Responsible Management for AI-augmented Knowledge: A Knowledge Ecosystem Approach

Belinda Wang
The University of Sydney
yichen.wang@sydney.edu.au

Barney Tan
University of New South Wales
barney.tan@unsw.edu.au

Sebastian Boell
The University of Sydney
sebastian.boell@sydney.edu.au

Jie Yu
University of Nottingham Ningbo
jie.yu@nottingham.edu.cn

Abstract

This study adopts a knowledge ecosystem approach to explore how organizations manage AI-augmented knowledge responsibly. Drawing from a case study on a large AI solution provider for manufacturing sector, our study reveals a dynamic knowledge management process that shapes and was shaped by the interdependent interactions among ecosystem actors. The preliminary findings in this study contribute to responsible AI and knowledge management literature, enhancing the understanding on actors' role to responsible knowledge management.

Keywords: Artificial Intelligence, Responsible AI, Knowledge Management, Knowledge Ecosystem

1. Introduction

Artificial Intelligence (AI) refers to the science of creating machines that simulate human intelligence, designed to mimic human cognitive functions and automate information and knowledge extraction through representation learning (Borges et al., 2021). From the knowledge management (KM) perspective, AI revolutionizes organizational knowledge production (Alavi et al., 2024), leading to innovations through new approaches to knowledge creation (Stahl, 2022). Studies indicate promise in uncovering new knowledge from big data that has previously eluded experienced experts (Benbya et al., 2024; Berente et al., 2021), potentially surpassing traditional human intelligence in knowledge creation (Bauer et al., 2023). While numerous studies highlight the business value of AI for predictive purposes, growing concerns about its trustworthiness and reliability pose challenges to its adoption in organizational KM (Dennehy et al., 2023; Jan et al., 2023). In response, research advocates for Responsible AI (RAI), which prioritizes accountability, transparency, and the alignment of AI with societal and human values throughout its use lifecycle (Mikalef et al., 2022).

Current RAI literature for managing AI-augmented knowledge primarily focuses on **explaining (XAI)** and **controlling** approaches. To ensure reliable outcomes, the XAI stream emphasizes using technical algorithms to interpret the mechanisms behind AI-generated knowledge in applications (Bauer et al., 2023). Meanwhile, the controlling approach aims to embed constraints in AI uses to enforce human oversight. However, the XAI approach often remains confined to technical discourse, lacking integration with domain experts who are the primary source of knowledge (Bauer et al., 2023; Jussupow et al., 2021). On the other hand, the controlling approach struggles to effectively translate expert knowledge into AI knowledge, leading to technical inaccuracies and unacceptable performance (Lebovitz et al., 2021; van den Broek et al., 2021). These challenges hinder the transfer of domain knowledge and the reuse of AI-generated knowledge. Therefore, our literature review identifies a research gap in the socio-technical management of AI-augmented knowledge to ensure responsible outcomes. Addressing these gaps can help organizations maximize the creation, transfer, and sharing of AI-augmented knowledge, thereby transforming knowledge management into a strategic asset and an inimitable resource for competitive advantage (Alavi & Leidner, 2001; Lee & Choi, 2003).

Organizational knowledge cannot be created solely internally. Hence, the rise of ecosystems as a new form of organizing organizational knowledge offers an opportunity to understand the role of AI in augmenting KM through interactions with diverse ecosystem actors (Wang, 2021). AI facilitates KM from multiple sources embedded in socio-technical interactions among actors, promoting continuous responsible knowledge creation with AI at the ecosystem level (Jarvaid et al., 2022). Therefore, this research aims to address the following question: *How can organizations ensure responsible management within their knowledge ecosystems when leveraging AI-augmented knowledge?*

URI: https://hdl.handle.net/10125/109511
978-0-9981331-8-8

2. Literature Review

2.1. Responsible AI Management

The growing interest in AI has been accompanied by a myriad of instances where negative or unintended consequences have surfaced. Unlike more transparent symbolic models, such as decision trees, which are explainable to users, machine learning (ML) models or deep learning (DL) models are often considered inscrutable (Mikalef et al., 2022). This inscrutability implies that the decision-making mechanisms employed in ML-based AI systems differ significantly from how knowledge was traditionally constructed in symbolic expert systems (Adadi & Berrada, 2018). Consequently, commercial adoption businesses may harbor skepticism and reluctance towards embracing AI systems due to lack transparency and trust in AI, even if these systems have demonstrated enhanced operational efficiency (Fernández-Loría et al., 2022). A pertinent example is the Robodebt scheme implemented by the Australian Government, which exposed procedural automation deficiencies, resulting in the issuance of half a million inaccurate debt notices to welfare recipients (Braithwaite, 2020; Rinta-Kahila et al., 2022). This failure eroded trust in the system and led to the suspension of the scheme, highlighting the importance of RAI (Braithwaite, 2020).

RAI refers to the use of AI systems that align with user expectations, organizational values, and social norms. RAI is often studied interchangeably with human-centered AI, ethical AI, and trustworthy AI (Gregor, 2024). It encompasses dimensions such as transparency, ethics, accountability, compliance, and robustness throughout the AI system's lifecycle (Mikalef et al., 2022). Through a review of current RAI and managing AI literature using an organizing review approach (Leidner, 2018), we have synthesized the management focus into two major streams.

The first stream focuses on the **explaining (XAI) approach**, which aims to make the inscrutable AI models and the associated decisions to be transparent. Within this stream, researchers propose various algorithms to address the opacity of AI models, attracting popularity in computer science literature for its role in enhancing transparency, thereby facilitating communication between stakeholders involved in AI-augmented knowledge (Mikalef et al., 2022). For instance, attention-based models interpret image content, while methods like LIME or SHAP explain and evaluate feature importance within models (Bauer et al., 2023; Fernández-Loría et al., 2022). Additionally, counterfactual approaches identify crucial features through evidence-based model explanations (Fernández-Loría et al., 2022). Hence, this stream aims to enhance transparency in understanding and interpreting AI generated knowledge that were previously opaque, thereby ensuring technical robustness in empirical applications.

The second stream focuses on the **controlling approach**, aims to ensure accountability and ethical AI use. This stream emphasizes applying constraint and implementing human-oversight to ensure responsible AI-augmented knowledge application. For instance, the European Union mandates human verification in every medical diagnosis made by AI systems. Constraints can also be applied to create a safe operating infosphere, ensuring that AI systems are effectively controlled through data training (Asatiani et al., 2021). As such, these approaches emphasize a human-AI augmentation, particularly in decision-support systems that typically require human expertise (Grønsund & Aanestad, 2020), to sustain socially fair (Teodorescu et al., 2021), procedurally compliant and technically accurate (Fügener et al., 2021; Lebovitz et al., 2021) in augmenting organizational knowledge.

Nevertheless, there are unaddressed challenges in above two approaches. XAI approaches primarily focus on technical aspects (Fernández-Loría et al., 2022), often neglecting a human science perspective. Consequently, they may fail to meet the needs of non-technical users within organizations, leading to the perception that explanations are not adequately understood by the actual users (Bauer et al., 2023; Herm et al., 2023). Moreover, the controlling approach lacks effectiveness in technically transferring between expert's knowledge and AI-generated knowledge, resulting in subpar machine performance and missed opportunities for knowledge transfer and reuse (Jussupow et al., 2021; Lebovitz et al., 2021; van den Broek et al., 2021). Consequently, our literature review has pinpointed a significant gap requiring further exploration on the responsible **socio-technical management** for AI-augmented knowledge.

2.2. Knowledge and Knowledge Ecosystem

Knowledge can be conceptualized in various ways, encompassing a state of mind, an object, a process, and a condition of having access to information and capability (Alavi & Leidner, 2001). Despite its multifaceted nature, the predominant

dimension for understanding how knowledge is created often revolves around tacit knowledge and explicit knowledge (Nonaka, 1994). Tacit knowledge refers to personal mental models, technical know-how, and skills that are localized within specific contexts and embedded in practice (Carlile, 2002). In contrast, explicit knowledge is codified learning and is transmittable in formal and systematic language (Nonaka, 1994). Knowledge creation constitutes a continuous management process involving individuals, communities, and functional boundaries.

From an organizational perspective, KM capabilities extend beyond managing internal knowledge and emphasize the need to effectively manage external knowledge acquisition and application (Gold et al., 2001). Technology serves as the enabler for organizational KM, facilitating knowledge flow through the KM process, integrating segmented knowledge, breaking down barriers to interaction, and aiding in the exploration of new knowledge creation (Lee & Choi, 2003). Both explicit and tacit knowledge can originate from diverse sources (Alavi & Leidner, 2001) and target towards diverse stakeholders, including AI (Harfouche et al., 2023), in a complementary manner (Tanriverdi, 2005). Therefore, it will be beneficial to investigate how knowledge interaction occurs at an ecosystem level to capture a holistic picture of how AI facilitates KM among various actors.

Ecosystems are groups of entities that collaborate on specific structures of relationships and are aligned to create value (Jacobides et al., 2018). In the literature on Knowledge Ecosystem, organizations aim to facilitate value creation through KM organized at the ecosystem level (Cobben et al., 2022). This necessitates multi-partner collaboration for knowledge creation, joint goal setting, and collective activities for KM goals (Järvi et al., 2018). Consequently, actors within the ecosystem, initially independent and distinct, engage in joint KM goal discovery and problem identification for mutual value creation in organizational knowledge. Given that knowledge resources within the ecosystem are distributed across entities (Dougherty & Dunne, 2011), domain knowledge boundaries (Bruns, 2013), and beyond individual organizational boundaries (Järvi et al., 2018), recent studies emphasize the emergence of the ecosystem with the presence of an anchor tenant (Clarysse et al., 2014). Anchor tenants have a significant role in developing the ecosystem by fostering disparate knowledge generation, assisting with establishing standards, and delegating different roles to actors across the value chain (Huang et al., 2018), therefore participating in multifaceted KM activities (Alavi & Leidner, 2001). To support these activities, AI, as a new actor in KM, shows potential in augmenting knowledge (Alavi et al., 2024; Harfouche et al., 2023). Consequently, AI-powered KM becomes a socio-technical process involving complex interactions between AI, the anchor tenant, and other ecosystem actors. However, there is still limited understanding of this intricate socio-technical process in achieving RAI (Stahl, 2022). Therefore, by utilizing knowledge ecosystems as a sensitizing concept, this study aims to unfold the socio-technical KM process for responsible management.

3. Method

We employed the case study method to address our research question, as this approach is well-suited for examining novel phenomena and their associated processes (Gephart, 2004), enabling the generation of rich insights necessary for inductive theory development (Walsham, 1995). Given that our research focuses on the responsible management of AI-augmented knowledge remains an area where existing theoretical frameworks are underdeveloped, the case study method is particularly appropriate (Eisenhardt & Graebner, 2007) for several reasons: First, this methodology allows us to explore the dynamic evolution of knowledge management processes through the interaction between AI and ecosystem actors. Second, the nature of our inquiry, which is analyzed at the ecosystem level, necessitates an in-depth examination of the anchor tenant's related ecosystem to gain a holistic understanding of the joint agency involved in the shared knowledge management practice (Järvi et al., 2018). Moreover, the complexity of socio-technical interactions among various actors within the knowledge ecosystem, combined with AI's knowledge creation capabilities, renders a positivist approach inadequate for capturing the intricacies of these phenomena (Walsham, 1995). Therefore, an interpretivist case study method offers a suitable inquiry device for describing and understanding these complex interactions (Eisenhardt & Graebner, 2007).

3.1. Data Collection

According to our research question, we identified three criteria for selecting our case study: 1) the case

shall use of AI in their knowledge management, helping the case firm to creating and transferring organizational knowledge; 2) functioning as an anchored actor interacting with multiple stakeholders for KM purposes; and 3) is considered as the leader in the industry to be qualified as the "extreme exemplar" for our target phenomenon (Eisenhardt & Graebner, 2007). Consequently, we negotiated access to SUPCON, a prominent provider of intelligent manufacturing solutions, as the "anchor tenant" for our case study. SUPCON met our rigorous relevance assessment criteria (Benbasat, 1999) and has positioned itself as a leading global provider committed to integrating AI technology with industry knowledge in their products. SUPCON has established a knowledge ecosystem in collaboration with complementor businesses, successfully co-developing and delivering AI solutions to numerous large size manufacturing firms across sectors. This makes it an ideal case for our research objectives.

Since its establishment in 1999, SUPCON has grown as a leader in AI-driven intelligent solutions for manufacturers. Publicly listed in 2000, SUPCON now commands approximately 21% of the market share in China for industrial digital solutions. Their key AI-based software offerings include the centralized intelligent IoT platform SupOS, the intelligent Process Control and Management platform i-OMC, the AI-driven Control platform Q-Lab, and the digital twin simulation platform APEX. To enhance its business ecosystem, SUPCON collaborates with industrial complementors as "channel partners" to drive innovation in knowledge management. The company serves over 30,000 customers across manufacturing sectors like chemical, petrochemical, oil and gas, and energy, all of whom have adopted SUPCON's commercialized intelligent systems.

Our current data collection to date comprised two phases. The initial preparation phase, conducted a month before fieldwork began, involved gathering secondary data from industry reports, corporate's social media documents, organizational materials, and news articles (Gioia et al., 2013). This data was crucial in informing the development of our interview protocols (Eisenhardt, 1989). In the second phase, conducted in October 2023, we visited SUPCON's commercial sites in Hangzhou, China, where we carried out 10 on-site interviews with average 90 minutes per session. This phase was part of a two-round interview process. In this round, we selected informants who could provide a comprehensive overview of the phenomenon under study (Pan & Tan, 2011) to validate our preliminary conceptualization. We focused on questions regarding how SUPCON manages both tacit and explicit knowledge from various stakeholders and ensures responsible design and deployment for client manufacturing businesses. The interviews included one director and one manager from the strategy and investment departments, four division managers and technical experts from the project management departments at SUPCON, and four representatives from SUPCON's complementors, including CyberInsights, specializing in industrial big data, and ProAIM, focusing on smart plant maintenance. All interviews were conducted in Chinese, audio-recorded, transcribed, and translated for analysis.

3.2. Data Analysis

We adopted grounded theory for coding, ensuring that our theoretical development was firmly rooted in the empirical data (Strauss & Corbin, 1998). We imported all transcribed data into NVivo 12, then conducted a first-order analysis of the interview data (Walsham, 1995), identifying 24 initial categories of concepts. These first-order concepts were then synthesized into more abstract second-order codes, leading to the emergence of nine themes from our data (Gioia et al., 2013). We also mapped the participating actors identified from the first-order concepts to the corresponding themes to our analysis. These second-order codes form the foundation for explaining the patterns in KM interactions, which we integrated into an aggregated dimension representing the three phases identified in the KM process (Langley, 1999). A sample node structure is depicted in Figure 1. From the identified KM dimensions, we inductively derived eight RAI dimensions, which emerged as the focal conceptual notions for each phase. Any disagreements in interpreting themes or RAI dimensions were resolved by revisiting interview data or consulting secondary sources, thereby triangulating our interpretations and iteratively refining our theory (Gioia et al., 2013). While our data collection is ongoing, we will continue to refine our theoretical framework to achieve theoretical saturation for a comprehensive understanding (Eisenhardt & Graebner, 2007). Based on our preliminary findings, we visualized our model and validated it with relevant informants and these results will be reported in the following section.

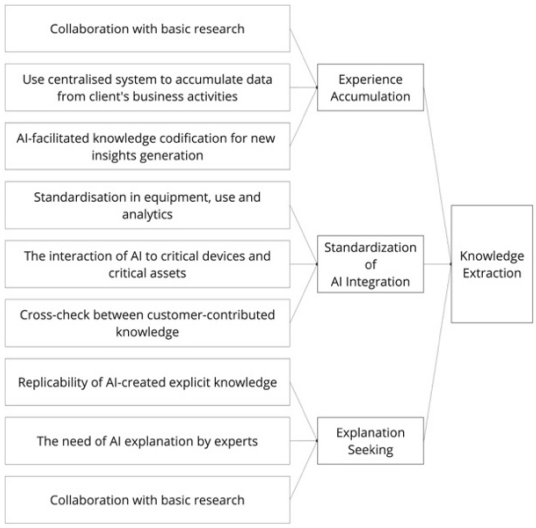

Figure 1. Sample code structure

4. Preliminary Findings

Our data identifies three distinct phases in SUPCON's KM process, highlighting key interactions among ecosystem actors. These interactions structure dynamic relationships that contribute to responsible AI management, involving nine key ecosystem actors collaborating with the focal tenant, SUPCON: customer manufacturers, suppliers, complementors, domain experts (internal and external to SUPCON), collaborating educational institutes, government (including industrial governing bodies), operators (e.g., machine operators at customers' sites), as well as the adopted AI systems and AI engineers. Therefore, we then develop the KM process model (see Figure 2) to demonstrate how interactions among these ecosystem actors drive responsible management of AI-augmented knowledge at different phases.

4.1. Knowledge Extraction

The first KM phase identified in our data is the Knowledge Extraction phase. This phase involves the ecosystem anchor coordinating the extraction of explicit knowledge from aggregated tacit knowledge. The adoption of AI enables organizations to capture dynamic knowledge continuously generated through business activities, which underpins the functions of knowledge extraction from AI implementations, leading to new knowledge co-creation between AI and multiple ecosystem actors. In facilitating effective knowledge extraction, our data reveals that key activities in this phase include **standardization of AI integration**, **experience accumulation**, and **explanation seeking**. Standardization for AI integration emphasize the need to attach sensors and data processing devices (e.g., IoT devices) to the critical assets of customer manufacturers to enable data collection with consistent formats. To ensure this consistency across multiple collections and analyses, SUPCON coordinated with suppliers and customers to standardize the integration process and provides a centralized operating system for customers to manage their collected data. A division manager explained, "*In the industrial sector, data silos are still very apparent. The ownership of data remains with the customer, who is very sensitive about this. We provide corresponding algorithms and models... However, even within the entire Customer A's system, data from different sites were not connected.*"

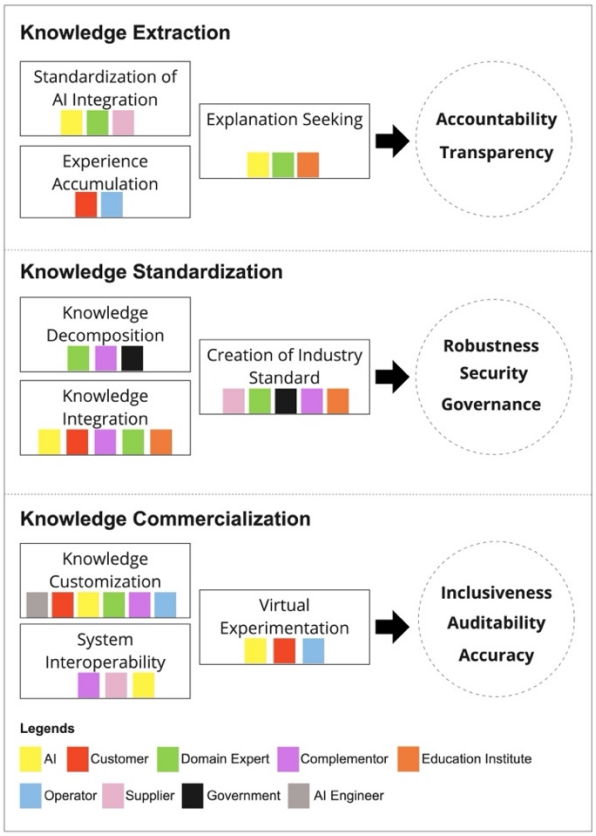

Figure 2. Responsible KM process model

Since tacit knowledge is inherently local and embedded in practice, making it difficult to articulate and retrieve (Carlile, 2002), the integration of AI allows for the quantified expression of such tacit knowledge. To facilitate the spread of tacit knowledge, SUPCON adopts a customer-facing business model, offering offline stores to customer in the ecosystem. These offline shops involve domain experts and AI engineers who test digital products, aiming to assist customers before adoption and after implementation.

As a medium, it helps the tacit knowledge expansion and integration. As one project manager at SUPCON noted: "…*If data cannot be shared, we can transfer experience. If a particular experience is successful in company A, company B can learn from it, and this [knowledge] can be passed down.*"

Additionally, AI-generated knowledge that spans across boundaries often becomes isolated due to temporal and physical constraints. SUPCON leverages its customer-facing business model and standardized operating system, SupOS, as a boundary-spanning tool. This approach maintains connections across engineering project sites and creates a shared field for knowledge creation (Levina & Vaast, 2005). As such, by tracking critical assets and integrating diverse sensor-collected data to operating systems, SUPCON accumulates data-based tacit knowledge from customers, particularly from the experience of factory operators, enabling AI to generate new explicit knowledge from tacit data. However, opaque AI decision-making raises significant concerns, especially when critical decisions impact human well-being and safety (Le et al., 2023). Our data suggests that converting tacit to explicit knowledge must be supported by domain theories and validated by domain experts, underpinning the explanation Seeking interaction from domain experts and affiliated education institutes. As the SUPCON's AI engineer noted, "*The best approach is the fusion of fundamental theory and data-driven algorithms. When data alone cannot provide explainability, domain knowledge is introduced. This translates into rules that the model recognizes, combining AI methods to produce accurate results*."

Hence, our analysis on actor interactions identifies **Accountability** and **Transparency** as focal dimensions of RAI derived from this phase. Research on XAI shows that local explanations require validation by human experts for relevance (Le et al., 2023) and accuracy (Saeed & Omlin, 2023), thereby enhancing the accountability of AI adoption consequences. Since explicit knowledge cannot be fully captured and created by AI without human validation, responsible management must ensure the interpretability of decision-making mechanisms and have them audited by domain experts using standardized data collection (Dwivedi et al., 2021). Thus, responsible knowledge extraction will require a bi-directional distillation between AI-generated knowledge and domain expertise from ecosystem actors (van den Broek et al., 2021), thereby promoting transparency in future empirical applications. However, this bi-directional exchange necessitates that early customer adopters conduct thorough evaluations of AI systems to fully understand their usage and the processes involved in deriving outputs (Lebovitz et al., 2021). Assessing the transferability and potential generalizability of these results across different use scenarios is therefore essential.

4.2. Knowledge Standardization

Knowledge standardization, understood as the convergence of knowledge to a "justified true belief" (Nonaka, 1994), is identified as the second phase in our data. This phase emphasizes the activities of **knowledge decomposition**, **knowledge integration**, and the **creation of industry standards**. Achieving this requires a collective effort from SUPCON, complementor businesses, customers, educational institutions, domain experts, and government agencies, with the goal of developing ecosystem-wide generalizable AI-augmented knowledge Our data indicates the implementation of a multi-sided platform, named i-OMC, to bring complementor businesses together to enable interactions and triangular knowledge exchanges. Localized explicit knowledge, even generated from data, often cannot offer generalizable solutions due to its situational limitations (Carlile, 2002). To enhance the transferability of knowledge between ecosystem actors, the challenge of knowledge generalizability can be addressed by decomposing complex knowledge into granular components or specific parts (i.e., Knowledge Decomposition). Decomposed approach facilitates the understanding and transfer of AI-generated knowledge across different assets and projects. As the CTO of CyberInsights explained, "*[We] break it down to a smaller component level as the failure types at the component level are highly similar, so I hope to accumulate models or algorithms for these components that can be reused.*"

With decomposed knowledge, the i-OMC platform facilitates the AI-augmented knowledge integration and cross-triangulation (i.e., Knowledge Integration) between complementor businesses, domain experts, and customers. However, ecosystem-level explicit knowledge generated by AI through big data analysis of aggregated local explicit knowledge is considered as "standards" for judging truthfulness only when it aligns with domain science justification (Nonaka, 1994). As one director stated: "*I can't only use big data learning or a general AI approach to solve problems. I must respect the experts, the industry,*

and simulations. This is a fundamental premise in the field of industrial AI."

As such, the validated explicit knowledge will be used for developing industrial standards (i.e. Creation of industry standard) to legitimize new knowledge generated or augmented by AI. However, privacy regulations and security concerns pose challenges in cross-customer data sharing. Our data underscore the role of the government in acting as a broker, facilitating knowledge sharing at the industry level. This involves working with domain experts to legitimize and accredit newly created standards, thereby supporting ongoing knowledge building and sharing within the ecosystem.

Consequently, our data analysis reveals that the relevant RAI dimensions in this phase are **Robustness, Security**, and **Governance**. Robustness is paramount for preventing adversarial perturbations in AI-driven knowledge creation, especially in scenarios with insufficient training data (Dwivedi et al., 2021). The complex and cost-intensive nature of manufacturing processes often limits the availability of extensive datasets (Demlehner et al., 2021), which hampers the application of AI in novel situations that fall outside the boundaries of existing training data. Hence, the validation from domain experts can extend the boundaries of robust uses. Security is crucial for protecting data privacy, addressing the concerns of large corporations and national security issues, which are central to IS ethics (Dennehy et al., 2023). The third dimension, Governance, focuses on risk management and quality assurance. Accredited solutions ensure the lawful validity and comparability of AI outputs, which therefore underscores the need for industry standards and regulatory "certifiability" to ensure responsible AI use, with potential involvement from insurance companies to cover AI-related risks (Martin, 2019). The certified newly knowledge will, in turn, contribute to basic research within educational institutions.

4.3. Knowledge Commercialization

Knowledge commercialization is the final phase, which focuses on offering standardized AI-augmented knowledge as a commodity at the ecosystem level, thereby capturing business value. Key components include **knowledge customization**, **system interoperability** and employing **virtual experimentation** to facilitate commercial adaptations. Maximizing commercialization value requires aligning adoption with business strategies and objectives, ensuring value-added benefits for customer manufacturers. SUPCON achieves this by customizing AI implementation by curating knowledge bases that target customers' operational needs and leverage AI for creating new knowledge insights. To enhance the curation of internalized knowledge, SUPCON's standardized factory operating system SupOS and AI control system Q-lab are designed for high interoperability with customers' existing systems regardless of vendors. This enables customers to leverage ecosystem-level explicit knowledge, integrate it with broader existing data, and effectively translate it into localized insights for more informed operational decisions augmented by AI. As the director from ProAIM noted: "*[The system] connects everything and serves as the core. It manages the entire lifecycle...this management system is not only a large industrial software system but also a management framework.*"

To improve the accuracy of their AI system, SUPCON offers the digital twin experimentation platform APEX, which facilitates virtual testing of AI performance for local implementation. Based on test results and practical requirements, iterative adjustments are made by AI engineers to harness AI learning capabilities and optimize performance. This is explained by the investment director from SUPCON: "*We have created a 3D simulation that accurately reproduces all the process equipment and pipelines of the site on a 1:1 scale. This allows us to display everything we investigate, approaching the concept of a digital twin that involves simulating the entire process of factory production from start to finish. Based on it, our customers can identify areas for improvement.*"

As a result, interactions revealed in this phase indicate that **Inclusiveness**, **Auditability** and **Accuracy** are key RAI dimensions. Inclusiveness is reflected in research on design principles for AI, which aims to reinforce equal access to AI-integrated applications. High compatibility in system design reduces access barriers and potentially increases inclusivity within the ecosystem. Customized knowledge with decomposed analysis also facilitates inclusivity for end users who are not IT specialists, thus enhancing transparency for real end-users. Second, AI functions as silent agents which can mislead managers in making critical operational decisions or result in the ethicality of AI being miscommunicated or omitted (Martin, 2019), hence it

underscores the need for professional auditability. Lastly, the accuracy of the AI system is enhanced through virtual experimentation, which improves model performance and validates new applications (Wang, 2021). This process leverages real production cycle data (Tao et al., 2024), providing customers with more reliable outcomes.

5. Discussion

5.1. Theoretical Contributions

Although our research is still in progress, our preliminary findings have already highlighted several important contributions. First, our study fills the research gaps in RAI by developing a framework (refers to Figure 2) derived from empirical evidence, addressing the current challenges in AI-powered responsible KM. To address the limitations identified in the explaining and controlling approaches, our findings underscore the collaborative configuration within the knowledge ecosystem actors. This fosters transparent, accurate and robust decision-makings, enhancing accountable value creation for organizational AI adopters (Someh et al., 2023; Sturm et al., 2021). As such, our study extends the current XAI approach studies from isolation in technical discourse (Bauer et al., 2023) into a socio-technical co-creation process for KM. Furthermore, our findings highlight the customization of KM as effective management that advances the controlling approaches, that aligns the organizational resources and business strategies to maximize the value creation enhanced by AI. Since value creation requires effective actions (Alavi et al., 2024), our findings indicate the additional need for knowledge decomposition and virtual experimentation to address data-misspecification-induced risks in AI implementation (Tao et al., 2024). The decomposed and verified AI-augmented knowledge, therefore, are more explainable and inclusive to end users, promoting greater control during the knowledge commercialization process, even for those lacking IT-technical skills. Additionally, our findings address reliability obstacle in RAI that are exacerbated by barriers in transferring and reusing domain knowledge (Lebovitz et al., 2021), thereby ensuring technically acceptable AI performance. We thus highlight the pivotal role of government in promoting standardization and compliance to foster robust knowledge accumulation at the industrial system level.

Secondly, our preliminary findings contribute to enhancing the conceptual clarity of RAI. Current studies often define its dimensions in a segmented manner and tend to adopt a laundry-listing approach to include every possible dimension. However, our findings indicate that RAI priorities vary depending on different KM goals and specific phases (see Figure 2). For example, transparency is a significant concern during the Knowledge Extraction phase, which aims to convert tacit knowledge and AI knowledge into explicit knowledge. Conversely, the ability to audit AI models and systems becomes more crucial when commercializing this extracted explicit knowledge, as it transitions back into tacit knowledge within customer organizations. This distinction highlights discrepancies in user perceptions regarding the need for explanations and contributes to the ongoing discourse on the desired format of explanations or interpretations in current XAI studies (Adadi & Berrada, 2018; Bauer et al., 2023). Therefore, our findings advocate for a nuanced understanding of AI-augmented organizational knowledge creation, demonstrating the need to align a firm's knowledge management activities with both the phases of KM and the associated goals of RAI.

Finally, our framework depicts KM ecosystem dynamics, highlighting the evolving interactions among actors that shape and redefine their goals and activities across difference KM phases. Current studies on KM often adopt static analyses, assuming independent and predefined roles of actors in knowledge creation and sharing (Alavi & Leidner, 2001). Furthermore, studies that consider dynamic processes often lack elaboration on the mechanisms driving these dynamics (Stahl, 2021). Our preliminary findings contribute to the missing discourse by identifying key interactions that drive KM activities in an interdependent manner, thus elucidating the structural dynamics within the KM ecosystem (Adner, 2017). For instance, the structure of the KM ecosystem shifts from a hub-and-spoke model centered around the focal-tenant in Knowledge Extraction to a more symbiotic model in Knowledge Standardization. This shift occurs due to the nest relationships derived from interdependent interactions to drive the creation of new standardized explicit knowledge. Such structural evolution illustrates how ecosystem actors adapt their activities to collaborate for RAI objectives, contributing to a dynamic understanding of how the KM ecosystem is organized and evolves collectively through actorhood (Järvi et al., 2018).

5.2. Practical Implications

Our study introduces empirical contributions to all identified actors within KM ecosystems, notably benefiting industry firm managers, government policymakers, educational institutions, and domain experts. We provide practical guidance on how actors can collaborate within the ecosystem by adapting their activities and identifying key partners for responsible KM. Our framework supports firm managers in devising effective AI investment strategies. For policymakers, our findings underscore the role of fostering knowledge-sharing and establishing standards to facilitate responsible knowledge co-creation. We encourage educational institutions and domain experts to engage in collaborative efforts with industry firms to ensure that responsible knowledge creation remains practical and actionable.

6. Conclusion and Future Research

This study has identified the responsible management for AI-augmented knowledge and key RAI dimensions associated with each KM phase. Our framework delineates KM dynamics and their corresponding RAI priorities driven by interactions among ecosystem actors. Moving forward, our future research will expand the interview data collection to encompass a broader spectrum of ecosystem actors, including customers, government, and educational institutions. With additional insights, we aim to further refine our framework to offer a comprehensive understanding of managing responsible knowledge generated by and co-created with AI.

7. References

Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, 52138–52160.

Adner, R. (2017). Ecosystem as Structure: An Actionable Construct for Strategy. *Journal of Management*, *43*(1), 39–58.

Alavi, M., & Leidner, D. E. (2001). Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. *MIS Quarterly, 25*(1), 107.

Alavi, M., Leidner, D. E., & Mousavi, R. (2024). Knowledge Management Perspective of Generative Artificial Intelligence. *Journal of the Association for Information Systems, 25*(1), 1–12.

Asatiani, A., Malo, P., Nagbøl, P. R., Penttinen, E., Rinta-Kahila, T., & Salovaara, A. (2021). Sociotechnical Envelopment of Artificial Intelligence: An Approach to Organizational Deployment of Inscrutable Artificial Intelligence Systems. *Journal of the Association for Information Systems*, *22*(2), 325–352.

Bauer, K., von Zahn, M., & Hinz, O. (2023). Expl(AI)ned: The Impact of Explainable Artificial Intelligence on Users' Information Processing. *Information Systems Research*, *34*(4), 1582–1602.

Benbasat, I. (1999). Empirical research in information systems: The practice of relevance. *MIS Quarterly, 23*(1), 3-16.

Benbya, H., Pachidi, S., & Jarvenpaa, S. L. (2021). Special Issue Editorial: Artificial Intelligence in Organizations: Implications for Information Systems Research. *Journal of the Association for Information Systems*, *22*(2), 281–303.

Benbya, H., Strich, F., & Tamm, T. (2024). Navigating Generative Artificial Intelligence Promises and Perils for Knowledge and Creative Work. *Journal of the Association for Information Systems*, *25*(1), 23–36.

Borges, A. F. S., Laurindo, F. J. B., Spínola, M. M., Gonçalves, R. F., & Mattos, C. A. (2021). The strategic use of artificial intelligence in the digital era: Systematic literature review and future research directions. *International Journal of Information Management, 57*, 102225.

Braithwaite, V. (2020). Beyond the bubble that is Robodebt: How governments that lose integrity threaten democracy. *Australian Journal of Social Issues*, *55*(3), 242–259.

Carlile, P. R. (2002). A Pragmatic View of Knowledge and Boundaries: Boundary Objects in New Product Development. *Organization Science, 13*(4), 442–455.

Clarysse, B., Wright, M., Bruneel, J., & Mahajan, A. (2014). Creating value in ecosystems: Crossing the chasm between knowledge and business ecosystems. *Research Policy, 43*(7), 1164–1176.

Cobben, D., Ooms, W., Roijakkers, N., & Radziwon, A. (2022). Ecosystem types: A systematic review on boundaries and goals. *Journal of Business Research, 142*, 138–164.

Dennehy, D., Griva, A., Pouloudi, N., Dwivedi, Y. K., Mäntymäki, M., & Pappas, I. O. (2023). Artificial Intelligence (AI) and Information Systems: Perspectives to Responsible AI. *Information Systems Frontiers, 25*(1), 1–7.

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., & Janssen, M. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management, 57*, 101994.

Dougherty, D., & Dunne, D. D. (2011). Organizing Ecologies of Complex Innovation. *Organization Science, 22*(5), 1214–1223.

Eisenhardt, K. M., and Graebner, M. E. (2007). Theory Building From Cases: Opportunities And Challenges. *Academy of Management Journal, 50*(1), 25–32.

Fernández-Loría, C., Provost, F., & Han. X. (2022). Explaining Data-Driven Decisions Made by Ai

Systems: The Counterfactual Approach. *MIS Quarterly*, *46*(3), 1635–1660.
Gephart, R. P. (2004). Qualitative Research and the Academy of Management Journal. *Academy of Management Journal, 47*(4), 454–462.
Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking Qualitative Rigor in Inductive Research: Notes on the Gioia Methodology. *Organizational Research Methods, 16*(1), 15–31.
Grønsund, T., & Aanestad, M. (2020). Augmenting the algorithm: Emerging human-in-the-loop work configurations. *Journal of Strategic Information Systems*, *29*(2).
Gold, A. H., Malhotra, A., & Segars, A. H. (2001). Knowledge Management: An Organizational Capabilities Perspective. *Journal of Management Information Systems, 18*(1), 185–214.
Herm, L.V., Heinrich, K., Wanner, J., & Janiesch, C. (2023). Stop ordering machine learning algorithms by their explainability! A user-centered investigation of performance and explainability. *International Journal of Information Management*, *69*, 102538.
Huang, P., Tafti, A., & Mithas, S. (2018). Platform Sponsor Investments and User Contributions in Knowledge Communities: The Role of Knowledge Seeding. *MIS Quarterly, 42*(1), 213–240.
Langley, A. (1991). Strategies for Theorizing from Process Data. *Academy of Management Review, 24*(4), 691-710.
Le, T.T.H., Prihatno, A. T., Oktian, Y. E., Kang, H., & Kim, H. (2023). Exploring Local Explanation of Practical Industrial AI Applications: A Systematic Literature Review. *Applied Sciences, 13*(9), 5809.
Lebovitz, S., Levina, N., & Lifshitz-Assa, H. (2021). Is AI Ground Truth Really True? The Dangers of Training and Evaluating AI Tools Based on Experts' Know-What. *MIS Quarterly, 45*(3), 1501–1526.
Lee, H., & Choi, B. (2003). Knowledge Management Enablers, Processes, and Organizational Performance: An Integrative View and Empirical Examination. *Journal of Management Information Systems, 20*(1), 179–228.
Leidner, D. (2018). Review and Theory Symbiosis: An Introspective Retrospective. *Journal of the Association for Information Systems*, *19*(06), 552–567.
Levina & Vaast. (2005). The Emergence of Boundary Spanning Competence in Practice: Implications for Implementation and Use of Information Systems. *MIS Quarterly, 29*(2), 335.
Jacobides, M. G., Cennamo, C., & Gawer, A. (2018). Towards a theory of ecosystems. *Strategic Management Journal, 39*(8), 2255–2276.
Jan, Z., Ahamed, F., Mayer, W., Patel, N., Grossmann, G., Stumptner, M., & Kuusk, A. (2023). Artificial intelligence for industry 4.0: Systematic review of applications, challenges, and opportunities. *Expert Systems with Applications, 216*, 119456.
Järvi, K., Almpanopoulou, A., & Ritala, P. (2018). Organization of knowledge ecosystems: Prefigurative and partial forms. *Research Policy, 47*(8), 1523–1537.
Martin, K. (2019). Ethical Implications and Accountability of Algorithms. *Journal of Business Ethics*, *160*(4), 835–850.
Mikalef, P., Conboy, K., Lundström, J. E., & Popovič, A. (2022). Thinking responsibly about responsible AI and 'the dark side' of AI. *European Journal of Information Systems, 31*(3), 257–268.
Nonaka, I. (1994). A Dynamic Theory of Organizational Knowledge Creation. *Organizational Science, 5*(1), 14-37.
Pan, S. L., & Tan, B. (2011). Demystifying case research: A structured-pragmatic-situational (SPS) approach to conducting case studies. *Information and Organization*, *21*, 161-176.
Rinta-Kahila, T., Someh, I., Gillespie, N., Indulska, M., & Gregor, S. (2022). Algorithmic decision-making and system destructiveness: A case of automatic debt recovery. *European Journal of Information Systems*, *31*(3), 313–338.
Saeed, W., & Omlin, C. (2023). Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities. *Knowledge-Based Systems, 263*, 110273.
Schwartz, D., & Te'eni, D. (2024). AI for Knowledge Creation, Curation, and Consumption in Context. *Journal of the Association for Information Systems*, *25*(1), 37–47.
Someh, I., Wixom, B., Davern, M., & Shanks, G. (2023). Configuring Relationships between Analytics and Business Domain Groups for Knowledge Integration. *Journal of the Association for Information Systems*, *24*(2), 592–618.
Strauss, A., & Corbin, J. (1998). Basics of qualitative research: Techniques and procedures for developing grounded theory (2nd ed.). Thousand Oaks, CA: Sage.
Stahl, B. C. (2022). Responsible innovation ecosystems: Ethical implications of the application of the ecosystem concept to artificial intelligence. *International Journal of Information Management*, *62*, 102441.
Tao, F., Zhang, H., & Zhang, C. (2024). Advancements and challenges of digital twins in industry. *Nature Computational Science, 4*(3), 169–177.
Tanriverdi. (2005). Information Technology Relatedness, Knowledge Management Capability, and Performance of Multibusiness Firms. *MIS Quarterly, 29*(2), 311.
Teodorescu, M., Morse, L., Awwad, Y., & Kane, G. (2021). Failures of Fairness in Automation Require a Deeper Understanding of Human-ML Augmentation. *MIS Quarterly*, *45*(3), 1483–1500.
Van Den Broek, E., Sergeeva, A., & Huysman Vrije, M. (2021). When the Machine Meets the Expert: An Ethnography of Developing AI for Hiring. *MIS Quarterly, 45*(3), 1557–1580.
Walsham, G. (1995). Interpretive case studies in IS research: Nature and method. *European Journal of Information Systems, 4*, 74–81.
Wang, P. (2021). Connecting the Parts with the Whole: Toward an Information Ecology Theory of Digital Innovation Ecosystems. *MIS Quarterly, 45*(1), 397–422.