ARM-AI: A HUMAN-CENTRED FRAMEWORK FOR AI-ASSISTED RISK MANAGEMENT IN AGILE SOFTWARE PRODUCT DEVELOPMENT

Claudiu Florin Iuonas [1][0009-0001-8031-3212], Diana Cristina Dragomir [2][0000-0001-6833-8815], Aurel Mihail Titu [3][0000-0002-0054-6535]
[1]National University of Science and Technology POLITEHNICA Bucharest, Faculty of Industrial Engineering and Robotics, 313 Splaiul Independenței, 6th District, Bucharest, Romania
[2]Department of Design Engineering and Robotics, Faculty of Industrial Engineering, Robotics and Production Management, Technical University of Cluj-Napoca, 103-105 Muncii Boulevard, 400641 Cluj-Napoca, Romania
[3]Industrial Engineering and Management Department, Faculty of Engineering, "Lucian Blaga" University of Sibiu, 10 Victoriei Street, 550024 Sibiu, Romania
E-mail(s): diana.dragomir@muri.utcluj.ro , mihail.titu@ulbsibiu.ro (✉)

Abstract - Agile software product development takes place in environments characterised by changing requirements, complex technical dependencies, cybersecurity concerns, resource constraints, and continuous pressure for rapid delivery. Although risk management is essential in such environments, existing practices are frequently fragmented, reactive, and dependent on the individual experience of product and delivery professionals. This paper proposes ARM-AI, a human-centred conceptual framework for integrating artificial intelligence into risk management activities throughout the software product lifecycle. The framework is designed to support, rather than replace, professional judgement by assisting teams in identifying emerging risks, organising relevant information, evaluating risk priority, recommending response alternatives, and monitoring changes over time. ARM-AI consists of four interconnected layers: organisational and project context, AI-assisted risk analysis, human decision and governance, and continuous learning and feedback. The framework also defines the responsibilities of Product Owners, Product Leads, Delivery Managers, technical specialists, stakeholders, and AI-enabled support mechanisms. Particular attention is given to transparency, explainability, accountability, data quality, and the prevention of excessive reliance on automated recommendations. The contribution of the paper is a structured and adaptable model that connects artificial intelligence capabilities with established risk management and Agile governance practices. As the proposed framework is conceptual, the paper does not report experimental or organisational performance data. Future research should validate ARM-AI through expert evaluation, case studies, and controlled implementation in real software development environments. The framework may provide a foundation for organisations seeking to introduce AI-supported decision-making while maintaining effective human oversight and clear managerial responsibility.

Keywords: Artificial intelligence, Risk management, Agile software development, Human-centred AI, Product management, Decision support, Explainable AI, AI governance.

1. Introduction

1.1 Context of Agile Software Product Development

Risk management is broadly defined as the coordinated set of activities and methods that an organisation uses to identify, assess, and treat sources of uncertainty that may affect the achievement of its objectives [1]. In Agile software product development, this generic logic operates within an environment characterised by iterative delivery cycles, evolving requirements, cross-functional collaboration, and continuous pressure to shorten time to market, conditions that are further complicated when delivery spans multiple teams, vendors, or cloud-based infrastructures [2].

Because Agile teams distribute planning and prioritisation decisions across multiple roles and

dx.doi.org/10.17683/ijomam/issue25/v2.13

short delivery cycles, risk-related information tends to be generated locally, at the level of individual sprints or product increments, and is not always consolidated into a coherent organisational picture [3]. Layers of risk exposure, spanning technical, organisational, and strategic levels, therefore need to be addressed concurrently rather than sequentially, which complicates traditional, document-centric risk management approaches [4].

1.2 Current Challenges in Risk Management

Current risk management practice in Agile software product development is frequently reported as fragmented across tools, backlogs, and informal conversations, reactive rather than anticipatory, and dependent on the individual experience of Product Owners, Delivery Managers, and senior engineers rather than on a shared organisational method [2], [3]. When delivery is distributed across several teams or scaled through frameworks that coordinate multiple Agile teams, these fragmentation effects tend to be amplified, since risks originating in one team or layer of the organisation may not be visible to decision-makers responsible for related but separately managed workstreams [4]. This combination of fragmentation, limited traceability, and uneven documentation makes it difficult to sustain a consistent view of risk exposure across a product's lifecycle.

1.3 Opportunities and Limitations of Artificial Intelligence

Interest in applying artificial intelligence to project and product management activities has grown considerably in recent years, with systematic reviews documenting an expanding body of work on AI-enabled planning, estimation, and monitoring support [5]. Forward-looking analyses similarly suggest that AI may influence several project management knowledge areas, including risk, schedule, and communication management, by supporting information processing tasks that would otherwise depend entirely on manual effort [6]. At the same time, systematic evidence on the practical integration of AI into software project planning highlights persistent limitations, including data quality constraints, organisational readiness gaps, unclear accountability for AI-informed decisions, and limited practitioner trust in automated outputs [7]. These findings suggest that the opportunity presented by AI is real but conditional on governance and human oversight mechanisms that are not yet consistently addressed in the existing literature.

1.4 Research Gap

A growing stream of work considers how AI techniques might support risk analysis more broadly, for example by assisting with the identification of patterns and dependencies in complex risk landscapes [8]. Closer to the present context, the TAI-PRM framework proposes a structured way of blending trustworthy AI principles with project risk management for Industry 5.0 manufacturing settings [9]. While TAI-PRM demonstrates that trustworthy AI principles can be operationalised within a project risk management structure, it is not designed for the iterative, role-distributed, and continuously evolving context of Agile software product development, and it does not explicitly formalise a layered human decision and governance structure oriented towards Agile roles and ceremonies. To the knowledge of the authors, no existing conceptual framework simultaneously (a) targets Agile software product development specifically, (b) treats AI strictly as a decision-support mechanism embedded within a human-in-the-loop structure, and (c) defines explicit governance, explainability, and accountability mechanisms alongside a proposed organisational adoption roadmap. This gap motivates the conceptual framework proposed in this paper.

1.5 Research Objective, Research Questions, and Contribution of the Paper

The objective of this paper is to develop and describe ARM-AI, a human-centred conceptual framework that structures how artificial intelligence capabilities can be organised, governed, and applied to support, rather than replace, human judgement in identifying, consolidating, classifying, prioritising, and monitoring risks throughout the Agile software product lifecycle, while preserving human authority, responsibility, and accountability over all final decisions.

The paper is guided by three research questions:

RQ1: How can artificial intelligence capabilities be integrated into a human-centred risk management framework for Agile software product development?

RQ2: What components, roles, and information flows are required to ensure effective human oversight within an AI-assisted risk management process?

RQ3: What governance principles and implementation stages are necessary for the responsible organisational adoption of the ARM-AI framework?

The contribution of the paper is twofold. First, it synthesises literature from risk management, Agile software development, product management, decision-support systems, explainable artificial

intelligence, human-in-the-loop systems, and AI governance into a single structured framework. Second, it proposes a concrete organisational vocabulary, comprising four framework layers, ten conceptually supported risk-related activities, defined roles and responsibilities, and a phased adoption roadmap, that can serve as a reference point for future empirical validation and organisational experimentation.

2. Theoretical Background

2.1 Risk Management in Software and Product Development

Responsibility for risk-related decisions in software product development is typically distributed across several roles rather than concentrated in a single risk function. Product Owners are expected to balance value delivery with technical and delivery constraints, a balancing act that becomes more demanding as organisations scale Agile practice across multiple teams and coordinate ownership decisions at a programme level [10]. Studies of how the Product Owner role evolves when organisations adopt scaled Agile frameworks show that owners must adapt their practices to increasingly complex governance and dependency structures, which directly affects how product-related risks are surfaced and prioritised [11]. Complementing this perspective, empirical work on the Scrum Master role emphasises the function's contribution to process facilitation, impediment removal, and coordination across teams, all of which are closely related to how operational risks are detected and communicated in practice [12].

2.2 Risk Management in Agile Environments

Beyond individual roles, empirical accounts of risk management in scaled Agile environments describe deliberate efforts to introduce lightweight, iteration-aligned risk practices without reverting to heavyweight, document-driven processes [13]. These accounts reinforce the observation, introduced above, that risk information in Agile settings is naturally distributed across teams, sprints, and delivery layers [3], and that this distribution can either be treated as an obstacle to consolidated risk visibility or, if properly structured, used as a source of richer, more contextual risk information [4]. This tension between distributed information and the need for a coherent organisational view is a central concern that the ARM-AI framework proposed in this paper is designed to address.

2.3 Artificial Intelligence as a Decision-Support Mechanism

Within the framework proposed in this paper, artificial intelligence is conceptualised strictly as a decision-support mechanism: a set of capabilities that can consolidate information, surface patterns, and generate candidate recommendations, without independently authorising actions or bypassing human evaluation. This framing is consistent with broader discussions of how AI can be applied to risk analysis tasks that traditionally required extensive manual effort, such as scanning large volumes of unstructured information for early risk indicators [8]. It also reflects the conditional nature of AI's contribution to project and product management activities noted in systematic reviews of the field, where the value of AI support depends heavily on data quality, organisational context, and the presence of appropriate human oversight structures [5], [7]. Framing AI as decision-support rather than as an autonomous decision-maker is therefore not only an ethical or governance preference; it is also consistent with the current state of evidence regarding the reliability of AI-generated recommendations in complex organisational settings [9].

2.4 Human-Centred and Explainable Artificial Intelligence

Explainable artificial intelligence (XAI) is concerned with making the reasoning behind AI-generated outputs interpretable to the people who must act on them. Recent surveys distinguish between different families of explainability techniques and the trade-offs they present between fidelity, simplicity, and computational cost [14]. Complementary surveys similarly catalogue the range of approaches available for producing explanations across different classes of AI models, underlining that no single technique is universally suitable [15]. From a user-centred perspective, however, the value of an explanation depends less on its technical sophistication than on whether it helps the intended user make a better decision, a point that has motivated calls for explainability research to more directly engage with the needs of specific professional audiences [16].

This user-centred concern connects directly to the broader literature on human-AI decision making, which examines how the presence of an AI recommendation changes the quality, speed, and consistency of human judgement [17]. Human-in-the-loop approaches, in which humans retain review and intervention rights over automated outputs, have been proposed as a way of preserving accountability while still benefiting from AI-generated support, although empirical evidence indicates that simply inserting a human reviewer

does not automatically guarantee improved decision quality, and can in some conditions reduce the accuracy of the final decision relative to fully automated processing [18]. A broader survey of human-in-the-loop machine learning practices further shows that the design of the interaction, not merely its presence, determines whether human oversight adds genuine value [19]. Design mechanisms such as cognitive forcing functions, which deliberately require the user to engage analytically with a recommendation before accepting it, have been shown in controlled studies to reduce overreliance on AI-generated suggestions [20]. Conversely, research on automation bias and selective adherence documents a persistent tendency for decision-makers to follow algorithmic recommendations that align with their prior expectations while dismissing those that do not, which can undermine the intended checks-and-balances role of human review [21]. Related experimental work on the impact of AI errors within human-in-the-loop processes shows that the way errors are presented and reviewed materially affects whether human oversight actually corrects them [22]. Taken together, this literature indicates that human-centred design choices, rather than the mere presence of a human reviewer, determine whether an AI-assisted process remains genuinely human-governed.

2.5 AI Governance, Accountability, and Responsible Use

Formal guidance for governing AI systems has matured considerably in recent years. The NIST AI Risk Management Framework provides a voluntary, function-based structure, organised around governing, mapping, measuring, and managing AI risk, that organisations can adapt to their own context [23]. ISO/IEC 42001 complements this guidance with a certifiable management-system standard for organisations that develop, provide, or use AI-based products and services, establishing requirements for policies, roles, and continual improvement processes [24]. At a more conceptual level, responsible AI has been described as a practice that must move from high-level principles towards concrete organisational routines if it is to influence real decision-making [25], a transition that a recent systematic review of responsible AI governance literature finds to be uneven across sectors and organisational maturity levels [26]. Related work on human-centricity in AI governance argues that governance structures should be designed around the needs, capabilities, and accountability of the humans who interact with AI systems, rather than treating human oversight as an afterthought bolted onto a primarily technical system [27]. A recently proposed roadmap for responsible AI systems similarly emphasises the interdependence of domain definition, trustworthy AI design, auditability, and governance as components that must be designed together rather than in isolation [28]. Finally, legal analyses of hybrid human-AI decision-making in public administration highlight that accountability cannot be meaningfully assigned unless the respective contributions of the AI component and the human decision-maker are clearly documented and traceable [29]. These governance perspectives collectively inform the design principles adopted for the ARM-AI framework in Section 4.

3. Research Approach

This paper adopts a conceptual and literature-informed research approach rather than an empirical or experimental one. The ARM-AI framework was developed through a conceptual synthesis of theoretical and applied literature drawn from several adjacent fields, including risk management, Agile software development, software product management, project governance, decision-support systems, explainable artificial intelligence, human-in-the-loop systems, responsible AI, and AI governance. Concepts, principles, and structural elements identified in this literature were interpreted, reorganised, and integrated through an interdisciplinary analysis process aimed at framework construction rather than empirical testing.

It is important to clarify that this research approach does not constitute a systematic literature review in the methodological sense. No formal database search protocol, inclusion or exclusion criteria, PRISMA-style screening process, or bibliometric analysis was applied, and no quantitative synthesis of reviewed publications is reported. The literature cited throughout this paper was instead selected and organised through targeted conceptual synthesis and theoretical integration, with the explicit purpose of grounding the proposed framework in verifiable, peer-reviewed, and institutional sources rather than producing a exhaustive mapping of the underlying research fields.

4. The Proposed ARM-AI Framework

4.1 Definition and Purpose of ARM-AI

ARM-AI (AI-Assisted Risk Management for Agile software product development) is defined in this paper as a human-centred conceptual framework that structures the way artificial intelligence capabilities are organised, governed, and applied to support, and not replace, human judgement in the identification, consolidation, classification, prioritisation, and monitoring of risks across the Agile product lifecycle, while keeping decision authority, traceability, and accountability with

designated human roles at every stage. The purpose of the framework is not to automate risk management decisions, but to structure how AI-generated analysis, human review, and organisational learning can be connected in a way that is transparent, explainable, and auditable.

4.2 Design Principles

The design of ARM-AI is guided by nine principles. Human control requires that AI-generated outputs remain recommendations subject to review rather than binding decisions. Transparency and explainability require that the basis for a recommendation be communicable to the responsible human role, consistent with user-centred explainability research [16]. Traceability requires that both AI-generated inputs and the resulting human decisions be recorded in a form that supports later review, in line with legal analyses of accountability in hybrid decision-making [29]. Accountability requires that responsibility for final decisions remain with designated human roles, never with the AI component itself. Adaptability requires that the framework be applicable across different organisational sizes, product types, and delivery models. Data quality requires that AI-assisted analysis be understood as conditional on the completeness and reliability of its inputs, echoing evidence on the limitations of AI adoption in software project planning [7]. Proportionality requires that the intensity of governance controls be matched to the significance of the risk under consideration. Finally, security and privacy require that risk-related information be processed in accordance with applicable data protection and information security requirements, consistent with the management-system obligations described in ISO/IEC 42001 [24].

4.3 Organisational and Project Context Layer

The Organisational and Project Context Layer provides the contextual grounding on which AI-assisted analysis and human governance depend. It encompasses product objectives and success criteria, stakeholder expectations, organisational constraints such as budget and staffing, technical architecture and its documented dependencies, delivery dependencies between teams or external suppliers, applicable regulatory requirements, cybersecurity considerations, available resources, and relevant historical project information. This layer does not perform analysis; it defines the boundary conditions within which the AI-Assisted Risk Analysis Layer operates and against which the Human Decision and Governance Layer interprets recommendations.

4.4 AI-Assisted Risk Analysis Layer

The AI-Assisted Risk Analysis Layer conceptually supports several activities: consolidating risk-related information originating from multiple sources such as backlogs, incident records, and technical documentation; assisting in the identification of candidate risks; supporting classification of risks into categories; detecting recurring patterns across historical project information; helping identify dependencies between risks; providing prioritisation support based on defined criteria; generating candidate mitigation alternatives for human consideration; and supporting continuous monitoring of changes in risk exposure. These capabilities correspond conceptually to the broader potential of AI to support pattern recognition and information processing in risk analysis contexts [8], and to the conditional opportunities documented for AI in project management more generally [5], [6]. Consistent with the strict research integrity requirements of this paper, none of these capabilities is presented as already implemented, tested, or empirically validated within ARM-AI; they are described as conceptual functions that a future implementation of the framework would need to realise and evaluate.

4.5 Human Decision and Governance Layer

The Human Decision and Governance Layer is where designated human roles evaluate AI-generated recommendations, add contextual knowledge that the AI component may not have access to, accept, reject, or modify recommendations, approve risk responses, assign ownership for mitigation actions, record the rationale for decisions, and remain accountable for outcomes. The design of this layer draws directly on human-in-the-loop and human-AI decision-making literature. Because evidence shows that automation bias and selective adherence can weaken the effectiveness of human review [21], and that the presence of a human reviewer does not by itself guarantee improved decision quality [18], [19], ARM-AI's governance layer is expected to incorporate design elements such as structured justification requirements and periodic escalation triggers, conceptually informed by evidence on cognitive forcing functions that encourage more deliberate engagement with AI-generated recommendations [20] and on how errors surface within human-in-the-loop processes [22].

4.6 Continuous Learning and Feedback Layer

The Continuous Learning and Feedback Layer conceptually supports the incorporation of outcomes from human decisions, mitigation actions, and observed project developments into future AI-assisted analysis, so that recurring risk patterns and previously effective responses can inform later recommendations. This layer is intended to operate within the boundaries of applicable data governance and privacy requirements and under continued human oversight, consistent with the management-system obligations for continual improvement described in ISO/IEC 42001 [24] and with the broader governance concern that feedback mechanisms should not erode human accountability over time [26].

4.7 Information Flow Between Framework Layers

Information flows through ARM-AI in a cyclical pattern rather than a linear one. Contextual information from the Organisational and Project Context Layer feeds the AI-Assisted Risk Analysis Layer, which generates candidate recommendations for the Human Decision and Governance Layer. Human decisions, approvals, and documented rationales generated in that layer are then captured by the Continuous Learning and Feedback Layer, which in turn updates the organisational and project context available for the next analysis cycle. Table 2 summarises the primary function of each layer together with the nature of AI and human involvement, and Figure 1 illustrates the overall cyclical architecture described above.

Table 2. Components and functions of the ARM-AI framework

Layer	Primary function	Nature of AI / human involvement
Organisational and Project Context Layer	Supplies contextual information (objectives, constraints, architecture, dependencies, resources)	Human-maintained; no AI analysis
AI-Assisted Risk Analysis Layer	Consolidates information; supports identification, classification, and prioritisation of risks	AI-assisted; generates candidate recommendations only
Human Decision and Governance Layer	Reviews, modifies, approves, or rejects recommendations; assigns ownership; records rationale	Human-controlled; retains final authority and accountability
Continuous Learning and Feedback Layer	Captures outcomes of decisions and actions to inform future analysis cycles	Human-governed with AI-assisted consolidation

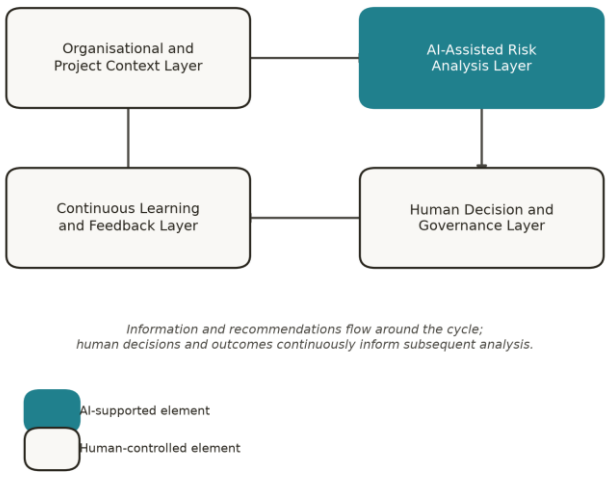

Figure 1: Overall architecture of the ARM-AI framework, showing the cyclical flow among the four layers and distinguishing AI-supported elements from human-controlled elements

5. Risk Categories Addressed by ARM-AI

ARM-AI is intended to conceptually support the identification and monitoring of risks across several categories relevant to Agile software product development: product risks related to market fit and value delivery; project and delivery risks related to schedule and scope; technical risks related to implementation quality; architectural risks related to structural design decisions and their long-term

consequences [4]; cybersecurity risks; operational risks affecting day-to-day delivery; organisational risks related to structure, staffing, and governance; stakeholder and communication risks; compliance and regulatory risks; and external dependency risks involving third parties, suppliers, or shared infrastructure [2].

Table 1 summarises these categories together with a short description and illustrative, non-empirical examples of the kind of situations each category may encompass. These categories are not presented as an exhaustive taxonomy but as a starting structure that organisations may adapt to their own product and delivery context.

Table 3. Main categories of risks addressed by ARM-AI

Risk category	Description	Illustrative examples
Product risks	Risks affecting product value and market fit	Misalignment with user needs; unclear value proposition
Project and delivery risks	Risks affecting schedule, scope, and delivery commitments	Delayed dependencies; scope volatility
Technical risks	Risks affecting implementation quality and technical feasibility	Unresolved technical debt; integration difficulties
Architectural risks	Risks related to structural design decisions and their consequences	Scalability constraints; coupling between components
Cybersecurity risks	Risks related to security of systems and data	Unpatched vulnerabilities; insecure data handling
Operational risks	Risks affecting day-to-day delivery activities	Resource unavailability; tooling failures
Organisational risks	Risks related to structure, staffing, and governance	Unclear ownership; staffing gaps
Stakeholder and communication risks	Risks related to alignment and information flow among stakeholders	Conflicting priorities; delayed feedback
Compliance and regulatory risks	Risks related to legal and regulatory obligations	Data protection requirements; industry standards
External dependency risks	Risks involving third parties, suppliers, or shared infrastructure	Vendor delays; shared platform outages

6. Roles and Responsibilities

ARM-AI defines responsibilities across a set of human roles and one AI-assisted component, reflecting the distributed nature of decision-making already established as a feature of Agile product development [10], [11]. The Product Owner is expected to weigh AI-generated risk information against product value and stakeholder priorities; the Product Lead to align risk-related decisions with the broader product strategy; the Delivery Manager to coordinate mitigation actions across teams; and the Scrum Master to facilitate the surfacing and discussion of risks within Agile ceremonies, consistent with the coordination-oriented view of the role documented in the literature [12].

The development team, software architects, and technical, cybersecurity, and compliance specialists are expected to validate the technical plausibility of AI-generated recommendations within their respective domains, while business stakeholders and organisational leadership retain approval authority for risks with strategic or organisational significance.

Table 3 clarifies the primary responsibility associated with each role, distinguishing between generating recommendations, validating recommendations, approving decisions, implementing mitigation actions, monitoring outcomes, and maintaining accountability. Consistent with the framework's design principles, the AI-assisted decision-support component is limited to generating candidate recommendations and is not assigned moral, legal, or managerial accountability for any outcome; this distinction follows directly from legal analyses showing that accountability in hybrid decision-making must remain traceable to identifiable human roles [29], and from evidence that unclear responsibility allocation is itself a source of governance risk in automation-supported decision environments [21].

Table 3. Human roles and responsibilities within the ARM-AI framework

Role	Primary responsibility
Product Owner	Weighs risk information against product value and stakeholder priorities
Product Lead	Aligns risk-related decisions with product strategy
Delivery Manager	Coordinates mitigation actions across teams
Scrum Master	Facilitates surfacing and discussion of risks within Agile ceremonies
Development team	Validates technical plausibility of recommendations
Software architects	Assesses architectural implications of identified risks
Technical, cybersecurity, and compliance specialists	Validates domain-specific recommendations
Business stakeholders	Provides approval for risks with strategic significance
Organisational leadership	Provides oversight and resourcing decisions
AI-assisted decision-support component	Generates candidate recommendations only; holds no moral, legal, or managerial accountability

7. Integration into the Agile Product Lifecycle

ARM-AI is intended to be integrated into existing Agile product lifecycle activities rather than operating as a separate process running alongside them. Conceptually, AI-assisted risk information could inform product discovery and strategy discussions by surfacing early technical or market signals; roadmap development and release planning by highlighting dependency and delivery risks; backlog refinement and sprint planning by flagging items with elevated technical or operational risk; daily delivery activities and dependency management by supporting the timely escalation of emerging issues; and sprint review, retrospective, release preparation, and post-release monitoring by consolidating observed outcomes for organisational learning. This pattern of integration is consistent with empirical accounts of introducing lightweight, iteration-aligned risk practices into scaled Agile environments without reverting to heavyweight, document-driven processes [13], and with broader evidence that risk-related activity in Agile settings is most effective when embedded within, rather than layered on top of, existing delivery rhythms [3], [4].

A central design requirement is that this integration should not create unnecessary bureaucracy or replace existing Agile events; ARM-AI is intended to inform the content of existing ceremonies, such as sprint planning or retrospectives, rather than introduce new mandatory meetings. Figure 2 illustrates this proposed integration by mapping ARM-AI touchpoints onto a representative Agile product lifecycle.

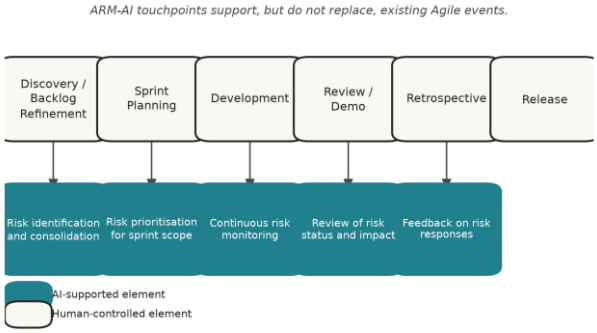

Figure 2: Proposed integration of ARM-AI touchpoints into a representative Agile product lifecycle, without introducing new mandatory ceremonies

8. Governance, Explainability, and Ethical Considerations

Effective governance of ARM-AI depends on maintaining human oversight over every AI-generated recommendation and on ensuring that the basis for each recommendation is explainable to the responsible human role, drawing on established XAI research on making AI-generated outputs interpretable to their intended users [14], [15], [16]. Governance practices consistent with the NIST AI Risk Management Framework's governing, mapping, measuring, and managing functions [23], and with the management-system requirements described in ISO/IEC 42001 [24], would need to be adapted to the specific context of Agile software product development rather than applied generically.

Data quality, algorithmic bias, privacy, information security, and access control are recurring concerns for any AI-assisted process and are treated in ARM-AI as governance requirements rather than as afterthoughts, consistent with the argument that responsible AI must operate through concrete organisational routines [25] and that governance maturity in practice remains uneven across contexts [26].

Overreliance on automated recommendations and automation bias represent a further governance risk. Evidence on cognitive forcing functions suggests that deliberately requiring analytical engagement before accepting a recommendation can reduce overreliance [20], while research on selective adherence shows that decision-makers may otherwise favour recommendations that confirm their expectations [21]; mechanisms for challenging or escalating AI-generated recommendations are therefore treated as a necessary governance feature rather than an optional addition, particularly given

evidence that the handling of AI errors within human-in-the-loop processes materially affects whether such errors are actually corrected [22].

Finally, responsibility allocation, regulatory compliance, and periodic human review are treated as ongoing governance obligations rather than one-time design choices, in line with the human-centred governance perspective [27] and the roadmap for responsible AI systems that emphasises the interdependence of governance, auditability, and accountability [28]. Figure 3 depicts the human-in-the-loop decision and governance process underlying these considerations, distinguishing the AI-supported analytical step from the human decision, escalation, and documentation steps that follow it.

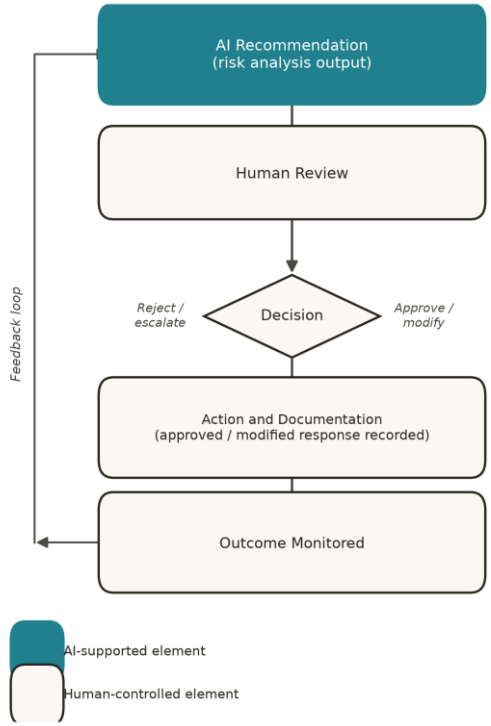

Figure 3: Human-in-the-loop decision and governance process, distinguishing the AI-supported analytical step from human decision, escalation, and documentation steps

9. Proposed Implementation Roadmap

This section proposes a conceptual, non-empirical organisational implementation roadmap for ARM-AI. The roadmap is intended as a structured starting point for organisations considering adoption of the framework and does not describe stages that have already been carried out; no organisation is reported in this paper to have implemented any part of ARM-AI. The roadmap comprises ten stages, summarised in Table 4, grouped conceptually into three broader phases: assessment and design, pilot and validation, and adoption and evaluation. This phased structure is broadly consistent with the incremental, feedback-oriented approach documented for introducing risk practices into scaled Agile environments [13] and with governance guidance that treats responsible AI adoption as an iterative rather than a one-off exercise [26], [28]. Figure 4 presents these ten stages positioned within the three adoption phases.

Table 4. Proposed stages for organisational implementation of ARM-AI

Stage	Description
1. Organisational readiness assessment	Assess organisational and technical readiness for AI-assisted risk management
2. Definition of risk categories and responsibilities	Define applicable risk categories and assign role responsibilities
3. Identification of relevant information sources	Identify data and information sources available for risk analysis
4. Establishment of data governance requirements	Define data quality, privacy, and security requirements
5. Selection of a limited pilot context	Select a bounded product or team context for initial piloting
6. Human validation of AI-generated recommendations	Establish review procedures for AI-generated outputs
7. Definition of governance and escalation controls	Define escalation paths and accountability mechanisms
8. Evaluation of usability, trust, and organisational fit	Assess practitioner usability, trust, and organisational fit
9. Gradual organisational adoption	Extend adoption incrementally beyond the initial pilot
10. Periodic framework evaluation and adjustment	Review and adjust the framework based on accumulated experience

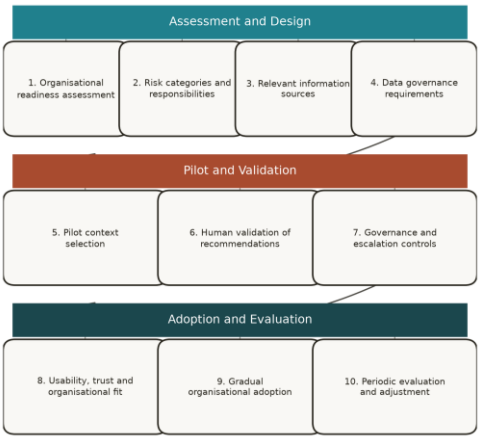

Figure 4: Proposed organisational adoption roadmap for ARM-AI, grouping the ten implementation stages into three broader phases

10. Discussion

As a conceptual framework, ARM-AI may support several potential organisational benefits, although none of these are demonstrated empirically in this paper. The framework could contribute to improved risk visibility by consolidating information that would otherwise remain fragmented across teams and tools [3]. It is expected to support greater consistency in risk-related activities across teams by structuring the same four-layer process regardless of team size or context. Earlier identification of emerging risks and more structured communication of risk information to relevant decision-makers are further plausible conceptual outcomes, consistent with the pattern-recognition potential attributed to AI in risk analysis contexts [8]. The framework also conceptually enables improved traceability of decisions, given its explicit design emphasis on documentation and accountability [29], and may contribute to stronger organisational learning through its Continuous Learning and Feedback Layer. Finally, the explicit role structure proposed in Section 6 has the potential to provide a clearer distribution of responsibilities than is typically documented in current fragmented practice. These potential contributions require future empirical validation before any claim of actual improvement can be made.

Several plausible barriers could limit the adoption or effectiveness of ARM-AI in practice. Poor data quality and fragmented information systems could constrain the reliability of AI-assisted analysis, echoing documented limitations of AI adoption in software project planning [7]. Limited trust in AI-generated recommendations, resistance to organisational change, and insufficient AI literacy among relevant roles may slow adoption regardless of the framework's technical design. Integration difficulties with existing tools and processes, lack of explainability in underlying AI components, unclear responsibility allocation, and privacy or security concerns represent additional plausible barriers, consistent with concerns raised throughout the responsible AI governance literature [26], [27]. Excessive dependence on automated recommendations, discussed in Section 8 in relation to automation bias [21], is a further barrier that governance mechanisms within ARM-AI are specifically designed to address, although their effectiveness in practice has not been tested.

11. Limitations and Future Research

This paper presents ARM-AI as a conceptual framework only. It has not yet been empirically validated in any organisational setting. No organisational dataset was analysed, no real software development project was evaluated, and no survey or interview study was conducted in support of this paper. Consequently, no quantitative performance improvement, whether expressed in terms of time, cost, quality, predictability, or risk reduction, is claimed at any point in this paper. The proposed framework may require adaptation to different organisational sizes, product types, and regulatory contexts, and its effectiveness in practice would depend on factors such as data quality, governance maturity, organisational readiness, and user acceptance, none of which are assessed here.

Future research should prioritise empirical validation of ARM-AI through methods appropriate to its conceptual and exploratory stage. These include structured interviews with Product Owners, Product Leads, Delivery Managers, and technical specialists; expert panel reviews; Delphi studies aimed at refining the framework's components and governance mechanisms; design science research cycles that iteratively build and evaluate framework artefacts; exploratory case studies within organisations willing to pilot elements of the framework; controlled pilot implementations limited to a defined product or team context; longitudinal organisational studies examining adoption over time; comparative evaluations against existing risk management practices; and usability and trust assessments focused on how practitioners actually interact with AI-generated recommendations under the framework's governance layer.

12. Conclusions

This paper has addressed the problem of fragmented, reactive, and experience-dependent risk management practice in Agile software product development, and has proposed ARM-AI as a human-centred conceptual framework for structuring how artificial intelligence can support risk management activities without displacing human judgement. The framework comprises four interconnected layers, organisational and project context, AI-assisted risk analysis, human decision and governance, and continuous learning and feedback, together with defined risk categories, human and AI roles, an integration approach for the Agile product lifecycle, governance and explainability mechanisms, and a proposed organisational adoption roadmap. Throughout the framework, human oversight, traceability, and accountability are treated as non-negotiable design requirements rather than optional safeguards, reflecting the position that artificial intelligence should function as a decision-support mechanism rather than an autonomous decision-maker in this domain. The principal conceptual contribution of the paper is a structured, adaptable model that connects AI capabilities with established risk management and Agile governance practice. As stated throughout this paper,

ARM-AI has not been empirically validated, and its practical value can only be established through the future research directions outlined in Section 11.

References

[1] International Organization for Standardization. (2018). *ISO 31000:2018 – Risk management — Guidelines.* ISO. https://www.iso.org/standard/65694.html

[2] Muntés-Mulero, V., Paladini, P., Solé, M., Manzoor, J., Gunka, A., Huérmann, C., & Gijsen, B. (2020). Agile risk management for multi-cloud software development. *IET Software, 14*(6), 691–699. https://doi.org/10.1049/iet-sen.2018.5295

[3] Noll, J., Lal, R. B., Beecham, S., & Clear, T. (2020). Do scaling agile frameworks address global software development risks? An empirical study. *Journal of Systems and Software, 171*, 110823. https://doi.org/10.1016/j.jss.2020.110823

[4] Dingsøyr, T., & Petit, Y. (2021). Managing layers of risk: A new challenge for large-scale Agile projects. In *Agile Processes in Software Engineering and Extreme Programming – Workshops (Lecture Notes in Business Information Processing)* (pp. 54–61). Springer. https://doi.org/10.1515/9783110652321-005

[5] Daneshpajouh, A., Taboada, F., Toledo, N., & De Vass, T. (2023). Artificial intelligence enabled project management: A systematic literature review. *Applied Sciences, 13*(8), 5014. https://doi.org/10.3390/app13085014

[6] Ingason, H. Þ., Jónasson, H. I., Jónsdóttir, S. A., & Fridgeirsson, T. V. (2021). Near future effect of artificial intelligence on project management knowledge areas. *Sustainability, 13*(4), 2345. https://doi.org/10.3390/su13042345

[7] Mohammad, A., & Chirchir, R. (2024). Challenges of integrating artificial intelligence in software project planning: A systematic literature review. *Digital, 4*(3), 28. https://doi.org/10.3390/digital4030028

[8] Stødle, K., Flage, R., Guikema, S. D., & Aven, T. (2024). Artificial intelligence for risk analysis: Considerations for improved implementation. *Risk Analysis*. Advance online publication. https://doi.org/10.1111/risa.14307

[9] Vyhmeister, E., & Castañé, G. G. (2025). TAI-PRM: Trustworthy AI—project risk management framework towards Industry 5.0. *AI and Ethics, 5*(2), 819–839. https://doi.org/10.1007/s43681-023-00417-y

[10] Bass, J. M., & Haxby, A. (2019). Tailoring product ownership in large-scale Agile projects: Managing scale, distance, and governance. *IEEE Software, 36*(2), 58–63. https://doi.org/10.1109/MS.2018.2885524

[11] Buchalcevová, A., & Remta, D. (2021). Product owner's journey to SAFe®: Adaptation options for the product owner role in scaled environments. *Information, 12*(3), 107. https://doi.org/10.3390/info12030107

[12] Noll, J., Razzak, M. A., Bass, J. M., & Beecham, S. (2017). A study of the Scrum Master's role. In *Product-Focused Software Process Improvement (Lecture Notes in Computer Science, vol. 10611)* (pp. 307–323). Springer. https://doi.org/10.1007/978-3-319-69926-4_22

[13] Jordan, S., Radtke, F., & Schön, E.-M. (2020). Improving risk management in a scaled Agile environment. In *Agile Processes in Software Engineering and Extreme Programming (Lecture Notes in Business Information Processing, vol. 383)* (pp. 122–130). Springer. https://doi.org/10.1007/978-3-030-49392-9_9

[14] Mersha, M., Lam, K., Wood, J., AlShami, A. K., & Kalita, J. (2024). Explainable artificial intelligence: A survey of needs, techniques, applications, and future direction. *Neurocomputing, 599*, 128111. https://doi.org/10.1016/j.neucom.2024.128111

[15] Islam, S. R., Eberle, W., Ghafoor, S. K., & Ahmed, M. (2021). Explainable artificial intelligence approaches: A survey. *arXiv preprint*. https://doi.org/10.48550/arXiv.2101.09429

[16] Haque, A. B., Islam, A. K. M. N., & Mikalef, P. (2022). Explainable Artificial Intelligence (XAI) from a user perspective: A synthesis of prior literature and problematizing avenues for future research. *Technological Forecasting and Social Change, 186*, 122120. https://doi.org/10.1016/j.techfore.2022.122120

[17] Lai, V., Chen, C., Liao, Q. V., Smith-Renner, A., & Tan, C. (2021). Towards a science of human-AI decision making: A survey of empirical studies. *arXiv preprint*. https://doi.org/10.48550/arXiv.2112.11471

[18] Sele, K., & Chugunova, M. (2023). Putting a human in the loop: Increasing uptake, but decreasing accuracy of automated decision-making. *PLOS ONE, 19*(7), e0298037. https://doi.org/10.1371/journal.pone.0298037

[19] Wu, X., Xiao, L., Sun, Y., Zhang, J., Ma, T., & He, L. (2022). A survey of human-in-the-loop for machine learning. *Future Generation Computer Systems, 135*, 364–381. https://doi.org/10.1016/j.future.2022.05.014

[20] Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction, 5*(CSCW1), Article 188. https://doi.org/10.1145/3449287

[21] Alon-Barkat, S., & Busuioc, M. (2022). Human–AI interactions in public sector decision-making: "Automation bias" and "selective adherence" to algorithmic advice. *Journal of Public Administration Research and Theory, 33*(1), 153–169. https://doi.org/10.1093/jopart/muac007

[22] Matute, H., Liberal, V., Arrese, C., & Agudo, N. (2024). Impact of AI errors in a human-in-the-loop process. *Cognitive Research: Principles and Implications, 9*, 1. https://doi.org/10.1186/s41235-023-00529-3

[23] Tabassi, E. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (NIST AI 100-1). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-1

[24] International Organization for Standardization. (2023). *ISO/IEC 42001:2023 — Information technology — Artificial intelligence — Management system.* ISO/IEC. https://www.iso.org/standard/42001

[25] Dignum, V. (2022). Responsible artificial intelligence: From principles to practice. *arXiv preprint*. https://doi.org/10.48550/arXiv.2205.10785

[26] Batool, A., Zowghi, D., & Bano, M. (2023). Responsible AI governance: A systematic literature review. *arXiv preprint*. https://doi.org/10.48550/arXiv.2401.10896

[27] Sigfrids, A., Koskimies, M., Salo-Pöntinen, H., & Leikas, J. (2023). Human-centricity in AI governance: A systemic approach. *Frontiers in Artificial Intelligence, 6*, 976887. https://doi.org/10.3389/frai.2023.976887

[28] Herrera-Poyatos, A., Del Ser, J., López de Prado, M., Wang, F.-Y., Herrera-Viedma, E., & Herrera, F. (2025). Responsible artificial intelligence systems: A roadmap to society's trust through trustworthy AI, auditability, accountability, and governance. *arXiv preprint*. https://doi.org/10.48550/arXiv.2503.04739

[29] Enqvist, L., Naarttijärvi, M., & Enarsson, T. (2021). Legal perspectives on hybrid human/AI decision-making systems in public administration. *Information & Communications Technology Law, 30*(3), 235–252. https://doi.org/10.1080/13600834.2021.1958860