Topic 1. AI-Supported Literature Search

Yong Cao

Postdoc
University of Tübingen
yong.cao@uni-tuebingen.de

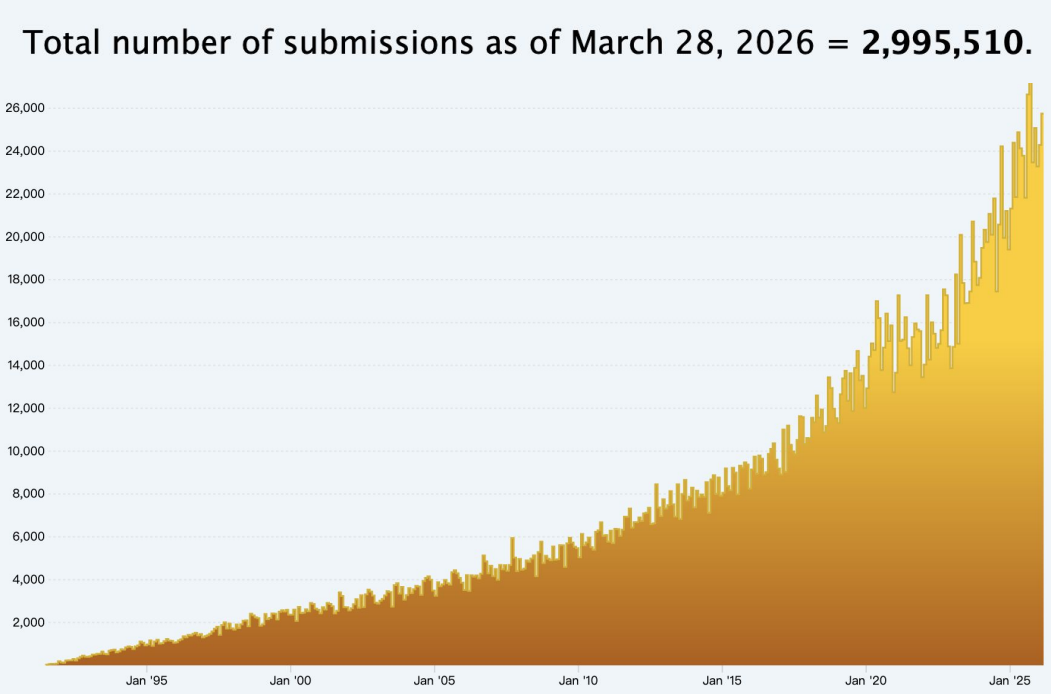
Agenda

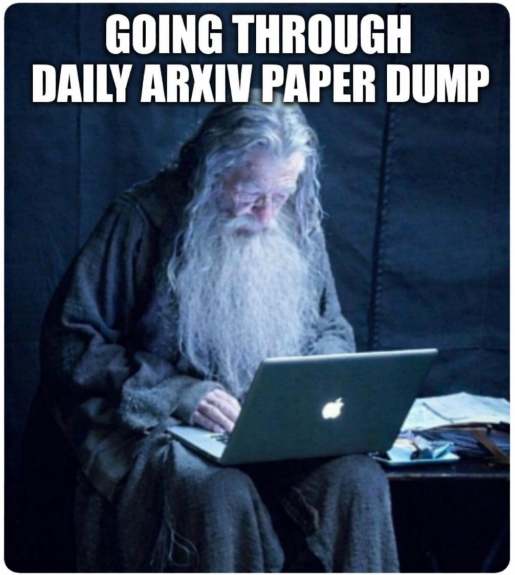
- ❏ Background
- ❏ AI-Enhanced Search System
- ❏ Paper Chat and Scientific QA
- ❏ Graph Based System
- ❏ Recommendation System
- ❏ Takeaway

Agenda

- ❑ **Background**
- ❑ AI-Enhanced Search System
- ❑ Paper Chat and Scientific QA
- ❑ Graph Based System
- ❑ Recommendation System
- ❑ Takeaway

Research Growth



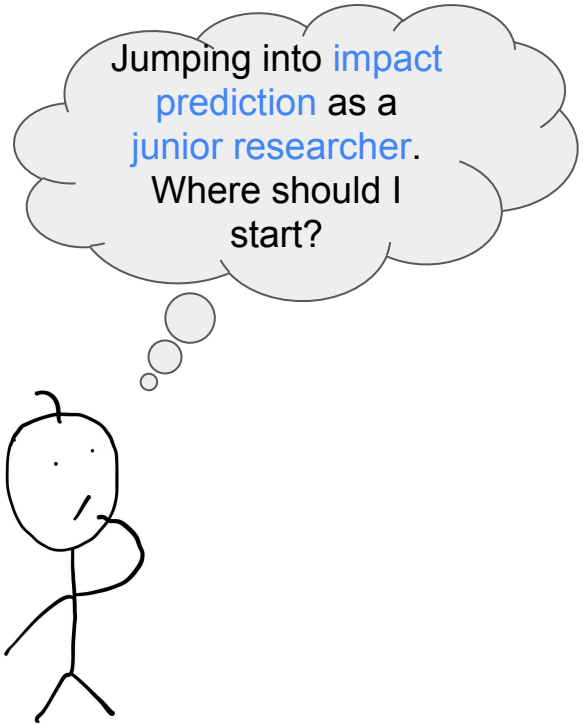


Arxiv Monthly Submission from 1995-2025, Arxiv, 2026.3.28.

*The real problem is **not information overload**, it's **filter failure**.*

— Clay Shirky

AI Changes the Way of Scholar's Search



1. Broad Keyword Input

Google 学术搜索 | impact prediction in AI domain

文章 — 找到约 3,450,000 条结果 （用时0.20秒）

时间不限 | 2026以来 | 2025以来 | 2022以来 | 自定义范围...

按相关性排序 | 按日期排序

不限语言 | 中文网页 | 简体中文网页

类型不限 | 评论性文章

包括专利 | 包含引用

创建快讯

Exploring the impact of artificial intelligence: Prediction versus judgment [PDF] nber.org
A Agrawal, JS Gans, A Goldfarb - Information Economics and Policy, 2019 - Elsevier
… -making as prediction technology improves. … prediction impacts the returns to different types of judgment in opposite ways. Hence, not all human judgment will be a complement to AI. …
☆ 保存 99 引用 被引用次数：546 相关文章 所有 20 个版本 »

AI-based protein structure prediction in drug discovery: impacts and challenges [PDF] acs.org
M Schauperl, RA Denny - Journal of Chemical Information and …, 2022 - ACS Publications
… predict individual domains with high confidence is an impressive step toward solving the protein prediction … However, we believe that this is just the starting point for AI-based methods to …
☆ 保存 99 引用 被引用次数：135 相关文章 所有 3 个版本

AI impact on various domain: an overview [PDF] ssrn.com
SC Kumain, K Kumain, P Chaudhary - International Journal of …, 2020 - papers.ssrn.com
… But in strong AI machines can do the calculation and predict the solution for the queries based on Intelligence provided during the training step. In general programming approach, …
☆ 保存 99 引用 被引用次数：7 相关文章 所有 3 个版本 »

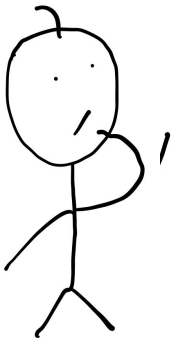
Domains and Impact of AI in IoT for Environmental Monitoring and Management: A Review [PDF] researchgate.net
DS Hammadi, DB Abdullah - 2025 International Conference on …, 2025 - ieeexplore.ieee.org
… The study in [62] used ML to predict heavy metal levels in soil faster than traditional methods, accelerating the soil rehabilitation process. The prediction of soil metal levels via AI is …
☆ 保存 99 引用 被引用次数：2 相关文章 所有 3 个版本

Using artificial intelligence tools in the judicial domain and the evaluation of their impact on the prediction of judgments [PDF] sciencedirect.com
S Yassine, M Esghir, O Ibrihich - Procedia Computer Science, 2023 - Elsevier
… artificial intelligence tools and measure their impact on the performance of the judicial system. We study its methods, its challenges, and its potential impacts, … results for the prediction of …
☆ 保存 99 引用 被引用次数：29 相关文章 所有 2 个版本

2. Manual Sifting required for relevance

AI Changes the Way of Scholar's Search

Jumping into impact prediction as a junior researcher. Where should I start?



1. Detailed Query

Impact prediction in AI domain, focusing on papers published at top AI conferences over the past three years

Pro · 2 steps · 20 sources

AI-based scientific impact prediction in recent literature

Recent work shows **growing and increasingly sophisticated AI methods for predicting scientific impact**, including in AI itself, but only a small subset focuses specifically on top AI conferences.

Main paradigms in scientific-impact prediction

1. Impact of ideas, topics, and papers

- **Knowledge-graph forecasting:** Gu & Krenn build an evolving, 21M-paper semantic+citation graph and use ML to forecast the future growth of research topics with high accuracy (AUC > 0.9), aiming to predict impact *before papers are written* [1].
- **Comprehensive survey:** Xia et al. review impact prediction across four entities (papers, scholars, venues, institutions), cataloging tasks (e.g., citation prediction, h-index forecasting), six feature groups (paper-, author-, venue-, institution-, network-, altmetrics-related), and methods (statistical, ML, deep, graph-based) [2].
- **Paper long-term impact:** Abramo et al. show that integrating early citations, Mendeley readership, journal impact factor, authorship diversity, and other non-scientific factors improves long-term citation prediction over citation-only models [8]. Content/NLP features and collaboration-network features also enhance accuracy [8].

2. Scholar and career impact

- FPIPN (graph neural network on heterogeneous academic graphs) predicts a scholar's **future publication impact after career moves** using hierarchical attention and knowledge distillation, outperforming prior models [4].

3. Methods, signals, and limitations

Aspect	Recent insights	Citations
Dominant targets	Future citation counts, topic growth, h-index, institutional/venue impact	[1] [2] [4] [8]
Strong predictors	Early citations, readership, venue quality, collaboration networks	[2] [8]
Advanced models	Deep learning, GNNs, graph-based temporal forecasting	[1] [2] [4] [8]

Ask a follow up...

Pro · Sources · Filter

2. More accurate papers

Results

1 **Forecasting high-impact research topics via machine learning on evolving knowledge graphs**
KEY TAKEAWAY · Machine learning on an evolving knowledge graph can accurately predict the impact of new research ideas, potentially aiding future artificial muses in inspiring new scientific ideas.
2024 · 21 citations · Xuemei Gu et al. · *Machine Learning: Science and Te...* · PDF

2 **A review of scientific impact prediction: tasks, features and methods**
KEY TAKEAWAY · This paper proposes a new framework for scientific impact prediction, categorizes input features into six groups, and classifies forecasting methods into mathematical statistics, traditional machine learning, deep learning, and graph-based.
LITERATURE REVIEW
2022 · 31 citations · Wanjun Xia et al. · *Scientometrics*

3 **Predictive modeling of climate change impacts using Artificial Intelligence: a review for equitable governance and sustainable outcome**
KEY TAKEAWAY · AI-based predictive models can accurately predict climate change impacts, providing valuable insights for scientists, policymakers, and communities.
LITERATURE REVIEW
2025 · 18 citations · K. Ukoba et al. · *Environmental Science and Pollut...* · PDF

4 **Where is the Next Step? Predicting the Scientific Impact of Research Career**
KEY TAKEAWAY · The Future Publication Impact Prediction Network (FPIPN) effectively predicts research scholar's scientific impact using a heterogeneous academic graph and hierarchical attention mechanism.
2025 · 0 citations · Hefu Zhang et al. · *IEEE Transactions on Big Data*

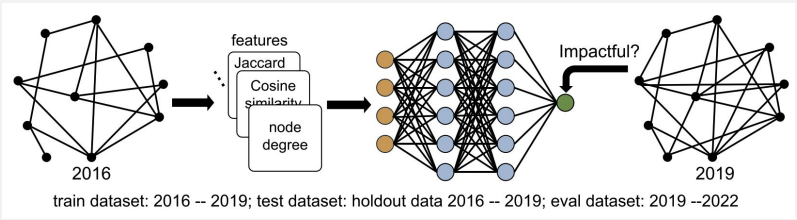
5 **Implementing Artificial Intelligence Techniques to Predict Environmental Impacts: Case of Construction Products**
KEY TAKEAWAY · AI techniques can accurately predict environmental impacts of construction products using environmental product declarations, aiding Life Cycle Assessment practitioners and verifiers.
2022 · 44 citations · Anish Koyamparambath et al. · *Sustainability* · PDF

6 **Advances in Blast-Induced Impact Prediction—A Review of Machine Learning Applications**
KEY TAKEAWAY · Machine learning methods are better predictors of blast impacts than empirical models, but currently mainly applied in academic research.
LITERATURE REVIEW
2021 · 50 citations · N. Dumakor-Dupey et al. · *Minerals* · PDF

3. Summarization Report

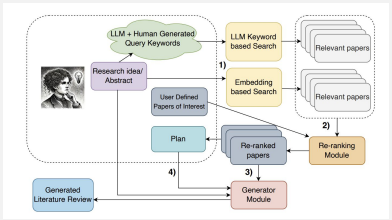
Why AI/LLMs for Literature Search?

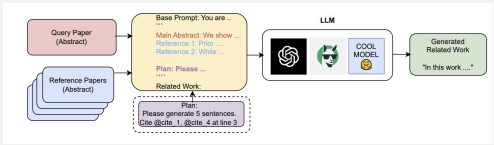
Uncover Emerging Trends



e.g., Forecasting high-impact research topics

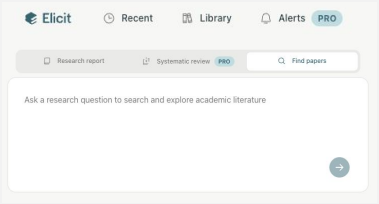
Semantic & Relation Understanding

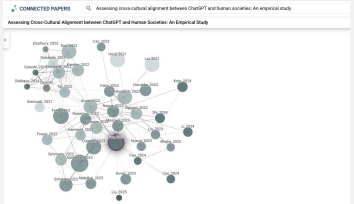


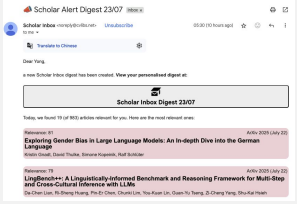


e.g., LLMs for literature review

Boost Research Efficiency and Increase Creativity



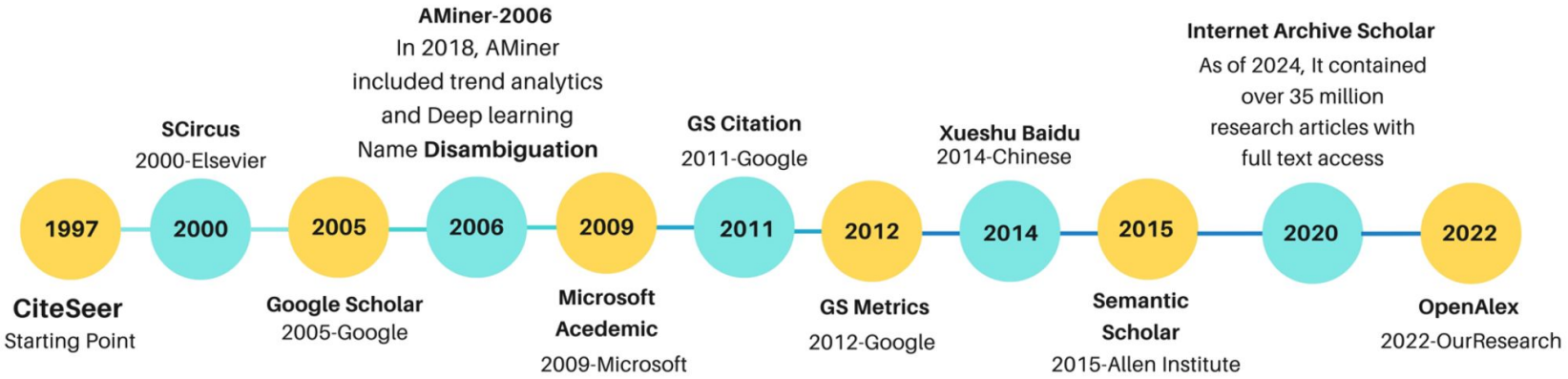




e.g., Scientific QA, Knowledge Graph Enhanced Search, Semantic Search, personalized Recommendation

Gu et al., Forecasting high-impact research topics via machine learning on evolving knowledge graphs, Machine Learning: Science and Technology, 2025.
Agarwal et al., LitLLMs, LLMs for Literature Review: Are we there yet? Transactions on Machine Learning Research, 2025.
Elicit, Connected Papers, Scholar Inbox, ResearchTrend.ai

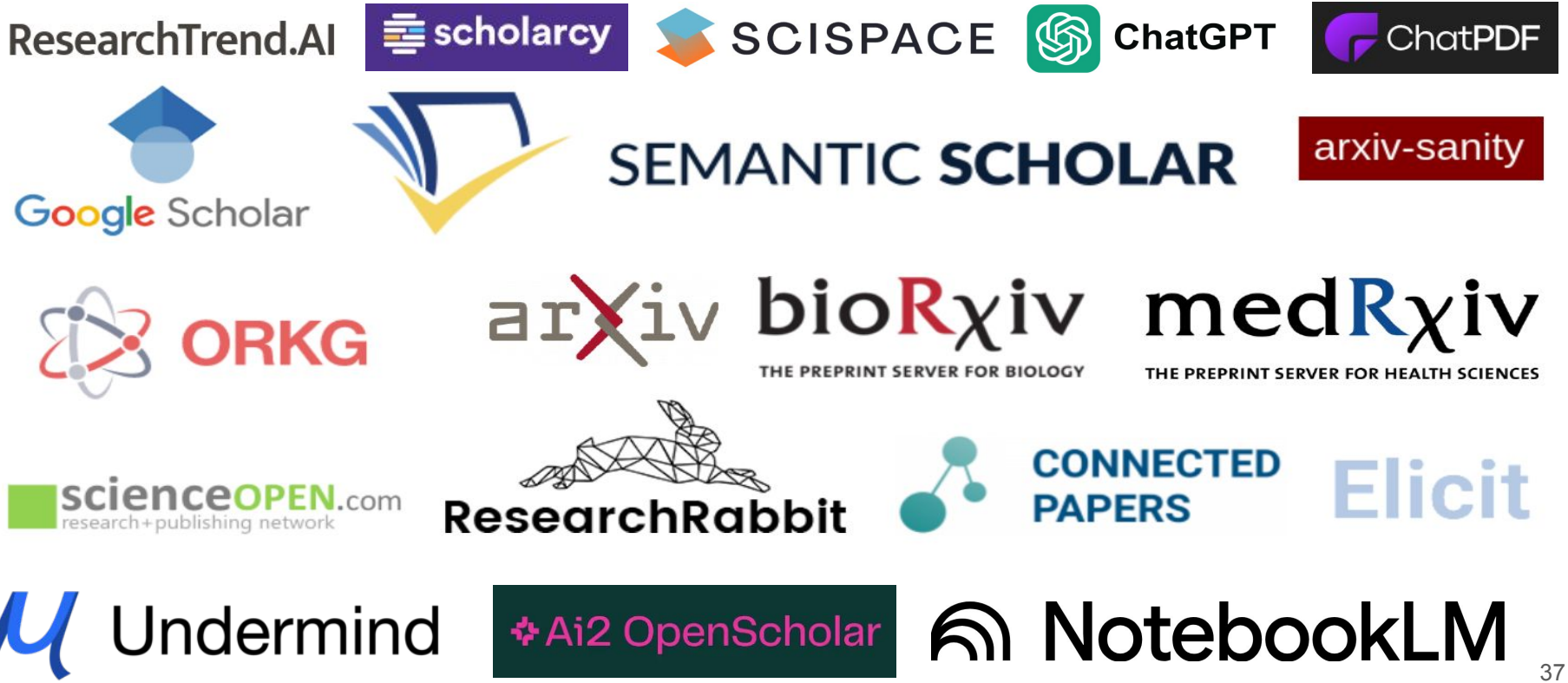
Evolution of academic search engines



A shift from simple document retrieval to more intelligent, structured, and open academic ecosystems.

Khalid et al., Comprehensive review of academic search systems: evolution, analysis, and future research directions, Social Network Analysis and Mining, 2025.

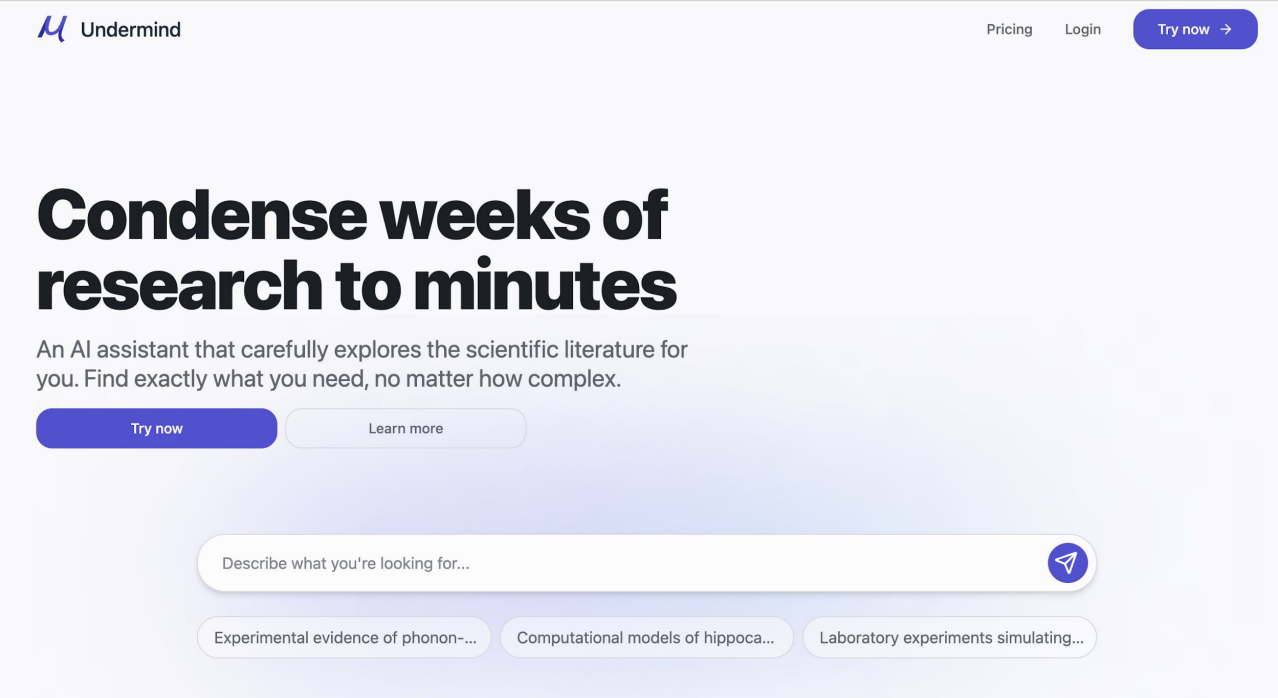
Data Source


ResearchTrend.AI
scholarcy
SCISPACE
ChatGPT
ChatPDF
Google Scholar
SEMANTIC SCHOLAR
arxiv-sanity
ORKG
arXiv
bioRxiv
THE PREPRINT SERVER FOR BIOLOGY
medRxiv
THE PREPRINT SERVER FOR HEALTH SCIENCES
scienceOPEN.com
research+publishing network
ResearchRabbit
CONNECTED PAPERS
Elicit
Undermind
Ai2 OpenScholar
NotebookLM

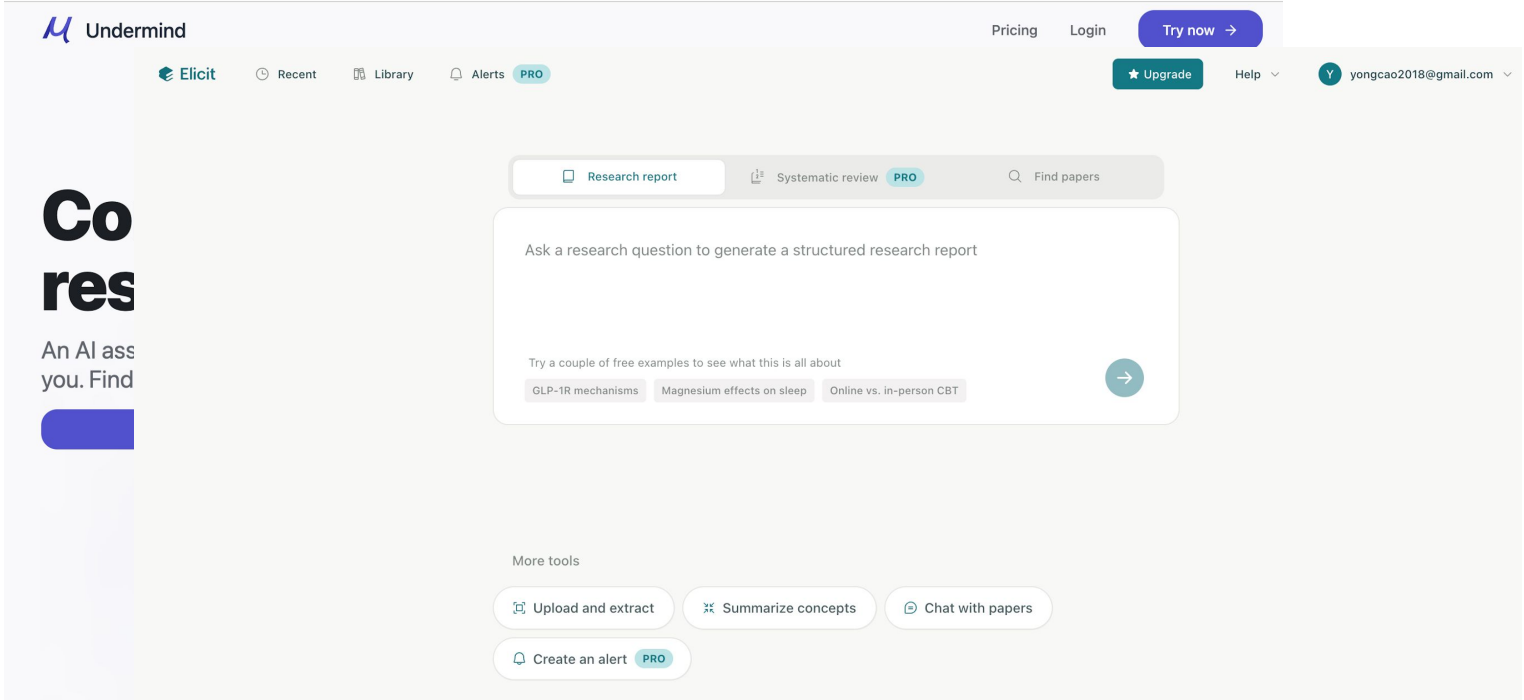
Agenda

- ❑ Background
- ❑ **AI-Enhanced Search System**
- ❑ Paper Chat and Scientific QA
- ❑ Graph Based System
- ❑ Recommendation System
- ❑ Takeaway

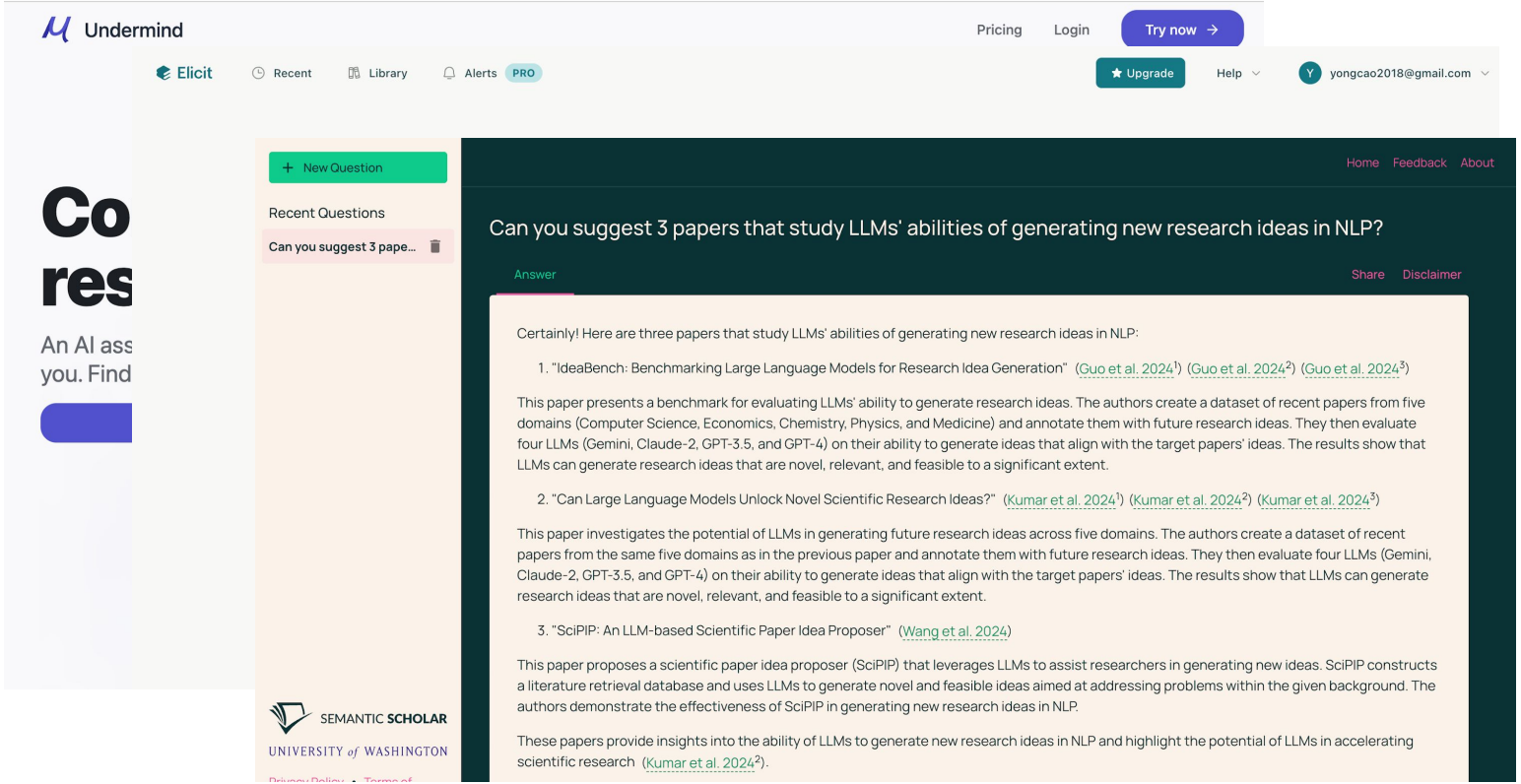
AI Enhanced Search System



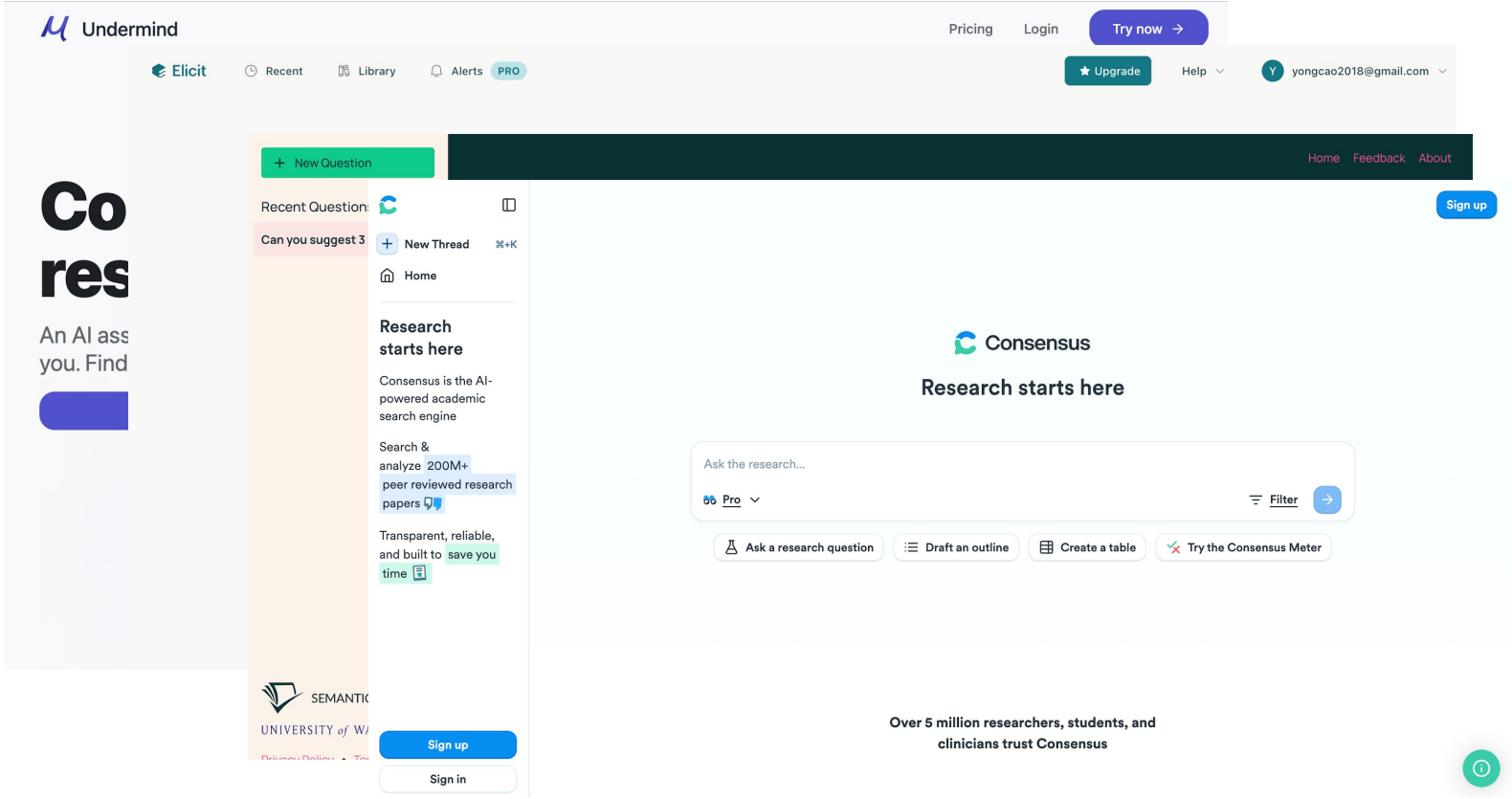
AI Enhanced Search System



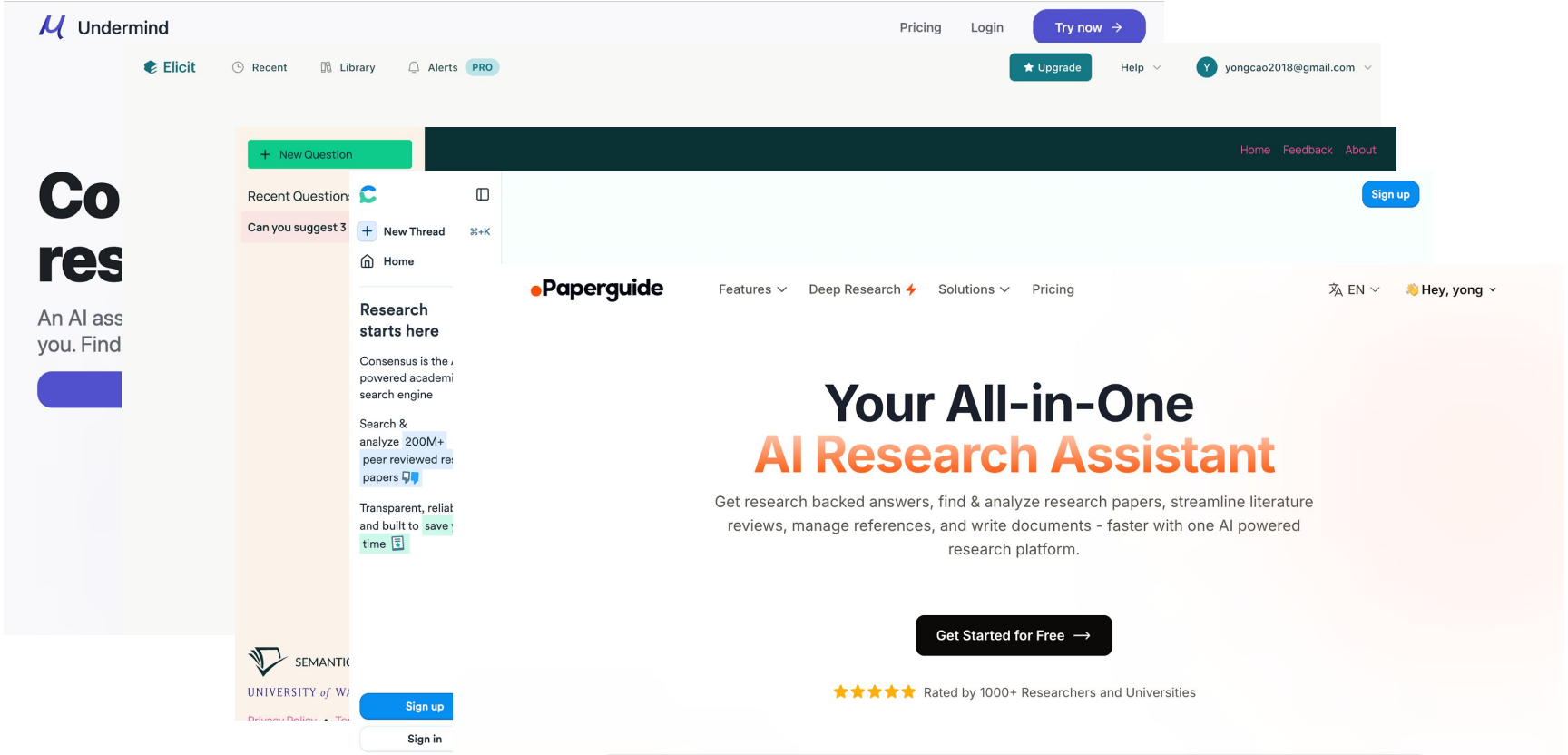
AI Enhanced Search System



AI Enhanced Search System



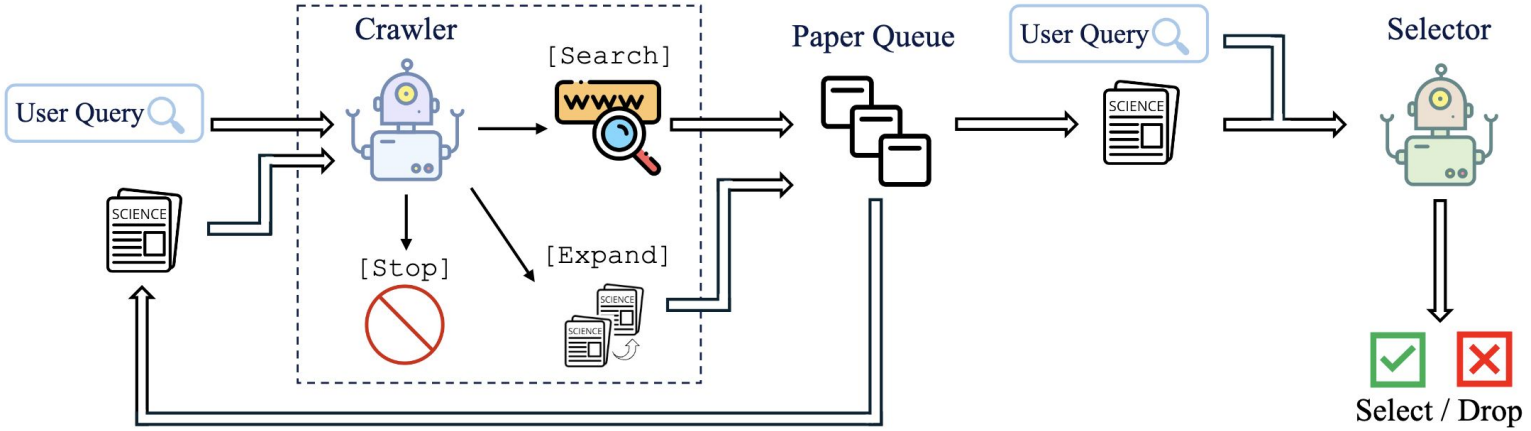
AI Enhanced Search System



How AI enhance literature search platforms?

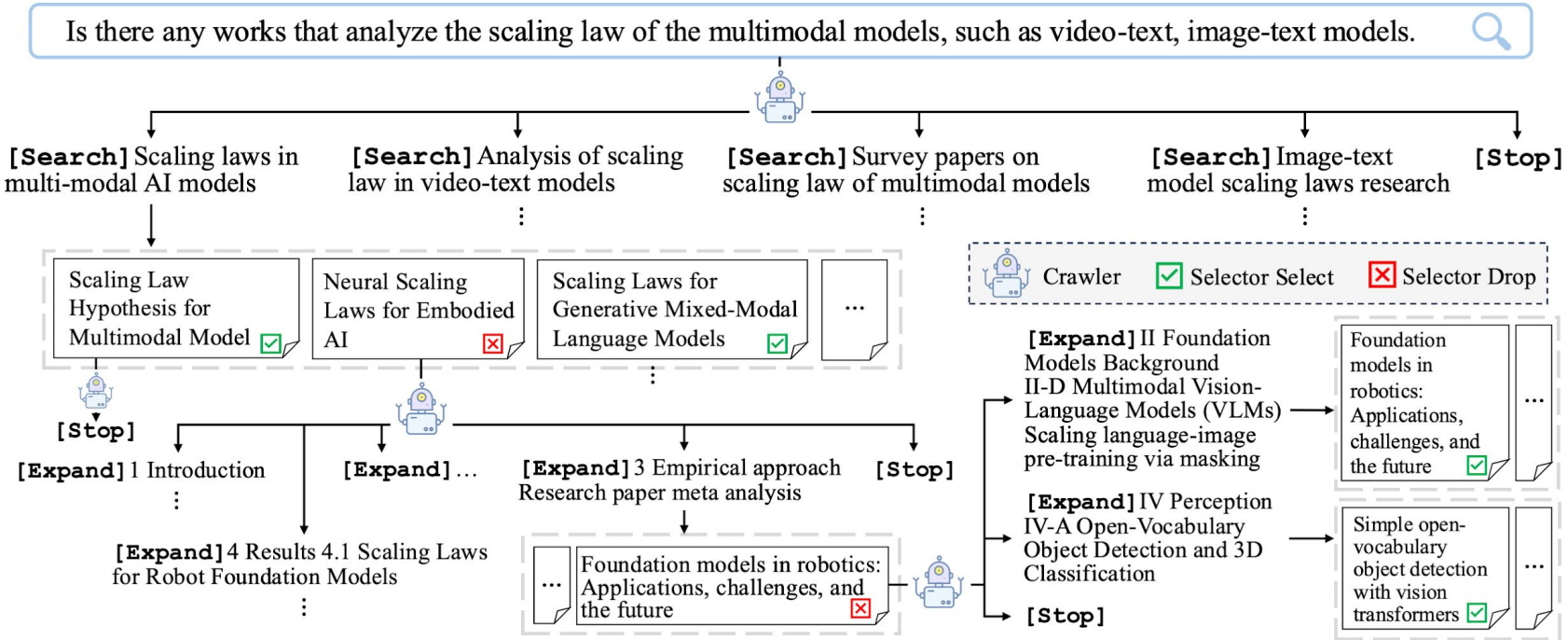
- Mimic human researchers workflows and expand search function.
- Two LLM agents: crawler and selector.

PaSa
Paper Search



He et al., PaSa: An LLM Agent for Comprehensive Academic Paper Search, ACL, 2025.

How AI enhance literature search platforms?



Performance

- Imitation Learning + Reinforcement Learning

$$\mathcal{L}_{\text{policy}}(\theta) = \mathbb{E}_{\tau' \sim \pi_\theta^{\text{old}}} \left[\min \left(\frac{\pi_\theta(a_t|s_t)}{\pi_\theta^{\text{old}}(a_t|s_t)} \hat{A}(s_t, a_t), \right. \right.$$
$$\left. \left. \text{clip}\left(\frac{\pi_\theta(a_t|s_t)}{\pi_\theta^{\text{old}}(a_t|s_t)}, 1-\epsilon, 1+\epsilon \right) \hat{A}(s_t, a_t) \right) \right]$$

$$\mathcal{L}_{\text{value}}(\phi) = \mathbb{E}_{\tau' \sim \pi_\theta^{\text{old}}} \left[\max \left(\left(\hat{\mathrm{R}}_t - \hat{\mathrm{V}}_\phi(\mathrm{s}_t) \right)^2, \right. \right.$$
$$\left. \left. \left(\hat{R}_t - \hat{V}_\phi^{\text{clip}}(s_t) \right)^2 \right) \right],$$

$$\mathcal{L}_{\text{RL}}(\theta, \phi) = \mathcal{L}_{\text{policy}}(\theta) + \eta \cdot \mathcal{L}_{\text{value}}(\phi)$$

Method	Crawler Recall	Precision	Recall	Recall@100	Recall@50	Recall@20
Google	-	-	-	0.2015	0.1891	0.1568
Google Scholar	-	-	-	0.1130	0.0970	0.0609
Google with GPT-4o	-	-	-	0.2683	0.2450	0.1921
ChatGPT*	-	0.0507	0.3046	-	-	-
GPT-o1	-	0.0413	0.1925	-	-	-
PaSa-GPT-4o	0.7565	**0.1457**	0.3873	-	-	-
PaSa-7b	**0.7931**	**0.1448**	**0.4834**	**0.6947**	**0.6334**	**0.5301**
PaSa-7b-ensemble	**0.8265**	0.1410	**0.4985**	**0.7099**	**0.6386**	**0.5326**

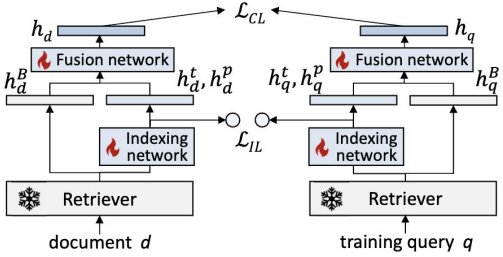
Performance

- Imitation Learning + Reinforcement Learning

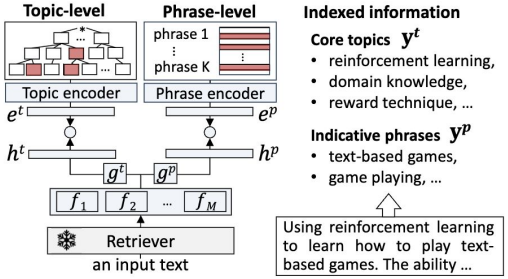
Method	Crawler Recall	Precision	Recall	Recall@100	Recall@50	Recall@20
Google	-	-	-	0.2015	0.1891	0.1568
Google Scholar	-	-	-	0.1130	0.0970	0.0609
Google with GPT-4o	-	-	-	0.2683	0.2450	0.1921
ChatGPT*	-	0.0507	0.3046	-	-	-
GPT-o1	-	0.0413	0.1925	-	-	-
PaSa-GPT-4o	0.7565	**0.1457**	0.3873	-	-	-
PaSa-7b	**0.7931**	**0.1448**	**0.4834**	**0.6947**	**0.6334**	**0.5301**
PaSa-7b-ensemble	**0.8265**	0.1410	**0.4985**	**0.7099**	**0.6386**	**0.5326**

Taxonomy-guided Index Construction

- The Problem: Beyond Surface-Level Text Matching
- TaxoIndex Framework
- Step 1: Constructing the Semantic Index
- Step 2: Index-grounded AI Training (Fine-tuning)
- Step 3: Enhanced Retrieval Process



(a) Index-grounded fine-tuning



(b) Index learning with the indexing network

		CSFCube						DORIS-MAE					
		N@5	N@10	M@5	M@10	R@50	R@100	N@5	N@10	M@5	M@10	R@50	R@100
	BM25	0.307	0.310	0.088	0.134	0.504	0.635	0.354	0.330	0.079	0.107	0.490	0.669
SPECTER-v2	no Fine-Tuning	0.352	0.337	0.108	0.151	0.524	0.680	0.385	0.360	0.079	0.113	0.551	0.709
	FFT	0.372	0.368	0.123	0.169	0.576	0.692	0.408	0.387	0.084	0.122	0.562	0.736
	aFT	0.378	0.344	0.119	0.160	0.578	0.696	0.400	0.372	0.080	0.115	0.558	0.714
	FFT w/ GRF	0.331	0.317	0.112	0.152	0.561	0.705	0.400	0.379	0.087	0.123	**0.586**	**0.756**
	FFT w/ ToTER	0.406	0.375	0.135	0.179	0.591	0.710	0.423	0.394	0.091	0.128	0.563	0.736
	JTR	0.379	0.352	0.118	0.157	0.598	0.699	0.395	0.380	0.080	0.118	0.548	0.713
	TaxoIndex	0.458$^{\dagger *}$	0.417$^{\dagger *}$	0.144$^{\dagger *}$	0.198$^{\dagger *}$	**0.633**$^{\dagger *}$	0.741$^{\dagger *}$	0.447$^{\dagger *}$	0.421$^{\dagger *}$	0.104$^{\dagger *}$	0.144$^{\dagger *}$	0.578†	**0.756**†
	TaxoIndex ++	**0.469**$^{\dagger *}$	**0.426**$^{\dagger *}$	**0.158**$^{\dagger *}$	**0.209**$^{\dagger *}$	0.621$^{\dagger *}$	**0.746**$^{\dagger *}$	**0.449**$^{\dagger *}$	**0.424**$^{\dagger *}$	**0.105**$^{\dagger *}$	**0.145**$^{\dagger *}$	0.581†	0.751†

Kang et al., Taxonomy-guided Semantic Indexing for Academic Paper Search, EMNLP, 2024.

Deep Research -- ChatGPT / Gemini

Model data and training

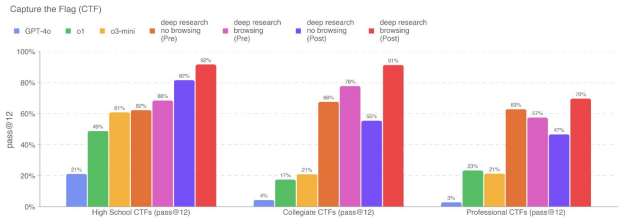
- Browsing datasets
- Graded against the ground truth answers or chain-of-thought model
- Safety datasets from o1 training

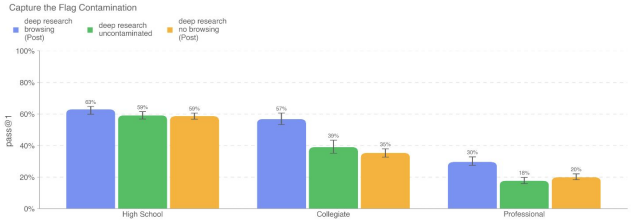
Risk and mitigation

- Prompt Injections
- Disallowed Content
- Privacy
- Ability to Run Code
- Bias
- Hallucinations

Overall Evaluation: Medium

OpenAI





Gemini

Planning

Deep Research transforms your prompt into a personalized multi-point research plan

Searching

Deep Research autonomously searches and deeply browses the web to find relevant, up-to-date information

Reasoning

Deep Research shows its thoughts as it reasons over information gathered iteratively and thinks before making its next move

Reporting

Deep Research provides comprehensive custom research reports with more detail and insights, generated in minutes and available as an Audio Overview, saving you hours of time

OpenAI, Deep Research System Card, 2025.
Gemini, Gemini Deep Research, 2026.

Takeaway – AI Enhanced Search System

- Main Function
  - Optimize scholarly information retrieval
  - Context-aware, semantically rich, and personalized search results

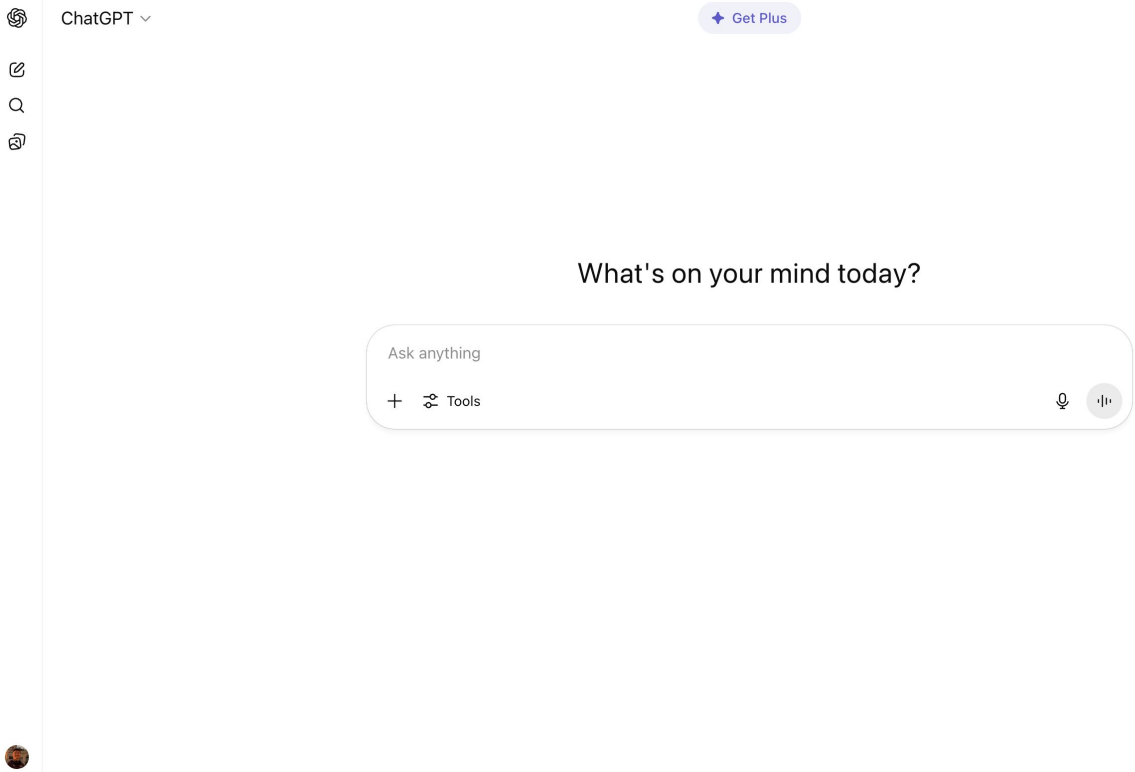
- Key Techniques
  - LLM-based agents
  - Embedding-based Retrieval
  - Personalization

- Challenges
  - Data heterogeneity
  - Limited handling of complex scientific content

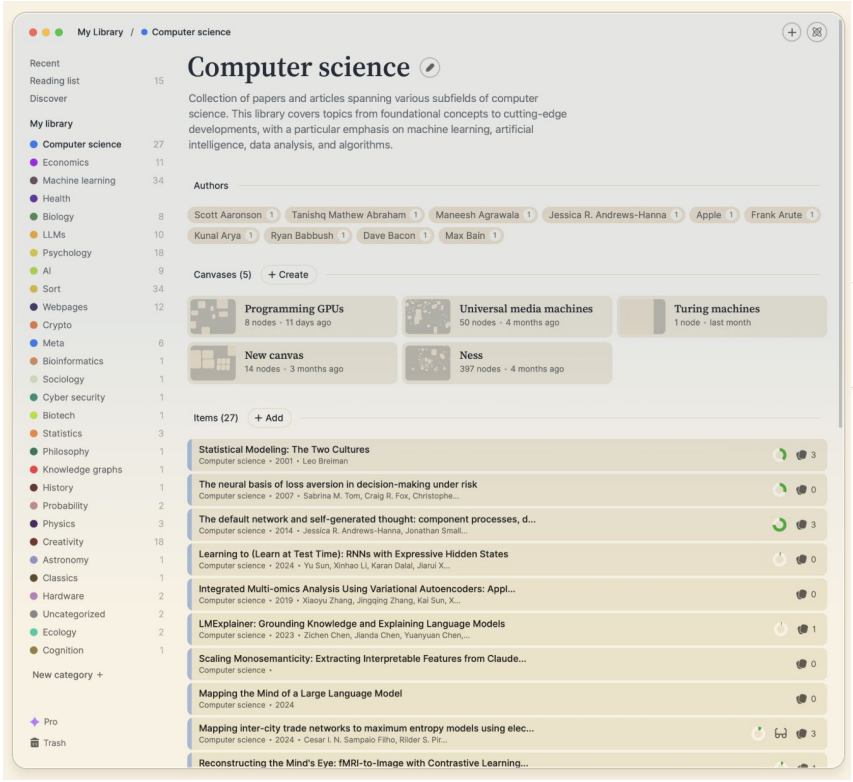
Agenda

- Background
- AI-Enhanced Search System
- **Paper Chat and Scientific QA**
- Graph Based System
- Recommendation System
- Takeaway

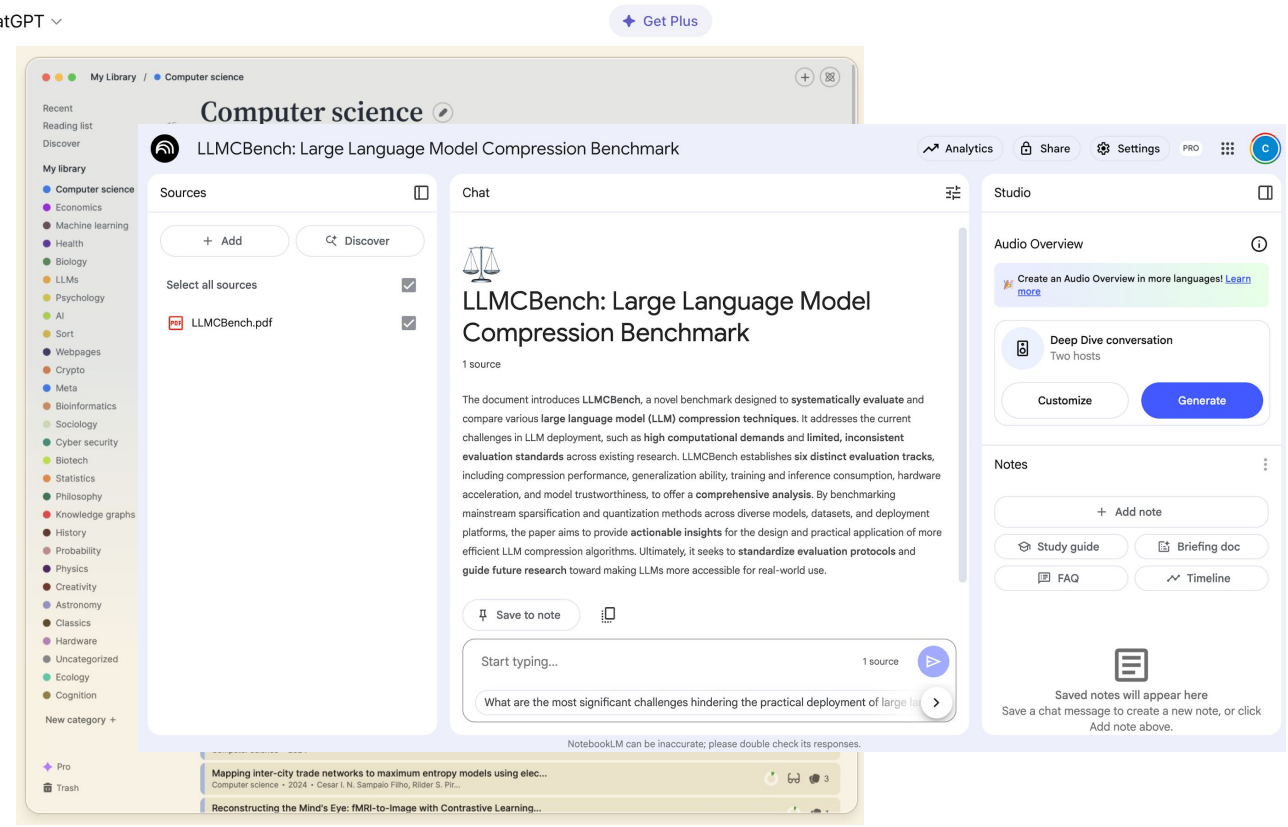
Paper Chat and Scientific QA



Paper Chat and Scientific QA

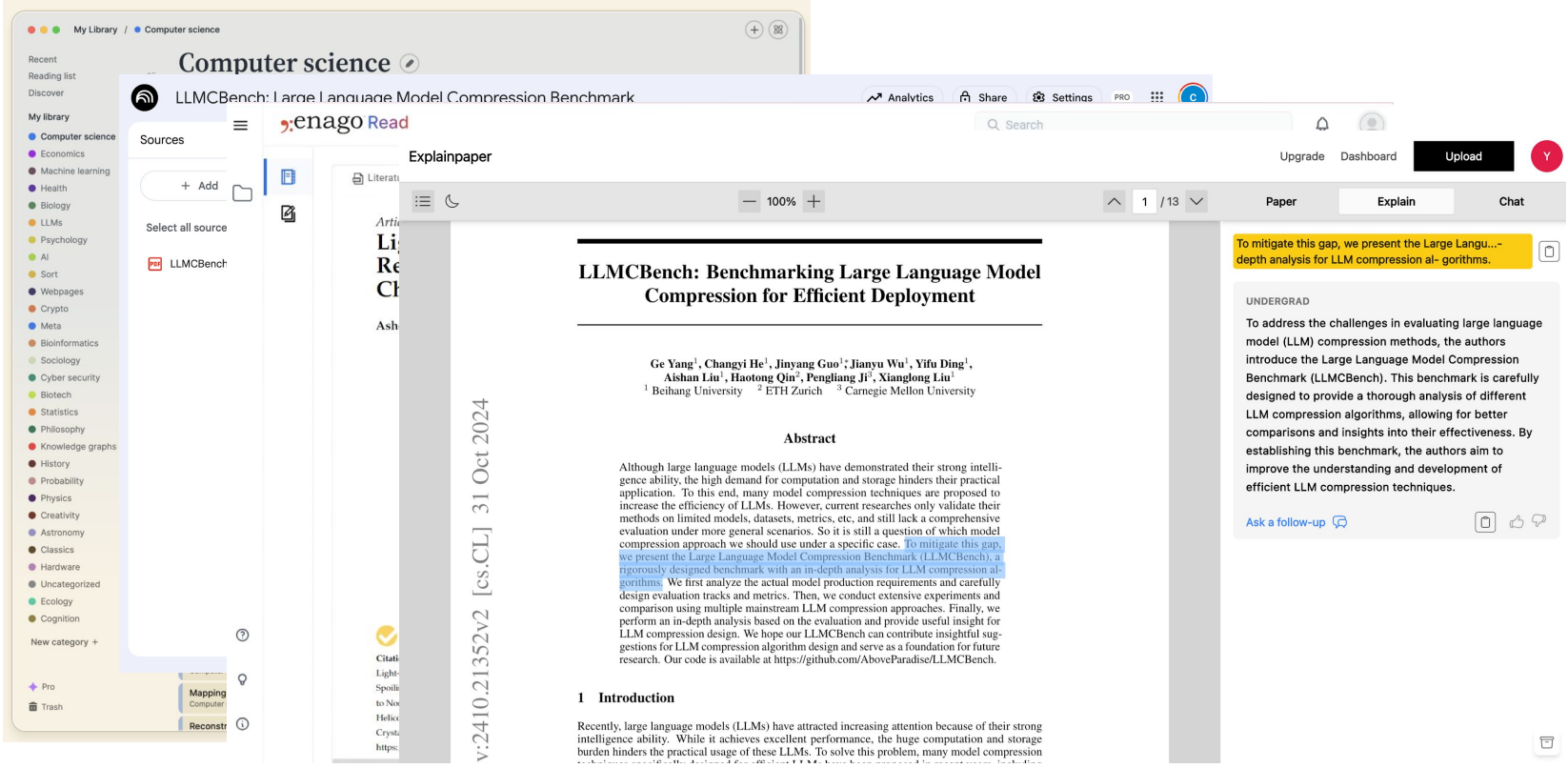


Paper Chat and Scientific QA



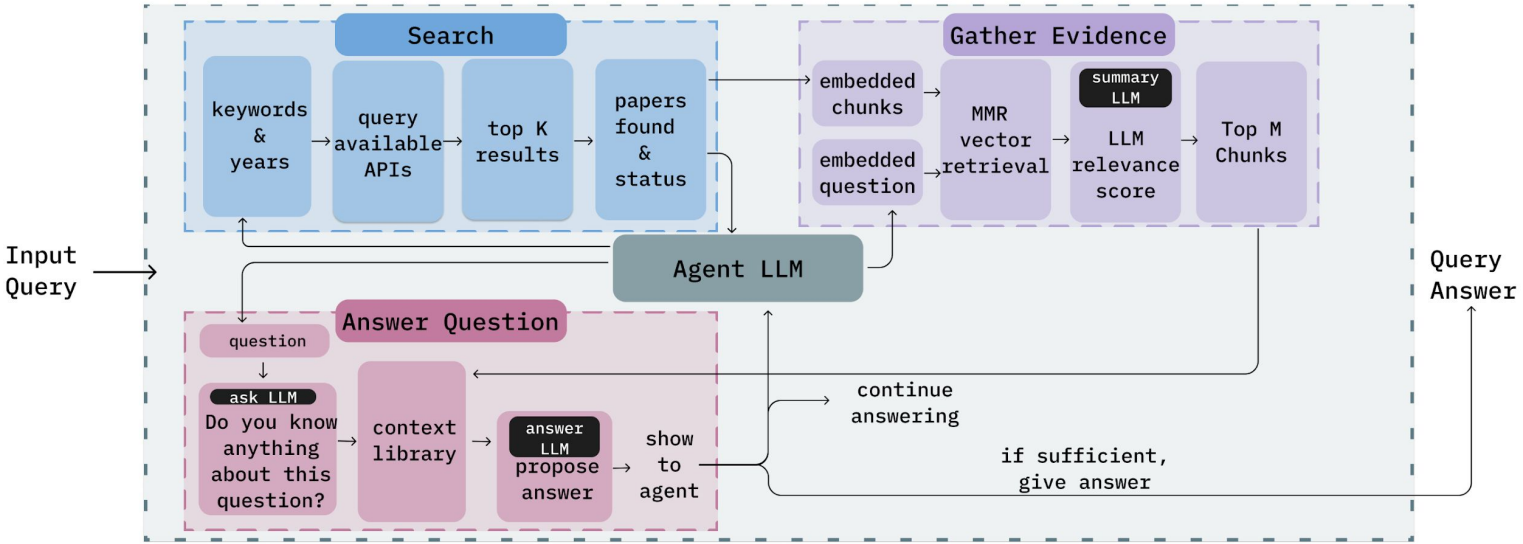
Paper Chat and Scientific QA

Paper Chat and Scientific QA



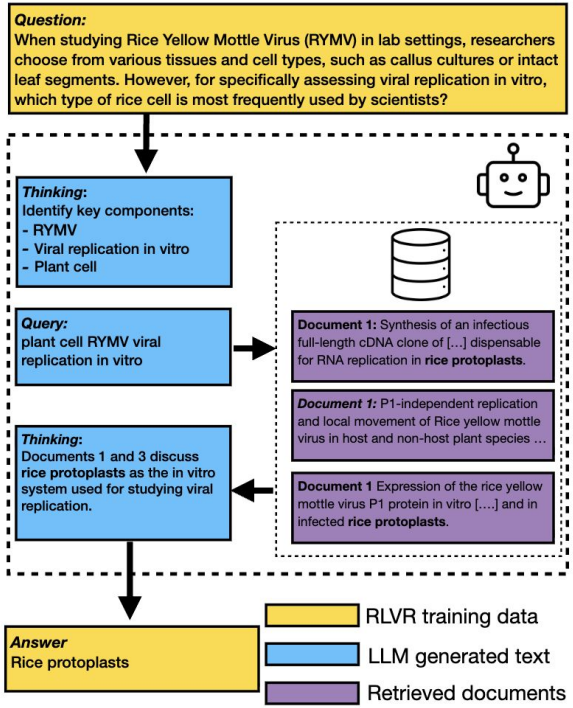
PaperQA – An Agent-Based RAG System

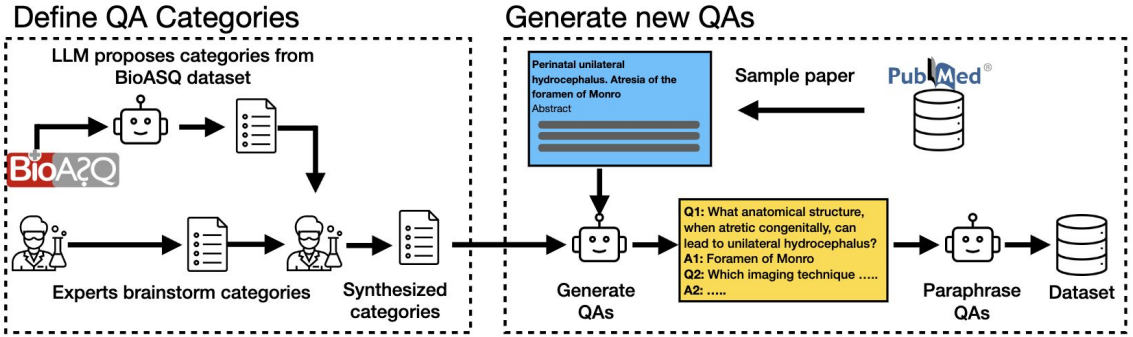
- Aims to address the hallucinations and knowledge update.
- Key Components:
  - Search, Gather Evidence, Answer Question



L´ala et al., PaperQA: Retrieval-Augmented Generative Agent for Scientific Research, arXiv, 2023.

Reinforcement Learning Helps

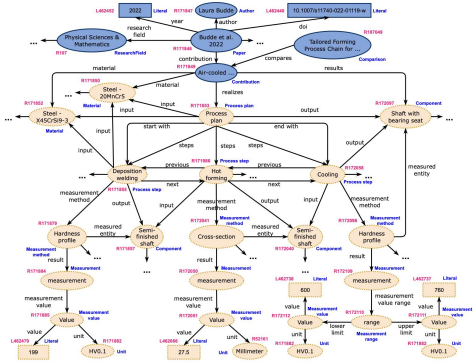




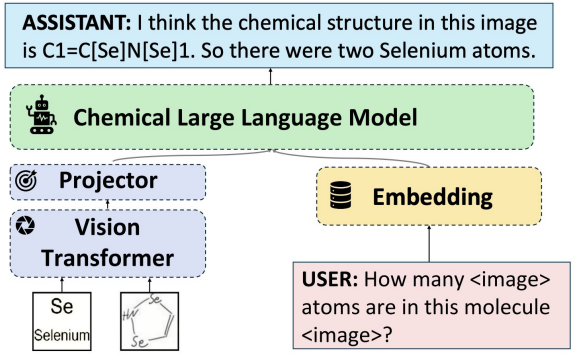
	PaperSearchQA	BioASQ
Qwen2.5-3b-Instruct		
Direct	16.7	15.8
CoT	20.3	16.5
RAG	32.0	30.0
Search-o1	30.8	29.4
PaperQA2	32.4	33.1
SearchR1	41.6	35.5
Qwen2.5-7b-Instruct		
Direct	27.5	24.9
CoT	29.7	23.4
RAG	36.5	29.7
Search-o1	36.5	31.5
PaperQA2	37.1	32.8
SearchR1	51.0	44.8

Burgess et al.,PaperSearchQA: Learning to Search and Reason over Scientific Papers with RLVR, EACL, 2026.

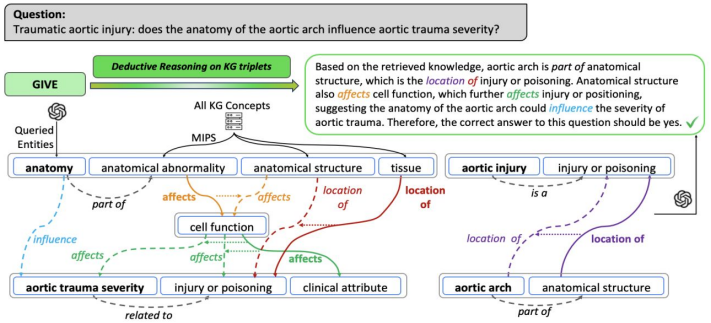
Mostly focus on…



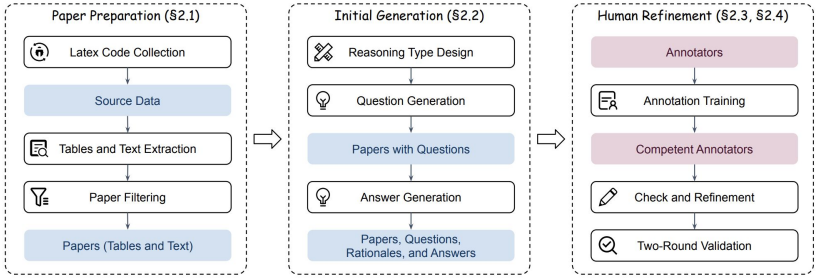
KG Integration



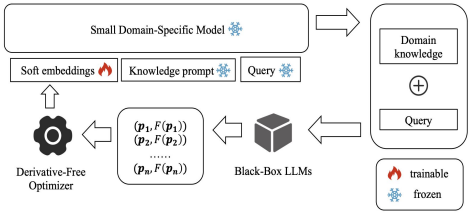
Multi-domal QA



Complex Reasoning



Benchmarking



Domain knowledge

Auer et al., The SciQA Scientific Question Answering Benchmark for Scholarly Knowledge, Scientific Reports, 2023.
Li et al., ChemVLM: Exploring the Power of Multimodal Large Language Models in Chemistry Area, AAAI, 2025.
Li et al., BLADE: Enhancing Black-box Large Language Models with Small Domain-Specific Models, AAAI, 2025.
Zhang et al., SCITAT: A Question Answering Benchmark for Scientific Tables and Text Covering Diverse Reasoning Types, ACL-findings, 2025.
He et al., GIVE: Structured Reasoning of Large Language Models with Knowledge-Graph-Inspired Veracity Extrapolation, ICML, 2025.

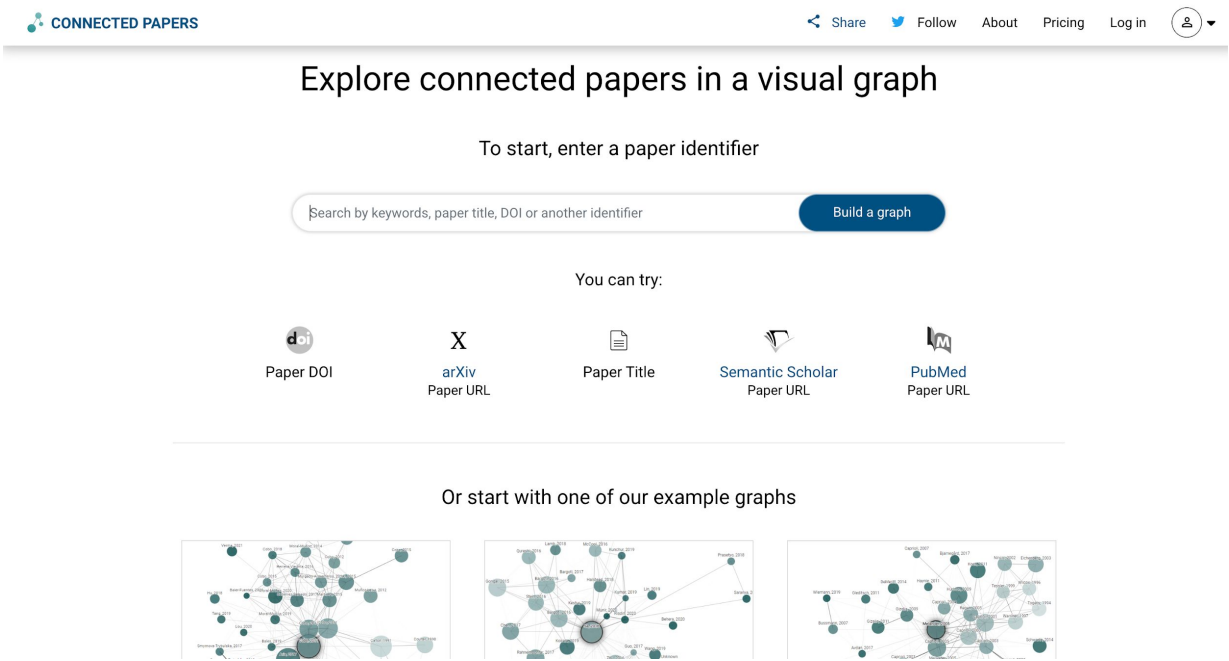
Takeaway – Paper Chat and Scientific QA

- Main Function
  - Interactive Q&A with research papers
- Key techniques
  - LLM-based agents
  - PDF Parsing & Contextual Chunking
  - Summarization & Key Point Extraction
  - Multi-turn Dialogue Tracking
- Challenges
  - PDF Quantity & Size Limits
  - Domain Knowledge
  - Reliability & Explainability

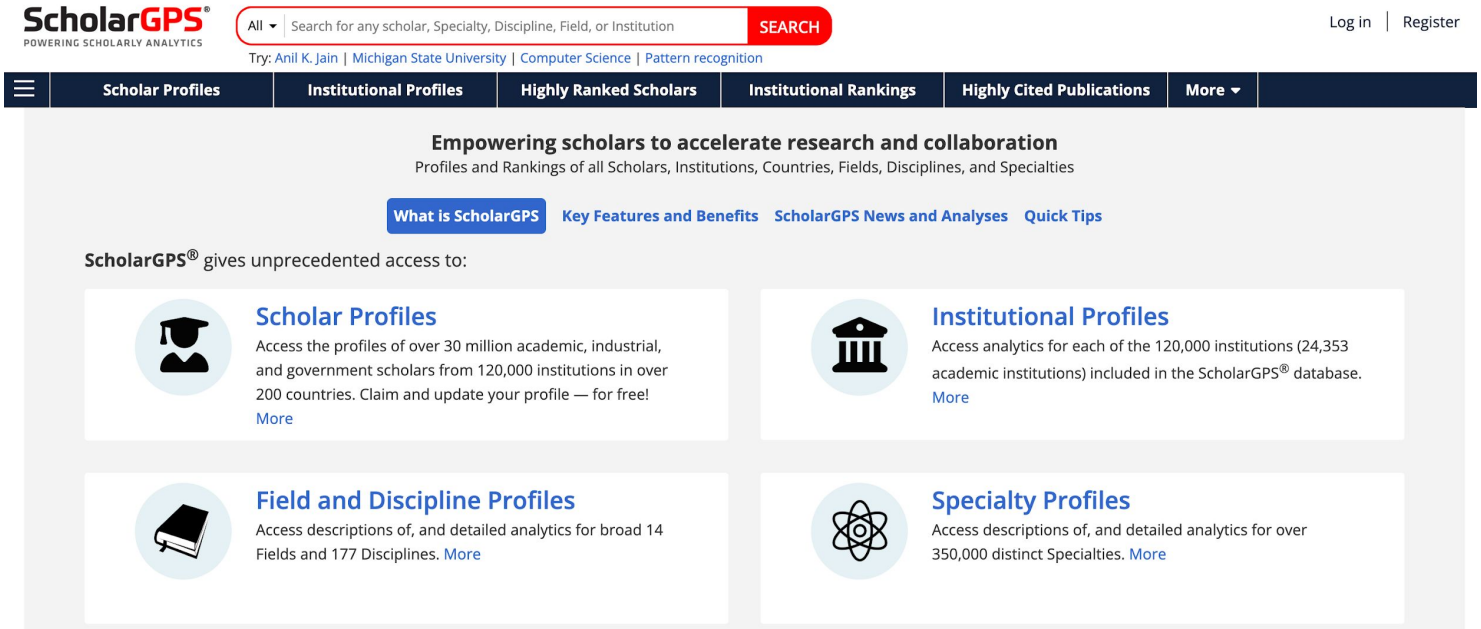
Agenda

- ❑ Background
- ❑ AI-Enhanced Search System
- ❑ Paper Chat and Scientific QA
- ❑ **Graph Based System**
- ❑ Recommendation System
- ❑ Takeaway

Graph Based System



Graph Based System



Graph Based System

CONNECTED PAPERS

Share Follow About Pricing Log in

ScholarGPS®
POWERING SCHOLARLY ANALYTICS

All | Search for any scholar, Specialty, Discipline, Field, or Institution | SEARCH

Try: Anil K. Jain | Michigan State University | Computer Science | Pattern recognition

Log in | Register

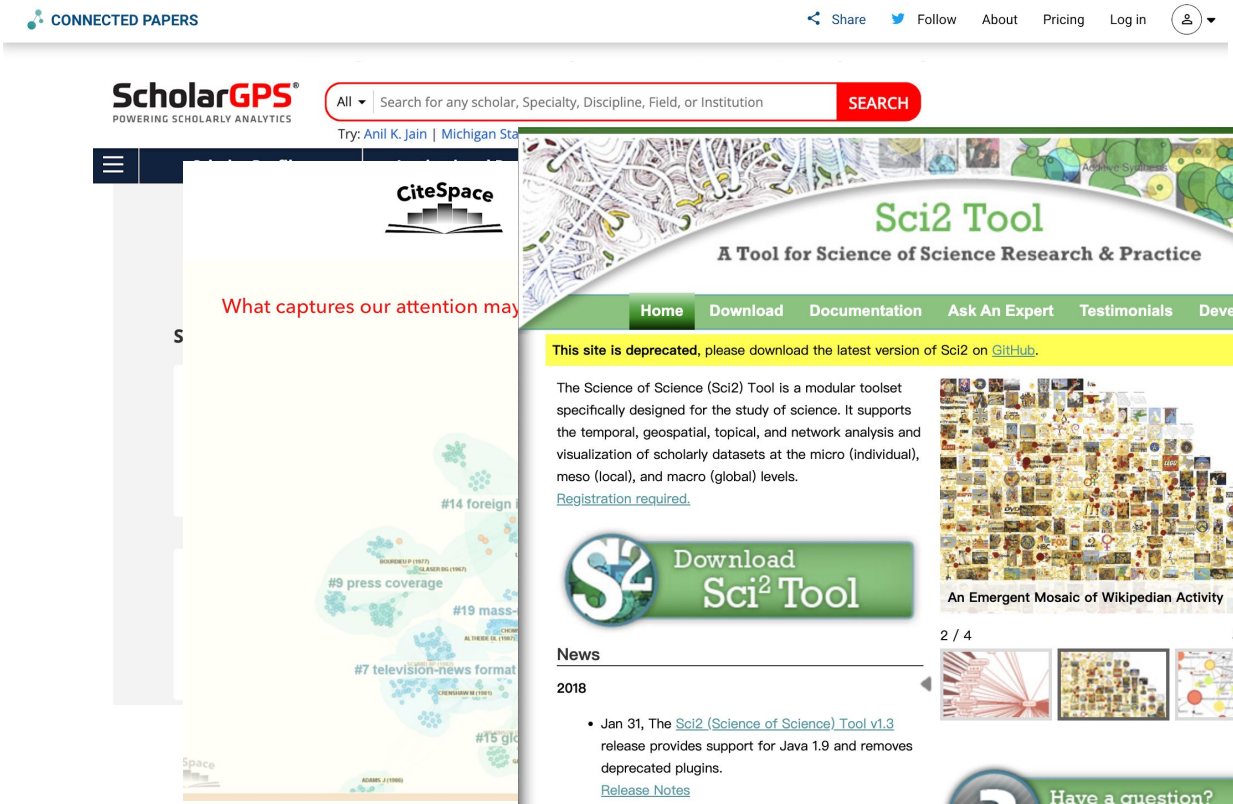
CiteSpace

Blogs Exemplars Gallery FAQs Standard Advanced Glossary Videos References

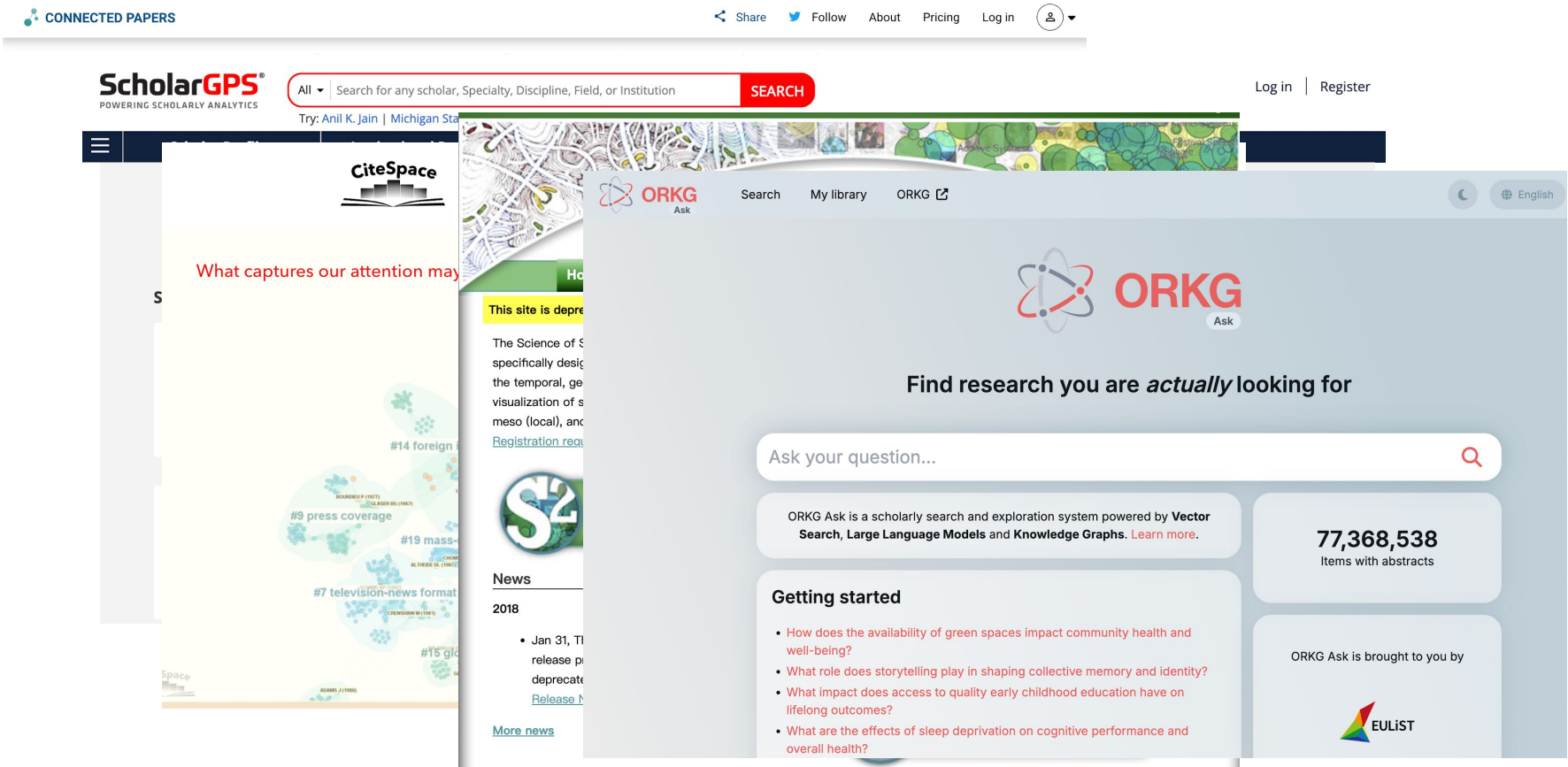
Login

What captures our attention may not always guide us to the best path, while what we fail to notice often makes the difference.

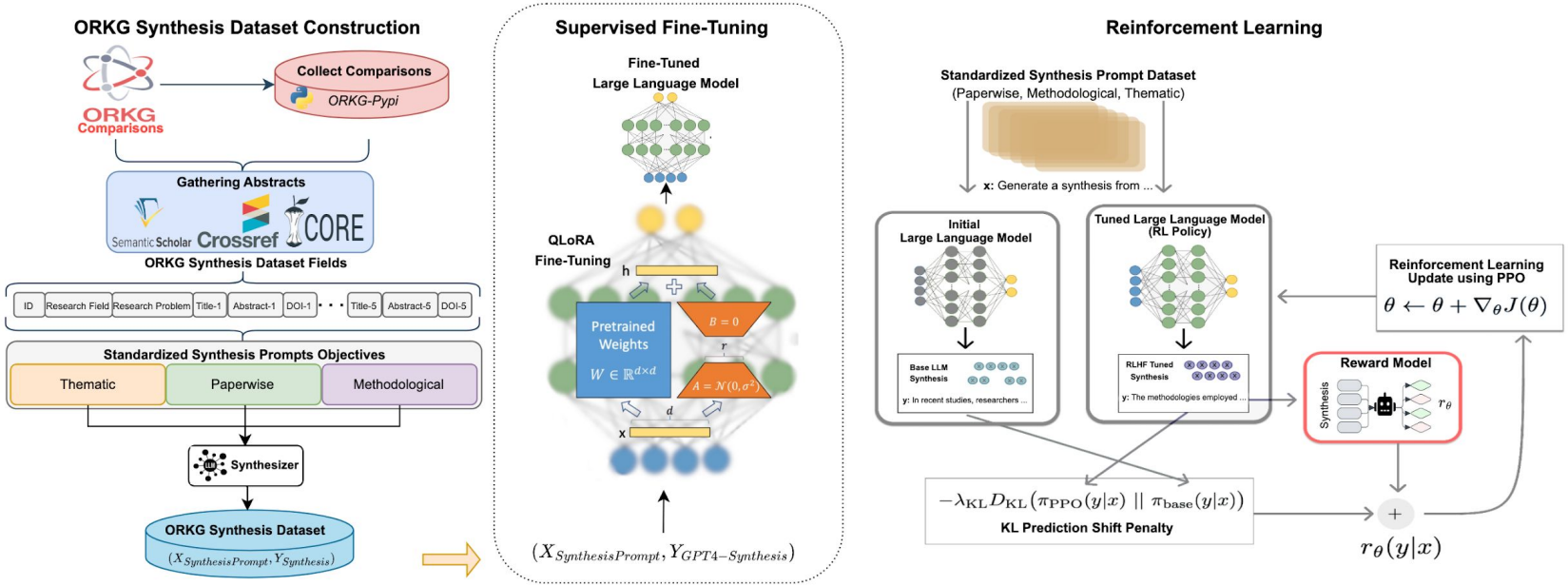
Graph Based System



Graph Based System



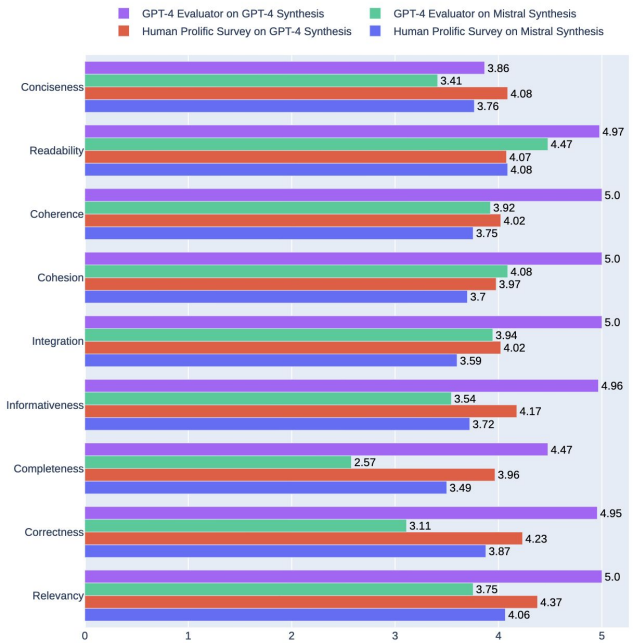
Graph Based System



Giglou et al., LLMs4Synthesis: Leveraging Large Language Models for Scientific Synthesis, JCSL, 2025.

Performance and Conclusion

- SFT+RLAIF performs the best.



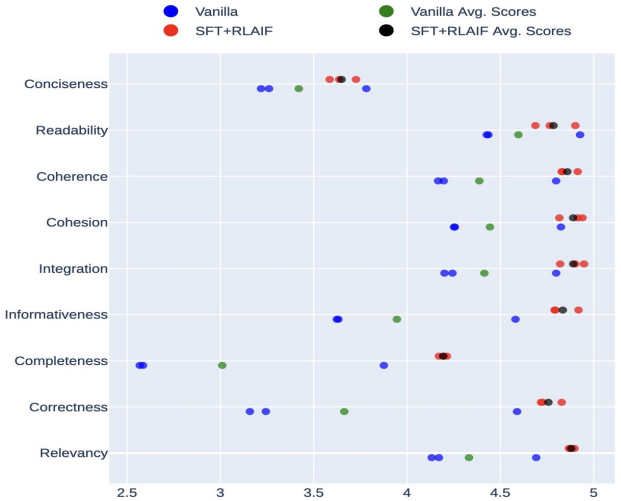
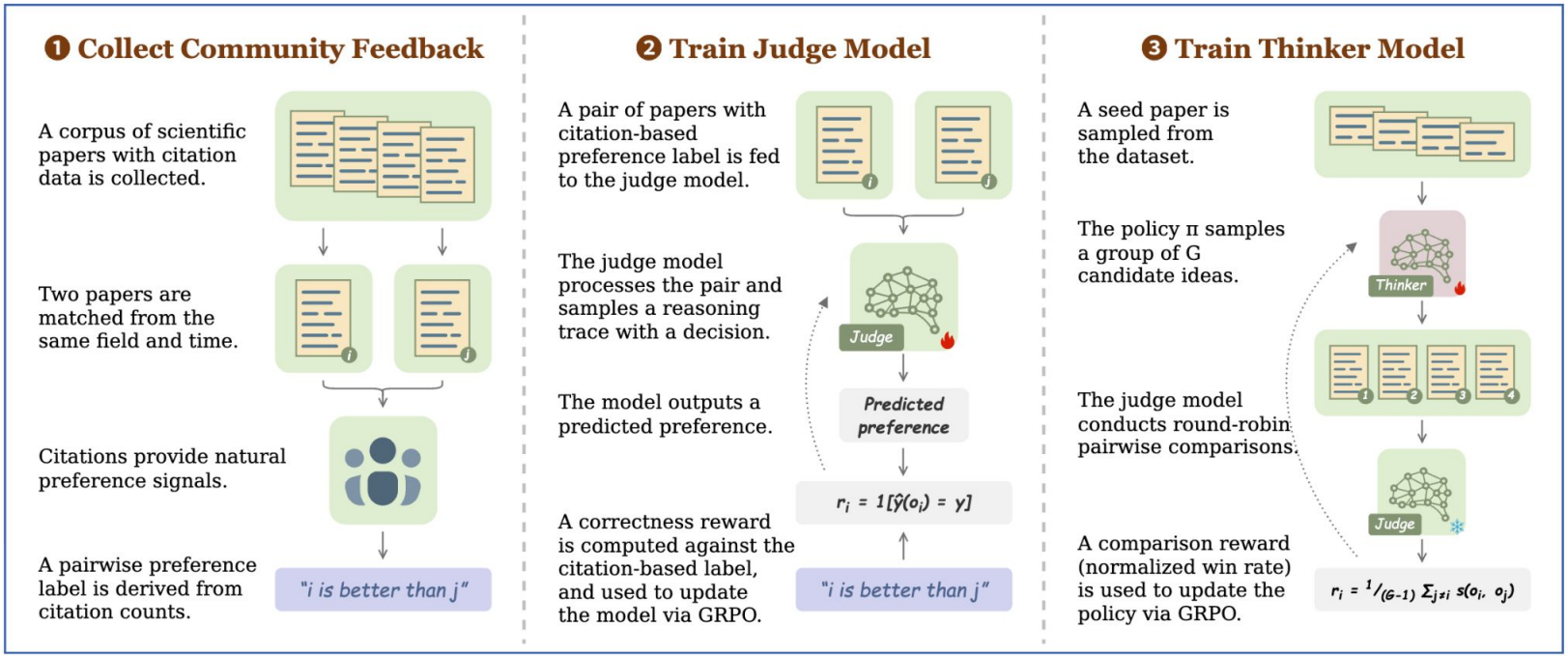


Figure 3: Consistency comparison of the GPT-4 evaluator between the *Vanilla* and *SFT+RLAIF (w/ GPT-4 Features)* models, assessed through three evaluations on the test set.

AI Can Learn Scientific Taste?



Tong et al.,AI Can Learn Scientific Taste, Arxiv, 2026.

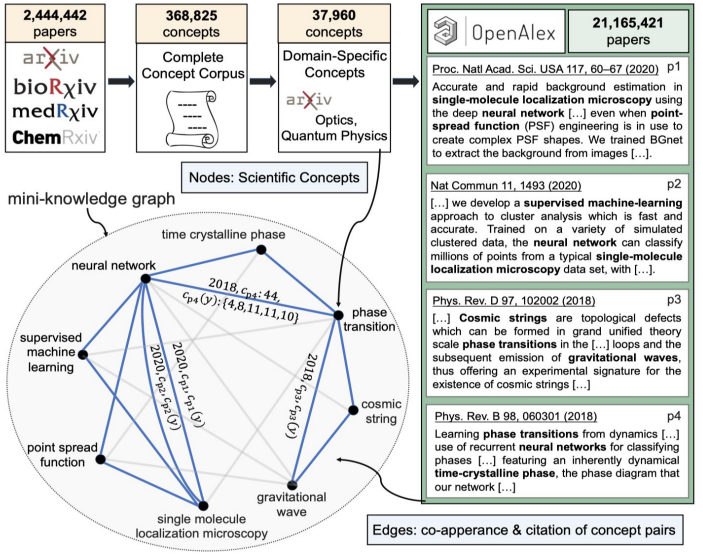
AI Can Learn Scientific Taste?

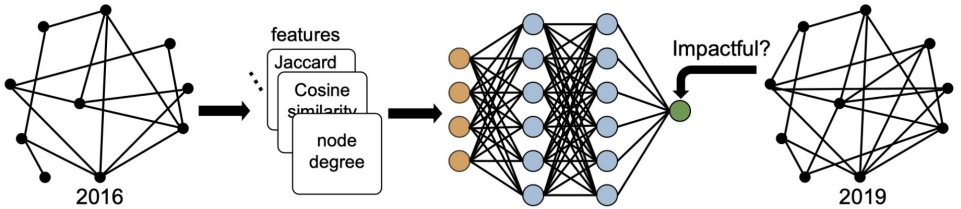
Model	CS	Math	Physics	Others	Avg.
Open-source Models					
Qwen3-4B-Instruct	66.5	65.6	54.8	57.1	60.3
SciJudge-Qwen3-4B	78.6 (+12.1)	74.6 (+9.0)	71.2 (+16.4)	79.8 (+22.7)	75.3 (+15.0)
Qwen3-30B-A3B-Instruct	73.8	70.5	59.4	65.5	66.3
SciJudge-Qwen3-30B	83.5 (+9.7)	78.7 (+8.2)	78.7 (+19.2)	82.3 (+16.8)	80.6 (+14.3)
Qwen2.5-1.5B-Instruct	6.3	10.7	6.0	6.7	7.0
SciJudge-Qwen2.5-1.5B	72.3 (+66.0)	73.0 (+62.3)	69.4 (+63.4)	77.3 (+70.6)	72.1 (+65.1)
Qwen2.5-3B-Instruct	16.5	36.9	23.8	21.0	23.5
SciJudge-Qwen2.5-3B	76.2 (+59.7)	76.2 (+39.3)	66.2 (+42.3)	81.5 (+60.5)	73.2 (+49.7)
Qwen2.5-7B-Instruct	57.3	37.7	37.0	51.3	45.2
SciJudge-Qwen2.5-7B	83.0 (+25.7)	68.8 (+31.1)	71.5 (+34.5)	87.4 (+36.1)	76.9 (+31.7)
Qwen2.5-14B-Instruct	64.1	63.9	54.5	56.3	59.1
SciJudge-Qwen2.5-14B	87.9 (+23.8)	78.7 (+14.8)	74.4 (+19.9)	84.9 (+28.6)	80.6 (+21.6)
Qwen2.5-32B-Instruct	71.4	61.5	55.9	62.2	62.2
SciJudge-Qwen2.5-32B	85.4 (+14.1)	77.9 (+16.4)	82.2 (+26.3)	89.9 (+27.7)	83.7 (+21.4)
Llama3.1-8B-Instruct	34.5	44.3	35.9	35.3	36.8
SciJudge-Llama3.1-8B	56.8 (+22.3)	59.8 (+15.6)	55.2 (+19.2)	60.5 (+25.2)	57.3 (+20.5)

Tong et al.,AI Can Learn Scientific Taste, Arxiv, 2026.

Trending Prediction

- Predict the impact of onsets of ideas.
- Extract 141 features for each pair are calculated.
  - 41 network features
  - 58 of these are node citation features
  - 42 features are about vertex pairs
- Graph Learning?

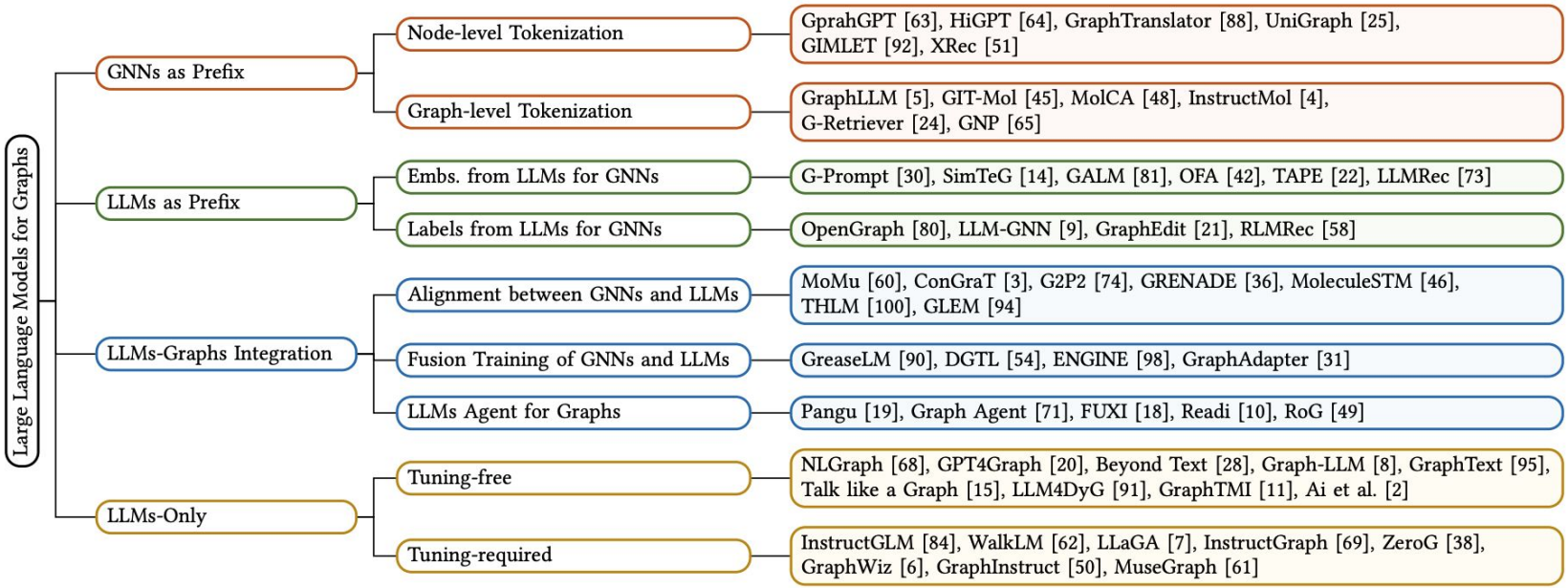




train dataset: 2016 -- 2019; test dataset: holdout data 2016 -- 2019; eval dataset: 2019 --2022

Gu et al., Forecasting high-impact research topics via machine learning on evolving knowledge graphs, Machine Learning: Science and Technology, 2025.

LLMs + Graph?



Ren et al., A Survey of Large Language Models for Graphs, KDD, 2024.

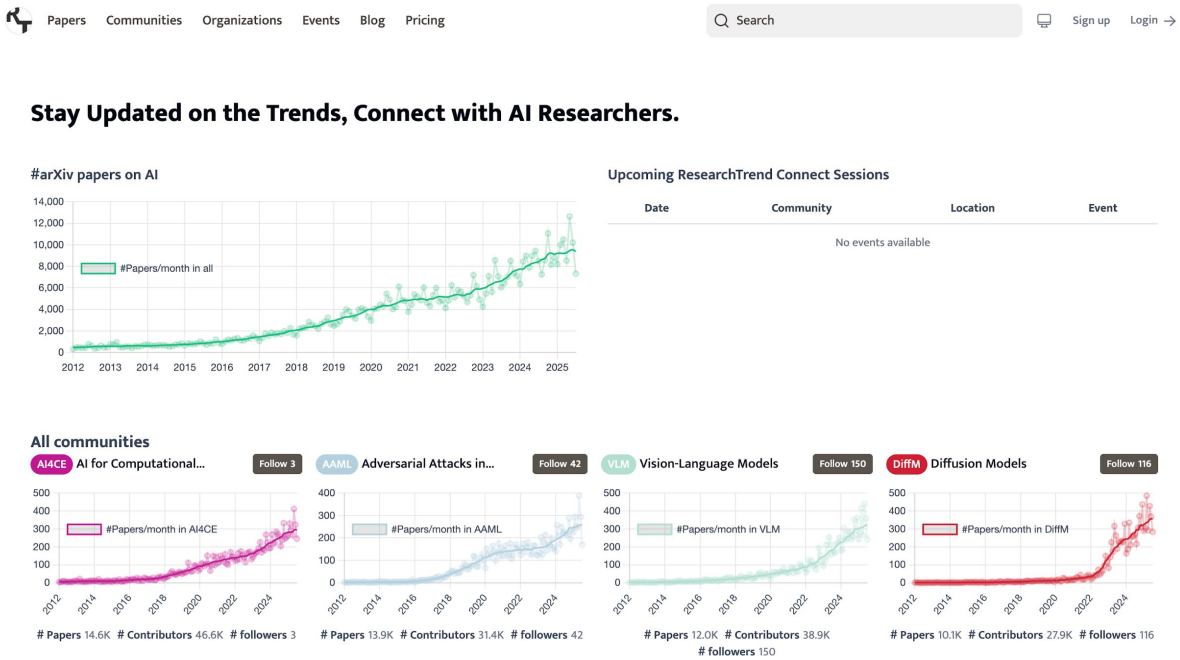
Takeaway – Graph Based System

- Main Function
  - Relationships between research papers
  - Explore knowledge structures
- Key techniques
  - Citation & Co-Authorship Networks
  - Graph Visualization & Navigation
  - Trending & Citation Analysis
- Challenges
  - Effective Graph Representation
  - Graph Update
  - Integration with LLMs

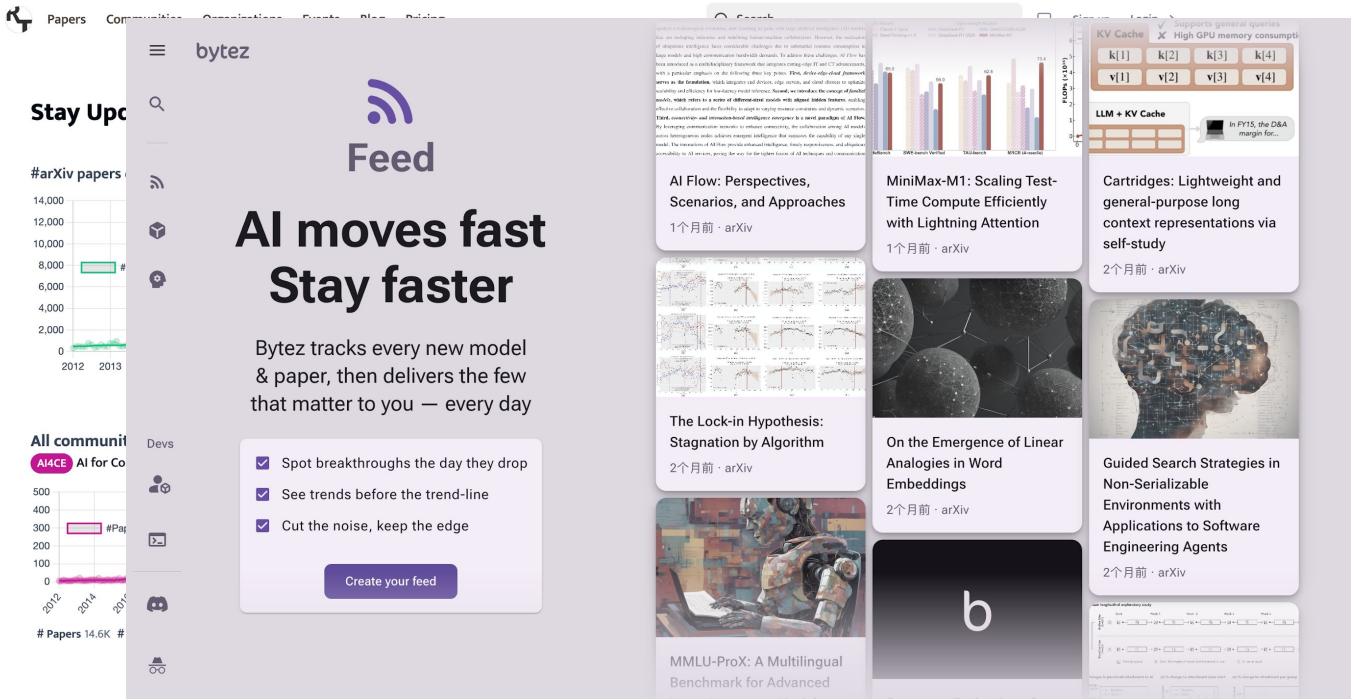
Agenda

- ❑ Background
- ❑ AI-Enhanced Search System
- ❑ Paper Chat and Scientific QA
- ❑ Graph Based System
- ❑ **Recommendation System**
- ❑ Takeaway

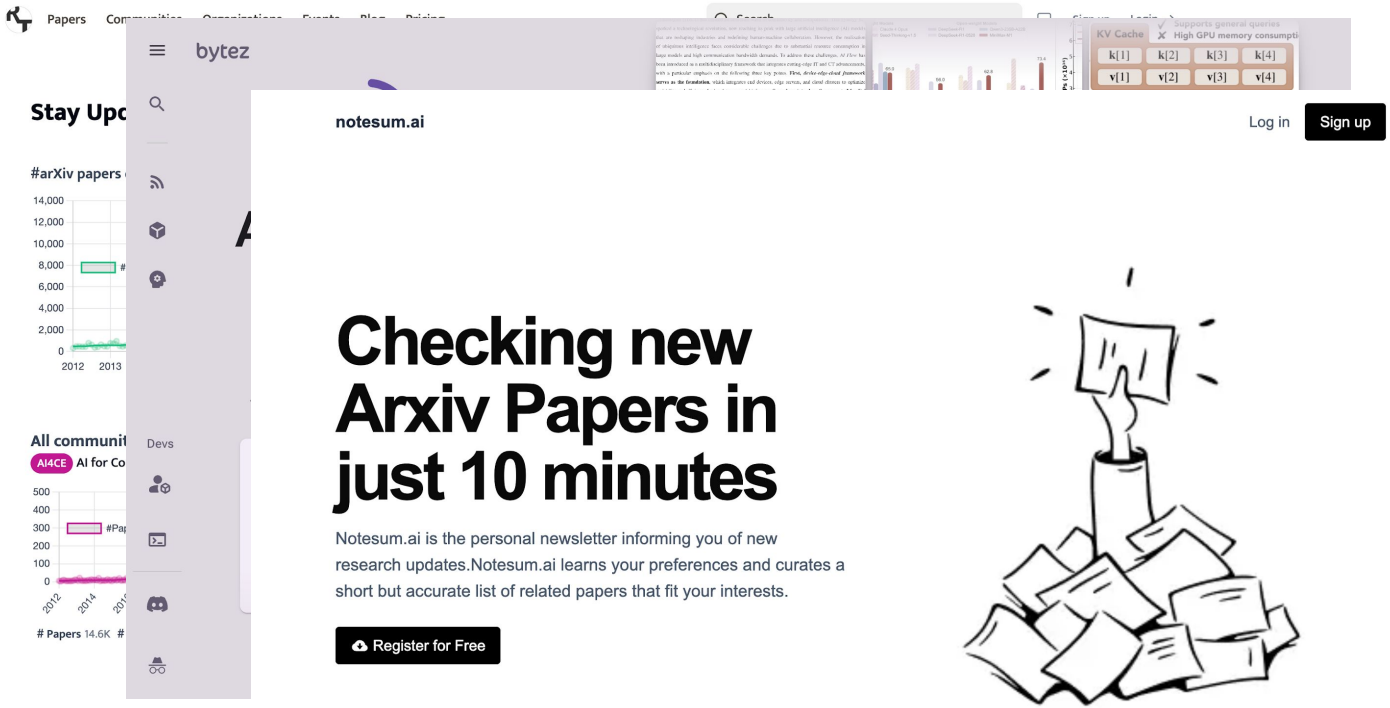
Recommendation System



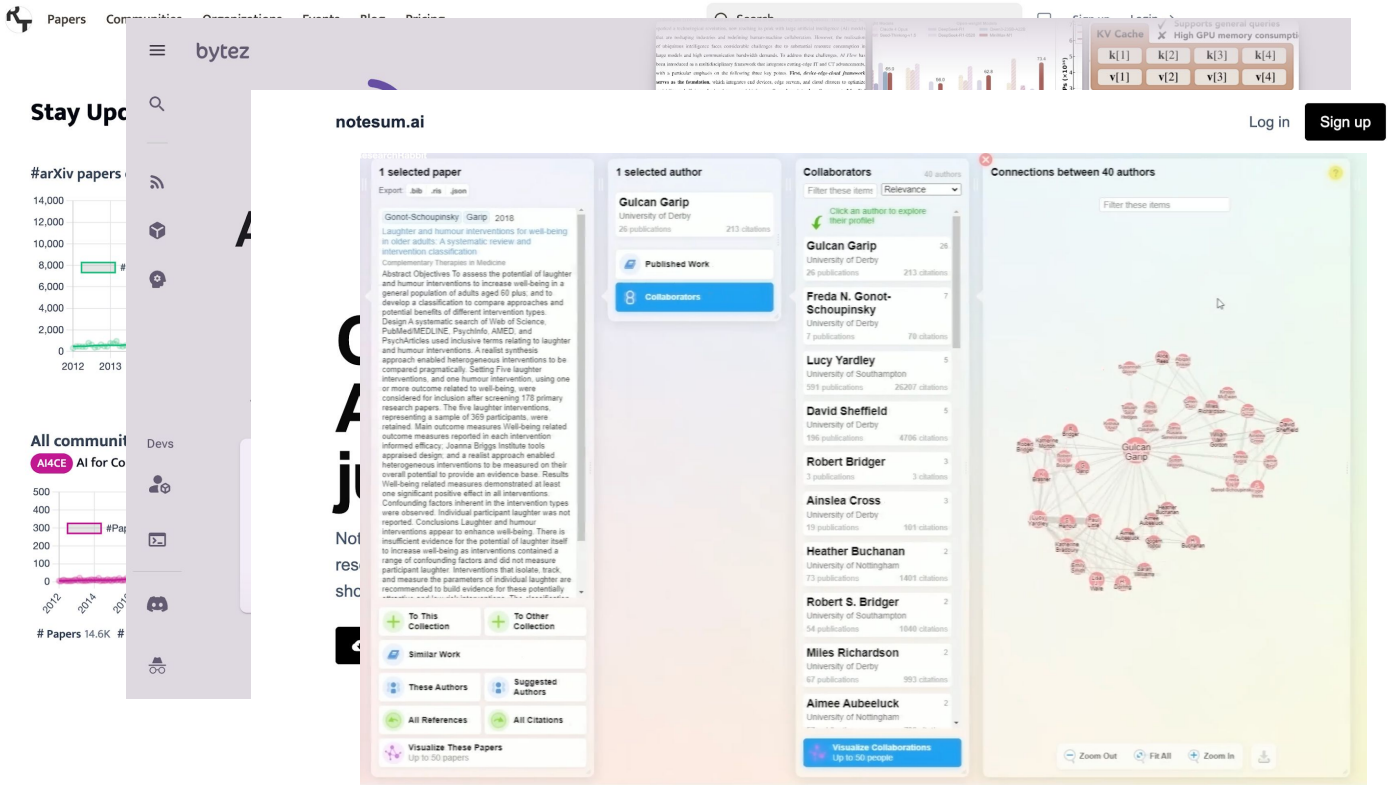
Recommendation System



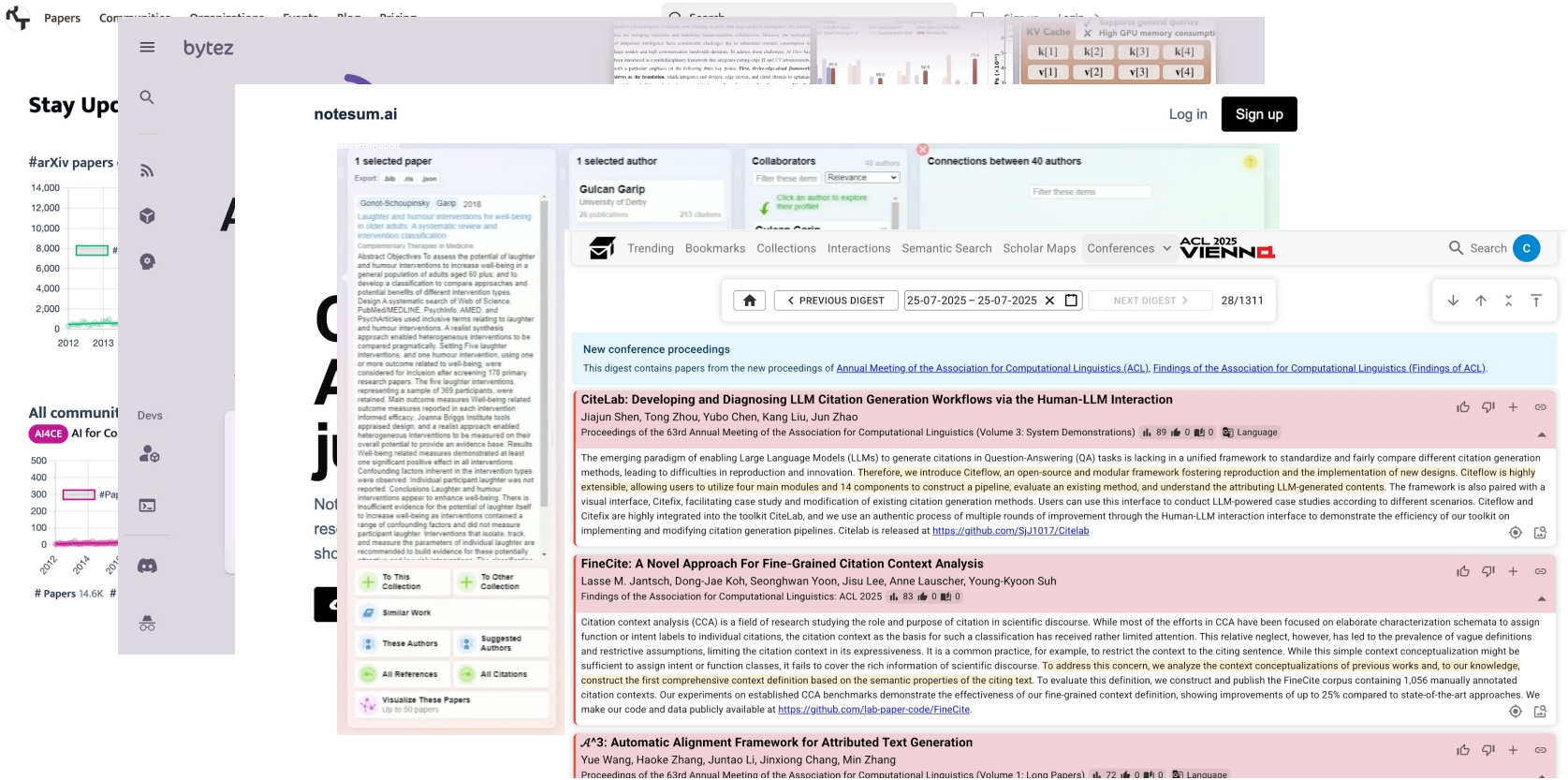
Recommendation System



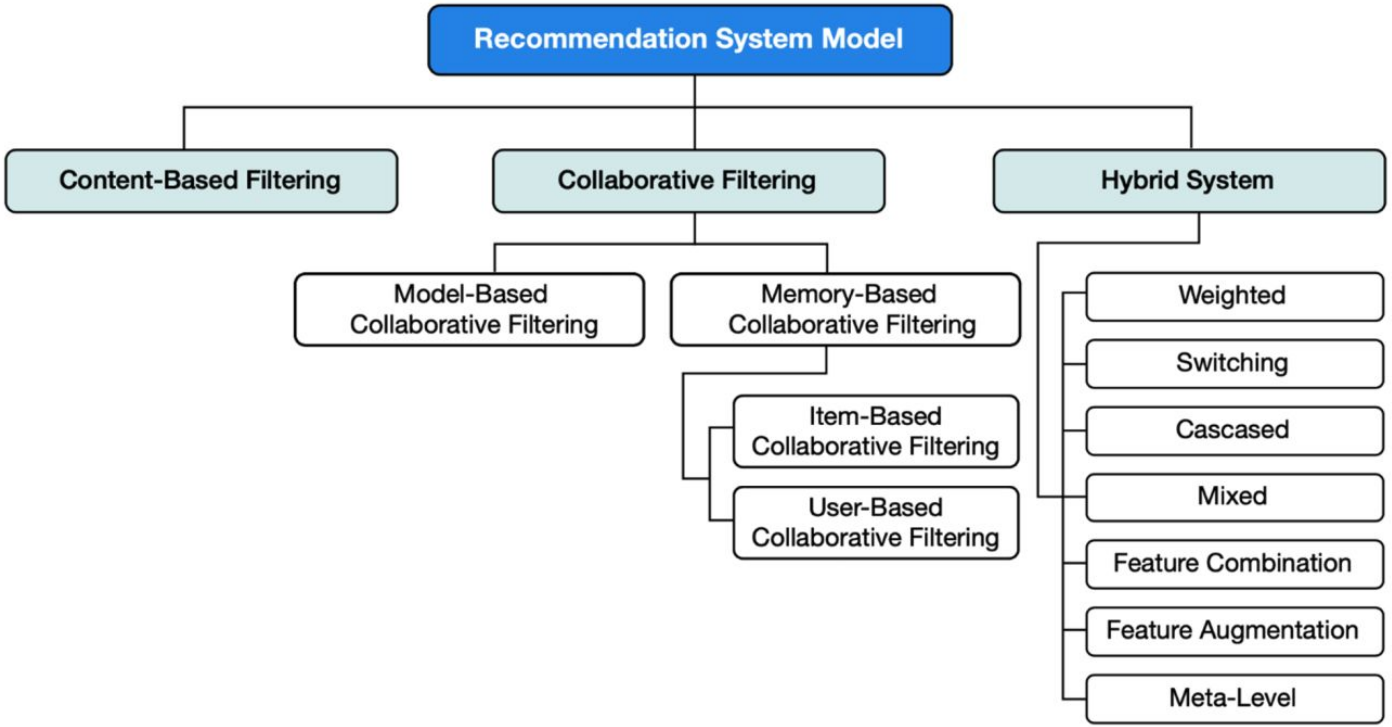
Recommendation System



Recommendation System

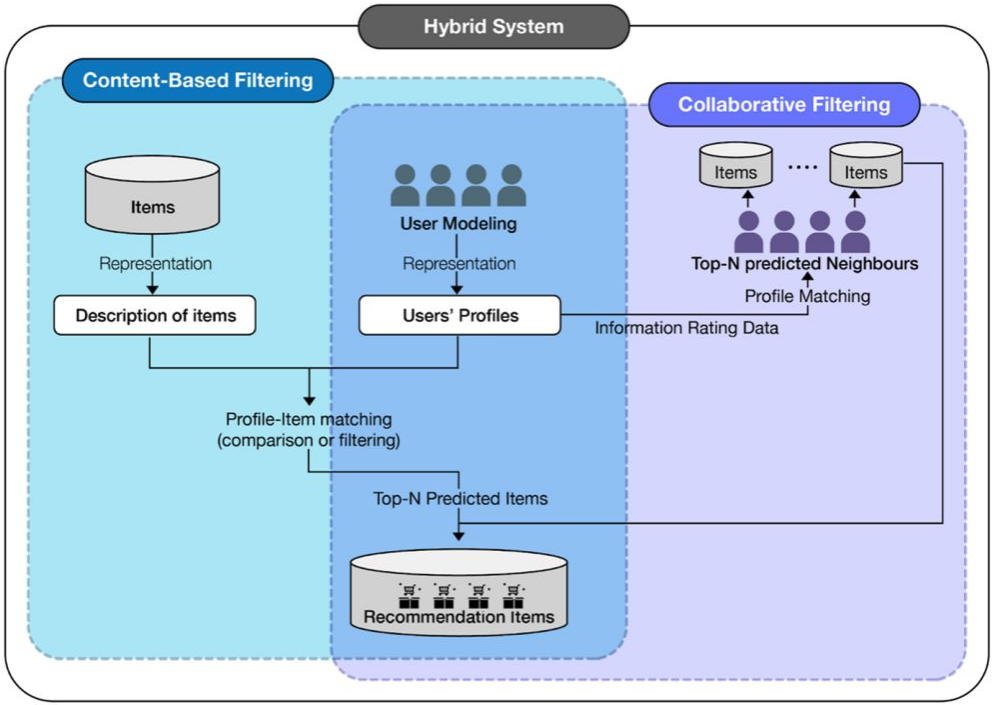


Recommendation System



Ko et al., A Survey of Recommendation Systems: Recommendation Models, Techniques, and Application Fields, electronics, 2022.

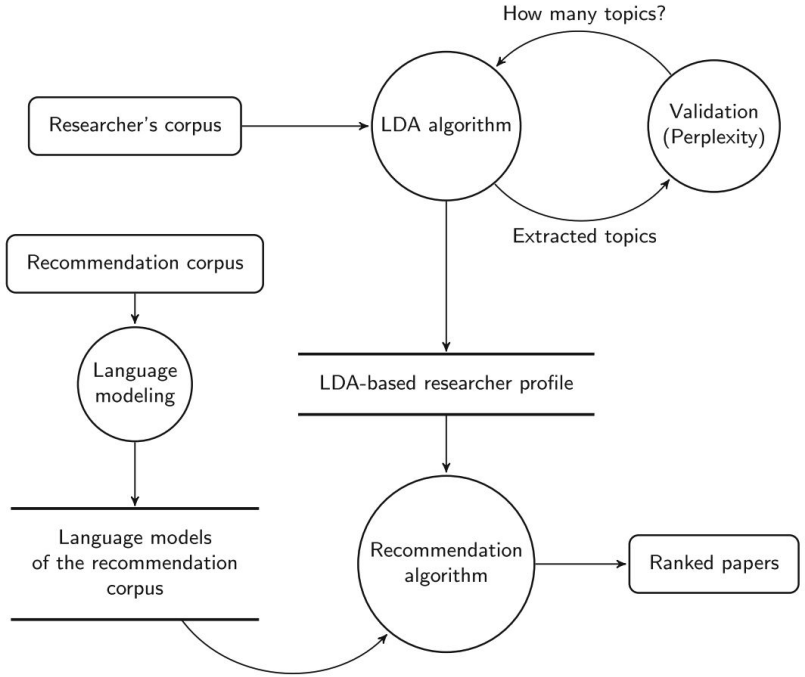
Pipeline of recommendation systems



Ko et al., A Survey of Recommendation Systems: Recommendation Models, Techniques, and Application Fields, electronics, 2022.

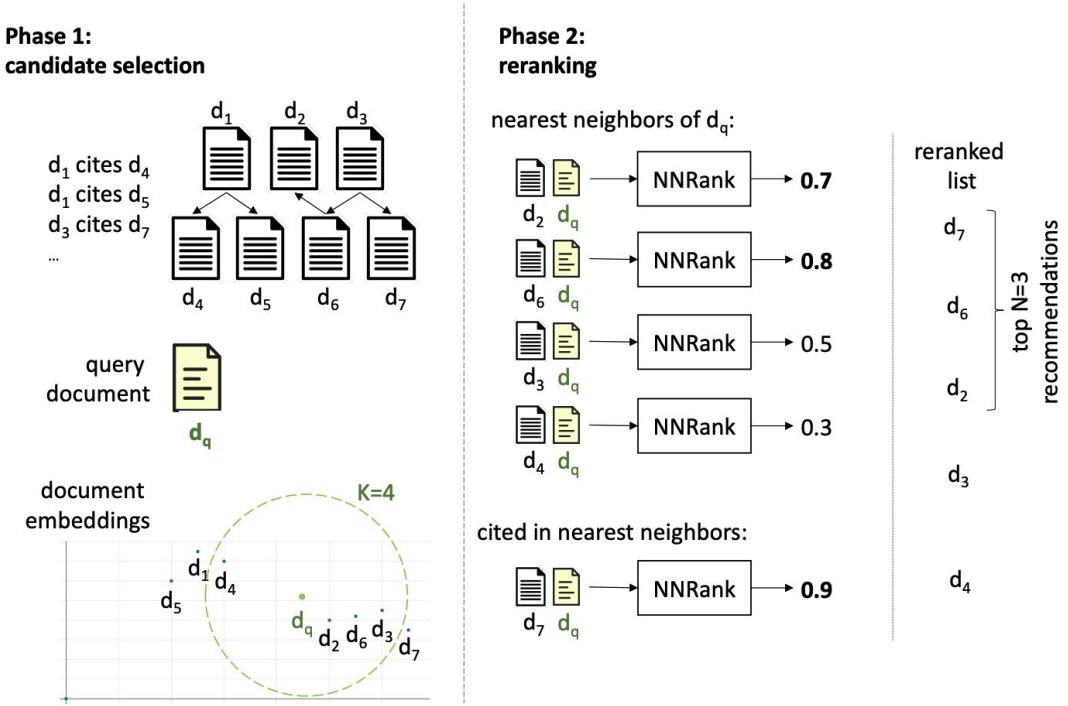
Content-based Methods

- Compares topics from the researcher's profile with the language models of unseen papers.
- Uses the symmetrized Kullback-Leibler divergence to measure similarity between probability distributions (topics and language models).
- **Limitations**
  - Shifted Cold-Start Problem
  - Potential for Limited/General Concepts
  - Lack of Contextual Citation Information



Amami et al., An LDA-Based Approach to Scientific Paper Recommendation, International Conference on Applications of Natural Language to Data Bases, 2016.

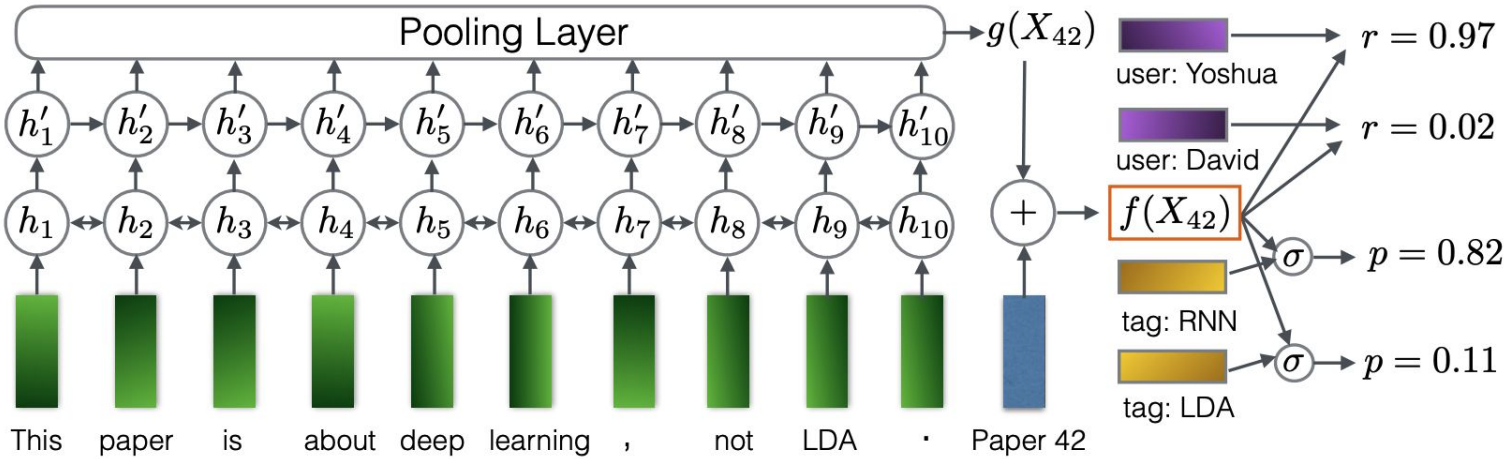
NNSelect+NNRank



Bhagavatula et al., Content-Based Citation Recommendation, NAACL, 2018.

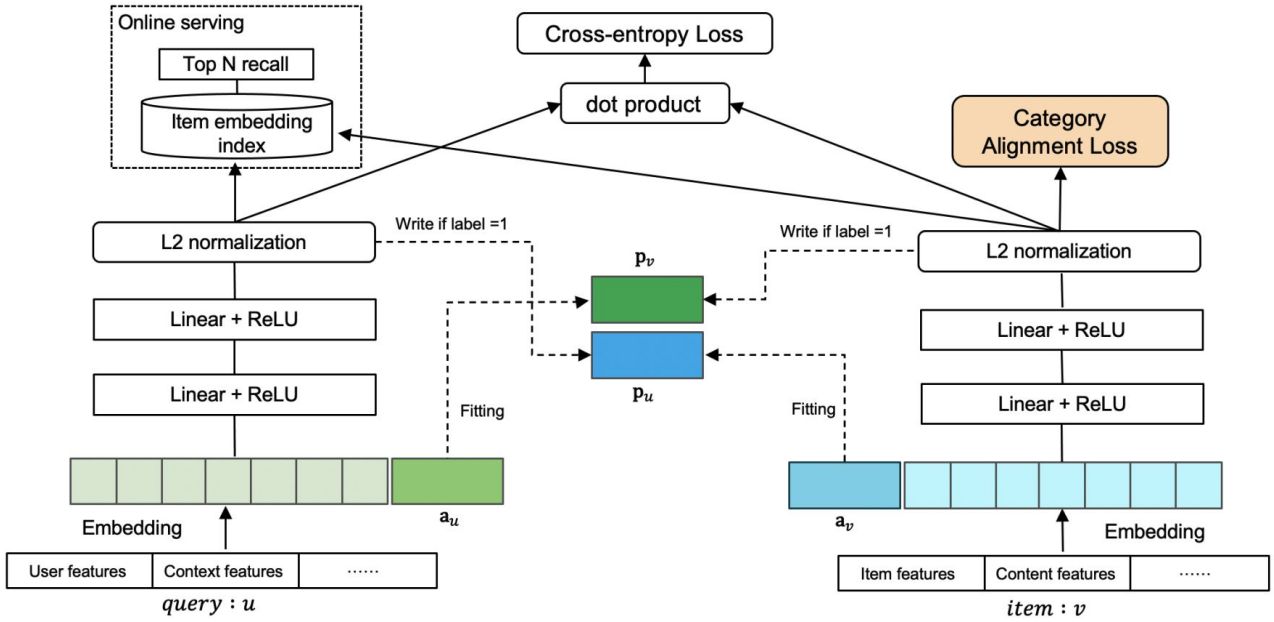
Collaborative Filtering

- By training a text encoder network as a combination of content recommendation and item metadata prediction (e.g., tag prediction)



Bansal et al., Ask the GRU: Multi-Task Learning for Deep Text Recommendations, RECSYS, 2016.

Hybrid Systems

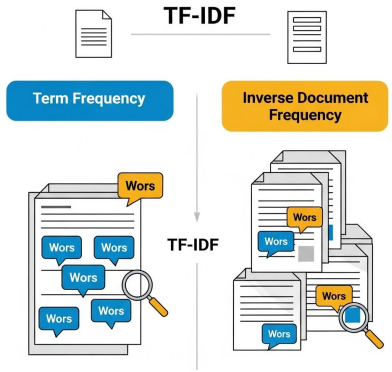


Yu et al., A Dual Augmented Two-tower Model for Online Large-scale Recommendation, KDD, 2021.

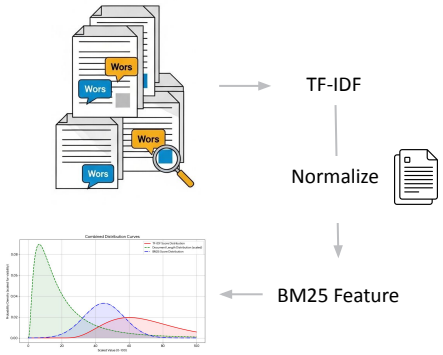
Comparsion of Recommendation Systems

Model Type	Pros	Cons
Content-Based Filtering	• Works well for **new users** • **Personalized** to individual preferences • Doesn't require large user base • Easy to **interpret**	• Needs detailed **item features** • Limited discovery of **new interests**
Collaborative Filtering	• Doesn't need item metadata • Leverages collective **user behavior**	• Struggles with **sparse data** • **Scalability** issues in memory-based
Hybrid Systems	• **Combines** best of all models • **Mitigates** individual **weaknesses** • Often **improves** accuracy	• Increased **complexity** • Requires **more computation**

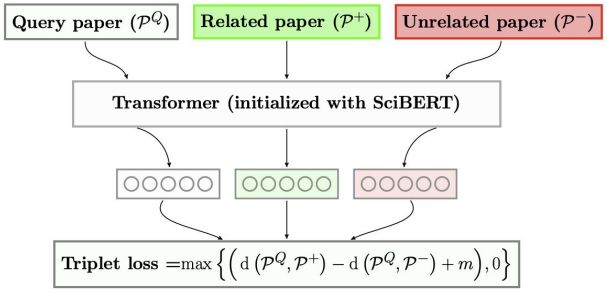
Document Embeddings



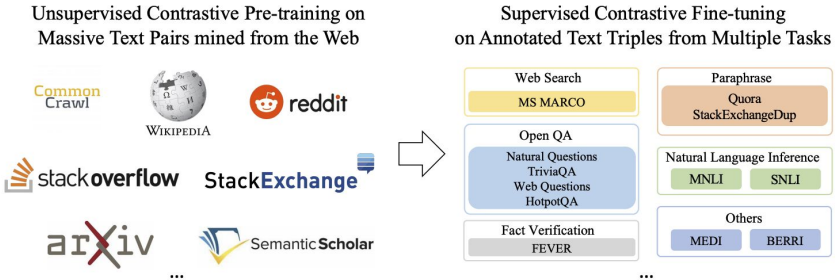
TF-IDF



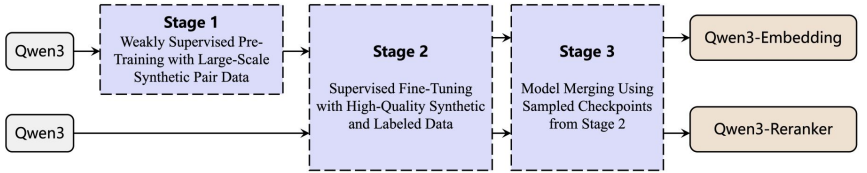
BM25



SPECTER



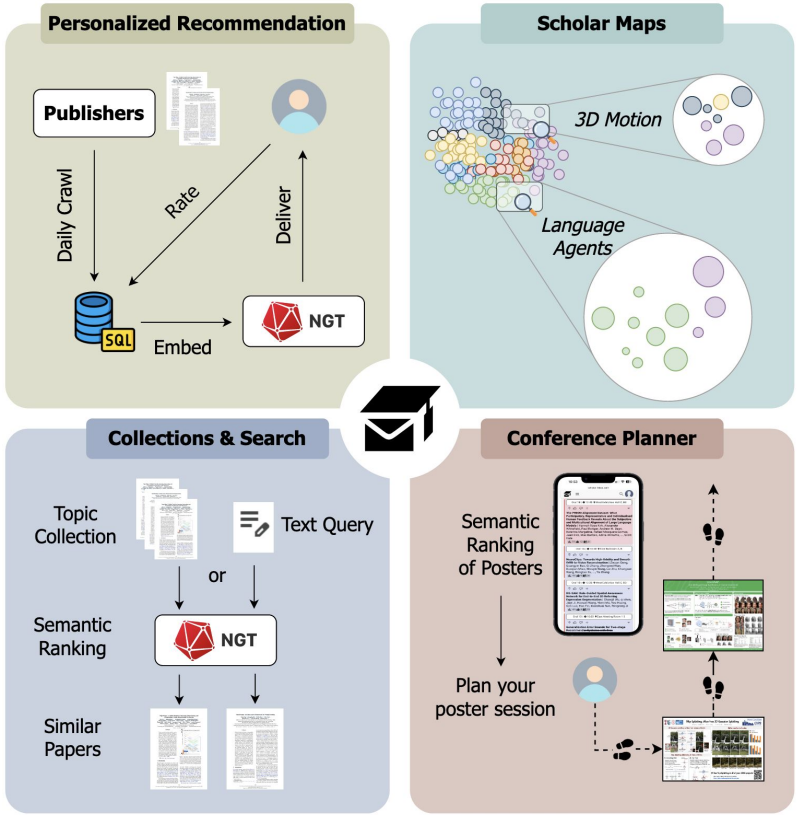
GTE



Qwen-3 Embedding

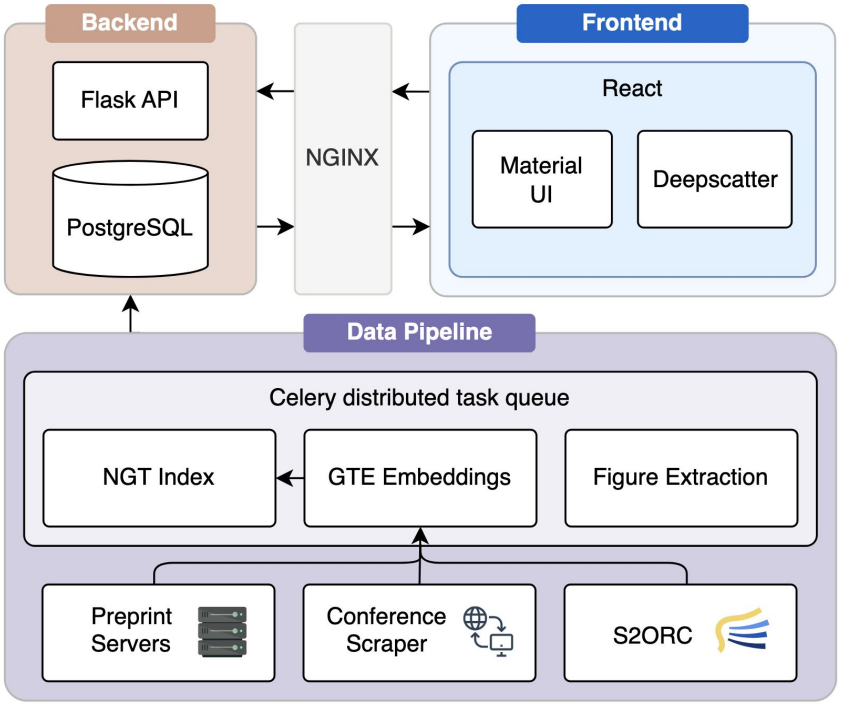
Scholar Inbox

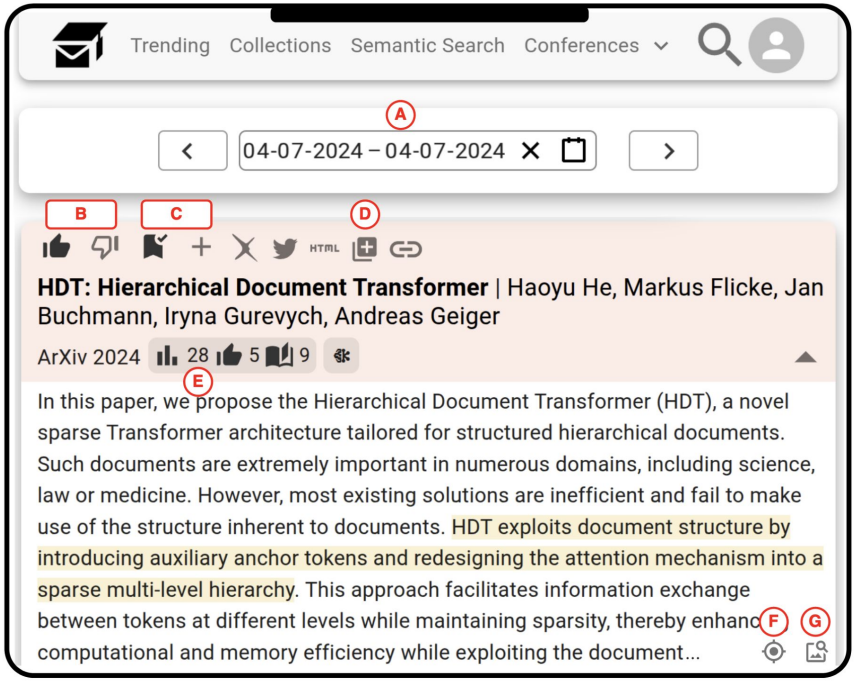
- Personalized Recommendations
- Scholar Maps
- Collections
- Conference Planner



Flicke et al., Scholar Inbox: Personalized Paper Recommendations for Scientists, ACL Demo, 2025.

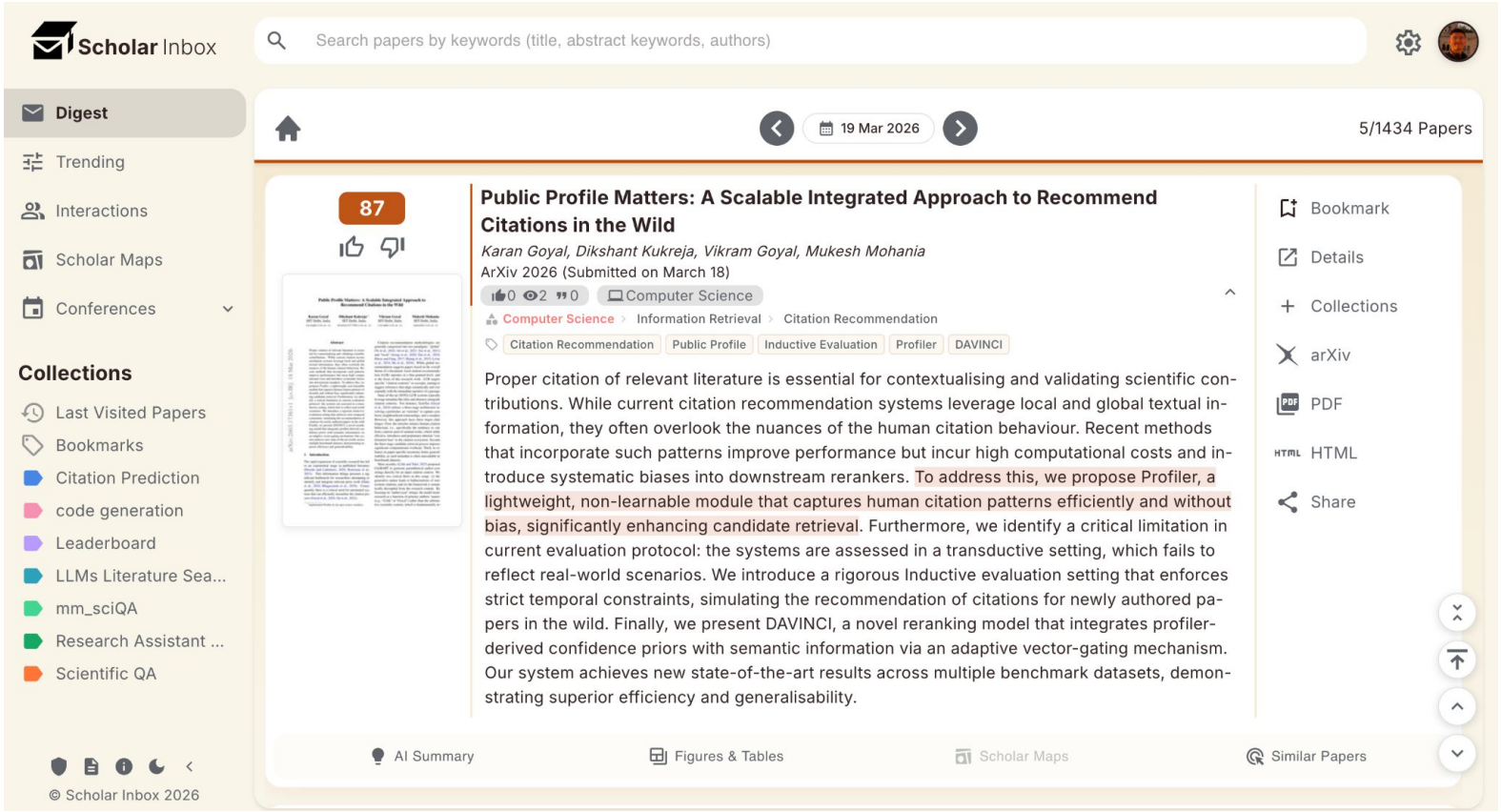
Software Flow



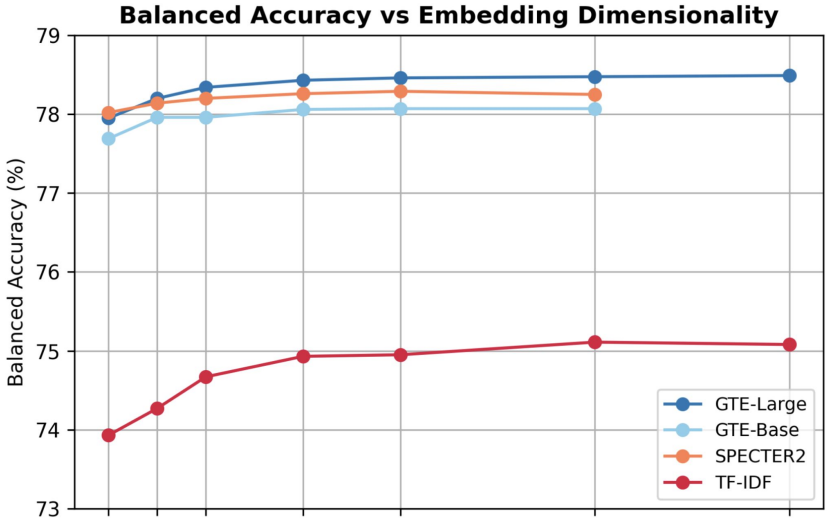


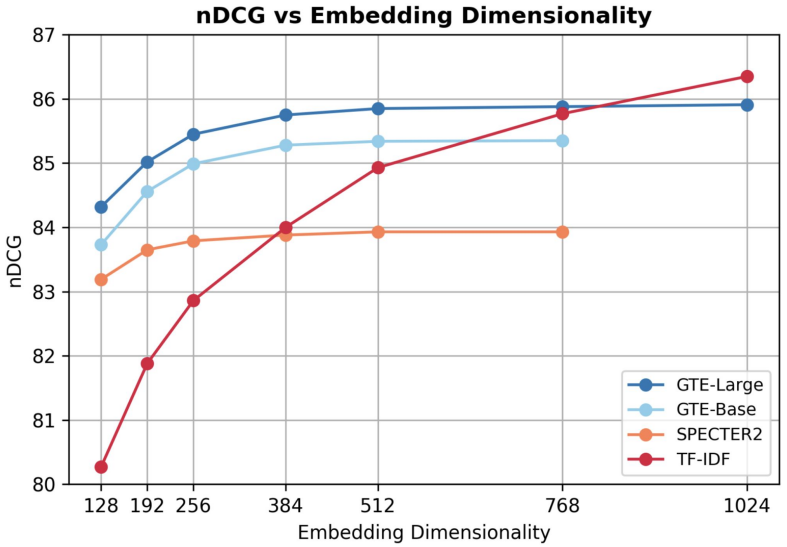
Flicke et al., Scholar Inbox: Personalized Paper Recommendations for Scientists, ACL Demo, 2025.

Recommender System

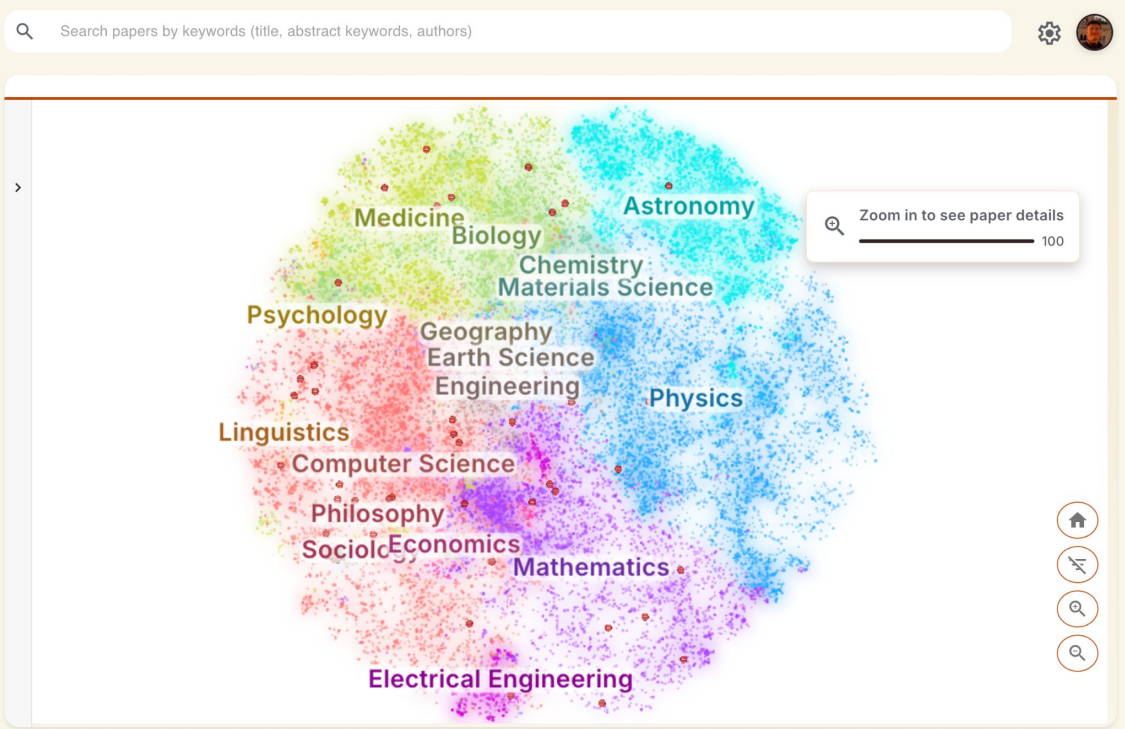


Embeddings


Balanced Accuracy vs Embedding Dimensionality
Balanced Accuracy (%)
79
78
77
76
75
74
73
GTE-Large
GTE-Base
SPECTER2
TF-IDF

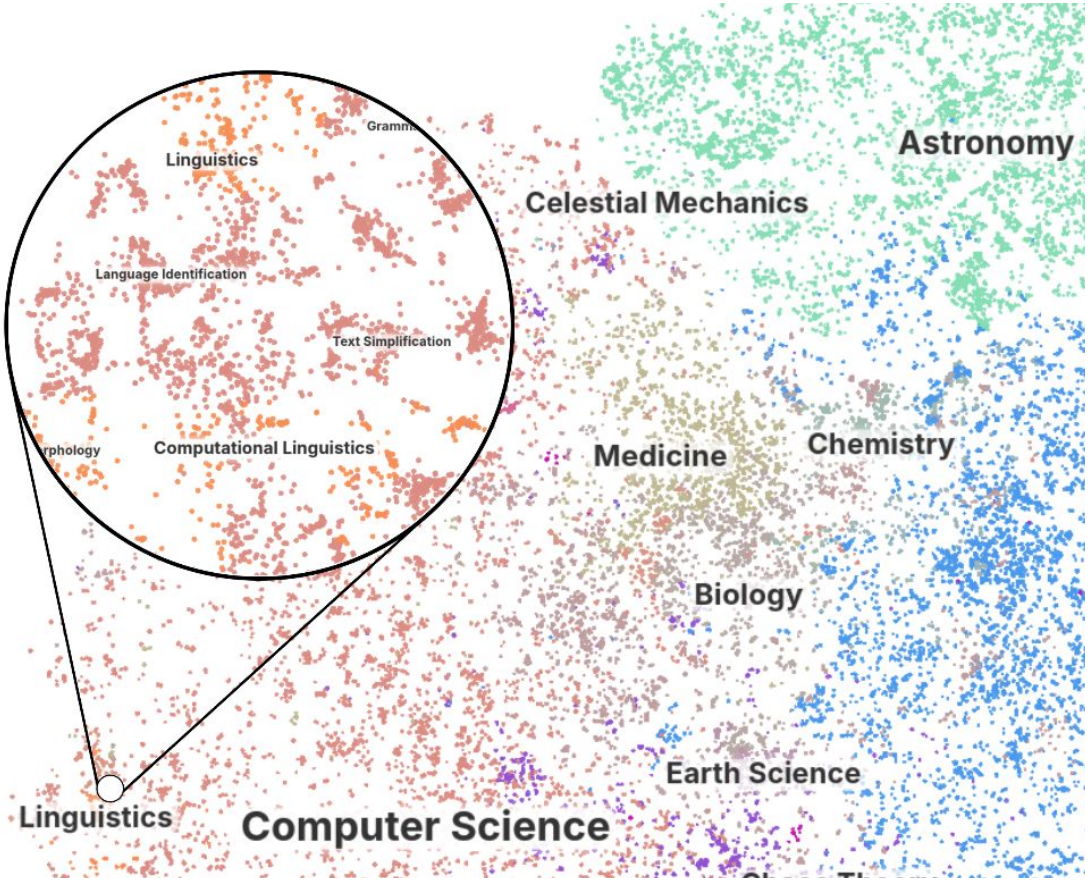

nDCG vs Embedding Dimensionality
nDCG
87
86
85
84
83
82
81
80
128
192
256
384
512
768
1024
Embedding Dimensionality
GTE-Large
GTE-Base
SPECTER2
TF-IDF

Scholar Maps



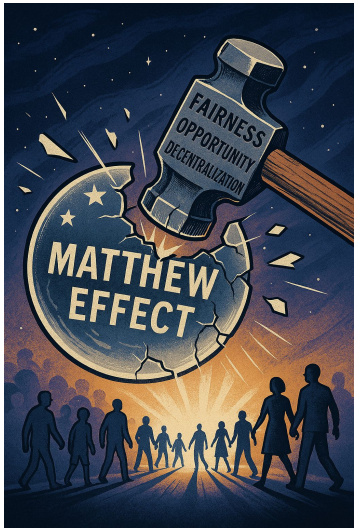
Flicke et al., Scholar Inbox: Personalized Paper Recommendations for Scientists, ACL Demo, 2025.

Scholar Maps

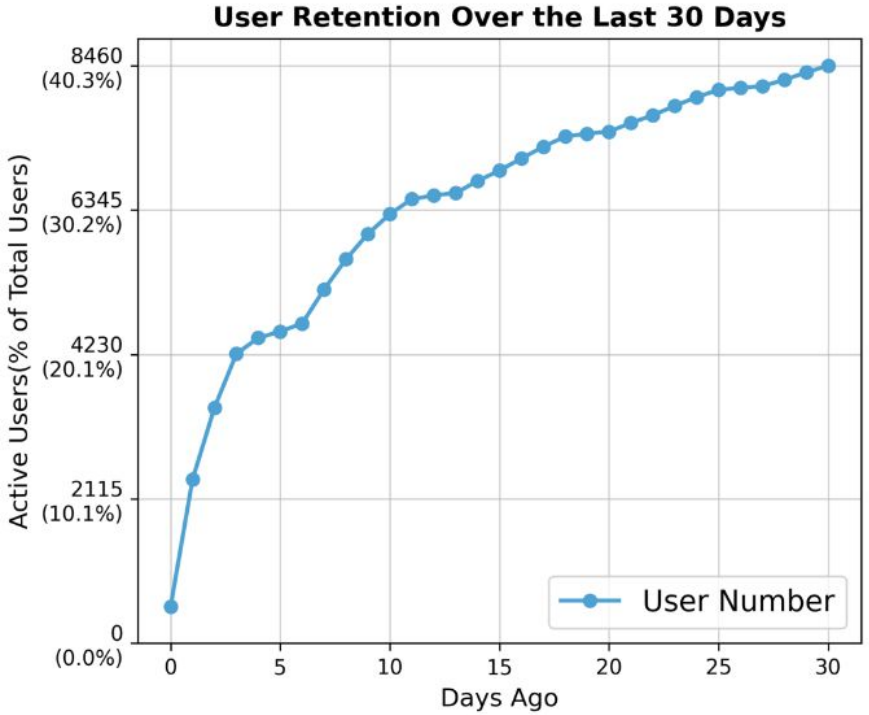


Flicke et al., Scholar Inbox: Personalized Paper Recommendations for Scientists, ACL Demo, 2025.

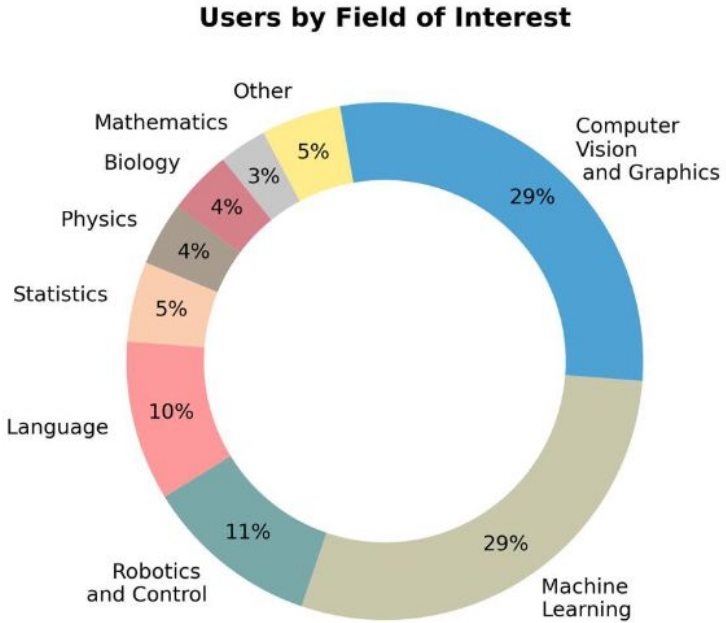
Why we create Scholar Inbox?



9,000 Users Come Back Again and Again

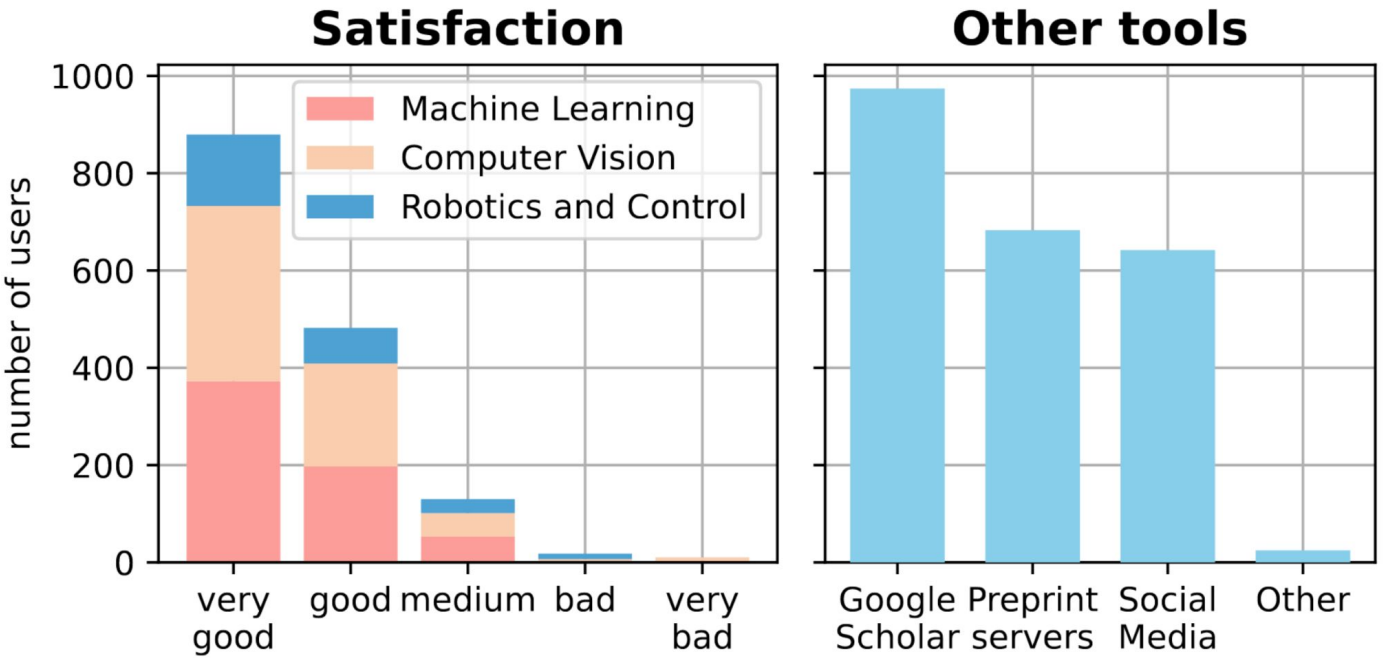


(a) User Retention



(b) User Domain Distribution

User Study


Satisfaction
Other tools
number of users
1000
800
600
400
200
0
Machine Learning
Computer Vision
Robotics and Control
very good
good
medium
bad
very bad
Google Scholar
Preprint servers
Social Media
Other

Takeaway – Recommendation System

- Main Function
  - Keep updated literatures
  - Personlized Recommendation
  - Collections
- Key techniques
  - Content-Based Filtering
  - Collaborative Filtering
  - Semantic Search & Embeddings
- Challenges
  - Cold Start Problem
  - Overpersonalization vs Matthrew Effect
  - Dynamic Interests of Researchers

Overview of popular literature search

	Platform	Search	Recommendations	Collections	Citation Analysis	Trending Analysis	Author Profiles	Visualization Tools	Paper Chat	Idea Generation	Paper Writing	Summarization	Paper Review	Datasets	Code Repositories	LLM Integration	Web API	Personalization	Cost	Data Source
AI-Enhanced Search	Elicit	✓						✓	✓		✓	✓			✓				Freemium	125 million
	OpenScholar	✓		✓				✓			✓				✓				Free	45 million
	Undermind	✓		✓				✓			✓				✓		✓		Premium	over 200 million
	Perplexity	✓						✓	✓		✓	✓			✓				Freemium	
	Consensus	✓		✓				✓			✓				✓	✓			Freemium	over 200 million
	SciSpace	✓		✓				✓	✓		✓	✓			✓				Freemium	
	scienceQA	✓		✓	✓			✓	✓		✓	✓			✓				Freemium	220 million
	PaperQA2							✓						✓	✓				Free	
	Paperguide	✓		✓				✓	✓		✓	✓			✓				Freemium	
	HyperWrite	✓						✓	✓	✓	✓	✓			✓				Premium	
	ResearchKick	✓						✓	✓	✓	✓	✓			✓		✓		Premium	
Graph-Based	Connected Papers	✓		✓			✓												Freemium	214 million
	ScholarGPS	✓			✓	✓	✓	✓											Free	over 200 million
	CiteSpace					✓		✓											Freemium	
	Sci2							✓											Free	
	NLP KG	✓		✓	✓		✓	✓											Free	
	ORKG ASK	✓		✓							✓				✓				Free	76 million
Paper Chat	ChatGPT	✓						✓	✓	✓	✓	✓			✓	✓			Freemium	10 pdf files
	Claude	✓						✓	✓	✓	✓	✓			✓	✓			Freemium	5 pdf files
	Deepseek	✓						✓	✓	✓	✓	✓			✓	✓			Free	
	Research			✓				✓	✓		✓	✓			✓				Freemium	1 pdf file
	NotebookLM							✓	✓		✓	✓			✓		✓		Freemium	50 pdf files
	Enago Read	✓		✓				✓	✓		✓	✓			✓		✓		Freemium	1 pdf file
	DocAnalyzer.AI			✓				✓	✓		✓	✓			✓	✓	✓		Premium	few pdf files
	CoralAI			✓				✓	✓		✓	✓			✓				Freemium	1 pdf file
	ExplainPaper							✓	✓		✓	✓			✓				Freemium	1 pdf file
	ChatPDF	✓		✓				✓	✓		✓	✓			✓				Premium	1 pdf file
Recommender	Arxiv Sanity	✓	✓	✓													✓		Free	
	Scholar Inbox	✓	✓	✓		✓		✓							✓		✓		Free	
	ResearchTrend.ai	✓				✓													Freemium	
	TrendingPapers	✓	✓			✓					✓				✓		✓		Free	
	Bytez	✓				✓			✓	✓		✓	✓		✓	✓			Freemium	
	Notesum.ai	✓	✓	✓							✓				✓		✓		Freemium	
	Research Rabbit	✓		✓			✓												Free	

Overview of popular literature search

	Platform	Search	Recommendations	Collections	Citation Analysis	Trending Analysis	Author Profiles	Visualization Tools	Paper Chat	Idea Generation	Paper Writing	Summarization	Paper Review	Datasets	Code Repositories	LLM Integration	Web API	Personalization	Cost	Data Source
Search Engines	Google Scholar	✓	✓	✓	✓		✓											✓	Free	
	Semantic Scholar	✓	✓	✓	✓	✓	✓		✓		✓				✓	✓	✓	✓	Free	214 million
	Baidu Scholar	✓	✓	✓	✓	✓	✓								✓			✓	Freemium	680 million
	BASE	✓		✓												✓			Free	415 million
	Internet Archive Scholar	✓														✓			Free	35 million
	Scilit	✓		✓	✓		✓												Free	172 million
	The Lens	✓		✓			✓									✓			Freemium	284 million
	Science.gov	✓						✓											Free	several million
	Academia.eu	✓		✓			✓												Freemium	55 million
	OpenAlex	✓					✓									✓			Freemium	
	AceMap	✓			✓	✓	✓	✓						✓					Free	260 million
	PubTator3	✓		✓	✓											✓			Free	6 million
Benchm.	Papers with Code	✓												✓	✓				Free	154 thousand
	ScienceAgentBench										✓			✓	✓	✓			Free	
	ORKG Benchmarks					✓		✓						✓					Free	
	Huggingface	✓		✓		✓								✓	✓				Freemium	

Agenda

- ❑ Background
- ❑ AI-Enhanced Search System
- ❑ Paper Chat and Scientific QA
- ❑ Graph Based System
- ❑ Recommendation System
- ❑ **Takeaway**

Takeaway

- AI is transforming literature search.

Takeaway

- AI is transforming literature search.

- AI tools boost discovery but still require oversight.

Takeaway

- AI is transforming literature search.
- AI tools boost discovery but still require oversight.
- Four AI paradigms jointly redefine research workflows.

Takeaway

- AI is transforming literature search.
- AI tools boost discovery but still require oversight.
- Four AI paradigms jointly redefine research workflows.
- Future directions point to smarter, multimodal systems.

Future Direction

- Multimodal literature search
  - Integrating text and figures/tables
- Event-oriented summarization
  - Extracting and organizing key research events (e.g., discoveries, methods, results) for clearer insights
- Real-time updates & knowledge tracking
  - Continuous integration of new findings
- Integration with scientific knowledge graphs
  - Structured, interconnected research data

References

1. Gemini, Gemini Deep Research, 2026.
2. Flicke et al., Scholar Inbox: Personalized Paper Recommendations for Scientists, ACL Demo, 2025.
3. Gu et al., Forecasting high-impact research topics via machine learning on evolving knowledge graphs, Machine Learning: Science and Technology, 2025.
4. Agarwal et al., LitLLMs, LLMs for Literature Review: Are we there yet? Transactions on Machine Learning Research, 2025.
5. Khalid et al., Comprehensive review of academic search systems: evolution, analysis, and future research directions, Social Network Analysis and Mining, 2025.
6. He et al., PaSa: An LLM Agent for Comprehensive Academic Paper Search, ACL, 2025.
7. OpenAI, Deep Research System Card, 2025.
8. Li et al., ChemVLM: Exploring the Power of Multimodal Large Language Models in Chemistry Area, AAAI, 2025.
9. Li et al., BLADE: Enhancing Black-box Large Language Models with Small Domain-Specific Models, AAAI, 2025.
10. Zhang et al., SCITAT: A Question Answering Benchmark for Scientific Tables and Text Covering Diverse Reasoning Types, ACL-findings, 2025.
11. He et al., GIVE: Structured Reasoning of Large Language Models with Knowledge-Graph-Inspired Veracity Extrapolation, ICML, 2025.
12. Giglou et al., LLMs4Synthesis: Leveraging Large Language Models for Scientific Synthesis, JCSL, 2025.
13. Mario et al., Forecasting the future of artificial intelligence with machine learning-based link prediction in an exponentially growing knowledge network, Nature machine intelligence, 2023.
14. Kang et al., Taxonomy-guided Semantic Indexing for Academic Paper Search, EMNLP, 2024.
15. Ren et al., A Survey of Large Language Models for Graphs, KDD, 2024.
16. L´ala et al., PaperQA: Retrieval-Augmented Generative Agent for Scientific Research, arXiv, 2023.
17. Auer et al., The SciQA Scientific Question Answering Benchmark for Scholarly Knowledge, Scientific Reports, 2023.
18. Ko et al., A Survey of Recommendation Systems: Recommendation Models, Techniques, and Application Fields, electronics, 2022.
19. Bhagavatula et al., Content-Based Citation Recommendation, NAACL, 2018.