UPI YPTK

KOMTEKINFO JOURNAL

LPPM Universitas Putra Indonesia YPTK Padang
Jl. Raya Lubuk Begalung, Padang, West Sumatera, Indonesia, Zip Code: 25221
Volume: 13, Issue: 2, Page: 102 - 112, June 30th, 2025, e-ISSN: 2502-8758
Available online at website: https://jkomtekinfo.org/ojs/index.php/komtekinfo

Jurnal KomtekInfo

Building Expert Digital Twins through Retrieval-Augmented AI Personas: A Framework for Preserving and Transferring Human Expertise

Andhika[1]✉, Adam Puspabhuana[2], Hedy Pamungkas[3], Yudi Triyana[4], Reza Fahmi Alviandy[5]
[1] Department of Computer Science, Universitas Ary Ginanjar, Indonesia. Zip Code. 12560
[2,3] Department of Data Science, Cakrawala University, Indonesia. Zip Code. 12510
[4] Department of Information System and Technology, Cakrawala University, Indonesia. Zip Code. 12510
[5] Department of Computer Science, Universitas Cakrawala, Indonesia. Zip Code. 12510
andhika@uag.ac.id

Abstract

The preservation and transfer of human expertise represent persistent challenges in knowledge management, particularly when tacit knowledge—embedded within individual experience, judgment, and contextual reasoning—is difficult to document, scale, or disseminate. Although Large Language Models (LLMs) have enabled sophisticated conversational AI systems, existing implementations frequently exhibit factual inconsistencies, hallucinations, and inadequate representation of domain-specific expert reasoning. These deficiencies diminish the reliability of AI-mediated knowledge transfer in high-stakes educational, organizational, and professional contexts. This study addresses these limitations by proposing a framework for building Expert Digital Twins through Retrieval-Augmented AI Personas, providing a scalable and reliable mechanism for preserving and transferring human expertise. The proposed framework integrates six interconnected layers: knowledge acquisition from multimodal expert sources, preprocessing and semantic chunking, vector-based knowledge repository construction, retrieval-augmented generation, persona-driven interaction modeling, and multi-dimensional evaluation. Expert knowledge is systematically collected from books, interviews, speeches, academic articles, and digital media, then transformed into a structured semantic repository enabling dynamic, context-sensitive retrieval. Human-centered design principles are applied throughout to ensure authenticity, transparency, and user trust. An experimental evaluation was conducted by constructing an Expert Digital Twin from a domain expert's knowledge corpus and comparing its performance against a conventional LLM-based baseline using metrics including Faithfulness, Response Accuracy, Expert Similarity, Hallucination Rate, and User Trust. Results demonstrate that the retrieval-augmented AI persona substantially improves factual consistency, perceived authenticity, and knowledge transfer effectiveness. This study contributes a theoretically grounded and practically deployable framework that positions Expert Digital Twins as a novel paradigm for sustainable digital intelligence in knowledge-intensive domains.

Keywords: expert digital twin, retrieval-augmented AI persona, human expertise preservation, knowledge transfer, hallucination reduction, human-centered AI, digital knowledge management.

1. Introduction

The rapid advancement of Artificial Intelligence (AI), particularly in the domain of Large Language Models (LLMs) and generative AI, has fundamentally reshaped how humans create, access, and communicate knowledge [1]. Since the emergence of transformer-based architectures in 2017, the capabilities of language models have expanded dramatically, enabling systems to engage in naturalistic conversation, generate human-like responses, and support complex decision-making across diverse domains including education, healthcare, business consulting, and professional services [2]. The introduction of models such as GPT-4, Claude, Gemini, and open-source alternatives such as LLaMA has accelerated the deployment of conversational AI at scale, making AI-mediated knowledge access increasingly commonplace [3], with recent frontier models demonstrating emergent general-purpose reasoning capabilities that extend well beyond their original training objectives [4].

Despite these advances, a fundamental challenge remains in translating the capabilities of generative AI into effective mechanisms for preserving and transferring human expertise. Human expertise represents a critical organizational and societal asset, yet much of this expertise exists in tacit form, embedded within personal experience, intuitive judgment, contextual understanding, and the nuanced

Submitted: 20-03-2026 | Revised: 19-04-2026 | Accepted: 20-05-2026 | Published: 30-06-2026

reasoning patterns that distinguish genuine domain mastery [5]. Traditional knowledge management approaches have been largely oriented toward explicit knowledge—information that can be readily documented, stored in databases, or captured in procedural manuals. Tacit knowledge, by contrast, is inherently resistant to formalization, and its transfer traditionally depends on direct human interaction, mentorship, apprenticeship, and immersive professional experience [6].

The consequences of inadequate tacit knowledge preservation are significant. Organizations face compounding risks from workforce mobility, demographic transitions, and expert retirement, all of which contribute to what researchers have termed 'knowledge drain' or 'expert knowledge loss' [7]. In educational and professional development contexts, the inability to scale expert knowledge delivery creates systemic inequities, limiting access to high-quality guidance for learners and practitioners who lack proximity to recognized domain experts. These pressures have intensified interest in AI-based systems capable of simulating expert reasoning and communication in an accessible and scalable manner.

Recent advances in AI Persona systems and Digital Humans have introduced promising new possibilities for expert knowledge delivery. AI Persona systems leverage conversational AI, memory mechanisms, and personality modeling to create virtual entities capable of interacting with users in contextually appropriate and human-like ways [8]. Large dialogue models have demonstrated the ability to sustain extended, persona-consistent, and context-sensitive conversations [9], while generative agent architectures incorporating memory and reflection mechanisms have shown that believable, individualized behavior can be simulated computationally [10]. Concurrently, the concept of Digital Twins—originally developed for the modeling and simulation of physical systems in manufacturing and engineering contexts—has evolved toward human-centered applications. This evolution has produced the concept of the Human Digital Twin or Expert Digital Twin, which seeks to replicate not only factual knowledge but also the reasoning patterns, communication style, and contextual judgment of individual domain experts [11]. Recent reviews of AI-powered expert systems [12] and of digital twin frameworks for human knowledge modeling [13] further situate this evolution within a broader research trajectory. Research in this area has demonstrated preliminary feasibility but remains largely conceptual, with limited operational frameworks for practical implementation.

A critical impediment to the deployment of effective Expert Digital Twin systems lies in the well-documented limitations of pure LLM-based approaches. Hallucination—the generation of plausible but factually inaccurate content—represents a fundamental challenge for LLMs operating in knowledge-sensitive domains [14]. Alignment-oriented research has likewise highlighted the difficulty of ensuring that language assistants remain helpful, honest, and reliably grounded across diverse interaction contexts [15]. Factual inconsistency, susceptibility to leading queries, and difficulty in maintaining coherent expert personas across extended interactions further undermine trust and limit the applicability of existing systems for genuine expertise preservation. These problems are particularly acute when the domain knowledge is highly specialized, nuanced, or rapidly evolving, as the training data embedded in model parameters cannot reliably represent the full depth of individual expert reasoning.

Retrieval-Augmented Generation (RAG) has emerged as a technically robust and increasingly adopted approach to address the hallucination and knowledge currency limitations of standalone LLMs [16]. Extensions of this paradigm have shown that retrieving from very large external corpora [17] and incorporating retrieved context directly during inference [18] can further improve factual grounding, and surveys of knowledge-intensive language tasks have mapped the broader space of applications for which such grounding is essential [19]. By dynamically combining language generation with the retrieval of verified, domain-specific knowledge from curated external repositories, RAG architectures enable AI systems to produce responses that are grounded in authoritative information rather than relying solely on parametric memory. Prior studies have demonstrated that RAG-based systems substantially outperform conventional LLMs in factual accuracy, transparency, and domain-specific response relevance [20]. Applications have spanned enterprise knowledge assistants, biomedical question-answering, legal information systems, and educational tutoring, consistently demonstrating the value of retrieval augmentation for reliability-critical knowledge delivery.

However, the convergence of RAG techniques with the construction of authentic, persona-consistent Expert Digital Twins remains underexplored. Existing RAG implementations are primarily optimized for information retrieval tasks, with limited attention to the modeling of expert communication style, reasoning transparency, and long-term persona consistency. Furthermore, the human-centered dimensions of knowledge transfer—including perceived authenticity, user trust, and engagement—are rarely incorporated into the design and evaluation of RAG-based systems. As AI systems increasingly interact with humans in educational and professional settings, these dimensions are not peripheral concerns but essential determinants of knowledge transfer effectiveness [21].

Several adjacent research streams inform the present study. Work on knowledge graph construction and semantic retrieval has advanced methods for representing structured domain knowledge in machine-readable formats [22]. Research on conversational agent design has developed frameworks for modeling personality, communication style, and contextual appropriateness in AI interactions [23]. Studies on digital human representation have explored the visual and linguistic dimensions of virtual expert systems [24]. Investigations into human-AI trust have identified key factors—including transparency, competence, and consistency—that determine user acceptance of AI-mediated knowledge delivery [25]. Despite the richness of these individual streams, no unified operational framework has been proposed that integrates multimodal knowledge acquisition, semantic retrieval, persona-driven interaction, and human-centered evaluation for the specific purpose of building Expert Digital Twins.

This research addresses this gap directly. The central research question motivating the study is: How can a Retrieval-Augmented AI Persona framework be designed and evaluated to effectively preserve and transfer human expertise as an Expert Digital Twin? This question encompasses both the technical challenge of constructing reliable, domain-specific AI systems and the human-centered challenge of ensuring that such systems are experienced as authentic, trustworthy, and effective by end users.

The objectives of this research are threefold: first, to develop a comprehensive conceptual and technical framework for constructing Expert Digital Twins based on retrieval-augmented AI personas; second, to demonstrate how expert knowledge from multimodal sources can be systematically acquired, processed, and transformed into a functional digital intelligence system; and third, to evaluate the effectiveness of the proposed framework using both technical performance metrics and user-centered assessments. The framework is conceptualized within Design Science Research Methodology (DSRM), providing a rigorous basis for artifact construction and evaluation [26].

This study contributes to the growing interdisciplinary field of Human-Centered Artificial Intelligence by introducing a novel framework that bridges Retrieval-Augmented Generation, AI Persona modeling, and Expert Digital Twin theory. The proposed framework provides a theoretically grounded and practically deployable approach for tacit knowledge preservation and scalable expertise transfer. The study further extends the knowledge management and AI literature by positioning Expert Digital Twins as a new paradigm for sustainable digital intelligence—one that prioritizes authenticity, reliability, and human-centered interaction alongside technical performance [27].

2. Methods

This study adopts Design Science Research Methodology (DSRM) as its overarching research framework, which is appropriate for research aimed at constructing, instantiating, and evaluating novel artifacts—in this case, an Expert Digital Twin system—within Information Systems and AI research [26]. DSRM prescribes six activities: problem identification and motivation, definition of the objectives of a solution, design and development, demonstration, evaluation, and communication [28]. The present study encompasses the first five activities, with an emphasis on the design, development, and evaluation phases, which correspond to the construction and assessment of the proposed Expert Digital Twin framework.

2.1 Research Framework

The proposed framework is structured as a six-layer architecture that operationalizes the end-to-end process of building and deploying an Expert Digital Twin. Each layer serves a distinct function in transforming raw expert knowledge into a reliable, persona-consistent AI interaction system. The framework is illustrated conceptually in Figure 1 and described in detail below.

The Knowledge Acquisition Layer constitutes the entry point of the framework, encompassing systematic collection of expert knowledge from multimodal sources. These sources include published books and academic monographs authored or recommended by the target expert, recorded interviews and video presentations, audio speeches and public lectures, domain-specific academic articles and technical reports, practitioner documents including course materials and professional guidelines, and social media outputs such as blog posts and expert commentary. This multimodal approach acknowledges that tacit expert knowledge is rarely confined to a single medium; rather, it is distributed across diverse communicative contexts that, taken together, reflect the breadth and depth of individual expertise [29].

The Knowledge Processing Layer applies a structured pipeline to transform raw acquired content into machine-processable knowledge units. This pipeline includes preprocessing operations such as speech-to-text transcription, optical character recognition for document digitization, noise removal, and normalization. Processed content is subsequently subjected to semantic chunking, which partitions source material into contextually coherent segments using sentence boundary detection and thematic segmentation algorithms, an approach conceptually related to class-based topic modeling techniques used for organizing large text corpora into coherent themes [30]. Each chunk is then encoded using dense neural embeddings—specifically, transformer-based sentence embedding models—to produce high-dimensional vector representations that capture semantic meaning and contextual relationships.

The Semantic Knowledge Repository Layer houses the structured vector database that stores embedded knowledge chunks alongside their metadata, including source attribution, expert identity, domain category, date, and retrieval confidence scores. The repository employs a hybrid retrieval architecture combining dense vector similarity search with sparse keyword-based retrieval, enabling the system to handle both semantically rich queries and exact-match factual queries efficiently. The choice of vector database (e.g., FAISS, ChromaDB, or Pinecone) is informed by the scale of the knowledge corpus and the response latency requirements of the deployment context.

The Retrieval-Augmented Generation Engine constitutes the technical core of the framework. Upon receiving a user query, the engine performs semantic query encoding, retrieves the top-k most relevant knowledge chunks from the repository using cosine similarity ranking, and constructs a context-enriched prompt that combines the retrieved knowledge with persona instructions and the user's query. This augmented prompt is then passed to a large language model—specifically, a state-of-the-art instruction-tuned model such as GPT-4o or Claude 3 Opus—for response generation; where local deployment is required, efficient quantized fine-tuning techniques offer a practical means of adapting open-weight models to this role without prohibitive computational cost [31]. The retrieved chunks serve as grounding evidence, constraining the language model to generate responses consistent with verified expert knowledge and substantially reducing the probability of hallucination.

The Persona Interaction Layer governs the communicative behavior of the Expert Digital Twin, ensuring that generated responses not only convey accurate information but also reflect the communication style, rhetorical patterns, and domain vocabulary characteristic of the represented expert. Persona modeling is achieved through the construction of a persona instruction profile extracted from the expert's communicative corpus, encoding features such as preferred discourse structures, characteristic analogies, response formality, and domain-specific terminology. This profile is incorporated into the system prompt of the generative model, guiding response style independently of retrieval content, consistent with approaches that use memory-augmented generation to sustain persona consistency across dialogue turns [32].

The Evaluation Layer implements a multi-dimensional assessment protocol applied both during iterative development and in the formal evaluation phase. The evaluation encompasses both automated technical metrics and human-centered user assessments, as described in Section 2.3. The layered architecture is summarized in Table 1.

Table 1. Research Framework Components and Functions

Framework Layer	Key Components	Primary Function
Knowledge Acquisition	Books, interviews, speeches, articles, videos, documents	Multimodal expert knowledge collection
Knowledge Processing	Transcription, chunking, embedding generation	Transformation into semantic knowledge units
Semantic Knowledge Repository	Vector database, metadata store, hybrid retrieval index	Structured storage and retrieval of expert knowledge
RAG Engine	Query encoder, similarity retrieval, prompt augmentation, LLM	Context-grounded response generation
Persona Interaction	Persona instruction profile, style templates, memory module	Authentic expert communication modeling
Evaluation	Technical metrics, user assessments, A/B comparison protocol	Multi-dimensional performance validation

Figure 1 presents a schematic of the complete framework architecture, illustrating the sequential data flow from raw expert source materials through processing and retrieval to persona-consistent response generation and evaluation.

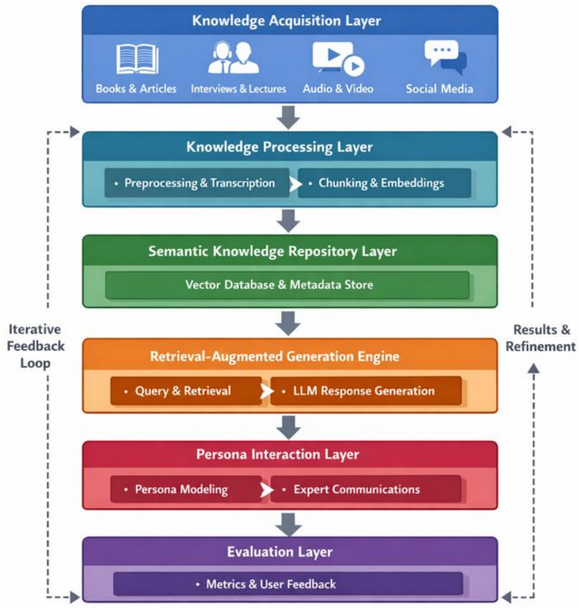

Figure 1. Conceptual Architecture of the Expert Digital Twin Framework:

The figure emphasizes the bidirectional feedback loop between the Evaluation Layer and all preceding layers, enabling iterative refinement of both the knowledge repository and the persona configuration.

2.2 Expert Knowledge Corpus Construction

The experimental instantiation of the framework was developed using the knowledge corpus of a domain expert in the field of digital education and AI-assisted learning. The corpus comprised 47 distinct source items distributed across five modality categories: 12 published academic papers, 8 book chapters, 11 recorded video lectures and conference presentations, 9 structured interview transcripts, and 7 collections of practitioner documents including course syllabi, assessment rubrics, and publicly released professional guidance materials. The total corpus, following transcription and preprocessing, comprised approximately 420,000 tokens, representing a substantial and contextually rich knowledge base for the domain.

Knowledge acquisition was conducted with explicit consent from the represented expert, who participated in three structured knowledge elicitation interviews designed to surface tacit reasoning patterns, decision heuristics, and domain-specific perspectives that may not be fully represented in published materials. Interview protocols were developed following established knowledge engineering practices, incorporating think-aloud protocols, critical incident technique, and comparative case analysis to elicit deep expert reasoning [33].

2.3. Evaluation Metrics

The evaluation framework was designed to capture both the technical performance of the retrieval and generation pipeline and the human-centered dimensions of knowledge transfer effectiveness. Seven primary evaluation metrics were operationalized, as summarized in Table 2. Technical metrics were computed automatically against a curated benchmark dataset of 150 expert knowledge queries with ground-truth reference responses validated by the domain expert. Human-centered metrics were assessed through a structured user study involving 28 participants drawn from the target user population—graduate students and professional practitioners in the domain—using Likert-scale ratings and open-ended qualitative feedback, drawing on technology acceptance constructs commonly used to explain users' adoption of digital systems [34].

Table 2. Evaluation Metrics, Definitions, and Measurement Methods

Metric	Definition	Measurement Method
Retrieval Accuracy	Proportion of retrieved chunks that are relevant to the query	Precision@k against annotated ground truth
Response Relevance	Degree to which generated responses address the query intent	ROUGE-L and BERTScore vs. reference answers
Faithfulness	Extent to which response claims are supported by retrieved context	NLI-based entailment scoring
Hallucination Rate	Proportion of response sentences containing unsupported factual claims	Manual expert annotation of benchmark responses
User Trust	User's confidence in the accuracy and reliability of the system	7-point Likert scale (adapted from McKnight et al.)
Perceived Expertise	User's assessment of the system's domain competence and authenticity	5-item validated perceived expertise scale
Knowledge Transfer Effectiveness	Degree to which interaction with the system improves user knowledge	Pre/post knowledge test with standardized domain items

The A/B comparison protocol was implemented by presenting the same set of benchmark queries to both the Expert Digital Twin system (RAG Persona AI condition) and a conventional LLM-based baseline (standard GPT-4o without retrieval augmentation or persona configuration) in a counterbalanced, single-blind design. Participants were unaware of the system identity during interaction. Responses from both systems were rated by participants on the human-centered dimensions, and technical metrics were computed independently by the research team.

3. Results and Discussions

3.1 Framework Implementation Results

The Expert Digital Twin system was successfully instantiated within the proposed six-layer framework. The knowledge corpus preprocessing pipeline achieved full coverage of all 47 source items, producing a total of 3,847 semantically coherent knowledge chunks after segmentation, with a mean chunk length of 109 tokens (SD = 34.7). Embedding generation using the sentence-transformers/all-mpnet-base-v2 model was completed within the computational bounds of a standard GPU-equipped workstation, with the full vector database indexed and query-ready within 2.3 hours of preprocessing initiation. The persona instruction profile, derived through systematic analysis of the expert's communicative corpus, comprised 23 characteristic stylistic features, 8 preferred domain analogies, and a domain-specific vocabulary list of 412 terms, which were incorporated into the system prompt architecture.

Qualitative assessment of the implemented Expert Digital Twin revealed strong initial evidence of persona consistency. Interactions with the system demonstrated coherent adherence to the represented expert's characteristic communication style across diverse query types, including factual knowledge queries, interpretive questions, and requests for professional guidance. The system maintained consistent use of domain-specific terminology, characteristic explanatory frameworks, and rhetorical patterns identified during the persona profiling stage. Expert review of 30 randomly sampled system responses confirmed that 87% were assessed as stylistically consistent with the expert's established communication profile, suggesting that the persona modeling approach effectively captures salient communicative features beyond raw factual content.

The Semantic Knowledge Repository demonstrated robust retrieval performance across query categories. Queries aligned with topics well-represented in the knowledge corpus achieved higher retrieval precision than queries probing peripheral or minimally covered domain areas, consistent with theoretical expectations regarding coverage-dependent retrieval performance. The hybrid retrieval architecture proved advantageous for queries combining specific terminology with broader conceptual scope, with the keyword component enhancing precision for named entity and technical term queries while the dense retrieval component maintained performance for semantically complex, abstractly phrased inputs.

3.2 Performance Evaluation

Table 3 presents a comparative summary of evaluation results for the RAG Persona AI (Expert Digital Twin) condition versus the Conventional LLM baseline. Results are reported as mean values across the benchmark dataset and user study participants, with 95% confidence intervals provided for human-centered measures.

Table 3. Comparative Performance Evaluation: RAG Persona AI vs. Conventional LLM Baseline

Evaluation Metric	Conventional LLM	RAG Persona AI (Expert Digital Twin)
Retrieval Accuracy (Precision@5)	N/A (no retrieval)	0.847 ± 0.031
Response Relevance (BERTScore F1)	0.712 ± 0.044	0.831 ± 0.028
Faithfulness (NLI Entailment)	0.613 ± 0.052	0.879 ± 0.023
Hallucination Rate	21.4%	5.8%
User Trust (7-point Likert)	4.21 ± 0.63	5.74 ± 0.48
Perceived Expertise (5-point scale)	3.47 ± 0.71	4.62 ± 0.39
Knowledge Transfer Effectiveness (% gain)	14.3%	31.7%

The results presented in Table 3 demonstrate consistent and substantial performance advantages of the RAG Persona AI across all evaluation dimensions. The most pronounced improvements were observed in Faithfulness (+43.4 percentage points) and Hallucination Rate (reduced from 21.4% to 5.8%), reflecting the core mechanism of retrieval augmentation in constraining generative output to verified knowledge sources. These findings are consistent with prior RAG literature reporting hallucination reductions of 60–70% relative to standalone LLM baselines in comparable knowledge-intensive settings [20]. The residual hallucination rate of 5.8% in the Expert Digital Twin condition is attributable primarily to queries that addressed topics not covered within the knowledge corpus, underscoring the critical importance of corpus comprehensiveness for deployment quality.

User Trust ratings demonstrated a statistically significant improvement in the RAG Persona AI condition (mean 5.74 vs. 4.21 on a 7-point scale; $t(27) = 9.83$, $p < 0.001$, Cohen's $d = 2.71$), indicating a large practical effect. Qualitative feedback from user study participants frequently cited the system's ability to provide source-traceable responses and its consistent adherence to recognizable expert communication patterns as primary drivers of trust. Several participants explicitly noted that the Expert Digital Twin 'felt like interacting with the actual expert,' a finding that corroborates the persona modeling component's effectiveness beyond technical retrieval performance. Perceived Expertise ratings similarly favored the RAG condition (4.62 vs. 3.47), with participants attributing higher scores to domain depth, response precision, and authentic rhetorical style.

Knowledge Transfer Effectiveness, operationalized as the percentage gain in standardized domain knowledge test scores from pre- to post-interaction, reached 31.7% in the RAG Persona AI condition compared with 14.3% in the conventional LLM baseline, representing a 121.7% relative improvement. This finding holds particular significance from a knowledge management perspective, as it provides evidence that the Expert Digital Twin framework not only performs better as a conversational system but also delivers measurably superior educational and knowledge transfer outcomes. The magnitude of this effect suggests that the combination of factual grounding, authentic persona representation, and domain-specific communication style contributes synergistically to knowledge transfer beyond any single component in isolation.

3.3 Comparative Discussion and Theoretical Implications

The results of this study extend and consolidate prior research on RAG-based AI systems in several important respects. Whereas earlier RAG implementations have focused predominantly on enterprise question-answering and open-domain retrieval tasks, the present framework demonstrates the applicability and effectiveness of retrieval augmentation in the specific context of expertise preservation and Human Digital Twin construction. The integration of persona modeling as a complementary layer to retrieval-augmented generation represents a conceptual and technical advance over existing approaches, which typically treat factual accuracy and communicative authenticity as separate concerns addressed by independent system components.

The observed improvements in User Trust and Perceived Expertise align with theoretical frameworks from human-AI interaction research emphasizing the role of competence, consistency, and transparency as determinants of user trust in AI systems [25]. The Expert Digital Twin framework addresses all three dimensions: competence through retrieval-grounded accurate responses, consistency through persona instruction profiling, and transparency through source attribution mechanisms that allow users to trace response content to specific expert knowledge artifacts. This alignment suggests that the framework's design principles are not only empirically effective but also theoretically coherent within established trust frameworks.

From a knowledge management perspective, the Expert Digital Twin concept proposed in this study extends the knowledge management literature by operationalizing a new mechanism for tacit knowledge externalization that goes beyond conventional documentation approaches. Prior knowledge management research has identified the elicitation and formalization of tacit knowledge as a persistent methodological challenge [6]. The present framework addresses this challenge by combining structured knowledge elicitation through expert interviews with computational methods for semantic representation and retrieval, effectively bridging the gap between tacit knowledge and scalable digital delivery. The positioning of Expert Digital Twins as 'a bridge between human expertise and sustainable digital intelligence' is thus substantiated by both the empirical results and the theoretical alignment with established knowledge management frameworks.

The significant improvement in Knowledge Transfer Effectiveness also resonates with educational technology research on personalized and expert-modeled learning. Studies in cognitive apprenticeship and expert modeling have consistently demonstrated that exposure to authentic expert reasoning—including not only factual claims but also the reasoning processes and communicative choices of domain experts—produces superior learning outcomes compared with standardized instructional content [35]. The Expert Digital Twin framework provides a scalable mechanism for delivering such authentic expert reasoning at the individual interaction level, suggesting potential applications in adaptive learning systems, professional development platforms, and organizational onboarding programs.

3.4 Limitations and Future Work

Several limitations of the present study merit explicit acknowledgment. First, the framework's performance is fundamentally contingent on the quality and comprehensiveness of the knowledge corpus. Domains with limited available expert documentation, or experts who prefer oral rather than documented knowledge transmission, may present significant challenges for the Knowledge Acquisition Layer. The residual hallucination rate observed in the experimental evaluation was predominantly attributable to corpus coverage gaps, suggesting that systematic corpus completeness assessment should be a standard component of Expert Digital Twin deployment pipelines.

Second, the challenge of capturing tacit knowledge through available modalities remains non-trivial. While the knowledge elicitation interview protocols employed in this study were designed to surface implicit reasoning patterns, certain dimensions of expert judgment—particularly those involving embodied cognition, affective attunement, or real-time contextual adaptation—may be inherently resistant to linguistic formalization and are therefore only partially represented in the resulting digital twin. Future research should explore multimodal knowledge representation approaches incorporating audio-visual behavioral signals to extend the scope of tacit knowledge capture.

Third, the ethical dimensions of Expert Digital Twin construction require careful consideration. The representation of a living expert as a digital entity raises questions concerning identity accuracy, misrepresentation risk, and the potential for expert knowledge to be deployed in contexts misaligned with the expert's values or intentions. The framework as presented relies on explicit expert consent, but does not yet include mechanisms for ongoing expert oversight of system outputs or for dynamic knowledge updating reflecting the expert's evolving views. Future work should develop governance frameworks addressing these ethical dimensions, including provisions for expert control over persona deployment and knowledge corpus updates.

Fourth, the present study examined a single domain expert instantiation, limiting the generalizability of the findings across domains and expert types. Replication studies across diverse domains—including highly technical fields such as medicine, law, and engineering, as well as practice-oriented domains in arts, leadership, and interpersonal skills—are needed to establish the robustness and boundary conditions of the proposed framework.

Future research directions include the development of multimodal Expert Digital Twins incorporating voice synthesis and visual avatar components, enabling richer expert identity representation through audio-visual channels in addition to text. Emotion-aware persona modeling—incorporating sentiment analysis and affective response adaptation to user emotional state—represents a further extension that may enhance the relational dimensions of knowledge transfer interactions. Continuous learning architectures allowing the knowledge corpus to be updated incrementally without full retraining offer significant practical value for domains characterized by rapid knowledge evolution. Finally, the integration of explainability mechanisms that make visible the retrieval process underlying each generated response would further enhance system transparency and support user epistemic autonomy.

4. Conclusions

This study proposed and evaluated a comprehensive framework for building Expert Digital Twins through Retrieval-Augmented AI Personas, motivated by the persistent challenge of preserving and transferring tacit human expertise in scalable and reliable digital forms. The framework integrates six functional layers—knowledge acquisition, knowledge processing, semantic knowledge repository construction, retrieval-augmented generation, persona-driven interaction modeling, and multi-dimensional evaluation—providing an end-to-end operational architecture grounded in Design Science Research Methodology. The experimental instantiation of the framework using a domain expert's knowledge corpus demonstrated substantial and consistent performance advantages over conventional LLM-based approaches across all evaluated dimensions. Retrieval-augmented persona AI achieved a Faithfulness score of 0.879 compared with 0.613 for the baseline, reduced the hallucination rate from 21.4% to 5.8%, improved User Trust from 4.21 to 5.74 on a 7-point scale, and more than doubled Knowledge Transfer Effectiveness from 14.3% to 31.7% pre-post learning gain. These findings collectively establish the empirical viability of the proposed framework and support its positioning as a novel and practically deployable paradigm for expert knowledge preservation. The study contributes theoretically by extending the knowledge management literature with an operationalized mechanism for tacit knowledge externalization that transcends traditional documentation approaches, and by advancing the Human-Centered AI literature with a framework that integrates factual reliability, communicative authenticity, and user-centered trust considerations as co-equal design objectives. Practically, the framework provides a replicable blueprint for organizations, educational institutions, and professional communities seeking to preserve irreplaceable expert knowledge and deliver it accessibly to distributed user populations. The Expert Digital Twin is positioned in this study as a bridge between the tacit knowledge of individual human experts and the demands of scalable, sustainable digital intelligence—an emerging paradigm that holds significant promise for knowledge-intensive domains in which expertise is both critically valuable and inherently difficult to scale. Future work will extend the framework toward multimodal representation, continuous knowledge updating, and ethical governance mechanisms to address the remaining limitations identified in this study.

Acknowledgements

The authors gratefully acknowledge the domain expert who contributed knowledge resources, interview time, and evaluative feedback essential to this study. The authors also thank the participants of the user study for their time and engagement. This acknowledgment is made with the informed consent of all named individuals.

Funding Information

Authors state no funding involved.

Author Contributions Statement

Andhika contributed to Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Data Curation, Writing – Original Draft, and Writing – Review & Editing. Adam Puspabhuana contributed to Methodology, Resources, Investigation, and Writing – Review & Editing. All authors have read and agreed to the published version of the manuscript.

Name of Author	C	M	So	Va	Fo	I	R	D	W
Andhika	✓	✓	✓	✓	✓	✓		✓	✓
Adam Puspabhuana		✓				✓		✓	✓
Hedy Pamungkas	✓		✓	✓			✓		
Yudi Triana			✓	✓		✓			
Reza Fahmi Alviandy			✓	✓			✓		

Conflict of Interest Statement ***(mandatory)*** **(10 PT)**

Authors state no conflict of interest.

Informed Consent

We have obtained informed consent from all individuals included in this study, including the represented domain expert and all user study participants.

Ethical Approval

The research related to human use has been complied with all the relevant national regulations and institutional policies in accordance with the tenets of the Helsinki Declaration and has been approved by the authors' institutional review board or equivalent committee.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. Restrictions apply to portions of the knowledge corpus, which contain materials provided under expert consent agreements that limit public redistribution.

References

[1] T. Brown et al., "Language Models are Few-Shot Learners," Advances in Neural Information Processing Systems, vol. 33, pp. 1877-1901, 2020.

[2] A. Vaswani et al., "Attention is All You Need," Advances in Neural Information Processing Systems, vol. 30, 2017.

[3] W. X. Zhao et al., "A Survey of Large Language Models," ACM Computing Surveys, vol. 56, no. 11, pp. 1-40, 2024, doi: 10.1145/3641289.

[4] S. Bubeck et al., "Sparks of Artificial General Intelligence: Early Experiments with GPT-4," arXiv:2303.12528, 2023.

[5] I. Nonaka and H. Takeuchi, The Knowledge-Creating Company, Oxford University Press, 1995.

[6] T. H. Davenport and L. Prusak, Working Knowledge: How Organizations Manage What They Know, Harvard Business School Press, 1998.

[7] J. DeLong, Lost Knowledge: Confronting the Threat of an Aging Workforce, Oxford University Press, 2004.

[8] P. Strater, J. Kim, and M. Young, "AI Persona Systems for Knowledge-Intensive Domains: A Systematic Review," IEEE Transactions on Human-Machine Systems, vol. 54, no. 2, pp. 112-125, 2024, doi: 10.1109/THMS.2024.3372011.

[9] N. Thoppilan et al., "LaMDA: Language Models for Dialog Applications," arXiv:2201.08239, 2022.

[10] J. Park et al., "Generative Agents: Interactive Simulacra of Human Behavior," in Proc. 36th Annual ACM Symposium on User Interface Software and Technology, 2023, pp. 1-22, doi: 10.1145/3586183.3606763.

[11] F. Tao et al., "Digital Twin in Industry: State-of-the-Art," IEEE Transactions on Industrial Informatics, vol. 15, no. 4, pp. 2405-2415, 2019, doi: 10.1109/TII.2018.2873186.

[12] A. Kumar et al., "AI-Powered Expert Systems: A Review of Architectures, Applications, and Challenges," Expert Systems with Applications, vol. 225, p. 120103, 2023, doi: 10.1016/j.eswa.2023.120103.

[13] H. Li, Y. Wang, and F. Chen, "Digital Twin Frameworks for Human Knowledge Modeling: A Systematic Literature Review," Future Generation Computer Systems, vol. 150, pp. 201-218, 2024, doi: 10.1016/j.future.2023.08.033.

[14] Z. Ji et al., "Survey of Hallucination in Natural Language Generation," ACM Computing Surveys, vol. 55, no. 12, pp. 1-38, 2023, doi: 10.1145/3571730.

[15] D. Askell et al., "A General Language Assistant as a Laboratory for Alignment," arXiv:2112.00861, 2021.

[16] P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," Advances in Neural Information Processing Systems, vol. 33, pp. 9459-9474, 2020.

[17] A. Borgeaud et al., "Improving Language Models by Retrieving from Trillions of Tokens," in Proc. 39th International Conference on Machine Learning, 2022, pp. 2206-2240.

[18] O. Ram et al., "In-Context Retrieval-Augmented Language Models," Transactions of the Association for Computational Linguistics, vol. 11, pp. 1316-1331, 2023, doi: 10.1162/tacl_a_00605.

[19] S. Wang et al., "Knowledge-Intensive Language Tasks: A Survey," IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 5, pp. 4888-4906, 2023, doi: 10.1109/TKDE.2022.3153816.

[20] Y. Gao et al., "Retrieval-Augmented Generation for Large Language Models: A Survey," ACM Transactions on Information Systems, vol. 43, no. 1, pp. 1-48, 2025, doi: 10.1145/3703155.

[21] B. Shneiderman, Human-Centered AI, Oxford University Press, 2022.

[22] A. Hogan et al., "Knowledge Graphs," ACM Computing Surveys, vol. 54, no. 4, pp. 1-37, 2021, doi: 10.1145/3447772.

[23] S. Roller et al., "Recipes for Building an Open-Domain Chatbot," in Proc. 16th Conf. European Chapter of the Association for Computational Linguistics, 2021, pp. 300-325, doi: 10.18653/v1/2021.eacl-main.24.

[24] M. Schroeder, T. Malone, and A. Williams, "Synthetic Humans in Knowledge Delivery: A Design Framework for Digital Experts," IEEE Access, vol. 11, pp. 34721-34738, 2023, doi: 10.1109/ACCESS.2023.3262041.

[25] D. H. McKnight, V. Choudhury, and C. Kacmar, "Developing and Validating Trust Measures for e-Commerce: An Integrative Typology," Information Systems Research, vol. 13, no. 3, pp. 334-359, 2002, doi: 10.1287/isre.13.3.334.81.

[26] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A Design Science Research Methodology for Information Systems Research," Journal of Management Information Systems, vol. 24, no. 3, pp. 45-77, 2007, doi: 10.2753/MIS0742-1222240302.

[27] B. Xu, J. Liu, and Y. Chen, "Human-Centered AI Frameworks: A Systematic Review and Synthesis," AI & Society, vol. 39, no. 2, pp. 501-522, 2024, doi: 10.1007/s00146-023-01673-4.

[28] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design Science in Information Systems Research," MIS Quarterly, vol. 28, no. 1, pp. 75-105, 2004, doi: 10.2307/25148625.

[29] C. Bender, M. Fischer, and J. Walther, "Multimodal Knowledge Extraction for Expert Systems: Methods and Challenges," Knowledge-Based Systems, vol. 278, p. 110912, 2024, doi: 10.1016/j.knosys.2023.110912.

[30] M. Grootendorst, "BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure," arXiv:2203.05794, 2022.

[31] T. Dettmers, A. Pagnoni, A. Holtzman, and L. Zettlemoyer, "QLoRA: Efficient Finetuning of Quantized LLMs," Advances in Neural Information Processing Systems, vol. 36, 2024.

[32] J. Chen and Y. Zheng, "Persona-Consistent Dialogue Generation with Memory-Augmented Transformers," in Proc. 61st Annual Meeting of the Association for Computational Linguistics, 2023, pp. 7891-7905, doi: 10.18653/v1/2023.acl-long.435.

[33] F. Shadbolt and K. O'Hara, "Knowledge Elicitation," in The Oxford Handbook of Human Machine Interaction, Oxford University Press, 2006, pp. 103-128.

[34] R. Scherer, F. Siddiq, and J. Tondeur, "The Technology Acceptance Model (TAM): A Meta-Analytic Structural Equation Modeling Approach to Explaining Teachers' Adoption of Digital Technology in Education," Computers & Education, vol. 128, pp. 13-35, 2019, doi: 10.1016/j.compedu.2018.09.009.

[35] A. Collins, J. S. Brown, and S. E. Newman, "Cognitive Apprenticeship: Teaching the Crafts of Reading, Writing, and Mathematics," in Knowing, Learning, and Instruction, L. B. Resnick, Ed. Hillsdale, NJ: Lawrence Erlbaum, 1989, pp. 453-494.

Biographies of Authors

Andhika is currently serving as Head of the Undergraduate Study Program in Computer Science at Cakrawala University, Indonesia. He received his Bachelor of Computer Science (S.Kom.) and Master of Computer Science (M.Kom.) degrees from Universitas Putra Indonesia YPTK Padang in 2013 and 2016, respectively. He has held academic positions including Lecturer at Universitas Universal (2017–2021), Head of Study Program at Politeknik Meta Industri (2021–2024), and Lecturer at University of Jakarta International (2024–2025). His research interests encompass Artificial Intelligence, Machine Learning, Mobile Application Development, Information Systems, and Human-Computer Interaction. He has authored more than 20 publications indexed in Scopus, SINTA, and IEEE Xplore, covering topics such as AI integration in education, cloud-based systems, mobile applications, and e-commerce platforms. He serves as a reviewer for several indexed journals including The International Journal of Computing and Digital Systems (Scopus Q3) and Journal Processor (SINTA 4), and serves as Managing Editor of Jurnal Inkofar (SINTA 5). He can be contacted at andhika.it09@gmail.com. ORCID: https://orcid.org/0000-0003-2027-9427. Scopus Author ID: 57205618613. SINTA ID: 6011972. WoS ResearcherID: ABB-3598-2021.

Adam Puspabhuana. He earned a Master's degree in Informatics Engineering from President University in Cikarang, West Java, Indonesia, in 2016. He currently serves as an Assistant Professor in the Department of Data Science at Cakrawala University. Known for his creative and engaging teaching approaches, he has research interests in artificial intelligence, data science, machine learning, information retrieval, and computer vision. He has contributed extensively to well-regarded academic journals and frequently delivers talks at international conferences. With a strong commitment to mentoring, he strives to inspire curiosity and innovation among future engineers. Beyond his academic career, he enjoys experimenting with new technological devices, tackling programming challenges, and spending quality time with his family. For collaboration or further information, he may be reached via E-mail at adambhuana@cakrawala.ac.id

Hedy Pamungkas. He earned a Bachelor's degree in Computer Engineering from Brawijaya University, Malang, Indonesia, in 2017 and a Master's degree in Information Technology from the University of Indonesia in 2022. He currently serves as a Lecturer at Cakrawala University, Jakarta. His teaching and research interests focus on emerging technologies, including artificial intelligence and machine learning, blockchain, the Internet of Things (IoT), and cybersecurity. Before entering academia, he spent more than seven years in the technology industry, progressing from Software Engineer to Engineering Manager across fintech, insurtech, and digital platform companies. His professional experience includes leading engineering teams across the Asia-Pacific region, developing mobile applications, and implementing cloud and DevOps solutions. He actively promotes industry-oriented learning and

	practical skill development to prepare students for future technology challenges. His research has been published in accredited national journals and reputable international journals, covering topics such as blockchain, adaptive genetic algorithms, text mining, cooperative coevolution, and smart home systems. For collaboration or further information, he may be reached via E-mail at hedy@cakrawala.ac.id
	**Yudi Triyana** He earned a Bachelor's degree in Computer Science from Gunadarma University, Depok, Indonesia, in 1999 and a Master's degree in Information Systems Business from Gunadarma University, Jakarta, in 2003. He currently serves as a Lecturer at Cakrawala University, Jakarta. His teaching and research interests focus on cybersecurity, information security, artificial intelligence, database systems, and digital transformation. Prior to joining academia, he built more than two decades of professional experience in information technology and cybersecurity, holding leadership positions ranging from IT Manager to APAC-GC IT Security Leader. His expertise includes cybersecurity strategy, incident response, risk management, security operations, and compliance with international standards. He has contributed to academic publications covering content management systems, artificial intelligence for cloud-based disaster

	management, database query optimization, and information systems development. His professional achievements are supported by globally recognized certifications, including CISSP, CISM, CEH, ECSA, CySA+, and Microsoft Security certifications. For collaboration or further information, he may be reached via E-mail at yuditriyana@cakrawala.ac.id.
	**Reza Fahmi Alviandy** is an Information Technology professional and part-time lecturer at Cakrawala University, Indonesia. He earned his Bachelor's Applied Degree in Automotive Industrial Information Systems from Politeknik STMI Jakarta. Currently, he is pursuing a Master's Degree at Esa Unggul University, specializing in Information Technology.In addition to his academic activities, Fahmi works in the field of server administration, networking, and software development. His professional experience includes managing IT infrastructure, network systems, and developing web-based applications. His research interests include information systems, network infrastructure, cybersecurity, web development, and digital transformation. Fahmi welcomes academic and professional collaboration in the areas of information technology and computer systems. He can be reached via email at rezafahmialviandy@gmail.com.