Jurnal Impresi Indonesia (JII)
p-ISSN: 2828-1284 e-ISSN: 2810-062x
Vol.5. No. 8. Agustus 2026

Knowledge Documentation Framework for AI Initiatives: Development and Validation Across Organizational AI Contexts

Fitria Handayani*, Finannisa Zhafira, Dana Indra Sensuse, Sofian Lusa
Universitas Indonesia, Indonesia
Email: fitria.handayani41@ui.ac.id*, finannisa.zhafira@ui.ac.id, dana@cs.ui.ac.id, sofian.lusa12@ui.ac.id

Keywords:
knowledge documentation; AI initiatives; knowledge management; design science research; content validity index.

Abstract

Despite the growing organizational reliance on artificial intelligence (AI) systems, knowledge documentation (KD) practices in AI initiatives remain largely ad hoc, unstandardized, and disconnected from project lifecycle management. This study addressed this gap by proposing and validating the Knowledge Documentation Framework for AI Initiatives (KDF-AI), a twelve-component, five-phase, maturity-tiered framework synthesized through a literature review of peer-reviewed studies on KD practices in organizations implementing AI. Using Design Science Research (DSR) as the methodological paradigm, the study completed two iterative cycles: a literature-based synthesis that produced KDF-AI v1 and an expert evaluation cycle that resulted in the refined KDF-AI v2. Expert content validation was conducted with three domain validators representing academic, governance, and AI practitioner perspectives using a mixed-method approach that combined quantitative Content Validity Index (CVI) assessment with deductive thematic analysis of semi-structured interviews. The results showed that 57 of 66 items (86.4%) achieved universal inter-rater agreement, producing S-CVI/UA = 0.864 and S-CVI/Ave = 0.955, both exceeding the recommended threshold of 0.80. The nine items that did not meet the threshold consistently reflected issues of clarity rather than relevance, indicating strong conceptual acceptance of the framework while highlighting the need for more operationally specific articulation in several Advanced-tier components. Nine targeted revisions resulted in the development of KDF-AI v2. The study contributed: (1) a validated lifecycle-integrated KD framework for AI initiatives; (2) a taxonomy of ten systematically identified gaps in current AI KD practices; and (3) a methodological demonstration of mixed-method CVI validation for framework development in information systems research.

INTRODUCTION

The adoption of artificial intelligence (AI) across organizations has grown substantially over the past decade, transforming how enterprises operate, make decisions, and manage organizational knowledge (Jarrahi et al., 2023; Kovačić et al., 2022; Olan et al., 2022; Oyekunle & Boohene, 2024; Song et al., 2025). Taherdoost and Madanchian (2023) describe AI as a major driver of contemporary knowledge management (KM), enabling organizations to automate routine knowledge capture, identify patterns in large datasets, and improve information-processing efficiency. Gelashvili-Luik, Vihma, and Pappel (2025), through a review of prior studies, observed that AI has shifted KM practices from static document repositories toward dynamic and real-time knowledge ecosystems. As AI systems become increasingly embedded in organizational operations, the knowledge generated throughout AI initiatives has also become more complex, distributed, iterative, and difficult to manage systematically.

Unlike traditional software systems, AI initiatives generate knowledge continuously throughout experimentation, dataset preparation, model iteration, deployment, monitoring, and organizational adaptation processes. This knowledge includes not only technical artifacts such as datasets, experiment logs, and model configurations but also tacit practitioner knowledge, decision rationales, governance considerations, operational assumptions, and lessons learned accumulated throughout AI development and deployment activities. As AI initiatives become increasingly iterative and exploratory, organizations face growing difficulties in preserving and reusing this knowledge consistently across teams and projects.

Prior studies have identified several recurring challenges related to knowledge documentation (KD) in AI initiatives. Gerlach and Lange (2026) identified model drift as a critical issue in machine learning systems, in which predictive models gradually lose alignment with the operational conditions they were originally trained to represent. Gelashvili-Luik et al. (2025) further emphasized that tacit knowledge held by AI practitioners, such as contextual judgment, practical experience, and design rationale, remains difficult to capture in forms that can be systematically reused across projects and teams. At the practitioner level, documentation practices in AI initiatives remain inconsistent and weakly integrated into project workflows. Chang and Custis (2022), based on interviews with machine learning practitioners, found that documentation was frequently treated as a secondary activity, dependent on individual preference rather than embedded organizational practice. Heger, Marquis, and Vorvoreanu (2022) similarly characterized existing data documentation approaches as largely ad hoc, where practitioners often failed to anticipate future reuse needs. Cha et al. (2023) explained that AI development workflows rely heavily on iterative and exploratory practices grounded in tacit knowledge, making standardization difficult without structured organizational mechanisms.

To contextualize these challenges within organizational AI practice, preliminary practitioner inquiries were conducted with professionals involved in AI development initiatives across different industries, including healthcare-oriented software development and telecommunications. The inquiries revealed recurring difficulties in preserving and organizing knowledge generated during AI experimentation and deployment activities.

A product manager involved in developing AI solutions for healthcare organizations described how conventional software development approaches were insufficient for AI projects characterized by frequent experimentation and iterative refinement. Important development knowledge, including design rationale and iteration history, was often undocumented or inconsistently recorded, causing project memory fragmentation and reducing alignment between expected and actual AI system outcomes. The practitioner further explained that more structured and iterative documentation practices subsequently improved knowledge continuity and supported subsequent AI development cycles.

Similarly, a data scientist in a telecommunications company reported that incomplete documentation of structured enterprise data—including metadata, column definitions, and dataset relationships—significantly increased experimentation costs during internal AI solution development. Because contextual knowledge regarding the data environment was poorly documented, teams repeatedly reinterpreted datasets throughout experimentation cycles, contributing to substantial inefficiencies and development costs that exceeded initial estimates. After improving data documentation and knowledge organization practices, the organization was able to reduce experimentation inefficiencies and align development costs more closely with planned budgets.

A synthesis of prior literature and practitioner observations identified recurring gaps in knowledge documentation practices within organizational AI initiatives. These gaps reflected persistent organizational, technical, and governance-related challenges that limited organizations' ability to systematically preserve, reuse, and govern knowledge generated throughout AI project lifecycles.

Existing documentation standards are also insufficient for addressing the unique characteristics of AI systems. Konigstorfer and Thalmann (2021) argued that traditional software engineering documentation approaches are poorly aligned with AI systems that continuously evolve through changing data and model behavior. AI documentation must therefore capture not only technical specifications but also organizational context, decision rationale, governance structures, and operational dependencies. Konigstorfer (2024) further identified the absence of AI-specific governance mechanisms as a key contributor to documentation failure, where unclear accountability structures cause documentation activities to become fragmented and inconsistent.

The consequences of inadequate knowledge documentation are significant across several organizational dimensions. First, accountability: Raji et al. (2020) argued that without systematic documentation throughout AI development, organizations struggle to trace model outputs back to their underlying assumptions, training conditions, and design decisions. Second, reproducibility: Gundersen and Kjensmo (2018) found that only a limited proportion of variables required to reproduce AI research outcomes were adequately documented in leading AI publications. Kapoor and Narayanan (2023) further demonstrated that reproducibility challenges were widespread across machine-learning-based research. Third, organizational learning: inadequate documentation reduces organizations' ability to preserve and reuse knowledge across AI initiatives, causing teams to repeatedly rediscover solutions that had previously been achieved (Chang & Custis, 2022).

Although these issues have become increasingly recognized, existing research still lacks a comprehensive and lifecycle-integrated framework that systematically supports knowledge documentation throughout AI initiatives. Existing KM frameworks were not originally designed to address the sociotechnical characteristics of AI development, including tacit knowledge externalization, model drift monitoring, governance accountability, and human-AI collaboration. Existing AI documentation standards, such as model cards and datasheets, primarily focus on individual artifacts rather than the broader organizational knowledge lifecycle surrounding AI initiatives.

This study addressed this gap by proposing the Knowledge Documentation Framework for AI Initiatives (KDF-AI), a twelve-component, five-phase, maturity-tiered framework synthesized through a literature review of peer-reviewed studies on knowledge documentation practices in AI initiatives. KDF-AI was designed as a domain-agnostic framework intended to support organizations that develop, deploy, or manage AI initiatives. Rather than using AI to support knowledge management processes, the framework focuses on documenting and governing knowledge generated throughout AI project lifecycles, including experimentation, model development, operational monitoring, and organizational learning activities.

To assess the conceptual validity and practical applicability of the framework prior to organizational deployment, this study conducted expert content validation using a mixed-method approach that combined quantitative Content Validity Index (CVI) assessment with qualitative deductive thematic analysis of semi-structured interviews. The study adopted Design Science Research (DSR) as its methodological paradigm (Hevner et al., 2004; Peffers et al., 2007), treating KDF-AI as a design artifact that underwent two iterative cycles: a literature-based synthesis cycle producing KDF-AI v1 and an expert evaluation cycle producing the refined KDF-AI v2.

METHOD

Design Science Research

This study employed Design Science Research (DSR) as its primary methodological paradigm. DSR was selected because the study aimed not only to analyze knowledge documentation (KD) challenges conceptually but also to design, refine, and validate a lifecycle-integrated framework capable of supporting organizations in preserving, governing, and reusing knowledge generated throughout AI development and deployment activities (Hevner et al., 2004; Peffers et al., 2007). Within this study, the Knowledge Documentation Framework for AI Initiatives (KDF-AI) was positioned as a prescriptive design artifact developed to address recurring KD challenges in organizational AI initiatives.

The study followed the interaction among the relevance cycle, rigor cycle, and design cycle proposed by Hevner et al. (2004) and operationalized through the DSR process model by Peffers et al. (2007). To enhance practical relevance, preliminary contextual inquiries were conducted with practitioners involved in organizational AI initiatives across multiple industry contexts, including healthcare-oriented software development and telecommunications. These inquiries identified recurring operational challenges, such as fragmented AI project knowledge, ad hoc documentation practices, weak governance accountability, and knowledge loss during iterative AI experimentation. The rigor cycle was supported through a synthesis of prior literature related to AI knowledge management, AI documentation practices, governance mechanisms, and DSR methodology.

Figure 1 Design Science Research process for KDF-AI development and validation adapted from Hevner et al. (2004) and Peffers et al. (2007), illustrating the interaction between relevance, rigor, and iterative design cycles.

As illustrated in Figure 1, the study was conducted through two iterative DSR cycles. The first cycle focused on literature-based synthesis. Step 1 involved identifying and synthesizing recurring KD gaps in AI initiatives based on prior studies, resulting in a taxonomy of ten recurring gaps related to documentation standardization, tacit knowledge capture, governance accountability, lifecycle integration, and knowledge drift monitoring. Step 2 defined the framework objectives and structural components required to address these identified gaps. Step 3 involved designing and developing KDF-AI v1 through the synthesis of concepts from AI governance, knowledge management, and AI documentation literature. The resulting artifact consisted of twelve framework components, five lifecycle phases, and three maturity tiers, which were subsequently operationalized into a 66-item validation instrument.

The second cycle focused on expert evaluation and refinement of the framework. Step 4 involved evaluating KDF-AI v1 through a structured validation instrument assessed by three domain experts representing academic, governance auditing, and AI practitioner perspectives. Step 5 evaluated the framework using quantitative Content Validity Index (CVI) analysis combined with qualitative deductive thematic analysis of expert feedback. Step 6 translated the evaluation results into targeted refinements, resulting in KDF-AI v2 with improved operational clarity, governance specification, and implementation guidance.

The iterative nature of the DSR process enabled continuous refinement of the framework while maintaining alignment between theoretical rigor and organizational relevance. The resulting KDF-AI v2 represented a validated prescriptive artifact designed to support organizations in systematically documenting and governing knowledge generated throughout AI initiative lifecycles.

Expert Content Validation and CVI

Content validity refers to the extent to which an instrument adequately represents and measures the construct of interest. In framework development research, content validity is commonly established through structured expert evaluation processes involving domain specialists. This study adopts the three-stage validation approach proposed by Almanasreh, Moles, and Chen (2022), consisting of: (1) instrument development, (2) expert evaluation and quantification, and (3) revision based on validation findings. The application of expert-based validation approaches in design and framework research is further supported by Beecham et al. who employed structured expert assessment in validating a software process improvement model.

The validation instrument consists of 66 items organized into four dimensions: (A) framework-level properties (8 items), including overall relevance, logical organization, and conceptual completeness; (B) phase-level coherence (7 items), evaluating the appropriateness of lifecycle sequencing and integration; (C) component-level assessment (36 items; 3 items per component), assessing relevance to AI knowledge documentation practice, clarity of definition, and practical applicability; and (D) implementation feasibility (15 items), covering organizational adoptability, workload appropriateness, and governance adequacy.

Each item was evaluated using a four-point relevance scale: (1) not relevant, (2) somewhat relevant, (3) relevant with minor revision, and (4) highly relevant. Consistent with established CVI methodology, ratings of 3 and 4 were categorized as relevant for index calculation purposes. Content validity was assessed at two levels:

1. **I-CVI (Item-level Content Validity Index):** the proportion of experts rating an individual item as relevant. With three validators, the recommended threshold of 0.78 effectively requires unanimous agreement, since two-of-three agreement produces an I-CVI value below the acceptable threshold.
2. **S-CVI/UA (Scale-level Content Validity Index/Universal Agreement):** the proportion of items achieving I-CVI = 1.00, representing the stricter scale-level validity indicator.
3. **S-CVI/Ave (Scale-level Content Validity Index/Average):** the mean I-CVI score across all instrument items, representing the more conservative scale-level validity measure. Both S-CVI indicators were evaluated against the recommended threshold of 0.80.

Although the number of validators was limited, prior CVI literature suggests that small expert panels remain acceptable for preliminary framework validation when universal agreement criteria are applied.

Expert Profiles

Validator selection followed established recommendations for expert-based content validation studies, emphasizing demonstrated expertise relevant to the framework domain and the ability to provide both theoretical and practical evaluation perspectives. Three complementary domains of expertise were identified as essential for evaluating KDF-AI: (1) knowledge management and AI systems research, (2) AI governance and accountability practices, and (3) operational experience in AI implementation environments. Table III summarizes the profiles of the selected validators. The combination of academic, governance, and practitioner perspectives was intended to support balanced evaluation of the framework's conceptual rigor, governance applicability, and operational relevance within organizational AI initiatives.

Table 1. Expert Validator Profiles

Code	Category	Role and Institution	Domain Expertise	Experience
V1	Academic	Lecturer, Public Polytechnic;	AI systems, IoT, biomedical engineering; interdisciplinary AI research projects with national research funding	6 years in AI research
V2	Professional (Governance)	Auditor, Public Accounting Firm	Auditing AI systems in regulated industries; assessing AI governance controls and compliance documentation	2 years in AI audit

V3	Professional (Practitioner)	Data Scientist, Telecommunication Industry	AI solutions for B2B clients (churn prediction, anomaly detection, customer segmentation); daily-use AI documentation practices	3 years in industry AI teams

Note— Validators are identified by code. V3 (MA) is identified only by initials per participant confidentiality agreement

Deductive Thematic Analysis

To complement the quantitative CVI assessment with explanatory qualitative insight, semi-structured interviews were conducted with each validator following questionnaire completion. The interviews explored three areas not fully captured through numerical CVI scoring alone: (1) the rationale underlying low-rated items or dimensions, (2) suggestions for framework refinement and clarification, and (3) perceived implementation challenges based on the validators' professional experience. Interview recordings were transcribed and analyzed using deductive thematic analysis following Braun and Clarke (2022). In this approach, the twelve KDF-AI components and four validation dimensions served as predefined analytical categories rather than allowing themes to emerge entirely inductively from the data. The analysis followed four stages: familiarization with the interview transcripts, indexing validator statements against framework components and rating dimensions, mapping patterns of convergence and divergence across validators, and interpretation of recurring themes relevant to framework refinement. Qualitative findings were triangulated with the quantitative CVI results to identify components requiring clarification, characterize recurring implementation concerns, and support evidence-based refinement of KDF-AI v1 into the finalized KDF-AI v2 framework artifact.

Data Collection Procedure

Data collection was conducted in three sequential stages. In Stage 1, the 66-item validation instrument was distributed electronically to all validators together with operational definitions and rating guidelines to support consistent interpretation of the four-point relevance scale. Validators completed the instrument independently to preserve the methodological independence required for CVI assessment. In Stage 2, semi-structured interviews were conducted individually with each validator following questionnaire completion. The interviews focused primarily on framework items receiving lower ratings, areas perceived as conceptually unclear, implementation feasibility concerns, and recommendations for refinement. All interviews were conducted online with participant consent and recorded for transcription and analysis purposes. In Stage 3, quantitative and qualitative analyses were conducted concurrently. CVI calculations were used to determine item-level and scale-level content validity, while interview transcripts were analyzed using deductive thematic analysis. Findings from both data streams were subsequently triangulated to identify recurring patterns, characterize validation disagreements, and formulate evidence-based refinements incorporated into KDF-AI v2.

Research Tools and Software

This study employed the following tools across research phases. (1) Literature search and retrieval: Scopus and Google Scholar databases were used to identify peer-reviewed studies published between 2018 and 2025, using search terms related to AI knowledge documentation, governance frameworks, and organizational knowledge management. (2) Validation instrument administration: The 66-item questionnaire was administered electronically through Google Forms, enabling structured data collection from geographically distributed validators. (3) Data management and CVI calculation: Raw validation scores were recorded and organized in Microsoft Excel, which served as the primary platform for computing I-CVI, S-CVI/UA, S-CVI/Ave, and for organizing validator response matrices. (4) Inter-rater reliability analysis: Fleiss' kappa, pairwise Cohen's kappa, and Gwet's AC1 were computed analytically from the raw rating matrix; calculations were performed analytically in Excel (5) Expert interview management: Semi-structured interviews were conducted via Google Meet; recordings were transcribed verbatim for qualitative analysis. (6) Qualitative analysis: Deductive thematic analysis was conducted manually following Braun and Clarke, using the twelve KDF-AI components and four validation dimensions as predefined analytical categories. (7) Framework visualization: The KDF-AI structural diagram and DSR process diagram were developed using digital diagramming tools and refined iteratively during the DSR design cycle.

RESULTS AND DISCUSSION

KDF-AI: The Proposed Framework

In response to the ten identified gaps, this study proposes the Knowledge Documentation Framework for AI Initiatives (KDF-AI), a lifecycle-integrated framework designed to support organizations in systematically documenting, governing, and reusing knowledge generated throughout AI initiatives. The framework consists of twelve interconnected components organized across five knowledge lifecycle phases: Capture, Organize, Share, Apply, and Evaluate. To support gradual organizational adoption, KDF-AI is structured into three maturity tiers: Basic, Intermediate, and Advanced. This tiered structure reflects the understanding that organizations typically require foundational governance and documentation practices before implementing more advanced and automated KD capabilities (Olan et al., 2022). KDF-AI is intended to function as an organizational documentation and governance framework rather than as a standalone technical documentation standard. The framework emphasizes that effective KD in AI initiatives requires not only documentation artifacts, but also supporting governance structures, collaboration mechanisms, knowledge reuse processes, and continuous monitoring capabilities integrated throughout the AI lifecycle.

Six design principles guide the construction of KDF-AI. First, lifecycle integration, where KD activities are embedded throughout AI project phases rather than performed retrospectively (Harfouche et al., 2022). Second, sociotechnical balance, emphasizing explicit coordination between human and AI roles in documentation processes (Gelashvili-Luik et al., 2025). Third, tacit-explicit integration, supporting structured mechanisms for capturing both explicit project artifacts and tacit practitioner knowledge (Chang & Custis, 2022; Harfouche et al., 2022). Fourth, configurational implementation, where framework components are designed to function as an interconnected system rather than as isolated practices (Olan et al., 2022). Fifth, continuous documentation maintenance, recognizing that AI documentation must evolve alongside changing models, datasets, and operational conditions (Gerlach & Lange, 2026). Finally, governance-first implementation emphasizes that accountability structures, responsibilities, and documentation policies should be established before advanced technical integration mechanisms are introduced.

Figure 2 KDF-AI: Lifecycle-Integrated Knowledge Documentation Framework

Table 2. Framework Components and Functions

Code	Component Name	Description	Phase	Tier	Gaps Addressed
C1	Knowledge Capture Layer	Systematic intake of explicit knowledge from AI project activities: experiment logs, decision records, and model artifacts	Capture	Intermediate	G3, G7
C2	AI/ML Documentation Protocol	Standardized templates and schemas for consistently documenting AI models, datasets, and experiments across projects and teams	Capture	Basic	G1, G7, G8
C3	Tacit Knowledge Elicitation	Structured methods for capturing tacit practitioner knowledge through interviews, retrospectives, and experience codification.	Capture	Advanced	G5
C4	Knowledge Taxonomy and Ontology Layer	Systematic classification system for AI knowledge that facilitates categorization, search, and retrieval across the organization	Organize	Intermediate	G4, G9
C5	Knowledge Graph Architecture	Semantic network representing relationships among AI knowledge entities, enabling cross-domain traversal and inference	Organize	Advanced	G10
C6	Collaborative Knowledge Sharing Platform	Infrastructure and mechanisms for sharing knowledge among AI project stakeholders, including across teams and projects at the enterprise level	Share	Intermediate	G9
C7	Human-AI Collaboration Protocol	Defined roles, interfaces, and workflows for human-AI co-documentation of knowledge artifacts	Share	Advanced	G6
C8	Decision Support and Knowledge Reuse Engine	Mechanisms for retrieving and reusing documented knowledge to	Apply	Intermediate	G3, G9

		support decision-making in subsequent AI initiatives			
C9	Real-time Knowledge Integration Layer	Direct integration of KD with operating AI systems, enabling automatic updates as models and contexts change	Apply	Advanced	G3
C10	KD Governance and Accountability Framework	Policies, role assignments, responsibilities, and accountability mechanisms governing all KD activities in the organization's AI portfolio	Evaluate	Basic	G8
C11	Knowledge Drift Monitor	Detection and management system for documentation validity degradation caused by model changes, data shifts, or operational context evolution	Evaluate	Advanced	G2
C12	Continuous Learning Loop	Organizational feedback mechanism for refining AI initiatives and KD practices using documented knowledge.	Evaluate	Advanced	G4

CVI Results: Overall Summary

Content validation of KDF-AI v1 was conducted using the 66-item instrument across four assessment dimensions. Of 66 evaluated items, 57 achieved full agreement among validators (I-CVI = 1.00), while 9 items did not meet the recommended threshold (I-CVI = 0.67). At the scale level, the framework achieved S-CVI/UA = 0.864 and S-CVI/Ave = 0.955, both exceeding the recommended acceptability threshold of 0.80 [33]. These findings indicate that KDF-AI demonstrates strong overall content validity across its framework structure, component definitions, and implementation dimensions. Table IV summarizes the CVI results by dimension, and Figure 3 provides a graphical representation.

Table 3. Results by Assessment Dimension

Dimension	Total Items	Valid	Failed	S-CVI/UA	S-CVI/Ave	Status
A - Framework-level Properties	8	8	0	1.000	1.000	Passed
B - Phase-level Coherence	7	4	3	0.571	0.857	Partial
C - Component-level Assessment	36	31	5	0.861	0.954	Passed
D - Implementation Feasibility	15	14	1	0.933	0.978	Passed
Total	66	57	9	0.864	0.955	Passed (Overall)

Recommended threshold: S-CVI/UA and S-CVI/Ave >= 0.80 [33]. Dimension B S-CVI/UA falls below threshold due to three localized failures; its S-CVI/Ave remains acceptable.

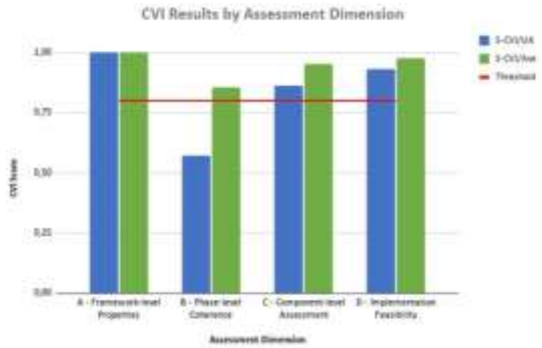

Figure 3 Content Validity Index (CVI) Results Across Assessment Dimensions of KDF-AI Framework Validation

CVI Results by Assessment Dimension

Substantive interpretation of these values provides additional insight beyond threshold reporting. The S-CVI/UA value of 0.864 indicates that 86.4% of instrument items (57 of 66) achieved unanimous agreement across all three validators. According to Polit and Beck, S-CVI/UA > 0.80 represents the stricter criterion for scale-level content validity, requiring universal agreement on the majority of items. A value of 0.864 demonstrates that the framework's structure, components, and implementation mechanisms were broadly endorsed by validators representing three distinct epistemic positions - academic research, governance auditing, and operational AI practice.

Similarly, the S-CVI/Ave value of 0.955 substantially exceeds the recommended threshold, indicating that on average approximately 95.5% of validator ratings considered the framework items relevant to AI knowledge documentation practice. The difference between S-CVI/UA (0.864) and S-CVI/Ave (0.955) reflects the presence of several items with partial agreement rather than widespread disagreement. Importantly, no item received an I-CVI score of 0.00 or 0.33, meaning that no framework component was considered broadly irrelevant by multiple validators simultaneously.

Dimension-level interpretation further strengthens these findings. Dimension A achieved perfect agreement (S-CVI/UA = 1.000), indicating unanimous acceptance of the framework's overall structure, including the five lifecycle phases, twelve components, and three-tier maturity architecture. Dimension C (S-CVI/Ave = 0.954) demonstrates strong component-level validity, suggesting that the proposed components were consistently perceived as relevant to organizational AI documentation practice. Dimension D achieved near-perfect agreement (S-CVI/Ave = 0.978), indicating that validators generally considered the framework feasible for organizational implementation. Although Dimension B produced a lower S-CVI/UA score, its S-CVI/Ave remained above the recommended threshold, indicating that disagreements were localized and associated primarily with clarity rather than conceptual relevance.

Inter-Rater Reliability Analysis

To complement the CVI validity assessment, inter-rater reliability was evaluated using three agreement statistics: Fleiss' kappa, pairwise Cohen's kappa, and Gwet's AC1. Results are summarized in Table 4.

Table 4. Inter Rater Reability Statistics

Statistic	Value	Interpretation
Fleiss' kappa (overall, 3 raters)	-0.048	See prevalence paradox note
Cohen's kappa V1-V2	1.000	Perfect (indeterminate form)
Cohen's kappa V1-V3	0.000	See prevalence paradox note
Cohen's kappa V2-V3	0.000	See prevalence paradox note
Gwet's AC1 (overall)	0.900	Almost perfect
Gwet's AC1 V1-V2	1.000	Perfect agreement
Gwet's AC1 V1-V3	0.844	Almost perfect
Gwet's AC1 V2-V3	0.844	Almost perfect

Threshold for 'almost perfect agreement': >= 0.80. Fleiss' kappa and pairwise Cohen's kappa are inappropriate primary measures in this dataset due to the prevalence paradox; see text below.

The near-zero and negative kappa values do not indicate poor inter-rater reliability, but rather reflect a well-documented statistical artifact commonly referred to as the kappa paradox or prevalence bias. In the present dataset, nearly all ratings were concentrated within the "relevant" category, producing highly skewed marginal distributions across validators. Under these conditions, chance-corrected agreement measures such as Cohen's kappa and Fleiss' kappa become unstable and may underestimate actual agreement despite high observed consistency among raters.

Specifically, Validators V1 and V2 rated all 66 items as relevant, while Validator V3 rated 57 items as relevant and 9 items as not relevant. As a result, the expected agreement probability under chance correction became disproportionately high, artificially suppressing the resulting kappa values despite strong substantive agreement across validators.

To address this limitation, Gwet's AC1 was adopted as the primary inter-rater reliability indicator because it is more robust under high-agreement and prevalence-skewed conditions. The overall Gwet's AC1 value of 0.900 indicates almost perfect agreement among validators, while pairwise AC1 values between validators ranged from 0.844 to 1.000. These findings confirm that the validation results reflect strong and reliable expert agreement rather than statistical artifact.

Collectively, the reliability findings reinforce the consistency of the expert validation process and support the stability of the overall framework evaluation despite partial divergence in operational clarity assessments identified during qualitative analysis.

Failed Items and Failure Pattern

The nine items that failed to meet the I-CVI threshold are summarized in Table VI. The failed items obtained an I-CVI score of 0.67 because only two out of three validators rated the items as relevant (scores ≥ 3), while one validator assigned a score below the relevance threshold. According to Polit and Beck, when three validators are involved, universal agreement is effectively required to achieve acceptable item-level content validity, since two-of-three agreement produces I-CVI = 0.67, which falls below the recommended threshold of 0.78.

Importantly, these failed items do not indicate conceptual rejection of the framework components. Instead, the disagreement pattern reflects concerns regarding operational clarity and implementation specificity, particularly from the practitioner validator perspective. Substantively, this finding suggests that the affected components were generally perceived as necessary and relevant for organizational AI knowledge documentation, but insufficiently articulated in terms of operational guidance and implementation detail.

Two consistent patterns emerged from the validation results. First, all failed items were associated with ratings from Validator 3 (industry practitioner), while Validators 1 and 2 rated all items above the relevance threshold. This pattern suggests that the primary concern was not the conceptual relevance of KDF-AI itself, but the operational interpretability of several framework components within real-world implementation environments.

Second, all failed items were concentrated within the “clarity of definition” dimension rather than the “relevance” dimension. No framework component was considered unnecessary or irrelevant by the validators. Instead, disagreements consistently reflected concerns regarding how several Advanced-tier mechanisms and governance-related processes were operationally described. This indicates that the framework structure and component selection were broadly accepted, while several components required more implementation-oriented articulation.

Consequently, the failed-item analysis functioned not as evidence of conceptual invalidity, but as a refinement mechanism within the DSR evaluation cycle. The failed items directly informed the targeted revisions incorporated into KDF-AI v2, particularly regarding governance specification, workflow clarification, and operational examples for Advanced-tier components.

Table 5. Nine Items Below I-CVI Threshold: Scores Per Validator

Item	Item Description	V1 (Acad.)	V2 (Audit)	V3 (DS)	I-CVI	Failure Dimension
B2	Organize phase supports systematic knowledge structuring	3	4	2	0.67	Clarity of definition
B3	Share phase facilitates knowledge dissemination across stakeholders	3	4	2	0.67	Clarity of definition
B4	Apply phase supports effective knowledge reuse	3	4	2	0.67	Clarity of definition
C4.2	Taxonomy and ontology concepts in C4 are clearly organized	3	3	2	0.67	Clarity of definition
C7.2	Human-AI collaboration protocol (C7) is well-defined and understandable	3	4	2	0.67	Clarity of definition
C9.2	Real-time integration process (C9) is described clearly and comprehensibly	4	4	1	0.67	Clarity of definition
C10.2	Governance structure and responsibilities in C10 are clearly defined	3	4	1	0.67	Clarity of definition

C11.2	Knowledge drift monitoring mechanism (C11) is clearly described	3	4	1	0.67	Clarity of definition
D11	Governance mechanisms in the framework are sufficiently clear	3	4	2	0.67	Clarity of definition

Scores below the relevance threshold (<3) are indicated in bold. V1 = N. (Academic); V2 = I.W.A. (Auditor); V3 = M.A. (Data Scientist). All failures are attributable to V3.

Deductive Thematic Analysis Findings

Semi-structured interviews with the three validators, analyzed using deductive thematic analysis based on the twelve KDF-AI components and four validation dimensions, produced four recurring themes relevant to framework refinement. First, all validators agreed that the framework addresses practical organizational needs associated with AI knowledge documentation. However, each validator emphasized different problem contexts based on their professional perspective, including academic research gaps (V1), governance and audit traceability requirements (V2), and operational documentation challenges in day-to-day AI project environments (V3). This convergence suggests that the framework possesses cross-domain relevance despite differences in evaluator background.

Second, the three-tier maturity structure was consistently perceived as practical and organizationally realistic. Validators noted that incremental adoption aligned with organizational readiness would likely be more sustainable than requiring immediate implementation of all framework components simultaneously. In particular, the practitioner validator (V3) identified the Basic-tier components, especially AI/ML Documentation Protocol (C2) and KD Governance and Accountability Framework (C10), as immediately applicable improvements over current documentation practice. Third, the interviews clarified the underlying cause of the CVI failure pattern. The practitioner validator (V3) consistently observed that several Advanced-tier component descriptions were conceptually understandable but insufficiently operationalized for implementation contexts, describing the issue as "the gap between aspiration and operationalization" [V3]. By contrast, Validators 1 and 2 considered the same level of abstraction acceptable for a prescriptive framework artifact. This difference reflects a broader tension between conceptual generalizability and practitioner demand for implementation specificity.

Fourth, all validators independently emphasized the foundational role of governance within the framework structure. In particular, C10 (KD Governance and Accountability Framework) was consistently viewed as a prerequisite for sustainable implementation of the remaining components. Validators noted that without clearly defined governance structures, accountability mechanisms, and organizational ownership, documentation practices would likely remain inconsistent regardless of technical capability. Collectively, these thematic findings directly informed the nine targeted revisions incorporated into KDF-AI v2.

KDF-AI v2: Refined Framework

Based on integrated quantitative CVI findings and qualitative thematic analysis, nine targeted revisions were applied to KDF-AI v1 to produce KDF-AI v2. All nine revisions involve clarification of articulation and addition of operational specification. No component was removed, added, or structurally changed. This confirms that the theoretical foundation of KDF-

AI v1 was sound; what was needed was more concrete explanation of operational mechanisms. Table 6 presents the revisions.

The revision from KDF-AI v1 to v2 targeted nine specific items identified as either too abstract in description or lacking operational mechanisms, without altering the framework's underlying architecture, which remains composed of twelve components, five lifecycle phases, and three maturity tiers (Fig. 2). At the phase level, the Organize phase (B2) was refined by adding concrete activities—taxonomy assignment, ontology mapping, and version control of knowledge artifacts—resulting in a searchable knowledge repository with structured metadata as its output. The Share phase (B3) was strengthened with clearer dissemination mechanisms, including collaborative platforms, structured knowledge transfer sessions, and stakeholder communication protocols. The Apply phase (B4) was enhanced by specifying how knowledge reuse occurs, namely through querying a reuse engine and integrating recommendations into ongoing AI project workflows.

At the component level, C4.2 (Knowledge Taxonomy and Ontology) was given operational definitions distinguishing taxonomy—a three-level hierarchical classification (category, subcategory, artifact)—from ontology, defined as a relational map of knowledge entities. C7.2 (Human-AI Collaboration Protocol) was clarified through role-based division of labor, where humans handle validation, contextual interpretation, and final decisions, while AI handles extraction, suggestions, and auto-fill, illustrated through a model card authorship workflow. C9.2 (Real-time Integration Layer) was made concrete through an event-driven mechanism, whereby model retraining automatically triggers a documentation update via API connection to the model registry.

Governance was reinforced through C10.2 and D11, which now define three core roles—KD Owner, KD Reviewer, and KD Auditor—along with a four-stage governance cycle (policy setting, implementation, audit, revision), supported by a practical implementation checklist for organizations. Finally, C11.2 (Knowledge Drift Monitor) was equipped with a four-step monitoring process to detect documentation validity degradation caused by model or contextual changes. The remaining seven components (C1, C2, C3, C5, C6, C8, C12), along with the Capture and Evaluate phases, were left unchanged, confirming that the v2 revisions are purely operational refinements rather than structural changes.

The validation findings carry three significant theoretical implications. First, CVI results confirm that the DSR approach is viable for developing and validating KM frameworks in AI contexts, where prescriptive artifacts (rather than predictive theories) can have their validity assessed through structured expert content judgment prior to empirical evaluation (Hevner et al., 2004). This contributes a methodological precedent for IS researchers working in framework development.

Second, the consistent failure pattern, all nine items failing on clarity rather than on relevance, provides empirical evidence that the KD gaps identified in the literature are real and experienced directly by practitioners. No validator rejected any component as irrelevant or unnecessary. Instead, they questioned how to implement them, a question that only arises when someone already accepts that the component is necessary. This aligns with the observation by Gelashvili-Luik et al. (2025) that the field of KD in AI is transitioning from problem recognition toward the need for concrete implementation guidance.

Third, the divergence between academic and auditor validators on one hand, and the practitioner validator on the other, reveals an inherent tension in KM framework development for AI contexts: a framework abstract enough to be universal tends to be insufficiently specific for field practitioners, while an overly operational framework risks losing cross-context transferability. KDF-AI v2 attempts to bridge this tension through more operational descriptions while preserving domain-agnostic design principles.

Two key practical implications emerge for organizations considering KDF-AI adoption. First, organizations should start from the Basic tier. The two Basic components, C2 (AI/ML Documentation Protocol) and C10 (KD Governance Framework), address four of five high-severity gaps and received unanimous acceptance from all three validators, including the field practitioner. Early investment in documentation standards and governance structures provides the foundation necessary for Intermediate and Advanced components to function effectively.

Second, organizations should pair Advanced-tier components with operational implementation guides. Findings from V3 demonstrate that components such as Real-time Integration Layer (C9), Knowledge Drift Monitor (C11), and Human-AI Collaboration Protocol (C7) require implementation playbooks alongside framework descriptions. For industry AI teams, this means each Advanced component should be accompanied by concrete role descriptions, illustrative scenarios, and specified trigger conditions.

CONCLUSION

This study addressed the fragmented and ad hoc nature of knowledge documentation (KD) practices in organizational AI initiatives. In response to RQ1, ten recurring gaps were identified, spanning standardization, governance, tacit knowledge capture, lifecycle integration, and drift monitoring, which informed the design of the Knowledge Documentation Framework for AI Initiatives (KDF-AI). Using Design Science Research (DSR) across two iterative cycles of literature synthesis and expert validation, the framework was developed to integrate twelve components across five lifecycle phases and three maturity tiers, enabling progressive adoption based on organizational readiness.

For RQ2, expert validation confirmed strong conceptual validity (S-CVI/UA = 0.864; S-CVI/Ave = 0.955), with all identified issues relating to operational clarity rather than conceptual soundness. These findings informed nine targeted refinements, resulting in KDF-AI v2.

The study contributed theoretically through the development of a taxonomy of KD gaps, practically through KDF-AI v2 as a validated governance-oriented artifact, and methodologically by integrating CVI analysis with thematic analysis within a DSR process. Overall, the study positioned KD in AI initiatives not as an administrative burden but as a core organizational capability supporting accountability, reproducibility, and sustainable knowledge governance, with KDF-AI v2 representing a validated foundation for advancing this direction.

REFERENCES

Almanasreh, E., Moles, R. J., & Chen, T. F. (2022). A practical approach to the assessment and quantification of content validity. In *Contemporary research methods in pharmacy and health services* (pp. 583–599). Elsevier.

Braun, V., Clarke, V., & Hayfield, N. (2022). ‘A starting point for your journey, not a map’: Nikki Hayfield in conversation with Virginia Braun and Victoria Clarke about thematic analysis. *Qualitative Research in Psychology*, *19*(2), 424–445.

Jarrahi, M. H., Askay, D., Eshraghi, A., & Smith, P. (2023). Artificial intelligence and knowledge management: A partnership between human and AI. *Business Horizons*, *66*(1), 87–99.

Kovačić, M., Mutavdžija, M., Buntak, K., & Pus, I. (2022). Using artificial intelligence for creating and managing organizational knowledge. *Tehnički Vjesnik*, *29*(4), 1413–1418.

Olan, F., Arakpogun, E. O., Suklan, J., Nakpodia, F., Damij, N., & Jayawickrama, U. (2022). Artificial intelligence and knowledge sharing: Contributing factors to organizational performance. *Journal of Business Research*, *145*, 605–615.

Oyekunle, D., & Boohene, D. (2024). Digital transformation potential: The role of artificial intelligence in business. *International Journal of Professional Business Review: Int. J. Prof. Bus. Rev.*, *9*(3), 1.

Song, Y., Qiu, X., & Liu, J. (2025). The impact of artificial intelligence adoption on organizational decision-making: An empirical study based on the technology acceptance model in business management. *Systems*, *13*(8), 683.

Cha, I., Oh, J., Park, C. Y., Han, J., & Lee, H. (2023). Unlocking the tacit knowledge of data work in machine learning. In *Extended abstracts of the 2023 CHI conference on human factors in computing systems (CHI EA '23)*. ACM.

Chang, J., & Custis, C. (2022). Understanding implementation challenges in machine learning documentation. In *Proceedings of the 2nd ACM conference on equity and access in algorithms, mechanisms, and optimization (EAAMO '22)*. ACM.

Gelashvili-Luik, T., Vihma, P., & Pappel, I. (2025). Navigating the AI revolution: Challenges and opportunities for integrating emerging technologies into knowledge management systems. *Frontiers in Artificial Intelligence, 8*, Article 1518744.

Gerlach, J. P., & Lange, D. (2026). Fading memories: The role of machine learning in organizational knowledge depreciation. *Academy of Management Review*. Advance online publication.

Gundersen, O. E., & Kjensmo, S. (2018). State of the art: Reproducibility in artificial intelligence. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 32, No. 1).

Harfouche, A., Kalika, M., Mouakket, S., & Ralyte, J. (2022). A framework for AI-supported knowledge creation and the KAM model. *Information Systems Frontiers, 24*, 1545–1562.

Heger, A. K., Marquis, L. B., & Vorvoreanu, M. (2022). Understanding machine learning practitioners' data documentation perceptions, needs, challenges, and desiderata. *Proceedings of the ACM on Human-Computer Interaction, 6*(CSCW2).

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly, 28*(1), 75–105.

Kapoor, S., & Narayanan, A. (2023). Leakage and the reproducibility crisis in machine-learning-based science. *Patterns, 4*(9), Article 100804.

Konigstorfer, F. (2024). A comprehensive review of techniques for documenting artificial intelligence. *Digital Policy, Regulation and Governance*. Advance online publication.

Konigstorfer, F., & Thalmann, S. (2021). Software documentation is not enough! Requirements for the documentation of AI. *Digital Policy, Regulation and Governance, 23*(3), 245–260.

Olan, F., Ogiemwonyi Arakpogun, E., Suklan, J., Nakpodia, F., Damij, N., & Jayawickrama, U. (2022). Artificial intelligence and knowledge sharing: Contributing factors to organizational performance. *Journal of Business Research, 145*, 605–615.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems, 24*(3), 45–77.

Raji, I. D., Smart, A., White, R. N., & Mitchell, M. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 ACM Conference on Fairness, Accountability, and Transparency (FAT* '20)*. ACM.

Taherdoost, H., & Madanchian, M. (2023). Artificial intelligence and knowledge management: Impacts, benefits, and implementation. *Computers, 12*(4), Article 72.