created insights and structured qualitative data. Layer 2 consists of local clinical and patient working groups providing continuous site-specific input. A parallel desk-research stream supports ethical and technical analysis, while a cross-layer elicitation process integrates top-down project requirements with bottom-up stakeholder contributions across the system lifecycle. HPR: healthcare professionals representatives; PR: patient representatives; ELSI: Ethical, Legal, and Societal Issues experts; HTA: Health Technology Assessment specialists; AIPP: AI Product Passport developers; GA: the Grant Agreement; WG: Working Groups; DoA: AI4HF Description of Action

The inputs collected were highly heterogeneous in origin, form, and level of detail, spanning patient narratives and feature requests, clinician workflow and interpretability concerns, and ELSI-articulated normative constraints, data-sovereignty requirements, and evaluation protocols [43]. The formalization phase (Section 2.2) was applied across the six stakeholder codes and the FR/NFR typology, pairing each requirement with a MoV drawn from automated audits, functional testing, structured empirical validation, or consensus-based assessment. The FUTURE-AI mapping (Section 2.3) was then applied to the full corpus, with inter-rater validation on a stratified subset assessing reproducibility.

3.2 Illustrative Translation Chain and Exemplars

To illustrate the chain from raw stakeholder concern to verifiable specification, we trace a single requirement through the three phases. During the participatory patient workshop, participants raised concerns that AI-based tools could produce unequal treatment between patients who accept AI-assisted care and those who do not [53, 54], articulating the risk that clinicians might unintentionally prefer algorithm-supported decisions over more conservative approaches. The formalization phase produced:

> **ELSI-18-FR:** *The system shall provide the user, as part of the output, with a notification about responsible decision-making irrespective of whether the clinician's decision relies on the system's prediction.*

This statement follows the ISO/IEC/IEEE 29148 structure: it includes a subject (*the system*), a mandatory modal verb (*shall*) and an implementation-independent description of the required behavior. The original concern, expressed as a negative constraint, was reformulated into a positive and testable requirement. Instead of stating what the system must not do (introduce bias based on consent), the requirement specifies an action the system must perform (provide a responsibility notification regardless of AI use). The requirement was mapped to the *Fairness 1* sub-principle, which focuses on identifying potential sources of bias early in the design process. The corresponding MoV involves evaluation by clinicians using the AI4HF System Usability Scale in an empirical study. The identifier (ELSI-18-FR) encodes the stakeholder source (ELSI), its sequence number, and its type (FR), ensuring full traceability from the original concern to the final specification. Although the concern originated from patients, the requirement was assigned to the ELSI category. This reflects the nature of the solution, which addresses ethical and normative safeguards, such as fairness and accountability, rather than a direct patient-facing functionality. Beyond this single trace, Additional

file 2 reports one functional and one non-functional exemplar for each FUTURE-AI domain, illustrating the syntactic uniformity of the translation across the full taxonomy.

3.3 Specification Corpus Overview

Table 2 presents the distribution of formal requirements by stakeholder constituency and type. The corpus includes 101 FRs and 72 NFRs. Although FRs are expected to dominate in interactive systems, the substantial share of NFRs, over 41% of the corpus, highlights the high regulatory and quality demands of clinical AI compared to conventional software. This distribution reflects the dual challenge of clinical AI development: systems must both perform complex computational tasks and comply with strict regulatory and quality standards required for clinical use.

The HTA group contributed the largest portion of the specification (59 requirements, 34.1%). This reflects its dual role in defining both clinical evaluation criteria and regulatory requirements, including those associated with the EU AI Act, the MDR [55], and the European Health Data Space [56]. HPR contributed 42 requirements (24.3%), with a majority of FRs (26 FRs vs 16 NFRs). Their contributions focus primarily on clinical interaction design, including risk prediction outputs, treatment assessment, comorbidity input, and decision-support notifications. Patients contributed 32 requirements (18.5%), with a relatively balanced distribution between FRs and NFRs, 18 and 14 respectively. This indicates that patient input addresses both system functionality (*e.g.,* personalized care plans) and quality aspects (*e.g.,* usability, accessibility, and informed consent). ELSI experts contributed 22 requirements (12.7%), mostly functional (18 FRs), focusing on data governance, consent management, and fairness mechanisms as concrete system behaviors. The AIPP/DoA and GA sources each contributed 9 requirements (5.2% each). As described in Section 3.1, the AIPP/DoA category combines requirements from the AI developers module and the project's DoA,

covering metadata management, provenance tracking, and interoperability. The GA category captures contractual obligations, including requirements related to federated learning infrastructure and clinical study design.

Group	FR	NFR	Total
Healthcare Professionals (HPR)	26	16	42
Patient-based requirement (PR)	18	14	32
Ethical, Legal, Societal Issues (ELSI)	18	4	22
Health Technology Assessment (HTA)	30	29	59
AI Product Passport / DoA (AIPP-related)	6	3	9
Grant Agreement (GA)	3	6	9
Total	**101**	**72**	**173**

Table 2 Distribution of formal requirements by stakeholder group and type. FR: Functional Requirement; NFR: Non-Functional Requirement.

3.4 Obligation-Level Distribution

During formalization, modal verbs were used to encode three levels of obligation, following ISO/IEC/IEEE 29148:2011[2]. Of the 173 requirements, 155 (89.6%) use *shall*, 15 (8.7%) use *should*, and 3 (1.7%) use *may*. The dominance of mandatory statements reflects the high-risk regulatory status of clinical AI under the EU AI Act [57], where most system behaviors must be explicitly enforced. Requirements expressed using *should* mainly originate from the ELSI and Patient groups. This suggests that ethical considerations and user experience needs are often context-dependent and less easily translated into strict, testable constraints [58, 59]. This distribution highlights a central challenge in multi-stakeholder AI design [60]: balancing nuanced human values with

[2]The 2011 edition was fixed by the AI4HF GA at the project's protocol-writing stage. The criteria formalized in this work, namely necessity, unambiguity, verifiability, feasibility, and traceability to a documented stakeholder source, are stable across the 2011 and 2018 editions.

the precise and deterministic requirements expected in software engineering. The use of a structured modal-verb scheme provides a transparent and auditable way to manage this balance while preserving the distinction between mandatory, recommended and optional requirements.

3.5 Trustworthiness Coverage Analysis

Table 3 reports the distribution of system requirements across the FUTURE-AI guiding principles, confirming exhaustive coverage of all 30 sub-principles. The taxonomy was applied as a normative coverage criterion rather than a post-hoc classification scheme. After the initial mapping of stakeholder-derived requirements, a gap analysis identified underrepresented sub-principles, and additional requirements were engineered to close those gaps. Inter-rater validation of the mapping was performed on a stratified subset of 35 requirements (20.2% of the corpus) drawn from the ELSI, HPR and HTA groups, covering both FRs and NFRs and the full range of guiding principles. The subsample size is consistent with established practice in content-analytic coding, where reliability subsets typically span 10-25% of the corpus [61, 62]. Cohen's κ was 0.72 at the guiding-principle level, indicating substantial agreement per the Landis-Koch benchmark [63] and 0.55 at the sub-principle level, indicating moderate agreement. Disagreements clustered within a single sub-guiding principle, reflecting the interpretative characteristic of mapping technical requirements to high-level taxonomies. The coverage distribution reflects the architecture of the target system rather than an imbalance in the specification process. Usability accounts for the largest share (67 requirements; 38.7%), driven by the clinical interface as the locus of clinician and patient interaction; within this category, the *Usability 1* sub-principle (early definition of intended use and user requirements) alone accounts for 55 requirements. The General category contributes 50 requirements (28.9%), concentrated on *General 2* (data privacy and security; 26 requirements) and *General 5* (regulatory compliance;

16 requirements), reflecting the regulatory weight of high-risk clinical AI systems [57]. Requirements derived from the EU AI Act, the GDPR and the MDR map naturally onto these general recommendations, indicating that FUTURE-AI integrates regulatory considerations within its taxonomic structure rather than treating them as external constraints. Traceability accounts for 26 requirements (15.0%), spanning logging, documentation, auditing, and human oversight. The model-level principles contribute between 7 and 8 requirements, meaning that individual model-level constraints (*e.g.,* bias mitigation in training) can apply across the entire system, whereas usability requirements must be re-specified for each user interaction.

FUTURE-AI Principle	**FR**	**NFR**	**Total**
Fairness	5	3	8
Universality	1	6	7
Traceability	16	10	26
Usability	38	29	67
Robustness	3	5	8
Explainability	4	3	7
General	34	16	50
Total	**101**	**72**	**173**

Table 3 Distribution of formal requirements by FUTURE-AI guiding principle and requirement type.

3.6 Means of Verification Distribution

Each of the 173 requirements is paired with at least one MoV, ensuring that every specification is both formally defined and testable. Table 4 reports their distribution across the five verification strategies.

Empirical User Validation is the single most common strategy, accounting for over half of the corpus. Most of these requirements rely on a structured instrument: the System Usability Scale [64], the Patient Satisfaction Scale [65], and the purpose-built

System Explainability Scale [66]; the last assessing whether end-users can interpret the system's explanations without additional support, whether those explanations aid clinical decision-making, and whether they increase decision confidence, while the remainder use other forms of empirical user evaluation, such as usability benchmarking and user-acceptance testing. Functional Testing, the second most common strategy, spans access-control and authentication checks, session and registration workflows, secure data-transmission and CRUD [67] operation validation, and stress, penetration, and cross-validation testing. The remaining requirements are verified through Automated System Audits of regulatory and standards conformity (*e.g.,* EU AI Act, MDR, GDPR), consensus-based assessment within the clinical working groups and consortium, and visual inspection of interface elements.

This distribution carries an important implication: verifying trustworthiness in this system rests substantially on human judgment. The strategies that require clinical or patient evaluators (*i.e.,* empirical validation and consensus-based assessment) together account for the majority of the corpus. Because trustworthiness in clinical AI depends on both technical performance and human experience, automated testing alone is insufficient [21, 68, 69]; validation must instead include structured empirical engagement with the stakeholders whose needs shaped the requirements.

Means of Verification	**Count**	**%**
Empirical User Validation	90	52.0
Functional Testing	52	30.1
Automated System Audits	18	10.4
Consensus-based assessment	9	5.2
Visual inspection	4	2.3
Total	**173**	**100**

Table 4 Distribution of requirements by MoV.

4 Discussion

This section interprets the broader implications of our multi-stakeholder framework.

4.1 Methodological Contribution

The framework presented in this paper addresses a critical gap in the literature. As highlighted in Table 1, no prior work has simultaneously applied a formal requirements engineering standard, integrated multi-stakeholder elicitation across six distinct groups, utilized a trustworthy AI taxonomy as an exhaustive coverage criterion, and formalized regulatory obligations as verifiable NFRs within a clinical AI context. The principal contribution of this work is not just the specific 173 requirements produced, which are bound to the AI4HF system, but the reproducible methodological framework that generated them. This integration is not a mechanical concatenation of existing instruments: ISO/IEC/IEEE 29148 and FUTURE-AI operate at different levels of abstraction (*i.e.,* engineering syntax versus ethical taxonomy), and making them interoperable required resolving translation problems neither addresses alone (Section 4.2). The methodological contribution lies in this translation layer, which renders trustworthy AI principles verifiable rather than merely classified. This three-phase pipeline (Section 2) offers a deterministic pathway from raw, heterogeneous stakeholder needs to deployable specifications, designed for adoption by any clinical AI development team.

This determinism rests on adopting ISO/IEC/IEEE 29148 as the requirements formalization standard, a choice that requires justification in light of the alternatives adopted in previous studies. Common alternatives include MoSCoW prioritization [29] and Planguage [22]. However, neither provides the syntactic rigor required for deterministic compliance verification. MoSCoW assigns priorities but does not prescribe a structured requirement syntax or verifiability criteria, often leading to ambiguous specifications. Planguage offers greater structure than unrestricted natural language

but does not enforce the strict subject-verb-complement syntax and modal verb taxonomy (*shall*, *should*, *may*) required to express technical obligations consistently. In our pipeline, MoSCoW is used solely as an upstream consensus and prioritization mechanism (Section 2.1), whereas ISO/IEC/IEEE 29148 is subsequently applied to formalize the selected requirements (Section 2.2). By defining both a formal property checklist and a strict syntactic template, ISO/IEC/IEEE 29148 constrains the solution space and substantially reduces the risk of misinterpretation during software implementation.

Although the framework was instantiated within the clinical domain, its three-phase pipeline is domain-agnostic and can be applied wherever AI systems are subject to high-risk regulatory classification and heterogeneous stakeholder constituencies such as finance and public resource allocation. Structured efforts to operationalize trustworthy AI requirements have begun to emerge outside healthcare, including in higher education [70], fintech [71], and the public sector [72]. These parallel efforts confirm both the generalizability of the gap this paper addresses and the methodological relevance of requirements engineering as the discipline best positioned to close it across high-stakes AI application domains.

4.2 Non-Deterministic Challenges in Clinical AI Specification

The formalization process showed several challenges that distinguish clinical AI specification from conventional software engineering.

The first challenge is the *formalization of inherently stochastic system properties*. Properties like model accuracy, fairness, or robustness cannot be specified as binary and dogmatic pass-or-fail constraints [73]. Our framework resolves this by separating the *mechanism* from the *outcome*. The mechanism is deterministic and formalized as a strict *shall*-statement. The stochastic outcome, however, is delegated to the corresponding MoV (*e.g.,* empirical evaluation against fairness metrics during clinical

trials). This separation preserves the deterministic syntax required by ISO engineering standards while accommodating the probabilistic reality of AI.

The second challenge involves the *conversion of negative stakeholder constraints into positive system assertions*. ISO/IEC/IEEE 29148 requires positive formulation because negative constraints (*e.g.,* "the system shall not expose patient data") are impossible to deploy or exhaustively test. Yet stakeholder inputs, particularly from patients and ethicists, are frequently expressed as prohibitions. Our scoping phase systematically translates these negative concerns into positive, testable assertions (*e.g.,* "the system shall restrict access to patient data to authenticated users").

The third challenge concerns the *tension between syntactic and functional atomicity*. Although ISO/IEC/IEEE 29148 recommends strictly atomic requirements, a small subset of the corpus ($< 12\%$) includes multiple clauses within a single statement. These cases typically encode conditionally linked or sequential behaviors, such as detection followed by a corresponding response, that form a single functional unit. Splitting such statements into separate requirements could sever this logical connection and weaken traceability to the original stakeholder need. We therefore distinguish between *syntactic atomicity*, the formal property of expressing one obligation per statement, and *functional atomicity*, the preservation of a coherent behavioral unit, retaining the latter when decomposition would fragment the underlying intent.

4.3 The Role of FUTURE-AI as a Coverage Criterion

A distinctive feature of our approach is using the FUTURE-AI taxonomy not merely as a post-hoc classification scheme but as an exhaustive *coverage criterion*: a classification scheme describes the distribution of requirements, whereas a coverage criterion enforces completeness. After the initial mapping, each of the 30 sub-principles was inspected for representation, and additional requirements were engineered to close the gaps the audit exposed. This iterative loop, rather than the upstream elicitation,

which was not itself taxonomy-driven, is what turns the taxonomy from a descriptive tool into a normative instrument that shapes the resulting architecture. We selected FUTURE-AI over broader alternatives (EU HLEG [15], WHO [16]) for its domain specificity and lifecycle-oriented structure spanning design, development, validation, regulation, deployment, and monitoring [9]. The constituency distribution is revealing: of the 59 HTA-contributed requirements, 37 (62.7%) map to the General category (*i.e.,* authentication, data governance, and regulatory compliance), surfacing a practical risk that a specification process fixated on algorithmic fairness or explainability will miss the socio-technical infrastructure required for safe clinical deployment.

4.4 Regulatory Formalization

A key methodological contribution of this work is the direct formalization of regulatory obligations into verifiable system requirements. In the existing digital health literature, legal instruments like the EU AI Act, the GDPR, or the MDR/IVDR are typically relegated to the background context or treated as downstream compliance steps [38, 39]. Instead, our framework treats these regulations as direct sources of requirements that shape the entire system lifecycle, whether or not market access is the immediate aim. For example, the EU AI Act's obligation for human oversight [8] was formalized into specific, functional human-in-the-loop mechanisms for the AI4HF system, complete with defined verification testing. This embeds regulatory compliance into the system's architecture from the outset, following the "trustworthy-by-design" principle.

Beyond compliance, because each requirement carries an explicit verification criterion traceable to its regulatory source, the resulting specification set can serve as a structured basis for auditing and benchmarking clinical AI systems against their

obligations. This suggests a path from requirements engineering toward standardized, regulation-grounded evaluation—an increasingly pressing need as generative and multi-modal systems enter clinical use.

4.5 Limitations

Several limitations constrain the generalizability of these findings. First, the framework was instantiated within a single HF-focused project. Its application to other clinical domains (*e.g.,* oncology), AI architectures (*e.g.,* LLM-based systems), or non-European regulatory jurisdictions remains to be empirically validated.

Second, the FUTURE-AI mapping was independently validated through dual-coding, but only on a stratified subset (Section 3.5) covering roughly one-fifth of the 173 requirements, not the full corpus. The substantial guiding-principle-level agreement supports the mapping's reproducibility, though full-corpus dual-coding remains a refinement for future work.

Third, this framework targets high-level *system* requirements; the subsequent decomposition into low-level software architecture and code-level specifications falls outside the scope of this paper.

Finally, while the MoV dictates *how* to test a requirement, they do not prescribe quantitative acceptance thresholds (*e.g.,* 95% accuracy, $<$ 200ms latency), as these are highly specific to the deployment environment. Moreover, while we successfully engaged six distinct groups, our elicitation phase did not include direct, dedicated representation from external regulatory bodies or health insurers, relying instead on our internal HTA and GA channels to proxy these perspectives.

5 Conclusions

This paper presented a systematic requirements engineering framework for translating the heterogeneous outputs of multi-stakeholder engagement into formally structured,

verifiable system specifications for clinical AI decision-support systems. The framework's three-phase pipeline, comprising stakeholder elicitation and scope definition, formalization via ISO/IEC/IEEE 29148, and systematic mapping to the FUTURE-AI sub-principle taxonomy, addressed a structural gap that, as demonstrated through comparative analysis, no prior work has resolved: the simultaneous satisfaction of formal engineering rigor, multi-stakeholder representativeness, exhaustive trustworthiness coverage, regulatory translation into verifiable requirements, and per-requirement verifiability.

Applied to the AI4HF Horizon Europe project for personalized HF risk assessment, the framework generated 173 formally expressed system requirements, each paired with a defined MoV. The resulting specification corpus, publicly available at [1], constituted both a deployable engineering artifact for the AI4HF system and a methodological exemplar for the clinical AI community. The analysis revealed that the translation from ethical principle to engineering specification is neither mechanical nor trivial: it requires systematic conversion of negative constraints into positive assertions, calibration of obligation levels across stakeholder groups with divergent professional vocabularies, and separation of deterministic mechanism specifications from stochastic outcome evaluations.

The framework is directly applicable to multi-stakeholder clinical AI projects under high-risk regulatory classification (*e.g.,* EU AI Act, MDR/IVDR), and because it relies on externally maintained standards rather than project-specific tooling, any team can instantiate it with their own stakeholder constituencies and clinical domain. It is most useful during early-stage specification, before architectural decisions foreclose downstream verifiability, and complements rather than replaces the implementation methodologies applied further downstream.

Future work will pursue two directions. First, a multi-project instantiation study will apply the framework to clinical AI systems in other medical domains to assess

generalizability and identify domain-dependent adaptations of the syntactic templates and coverage criteria. Second, the framework will be extended downstream (from system requirements into software level specifications and architectural design decisions) to cover the full ISO/IEC/IEEE 29148 lifecycle while preserving end-to-end traceability from FUTURE-AI principles through stakeholder needs to implementation-level artifacts.

6 List of abbreviations

AI	Artificial Intelligence
ML	Machine Learning
CDSS	Clinical Decision Support System
HF	Heart Failure
ISO/IEC/IEEE 29148	International Standard for Requirements Engineering
MoV	Means of Verification
FR	Functional Requirement
NFR	Non-Functional Requirement
GA	Grant Agreement
DoA	Description of Action
ELSI	Ethical, Legal and Societal Issues
HPR	Healthcare Professionals
PR	Patient Requirements
HTA	Health Technology Assessment
AIPP	AI Product Passport
WG	Working Group
EU	European Union
MDR	Medical Device Regulation
GDPR	General Data Protection Regulation
SUS	System Usability Scale
MLOps	Machine Learning Operations
ModelOps	Model Operations

Declarations

Ethics approval and consent to participate

The stakeholder engagement activities underlying the requirements elicitation reported in this article were conducted as part of the AI4HF project's participatory requirements-gathering activities and are described in detail in [43–46]. The present article does not report a new study involving human participants; rather, it presents a requirements engineering framework and uses previously generated stakeholder inputs to demonstrate its application. No clinical data, biological samples, or identifiable patient information were collected or analyzed as part of the present framework study.

Consent for publication

Not applicable.

Availability of data and materials

The system requirements dataset used to instantiate the proposed framework is publicly available in the AI4HF project repository on Zenodo [1].

Competing interests

SR is an editor for BMJ Heart and reports consultancy fees from Lucem Health and grants from Oxford University Hospital Trust, the Medical Research Council, and F. Hoffmann-La Roche AG (Roche). The remaining authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This work received funding from the European Union's Horizon Europe research and innovation program under Grant Agreement No. 101080430 (AI4HF project).

Authors' contributions

AA carried out the conceptualization, methodology, investigation, writing of the original draft, writing, review and editing, and visualization. JPL, IS, MC, LAO, AAS, SP, CD, MJB, SH, AT, SR, SS, and FWA contributed to the methodology, investigation, validation, writing, review and editing, as applicable to their respective roles. MJB, SH, AT, SR, and SS contributed to the investigation and validation. FWA and KL contributed to funding acquisition, supervision, and project administration. All authors read and approved the final manuscript.

Acknowledgements

The authors gratefully acknowledge all individuals involved in the system requirements definition process, as well as all partners of the AI4HF consortium. The authors also thank Stefan Buijsman for his contributions as a member of the AI4HF ethics advisory board in the review and validation of the ELSI-related requirements.

References

[1] Angi A, Perramon Llussà J, Stepin I, Cabrita M, ten Broeke J, Sinaci AA, et al.: AI4HF System Requirements: A FUTURE-AI-Aligned, ISO/IEC/IEEE 29148 Requirements Dataset for Trustworthy Heart Failure Decision Support. Zenodo. Accessed: 2026-06-01.

[2] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nature medicine. 2019;25(1):44–56.

[3] Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ digital medicine. 2020;3(1):17.

[4] Bürger VK, Amann J, Bui CK, Fehr J, Madai VI. The unmet promise of trustworthy AI in healthcare: why we fail at clinical translation. Frontiers in Digital Health. 2024;6:1279629.

[5] Goktas P, Grzybowski A. Shaping the future of healthcare: ethical clinical challenges and pathways to trustworthy AI. Journal of Clinical Medicine. 2025;14(5):1605.

[6] Gkontra P, Quaglio G, Garmendia AT, Lekadir K. Challenges of machine learning and AI (what is next?), responsible and ethical AI. In: Clinical applications of artificial intelligence in real-world data. Springer; 2023. p. 263–285.

[7] Perramon-Llussà J, Skorupko G, Rao S, Pujadas ER, Kaderi SJE, Stepin I, et al. Big Data and Trustworthy AI for Heart Failure: A Review. Circulation: Heart Failure. 2026;https://doi.org/10.1161/CIRCHEARTFAILURE.125.013823.

[8] Busch F, Kather JN, Johner C, Moser M, Truhn D, Adams LC, et al. Navigating the European Union Artificial Intelligence Act for Healthcare. npj Digital Medicine. 2024;7(1):210.

[9] Lekadir K, Frangi AF, Porras AR, Glocker B, Cintas C, Langlotz CP, et al. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. bmj. 2025;388.

[10] Elechi U, Orobator ET, Udoh K, Ngozi EO, Uzoma CAE, Forson KAAM, et al. Artificial intelligence in healthcare: A narrative review of recent clinical applications, implementation strategies, and challenges. Journal of Healthcare Leadership. 2025;p. 863–876.

[11] IEEE. ISO/IEC/IEEE International Standard - Systems and software engineering – Life cycle processes –Requirements engineering. ISO/IEC/IEEE 29148:2011(E).

2011;p. 1–94. https://doi.org/10.1109/IEEESTD.2011.6146379.

[12] Consortium A. Trustworthy Artificial Intelligence for Personalised Risk Assessment in Chronic Heart Failure (AI4HF); 2023. Accessed: 09-04-2026. Available from: https://www.ai4hf.com/.

[13] Mora-Cantallops M, García-Barriocanal E, Sicilia MÁ. Trustworthy AI guidelines in biomedical decision-making applications: a scoping review. Big Data and Cognitive Computing. 2024;8(7):73.

[14] Goirand M, Austin E, Clay-Williams R. Implementing ethics in healthcare AI-based applications: a scoping review. Science and engineering ethics. 2021;27(5):61.

[15] Smuha NA. The EU approach to ethics guidelines for trustworthy artificial intelligence. Computer Law Review International. 2019;20(4):97–106.

[16] Organization WH. Ethics and governance of artificial intelligence for health: large multi-modal models. WHO guidance. World Health Organization; 2024.

[17] Morley J, Kinsey L, Elhalal A, Garcia F, Ziosi M, Floridi L. Operationalising AI ethics: barriers, enablers and next steps. AI & SOCIETY. 2023;38(1):411–423.

[18] Chan A, Rahimi-Ardabilli H, Rogers WA, Coiera E. The real-world impact of artificial intelligence ethics frameworks across a decade in healthcare: a scoping review. Journal of the American Medical Informatics Association. 2025;32(11):1767–1777.

[19] Labkoff S, Oladimeji B, Kannry J, Solomonides A, Leftwich R, Koski E, et al. Toward a responsible future: recommendations for AI-enabled clinical decision support. Journal of the American Medical Informatics Association.

2024;31(11):2730–2739.

[20] Bergquist M, Rolandsson B, Gryska E, Laesser M, Hoefling N, Heckemann R, et al. Trust and stakeholder perspectives on the implementation of AI tools in clinical radiology. European radiology. 2024;34(1):338–347.

[21] Goisauf M, Cano Abadía M, Akyüz K, Bobowicz M, Buyx A, Colussi I, et al. Trust, trustworthiness, and the future of medical AI: outcomes of an interdisciplinary expert workshop. Journal of Medical Internet Research. 2025;27:e71236.

[22] Lopez-Ramos LM, Pluktaite G, Bui CK, Amann J, Haven T, Madai VI, et al. Operationalizing AI ethics in medicine—a co-creation workshop study. BMC Medical Ethics. 2025;26(1):150.

[23] Gilb T. 8.1. 1 The Use of Planguage to Improve Requirement Specifications. In: INCOSE International Symposium. vol. 14. Wiley Online Library; 2004. p. 1604–1614.

[24] Dantas C, Cabrita M, Bobowicz M, op den Akker H, Rafael-Palou X, Hakkarainen T, et al. Integrating Participatory Social Innovation Into Requirements Engineering for AI Health Care Solutions: Case Study. Journal of Medical Internet Research. 2026;28:e87984.

[25] Muda N, Sulaiman MH. Adapting quality function deployment to translate patient feedback into prioritized technical requirements for healthcare artificial intelligence. Scientific Reports. 2026;.

[26] Koutra K, Kondylakis H, Kilintzis V, Guerra B, Triantafyllidis A, Tziraki C, et al. MapStakes-PH: A Methodological Framework for Co-Creation in Preventive Public Health. International Journal of Qualitative Methods. 2026;25:16094069261435583.

[27] Barquet K, Segnestam L, Dickin S. MapStakes: a tool for mapping, involving and monitoring stakeholders in co-creation processes. Stockholm Environment Institute; 2022.

[28] Economou-Zavlanos NJ, Bessias S, Cary Jr MP, Bedoya AD, Goldstein BA, Jelovsek JE, et al. Translating ethical and quality principles for the effective, safe and fair development, deployment and use of artificial intelligence technologies in healthcare. Journal of the American Medical Informatics Association. 2024;31(3):705–713.

[29] Moreno-SÃ PA, Del Ser J, Van Gils M, Hernesniemi J, et al. A design framework for operationalizing trustworthy artificial intelligence in healthcare: Requirements, tradeoffs and challenges for its clinical adoption. Information Fusion. 2025;p. 103812.

[30] Agile Business Consortium. What is MoSCoW Prioritization?; 2026. Accessed: 2026-07-08.

[31] Díaz-Rodríguez N, Del Ser J, Coeckelbergh M, De Prado ML, Herrera-Viedma E, Herrera F. Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. Information Fusion. 2023;99:101896.

[32] Solanki P, Grundy J, Hussain W. Operationalising ethics in artificial intelligence for healthcare: a framework for AI developers. AI and Ethics. 2023;3(1):223–240.

[33] Amugongo LM, Kriebitz A, Boch A, Lütge C. Operationalising AI ethics through the agile software development lifecycle: a case study of AI-enabled mobile health applications. AI and Ethics. 2025;5(1):227–244.

[34] Ahmad K, Abdelrazek M, Arora C, Baniya AA, Bano M, Grundy J. Requirements framework for engineering human-centered artificial intelligence-based software systems. arXiv preprint arXiv:230302920. 2023;.

[35] Habibullah KM, Gay G, Horkoff J. Non-functional requirements for machine learning: Understanding current use and challenges among practitioners. Requirements Engineering. 2023;28(2):283–316.

[36] Festor P, Jia Y, Gordon AC, Faisal AA, Habli I, Komorowski M. Assuring the safety of AI-based clinical decision support systems: a case study of the AI Clinician for sepsis treatment. BMJ health & care informatics. 2022;29(1):e100549.

[37] Jia Y, McDermid J, Lawton T, Habli I. The role of explainability in assuring safety of machine learning in healthcare. IEEE Transactions on Emerging Topics in Computing. 2022;10(4):1746–1760.

[38] Fraser AG, Biasin E, Bijnens B, Bruining N, Caiani EG, Cobbaert K, et al. Artificial intelligence in medical device software and high-risk medical devices–a review of definitions, expert recommendations and regulatory initiatives. Expert review of medical devices. 2023;20(6):467–491.

[39] Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis. The Lancet Digital Health. 2021;3(3):e195–e203.

[40] Niemiec E, Davis PA, Hauglid MK. Will the EU AI Act help to mitigate dataset bias in medical AI? Journal of Law and the Biosciences. 2026;13(2):lsag001.

[41] Shojaei S, Yakar D, Vellinga N, Bozgo V, Kwee T, Huisman H, et al. The AI Act and the MDR post-market requirements for semiautonomous AI SaMD: a

radiology case study in prostate cancer. Abdominal Radiology. 2026;p. 1–12.

[42] Camacho M, Perramon J, Puig-Bosch X, Dang VN, Díaz O, Lekadir K. Stakeholder engagement: the path to trustworthy AI in healthcare. In: Trustworthy AI in Medical Imaging. Elsevier; 2025. p. 471–493.

[43] Cabrita M, Machado N, op den Akker H, Cinà I, Llussà JP, Haitjema S, et al. Integrating participatory research and literature review to identify social and ethical requirements for responsible AI in heart failure treatment. In: 2025 10th International Digital Public Health Conference (DPH). IEEE; 2025. p. 1–6.

[44] Puig-Bosch X, Boonstra MJ, Cabrita M, Perramon J, Munive S, Guala A, et al. Requirements for Human-Centered Artificial Intelligence: A Heart Failure Study Across Europe and Latin America. In: 2024 20th International Symposium on Medical Information Processing and Analysis (SIPAIM). IEEE; 2024. p. 1–4.

[45] Smorenburg, S , Schuuring M , Cabrita, M . Deliverable D1.1; 2024. Project deliverable. Accessed: 2026-05-05.

[46] Cabrita, M , Machado, N , Freitas, A , Marinho, A , op den Akker, H , Dantas, C , Smorenburg, S , Stepin, I . Deliverable D1.2; 2024. Project deliverable. Accessed: 2026-05-05.

[47] Belhajjame K, B'Far R, Cheney J, Coppens S, Cresswell S, Gil Y, et al. Prov-dm: The prov data model. W3C Recommendation. 2013;14:15–16.

[48] Souza R, Azevedo L, Lourenço V, Soares E, Thiago R, Brandão R, et al. Provenance data in the machine learning lifecycle in computational science and engineering. In: 2019 IEEE/ACM Workflows in Support of Large-Scale Science (WORKS). IEEE; 2019. p. 1–10.

[49] Arnold M, Bellamy RK, Hind M, Houde S, Mehta S, Mojsilović A, et al. FactSheets: Increasing trust in AI services through supplier's declarations of conformity. IBM Journal of Research and Development. 2019;63(4/5):6–1.

[50] Kreuzberger D, Kühl N, Hirschl S. Machine Learning Operations (MLOps): Overview, Definition, and Architecture. IEEE access. 2023;11:31866–31879.

[51] Hummer W, Muthusamy V, Rausch T, Dube P, El Maghraoui K, Murthi A, et al. Modelops: Cloud-based lifecycle management for reliable and trusted ai. In: 2019 IEEE International Conference on Cloud Engineering (IC2E). IEEE; 2019. p. 113–120.

[52] Zaldivar A, Hutchinson B, Spitzer E, Raji ID, Vasserman L, Mitchell M, et al., editors. Model Cards for Model Reporting; 2019. Available from: https://dl.acm.org/citation.cfm?id=3287596.

[53] Burgess ER, Jankovic I, Austin M, Cai N, Kapuścińska A, Currie S, et al. Healthcare AI treatment decision support: Design principles to enhance clinician adoption and trust. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems; 2023. p. 1–19.

[54] Tsoupras G, Syed ZA. AI-Driven Decision Support Systems for Early Breast Cancer Detection: Adoption Implications in Healthcare Contexts [Master's Thesis]. Lund University; 2025. Available from: https://lup.lub.lu.se/student-papers/search/publication/9203488.

[55] Clemens N. The European medical device regulation 2017/745/EU: changes and impact on stakeholders. Journal of Clinical Research Best Practices. 2018;14(9):1–7.

[56] Fåhraeus D, Reichel J, Slokenberga S. The European Health Data Space: Challenges and Opportunities. Sieps–Swedish Institute for European Policy Studies; 2024. Available from: https://sieps.se/en/publications/2024/the-european-health-data-space/.

[57] Aboy M, Minssen T, Vayena E. Navigating the EU AI Act: implications for regulated digital medical products. NPJ Digital Medicine. 2024;7(1):237.

[58] Poszler F, Portmann E, Lütge C. Formalizing ethical principles within AI systems: experts' opinions on why (not) and how to do it. AI and Ethics. 2025;5(2):937–965.

[59] Mittelstadt B. Principles alone cannot guarantee ethical AI. Nature machine intelligence. 2019;1(11):501–507.

[60] Fordyce S. Value Sensitive Design: Shaping Technology with Moral Imagination: by Batya Friedman, and David G. Hendry, Cambridge, MA: The MIT Press, 2019, 229 pp. ISBN 9780262039536. N/A. 2020;.

[61] Lombard M, Snyder-Duch J, Bracken CC. Content analysis in mass communication: Assessment and reporting of intercoder reliability. Human communication research. 2002;28(4):587–604.

[62] Neuendorf KA. The content analysis guidebook. sage; 2017.

[63] Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics. 1977;p. 363–374.

[64] Hajesmaeel-Gohari S, Khordastan F, Fatehi F, Samzadeh H, Bahaadinbeigy K. The most used questionnaires for evaluating satisfaction, usability, acceptance,

and quality outcomes of mobile health. BMC Medical Informatics and Decision Making. 2022;22(1):22.

[65] La Monica EL, Oberst MT, Madea AR, Wolf RM. Development of a patient satisfaction scale. Research in nursing & health. 1986;9(1):43–50.

[66] Kibria MG, Kucirka L, Mostafa J. Assessing AI Explainability: A Usability Study Using a Novel Framework Involving Clinicians. In: 2025 IEEE 13th International Conference on Healthcare Informatics (ICHI). IEEE; 2025. p. 553–564.

[67] Martin J. Managing the data base environment. Prentice Hall PTR; 1983.

[68] Rees N, Holding K, Sujan M. Information governance as a socio-technical process in the development of trustworthy healthcare AI. Frontiers in Computer Science. 2023;5:1134818.

[69] Khashu K. Operationalizing trustworthy artificial intelligence in clinical and operational workflows. Frontiers in Digital Health. 2026;8:1779041.

[70] Tirado AM, Mulholland P, Fernandez M. Towards an operational responsible AI framework for learning analytics in higher education. arXiv preprint arXiv:241005827. 2024;.

[71] Pasha MZ. AI Governance Framework for Fintech: A Structured Approach for Responsible Artificial Intelligence. Available at SSRN 5372304. 2025;.

[72] Ologun A, Ikeakaonwu OM, Ibidunmoye AF, Eneano GA, Chuwa CC, Kolawole LO. AI governance and ethical frameworks for public-sector AI decision systems: an institutional operationalisation model. AI and Ethics. 2026;6(3):294.

[73] Vogelsang A, Borg M. Requirements engineering for machine learning: Perspectives from data scientists. In: 2019 IEEE 27th international requirements

engineering conference workshops (REW). IEEE; 2019. p. 245–251.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1FUTUREAIframework.csv
- Additionalfile2Domainexemplars.csv