

AI-Supported Literature Search

TRANSFORMING SCIENCE WITH LARGE LANGUAGE MODEL

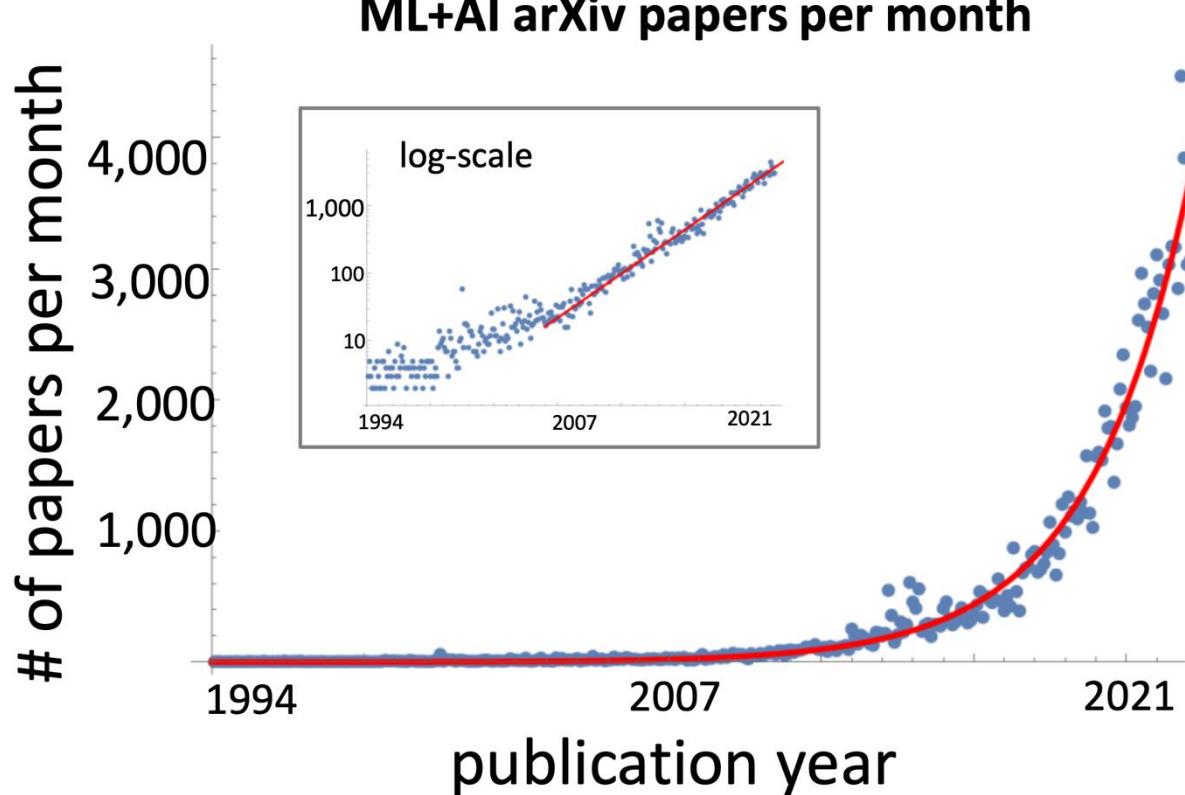


Tübingen.



Dr. Yong Cao
PostDoc, University of Tübingen
2025-07-26

Research Growth



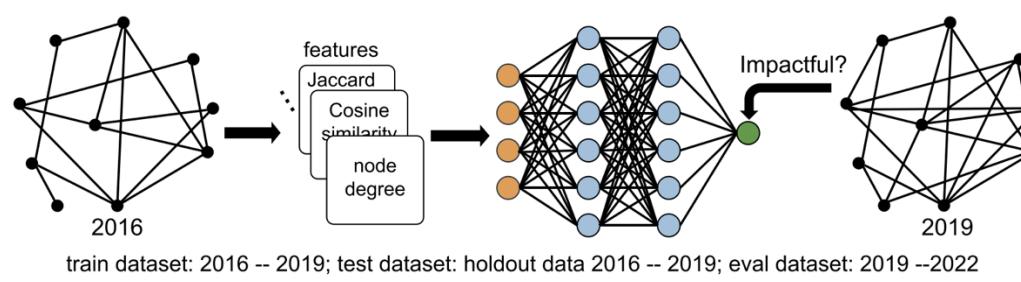
[Predicting the Future of AI with AI: High-Quality link prediction in an exponentially growing knowledge network, 2022](#)

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

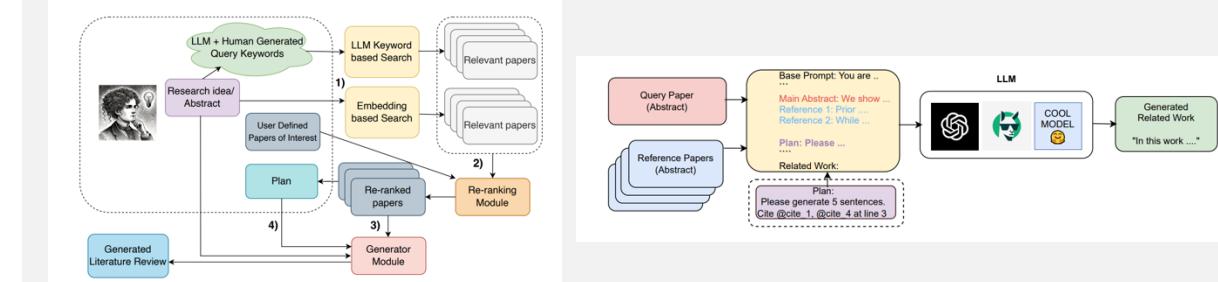
— Clay Shirky

Why AI/LLMs for Literature Search?

Uncover Emerging Trends



Semantic & Relation Understanding



Boost Research Efficiency and Increase Creativity

The Elicit platform interface includes a search bar, a library section, and a 'PRO' tab. A central feature is a knowledge graph visualization where nodes represent papers and edges represent connections like 'Connected Papers' or 'Assessing cross-cultural alignment between ChatGPT and human societies: An empirical study'.

The Scholar Inbox interface displays a 'Scholar Alert Digest 23/07' with a message to the user. It lists relevant articles such as 'Exploring Gender Bias in Large Language Models: An In-depth Dive into the German Language' and 'LingBench++: A Linguistically-Informed Benchmark and Reasoning Framework for Multi-Step Logic'. The inbox also includes a 'Translate to Chinese' option.



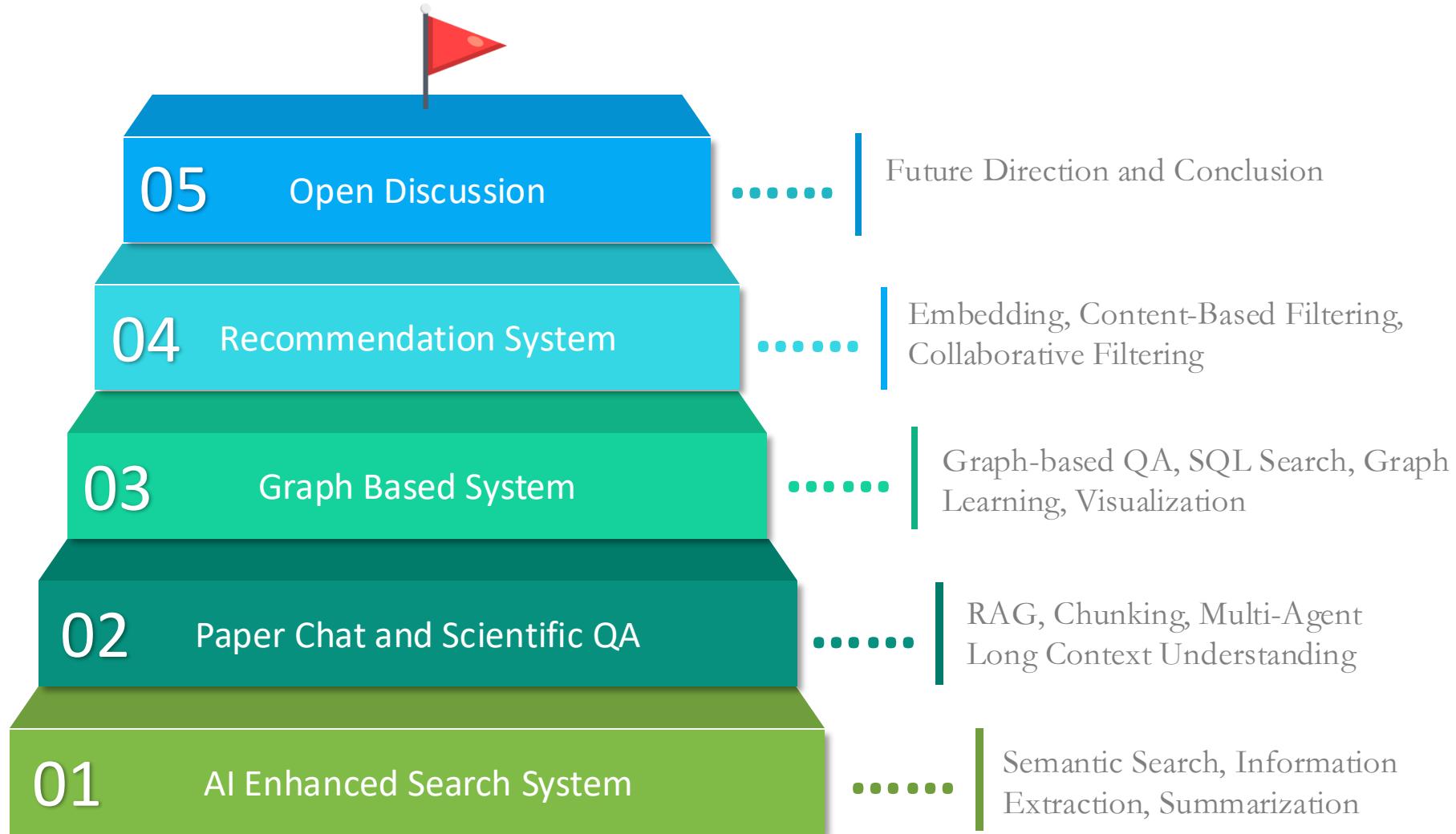
e.g., Scientific QA, Knowledge Graph Enhanced Search, Semantic Search, personalized Recommendation

[Forecasting high-impact research topics via machine learning on evolving knowledge graphs, 2025](#)

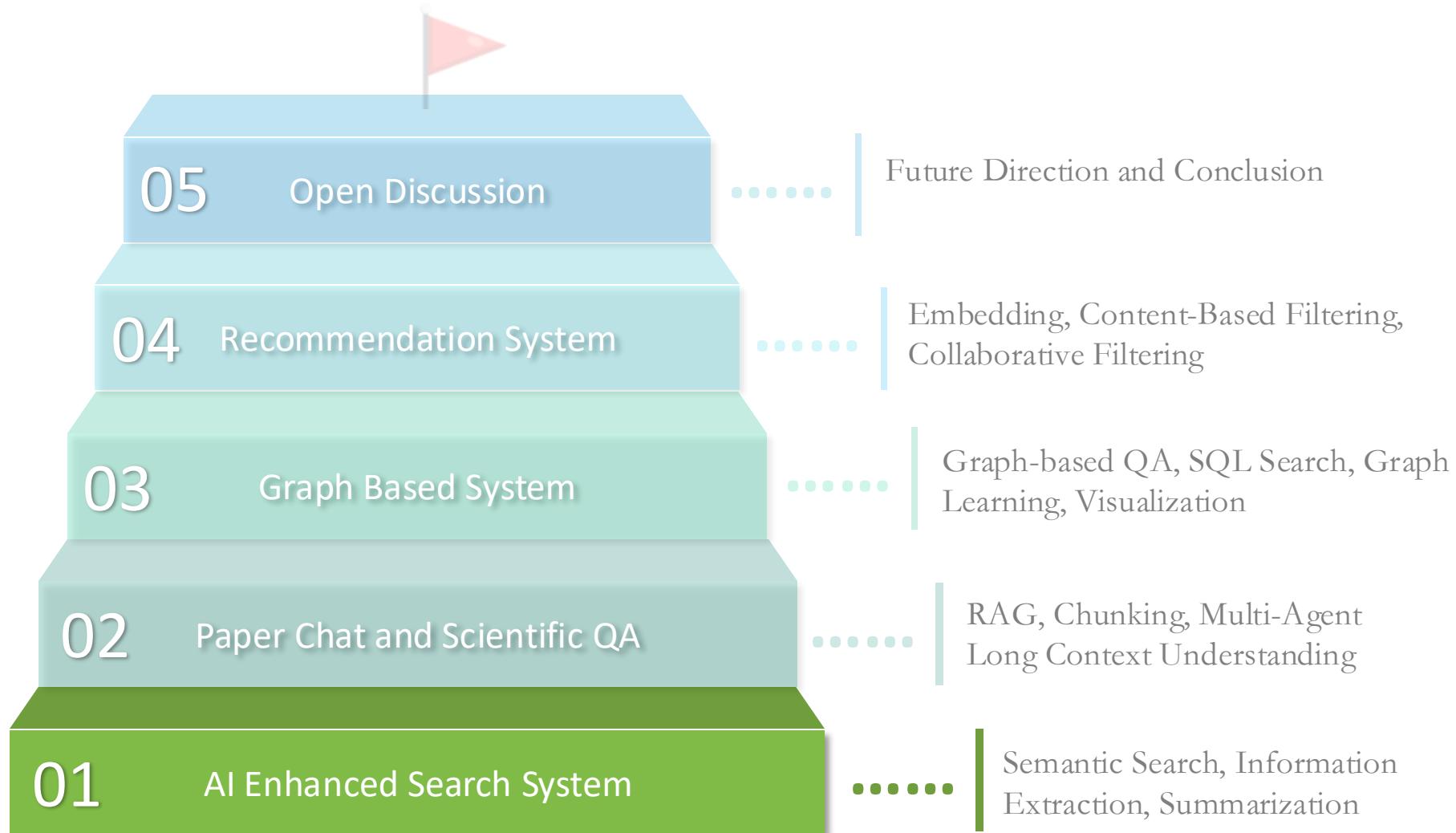
[LitLLMs, LLMs for Literature Review: Are we there yet? 2025](#)

[Elicit, Connected Papers, Scholar Inbox, ResearchTrend.ai](#)

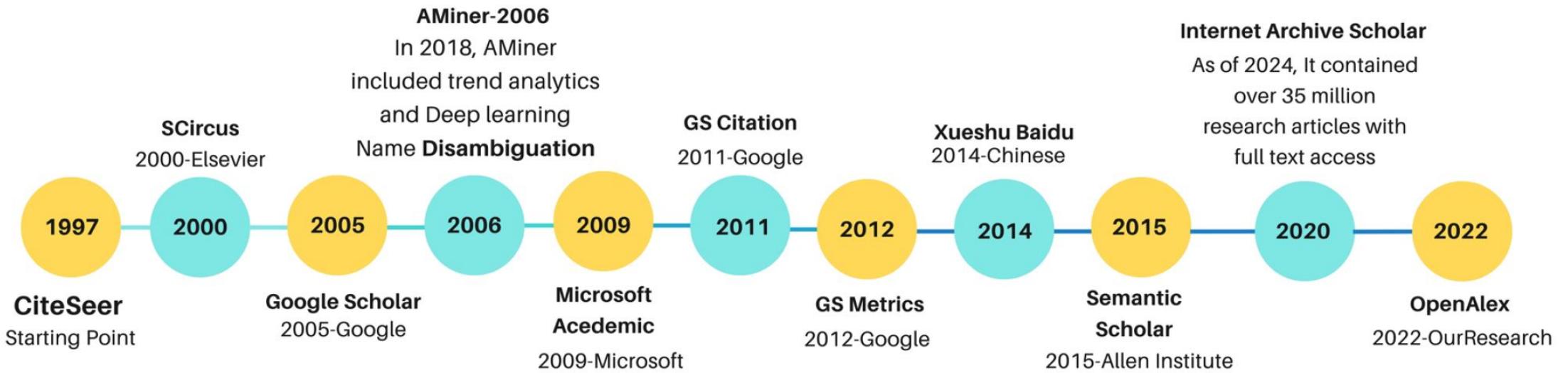
Outline



Outline



Evolution of academic search engines



Data Source

ResearchTrend.AI



SEMANTIC SCHOLAR

arxiv-sanity



arXiv

bioRxiv

THE PREPRINT SERVER FOR BIOLOGY

medRxiv

THE PREPRINT SERVER FOR HEALTH SCIENCES



Elicit



Ai2 OpenScholar

NotebookLM

1 - AI Enhanced Search System

The screenshot shows the homepage of the Undermind AI Enhanced Search System. At the top left is the Undermind logo, which consists of a stylized purple Greek letter μ followed by the word "Undermind". To the right are links for "Pricing" and "Login", and a prominent blue button labeled "Try now →". Below the header, a large section features the headline "Condense weeks of research to minutes" in bold black text. A subtext below it reads: "An AI assistant that carefully explores the scientific literature for you. Find exactly what you need, no matter how complex." There are two buttons at the bottom of this section: a blue "Try now" button and a white "Learn more" button. Further down, there's a search bar with the placeholder "Describe what you're looking for..." and a blue search icon. Below the search bar are three category suggestions: "Experimental evidence of phonon-...", "Computational models of hippoc...", and "Laboratory experiments simulating...".

μ Undermind

Pricing Login Try now →

Condense weeks of research to minutes

An AI assistant that carefully explores the scientific literature for you. Find exactly what you need, no matter how complex.

Try now Learn more

Describe what you're looking for...

Experimental evidence of phonon-... Computational models of hippoc... Laboratory experiments simulating...

1 - AI Enhanced Search System

The screenshot shows the Undermind AI Enhanced Search System interface. At the top left is the Undermind logo (μ) followed by the word "Undermind". To the right are "Pricing", "Login", and a "Try now →" button. Below the header are navigation links: "Elicit", "Recent", "Library", "Alerts" (with a "PRO" badge), "Upgrade" (with a star icon), "Help", and a user account link "yongcao2018@gmail.com".

The main search area features three tabs: "Research report" (selected), "Systematic review" (with a "PRO" badge), and "Find papers". Below these tabs is a large input field with placeholder text: "Ask a research question to generate a structured research report". Underneath the input field is a sub-instruction: "Try a couple of free examples to see what this is all about". A row of three buttons follows: "GLP-1R mechanisms", "Magnesium effects on sleep", and "Online vs. in-person CBT". To the right of these buttons is a teal-colored button with a white arrow pointing right.

At the bottom of the main search area is a "More tools" section containing four buttons: "Upload and extract", "Summarize concepts", "Chat with papers", and "Create an alert" (with a "PRO" badge). There is also a large blue rectangular button at the bottom left of the main search area.

1 - AI Enhanced Search System

The screenshot shows the Undermind platform interface. At the top, there's a navigation bar with the Undermind logo, Pricing, Login, Try now →, Upgrade, Help, and a user account (yongcao2018@gmail.com). On the left, a sidebar has Elicit, Recent, Library, Alerts, PRO, + New Question, and Recent Questions (listing "Can you suggest 3 paper..."). The main content area displays a question: "Can you suggest 3 papers that study LLMs' abilities of generating new research ideas in NLP?". Below it, an Answer section provides three references:

1. "IdeaBench: Benchmarking Large Language Models for Research Idea Generation" (Guo et al. 2024¹) (Guo et al. 2024²) (Guo et al. 2024³)
This paper presents a benchmark for evaluating LLMs' ability to generate research ideas. The authors create a dataset of recent papers from five domains (Computer Science, Economics, Chemistry, Physics, and Medicine) and annotate them with future research ideas. They then evaluate four LLMs (Gemini, Claude-2, GPT-3.5, and GPT-4) on their ability to generate ideas that align with the target papers' ideas. The results show that LLMs can generate research ideas that are novel, relevant, and feasible to a significant extent.
2. "Can Large Language Models Unlock Novel Scientific Research Ideas?" (Kumar et al. 2024¹) (Kumar et al. 2024²) (Kumar et al. 2024³)
This paper investigates the potential of LLMs in generating future research ideas across five domains. The authors create a dataset of recent papers from the same five domains as in the previous paper and annotate them with future research ideas. They then evaluate four LLMs (Gemini, Claude-2, GPT-3.5, and GPT-4) on their ability to generate ideas that align with the target papers' ideas. The results show that LLMs can generate research ideas that are novel, relevant, and feasible to a significant extent.
3. "SciPIP: An LLM-based Scientific Paper Idea Proposer" (Wang et al. 2024)
This paper proposes a scientific paper idea proposer (SciPIP) that leverages LLMs to assist researchers in generating new ideas. SciPIP constructs a literature retrieval database and uses LLMs to generate novel and feasible ideas aimed at addressing problems within the given background. The authors demonstrate the effectiveness of SciPIP in generating new research ideas in NLP.

At the bottom, there are links for Privacy Policy and Terms of Use, and logos for Semantic Scholar and University of Washington.

1 - AI Enhanced Search System

The screenshot shows the Undermind platform interface. At the top, there's a navigation bar with 'Undermind' logo, 'Pricing', 'Login', 'Try now →', 'Upgrade', 'Help', and a user account section. Below the navigation is a main dashboard area with a green button '+ New Question', a sidebar with 'Recent Question' (a link to 'Can you suggest 3'), and a central panel titled 'Research starts here' which describes Consensus as an AI-powered academic search engine. The sidebar also includes links for 'New Thread' and 'Home'. On the right side of the dashboard, there's a 'Consensus' section with its logo, a search bar 'Ask the research...', and several buttons for 'Ask a research question', 'Draft an outline', 'Create a table', and 'Try the Consensus Meter'. The bottom of the dashboard features a 'SEMANTIC' logo and a 'Sign up' button.

**Co
res**

An AI ass
you. Find

+ New Question

Recent Question: Can you suggest 3

New Thread

Home

Research starts here

Consensus is the AI-powered academic search engine

Search & analyze 200M+ peer reviewed research papers

Transparent, reliable, and built to save you time

SEMANTIC

Ask the research...

Pro Filter

Ask a research question Draft an outline Create a table Try the Consensus Meter

Over 5 million researchers, students, and clinicians trust Consensus

Sign up Sign in

1 - AI Enhanced Search System

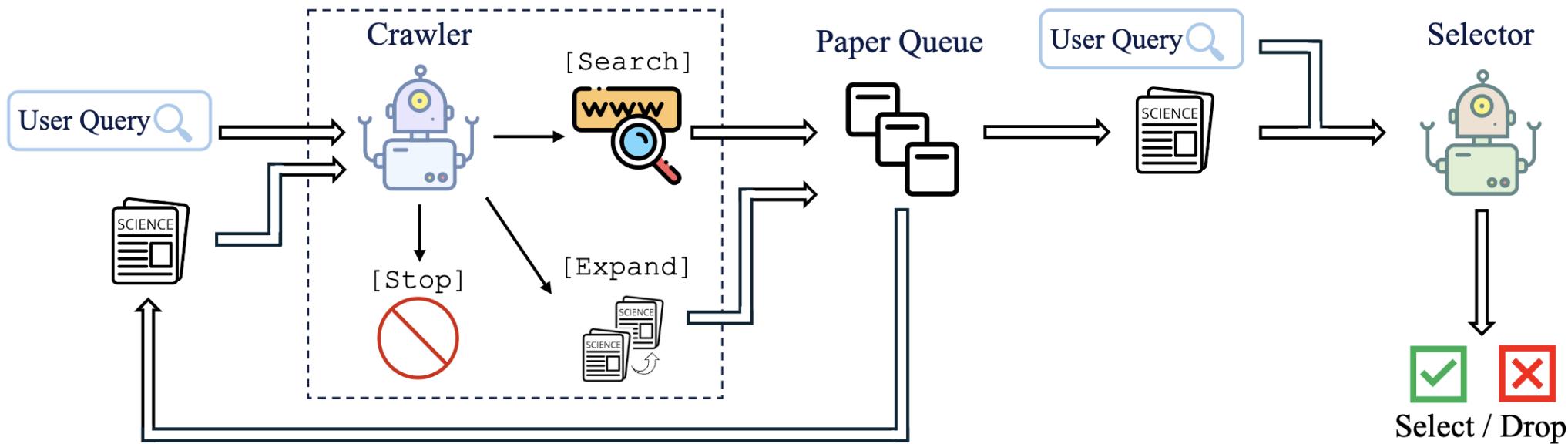
The image is a collage of screenshots from several AI research platforms:

- Undermind:** A dashboard with a green header button "+ New Question". It shows "Recent Question: C" and "Can you suggest 3". Buttons include "Elicit", "Recent", "Library", "Alerts", "PRO", "Pricing", "Login", "Try now →", "Upgrade", "Help", and "yongcao2018@gmail.com".
- Paperguide:** A dark-themed interface with a red header button "+ New Thread". It features a sidebar with "Research starts here" and "Consensus is the...". Buttons include "Home", "Sign up", "Features", "Deep Research", "Solutions", "Pricing", "EN", and "Hey, yong".
- Semantic Scholar:** A sidebar with "SEMANTIC", "UNIVERSITY of WI", "Privacy Policy", "Terms", "Sign up", and "Sign in".
- ResearchGate:** A sidebar with "Your All-in-One AI Research Assistant" and "Get Started for Free →". It includes a star rating "★★★★★ Rated by 1000+ Researchers and Universities".

On the left side of the collage, there is vertical text: "Co" and "res" stacked vertically, followed by "An AI ass" and "you. Find". Below these words is a large blue rectangular button.

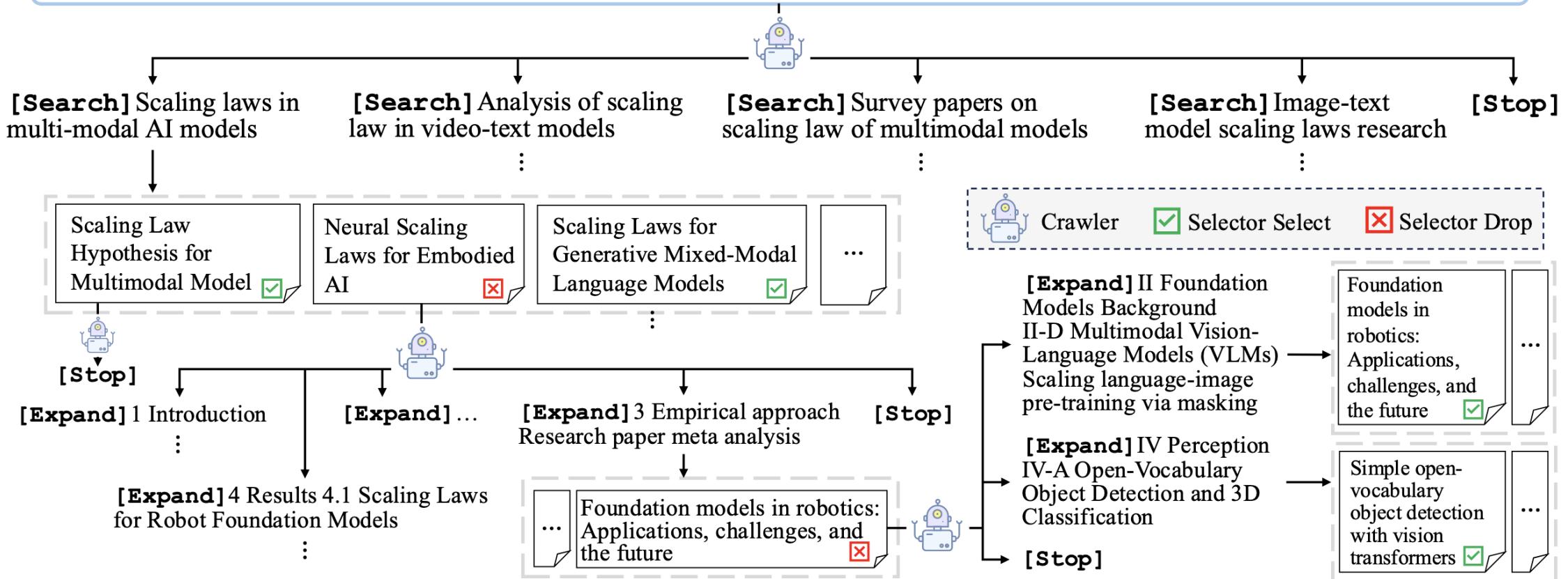
How AI enhance literature search platforms?

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



How AI enhance literature search platforms?

Is there any works that analyze the scaling law of the multimodal models, such as video-text, image-text models.



Performance

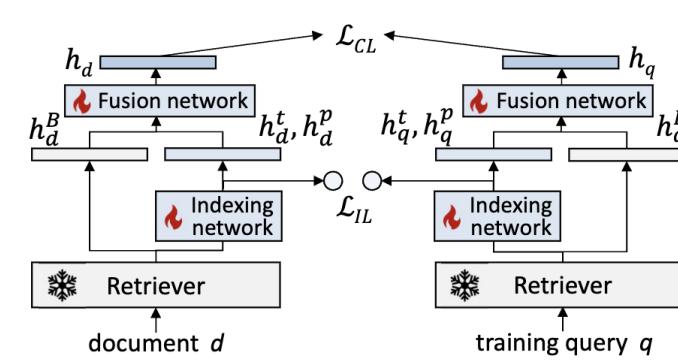
- Imitation Learning + Reinforcement Learning

$$\begin{aligned}\mathcal{L}_{\text{policy}}(\theta) = & \mathbb{E}_{\tau' \sim \pi_{\theta}^{\text{old}}} \left[\min \left(\frac{\pi_{\theta}(a_t | s_t)}{\pi_{\theta}^{\text{old}}(a_t | s_t)} \hat{A}(s_t, a_t), \right. \right. \\ & \left. \left. \text{clip} \left(\frac{\pi_{\theta}(a_t | s_t)}{\pi_{\theta}^{\text{old}}(a_t | s_t)}, 1 - \epsilon, 1 + \epsilon \right) \hat{A}(s_t, a_t) \right) \right], \\ \mathcal{L}_{\text{value}}(\phi) = & \mathbb{E}_{\tau' \sim \pi_{\theta}^{\text{old}}} \left[\max \left((\hat{R}_t - \hat{V}_{\phi}(s_t))^2, \right. \right. \\ & \left. \left. (\hat{R}_t - \hat{V}_{\phi}^{\text{clip}}(s_t))^2 \right) \right], \\ \mathcal{L}_{\text{RL}}(\theta, \phi) = & \mathcal{L}_{\text{policy}}(\theta) + \eta \cdot \mathcal{L}_{\text{value}}(\phi)\end{aligned}$$

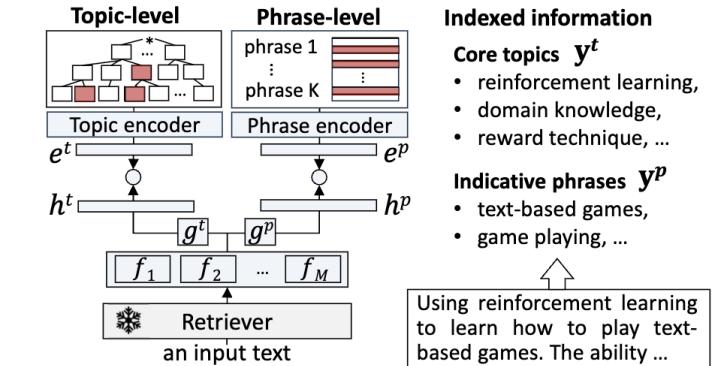
Method	Crawler Recall	Precision	Recall	Recall@100	Recall@50	Recall@20
Google	-	-	-	0.2015	0.1891	0.1568
Google Scholar	-	-	-	0.1130	0.0970	0.0609
Google with GPT-4o	-	-	-	0.2683	0.2450	0.1921
ChatGPT*	-	0.0507	0.3046	-	-	-
GPT-o1	-	0.0413	0.1925	-	-	-
PaSa-GPT-4o	0.7565	0.1457	0.3873	-	-	-
PaSa-7b	0.7931	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



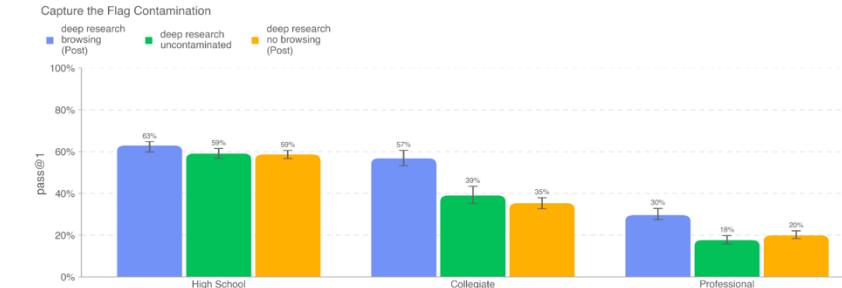
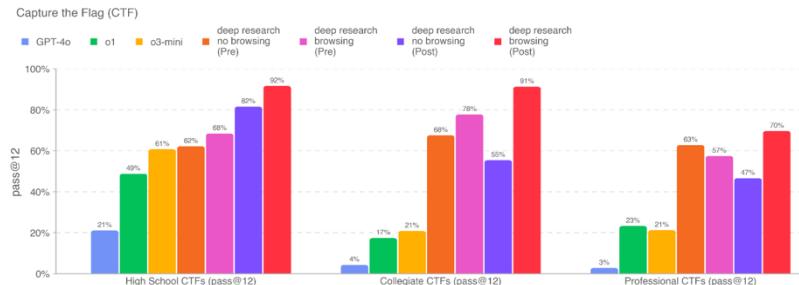
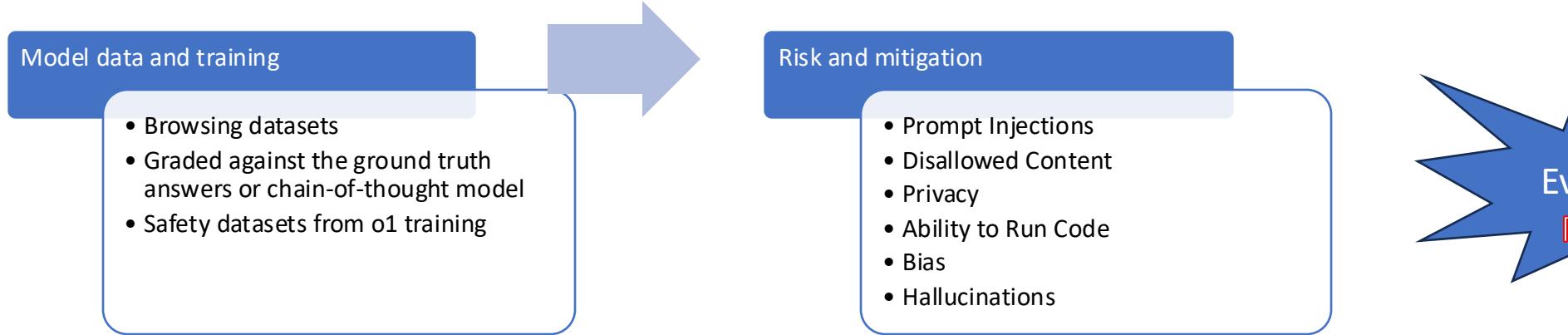
(a) Index-grounded fine-tuning



(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
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^{†*}	0.417^{†*}	0.144^{†*}	0.198^{†*}	0.633^{†*}	0.741^{†*}	0.447^{†*}	0.421^{†*}	0.104^{†*}	0.144^{†*}	0.578[†]	0.756[†]
TaxoIndex ++	0.469^{†*}	0.426^{†*}	0.158^{†*}	0.209^{†*}	0.621^{†*}	0.746^{†*}	0.449^{†*}	0.424^{†*}	0.105^{†*}	0.145^{†*}	0.581[†]	0.751[†]

Deep Research -- ChatGPT / 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

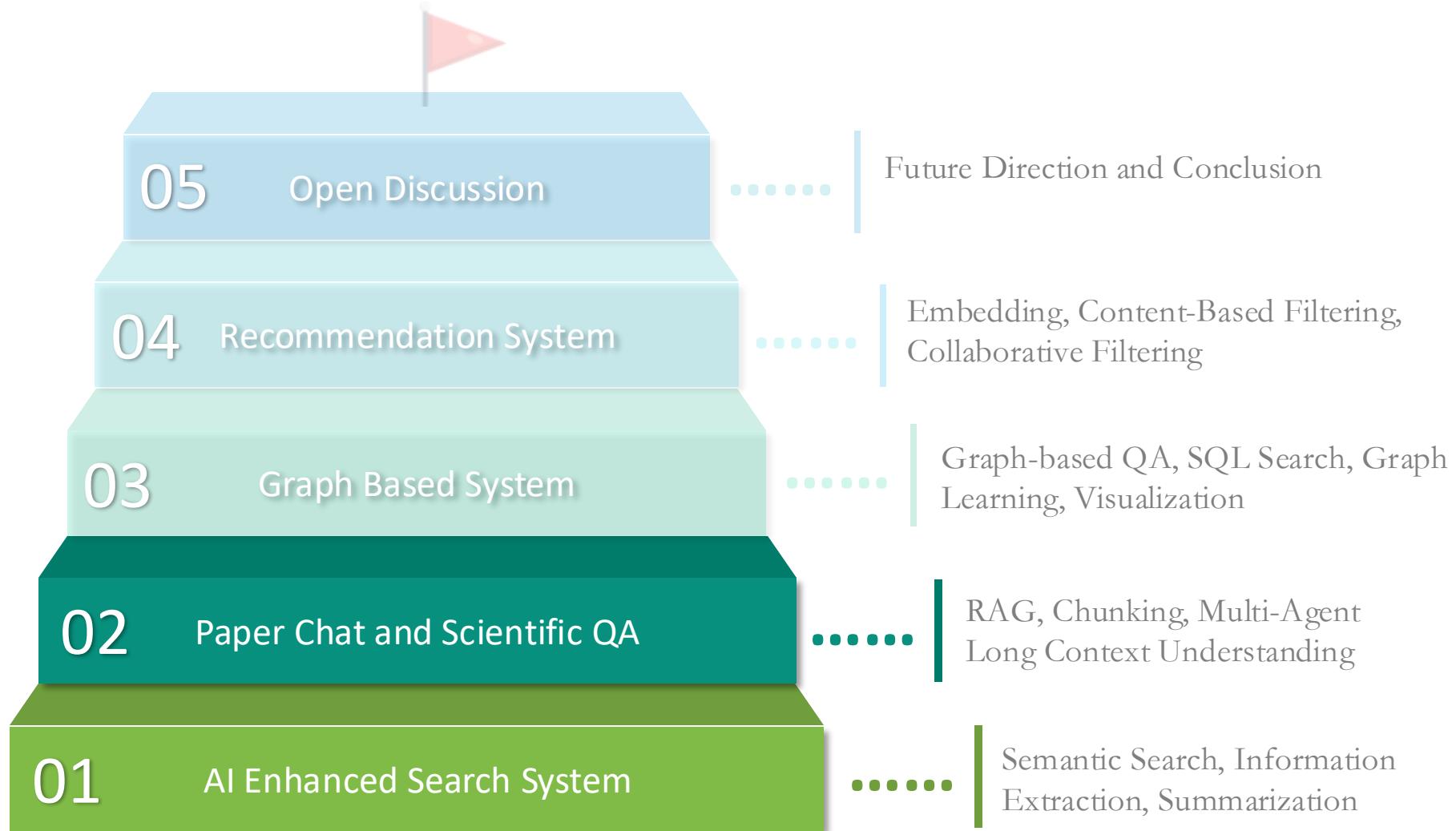
Reporting

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

1 (Summary) - 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

Outline



2 - Paper Chat and Scientific QA

The image shows a screenshot of the ChatGPT interface. At the top left is a navigation bar with icons for Home, ChatGPT (selected), AI Assistant, and Help. To the right of the ChatGPT icon is a dropdown menu labeled "ChatGPT" with a downward arrow. Further to the right is a "Get Plus" button. On the far right of the top bar are three small circular icons: a question mark, a gear, and a refresh symbol.

The main area features a large input field with the placeholder text "What's on your mind today?". Below this input field is a rounded rectangular button with the text "Ask anything" and a plus sign icon. To the right of this button are two more small circular icons: a question mark and a gear.

At the bottom left of the main area, there is a small circular profile picture of a person with dark hair and a beard. The background of the slide is white, and there is a thin red horizontal line at the very top.

2 - Paper Chat and Scientific QA

The screenshot shows a digital library interface with a sidebar and a main content area.

Sidebar:

- ChatGPT (dropdown)
- Recent
- Reading list (15)
- Discover
- My library**
- Computer science (27)
- Economics (11)
- Machine learning (34)
- Health
- Biology (8)
- LLMs (10)
- Psychology (18)
- AI (9)
- Sort (34)
- Webpages (12)
- Crypto
- Meta (6)
- Bioinformatics (1)
- Sociology (1)
- Cyber security (1)
- Biotech
- Statistics (3)
- Philosophy (1)
- Knowledge graphs
- History
- Probability (2)
- Physics
- Creativity (18)
- Astronomy
- Classics
- Hardware
- Uncategorized (2)
- Ecology (2)
- Cognition (1)
- New category +
- Pro
- Trash

Main Content Area:

Computer science (eye icon)

Collection of papers and articles spanning various subfields of computer science. This library covers topics from foundational concepts to cutting-edge developments, with a particular emphasis on machine learning, artificial intelligence, data analysis, and algorithms.

Authors:

- Scott Aaronson 1
- Tanishq Mathew Abraham 1
- Maneesh Agrawala 1
- Jessica R. Andrews-Hanna 1
- Apple 1
- Frank Arute 1
- Kunal Arya 1
- Ryan Babbush 1
- Dave Bacon 1
- Max Bain 1

Canvases (5) (+ Create):

- Programming GPUs (8 nodes · 11 days ago)
- Universal media machines (50 nodes · 4 months ago)
- Turing machines (1 node · last month)
- New canvas (14 nodes · 3 months ago)
- Ness (397 nodes · 4 months ago)

Items (27) (+ Add):

- Statistical Modeling: The Two Cultures (Computer science · 2001 · Leo Breiman)
- The neural basis of loss aversion in decision-making under risk (Computer science · 2007 · Sabrina M. Tom, Craig R. Fox, Christopher...
- The default network and self-generated thought: component processes, d... (Computer science · 2014 · Jessica R. Andrews-Hanna, Jonathan Small...
- Learning to (Learn at Test Time): RNNs with Expressive Hidden States (Computer science · 2024 · Yu Sun, Xinhao Li, Karan Dalal, Jiarui X...
- Integrated Multi-omics Analysis Using Variational Autoencoders: Appl... (Computer science · 2019 · Xiaoyu Zhang, Jingqing Zhang, Kai Sun, X...
- LMExplainer: Grounding Knowledge and Explaining Language Models (Computer science · 2023 · Zichen Chen, Jianta Chen, Yuanyuan Chen,...)
- Scaling Monosematicity: Extracting Interpretable Features from Claude... (Computer science ·)
- Mapping the Mind of a Large Language Model (Computer science · 2024)
- Mapping inter-city trade networks to maximum entropy models using elec... (Computer science · 2024 · Cesar I. N. Sampai Filho, Rilder S. Pir...
- Reconstructing the Mind's Eye: fMRI-to-Image with Contrastive Learning... (Computer science ·)

2 - Paper Chat and Scientific QA

The screenshot shows the ChatGPT interface with a sidebar on the left and a main content area on the right.

Left Sidebar:

- ChatGPT ▾
- Recent
- Reading list
- Discover
- My library
 - Computer science
 - Economics
 - Machine learning
 - Health
 - Biology
 - LLMs
 - Psychology
 - AI
 - Sort
 - Webpages
 - Crypto
 - Meta
 - Bioinformatics
 - Sociology
 - Cyber security
 - Biotech
 - Statistics
 - Philosophy
 - Knowledge graphs
 - History
 - Probability
 - Physics
 - Creativity
 - Astronomy
 - Classics
 - Hardware
 - Uncategorized
 - Ecology
 - Cognition
- New category +
- Pro
- Trash

Main Content Area:

Top bar: My Library / Computer science. Get Plus button.

Header: LLMCBench: Large Language Model Compression Benchmark

Left panel (Sources):

- + Add
- Discover
- Select all sources
- LLMCBench.pdf

Right panel (Chat):

- LLMCBench: Large Language Model Compression Benchmark
- 1 source
- The document introduces LLMCBench, a novel benchmark designed to systematically evaluate and compare various large language model (LLM) compression techniques. It addresses the current challenges in LLM deployment, such as high computational demands and limited, inconsistent evaluation standards across existing research. LLMCBench establishes six distinct evaluation tracks, including compression performance, generalization ability, training and inference consumption, hardware acceleration, and model trustworthiness, to offer a comprehensive analysis. By benchmarking mainstream sparsification and quantization methods across diverse models, datasets, and deployment platforms, the paper aims to provide actionable insights for the design and practical application of more efficient LLM compression algorithms. Ultimately, it seeks to standardize evaluation protocols and guide future research toward making LLMs more accessible for real-world use.
- Save to note
- Start typing... 1 source
- NotebookLM can be inaccurate; please double check its responses.
- What are the most significant challenges hindering the practical deployment of large language models?

Bottom panel (Studio):

- Analytics
- Share
- Settings
- PRO
- c

Audio Overview: Create an Audio Overview in more languages! Learn more

Deep Dive conversation: Two hosts. Customize Generate

Notes:

- + Add note
- Study guide
- Briefing doc
- FAQ
- Timeline

Saved notes will appear here. Save a chat message to create a new note, or click Add note above.

2 - Paper Chat and Scientific QA

The screenshot shows a user interface for a scientific paper reading and discussion platform. At the top left is a navigation bar with icons for ChatGPT, My Library, Recent, Reading list, Discover, Sources, and a search bar. The main title is "Computer science". A "Get Plus" button is visible at the top right.

The central area displays a paper titled "LLMCBench: Large Language Model Compression Benchmark" from "enago Read". The article is by Ashot H. Gevorgyan and Francesco Simoni. The abstract discusses self-induced oscillations and Bragg resonance in heliconical cholesteric liquid crystals. The text is annotated with several blue callout boxes containing AI-generated summaries and questions. One box on the right side reads: "the light wavelength is close to the Bragg resonance. The study demonstrates that self-induced oscillations take place at lower intensities on the red side edge of the Bragg resonance while stable values of transmittivity are still observed on the blue side edge. A further increase in intensity leads to oscillations of lower amplitude on the blue side and an irregular behavior of the transmission on the red side. At higher intensities, the Bragg resonance disappears and transmission becomes unstable for any light wavelength. A simple phenomenological model is proposed to account for the onset of the oscillations and the asymmetry of the behavior at the opposite side of the Bragg resonance. The study also points out that the static electric field is a driving parameter to switch from stable to oscillatory to irregular behavior in the transmittivity at a given light wavelength. The research received no external funding and the authors declare no conflict of interest. Data are available from the authors." Another box at the bottom left says: "The research reports the results of a study on the nonlinear light propagation of a beam traveling along the helix...".

At the bottom right, there is a text input field labeled "Ask your question..." with a blue "Ask" button.

2 - Paper Chat and Scientific QA

The screenshot shows a digital library interface with a sidebar on the left and a main content area on the right.

Left Sidebar:

- ChatGPT dropdown
- Recent, Reading list, Discover buttons
- My library section:
 - Computer science (selected)
 - Economics
 - Machine learning
 - Health
 - Biology
 - LLMs
 - Psychology
 - AI
 - Sort
 - Webpages
 - Crypto
 - Meta
 - Bioinformatics
 - Sociology
 - Cyber security
 - Biotech
 - Statistics
 - Philosophy
 - Knowledge graphs
 - History
 - Probability
 - Physics
 - Creativity
 - Astronomy
 - Classics
 - Hardware
 - Uncategorized
 - Ecology
 - Cognition
- New category +
- Pro, Mapping Int., Reconstruct buttons

Main Content Area:

Top bar: My Library / Computer science, Get Plus button, Analytics, Share, Settings, PRO, User icon.

Search bar: Search, Filter, Edit, Info, Upload buttons.

Section: Explainpaper

PDF preview: LLMCBench.pdf, 100% zoom, page 1 / 13.

Article Title: **LLMCBench: Benchmarking Large Language Model Compression for Efficient Deployment**

Authors: Ge Yang¹, Changyi He¹, Jinyang Guo^{1*}, Jianyu Wu¹, Yifu Ding¹, Aishan Liu¹, Haotong Qin², Pengliang Ji³, Xianglong Liu¹

Institutions: ¹ Beihang University, ² ETH Zurich, ³ Carnegie Mellon University

Abstract

Although large language models (LLMs) have demonstrated their strong intelligence ability, the high demand for computation and storage hinders their practical application. To this end, many model compression techniques are proposed to increase the efficiency of LLMs. However, current researches only validate their methods on limited models, datasets, metrics, etc., and still lack a comprehensive evaluation under more general scenarios. So it is still a question of which model compression approach we should use under a specific case. To mitigate this gap, we present the Large Language Model Compression Benchmark (LLMCBench), a rigorously designed benchmark with an in-depth analysis for LLM compression algorithms. We first analyze the actual model production requirements and carefully design evaluation tracks and metrics. Then, we conduct extensive experiments and comparison using multiple mainstream LLM compression approaches. Finally, we perform an in-depth analysis based on the evaluation and provide useful insight for LLM compression design. We hope our LLMCBench can contribute insightful suggestions for LLM compression algorithm design and serve as a foundation for future research. Our code is available at <https://github.com/AboveParadise/LLMCBench>.

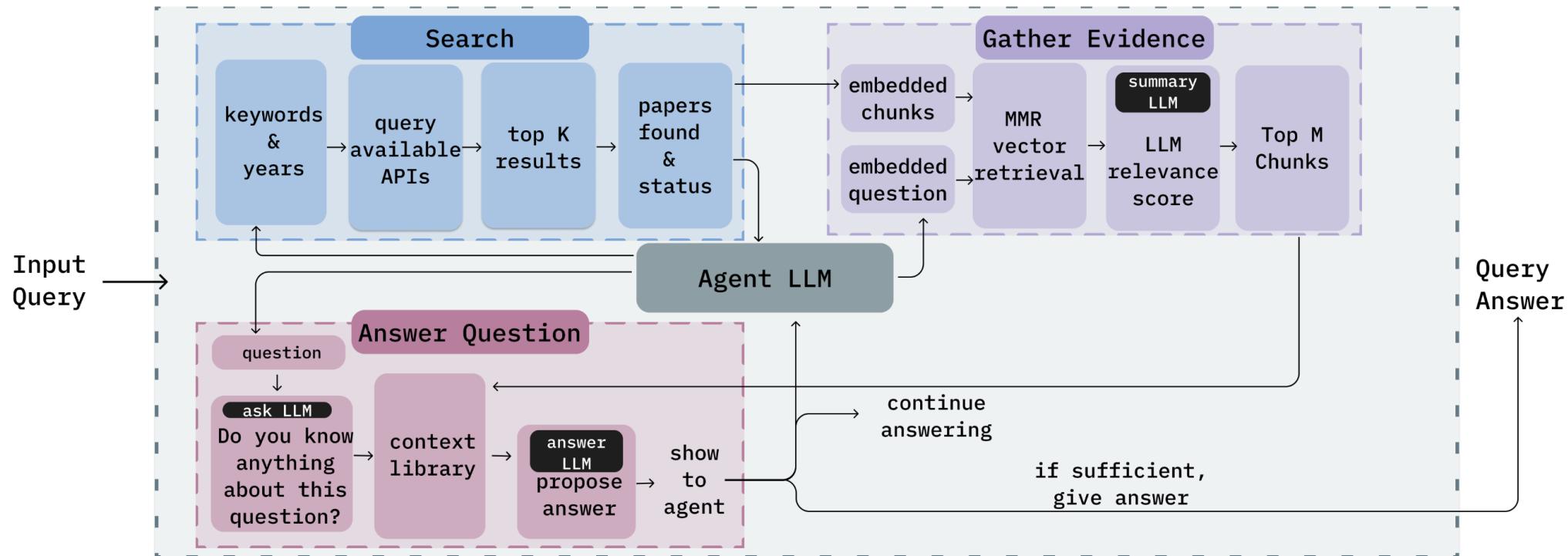
Section: 1 Introduction

Text: Recently, large language models (LLMs) have attracted increasing attention because of their strong intelligence ability. While it achieves excellent performance, the huge computation and storage burden hinders the practical usage of these LLMs. To solve this problem, many model compression

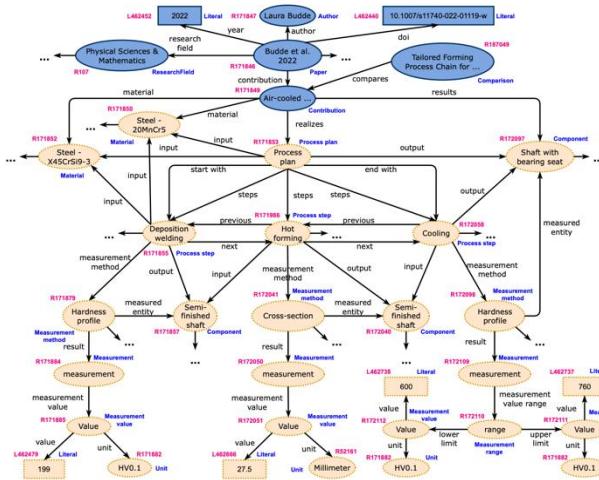
Bottom right: UNDERGRAD, Ask a follow-up, like, dislike buttons.

PaperQA – An Agent-Based RAG System

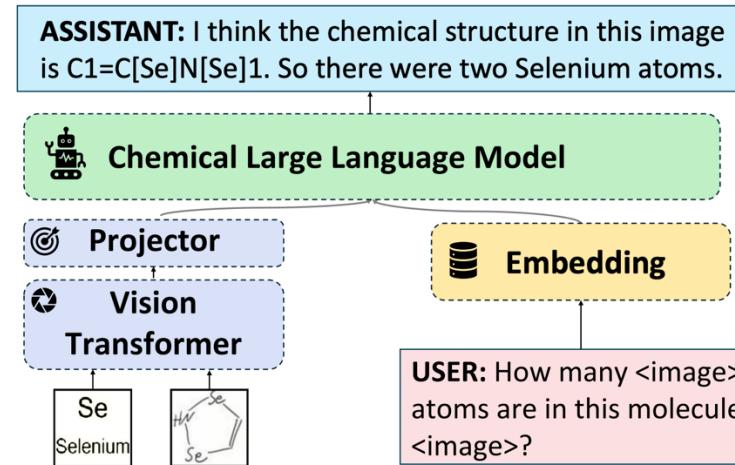
- Aims to address the hallucinations and knowledge update.
- Key Components:
 - Search, Gather Evidence, Answer Question



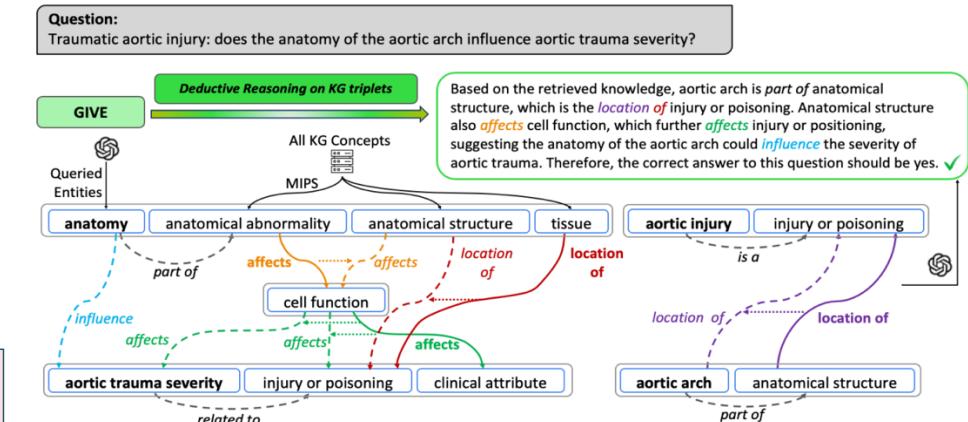
Mostly focus on...



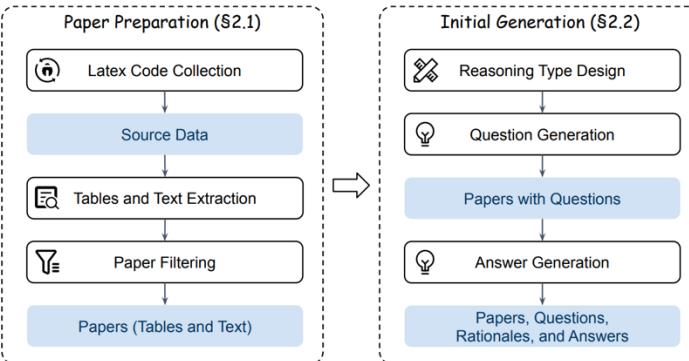
KG Integration



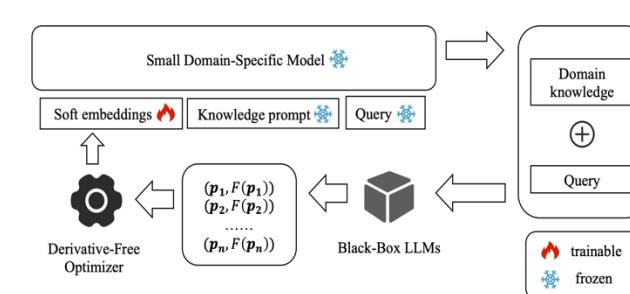
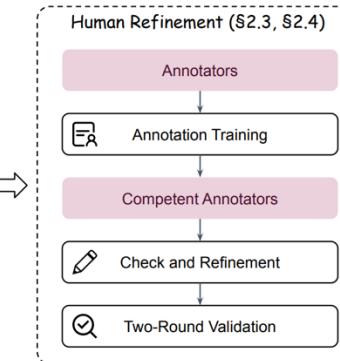
Multi-domal QA



Complex Reasoning



Benchmarking



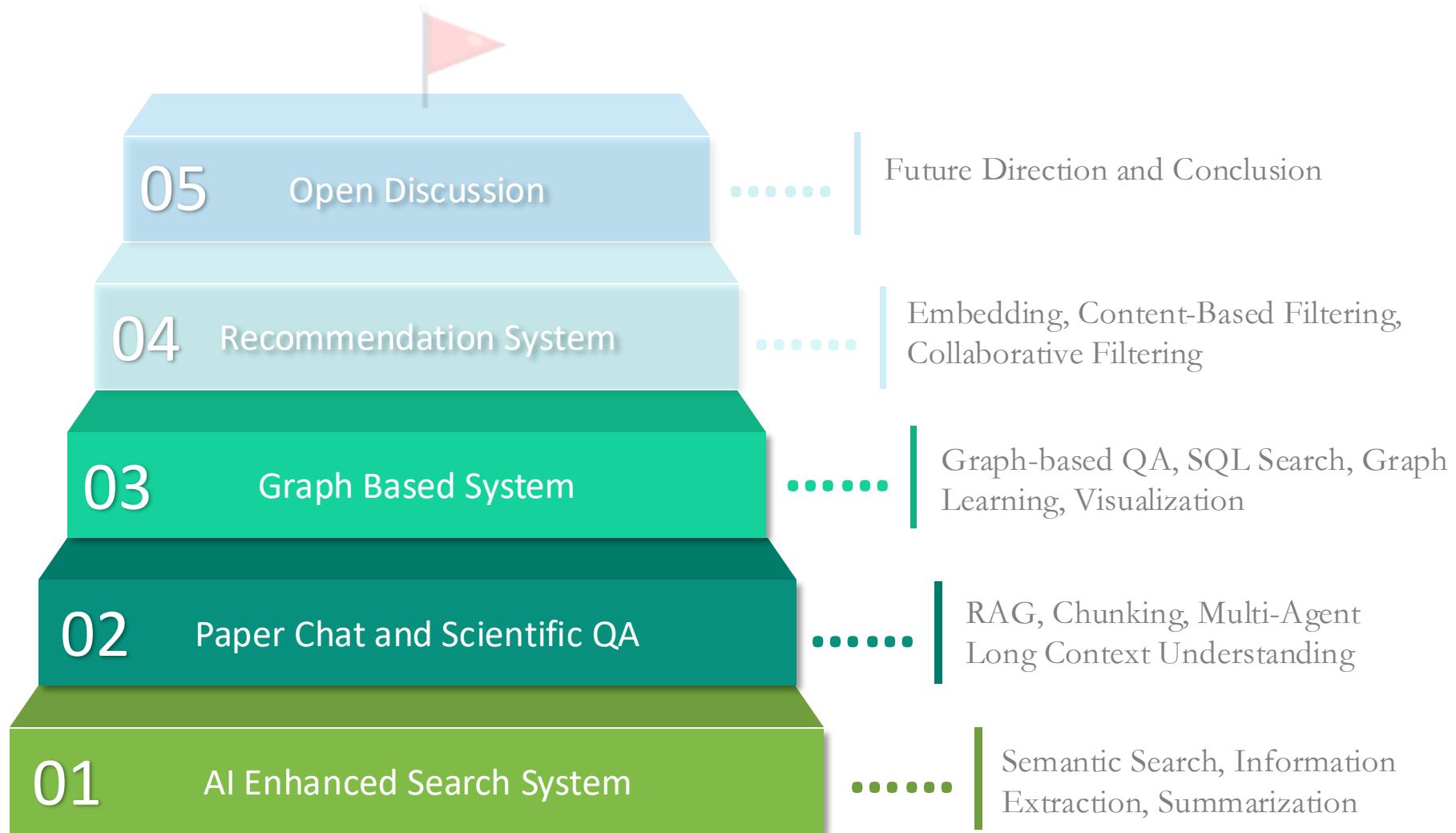
Domain knowledge

- [The SciQA Scientific Question Answering Benchmark for Scholarly Knowledge, 2023](#)
- [ChemVLM: Exploring the Power of Multimodal Large Language Models in Chemistry Area, 2025](#)
- [BLADE: Enhancing Black-box Large Language Models with Small Domain-Specific Models, 2024](#)
- [SCITAT: A Question Answering Benchmark for Scientific Tables and Text Covering Diverse Reasoning Types, 2024](#)
- [GIVE: Structured Reasoning of Large Language Models with Knowledge-Graph-Inspired Veracity Extrapolation, 2025](#)

2 (Summary) - 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

Outline



3 - Graph Based System

The image shows the homepage of Connected Papers. At the top left is the logo 'CONNECTED PAPERS'. To the right are social sharing icons for 'Share' and 'Follow', and links for 'About', 'Pricing', and 'Log in'. The main title 'Explore connected papers in a visual graph' is centered above a search bar. Below the search bar is a button labeled 'Build a graph'. A section titled 'You can try:' lists five ways to start a graph: 'Paper DOI' (with a DOI icon), 'arXiv Paper URL' (with an arXiv icon), 'Paper Title' (with a document icon), 'Semantic Scholar Paper URL' (with a magnifying glass icon), and 'PubMed Paper URL' (with a PubMed icon). Below this is a horizontal line. Further down, there are three examples of visual graphs: one for 'arXiv Paper URL' showing a network of papers from 2017-2018, one for 'Semantic Scholar Paper URL' showing a network from 2017-2020, and one for 'PubMed Paper URL' showing a network from 1994-2018.

3 - Graph Based System

CONNECTED PAPERS

Share Follow About Pricing Log in

Explore connected papers in a visual graph

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

Scholar Profiles Institutional Profiles Highly Ranked Scholars Institutional Rankings Highly Cited Publications More ▾

Empowering scholars to accelerate research and collaboration
Profiles and Rankings of all Scholars, Institutions, Countries, Fields, Disciplines, and Specialties

What is ScholarGPS Key Features and Benefits ScholarGPS News and Analyses Quick Tips

ScholarGPS® gives unprecedented access to:

Scholar Profiles
Access the profiles of over 30 million academic, industrial, and government scholars from 120,000 institutions in over 200 countries. Claim and update your profile — for free!
[More](#)

Institutional Profiles
Access analytics for each of the 120,000 institutions (24,353 academic institutions) included in the ScholarGPS® database.
[More](#)

Field and Discipline Profiles
Access descriptions of, and detailed analytics for broad 14 Fields and 177 Disciplines. [More](#)

Specialty Profiles
Access descriptions of, and detailed analytics for over 350,000 distinct Specialties. [More](#)

3 - Graph Based System

CONNECTED PAPERS

Share Follow About Pricing Log in

Explore connected papers in a visual graph

ScholarGPS®
POWERING SCHOLARLY ANALYTICS

All ▾ Search for any scholar, Specialty, Discipline, Field, or Institution

SEARCH

Trv: Anil K. Iain | Michigan State University | Computer Science | Pattern recognition

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.

Highly Cited Papers

The deeper, more inspirational, and timely signals often hide beneath the surface, where intriguing ideas emerge and groundbreaking research begins.

CiteSpace brings them to light.

The screenshot shows the ScholarGPS website interface. At the top, there's a navigation bar with 'CONNECTED PAPERS' on the left, followed by social sharing icons (Share, Follow), 'About', 'Pricing', and 'Log in'. Below this is a search bar with the placeholder 'Search for any scholar, Specialty, Discipline, Field, or Institution' and a red 'SEARCH' button. Underneath the search bar, it says 'Trv: Anil K. Iain | Michigan State University | Computer Science | Pattern recognition'. The main content area features the 'CiteSpace' logo and a navigation menu with links to 'Blogs', 'Exemplars', 'Gallery', 'FAQs', 'Standard', 'Advanced', 'Glossary', 'Videos', and 'References'. To the right of the menu is a purple 'Login' button. The central part of the page has a large, colorful visualization of a research landscape with various clusters of data points and text labels like '#11 critical legal studies', '#49 war trauma survivor', '#0 biological weapon', '#20 latin', and '#10 military commission'. Overlaid on this visualization is a dark blue graphic for 'CiteSpace' with the text 'Highly Cited Papers', a quote about hidden signals, and the 'CiteSpace' logo. On the far left, there's a vertical sidebar with sections for 'Scholar', 'Space', and 'Time'.

3 - Graph Based System

CONNECTED PAPERS

Share Follow About Pricing Log in

Explore connected papers in a visual graph

ScholarGPS® POWERING SCHOLARLY ANALYTICS

All ▾ Search for any sci

Trv: Anil K. Jain | Michigan

CiteS

What captures our attention

Scho

#9 press coverage

#7 television-new

Space

BOURDIEU P (1977) GLASSRING C (1996)

ADAMS J (1986)

Sci2 Tool

A Tool for Science of Science Research & Practice

Home Download Documentation Ask An Expert Testimonials Developers

This site is deprecated, please download the latest version of Sci2 on [GitHub](#).

The Science of Science (Sci2) Tool is a modular toolset specifically designed for the study of science. It supports the temporal, geospatial, topical, and network analysis and visualization of scholarly datasets at the micro (individual), meso (local), and macro (global) levels.

[Registration required](#).

Download Sci² Tool

An Emergent Mosaic of Wikipedian Activity

2 / 4 Start Stop (3)

News

2018

- Jan 31, The [Sci2 \(Science of Science\) Tool v1.3](#) release provides support for Java 1.9 and removes deprecated plugins.

[Release Notes](#)

[More news](#)

Have a question?
Ask an Expert!

Log in | Register

The screenshot shows the Sci2 Tool homepage. At the top, there's a navigation bar with links for Share, Follow, About, Pricing, and Log in. Below it is a search bar with the placeholder 'Search for any sci'. The main header reads 'Explore connected papers in a visual graph' and features the 'ScholarGPS®' logo. To the left, there's a sidebar with sections for 'What captures our attention' and 'Space', displaying network graphs for 'press coverage' and 'television-new'. The central content area has a large banner for 'Sci2 Tool' with the subtitle 'A Tool for Science of Science Research & Practice'. A yellow bar below the banner says 'This site is deprecated, please download the latest version of Sci2 on GitHub.' Below this, there's a detailed description of the tool's purpose and a 'Registration required' link. A large green button labeled 'Download Sci² Tool' is prominently displayed. To the right, there's a large network graph titled 'An Emergent Mosaic of Wikipedian Activity'. At the bottom, there's a 'News' section for 2018, a 'More news' link, and a 'Have a question? Ask an Expert!' button with a question mark icon.

3 - Graph Based System

CONNECTED PAPERS

Share Follow About Pricing Log in

Explore connected papers in a visual graph

ScholarGPS® POWERING SCHOLARLY ANALYTICS

All Search for any sci... Trv: Anil K. Jain | Michigan

CiteScape ORKG Ask Search My library ORKG Ask

Log in Register English

What captures our attention?

This site is designed for the Science of Cities. It is specifically designed to support the temporal, spatial, and visual exploration of urban systems at the meso (local), macro (regional), and global levels.

Registration required

#9 press coverage #7 television-news

Space

ORKG Ask

Find research you are *actually* looking for

Ask your question...

ORKG Ask is a scholarly search and exploration system powered by **Vector Search, Large Language Models and Knowledge Graphs**. [Learn more.](#)

77,368,538 Items with abstracts

77,368,538 Items with abstracts

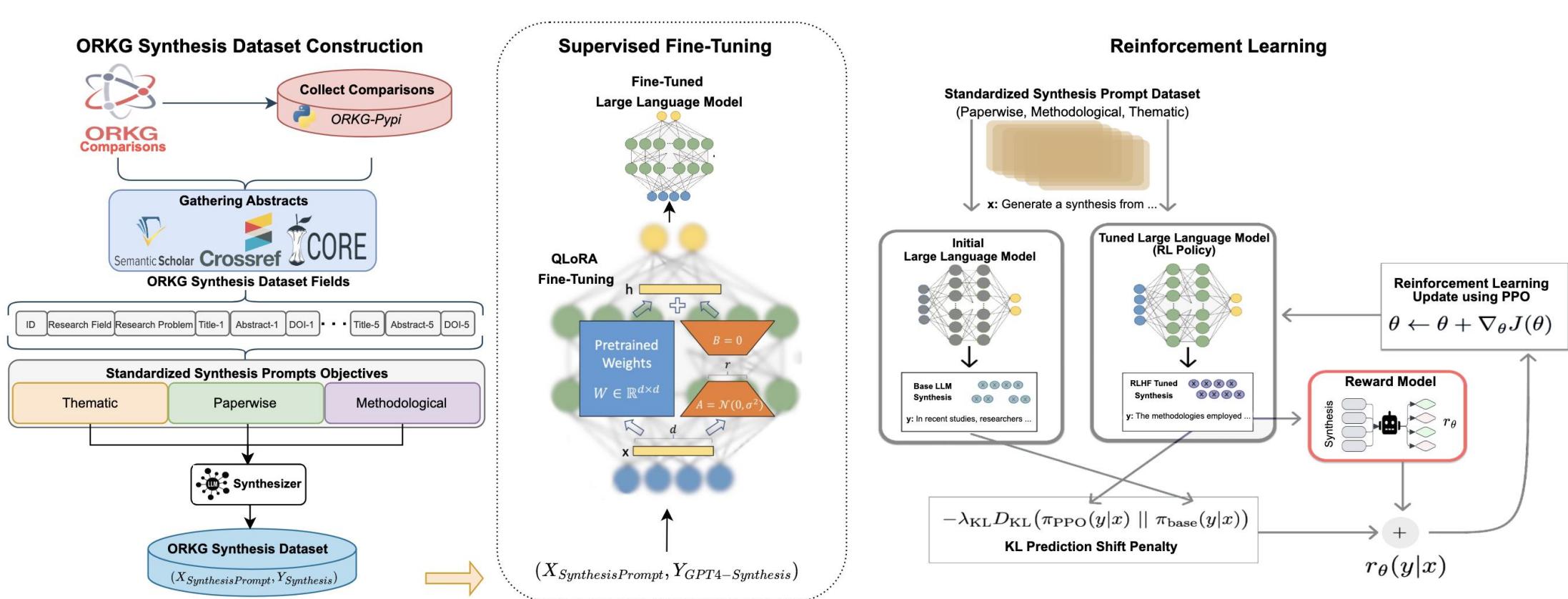
ORKG Ask is brought to you by

EULiST

Getting started

- How does the availability of green spaces impact community health and well-being?
- What role does storytelling play in shaping collective memory and identity?
- What impact does access to quality early childhood education have on lifelong outcomes?
- What are the effects of sleep deprivation on cognitive performance and overall health?

3 - Graph Based System



Performance and Conclusion

- SFT+RLAIF performs the best.

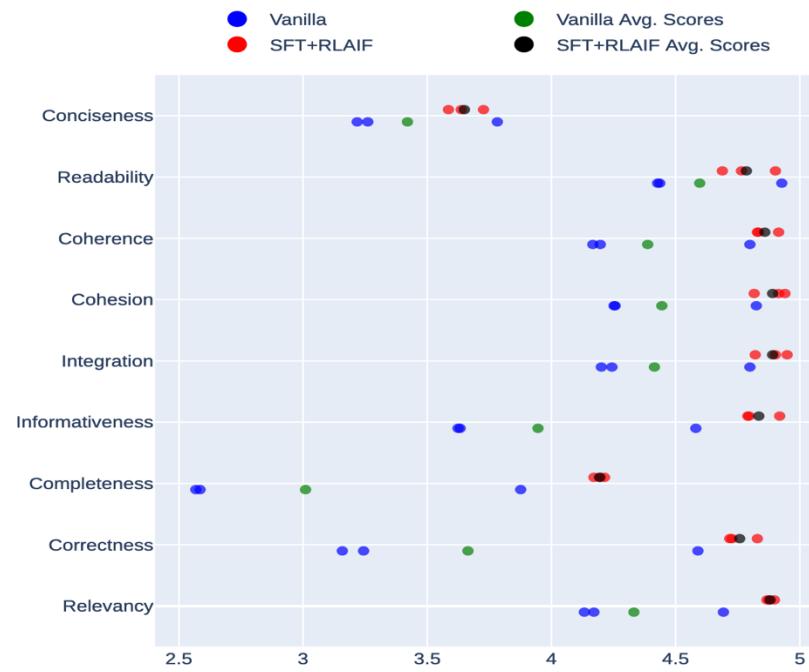
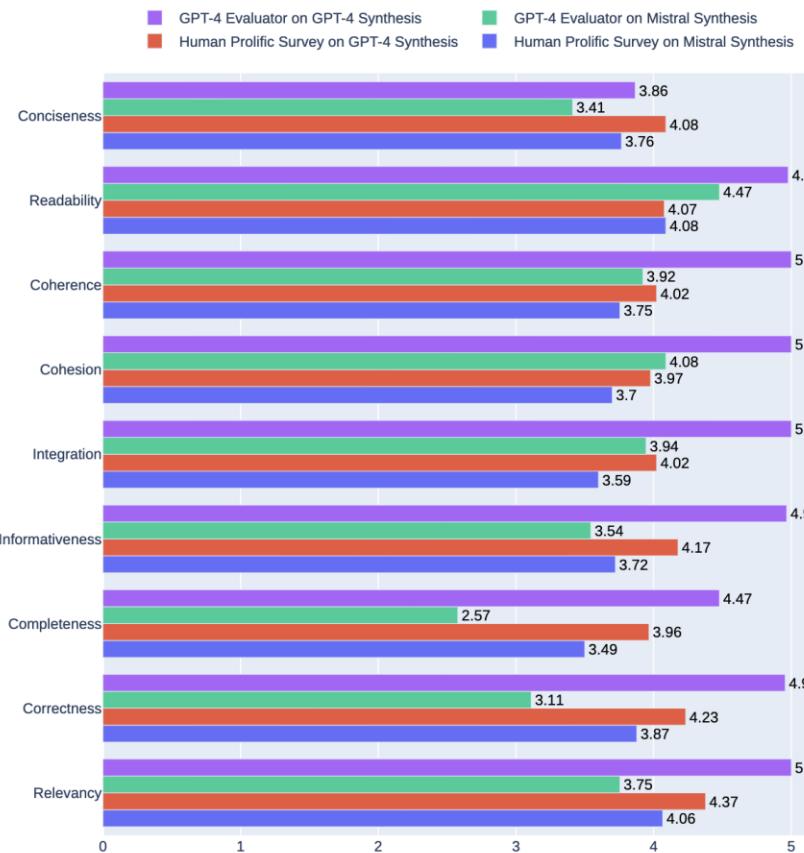
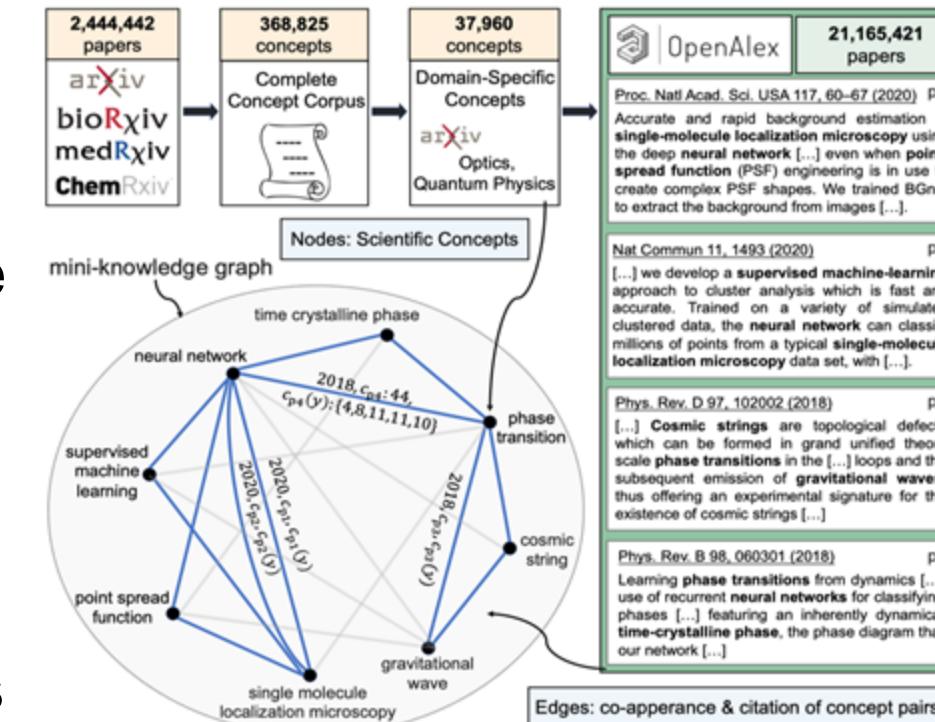


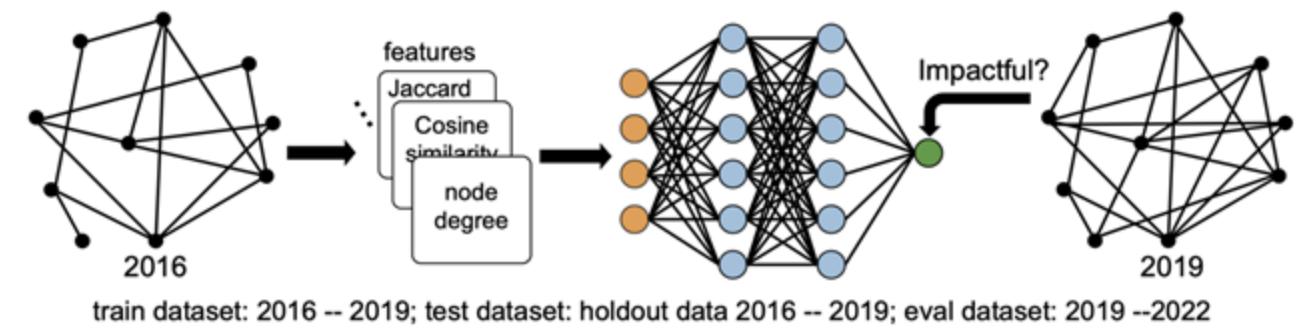
Figure 3: Consistency comparison of the GPT-4 evaluator between the *Vanilla* and *SFT+RLAIF* (w/ GPT-4 Features) models, assessed through three evaluations on the test set.

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

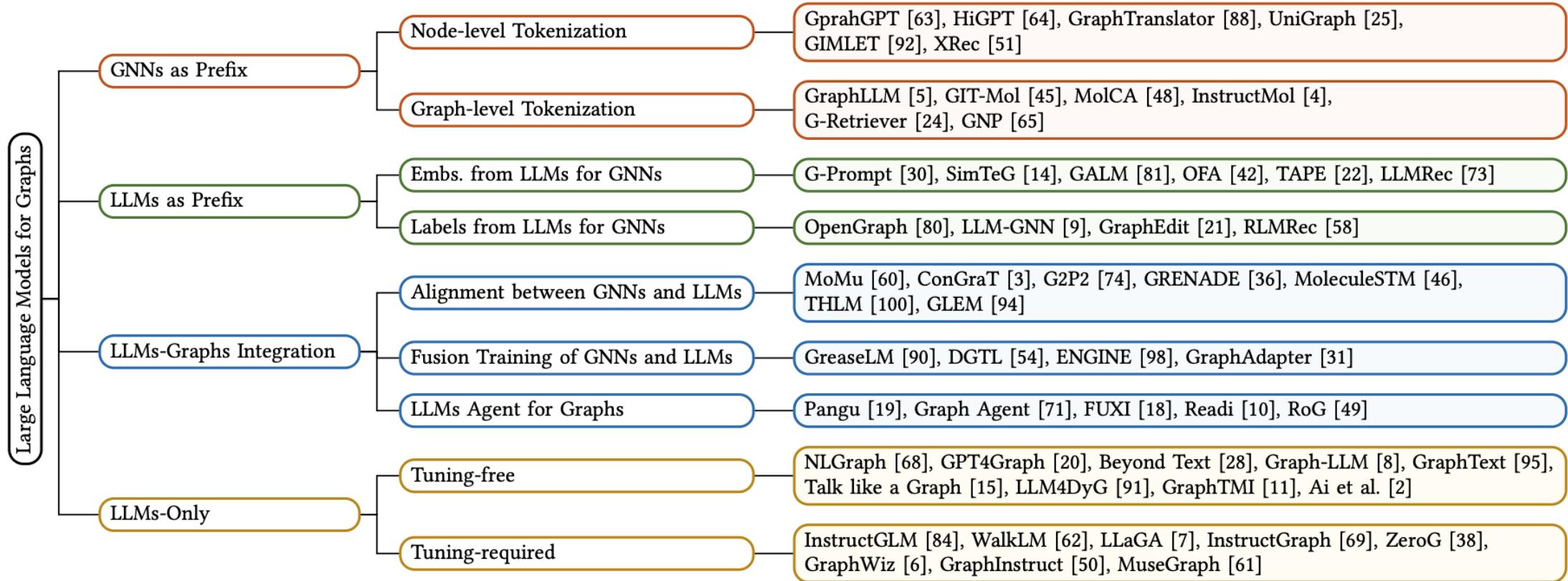


- Graph Learning?



Forecasting high-impact research topics via machine learning on evolving knowledge graphs, 2025

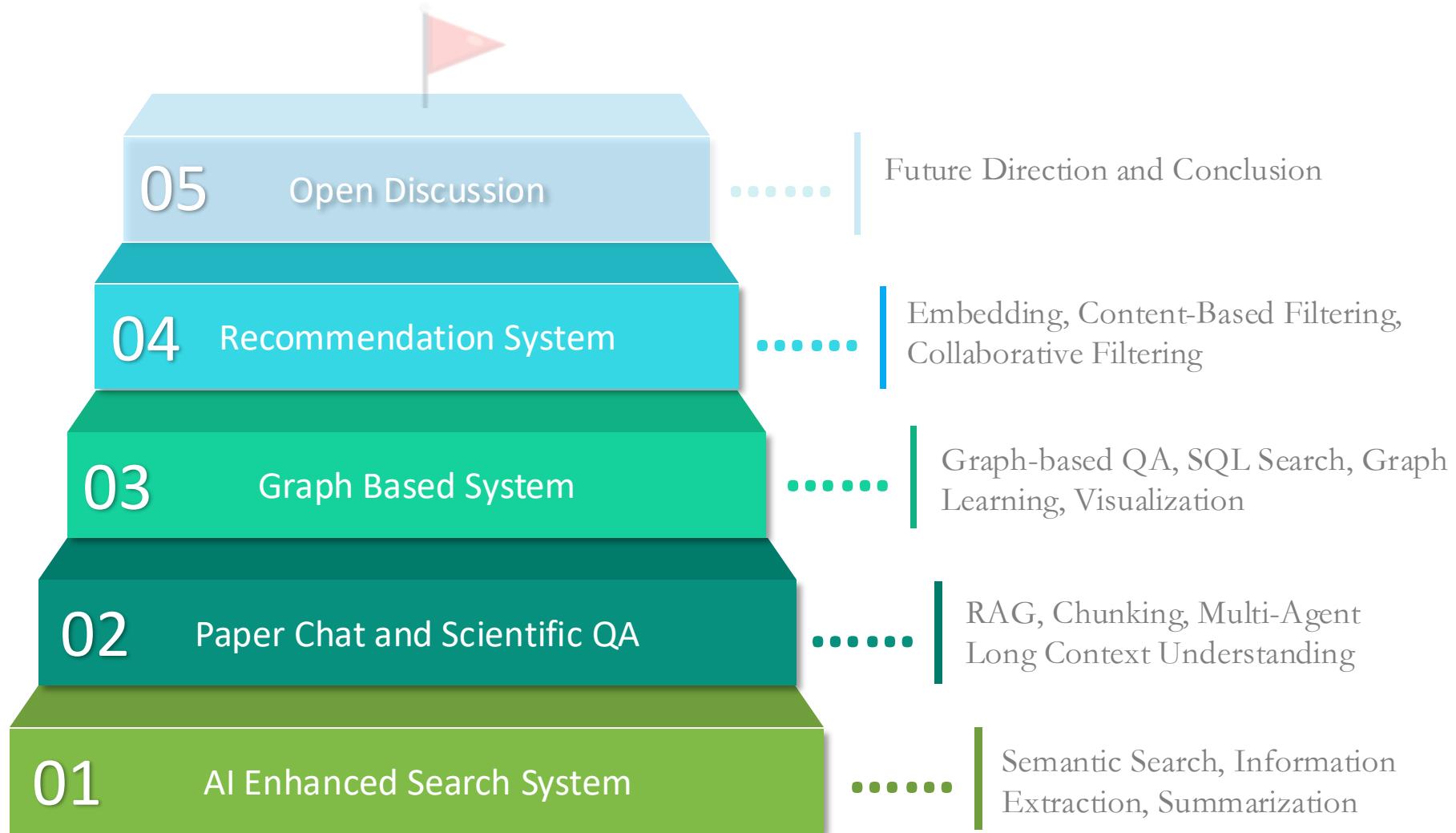
LLMs + Graph?



3 (Summary) - 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

Outline

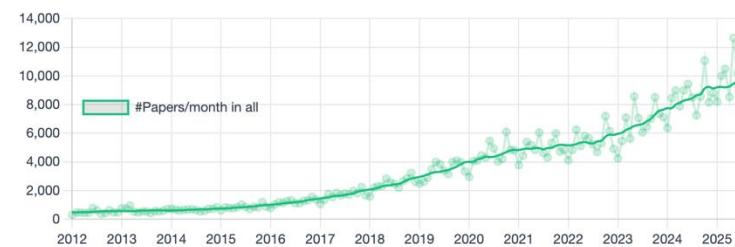


4 - Recommendation System

[Papers](#) [Communities](#) [Organizations](#) [Events](#) [Blog](#) [Pricing](#) Search Sign up Login →

Stay Updated on the Trends, Connect with AI Researchers.

#arXiv papers on AI



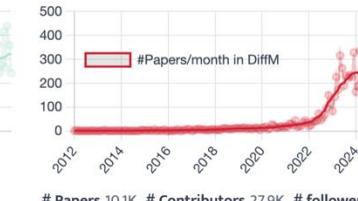
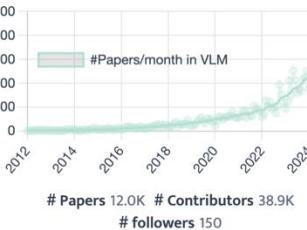
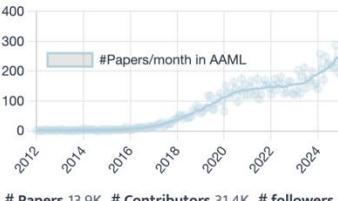
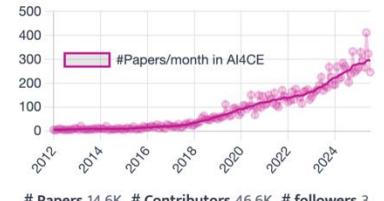
Upcoming ResearchTrend Connect Sessions

Date	Community	Location	Event
------	-----------	----------	-------

No events available

[Feedback](#)

All communities

[AI4CE](#) [AI for Computational...](#)[Follow 3](#)[AAML](#) [Adversarial Attacks in...](#)[Follow 42](#)[VLM](#) [Vision-Language Models](#)[Follow 150](#)[DiffM](#) [Diffusion Models](#)[Follow 116](#)

Papers 14.6K # Contributors 46.6K # followers 3

Papers 13.9K # Contributors 31.4K # followers 42

Papers 12.0K # Contributors 38.9K # followers 150

Papers 10.1K # Contributors 27.9K # followers 116

4 - Recommendation System



Papers Communities Organizations Events Blog Pricing

Search

Sign up Login →

Stay Up-to-Date with Bytez

AI moves fast Stay faster

Bytez tracks every new model & paper, then delivers the few that matter to you — every day

- Spot breakthroughs the day they drop
- See trends before the trend-line
- Cut the noise, keep the edge

Create your feed

#arXiv papers

All communities

AI4CE AI for Co

Devs

Papers 14.6K

Recent Posts

AI Flow: Perspectives, Scenarios, and Approaches 1个月前 · arXiv

MiniMax-M1: Scaling Test-Time Compute Efficiently with Lightning Attention 1个月前 · arXiv

The Lock-in Hypothesis: Stagnation by Algorithm 2个月前 · arXiv

On the Emergence of Linear Analogies in Word Embeddings 2个月前 · arXiv

MMLU-ProX: A Multilingual Benchmark for Advanced 2个月前 · arXiv

Technologies

KV Cache ✓ Supports general queries
k[1] k[2] k[3] k[4]
v[1] v[2] v[3] v[4]

LLM + KV Cache In FY15, the D&A margin for...

Cartridges

Lightweight and general-purpose long context representations via self-study

2个月前 · arXiv

Scaling Test

FLOPs ($\times 10^9$)

Method	SWF-search Verified	TAU-search	MRCR (4 needles)	Classical 4-Point
DeepRank HT-OSCAR	56.0	42.8	37.4	65.0
MobileNet M1	56.0	42.8	37.4	65.0

4 - Recommendation System



Papers Communities Organizations Events Blog Pricing

Search

Sign up Login →

Stay Up-to-Date by bytez

#arXiv papers

notesum.ai

Log in Sign up

All communities

AI4CE AI for Co

Devs

Papers 14.6K

Register for Free

Checking new Arxiv Papers in just 10 minutes

Notesum.ai is the personal newsletter informing you of new research updates. Notesum.ai learns your preferences and curates a short but accurate list of related papers that fit your interests.



4 - Recommendation System

Papers Communities Organizations Events Blog Pricing

Search Sign up Login →

Stay Up-to-Date

#arXiv papers

2012 2013

All communities

AI4CE AI for Co

2012 2014 2015

Devs

Papers 14.6K

Ch

Ar

jus

Notesum. research short but

Register

notesum.ai

1 selected paper

Gonot-Schoupsky Garip 2018 Laughter and humour interventions for well-being in older adults: A systematic review and intervention classification. Complementary Therapies in Medicine Abstract Objectives To assess the potential of laughter and humour interventions to increase well-being in a general population of adults aged 60 plus, and to develop a classification to compare approaches and potential benefits of different intervention types. Design A systematic search of Web of Science, PubMed/MEDLINE, PsycInfo, AMED, and PsychArticle using inclusive terms relating to laughter and humour interventions. A realist synthesis approach was used to examine how interventions to be compared pragmatically. Setting Five studies of laughter interventions, and one humour intervention, using one or more outcome related to well-being, were considered for inclusion after screening 178 primary research papers. The five laughter interventions, representing a sample of 369 participants, were retained. Main outcome measures Well-being related outcome measures reported in each intervention informed efficacy. Joanna Briggs Institute tools appraised designs as relevant and included interventions interventions to be measured on their overall potential to provide an evidence base. Results Well-being related measures demonstrated at least one significant positive effect in all interventions. Confounding factors inherent in the intervention types were observed. Individual participant laughter was not reported. Conclusions Laughter and humour interventions appear to enhance well-being. There is insufficient evidence for the potential of laughter itself to increase well-being. Future research should consider a range of confounding factors and did not measure participant laughter. Interventions that isolate, track, and measure the parameters of individual laughter are recommended to build evidence for these potentially important interventions.

To This Collection To Other Collection

Similar Work These Authors Suggested Authors

All References All Citations

Visualize These Papers Up to 50 papers

Collaborators 40 authors

Gulcan Garip University of Derby 26 publications 213 citations

Published Work Collaborators

Click an author to explore their profile

Gulcan Garip 26 publications 213 citations

Freda N. Gonot-Schoupsky University of Derby 7 publications 70 citations

Lucy Yardley University of Southampton 591 publications 26267 citations

David Sheffield University of Derby 196 publications 4706 citations

Robert Bridger 3 publications 3 citations

Ainsley Cross University of Derby 19 publications 101 citations

Heather Buchanan University of Nottingham 73 publications 1401 citations

Robert S. Bridger University of Southampton 54 publications 1040 citations

Miles Richardson University of Derby 67 publications 993 citations

Aimee Aubeeluck University of Nottingham 2 publications 2 citations

Connections between 40 authors

Filter these items

Gulcan Garip

Visualize Collaborations Up to 50 people

Zoom Out Fit All Zoom In

Log in Sign up

4 - Recommendation System



Papers Communities Organizations Events Blog Pricing

Search

Sign up Login →

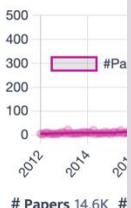
Stay Up to Date

#arXiv papers



All communities

AI4CE AI for Co



Papers 14.6K

bytez

notesum.ai

Change
And
just

Notesum.
research
short but

Register

reported a technological evolution, now focusing on post with large-scale intelligent (AI) models that are reshaping industries and nurturing human-machine collaboration. However, the evolution of ubiquitous intelligence faces constraints due to the inherent limitations of computation in large-scale systems. In this paper, we propose a framework for the design of AI-driven systems. This framework is based on a multi-layered architecture that integrates machine learning (ML) and AI technologies, with a particular emphasis on the following three key points. First, *decentralized framework* serves as the foundation, which integrates end devices, edge servers, and cloud clusters to optimize

resource utilization and reduce latency. Second, *intelligent optimization* is achieved by combining ML and AI techniques to handle complex optimization problems.

Third, *interoperability* is ensured through standard interfaces and protocols that facilitate communication between different components of the system.

Overall, this framework aims to address the challenges of AI-driven systems and enable their widespread adoption in various industries.

Log in

Sign up

1 selected paper

Gulcan Garip University of Derby

Collaborators 40 authors

Connections between 40 authors

Trending Bookmarks Collections Interactions Semantic Search Scholar Maps Conferences ACL 2025 VIENNA

25-07-2025 – 25-07-2025

28/1311

New conference proceedings

This digest contains papers from the new proceedings of [Annual Meeting of the Association for Computational Linguistics \(ACL\)](#), [Findings of the Association for Computational Linguistics \(Findings of ACL\)](#).

CiteLab: Developing and Diagnosing LLM Citation Generation Workflows via the Human-LLM Interaction

Jiajun Shen, Tong Zhou, Yubo Chen, Kang Liu, Jun Zhao

Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)

Language

The emerging paradigm of enabling Large Language Models (LLMs) to generate citations in Question-Answering (QA) tasks is lacking in a unified framework to standardize and fairly compare different citation generation methods, leading to difficulties in reproduction and innovation. Therefore, we introduce Citeflow, an open-source and modular framework fostering reproduction and the implementation of new designs. Citeflow is highly extensible, allowing users to utilize four main modules and 14 components to construct a pipeline, evaluate an existing method, and understand the attributing LLM-generated contents. The framework is also paired with a visual interface, Citefix, facilitating case study and modification of existing citation generation methods. Users can use this interface to conduct LLM-powered case studies according to different scenarios. Citeflow and Citefix are highly integrated into the toolkit CiteLab, and we use an authentic process of multiple rounds of improvement through the Human-LLM interaction interface to demonstrate the efficiency of our toolkit on implementing and modifying citation generation pipelines. CiteLab is released at <https://github.com/SJ1017/CiteLab>

FineCite: A Novel Approach For Fine-Grained Citation Context Analysis

Lasse M. Jantsch, Dong-Jae Koh, Seonghwan Yoon, Jisu Lee, Anne Lauscher, Young-Kyoon Suh

Findings of the Association for Computational Linguistics: ACL 2025

Language

Citation context analysis (CCA) is a field of research studying the role and purpose of citation in scientific discourse. While most of the efforts in CCA have been focused on elaborate characterization schemata to assign function or intent labels to individual citations, the citation context as the basis for such a classification has received rather limited attention. This relative neglect, however, has led to the prevalence of vague definitions and restrictive assumptions, limiting the citation context in its expressiveness. It is a common practice, for example, to restrict the context to the citing sentence. While this simple context conceptualization might be sufficient to assign intent or function classes, it fails to cover the rich information of scientific discourse. To address this concern, we analyze the context conceptualizations of previous works and, to our knowledge, construct the first comprehensive context definition based on the semantic properties of the citing text. To evaluate this definition, we construct and publish the FineCite corpus containing 1,056 manually annotated citation contexts. Our experiments on established CCA benchmarks demonstrate the effectiveness of our fine-grained context definition, showing improvements of up to 25% compared to state-of-the-art approaches. We make our code and data publicly available at <https://github.com/lab-paper-code/FineCite>.

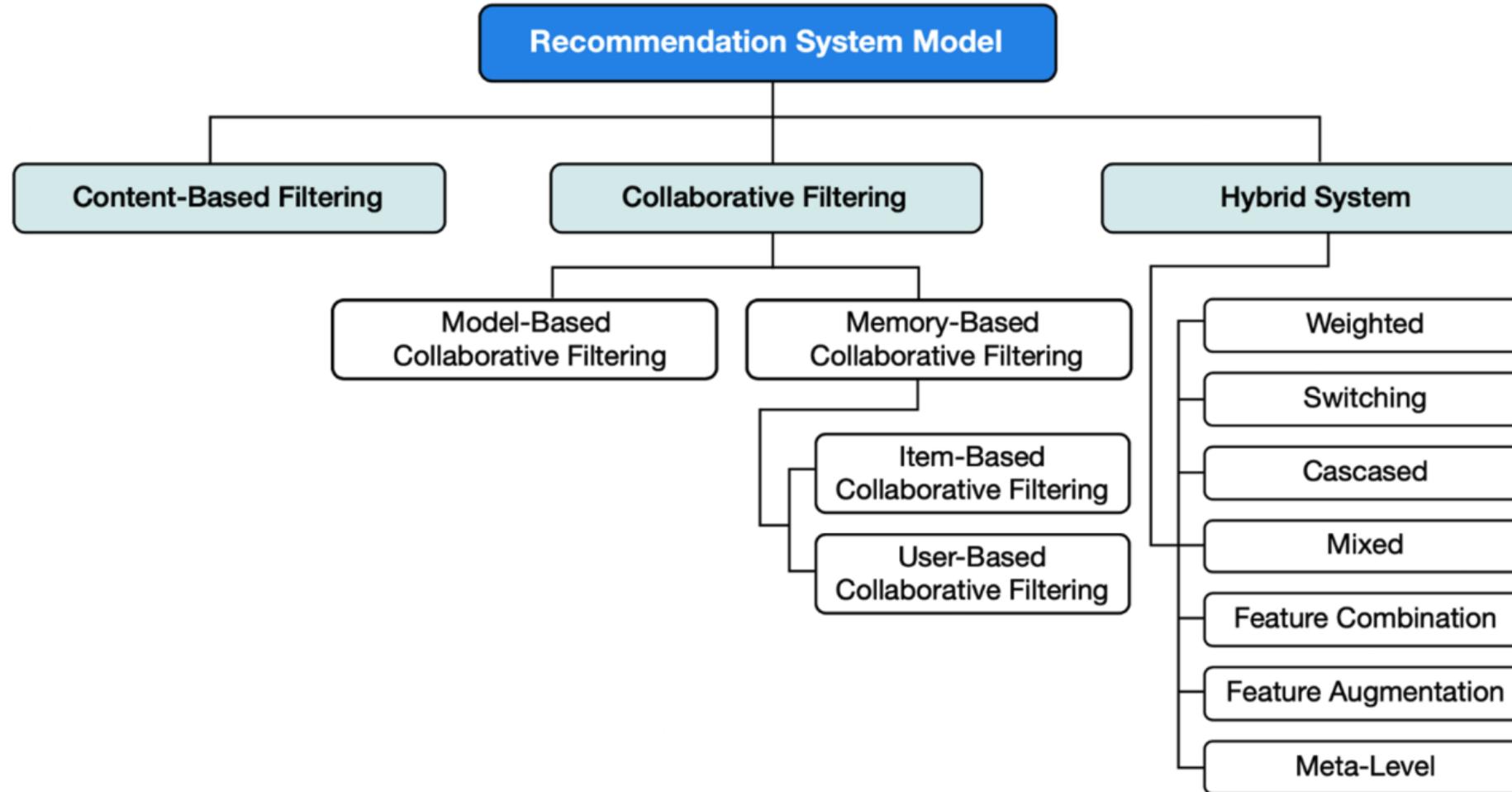
A^3: Automatic Alignment Framework for Attributed Text Generation

Yue Wang, Haoke Zhang, Juntao Li, Jinxiang Chang, Min Zhang

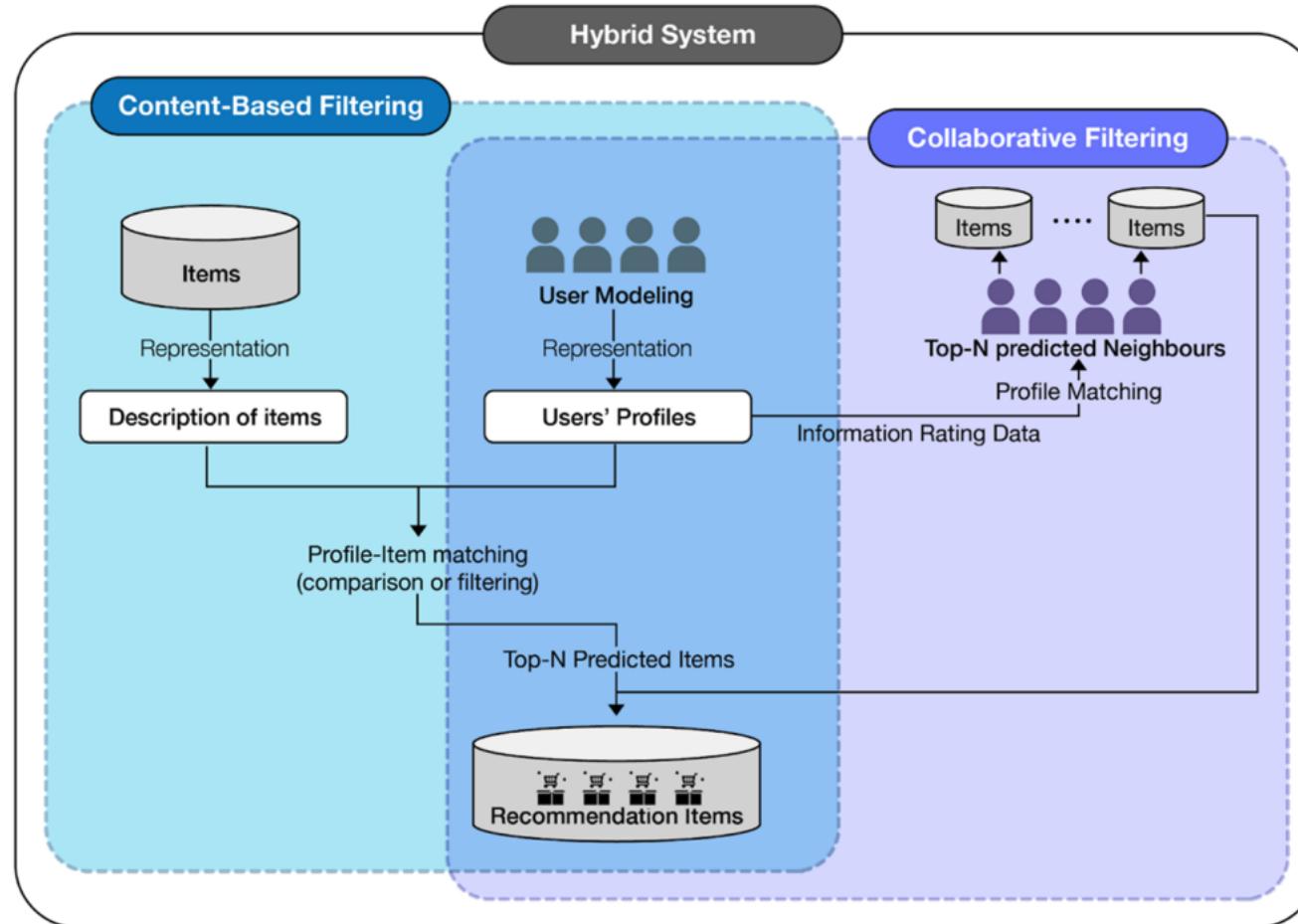
Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)

Language

Recommendation System

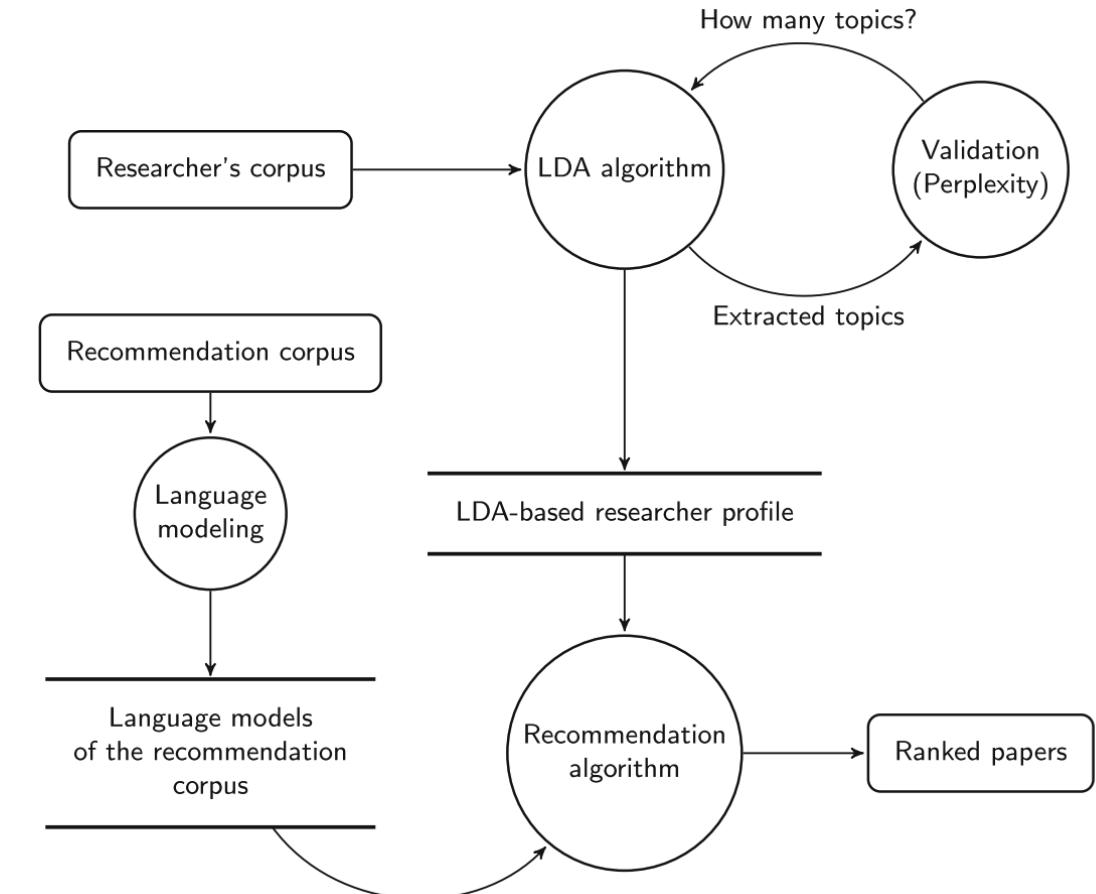


Pipeline of recommendation systems



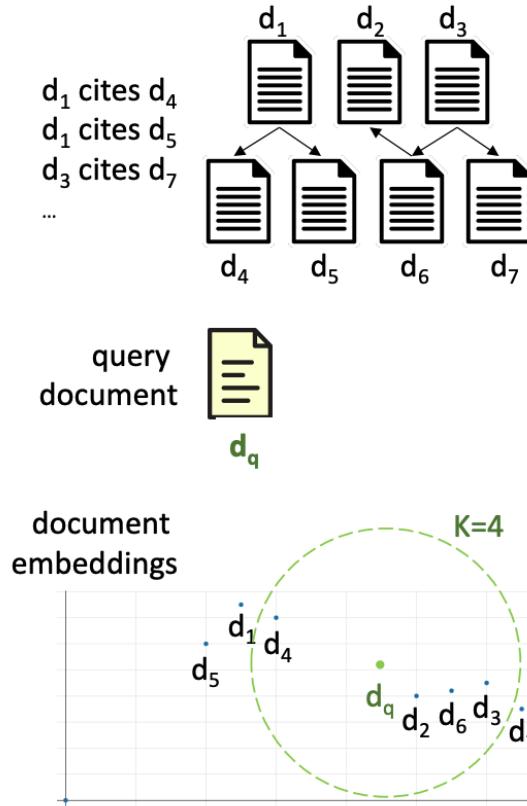
Content-based Methods

- 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



NNSelect+NNRank

Phase 1: candidate selection



Phase 2: reranking

nearest neighbors of d_q :

	→ NNRank	→ 0.7
d_q	→ NNRank	→ 0.8
d_q	→ NNRank	→ 0.5
d_q	→ NNRank	→ 0.3

cited in nearest neighbors:

d_q	→ NNRank	→ 0.9
-------	----------	-------

reranked list

d_7

d_6

d_2

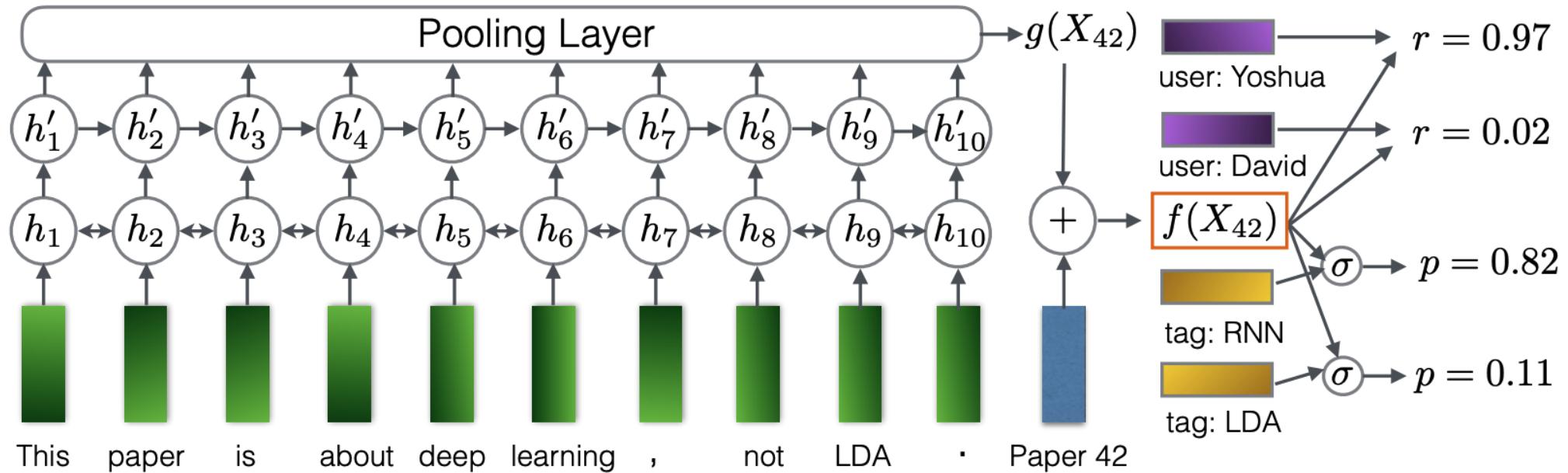
d_3

d_4

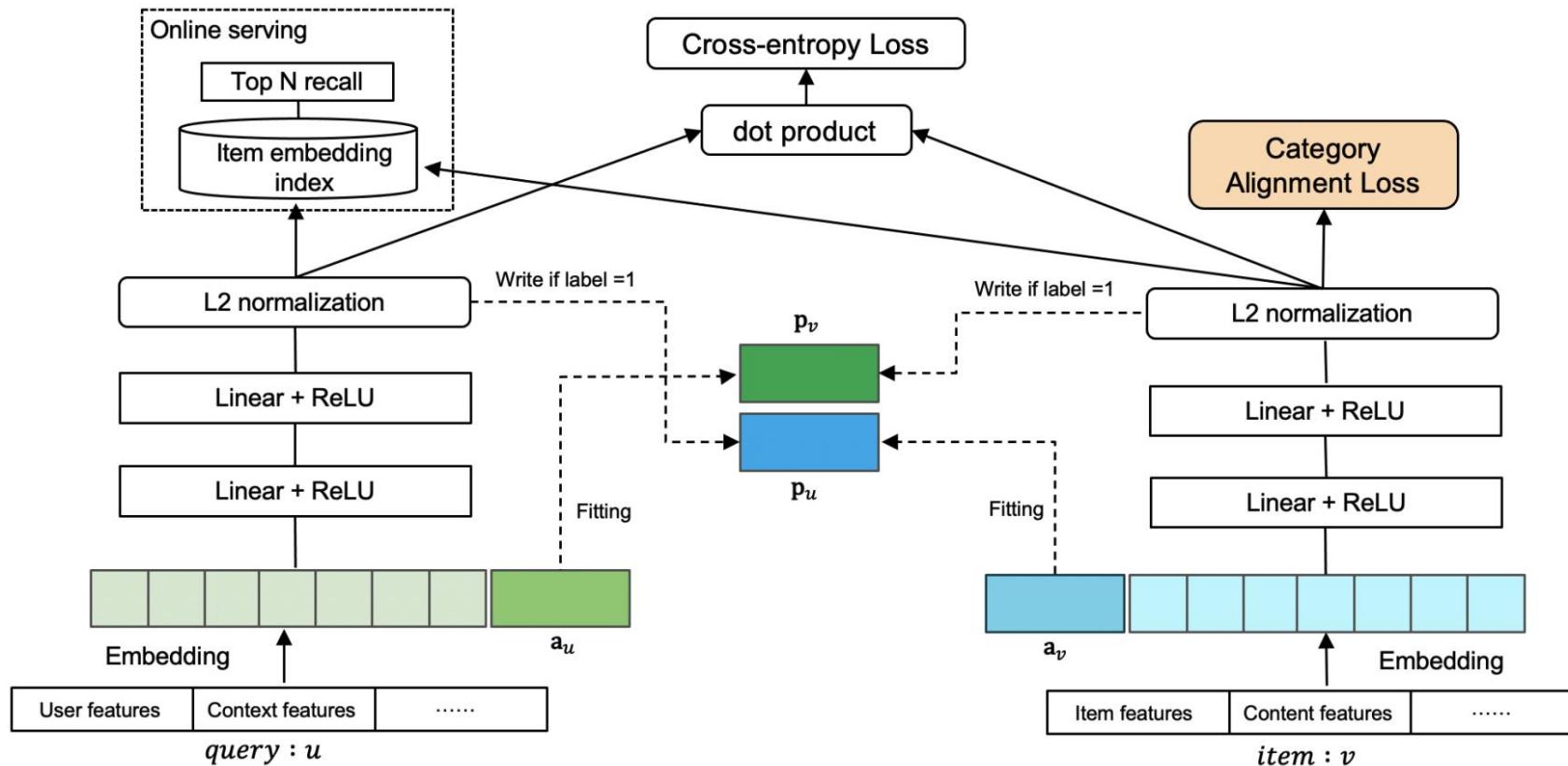
top N=3 recommendations

Collaborative Filtering

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



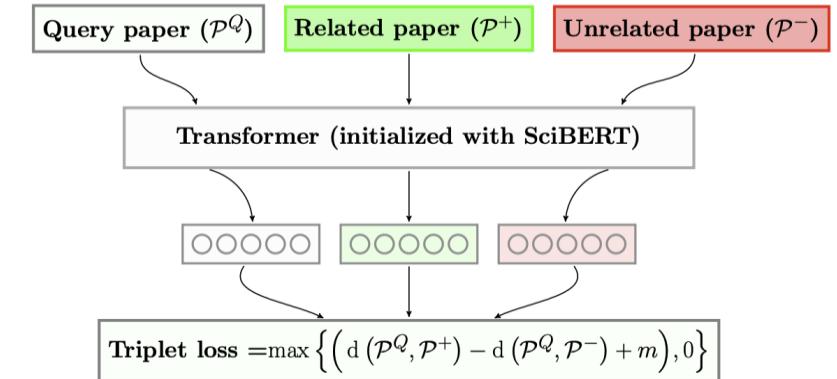
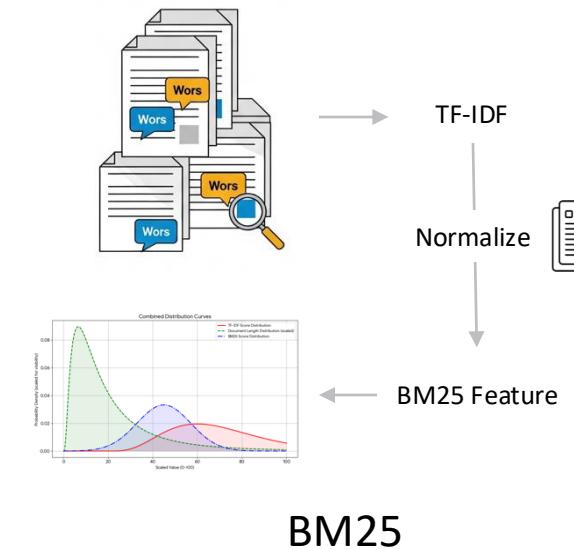
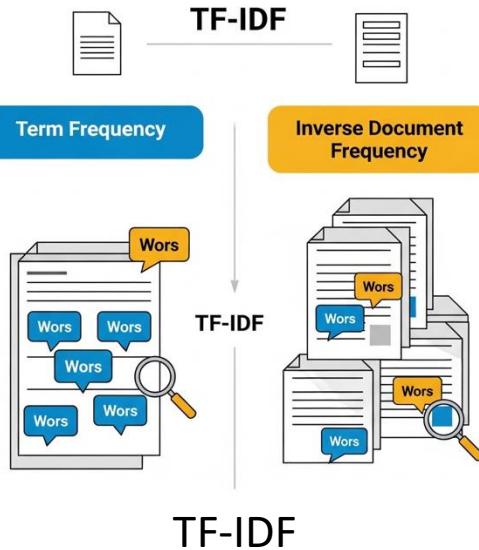
Hybrid Systems



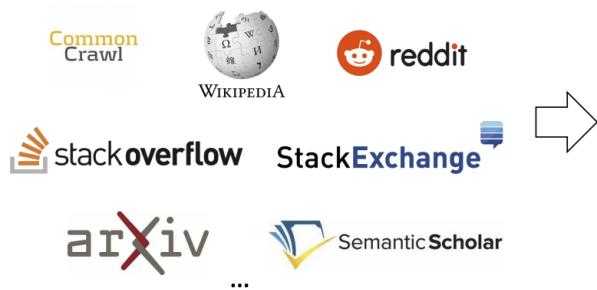
Comparison of Recommendation Systems

Model Type	Pros	Cons
Content-Based Filtering	<ul style="list-style-type: none">• Works well for new users• Personalized to individual preferences• Doesn't require large user base• Easy to interpret recommendations	<ul style="list-style-type: none">• Needs detailed item features• Limited discovery of new interests
Collaborative Filtering	<ul style="list-style-type: none">• Doesn't need item metadata• Leverages collective user behavior	<ul style="list-style-type: none">• Struggles with sparse data• Scalability issues in memory-based
Hybrid Systems	<ul style="list-style-type: none">• Combines best of all models• Mitigates individual weaknesses• Often improves accuracy	<ul style="list-style-type: none">• Increased complexity• Requires more computation

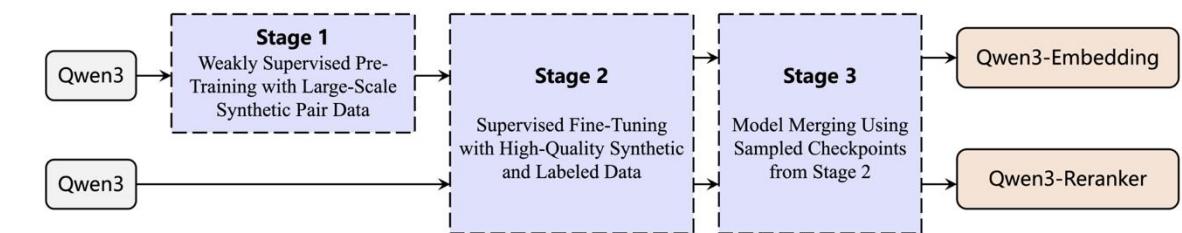
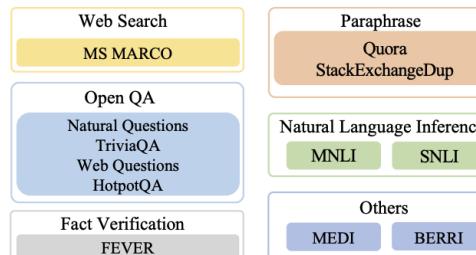
Document Embeddings



Unsupervised Contrastive Pre-training on Massive Text Pairs mined from the Web

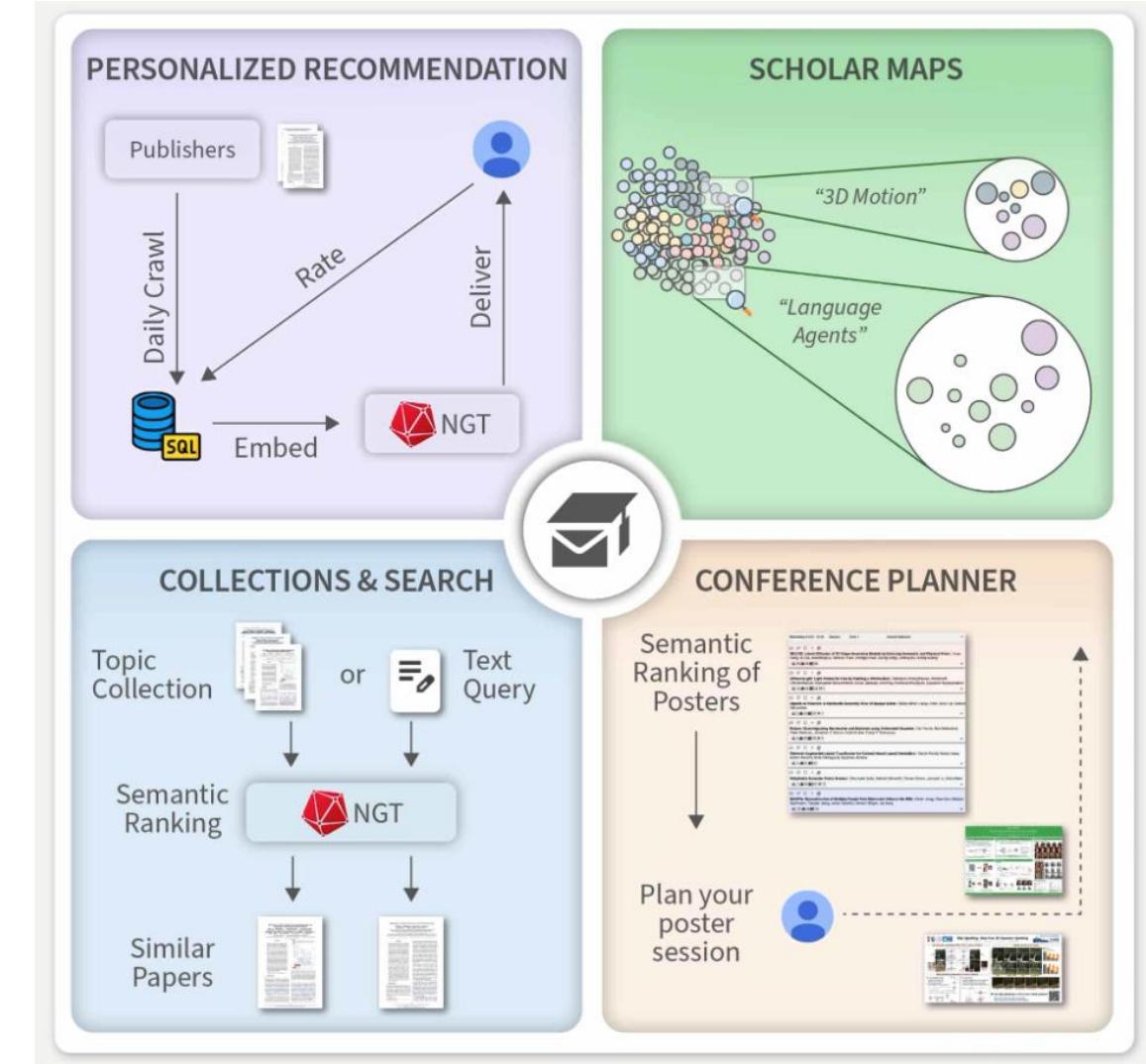


Supervised Contrastive Fine-tuning on Annotated Text Triples from Multiple Tasks

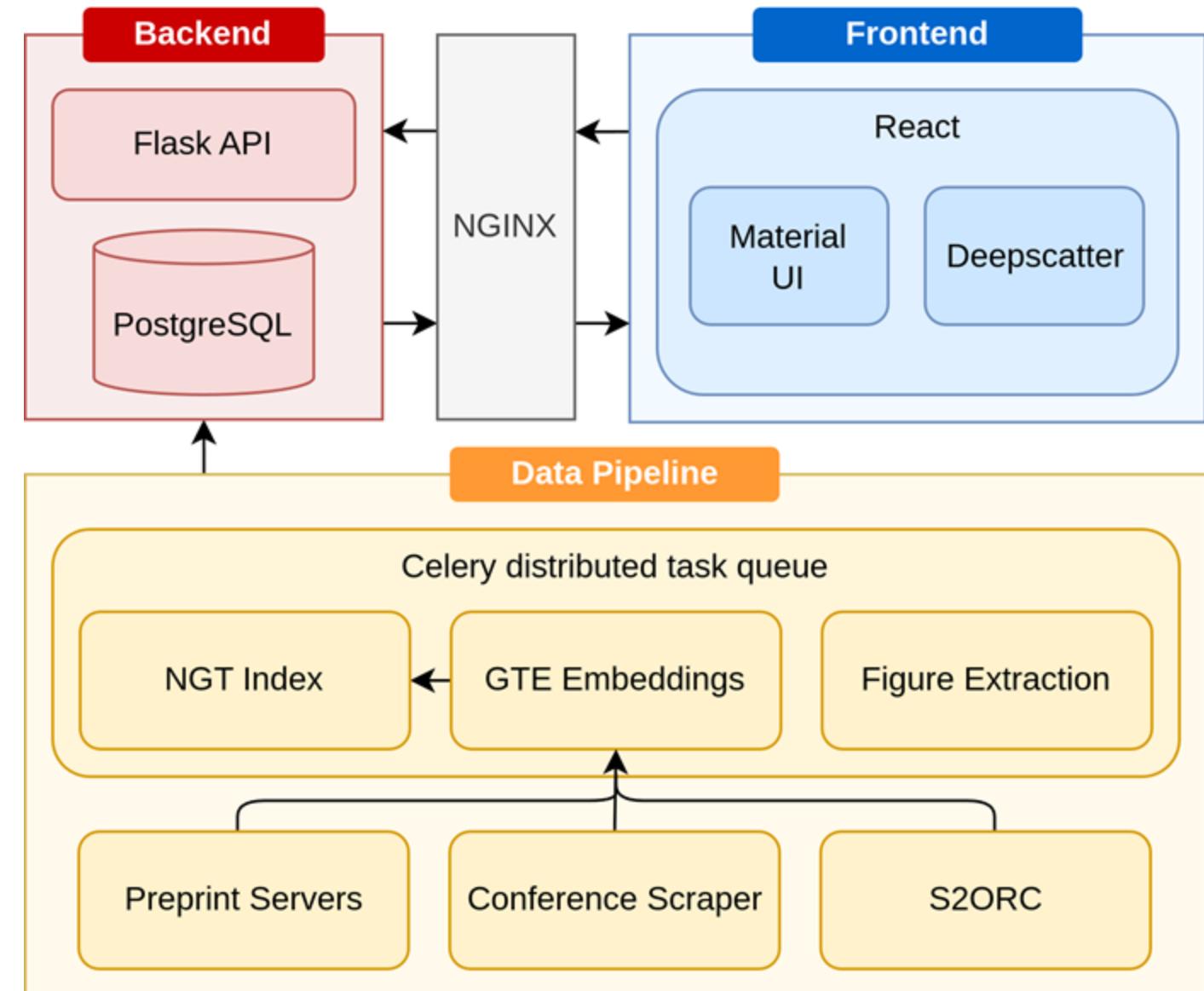


Scholar Inbox – 2025 Demo

- Personalized Recommendations
- Scholar Maps
- Collections
- Conference Planner



Software Flow



Recommender system

Scholar Inbox

Search

Digest

56/1327 Papers

20-05-2025 - 20-05-2025

Trending

Interactions

Scholar Maps

Conferences

Labels

- Bookmarks
- Deep Learning
- Neural Radiance F...
- Generative Models
- Large Language...
- Optimization
- Self-Driving Cars
- More

93 Relevance

GIRAFFE: Representing Scenes as Compositional Generative Neural Feature Fields

Michael Niemeyer, Andreas Geiger

Computer Vision and Pattern Recognition (CVPR), 2021

19 23 1013 Computer Vision and Graphics Generative Models

Deep generative models allow for photorealistic image synthesis at high resolutions. But for many applications, this is not enough: content creation also needs to be controllable. While several recent works investigate how to disentangle underlying factors of variation in the data, most of them operate in 2D and hence ignore that our world is three-dimensional. Further, only few works consider the compositional nature of scenes. Our key hypothesis is that incorporating a compositional 3D scene representation into the generative model leads to more controllable image synthesis. Representing scenes as compositional generative neural feature fields allows us to disentangle one or multiple objects from the background as well as individual objects' shapes and appearances while learning from unstructured and unposed image collections without any additional supervision. Combining this scene representation with a neural rendering pipeline yields a fast and realistic image synthesis model. As evidenced by our experiments, our model is able to disentangle individual objects and allows for translating and rotating them in the scene as well as changing the camera pose.

Figures & Tables

AI Summary

Similar Papers

Bookmark

Labels

ArXiv

Citation

Project page

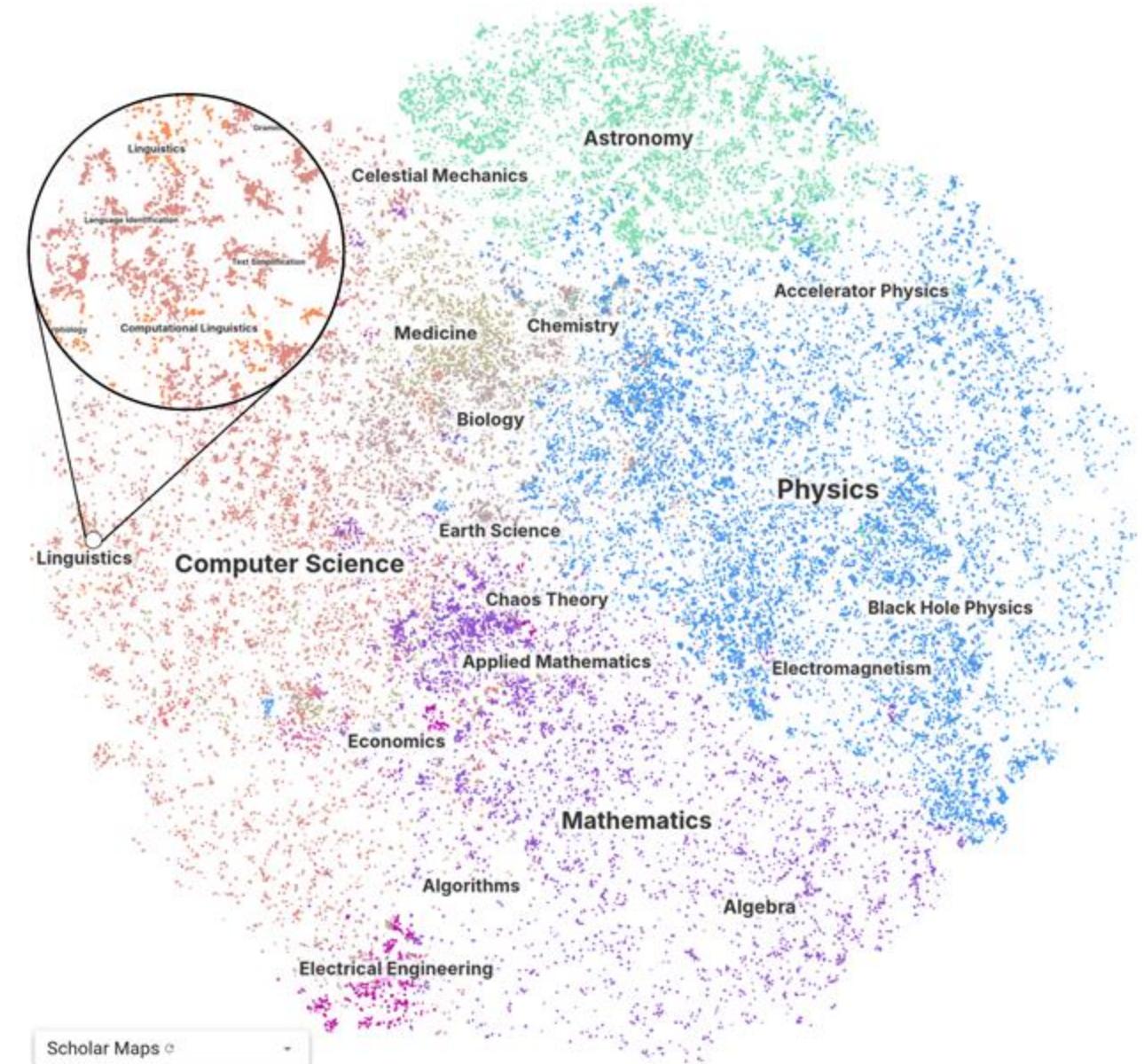
Share

GIRAFFE: Representing Scenes as Compositional Generative Neural Feature Fields

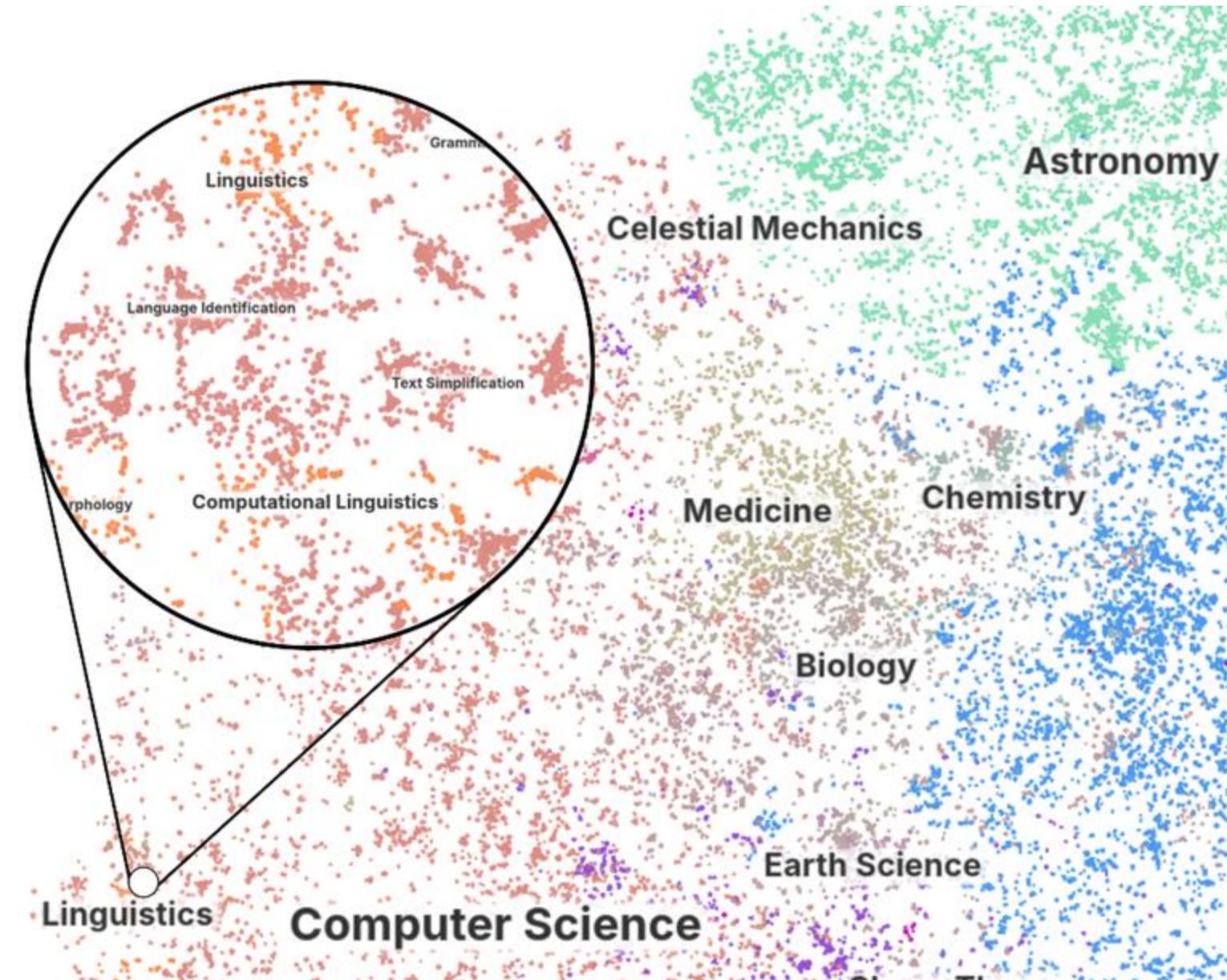
Michael Niemeyer, Andreas Geiger

Bookmark

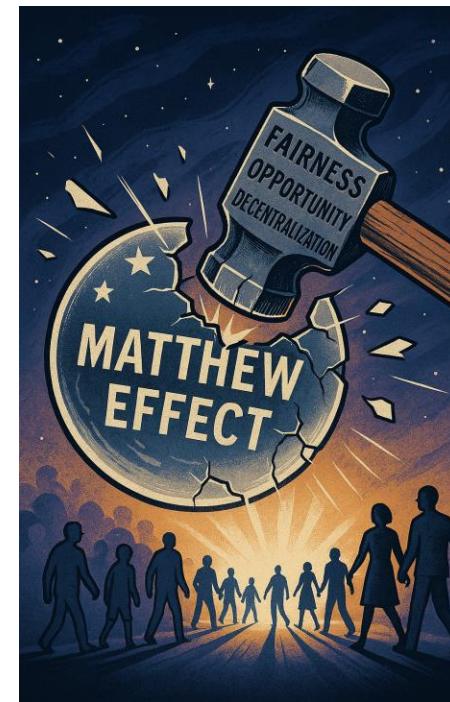
Scholar Maps



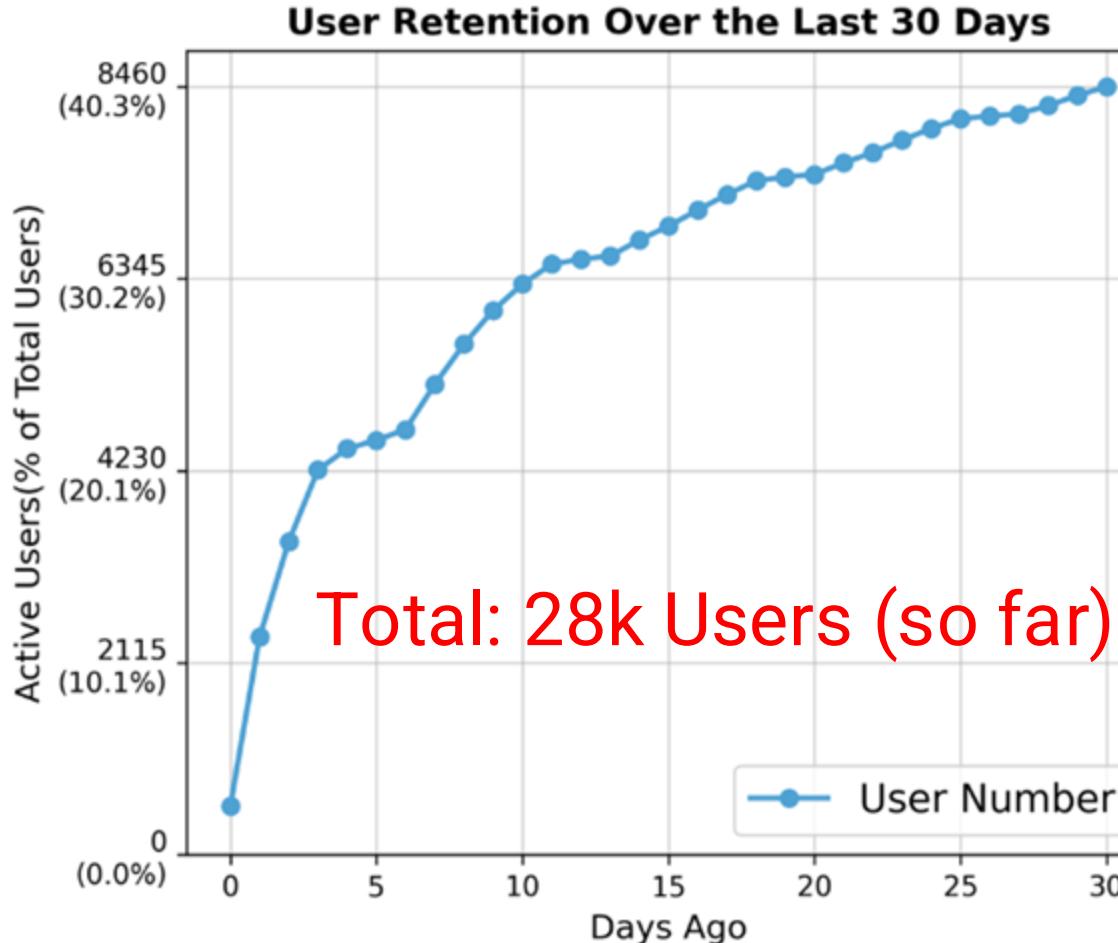
Scholar Maps



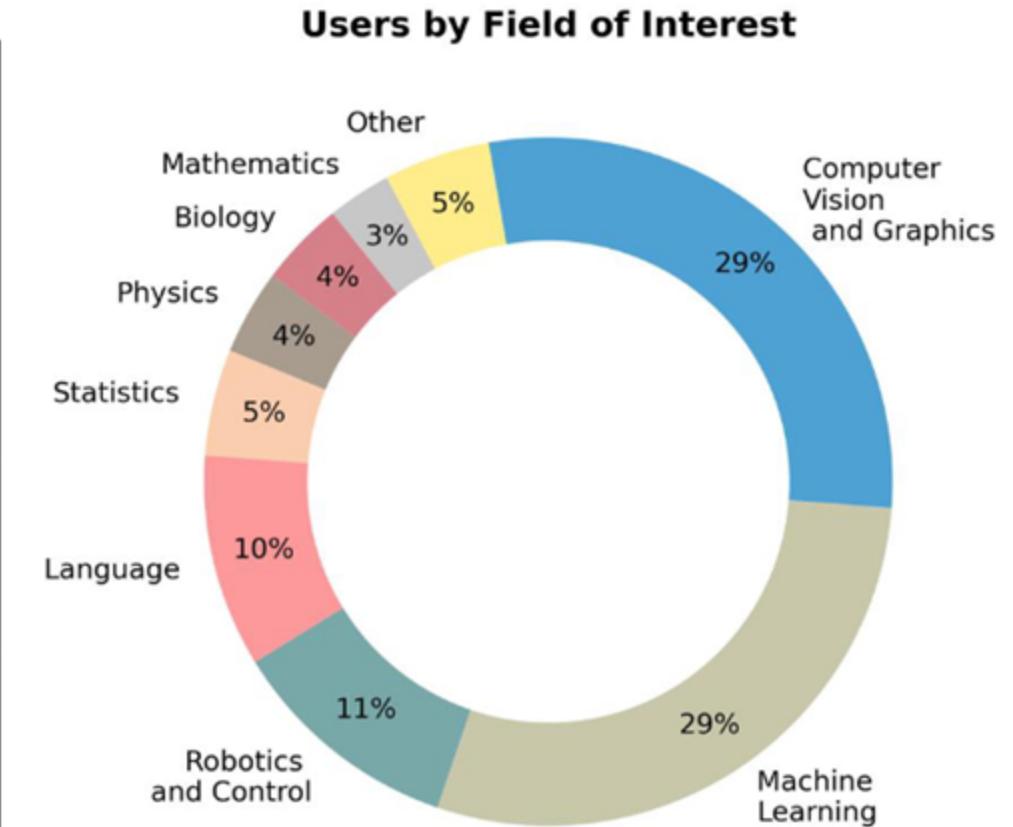
Why we create Scholar Inbox?



9.000 users come back again and again



(a) User Retention



(b) User Domain Distribution

Welcome to check our paper

- First time serving NLP conference.
- Hall 5X, 28th July, 11:00-12:30, ACL 2025.



Scholar Inbox supports **ACL25**

With your personalised relevance:

- Explore the ACL25 timetable
- Prioritise within poster sessions
- Plan your visit

Make this year your most efficient
ACL visit ever!



Join our 30k users



www.scholar-inbox.com

✉ Scholar Inbox: Personalized Paper Recommendations for Scientists

Markus Flicke Glenn Angrabeit Madhav Iyengar Vitalii Protsenko
Illia Shakun Jovan Cicvaric Bora Kargi Haoyu He Lukas Schuler
Lewin Scholz Kavyanjali Agnihotri Yong Cao Andreas Geiger

University of Tübingen, Tübingen AI Center

www.scholar-inbox.com



4 (Summary) - Recommendation System

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

Overview of popular literature search

Other AI-enhanced Literature Search

	Platform	AI Features												Cost	Data Source	
		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
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	

Conclusion

- AI is transforming literature search.

Conclusion

- AI is transforming literature search.
- AI tools boost discovery but still require oversight.

Conclusion

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

Conclusion

- 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

Thank you! Any questions?

