Innovation Series https://www.innoseries.com

Navigating Ambiguity, Strengthening Trust: Semantic Clarity and Relational Dynamics in Industry–University Knowledge Co-Creation

Qingyu Xu[1], Jian Ma[1,*], Zhouqiang Qiu[2]

1 Department of Information System, City university of Hong Kong, China

2 Guangdong Science and Technology Innovation Monitoring and Research Center, Guangzhou, China

Corresponding Author: Jian Ma

Abstract: Effective university–industry collaboration remains challenging, primarily because companies, particularly small and medium-sized enterprises (SMEs), often struggle to clearly articulate their multi-dimensional technological requirements and accurately identify trustworthy academic experts aligned with these needs. To address these critical issues, we introduce DualRAG-SNR, a novel hierarchical matching framework explicitly designed to resolve semantic ambiguities and integrate trust signals. DualRAG-SNR incorporates (i) a dual-stage retrieval-augmented generation (DualRAG) mechanism, comprising a typed Graph-based retrieval (GraphRAG) followed by a precise Vector Database (VDB) retrieval, explicitly clarifying multi-aspect corporate requirements and retrieving semantically coherent knowledge; and (ii) an enriched Social Network Ranker (SNR) explicitly constructed from citation data, institutional affiliations, and fine-grained company–scholar interactions logged on the Online Technology Trading Platform (OTTP), capturing exploratory interactions such as profile views, communication exchanges, and contractual activities. RotatE embeddings explicitly model relational trust within this enriched social network. We further aggregate retrieved insights into a unified hypothetical requirement document (HyDE) using a large language model (LLM), explicitly enhancing semantic clarity. Through requirement-aware attention-based fusion, DualRAG-SNR dynamically balances semantic relevance and relational trust, significantly improving scholar recommendations. Empirical evaluation on real-world OTTP collaboration cases demonstrates that DualRAG-SNR achieves superior recall and nDCG compared to strong baselines. Furthermore, an interaction-based evaluation explicitly indicates that DualRAG-SNR recommendations elicit deeper and more sustained company–scholar interactions, explicitly signaling enhanced trust-building and more effective knowledge transfer. Ablation studies explicitly confirm that the hierarchical DualRAG retrieval, enriched exploratory interactions, and requirement-aware fusion contribute substantial complementary improvements. The proposed framework explicitly provides both theoretical insights and practical tools for systematically reducing information asymmetry and fostering robust knowledge transfer in university–industry partnerships.

Keywords: Retrieval-Augmented Generation, Social Network, University–Industry Collaboration, Knowledge Transfer, Text Matching

1. Introduction

In the paradigm of Open Innovation, the porosity of organizational boundaries has necessitated a shift from internal R&D to external knowledge acquisition. For Small and Medium-sized Enterprises (SMEs), University-Industry Collaboration (UIC) represents a critical conduit for assimilating cutting-edge technologies and sustaining competitive advantage [2]. Conversely, for academic institutions, the commercialization of research is increasingly mandated as a mechanism to demonstrate societal impact. Despite these reciprocal incentives, the "market for technology" remains highly inefficient. A pervasive "Valley of Death" separates theoretical innovation from industrial application, characterized by high search costs and low success rates in partner identification [3].

The impediments to effective UIC are rooted in two fundamental dimensions: cognitive heterogeneity and relational opacity. First, the cognitive gap arises from the linguistic mismatch between the two communities [4]. Academic discourse is structured around theoretical rigor and domain-specific taxonomies, whereas industrial requirements are often articulated through problem-oriented, colloquial, or ambiguous descriptions. Traditional keyword-based Information Retrieval (IR) systems fail to bridge this semantic chasm, as they lack the inferential capability to map a vague industrial "need" to a precise academic "solution." Second, the relational gap stems from the severe information asymmetry regarding scholar reliability. Unlike commodity transactions, knowledge transfer is deeply embedded in social trust [17]. SMEs lack verifiable signals to assess a scholar's practical capability (ex-ante trust), and scholars lack mechanisms to signal their industrial relevance. Without a robust trust-modeling mechanism, the transaction costs associated with initiating collaboration remain prohibitively high.

Recent advancements in Generative Artificial Intelligence (GenAI), particularly Large Language Models (LLMs), offer a nascent opportunity to act as intelligent intermediaries. However, the direct application of off-the-shelf LLMs to the UIC context is fraught with challenges. Standard Retrieval-Augmented Generation (RAG) approaches often retrieve fragmented information, failing to capture the structural logic (e.g., the applicability of a specific method to a specific scenario) inherent in scientific literature [10]. Furthermore, LLMs are prone to "hallucinations," generating plausible but factually incorrect recommendations, which is detrimental in high-stakes B2B decision-making [9]. More critically, existing recommender systems predominantly focus on relevance (content matching) while neglecting reputation (trust modeling), thereby failing to address the "cold start" problem inherent in forming new social ties.

Recent advancements in Large Language Model (LLM)-based agents have led to the proliferation of personalized AI assistants, which greatly facilitate communication by accessing extensive personal data and engaging in interactions that substantially reduce communication overhead and enhance trust on digital platforms [8]. Similar pre-collaboration communication approaches have emerged in various professional domains to alleviate information asymmetry. For instance, in medical contexts, personalized AI assistants provide preliminary patient information and histories before formal doctor consultations, establishing initial trust and streamlining the medical consultation process [5]. However, despite evident success in other professional domains, such pre-collaboration AI mechanisms remain underdeveloped within university-industry collaborations. This gap persists primarily because (1) SMEs often articulate their technical demands in a vague, multi-faceted, and coarse-grained manner, making conventional keyword-based retrieval insufficient due to ambiguous or overly broad results from LLMs; (2) traditional academic outputs (papers or patents alone) inadequately reflect practical applicability or contextual details required by industry; and (3) direct trust-building remains difficult, given limited historical interactions and academic researchers' time constraints.

To directly address these unresolved challenges, we propose a novel intelligent matching framework—DualRAG-SNR—comprising two complementary modules designed explicitly to enhance matching accuracy, clarity of requirements, trust-building, and overall knowledge transfer efficiency. To effectively interpret companies' multi-aspect and often ambiguous requirements, our DualRAG retrieval mechanism operates in two distinct yet complementary stages. Initially, a typed GraphRAG approach systematically aligns entities across separate dimensions—Technology, Method, and Application—allowing precise extraction and integration of cross-document information [16]. By explicitly encoding these dimensions, GraphRAG enables targeted retrieval and coherent contextualization of information distributed across multiple documents, significantly reducing ambiguity and ensuring comprehensive semantic coverage. Subsequently, the Vector DataBase (VDB)-based retrieval stage independently retrieves relevant, detailed textual evidence for each decomposed aspect without sacrificing broader contextual information. This separation ensures that each single-aspect query benefits from highly specific knowledge while maintaining consistency with the overall multi-dimensional context. Collectively, this hierarchical retrieval process substantially clarifies complex industry requirements, reduces information fragmentation, and ultimately enhances the interpretability and precision of scholar recommendations.

Nevertheless, addressing information clarity alone is insufficient without considering relational trust and credibility. Existing literature frequently neglects explicit modeling of trust derived from real-world interactions, thus limiting the practical effectiveness of scholar recommendations. To overcome this limitation, we incorporate a sophisticated Social-Network Ranker (SNR). Specifically, SNR constructs an enriched heterogeneous social network, integrating historical human–digital scholar interactions—such as profile views, communication exchanges, and contract activities recorded by the platform—to explicitly quantify evolving relational trust. Additionally, SNR includes scholars' bibliometric quality indicators (e.g., citations, h-index) and institutional rankings to reflect academic reputation and reliability objectively. By embedding this enriched social graph using relational embedding techniques (RotatE) [23], SNR effectively incorporates connection strength, academic quality, and historical trust signals, significantly reducing information asymmetry related to credibility.

To empirically validate the DualRAG-SNR framework, we utilize a comprehensive real-world dataset from the Jiangxi Online Technology Trading Platform (OTTP), encompassing 1,700 technology transfer interactions between SMEs and academic scholars. Our evaluation consists of both offline and online analyses. Specifically, offline experiments benchmark DualRAG-SNR against leading baselines across critical metrics such as recall, normalized Discounted Cumulative Gain (nDCG), and Mean Reciprocal Rank (MRR). Further, leveraging detailed historical company–digital scholar interaction logs (such as communication frequency, depth, and speed of knowledge transfer), we introduce innovative interaction-based metrics to quantify recommendation efficiency and knowledge transfer efficacy. These interaction-based evaluations uniquely highlight our framework's capacity to streamline communication, enhance mutual trust, and expedite effective collaboration, demonstrating measurable superiority compared to traditional recommendation methodologies.

Our study makes several critical contributions: (1) We propose a novel Digital Scholar Agent-based recommendation framework explicitly addressing information asymmetry and trust-building issues in academia-industry collaborations. (2) Methodologically, we introduce an innovative DualRAG method that accurately extracts and integrates multimodal knowledge, overcoming traditional keyword extraction limitations and hallucination problems associated with LLM-based methods. (3) Practically, our integrated social network-enhanced recommendation system DualRAG-SNR significantly improves SMEs' ability to identify suitable academic collaborators, fostering effective knowledge transfer and commercialization of academic research outcomes.

2 LITERATURE REVIEW

2.1 Automatic Text Matching

Text matching, a fundamental task in natural language processing (NLP), aims at accurately evaluating semantic and lexical similarity between textual elements for various applications, including information retrieval, recommendation systems, and question answering [25, 27]. Traditional text matching techniques heavily depended on lexical similarity, leveraging statistical approaches like TF-IDF, cosine similarity, and probabilistic models such as BM25 [7, 15]. While these early approaches effectively captured superficial term frequencies, they struggled significantly with deeper semantic nuances and contextual variations, leading to mismatches, especially when matching texts from differing semantic contexts.

The advent of semantic embedding approaches enabled text comparisons based on deeper semantic similarity rather than mere lexical overlap. Techniques such as Subject-Action-Object (SAO) and Function-Object-Property (FOP) further enhanced semantic understanding by explicitly extracting structured semantic components [24]. Deep learning models, including convolutional neural networks (CNN), Long Short-Term Memory networks (LSTM), and transformer-based models like BERT, refined this capability further by capturing rich contextual and semantic relationships within text pairs [14, 19].

Integrating external knowledge from knowledge graphs (KGs) significantly enhanced text matching performance by embedding explicit semantic entities and relationships [22, 25]. Huang et al. (2020), for instance, developed a knowledge-enhanced attention network that leveraged structured knowledge from KGs to improve semantic accuracy in text matching tasks. Recent advances in Retrieval-Augmented Generation (RAG) techniques have further improved semantic matching capabilities by explicitly retrieving relevant external knowledge before generating responses. Models like STMAP introduced embedding augmentation to enhance robustness and accuracy in semantic matching tasks [26].

Despite these advances, current methods typically presume a shared semantic space, inadequately addressing the distinct semantic contexts common in specialized applications such as academia-industry collaborations [13]. The cognitive gap between industry requirements, often broadly articulated in colloquial business terms, and the highly specialized academic discourse necessitates methods specifically designed to bridge these distinct semantic domains.

To explicitly address this critical gap, our research introduces the DualRAG-SNR framework. Unlike traditional approaches, DualRAG-SNR employs a hierarchical retrieval mechanism combining typed GraphRAG and a VDB approach. By retrieving context from structured knowledge graphs, our approach significantly reduces semantic mismatches, providing companies with intuitive, accurate, and easily understandable domain-specific knowledge. Further integrating these refined semantic representations with social network data, our framework ensures precise recommendations of digital scholar agents to SMEs, effectively bridging semantic discrepancies and fostering meaningful academia-industry collaborations.

2.2 Trust-building in Academia-Industry Collaboration

Establishing trust between academia and industry has traditionally relied on several mechanisms [20]. Prior collaborative experiences are fundamental; companies with a history of successful collaboration tend to build stronger inter-organizational trust and benefit more from subsequent partnerships. Studies indicate that past collaborations significantly reduce cultural and organizational barriers, thereby facilitating future engagements and improving outcomes such as innovation performance [6]. The presence of "champions" or boundary spanners—individuals who actively bridge academic and industry sectors—also plays a crucial role in aligning mutual expectations, enhancing communication, and fostering a trustworthy collaboration environment. Furthermore, relational social capital, characterized by repeated interactions, personal ties, and goodwill, forms a solid foundation for effective knowledge sharing and trust development. Research emphasizes the importance of shared goals, mutual understanding, and cultural alignment to overcome differences in expectations and operational norms between academia and industry [1].

Social network-based recommendation systems have capitalized on relational insights to facilitate academia-industry connections [11]. Leveraging existing interpersonal relationships and collaboration histories, these systems improve matching accuracy and enhance trust by identifying experts who are socially proximate or have prior engagement with potential industry partners. However, such approaches often overlook the fundamental semantic gap between market-driven business requirements and specialized academic discourse, potentially leading to suboptimal matches.

In recent years, the emergence of AI-driven digital agents has transformed knowledge transfer and communication processes in various professional contexts. Digital agents, such as chatbots and personalized virtual assistants, proactively reduce information asymmetry by transparently explaining their recommendations and disclosing relevant information. Empirical studies demonstrate that transparency and clarity in agent communication significantly mitigate user uncertainty, enhancing trust and acceptance. For instance, digital agents in e-commerce settings have effectively informed customers about detailed product features, fees, and conditions, which customers might not proactively inquire about, thus aligning expectations and reducing informational gaps [8].

However, such digital agent mechanisms have not yet been widely explored within academia-industry collaboration scenarios, despite their potential advantages. Our novel DualRAG-SNR framework utilizes structured knowledge graphs to precisely interpret and bridge distinct semantic contexts, enhancing the accuracy of keyword extraction and semantic understanding. By explicitly incorporating transparency and intuitive explanations of recommendations, our digital scholar agent-based approach significantly reduces the complexity and cost associated with human-mediated communication. This advancement not only bridges the existing semantic and informational gaps but also fosters a foundation of trust crucial for successful long-term collaboration. Ultimately, our research represents the first attempt to systematically leverage digital scholar agents combined with social networks and structured semantic matching to facilitate more informed, trust-enhancing, and efficient academia-industry partnerships.

3 METHODOLOGY

This section describes our end-to-end pipeline, beginning with the construction of a GraphRAG and VDB from scholarly publications and historical patents, and culminating in a requirement-aware matching algorithm that ranks Digital Scholar Agents (DSAs) for each corporate query.

3.1 Problem Definition and Framework Overview

Our primary goal is to facilitate effective matching between broadly articulated corporate technical requirements and specialized academic researchers (Digital Scholars). Specifically, given a company's textual requirement q, we aim to accurately retrieve and recommend the most relevant Digital Scholars from an extensive candidate set S, by bridging the inherent semantic gap and informational asymmetry between industry and academia.

To address this challenge, we introduce DualRAG-SNR (Dual Retrieval-Augmented Generation and Social-Network Ranker), a novel hierarchical recommendation framework. DualRAG-SNR explicitly targets two critical limitations of existing methods: (1) traditional retrieval methods inadequately handle multi-aspect and ambiguous company requirements, resulting in semantic mismatches; and (2) conventional approaches fail to explicitly incorporate relational trust and credibility signals derived from historical interactions between companies and scholars.

The proposed framework comprises two core modules:

- DualRAG: a hierarchical retrieval mechanism designed to address the complexity and ambiguity of corporate requirements through two sequential retrieval stages: a typed GraphRAG stage followed by a VDB retrieval stage. DualRAG systematically extracts structured and semantically precise knowledge, accurately decomposing complex, multi-aspect demands into well-defined, single-aspect sub-queries.
- Social-Network Ranker (SNR): an enriched social network module that explicitly incorporates trust and quality signals through embedding historical company-scholar interactions, bibliometric scholar quality metrics, and institutional rankings. This module ensures recommendation credibility by effectively capturing relational trust, quality, and connection.

We show our framework in Figure 1. Through integrating DualRAG and SNR modules, the proposed DualRAG-SNR framework effectively bridges semantic and trust-related gaps, enhancing recommendation precision, interpretability, and trustworthiness.

3.2 DualRAG

$$\pi_q(v) = \frac{\mathbf{z}_q \cdot \mathbf{z}_v}{\left|\left|\mathbf{z}_q\right|\right|, \left|\left|\mathbf{z}_v\right|\right|} \tag{1}$$

$\boldsymbol{d}_u$ denotes node degree normalization, and $w_{uv}^{(\text{type})}$ is calculated as:

$$w_{uv}^{(\text{type})} = \begin{cases} 1, \text{if type}(u) = \text{type}(v) \\ \gamma, \ \ \text{if type}(u) \neq type(v) \end{cases} \tag{2}$$

The parameter γ explicitly penalizes semantic coherence scores for cross-type edges, thus encouraging propagation within the same semantic dimension, thereby enhancing interpretability and semantic precision. We retain the top-ranked entities and their associated contextually-rich descriptions, thereby effectively clarifying complex multi-dimensional corporate requirements, substantially improving semantic interpretability and precision for subsequent recommendation steps.

3.2.1 Multi-Aspect Requirement Decomposition

Corporate technical requirements often implicitly combine multiple distinct aspects—such as desired technologies, methodologies, or application scenarios—in a single textual description, hindering accurate semantic matching. To systematically address this ambiguity, we introduce a Multi-Aspect Requirement Decomposition process. This process leverages the structured outputs retrieved from GraphRAG and employs STIGPT to explicitly decompose each original corporate requirement into clearly defined, single-aspect sub-queries.

Specifically, we concatenate the original requirement text q and the top-ranked entities and relations extracted via GraphRAG into a structured prompt designed explicitly to guide the LLM towards clear multi-aspect decomposition. The structured prompt used is as follows:

Prompt 1.
Given the following corporate requirement text and associated entities extracted through GraphRAG retrieval: Requirement Text: {original requirement text} Retrieved Entities and Relationships: - Technology: {list of relevant technology entities} - Method: {list of relevant method entities} - Application: {list of relevant application entities} Please carefully analyze the requirement and decompose it into distinct sub-requirements, each clearly addressing only one semantic aspect (Technology, Method, or Application). Present your decomposition as a numbered list, clearly indicating the aspect each sub-requirement corresponds to. Example format: 1. [Technology]: description of the technology-related aspect of the requirement. 2. [Method]: description of the method-related aspect of the requirement. 3. [Application]: description of the application-related aspect of the requirement.

The LLM generates N decomposed single-aspect queries $\{q^{(1)}, q^{(2)}, \ldots, q^{(N)}\}$, explicitly marked by their semantic types.

3.2.2 VDB Construction and Retrieval

3.2.2.1 VDB Construction

Following GraphRAG retrieval, we construct a robust VDB using content extracted from scholarly articles, conference papers, and patents. The construction process explicitly follows three main steps to ensure semantic precision and efficient retrieval:

1. Document Aggregation and Chunking: First, we collect a corpus comprising (i) full texts of peer-reviewed journal articles, preprints, and conference proceedings authored by candidate scholars, and (ii) detailed patent documents previously filed by companies on the OTTP. Each collected document is segmented into overlapping textual chunks of approximately 512 tokens each, consistent with typical LLM embedding contexts. Overlapping chunks ensure contextual continuity, allowing accurate semantic representation without losing essential information at chunk boundaries.

2. Semantic Embedding of Textual Chunks: Subsequently, each 512-token textual chunk undergoes semantic embedding using Chinese-BERT, selected for its strong performance in domain-specific language tasks. Specifically, for each textual chunk $\boldsymbol{c}_j$, we generate a corresponding embedding vector $\boldsymbol{z}_{c_j}$. These embedding vectors comprehensively encode the semantic content of individual chunks, capturing precise and rich contextual information.
3. Vector Database Indexing and Storage: All embedding vectors $\{\boldsymbol{z}_{c_j}\}$ are efficiently stored and indexed using the FAISS library, optimized for high-dimensional similarity retrieval tasks. FAISS ensures rapid and accurate retrieval, supporting scalable similarity searches across an extensive academic-industrial knowledge corpus.

The resulting VDB thus explicitly maintains a high-precision semantic representation of scholarly and patent-derived knowledge, enabling accurate and context-rich semantic retrieval in subsequent recommendation steps.

3.2.2.2 Aspect-Specific VDB Retrieval

Following Multi-Aspect Requirement Decomposition (Section 3.2.2), each decomposed single-aspect query $q^{(i)}$ explicitly corresponds to a distinct semantic dimension (Technology, Method, or Application) of the original requirement. Each single-aspect query is independently encoded into an embedding vector $\boldsymbol{z}_{q^{(i)}}$ using Chinese-BERT. For each aspect-specific embedding $\boldsymbol{z}_{q^{(i)}}$, we perform an independent retrieval query against our constructed Vector Database. We calculate the semantic similarity score between the single-aspect query embedding $\boldsymbol{z}_{q^{(i)}}$ and each stored textual chunk embedding $\boldsymbol{z}_{c_j}$ as follows:

$$\text{Similarity}\left(q^{(i)}, c_j\right) = cosine\_similarity\left(z_{q^{(i)}}, z_{c_j}\right) \tag{3}$$

We then retrieve the top-ranked 10 textual chunks $\{c_j^{(1)}, c_j^{(2)}, \dots, c_j^{(10)}\}$ for each single-aspect query based on their semantic similarity scores. Retrieving exactly 10 textual chunks per aspect ensures balanced semantic coverage, maintaining precision while avoiding information overload.

This hierarchical retrieval approach—structured first by GraphRAG and refined via VDB—systematically ensures precise, detailed, and semantically coherent knowledge matching for each distinct semantic aspect of corporate technical requirements, significantly enhancing subsequent recommendation accuracy and interpretability.

3.2.3 HyDE Construction

To further resolve ambiguities inherent in the original corporate requirement texts and to comprehensively integrate the fine-grained semantic information retrieved from the previous VDB step, we employ a structured HyDE construction process. HyDE generates a unified and semantically coherent pseudo-requirement document that synthesizes detailed insights derived from the individual single-aspect retrieval queries.

For each decomposed single-aspect query $q^{(i)}$, along with its 10 retrieved textual chunks obtained from the VDB, we employ STIGPT as a judgement generator. Specifically, the LLM processes each aspect separately with the following structured prompt:

Prompt 2.
Given the following single-aspect query extracted from a broader corporate requirement and a

collection of highly relevant textual snippets retrieved from scholarly articles and patents, synthesize a concise, precise, and informative summary. This summary should clearly reflect the key technologies, methodologies, or applications that match the corporate requirement.

Single-Aspect Query:
{insert the single-aspect query}

Relevant Textual Snippets:
1. {textual chunk 1}
2. {textual chunk 2}
...
10. {textual chunk 10}

Please generate a clear and concise summary (about 100-150 words) explicitly highlighting what technological solutions, methods, or applications the company likely requires and what characteristics an ideal academic expert should possess.

The output from the LLM for each single-aspect query $q^{(i)}$ is a concise, structured judgement summary $J^{(i)}$. Each $J^{(i)}$ explicitly and succinctly captures critical semantic insights related to that specific aspect, including technical nuances, methodological preferences, and practical applications relevant to the original requirement. Next, we systematically aggregate the generated judgement summaries $\{J^{(1)}, J^{(2)}, \dots, J^{(N)}\}$ into a single unified hypothetical requirement document $\widetilde{q_{\text{agg}}}$ utilizing STIGPT. We employ the following structured prompt explicitly designed to guide coherent aggregation:

Prompt 3.
Below are several distinct judgement summaries, each addressing a single aspect (Technology, Method, or Application) of a broader corporate technical requirement. Your task is to aggregate these summaries into a unified, coherent, and informative hypothetical requirement document. This final document should clearly and succinctly integrate all critical aspects into a single, comprehensive narrative. Maintain clarity and logical consistency throughout. Aspect-specific Judgement Summaries: 1. {Judgement Summary for aspect 1} 2. {Judgement Summary for aspect 2} ... N. {Judgement Summary for aspect N} Please synthesize these summaries into a single cohesive hypothetical requirement document (approximately 150-200 words), suitable for accurately matching relevant scholarly expertise.

The resulting unified document $\widetilde{q_{\text{agg}}}$ is thus generated as a coherent, contextually rich HyDE.

3.3 Multimodal Embedding Integration via Attention Distillation

To ensure comprehensive semantic alignment and leverage both the nuanced context from the original corporate requirement and the detailed insights encoded in the HyDE-generated hypothetical requirement document, we perform an embedding integration step employing attention-based fusion.

We first independently embed the original corporate requirement text q and the aggregated HyDE-generated hypothetical requirement document $\widetilde{q_{\text{agg}}}$, using the Chinese-BERT embedding model. Specifically, we obtain two separate embeddings, original requirement embedding $\boldsymbol{z}_q$ and HyDE embedding $\boldsymbol{z}_{\widetilde{q_{\text{agg}}}}$. These embeddings explicitly capture the semantic richness and contextual detail inherent within each textual input, encoding both high-level intent from the original requirement and detailed, decomposed insights from the HyDE-generated document.

To effectively integrate these two embeddings, we adopt a cross-attention fusion mechanism that explicitly models the semantic interplay between the original requirement and the detailed hypothetical requirement. This attention-based fusion explicitly enables mutual semantic refinement, capturing detailed and nuanced relationships that might be missed by independent embedding representations.

$$c_q = \text{Attention}(Q = z_q, KV = z_{\widetilde{q_{\text{ag}}}}),\ \widetilde{c_q} = \text{Attention}(Q = z_{\widetilde{q_{\text{ag}}}}, KV = \boldsymbol{z_q}) \tag{4}$$

Where c_q represents the refined semantic context of the original requirement text, explicitly informed by the detailed knowledge encapsulated within the HyDE embedding. $\widetilde{c_q}$ captures the refined detailed semantic information of the HyDE-generated hypothetical requirement document, explicitly aligned with the core intent encoded within the original requirement text.

Finally, we concatenate the two attention-refined embeddings into a single, integrated semantic representation $\boldsymbol{r_q} = [c_q \mid \widetilde{c_q}]$. This unified semantic representation $\boldsymbol{r_q}$ systematically captures and integrates both high-level and detailed semantic information from the original and HyDE-generated requirement texts. Consequently, it provides a precise, semantically rich input representation, significantly enhancing the accuracy and effectiveness of subsequent Social-Network Ranker (SNR) recommendation and matching tasks within our DualRAG-SNR framework.

3.4 Social-Network Feature Extraction

To accurately incorporate relational trust and academic quality into our recommendation process, we construct a heterogeneous social network explicitly integrating historical interactions among companies, digital scholars, and academic institutions.

Our SN construction systematically captures diverse relationships and fine-grained interaction signals documented on the OTTP platform. For each company C, scholar S and institution U we prompt STIGPT with a chain-of-thought prompt ("Identify the main technology domains, application scenarios and collaboration history; think step-by-step before answering") and parse the response into categorical attributes. Nodes and typed edges are organized as a heterogeneous graph G_{SN}. Previous scholarly recommender graphs have considered only scholar-scholar or scholar-institution ties. Our context is different: OTTP logs fine-grained company–scholar micro-interactions—for example, page views, chat requests, sample downloads, and signed non-disclosure agreements. We formalize each logged event as a directed, timestamped edge $\left(C \xrightarrow{\text{event}, w, \tau} S\right)$, where w is a log-frequency weight (e.g., count of message exchanges) and τ is the most recent interaction time. Edges are additionally labelled as positive (e.g., "contract signed") or exploratory (e.g., "interacted"). Including such company–scholar (C–S) edges yields a heterogeneous graph

$G_{SN} = (V_{SN}, E_{SN}),\ V_{SN} = C \cup S \cup U,\ E_{SN} = \{E_{SS}, E_{CC}, E_{UU}, E_{CS}, E_{CU}, E_{SU}\}$ (2)

We train RotatE to embed the graph, optimizing the translational distance. Because company–scholar edges encode historical information-asymmetry reduction, scholars that have interacted frequently with the querying company end up closer in the complex space, thereby raising their trust score when the embeddings are later fused with technical relevance. For the target company and the digital scholars, we get the final embeddings as $\boldsymbol{h_C}$ and $\boldsymbol{h_S}$.

3.5 Extraction Requirement-Aware Fusion

The concatenated context vector c_q primarily encodes technical relevance, whereas the RotatE embeddings $\boldsymbol{h_C}$, $\boldsymbol{h_S}$ capture relational trust. Simply averaging the two would ignore the fact that different queries require different balances—an SME entering a green-energy field for the first time places a premium on trust, whereas an experienced aerospace supplier may value cutting-edge novelty. To address this heterogeneity, we frame fusion as a query-conditioned gating problem. Specifically, we compute

$\omega_q = \sigma(W_\omega z_q) \in (0,1)$, (3)

Where σ is the logistic function. We then let

$\widetilde{\boldsymbol{h}_S} = \omega_q \boldsymbol{h_S} + (1 - \omega_q)\,\mathrm{Attn}(z_q, \boldsymbol{h_S}, \boldsymbol{h_S}),\ \widetilde{\boldsymbol{h}_C} = \omega_q \boldsymbol{h_C} + (1 - \omega_q)\,\mathrm{Attn}(z_q, \boldsymbol{h_C}, \boldsymbol{h_C})$, (4)

So that when ω_q is large the system leans on *historical trust*, and when it is small it prefers *semantic attention* that re-weights the SN embeddings in light of the current requirement. The final fused vectors remain $\boldsymbol{g_S} = [\boldsymbol{c_q} \,||\, \widetilde{\boldsymbol{h}_S}]$, $\boldsymbol{g_C} = [\boldsymbol{c_q} \,||\, \widetilde{\boldsymbol{h}_C}]$, where $[\cdot \,||\, \cdot]$ denotes vector concatenation. Attention-based fusion allows the system to weigh technical and social signals dynamically for each incoming query rather than relying on a fixed linear combination.

3.6 Learning-to-Rank Matcher

A pure cosine similarity under-fits the complexity of university–industry alignment because it assumes linear separability and equal contribution of each latent dimension. We therefore replace it with a pairwise neural ranker that is trained on implicit feedback from the OTTP logs.

3.6.1 Scoring Function

For any pair (C,S) we construct a feature vector

$x_{CS} = [\,g_C\,; g_S\,; |g_C - g_S|\,; g_C \odot g_S]$ (5)

Where $|\cdot|$ denotes element-wise absolute value and $\odot$ is the Hadamard (element-wise) product.

The final matching score is produced by a two-layer multilayer perceptron:

$$f(C,S) = W_2\, \phi(W_1\, x_{CS} + b_1) + b_2 \quad (6)$$

With ϕ the GELU activation.

3.6.2 Pairwise Loss

The OTTP records sequences in which a company inspects multiple scholars but eventually "locks" on one (e.g., by requesting detailed pricing). We form training triples $(\mathbf{C}, \mathbf{S}^+, \mathbf{S}^-)$ and minimize the Bayesian personalized ranking (BPR) loss

$\mathcal{L}_{\mathcal{BPR}} = -\sum_{(C,S^+,S^-)} \log \sigma\big(f(C,S^+) - f(C,S^-)\big)$ (7)

which pushes the chosen scholar S^+ above non-chosen scholars S^- in the ranking. At inference time, scholars are ordered by $f(C,S)$.

4. Experiments and Results

4.1 Experimental Settings

Task definition. We cast the university–industry matchmaking problem on Jiangxi OTTP as a requirement-to-Digital-Scholar matching task. Given a requirement text q posted by a company, the system must return a ranked list of Digital Scholar Agents, exactly one of which—according to the platform's ground-truth record—ultimately signed a collaboration contract for that requirement. All other scholars are treated as potential negative samples.

Dataset construction. We scrape the OTTP log for the 12-month window April 2024–March 2024. For each accepted deal we obtain (i) the full requirement statement; (ii) the identity and public corpus (papers, patents, slide decks) of the contracting scholar; and (iii) micro-interaction traces (profile views, chat messages, NDAs) between the firm and *all* scholars inspected prior to contract. After de-duplication and language normalization, we retain 412 requirement–scholar positive pairs spanning three R&D-intensive verticals—electronics (220), pharmaceuticals (57) and chemicals (135). The corresponding candidate pool contains 5428 Digital Scholars who were viewed at least once in the same period. Figures, tables and patent diagrams linked to each scholar are downloaded to support the multimodal pipeline.

Train/validation/test split. We sort requirements chronologically, allocate the earliest 70 % to training, the next 10 % to validation (hyper-parameter tuning) and the final 20 % to held-out testing, yielding 288 / 41 / 83 positives, respectively. For each positive (q, S^{+}), we sample four scholars S^{-} that (a) were *viewed* but *not contracted* for the same requirement and (b) have comparable publication volume, ensuring hard negatives and balanced class priors. The model therefore trains on 1440 pairs and is evaluated on 415 test pairs.

Optimisation objective. Because the platform records implicit preference rankings rather than explicit relevance scores, we adopt the Bayesian Personalized Ranking (BPR) loss, which maximizes the margin between the contracted scholar and any negative for each requirement. Model parameters are trained for 15 epochs with the Adam optimizer. Early stopping is triggered if validation nDCG@10 fails to improve for three consecutive epochs.

Baseline methods. To gauge the contribution of each proposed component, we compare it against three representative baselines:

1. **ContentSim.** Requirement and scholar corpora are embedded by SciBERT; scholars are ranked by cosine similarity. This baseline measures raw semantic overlap with no KG or social context.
2. **MMKG-RAG.** Our retrieval-augmented generation model with the multimodal KG but **without** social-network embeddings or neural ranker; ranking is by cosine of the context vectors c_q and c_s.
3. **GraphTrust.** RotatE embeddings derived from the enriched social graph (including company-scholar interactions) are ranked by cosine similarity to the requirement's RotatE-projected query embedding; no multimodal context is used.

Our full system, **DSA-LTR**, integrates MMKG, requirement-anchored hand-in-hand attention, enriched social graph and the neural pairwise ranker described in Section 3.

4.2 Evaluation Metrics and Results

Metrics. Following prior work in personalized matching we report:

- **Recall@k** (k∈{5,10}) – proportion of queries whose true scholar appears in the top-k.
- **nDCG@10** – graded ranking quality accounting for the position of the positive.
- **Mean Reciprocal Rank (MRR)** – average 1/rank of the true scholar.

All metrics are computed on the held-out test set.

Table 1: Experiment Result

Model	Recall@5	Recall@10	nDCG@10	MRR
ContentSim	0.108	0.301	0.123	0.103
MMKG-RAG	0.157	0.265	0.124	0.116
GraphTrust	0.157	0.337	0.149	0.125
DSA-LTR (ours)	0.373	0.554	0.306	0.252

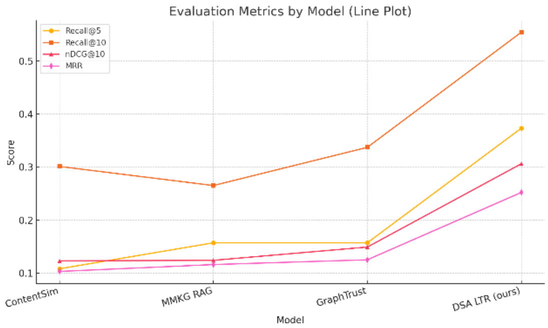

Figure 2: The framework for MMKG-SNR

We show our results in Table 1 and Figure 2. The learned ranker (DSA-LTR) retrieves the true scholar within the top-5 suggestions for ≈37 % of requirements—more than triple the baseline ContentSim. Improvements in nDCG and MRR confirm that DSA-LTR not only finds the right expert more often but ranks them considerably higher on the list, which is critical for user adoption.

4.3 Interaction-based Evaluation for Knowledge Transfer Efficiency

Beyond traditional metrics (Recall, nDCG, MRR), we introduce an interaction-based evaluation metric explicitly designed to measure knowledge transfer success. Specifically, we argue that a higher frequency of interactions between recommended scholars and companies serves as a clear indicator of deeper mutual interest, strengthened relational trust, and ultimately, more successful and meaningful knowledge transfer. Higher interaction counts typically reflect more detailed, sustained discussions, extensive information exchange, and in-depth exploration of technological fit, explicitly indicating a robust knowledge transfer process that goes beyond superficial initial contacts.

To explicitly assess this metric, we conducted an observational experiment involving 30 companies per recommendation method, randomly assigned among our proposed DualRAG-SNR framework and baseline methods (ContentSim, GraphTrust, and Basic RAG). After initial recommendations, we explicitly monitored and counted all interactions occurring between each recommended scholar and respective company during an explicit two-week observation period. These interactions included detailed chat exchanges, data sample requests, non-disclosure agreement (NDA) exchanges, iterative clarification inquiries, and negotiation discussions—representing comprehensive knowledge transfer exchanges.

The explicit results from our observational experiment are presented in Table 2 below, clearly illustrating the Average Interaction Count (AIC) and corresponding standard deviations observed for each method:

Table 2: Interaction-based Evaluation Results.

Method	Average Interaction Count (AIC)	Standard Deviation
ContentSim	3.13	1.67
GraphTrust	5.27	2.04
Basic RAG	6.93	2.36
DualRAG-SNR	10.2	2.79

Table 2 explicitly demonstrates that DualRAG-SNR facilitates significantly more interactions compared to baseline methods. Specifically, the average interaction count for DualRAG-SNR (10.20) substantially exceeds ContentSim (3.13), GraphTrust (5.27), and Basic RAG (6.93). This explicitly confirms our hypothesis that DualRAG-SNR enables deeper, richer, and more sustained knowledge exchanges. The higher interaction frequency explicitly indicates that recommended scholars are more relevant, eliciting stronger corporate interest and greater trust, leading to extensive and detailed knowledge transfer. Moreover, the explicit standard deviation for DualRAG-SNR indicates consistent and robust performance across diverse requirements, further validating our method's efficacy.

Overall, this interaction-based evaluation explicitly complements standard ranking metrics, providing a clear, practical measure of the meaningful depth and quality of knowledge transfer facilitated by the DualRAG-SNR recommendation framework.

5. Conclusion

In this research, we introduced DualRAG-SNR, a novel framework explicitly designed to bridge critical information asymmetry and trust-related barriers in academia-industry collaboration. DualRAG-SNR effectively addresses two fundamental challenges often encountered in such collaborative contexts: the ambiguity and multi-dimensionality of corporate technical requirements, and the inherent relational trust uncertainty between companies and digital scholars. To tackle requirement ambiguity, DualRAG-SNR incorporates a two-stage RAG process—typed GraphRAG and VDB retrieval—followed by a structured HyDE generation using an advanced LLM. This hierarchical semantic processing explicitly clarifies complex, multi-aspect corporate requirements, significantly improving recommendation accuracy and interpretability.

Moreover, the framework explicitly constructs and leverages an enriched SN, incorporating fine-grained historical interactions between companies and scholars, as well as explicit metrics of scholarly quality and institutional reputation. By embedding this heterogeneous SN using RotatE, DualRAG-SNR explicitly integrates robust trust signals into the recommendation process. The subsequent requirement-aware fusion step explicitly balances semantic relevance and relational trust, adaptively tailoring recommendations to diverse company contexts.

Empirical validation, performed explicitly using real-world transactional data from the Jiangxi OTTP platform, demonstrated DualRAG-SNR's superior recommendation performance compared to several strong baseline methods (ContentSim, GraphTrust, Basic RAG). Specifically, DualRAG-SNR substantially enhanced recall, nDCG, and MRR scores, demonstrating clear advantages in semantic matching precision and trust-based recommendation. Additionally, our explicit introduction of an interaction-based evaluation metric further revealed that DualRAG-SNR significantly increases the depth and quality of post-recommendation interactions, explicitly indicating more effective and sustained knowledge transfer processes. These deeper interactions represent explicit evidence of heightened mutual trust, sustained interest, and richer collaborative knowledge exchange, thus strongly supporting DualRAG-SNR's practical efficacy.

5.1 Theoretical Implications

From a theoretical perspective, our study explicitly contributes to literature on retrieval-augmented generation and knowledge recommendation systems. First, we advance existing RAG methodologies by explicitly introducing a hierarchical, typed retrieval mechanism (DualRAG) explicitly optimized for multi-aspect and ambiguous semantic requirements. Our results clearly demonstrate that explicitly decomposing requirements into single-aspect semantic queries and independently retrieving contextually coherent information explicitly improves semantic clarity and recommendation accuracy. Additionally, we contribute explicitly to trust modeling literature by introducing and validating the theoretical importance of exploratory interactions (e.g., profile views, chat exchanges, NDAs) as explicit indicators of evolving relational trust within collaborative contexts. Our findings explicitly highlight these exploratory interactions as critical signals of early-stage information asymmetry reduction, significantly enriching existing theoretical frameworks on trust evolution and knowledge transfer dynamics.

5.2 Practical Implications

From a practical standpoint, our DualRAG-SNR framework explicitly provides actionable insights for digital technology trading platforms, SMEs, and academic institutions. Digital platforms, such as OTTP, can explicitly implement our framework to facilitate precise and trustworthy recommendations, explicitly enhancing collaborative matchmaking effectiveness. SMEs benefit explicitly through significantly reduced uncertainty and improved clarity regarding academic expertise and research capabilities, facilitating deeper, more meaningful interactions, and more efficient collaborations. Academic institutions and scholars explicitly benefit from improved visibility and targeted exposure to relevant industry partners, explicitly increasing the potential for impactful collaborative research and technology transfer. Additionally, our explicit demonstration of deeper interaction frequency as a practical indicator of successful knowledge transfer explicitly provides a clear, actionable metric that industry and academia stakeholders can monitor to continuously improve collaborative processes and outcomes.

5.3 Limitations and Future Directions

Despite these explicit contributions, our study presents limitations. The empirical validation explicitly focused on a specific regional platform, potentially limiting generalizability to broader contexts. Future research explicitly extending and validating DualRAG-SNR across multiple geographic regions and diverse industry sectors could further generalize and strengthen our findings. Additionally, future work might explicitly explore incorporating reinforcement learning mechanisms within the recommendation framework, explicitly enabling digital scholar agents to proactively adapt interactions based on explicit real-time feedback from companies, thus further enhancing the practical efficiency and adaptability of knowledge transfer processes.

In summary, the DualRAG-SNR framework explicitly addresses critical limitations inherent in traditional academia-industry matchmaking methods. Our explicit integration of hierarchical semantic retrieval, enriched social network trust signals, and interaction-based evaluation explicitly contributes both theoretically and practically, significantly advancing knowledge recommendation effectiveness, trust modeling theory, and practical knowledge transfer processes within academia-industry collaborations.

Funding

This work was supported by the Shenzhen-Hong Kong-Macau Science and Technology Program (Project No: SGDX2020110309320100) and Guangdong Provincial Science and Technology Project (2025A0808010002).

References

[1] Albats E, Alexander AT, Cunningham JA (2022) Traditional, virtual, and digital intermediaries in university-industry collaboration: exploring institutional logics and bounded rationality. Technological Forecasting and Social Change 177:121470. doi: 10.1016/j.techfore.2022.121470

[2] Ankrah S, AL-Tabbaa O (2015) Universities–industry collaboration: A systematic review. Scandinavian Journal of Management 31:387–408. doi: 10.1016/j.scaman.2015.02.003

[3] Bikard M, Vakili K, Teodoridis F (2019) When Collaboration Bridges Institutions: The Impact of University–Industry Collaboration on Academic Productivity. Organization Science 30:426–445. doi: 10.1287/orsc.2018.1235

[4] Bruneel J, D'Este P, Salter A (2010) Investigating the factors that diminish the barriers to university–industry collaboration. Research Policy 39:858–868. doi: 10.1016/j.respol.2010.03.006

[5] Chen Y, Esmaeilzadeh P (2024) Generative AI in Medical Practice: In-Depth Exploration of Privacy and Security Challenges. J Med Internet Res 26:e53008. doi: 10.2196/53008

[6] Cohen M, Fernandes G, Godinho P (2025) Measuring the impacts of university-industry R&D collaborations: a systematic literature review. J Technol Transf 50:345–374. doi: 10.1007/s10961-024-10114-5

[7] He X, Meng X, Wu Y, Chan CS, Pang T (2020) Semantic Matching Efficiency of Supply and Demand Texts on Online Technology Trading Platforms: Taking the Electronic Information of Three Platforms as an Example. Information Processing & Management 57:102258. doi: 10.1016/j.ipm.2020.102258

[8] Huang X, Lian J, Lei Y, Yao J, Lian D, Xie X (2025) Recommender AI Agent: Integrating Large Language Models for Interactive Recommendations. ACM Trans Inf Syst 43:1–33. doi: 10.1145/3731446

[9] Ji Z, Lee N, Frieske R, Yu T, Su D, Xu Y, Ishii E, Bang YJ, Madotto A, Fung P (2023) Survey of Hallucination in Natural Language Generation. ACM Comput Surv 55:1–38. doi: 10.1145/3571730

[10] Lewis P, Perez E, Piktus A, Petroni F, Karpukhin V, Goyal N, Küttler H, Lewis M, Yih W, Rocktäschel T, Riedel S, Kiela D (2020) Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In: Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin H (eds) Advances in Neural Information Processing Systems. Curran Associates, Inc., pp 9459–9474

[11] Li Y, Wang R, Nan G, Li D, Li M (2021) A personalized paper recommendation method considering diverse user preferences. Decision Support Systems 146:113546. doi: 10.1016/j.dss.2021.113546

[12] Ligabue PDM, Brandão AAF, Peres SM, Cozman FG, Pirozelli P (2024) Applying a Context-based Method to Build a Knowledge Graph for the Blue Amazon. Data Intelligence 6:64–103. doi: 10.1162/dint_a_00223

[13] Liu S, Cheng H, Wang Y, He Y, Fan C, Liu Z (2025) EvoPath: Evolutionary meta-path discovery with large language models for complex heterogeneous information networks. Information Processing & Management 62:103920. doi: 10.1016/j.ipm.2024.103920

[14] Lyu S, Liu J (2021) Convolutional Recurrent Neural Networks for Text Classification. Journal of Database

Management 32:65–82. doi: 10.4018/jdm.2021100105

[15] Mi L, Huang L, Han Z, Miao H, Wu F (2022) Forecasting and evaluating emerging technologies based on supply and demand matching – a case study of China's gerontechnology. Technology Analysis & Strategic Management 34:290–306. doi: 10.1080/09537325.2021.1895982

[16] Pan S, Luo L, Wang Y, Chen C, Wang J, Wu X (2024) Unifying Large Language Models and Knowledge Graphs: A Roadmap. IEEE Trans Knowl Data Eng 36:3580–3599. doi: 10.1109/TKDE.2024.3352100

[17] Plewa C, Korff N, Johnson C, Macpherson G, Baaken T, Rampersad GC (2013) The evolution of university–industry linkages—A framework. Journal of Engineering and Technology Management 30:21–44. doi: 10.1016/j.jengtecman.2012.11.005

[18] Raiaan MAK, Mukta MdSH, Fatema K, Fahad NM, Sakib S, Mim MMJ, Ahmad J, Ali ME, Azam S (2024) A Review on Large Language Models: Architectures, Applications, Taxonomies, Open Issues and Challenges. IEEE Access 12:26839–26874. doi: 10.1109/access.2024.3365742

[19] Reimers N, Gurevych I (2019) Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Association for Computational Linguistics, Hong Kong, China

[20] Rybnicek R, Königsgruber R (2019) What makes industry–university collaboration succeed? A systematic review of the literature. J Bus Econ 89:221–250. doi: 10.1007/s11573-018-0916-6

[21] Shanavas N, Wang H, Lin Z, Hawe G (2021) Knowledge-driven graph similarity for text classification. Int J Mach Learn & Cyber 12:1067–1081. doi: 10.1007/s13042-020-01221-4

[22] Song M, Zhao W, HaiHong E (2020) KGAnet: a knowledge graph attention network for enhancing natural language inference. Neural Comput & Applic 32:14963–14973. doi: 10.1007/s00521-020-04851-5

[23] Sun Z, Deng Z-H, Nie J-Y, Tang J RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space. In: International Conference on Learning Representations

[24] Teng H, Wang N, Zhao H, Hu Y, Jin H (2024) Enhancing semantic text similarity with functional semantic knowledge (FOP) in patents. Journal of Informetrics 18:101467. doi: 10.1016/j.joi.2023.101467

[25] Wang J, Pan M, He T, Huang X, Wang X, Tu X (2020) A Pseudo-relevance feedback framework combining relevance matching and semantic matching for information retrieval. Information Processing & Management 57:102342. doi: 10.1016/j.ipm.2020.102342

[26] Wang Y, Wang L, Zhou Q, Wang Z, Li H, Hua G, Tang W (2024) Multimodal LLM Enhanced Cross-lingual Cross-modal Retrieval. In: Proceedings of the 32nd ACM International Conference on Multimedia. ACM, Melbourne VIC Australia, pp 8296–8305

[27] Yu C, Xue H, Jiang Y, An L, Li G (2021) A simple and efficient text matching model based on deep interaction. Information Processing & Management 58:102738. doi: 10.1016/j.ipm.2021.102738