

TECHMATCH: Matching Enterprise Technology Demands to Academic Resources with Graph-Augmented Retrieval

Rui Chai¹, Shiyun Huang¹, Yubing Su¹, Yuhan Zhen¹, Yi Sun¹,
Maozhe Zhao², Rui Shen², Shengzhong Liu², Ruijie Wang^{1*}, Jianxin Li¹

¹Beihang University, Beijing, China

²Shanghai Jiao Tong University, Shanghai, China

{24371071, 24373386, 24373213, 24373397, 24371286, ruijiiew, lijx}@buaa.edu.cn

{larval_roc, shenrui2023, shengzhong}@sjtu.edu.cn

Abstract

Matching enterprise technology demands with academic resources is critical for technology transfer, yet remains challenging because enterprise requests are often vague, constraint-sensitive, and expressed differently from scholarly outputs. Existing search and RAG systems may retrieve topically relevant results but often provide unsupported recommendations or insufficient evidence. We present TECHMATCH, an interactive system for evidence-constrained matching of enterprise demands to papers, scholars, patents, and projects. Its key novelty is a trustworthy workflow that combines demand clarification, structured constraint extraction, graph-augmented retrieval, evidence-aware reranking, and traceable report generation. Each recommendation is grounded in retrieved source evidence, reducing the risk of hallucinated rationales and exposing unresolved constraints and feasibility risks. To support efficient interaction, TECHMATCH uses hierarchical scholar representations, lightweight profile indexing, parallel reranking, and context-budgeted evidence selection. We construct a large-scale academic corpus comprising nearly four million research resources. On a public-source benchmark, TECHMATCH outperforms Standard RAG by 9.9% on P@5, 10.1% on NDCG@5, and 9.3% on evidence support, while enabling users to inspect evidence and generate actionable technology-transfer reports.

Keywords: Technology Transfer; Enterprise Demand Analysis; GraphRAG

1 Introduction

Technology transfer depends on connecting enterprise demands with academic capabilities. This discovery need is bidirectional: enterprises must find papers, scholars, patents, and projects that can

address concrete operational problems, while universities and researchers need to make their outputs discoverable to enterprises whose demands they can realistically address. However, enterprises describe application scenarios, constraints, and target indicators, whereas academic resources emphasize methods, experiments, authors, venues, and technical contributions.

Meeting this need is difficult because enterprise requests are often vague, constraint-sensitive, and expressed differently from scholarly outputs. Keyword systems may miss semantically related solutions, while dense retrieval and general-purpose RAG can return broadly topical documents without explicitly representing technology entities, numerical targets, or method-task relations. Their recommendations and rationales may also lack source-grounded evidence, leaving users unable to verify the match, identify unresolved constraints, or assess feasibility risks. Technology-transfer matching therefore requires both ranked resources and traceable, task-specific evidence.

We present TECHMATCH, an interactive system for evidence-constrained enterprise-demand-academic-resource matching. Unlike standard RAG, TECHMATCH combines demand clarification, structured constraint extraction, graph-augmented retrieval (Edge et al., 2024), evidence-aware reranking, and traceable report generation. It clarifies vague requests through dialogue, extracts technical directions and numerical requirements, retrieves candidates across papers, scholars, patents, and projects, and reports supporting evidence and remaining feasibility risks. As a decision-support system rather than a replacement for domain experts, it lets users inspect evidence and compare candidates before contacting potential collaborators.

To support this workflow at scale, we collect 2016–2026 AMiner records covering 33,891 scholars and nearly four million academic resources:

*Ruijie Wang is the corresponding author.

3.25M papers, 682K patents, and 59K projects across ten fields. On a public-source benchmark, TECHMATCH achieves relative gains over Standard RAG of approximately 10% in top-ranked precision, ranking quality, and actionability, together with a 9.3% gain in evidence support. Further analyses show that explicit constraints are important for actionable matching and that resource-type routing supports complementary exploration of papers, scholars, patents, and projects.

Our contributions are as follows:

- We introduce TECHMATCH, an interactive system for evidence-grounded matching between enterprise technology demands and academic resources.
- We implement a GraphRAG-based pipeline that combines structured demand extraction, graph-enhanced retrieval, LLM-based reranking, and traceable match-report generation.
- We construct a public-source benchmark and evaluate TECHMATCH against sparse retrieval, direct E^2 GraphRAG retrieval, and a Standard RAG baseline. We also analyze demand representation, resource types, report quality, and latency.

2 TECHMATCH: A Demand–Research Matching System

2.1 Data

We collect scholars’ papers, patents, and projects (hereinafter referred to as resources), together with scholar profile metadata, from AMiner (Wan et al., 2019). We focus on scholars affiliated with China’s Project 985 and Project 211 universities. To retain active researchers with sufficient evidence, we keep only scholars with at least 50 papers published during 2016–2026 and an H-index of at least 10. The resulting dataset contains 33,891 scholar profiles and covers 3.25M papers, 682K patents, and 59K projects from 2016–2026 across ten broad research fields. Detailed corpus statistics are provided in Appendix B.

2.2 Knowledge Base Construction

We organize the knowledge base using GraphRAG (Edge et al., 2024) and a vector database. Unlike plain RAG, which relies mainly on text similarity, GraphRAG represents entities and their relations in a graph and uses hierarchical

summaries to support relation-aware retrieval. We use spaCy (Honnibal et al., 2020) to extract technical entities efficiently and LLMs to build summary trees. Each scholar is represented by an individual graph and summary tree. Each resource (paper, patent, or project) is represented as a text chunk, from which entities are extracted to construct the graph. The leaf nodes of each summary tree correspond to these chunks. We refine the tree-construction prompt so that the top-level node represents the scholar’s comprehensive academic profile, and we use FAISS (Johnson et al., 2021) to index scholar-profile embeddings.

2.3 Retrieval

Scholar Retrieval. Scholar retrieval first uses semantic similarity between the request and profile embeddings to select candidates. For each candidate, it then queries GraphRAG for the paper chunks most relevant to the request. Following the AMiner AI 2000 weighting strategy (AMiner, 2026), the contribution weight of the i -th author in a paper with n authors is $w_i = (1/i) / \sum_{k=1}^n (1/k)$. For a scholar with M retrieved chunks, the academic score is $s_{\text{faiss}} \cdot \sum_{m=1}^M w_i^{(m)} \times \text{citation}_m$. Here, s_{faiss} denotes semantic similarity. We select the top- N scholars for recommendation based on these scores. This design favors scholars who are semantically relevant to the demand and have highly cited outputs that support the match.

Resource Retrieval. For the top- N scholars, the system returns the resources retrieved via GraphRAG.

2.4 Model

Beyond knowledge base construction and retrieval, TECHMATCH incorporates three additional model components: structured demand extraction, LLM-based reranking, and evidence-grounded report generation.

Structured Demand Extraction. Given an industrial demand document d , a Qwen3-32B-AWQ requirement-analysis model (Yang et al., 2025) extracts a structured representation $z_d = \{p, s, c, t, r\}$. Here, p , s , c , t , and r denote the pain point, application scenario, technical constraints, target technology area, and readiness or pilot-deployment requirements, respectively. The prompt explicitly asks the model to retain only source-grounded facts. When the input is vague, the conversational interface asks clarifying questions about missing scenario details, constraints,

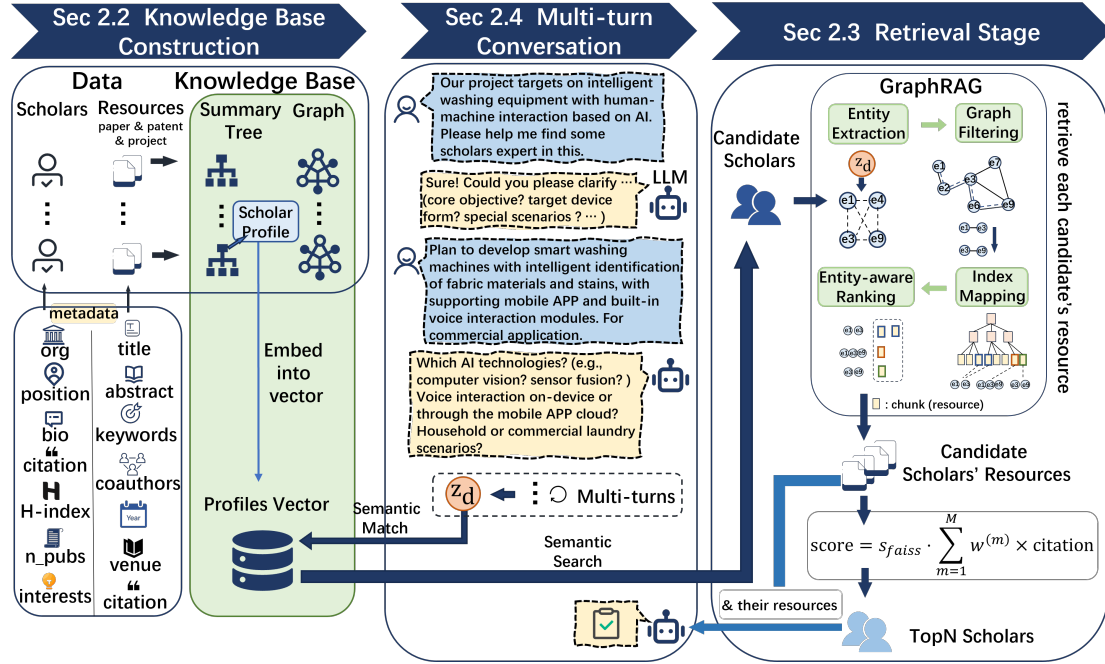


Figure 1: Overview of TECHMATCH. The system converts heterogeneous technology-transfer documents into structured representations, retrieves supporting evidence, ranks candidate matches, and generates explainable matching reports.

or target metrics. After user confirmation, the finalized representation is converted into technical directions and numerical constraints for retrieval and reporting.

LLM-Based Reranking. After retrieval, each candidate scholar or resource is evaluated by a scoring LLM (Qwen3-32B-AWQ, temperature = 0). The prompt contains the full demand, candidate metadata, source evidence, and, for scholars, profile information and representative papers. The model assigns a 0–10 match score and a one-sentence rationale. Scores of 8–10 indicate high alignment, 6–7 partial but useful alignment, and 0–5 weak relevance. Candidates are sorted by score, with the original retrieval rank as a tiebreaker. To maintain throughput, scoring calls run in parallel with up to 30 workers.

Evidence-Grounded Report Generation. The final stage generates a match report for the selected candidates. Each report includes the matched demand element, recommended scholar or resource, evidence excerpts from retrieved chunks, feasibility risks, and suggested engagement steps. When retrieved chunks exceed the context window, a context-budget manager retains the highest-scoring evidence and a small tail segment that may contain caveats, subject to a 15,000-token limit. The system can also generate a connection suggestion that orders matched scholars and outlines outreach and pilot-validation steps.

3 Demonstration

We demonstrate TECHMATCH with an end-to-end university–industry matching scenario in which an enterprise seeks academic resources for data-center liquid-cooling technology. The source code, demonstration video, benchmark artifacts, and access notes for restricted data are available at https://anonymous.4open.science/r/techmatch_emnlp-151E. In the workflow, the user selects a target resource type, enters a brief demand, answers clarification questions about cooling route, infrastructure compatibility, coolant preference, and target metrics, and then triggers resource matching. Figure 2 shows the main interface. Appendix D presents the full system demonstration. TECHMATCH returns ranked candidates with fit scores, institutions, representative outputs, recommendation rationales, and evidence snippets. The user can select promising resources for report generation, after which the system summarizes the clarified demand, selected resources, supporting evidence, feasibility risks, and connection suggestions. This demonstration shows how TECHMATCH supports an interactive workflow from vague industrial demand to evidence-grounded resource matching and actionable technology-transfer follow-up.

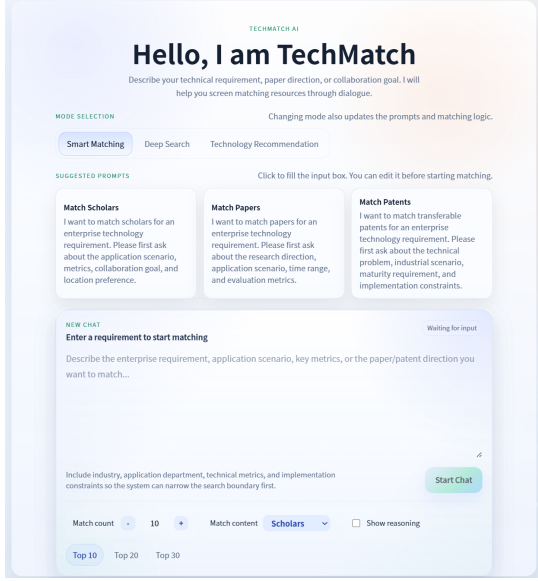


Figure 2: Main interface of TECHMATCH for entering a demand and selecting the target resource type.

4 Evaluation

We evaluate TECHMATCH with a public-source enterprise-demand benchmark. The benchmark represents industrial demands as natural-language requirements and, when available, as structured descriptions containing the domain, application scenario, demand title, technical constraints, target indicators, expected resource types, and technology tags. For each demand, a method returns the top 10 candidate academic resources.

Compared methods. The main comparison includes six methods: BM25 (Robertson and Zaragoza, 2009), TF-IDF (Salton and Buckley, 1988), BM25+TF-IDF hybrid retrieval, direct E^2 GraphRAG retrieval (Zhao et al., 2025), a Standard RAG baseline, and the full TECHMATCH pipeline. All methods use the same hybrid demand description, which combines natural-language text with structured fields. The sparse baselines rank papers using lexical matching over the paper corpus. Direct E^2 GraphRAG is included as an external graph-retrieval baseline and does not use the TECHMATCH demand-processing or matching stages. The Standard RAG baseline uses retrieval-augmented matching without the full TECHMATCH workflow. The full system uses structured demand analysis, GraphRAG candidate retrieval, LLM-based reranking, and evidence-grounded match rationales. The main comparison is therefore a controlled-input, preliminary system-level evaluation; raw natural-language inputs are separately analyzed in the demand representation ablation.

Evaluation protocol. We pool the top-ranked results from all compared methods and conduct blind, LLM-assisted evaluation using Qwen3-32B-AWQ served through vLLM (Kwon et al., 2023), following recent LLM-as-judge and reference-free evaluation practices (Liu et al., 2023; Zheng et al., 2023). For each demand–resource pair, the judge assigns three integer scores from 0 to 3: relevance, evidence support, and actionability. The relevance rubric assigns scores of 0, 1, 2, and 3 to unrelated, topical, partially matched, and highly matched results, respectively. For precision metrics, a result is counted as relevant when its relevance score is at least 2. The same protocol is used for the ablation studies. To assess judging stability, we repeat the scoring on a 10% sample of the pooled candidates. Exact agreement is 0.8159 for relevance, 0.8434 for evidence support, and 0.7088 for actionability, while agreement within one point is 1.0, 0.9973, and 1.0, respectively. Because LLM evaluators can exhibit position, self-preference, and other systematic biases (Wang et al., 2024; Panickssery et al., 2024), we treat these scores as preliminary evidence rather than expert ground truth.

Metrics. We report AvgRel@5, P@5, NDCG@5, and NDCG@10 as ranking metrics. We additionally report EvidenceSupport@5 and Actionability@5. AvgRel@5 averages the graded relevance scores of the top five results. P@5 measures the fraction of the top five results with a relevance score of at least 2. NDCG uses graded relevance gains. EvidenceSupport@5 and Actionability@5 average the corresponding 0–3 judge scores over the top five results.

5 Results and Discussion

Table 1 reports the main comparison. TECHMATCH achieves the best result on every metric. The sparse baselines perform poorly: BM25, TF-IDF, and their hybrid variant obtain P@5 scores of around 0.30, showing that surface-form matching is insufficient for aligning enterprise requirements with academic papers. Direct E^2 GraphRAG substantially improves over lexical retrieval, supporting the value of graph-enhanced retrieval. The Standard RAG baseline further improves the top-ranked results, but the full TECHMATCH pipeline remains strongest, reaching 0.952 P@5 and 0.783 NDCG@5. Relative to Standard RAG, these scores correspond to improvements of 9.9% on P@5, 10.1% on NDCG@5, 7.2% on AvgRel@5, 9.3%

Method	AvgRel@5	P@5	NDCG@5	NDCG@10	Evid.@5	Act.@5
BM25	0.806	0.244	0.229	0.211	0.676	0.576
TF-IDF	1.100	0.360	0.322	0.298	0.918	0.766
BM25+TF-IDF	0.990	0.316	0.289	0.264	0.828	0.696
Direct E^2 GraphRAG	1.820	0.750	0.578	0.551	1.572	1.442
Standard RAG	2.044	0.866	0.711	0.638	1.936	1.780
TECHMATCH Full	2.191	0.952	0.783	0.710	2.117	1.960

Table 1: Main comparison on the public-source enterprise-demand benchmark under the controlled hybrid-input setting. Evid.@5 and Act.@5 denote EvidenceSupport@5 and Actionability@5.

on EvidenceSupport@5, and 10.1% on Actionability@5.

A bootstrap comparison against the strongest baseline, Standard RAG, also shows consistent gains. The relative gains above correspond to positive paired differences on the original metric scale: 0.086 for P@5, 0.0719 for NDCG@5, 0.1465 for AvgRel@5, 0.1805 for EvidenceSupport@5, and 0.180 for Actionability@5. The 95% confidence intervals for these paired differences do not cross zero; for example, the NDCG@5 interval is [0.0248, 0.1217], indicating that the full pipeline outperforms the strongest baseline under this LLM-assisted evaluation protocol.

Demand representation ablation. Table 2 studies five demand formulations: raw natural language, natural language plus structured fields, oracle structured input, structured input without target indicators, and structured input without technical constraints. This ablation varies the information available in the input demand rather than disabling an internal backend module. The first three variants perform similarly, suggesting that the system is robust to raw enterprise descriptions. The clearest degradation occurs when constraints are removed: relative to Hybrid NL+Struct., P@5 decreases by 3.6%, NDCG@5 by 3.9%, and Actionability@5 by 5.0%. This result indicates that application constraints, operating conditions, and boundary requirements are particularly important for technology-transfer matching. By contrast, removing target indicators causes little degradation on this benchmark, possibly because many matches are driven by evidence about the technical direction and application scenario rather than by numerical targets alone.

Resource-type analysis. Table 3 compares retrieval over papers, scholars, patents, and funded projects using the same hybrid demands. Paper retrieval performs best across all ranking and evidence metrics. Scholar retrieval remains useful for identifying potential experts, but its evidence support is lower because scholar profiles contain more

Variant	AvgRel@5	P@5	NDCG@5	Act.@5
Raw NL	2.186	0.950	0.763	1.976
Hybrid NL+Struct.	2.189	0.954	0.763	1.960
Oracle Struct.	2.197	0.944	0.772	1.967
w/o Indicators	2.193	0.952	0.766	1.986
w/o Constraints	2.113	0.920	0.733	1.862

Table 2: Demand representation ablation. The largest drop is observed when technical constraints are removed.

Type	AvgRel@5	P@5	NDCG@5	Evid.@5
Paper	2.158	0.944	0.799	2.104
Scholar	1.948	0.808	0.696	1.234
Patent	1.822	0.798	0.623	0.894
Project	1.744	0.746	0.571	0.728

Table 3: Resource-type analysis. Papers provide the strongest evidence-grounded matches; other resource types support broader technology-transfer exploration.

aggregated evidence than individual papers. Patent and project retrieval show lower evidence support and actionability, reflecting a mismatch between enterprise demands and shorter or less descriptive resource records. Thus, these results should be interpreted as a resource-specific capability analysis rather than a single fair comparison among resource types.

Report quality. We also evaluate 20 generated matching reports. The report-level average scores are 1.85/3 for relevance, 1.20/3 for evidence support, 1.15/3 for completeness, 0.30/3 for risk awareness, and 1.25/3 for actionability. These results show that the reports achieve usable scores for a demonstration system, although their evaluation scores remain below those of the top- K matches. In particular, risk awareness is low: reports often summarize retrieved resources but do not sufficiently discuss feasibility risks, deployment constraints, or concrete next steps. This finding motivates future work on risk-sensitive report generation and more rigorous citation of supporting evidence. Supplementary evaluation visualizations are provided in Appendix C.

Latency and stability. Table 4 reports end-to-end runtime and success rates for the backend methods. Direct E^2 GraphRAG retrieval is fastest because it only performs retrieval. The full TECHMATCH workflow adds LLM-based screening and evidence-grounded explanation generation, but remains 24.1% faster than Standard RAG while achieving stronger matching quality. This suggests a practical latency-quality trade-off for interactive technology-transfer matching.

Method	Cases	Success	Avg. time
Direct E^2 GraphRAG	100	100/100	39.73s
TECHMATCH Full	100	100/100	63.87s
Standard RAG	100	100/100	84.16s

Table 4: Runtime and stability of backend methods on the public-source benchmark.

6 Related Work

TECHMATCH is related to three lines of work. First, retrieval-augmented generation grounds outputs in external documents and supports modular, updatable, and evidence-grounded systems for knowledge-intensive NLP, including open-domain QA systems and large retrieval-augmented language models (Guu et al., 2020; Lewis et al., 2020; Izacard et al., 2023). Recent work advances this paradigm through late-interaction and zero-shot dense retrieval (Khattab and Zaharia, 2020; Gao et al., 2023), in-context retrieval (Ram et al., 2023), hierarchical indexing and self-reflection (Sarathi et al., 2024; Asai et al., 2024), and graph-based retrieval (Jimenez Gutierrez et al., 2024; He et al., 2024). Unlike general QA, TECHMATCH returns ranked academic resources with evidence, fit rationales, and follow-up suggestions.

Second, scientific search, dense retrieval, and expert finding connect information needs with papers, authors, and institutions. SciBERT and SPECTER improve paper representations with scientific corpora or citation graphs (Beltagy et al., 2019; Cohan et al., 2020), while AMiner provides researcher profiles, publication graphs, and expert discovery (Wan et al., 2019). Dense retrieval supports matching beyond keyword overlap (Karpukhin et al., 2020; Reimers and Gurevych, 2019); large scientific corpora and benchmarks support representation, citation recommendation, and evidence verification (Lo et al., 2020; Singh et al., 2023; Bhagavatula et al., 2018; Wadden et al., 2020); and expert-finding models link authors, documents, topics, and assignments (Tang et al., 2008; Balog et al., 2006; Mimno and McCallum, 2007). TECHMATCH builds on this work with GraphRAG candidate retrieval (Edge et al., 2024) that explicitly aligns industrial constraints with academic outputs.

Third, university–industry research spans academic engagement, collaboration channels, public-research spillovers, organizational practices, and patent licensing (Perkmann et al., 2013; D’Este and Patel, 2007; Cohen et al., 2002; Siegel et al., 2003; Bozeman, 2000; Mowery et al., 2001). Because

matches affect resource allocation, intellectual-property negotiation, and pilot planning, platforms need human-in-the-loop support that exposes the evidence behind recommendations and any remaining risks. Human–AI research similarly emphasizes understandable recommendations and calibrated reliance (Amershi et al., 2019; Bansal et al., 2021). Accordingly, TECHMATCH combines RAG, graph-enhanced retrieval, and interactive review for evidence-grounded demand–research matching rather than generic QA or paper search.

7 Conclusion

We presented TECHMATCH, an interactive system integrating demand clarification, constraint extraction, graph-enhanced retrieval, LLM-based reranking, and evidence-grounded reporting to match industrial demands with papers, scholars, patents, and projects. On a public-source benchmark, it outperforms sparse retrieval baselines, direct E^2 GraphRAG, and Standard RAG in top-ranked quality, evidence support, and actionability. Ablations show that explicit constraints improve the actionability of matches; papers provide the strongest evidence, while other resource types broaden discovery. Overall, TECHMATCH turns vague industrial requests into traceable candidate resources for expert review.

8 Limitations

The corpus inherits AMiner’s coverage and includes 2016–2026 outputs from scholars at selected Chinese universities; missing or outdated records can exclude candidates. The public-source benchmark and LLM-assisted judging cannot replace expert annotations, user studies, or confidential-demand evaluation. Strong retrieval does not guarantee technical, commercial, or deployment feasibility. Limited risk awareness and actionability call for stronger citations, calibrated uncertainty, and checks of target metrics and intellectual-property or legal constraints. Expert review remains necessary; future work will broaden coverage and strengthen risk awareness.

Acknowledgments

This work is partially supported by NSFC program (No. 62672021). Ruijie Wang is supported by the Fundamental Research Funds for the Central Universities. Ruijie Wang is the corresponding author.

References

- Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. *Guidelines for human–AI interaction*. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–13. Association for Computing Machinery.
- AMiner. 2026. *AMiner AI 2000 most influential scholar annual list*. <https://www.aminer.cn/ai2000>. Accessed 2026-07-08.
- Akari Asai, Zeqiu Wu, Yizhong Wang, Avirup Sil, and Hannaneh Hajishirzi. 2024. *Self-RAG: Learning to retrieve, generate, and critique through self-reflection*. In *The Twelfth International Conference on Learning Representations*. OpenReview.net.
- Krisztian Balog, Leif Azzopardi, and Maarten de Rijke. 2006. *Formal models for expert finding in enterprise corpora*. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 43–50. Association for Computing Machinery.
- Gagan Bansal, Tongshuang Wu, Joyce Zhou, Raymond Fok, Besmira Nushi, Ece Kamar, Marco Tulio Ribeiro, and Daniel S. Weld. 2021. *Does the whole exceed its parts? the effect of AI explanations on complementary team performance*. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–16. Association for Computing Machinery.
- Iz Beltagy, Kyle Lo, and Arman Cohan. 2019. *SciBERT: A pretrained language model for scientific text*. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 3615–3620. Association for Computational Linguistics.
- Chandra Bhagavatula, Sergey Feldman, Russell Power, and Waleed Ammar. 2018. *Content-based citation recommendation*. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 238–251. Association for Computational Linguistics.
- Barry Bozeman. 2000. *Technology transfer and public policy: A review of research and theory*. *Research Policy*, 29(4–5):627–655.
- Arman Cohan, Sergey Feldman, Iz Beltagy, Doug Downey, and Daniel S. Weld. 2020. *SPECTER: Document-level representation learning using citation-informed transformers*. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2270–2282. Association for Computational Linguistics.
- Wesley M. Cohen, Richard R. Nelson, and John P. Walsh. 2002. *Links and impacts: The influence of public research on industrial R&D*. *Management Science*, 48(1):1–23.
- Pablo D’Este and Pari Patel. 2007. *University–industry linkages in the UK: What are the factors underlying the variety of interactions with industry?* *Research Policy*, 36(9):1295–1313.
- Darren Edge, Ha Trinh, Newman Cheng, Joshua Bradley, Alex Chao, Apurva Mody, Steven Truitt, Dasha Metropolitansky, Robert Osazuwa Ness, and Jonathan Larson. 2024. *From local to global: A graph RAG approach to query-focused summarization*. *arXiv preprint arXiv:2404.16130*.
- Luyu Gao, Xueguang Ma, Jimmy Lin, and Jamie Callan. 2023. *Precise zero-shot dense retrieval without relevance labels*. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1762–1777. Association for Computational Linguistics.
- Kelvin Guu, Kenton Lee, Zora Tung, Panupong Pasupat, and Ming-Wei Chang. 2020. *REALM: Retrieval-augmented language model pre-training*. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 3929–3938. PMLR.
- Xiaoxin He, Yijun Tian, Yifei Sun, Nitesh V. Chawla, Thomas Laurent, Yann LeCun, Xavier Bresson, and Bryan Hooi. 2024. *G-Retriever: Retrieval-augmented generation for textual graph understanding and question answering*. In *Advances in Neural Information Processing Systems*, volume 37.
- Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. *spaCy: Industrial-strength natural language processing in python*. Zenodo. Software.
- Gautier Izacard, Patrick Lewis, Maria Lomeli, Lucas Hosseini, Fabio Petroni, Timo Schick, Jane Dwivedi-Yu, Armand Joulin, Sebastian Riedel, and Edouard Grave. 2023. *Atlas: Few-shot learning with retrieval augmented language models*. *Journal of Machine Learning Research*, 24(251):1–43.
- Bernal Jimenez Gutierrez, Yiheng Shu, Yu Gu, Michihiro Yasunaga, and Yu Su. 2024. *HippoRAG: Neurobiologically inspired long-term memory for large language models*. In *Advances in Neural Information Processing Systems*, volume 37.
- Jeff Johnson, Matthijs Douze, and Hervé Jégou. 2021. *Billion-scale similarity search with GPUs*. *IEEE Transactions on Big Data*, 7(3):535–547.
- Vladimir Karpukhin, Barlas Oguz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen-tau Yih. 2020. *Dense passage retrieval for open-domain question answering*. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pages 6769–6781. Association for Computational Linguistics.

- Omar Khattab and Matei Zaharia. 2020. [ColBERT: Efficient and effective passage search via contextualized late interaction over BERT](#). In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 39–48. Association for Computing Machinery.
- Woosuk Kwon, Zhuohan Li, Siyuan Zhuang, Ying Sheng, Lianmin Zheng, Cody Hao Yu, Joseph E. Gonzalez, Hao Zhang, and Ion Stoica. 2023. [Efficient memory management for large language model serving with PagedAttention](#). In *Proceedings of the 29th Symposium on Operating Systems Principles*, pages 611–626. Association for Computing Machinery.
- Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Kuttler, Mike Lewis, Wen-tau Yih, Tim Rocktaschel, Sebastian Riedel, and Douwe Kiela. 2020. [Retrieval-augmented generation for knowledge-intensive NLP tasks](#). In *Advances in Neural Information Processing Systems*, volume 33, pages 9459–9474.
- Yang Liu, Dan Iter, Yichong Xu, Shuohang Wang, Ruochen Xu, and Chenguang Zhu. 2023. [G-Eval: NLG evaluation using GPT-4 with better human alignment](#). In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 2511–2522. Association for Computational Linguistics.
- Kyle Lo, Lucy Lu Wang, Mark Neumann, Rodney Kinney, and Daniel S. Weld. 2020. [S2ORC: The semantic scholar open research corpus](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4969–4983. Association for Computational Linguistics.
- David Mimno and Andrew McCallum. 2007. [Expertise modeling for matching papers with reviewers](#). In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 500–509. Association for Computing Machinery.
- David C. Mowery, Richard R. Nelson, Bhaven N. Sampat, and Arvids A. Ziedonis. 2001. [The growth of patenting and licensing by U.S. universities: An assessment of the effects of the Bayh–Dole act of 1980](#). *Research Policy*, 30(1):99–119.
- Arjun Panickssery, Samuel R. Bowman, and Shi Feng. 2024. [LLM evaluators recognize and favor their own generations](#). In *Advances in Neural Information Processing Systems*, volume 37.
- Markus Perkmann, Valentina Tartari, Maureen McKelvey, Erkkö Autio, Anders Broström, Pablo D’Este, Riccardo Fini, Aldo Geuna, Rosa Grimaldi, Alan Hughes, Stefan Krael, Michael Kitson, Patrick Llerena, Francesco Lissoni, Ammon Salter, and Maurizio Sobrero. 2013. [Academic engagement and commercialisation: A review of the literature on university–industry relations](#). *Research Policy*, 42(2):423–442.
- Ori Ram, Yoav Levine, Itay Dalmedigos, Dor Mühlgay, Amnon Shashua, Kevin Leyton-Brown, and Yoav Shoham. 2023. [In-context retrieval-augmented language models](#). *Transactions of the Association for Computational Linguistics*, 11:1316–1331.
- Nils Reimers and Iryna Gurevych. 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*, pages 3982–3992. Association for Computational Linguistics.
- Stephen Robertson and Hugo Zaragoza. 2009. [The probabilistic relevance framework: BM25 and beyond](#). *Foundations and Trends in Information Retrieval*, 3(4):333–389.
- Gerard Salton and Christopher Buckley. 1988. [Term-weighting approaches in automatic text retrieval](#). *Information Processing & Management*, 24(5):513–523.
- Parth Sarthi, Salman Abdullah, Aditi Tuli, Shubh Khanna, Anna Goldie, and Christopher D. Manning. 2024. [RAPTOR: Recursive abstractive processing for tree-organized retrieval](#). In *The Twelfth International Conference on Learning Representations*. OpenReview.net.
- Donald S. Siegel, David Waldman, and Albert Link. 2003. [Assessing the impact of organizational practices on the relative productivity of university technology transfer offices: An exploratory study](#). *Research Policy*, 32(1):27–48.
- Amanpreet Singh, Mike D’Arcy, Arman Cohan, Doug Downey, and Sergey Feldman. 2023. [SciRepEval: A multi-format benchmark for scientific document representations](#). In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 5548–5566. Association for Computational Linguistics.
- Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. 2008. [ArnetMiner: Extraction and mining of academic social networks](#). In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 990–998. Association for Computing Machinery.
- David Wadden, Shanchuan Lin, Kyle Lo, Lucy Lu Wang, Madeleine van Zuylen, Arman Cohan, and Hannaneh Hajishirzi. 2020. [Fact or fiction: Verifying scientific claims](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pages 7534–7550. Association for Computational Linguistics.
- Huaiyu Wan, Yutao Zhang, Jing Zhang, and Jie Tang. 2019. [AMiner: Search and mining of academic social networks](#). *Data Intelligence*, 1(1):58–76.
- Peiyi Wang, Lei Li, Liang Chen, Zefan Cai, Dawei Zhu, Binghuai Lin, Yunbo Cao, Lingpeng Kong, Qi Liu,

- Tianyu Liu, and Zhifang Sui. 2024. [Large language models are not fair evaluators](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 9440–9450. Association for Computational Linguistics.
- An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, et al. 2025. [Qwen3 technical report](#). *arXiv preprint arXiv:2505.09388*.
- Yibo Zhao, Jiapeng Zhu, Ye Guo, Kangkang He, and Xiang Li. 2025. [E²GraphRAG: Streamlining graph-based RAG for high efficiency and effectiveness](#). *arXiv preprint arXiv:2505.24226*.
- Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric P. Xing, Hao Zhang, Joseph E. Gonzalez, and Ion Stoica. 2023. [Judging LLM-as-a-judge with MT-Bench and chatbot arena](#). In *Advances in Neural Information Processing Systems*, volume 36, pages 46595–46623.

A Benchmark and Evaluation Details

Benchmark. The benchmark contains public-source enterprise demands covering intelligent manufacturing, materials, energy, electronics, transportation, agriculture, medicine, and other applied technology domains. Each demand is represented as a raw natural-language description and, when available, a hybrid description with structured fields such as domain, industry, application scenario, technical constraints, target indicators, expected resource types, and technology tags.

Controlled input. For the controlled main comparison, all methods use the hybrid demand representation; raw natural-language inputs are used only in the demand representation ablation. For each method, we collect the top 10 resources and pool them by demand and resource identity before blind judging. Relevance, evidence support, and actionability are scored from 0 to 3; a resource is treated as relevant when its relevance score is at least 2.

Data construction. Academic resources are collected from AMiner and organized into scholar profiles, papers, patents, and projects. The current corpus focuses on scholars affiliated with China’s Project 985 and Project 211 universities and includes their outputs from 2016–2026. Scholars are filtered by publication count and H-index to retain active researchers with sufficient academic evidence. Each paper, patent, or project is converted into a typed resource chunk and indexed by the GraphRAG pipeline. Enterprise demands are compiled from public-source technology-transfer descriptions and rewritten as controlled benchmark inputs for evaluation. The benchmark is intended as a preliminary public-source evaluation set rather than an official expert-annotated benchmark, and sensitive private enterprise documents are not included.

B Corpus Statistics

Table 5 reports corpus-level statistics for scholars and academic outputs by field. The same hybrid demands are routed to papers, scholars, patents, and projects in the resource-type analysis. Papers provide the strongest evidence-rich matches, while the other resource types support expert discovery, intellectual-property exploration, and project-level opportunity analysis.

Field	Scholars	Papers	Patents	Projects
AI & CS	8,879	922,046	204,754	14,748
Materials Science	12,272	1,194,246	242,721	23,670
Biology	2,203	193,973	32,381	4,855
Electronic Engineering	2,312	241,501	69,839	3,637
Mechanical Engineering	2,142	199,961	57,048	3,700
Civil Engineering	733	62,418	13,447	1,494
Environment & Energy	719	61,299	8,954	1,210
Agriculture & Food Science	474	38,393	6,975	775
Medicine & Public Health	1,678	149,983	8,978	2,415
Others	2,479	190,456	37,267	3,053
Total	33,891	3,254,276	682,364	59,557

Table 5: Corpus statistics by research field.

C Supplementary Evaluation Visualizations

Figure 3 visualizes the bootstrap confidence intervals, demand representation ablation, resource-type profile, and report-quality evaluation discussed in Section 5.

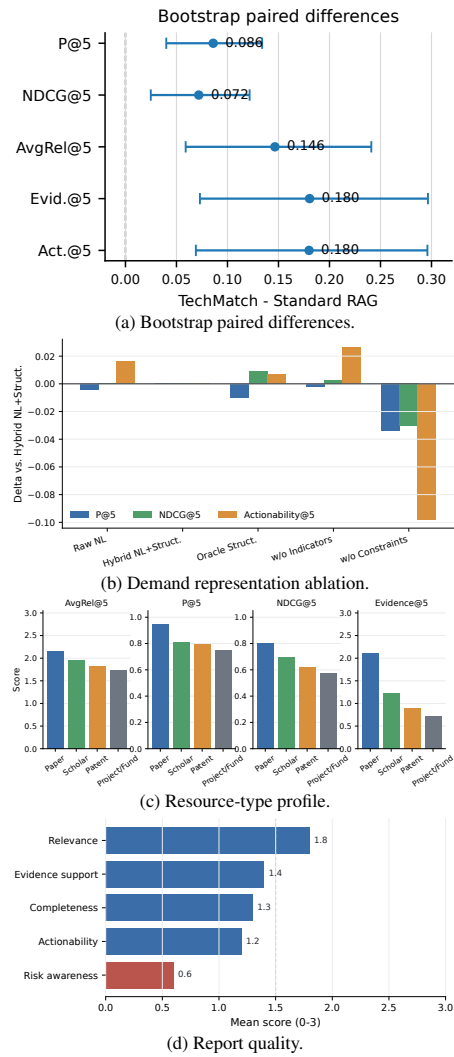


Figure 3: Supplementary evaluation visualizations.

D Demonstration Workflow

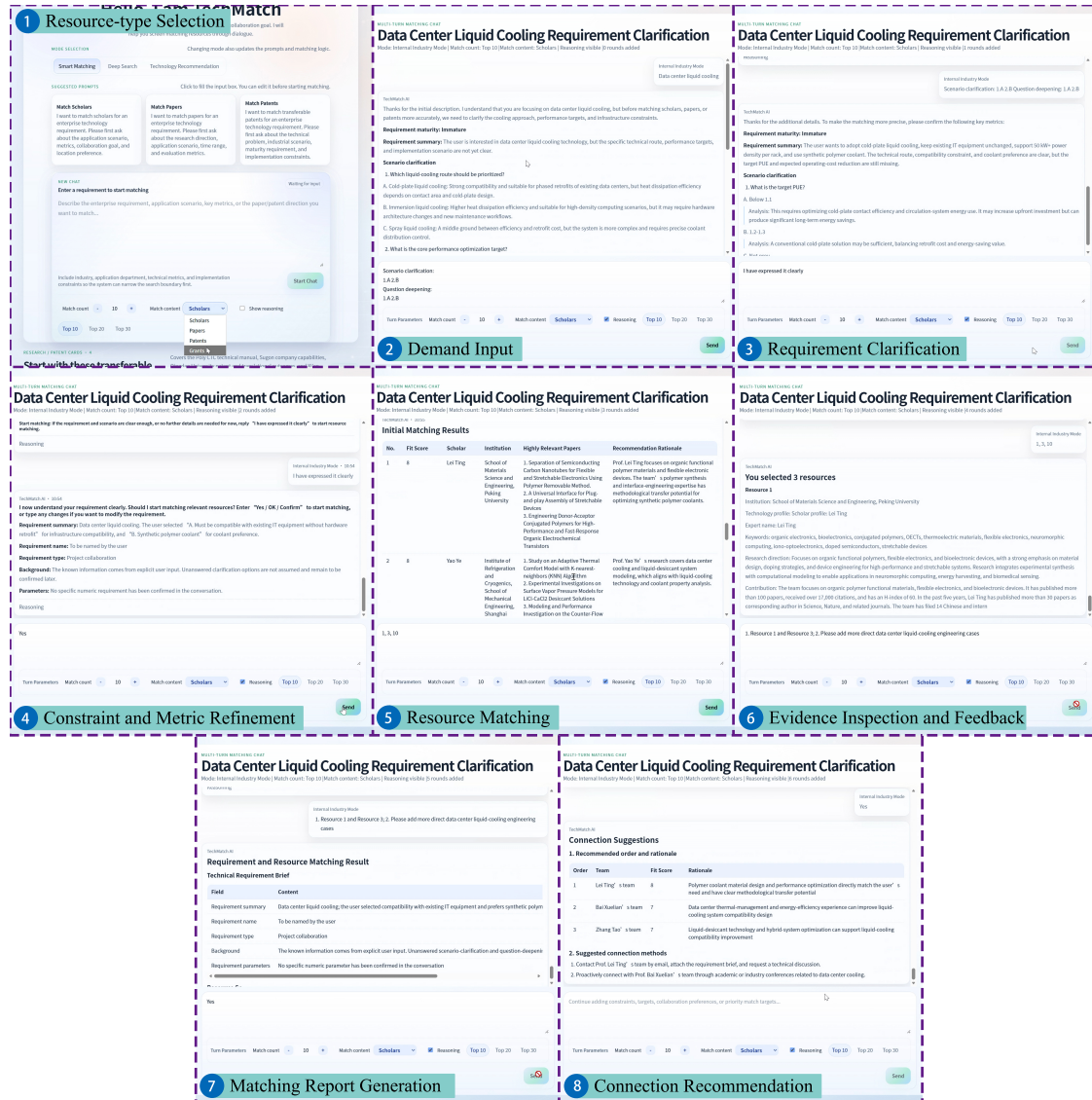


Figure 4: Eight-step workflow of TECHMATCH, from demand clarification to resource matching and connection recommendations.

E Ethics Statement

Technology-transfer matching systems can influence resource allocation among universities, enterprises, experts, and regions. To reduce automation bias, TECHMATCH presents evidence-backed recommendations as decision support rather than as final decisions. Sensitive enterprise documents, unpublished research outputs, and intellectual-property materials should be processed with appropriate access controls, data-use agreements, and audit logs. The system should also expose uncertainty and allow users to challenge or correct generated explanations.