THEORY AND PRACTICE OF DIGITAL LIBRARIES

BIRDS 2025

Beyond Retrieval: A Vision of Intelligent Digital Libraries in the Large Language Model Era

DR. JIAN WU

ASSOCIATE PROFESSOR OF COMPUTER SCIENCE

OLD DOMINION UNIVERSITY, VIRGINIA, UNITED STATES

OLD DOMINION UNIVERSITY®

Self-Introduction

2004: B.S. in Physics and Astronomy

2011: Ph.D. in Astronomy and Astrophysics

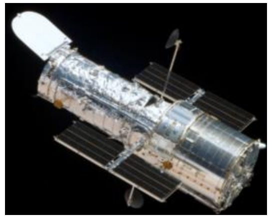
Big Data

2011-2017: Postdoctoral fellow, Information Sciences and Technology

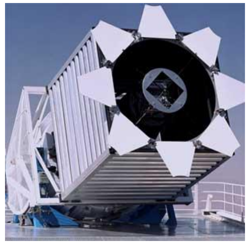
2017-2018: Assistant Teaching Professor, Information Sciences and Technology

2018-2025: Assistant Professor (tenure track), Computer Science

2025-: Associate Professor (effective July 25), Computer Science



Hubble Space Telescope



Sloan Digital Sky Survey

CiteSeer^X

Scholarly Big Data + AI

Conference Proceedings | ETDs | KGs | Tables

Journals | Technical Drawings | Keyphrases | Figures

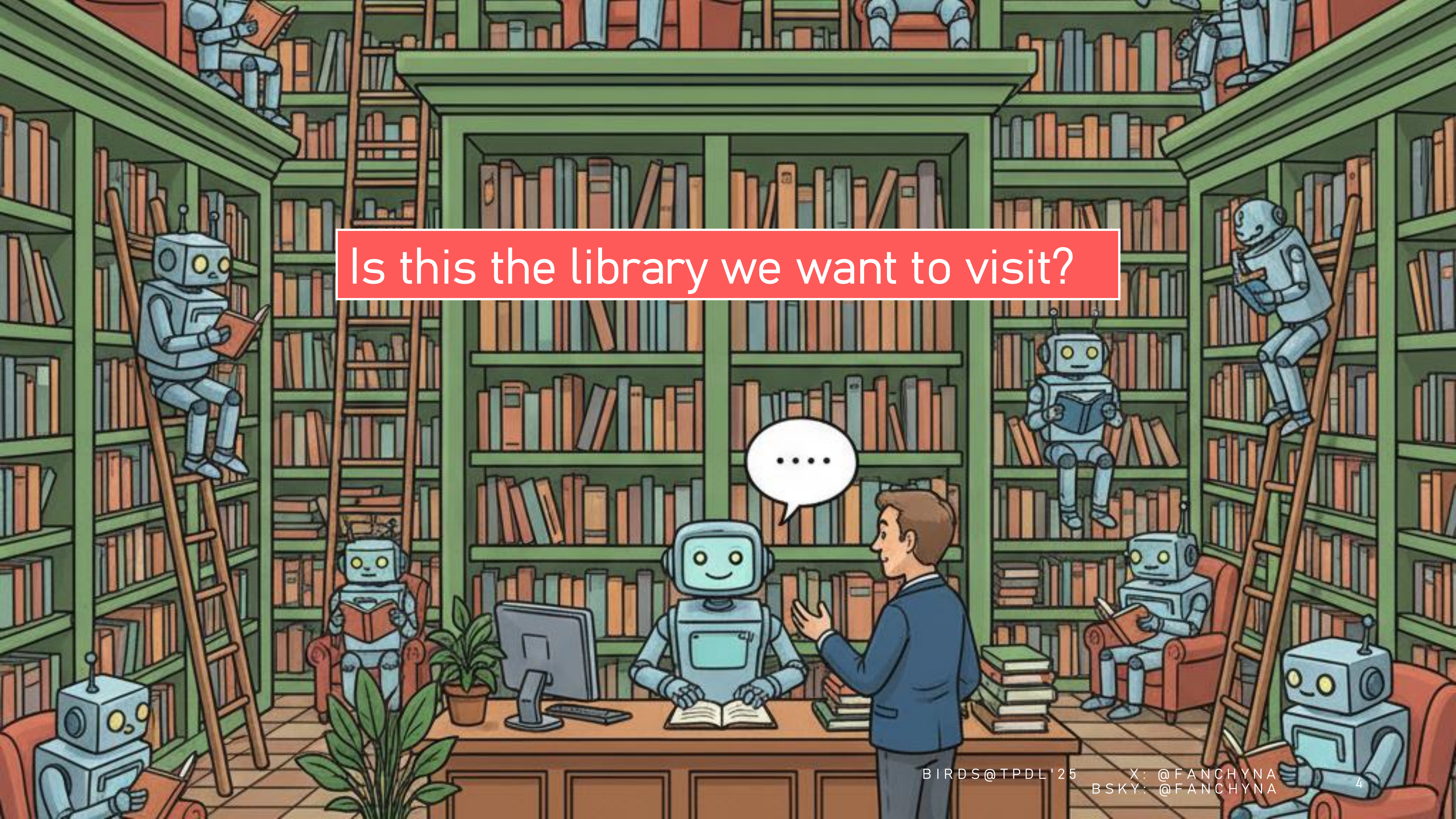
Acknowledgments

CiteSeerX

- Dr. C. Lee Giles (Pennsylvania State University)
  - PI of CiteSeerX
  - Eminent David Reese Professor of Information Sciences and Technology
  - Professor of Supply Chain and Information Systems
  - Director of the Intelligent Systems Research Laboratory

NDLTD

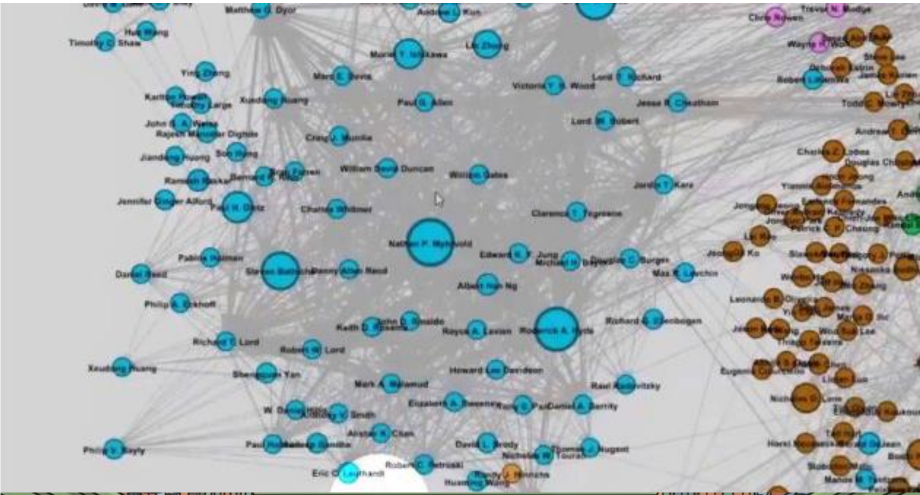
- Dr. Edward A. Fox (Virginia Tech)
  - Director of NDLTD
  - Eminent Professor of Computer Science
  - Director, Digital Library Research Laboratory

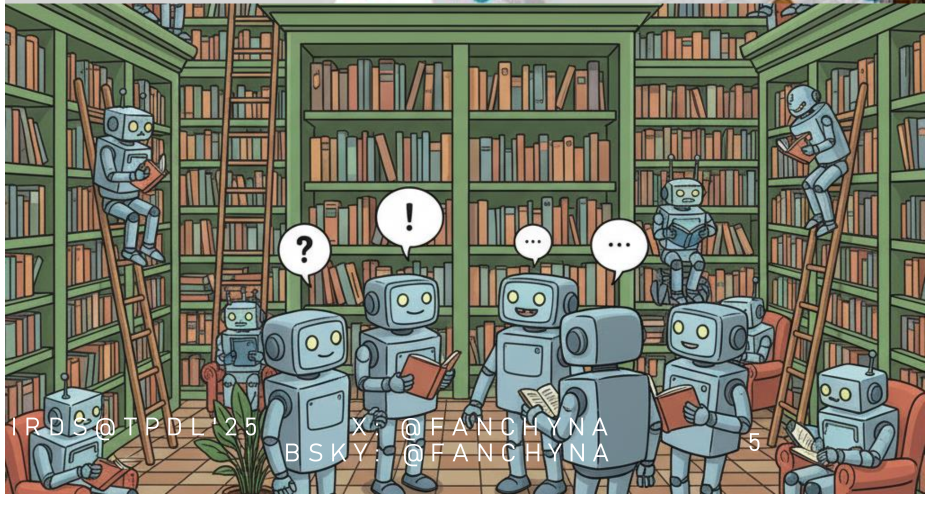


Is this the library we want to visit?

Answer the question from the retrospective point of view

- Digital Repositories
- Digital Library Search Engines
- Intelligent Digital Libraries (IDLs)

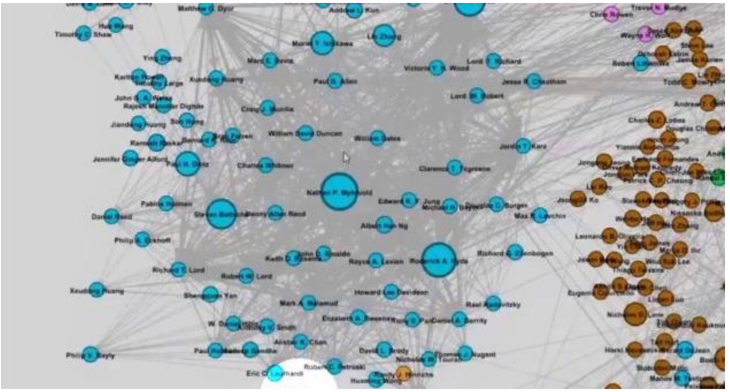




Digital Repositories

- A simple catalog of documents with basic metadata records with or without the full text and a simple boolean query interface
- Can be implemented using a relational database with a single lookup interface
  - University ETD repositories
  - arXiv and sister repos (bioRxiv, medRxiv, PsyArXiv)
  - etc.

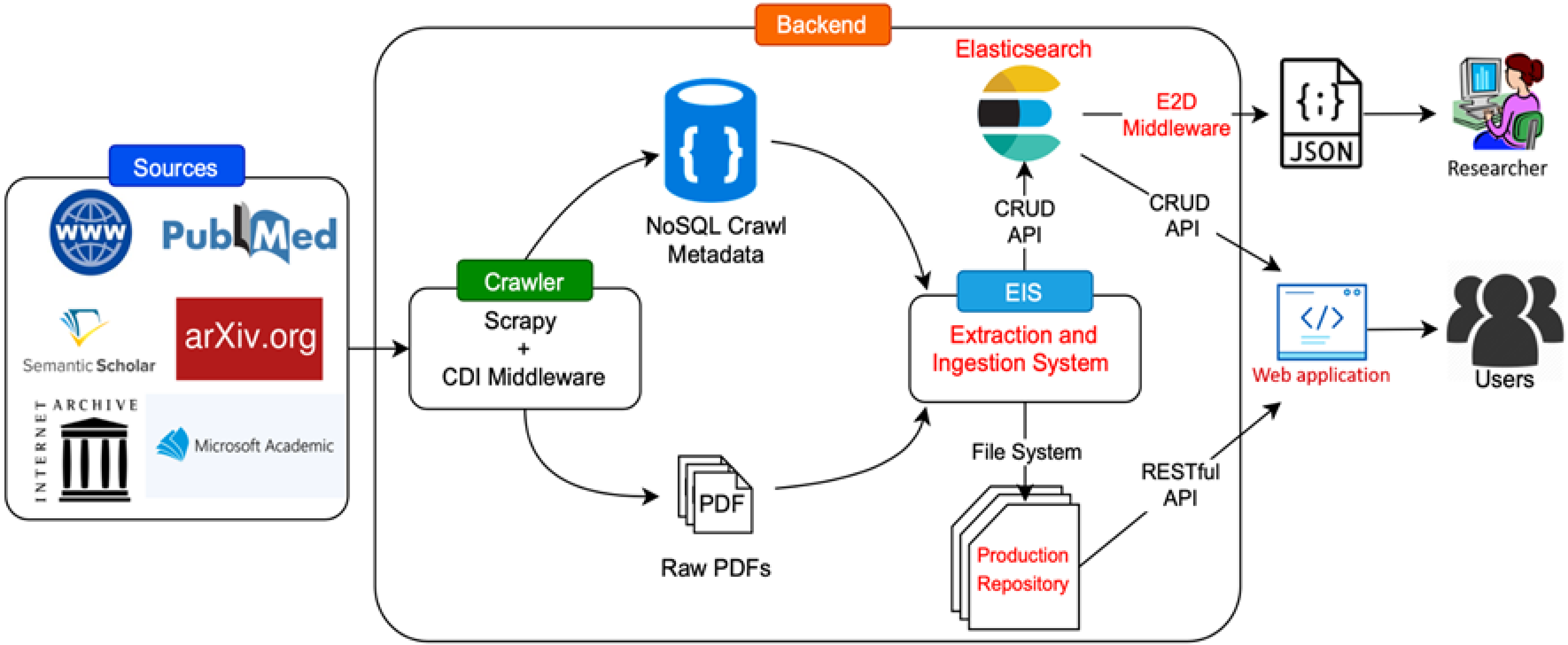
Digital Library Search Engines



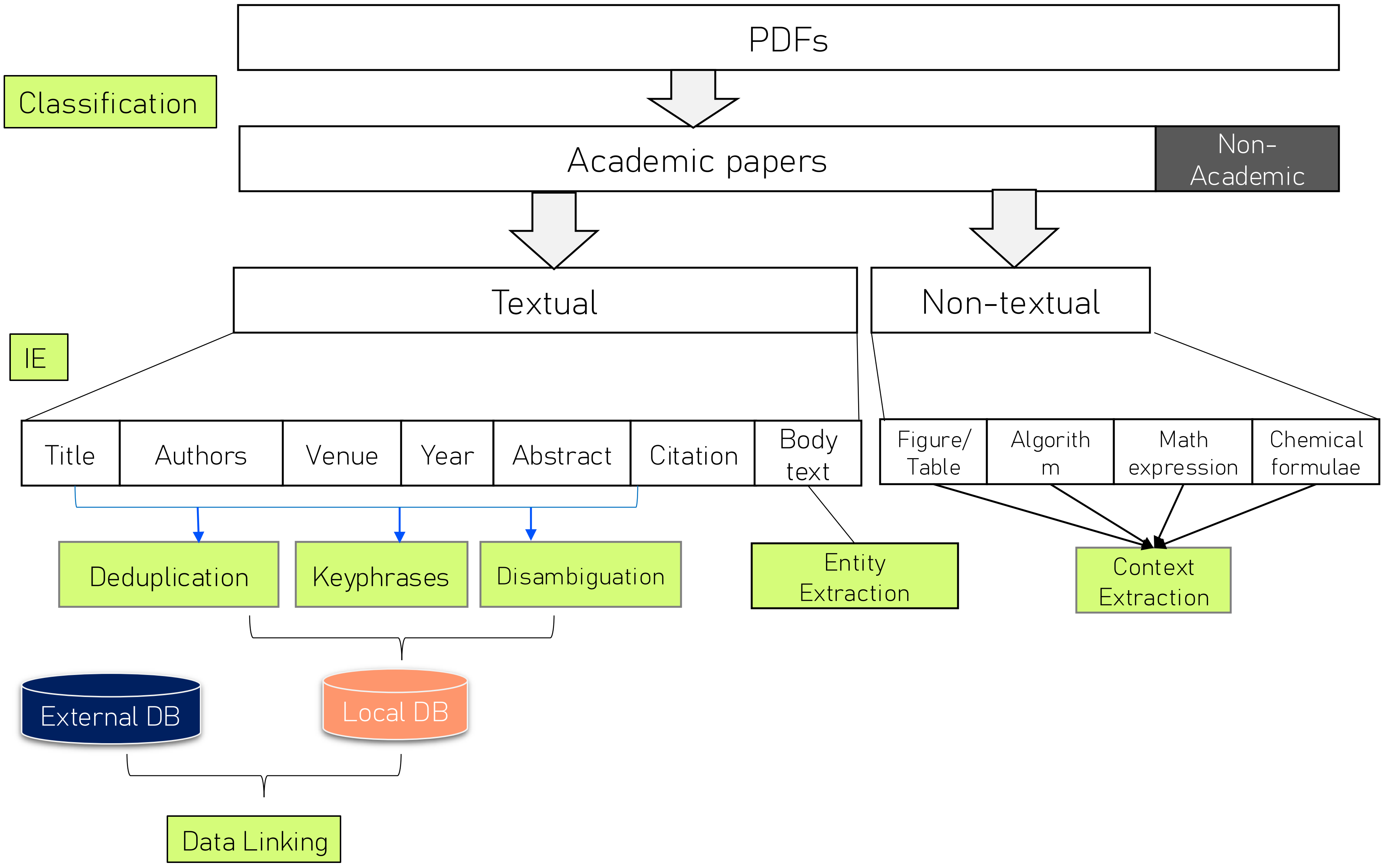
- Documents are organized in a connected manner, by inverted index, citation networks, co-author networks, or other structures so that they are more findable, navigable, and usually provide meta-level knowledge.
  - CiteSeerX
  - Google Scholar
  - Semantic Scholar
  - etc.
- Require
  - Backend (crawling, extraction and ingestion, index, repository)
  - Frontend (search, browsing, download)

The Architecture of CiteSeerX (2022-present)

- A **focused web crawler** for open access academic PDFs
- Elasticsearch for a search engine and a data storage
- Integrated and **parallelized** extraction and ingestion
- Redundancy for high availability and resilience



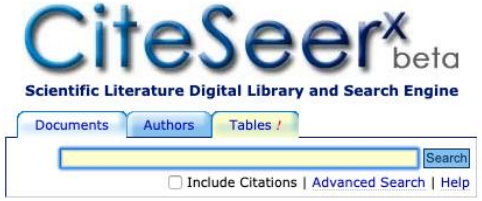
AI in CiteSeerX



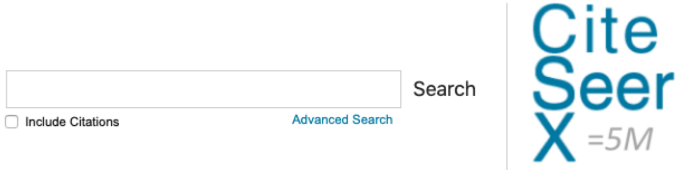
Retrospect and Status of CiteSeerX

2003

2006



2011



2015

- 15+ million full text English documents and metadata.
- 1 billion hits and 180 million downloads annually.
- Googling "CiteSeerX OR CiteSeer" returns 6 million results (Sep 6, 2025).

Semanti Scholar

- Basic functionalities
  - Search
  - Browse
  - Download

Semantic **Scholar**

- AI-powered functionalities are built based on documents in the repository
  - TLDR summarization
  - Citation intent and influence classifications
  - Field of study classification
  - Paper recommendation
  - Metrics (most influential citations, etc.)

Recently Emerged Digital Library Search Engines

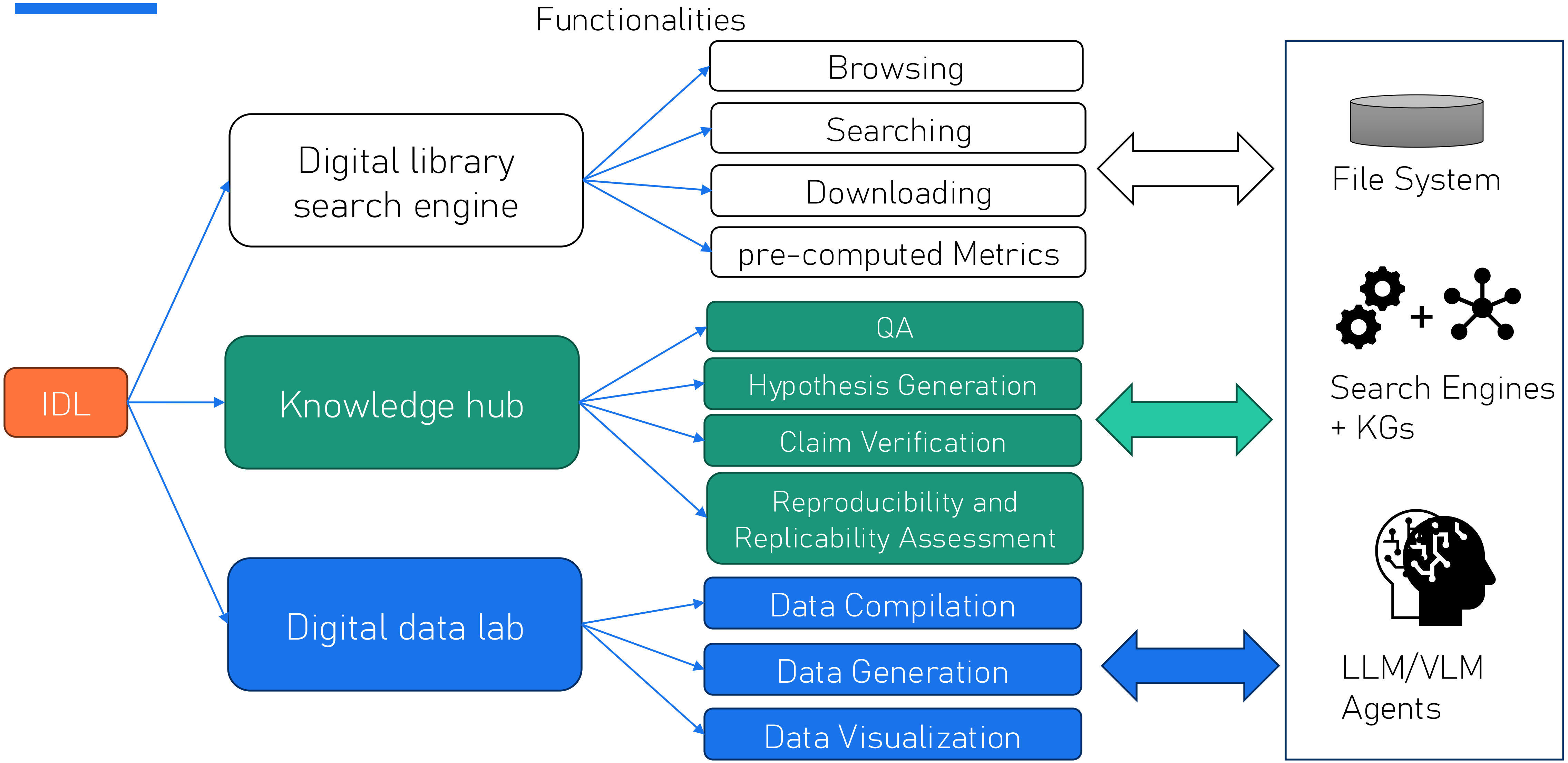
- AllSci: hypothesis-centric, AI-powered
  - Atomized more than 12 million scientific hypotheses into a knowledge graph
  - Using AI-guided tools to help researchers formulate better hypotheses
- Scite: using citation context for QA and Table search
- Consensus: search engine + QA
- ResearchRabbit
  - Allowing users to visualize networks

https://guides.pnw.edu/AISearch

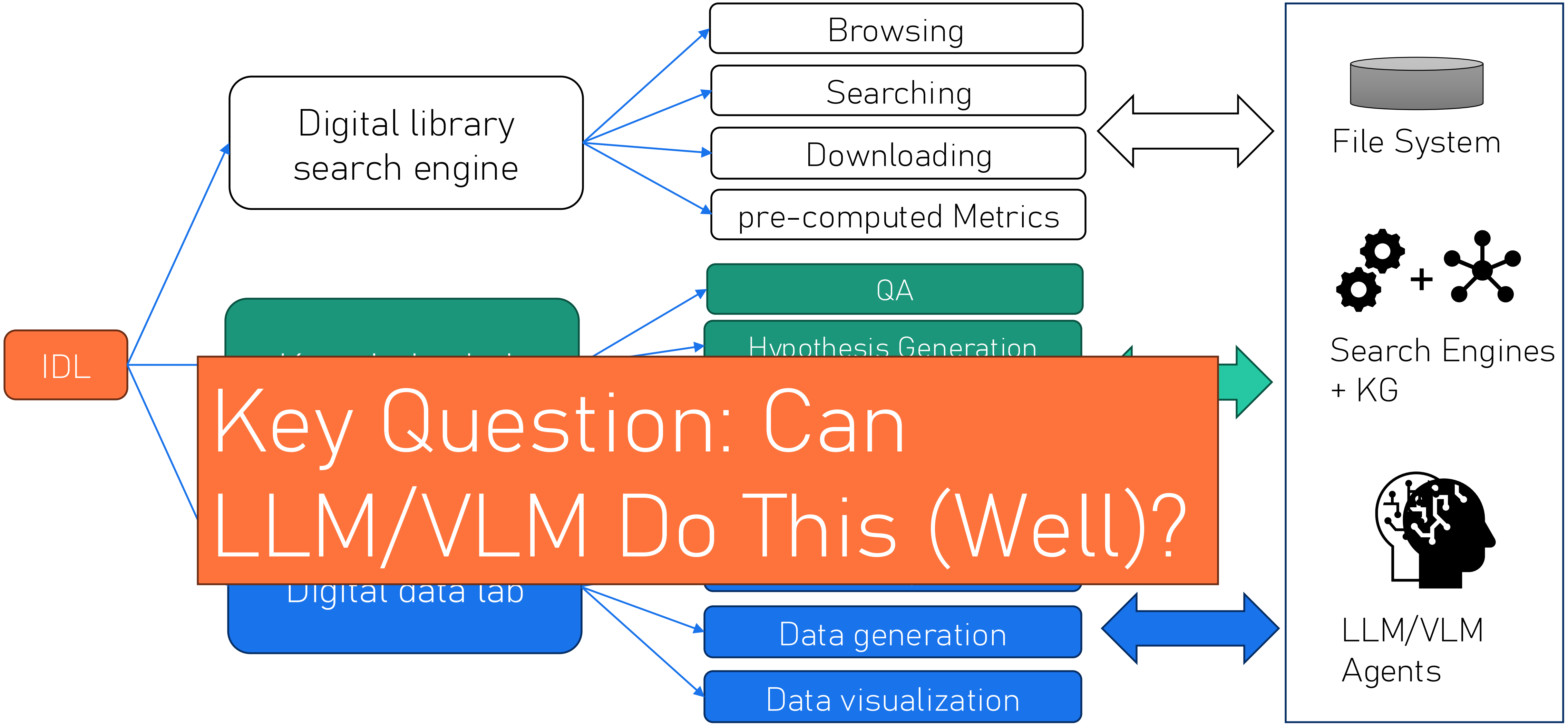
What’s Next? Intelligent Digital Libraries (IDLs)

- Providing digital services that require semantic, multigranularity reasonings about a variety of documents and their derivatives.
- Implemented by
  - Backend (document acquisition, IE, index, repository (document, data, software, etc.), LLM/VLM agents, KGs, internal and external connections)
  - Frontend (more interfaces will be created to support user-user, user-document, user-data, user-knowledge **interactions**)

A Vision of the Intelligent Digital Libraries (IDLs)



A Vision of the Intelligent Digital Libraries (IDLs)



Key Question: Can LLM/VLM Do This (Well)?

IDL as a Knowledge Hub
IDL as a Digital Lab

Recent Research on LLM/VLM-based Functionalities

QA – Offloading Reading to Bots

- Single document QA
  - Answer questions after reading a single scholarly document provided by a user
- Multi-document QA
  - Answer questions after reading multiple documents provided by a user
- Users
  - Any one to need to read papers (students, professors, researchers, reviewers, etc.)

ChatPDF

Humata

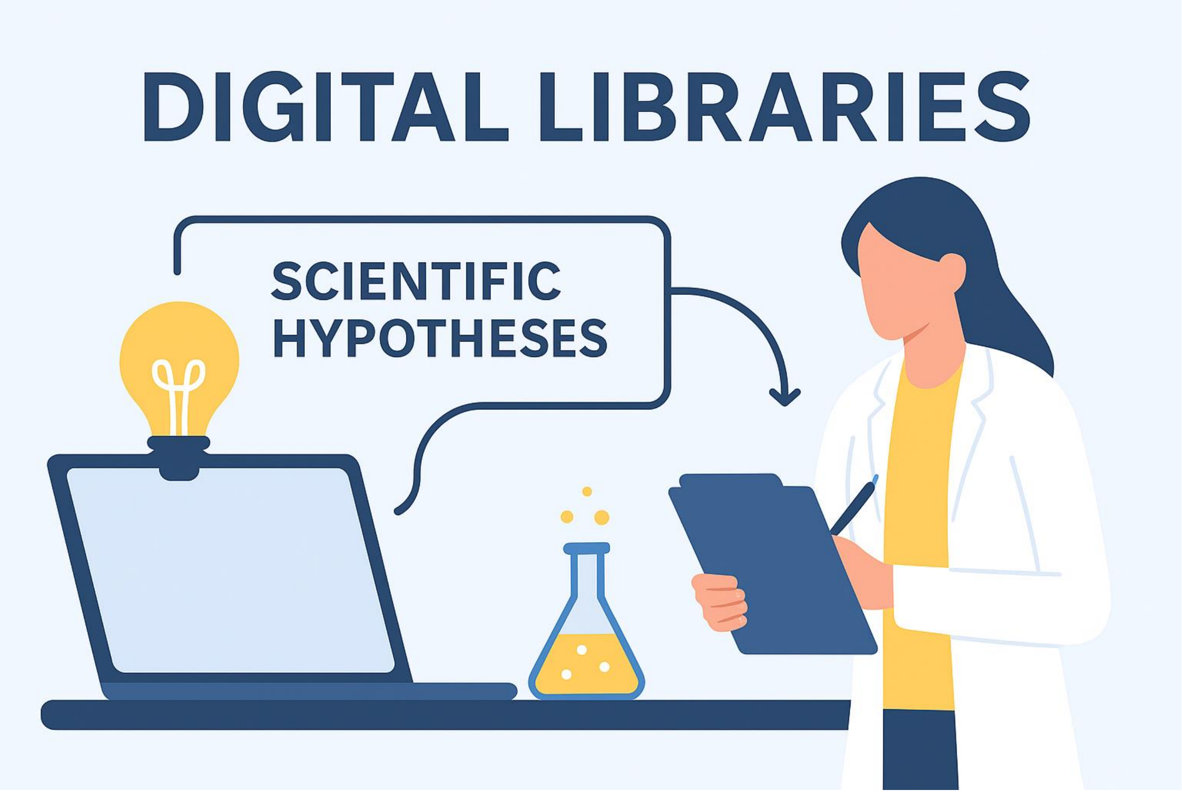
SciSummary

AskYourPDF

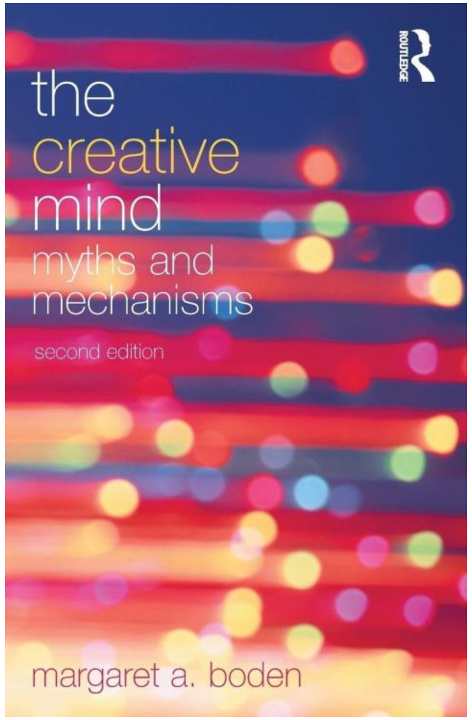
Limitations of Existing Tools

- Limitation: users must know which documents to use to find the answers
- Example: what evidence has been reported to refute that Covid was produced in the lab?
- Vision: automatically reason on hundreds of millions of documents to answer the questions (as opposed to simply search for relevant documents)

Hypothesis Generation

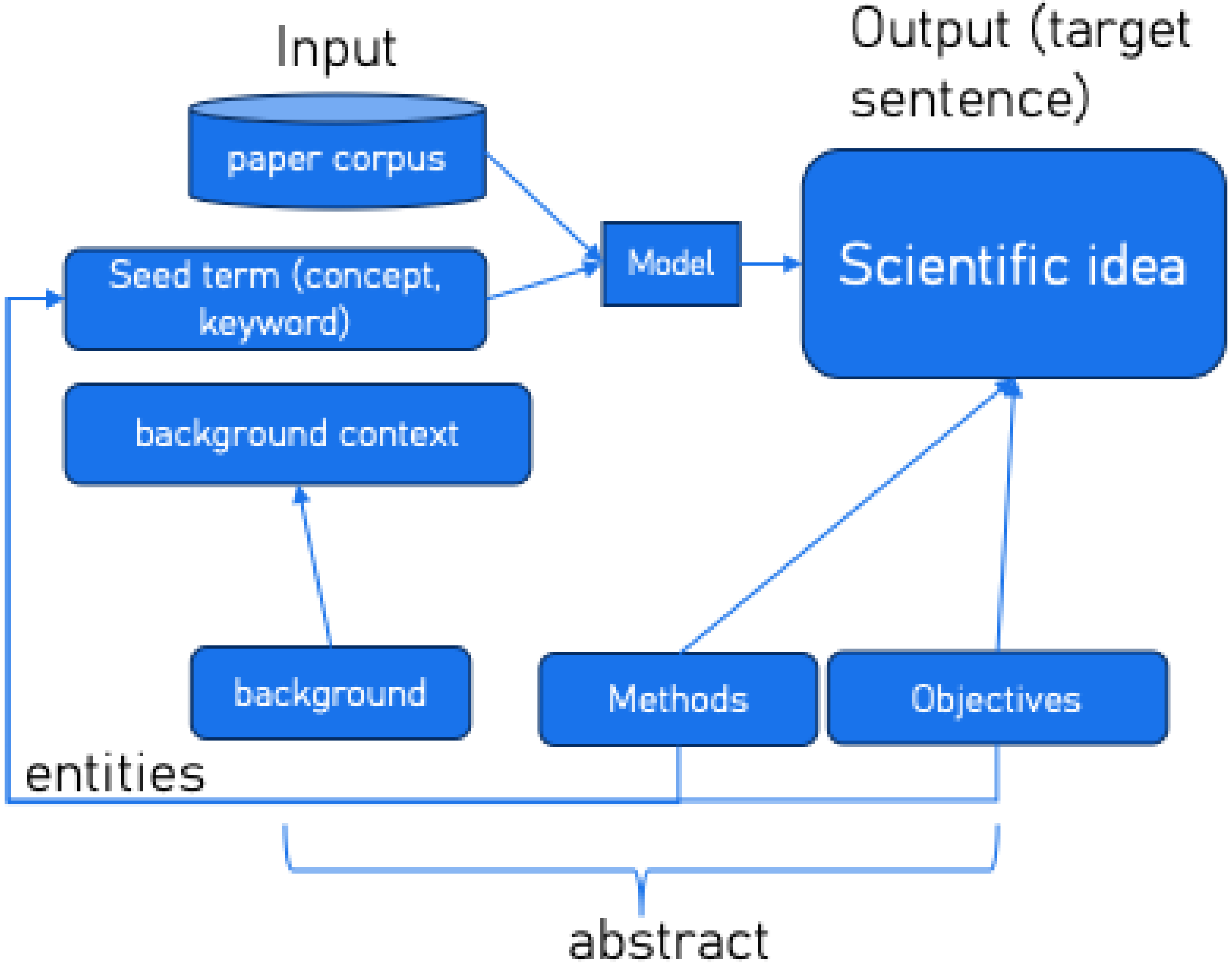


- Can IDLs help scientist propose novel and feasible scientific hypotheses and then plan experiments to verify the them?

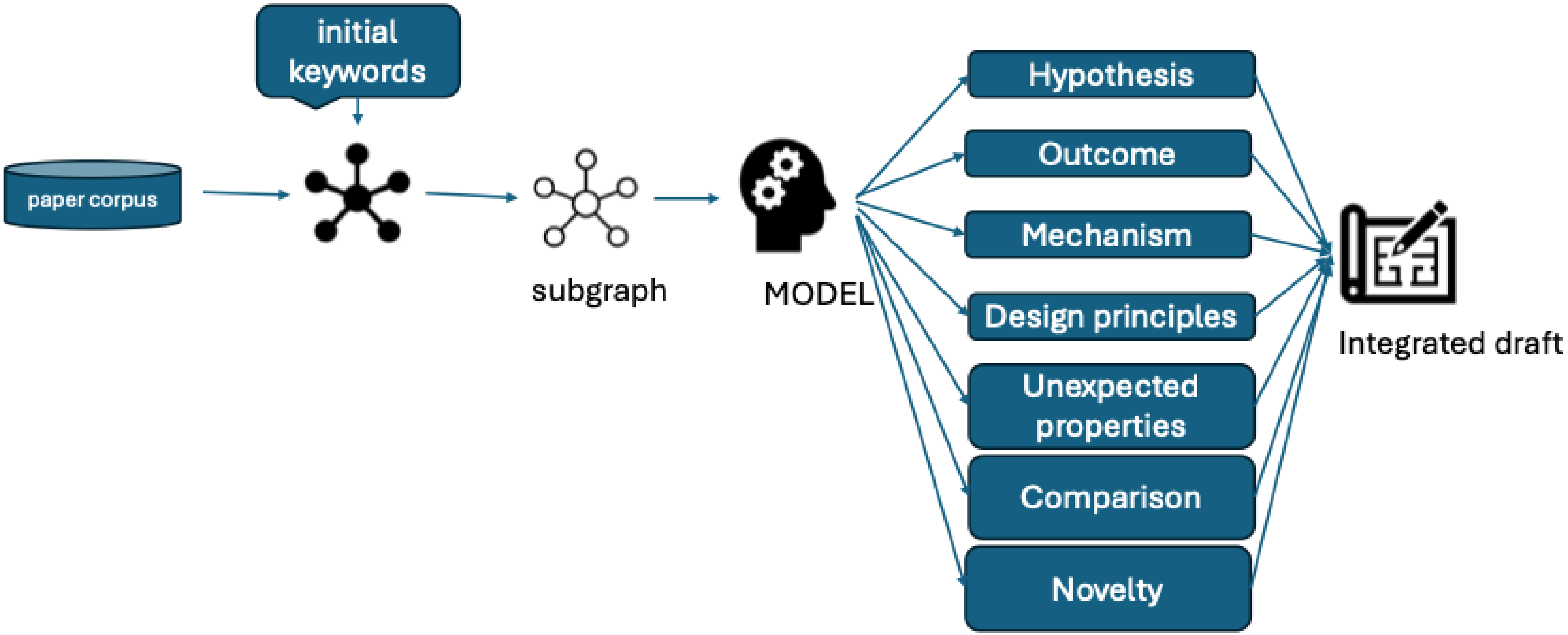


Boden, M. A. (2004). The Creative Mind: Myths and Mechanisms

SciMon (Wang et al. ACL 2024)



SciAgents (Ghafarollahi et al., Advanced Materials, 2025)



Hypothesis Generation

What Are Research Hypotheses?

Jian Wu[1,*], Sarah Rajtmajer[2,*]

[1]Old Dominion University

[2]The Pennsylvania State University

SciK 2025

Title	Input	Output	Reference
SciMon	Background + Keywords	Ideas	Wang et a. (ACL'24)
SciAgents	Keywords	Proposal	Ghafarollahi et al. (Advanced Materials, 2025)
Hypothesis generation with large language models	Data	Hypothesis	Zhou et al. (NLP4Science'24)
AI-CoScientist	Research Goal	Proposal	Gottweis et al., (arXiv:2502.18864)

Evaluation is a major challenge!

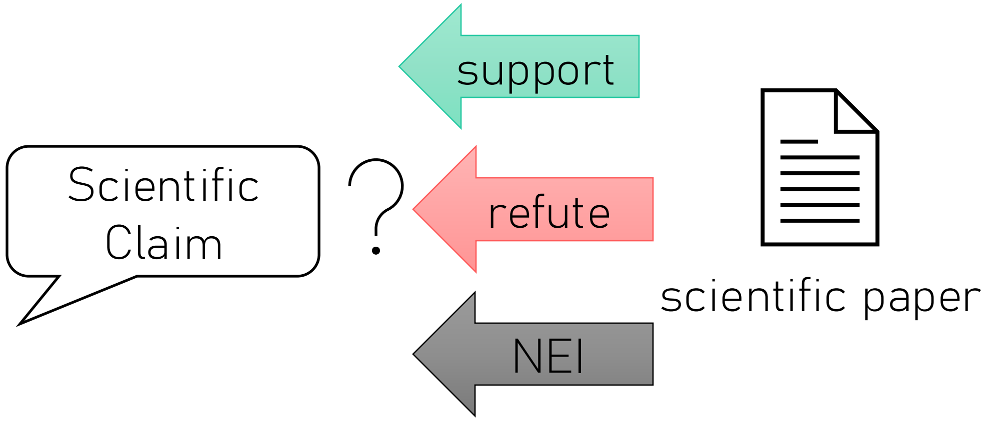
- Models have different input and output.
- A lack of benchmark and expert evaluation standard.

Our Proposed Work

- We propose generating highly novel and interdisciplinary citation-enriched hypothesis proposals through iterative interactions between expert LLMs.
- One challenge is evaluation. How to automatically evaluate the quality of the generated hypotheses?
  - Novelty
  - Feasibility

OLD DOMINION UNIVERSITY
OAK RIDGE National Laboratory
PennState

Scientific Claim Verification



- Problem definition: Given a claim and a scientific paper, can AI tell us if the paper supports or refute claim (or does not provide enough information (NEI))?
- Example A (claim in a scientific paper): Is there an association between social media use and bad mental health outcomes?
- Example B (claim in news or social media): Use of hand sanitizer can seriously mess with breath alcohol test results.
- Use cases
  - Scientific review
  - Misinformation and disinformation

Discern Claims (hypotheses) in Scientific Papers

Can Large Language Models Discern Evidence for Scientific Hypotheses? Case Studies in the Social Sciences

Sai Koneru[1], Jian Wu[2], Sarah Rajtmajer[1]
[1] Pennsylvania State University, State College, PA
[2] Old Dominion University, Norfolk, VA
{sdk96, smr48}@psu.edu, j1wu@odu.edu

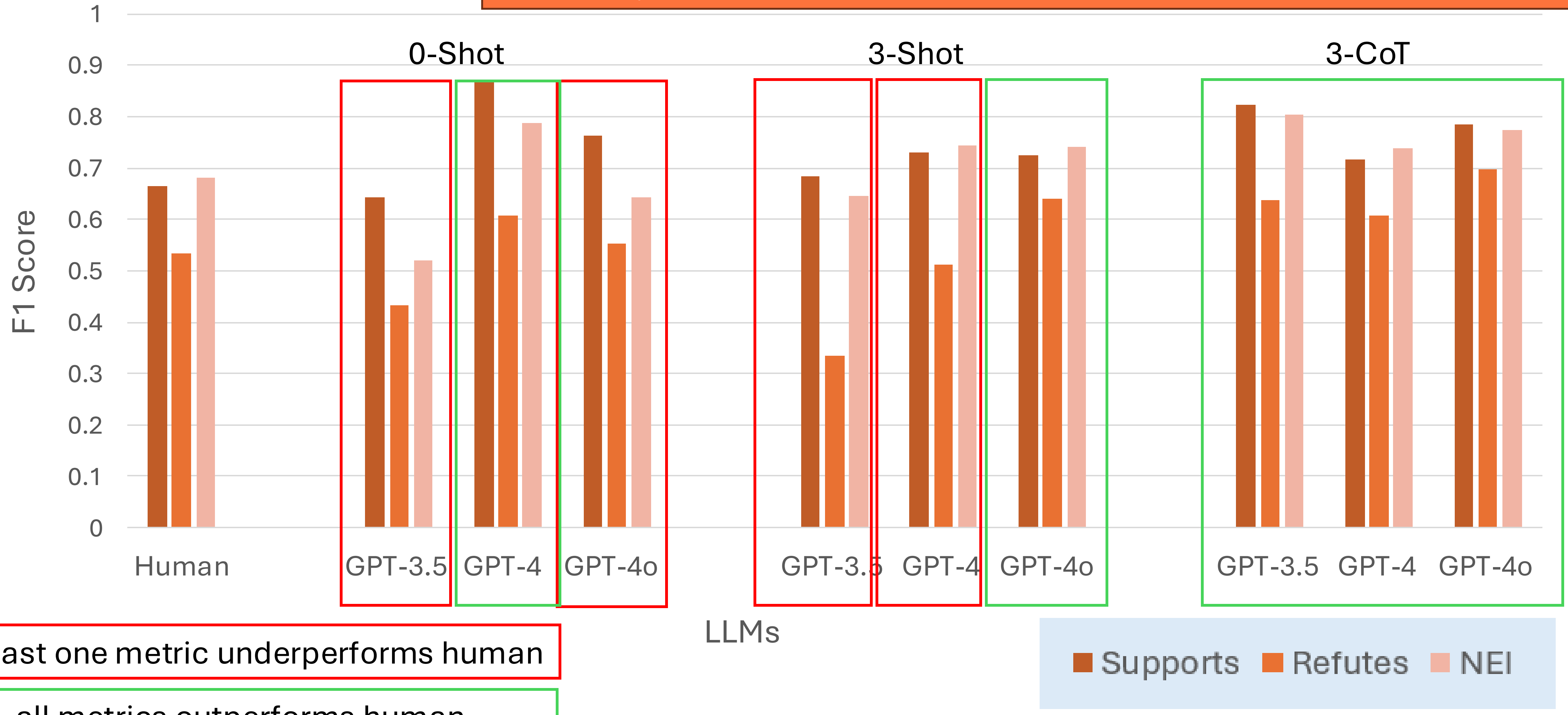
LREC COLING 2024

AI underperforms domain experts on discerning claims (hypotheses) in scientific papers.

Best model of each type	Accuracy	Macro F1
embedding + supervised classification	70.31%	0.615
Transfer learning	67.97%	0.523
GPT3.5 few-shot	66.57%	0.576
PaLM2	62.87%	0.536

Discern Claims in News and Social Media posts

- AI beats college students on discerning claims from news and social media posts. (Evans et al. 2024, Dzhaman et al. 2025 NCUR)



at least one metric underperforms human

all metrics outperforms human

Automatic Reproducibility and Replicability Assessment

- Reproducibility: same data, same method
- Replicability: different data, same method
- Reproducibility and replicability crisis in
  - Social and Behavioral Science (SBS) (Camerer 2016 Nature; Camerer 2018 Nature)
  - Computer Science (Moraila et al. 2014 PloS; Collberg et al. 2016)
  - Artificial Intelligence (Raff et al. 2019 NeurIPS; Gundersen et al. 2018 AAAI; Haibe-Kains et al. 2020 Nature; Ajayi et al. 2023 ICDAR)
  - Biomedical Science (Gentleman et a. 2005)

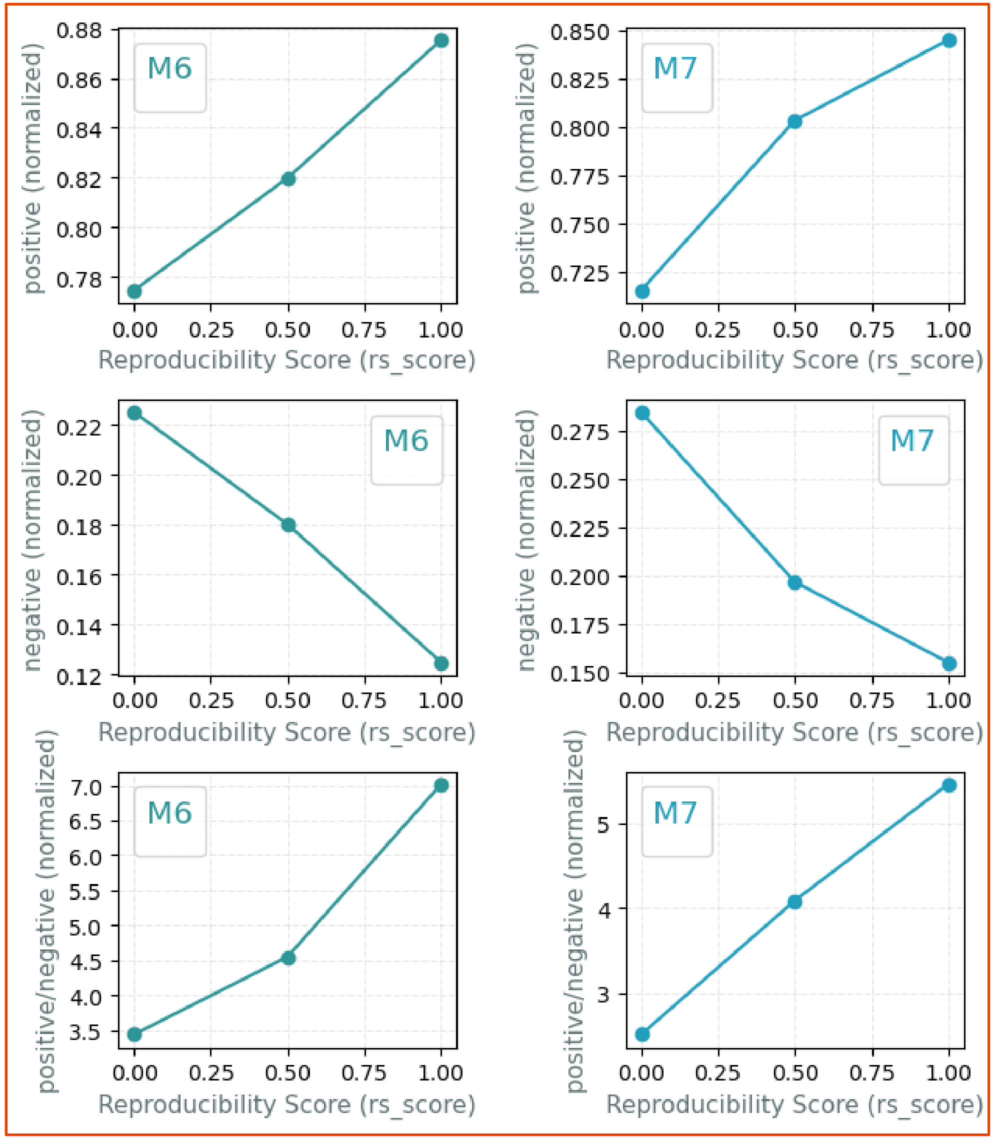
The Challenge of Reproducibility and Replicability

- Manually reproducing reported results is time-consuming and not scalable
- Average time to reproduce the main results in one paper
  - Table Structure Recognition (an AI task): 8 hours (using code and data provided by the original authors; Ajayi et al. 2023 ICDAR)
  - General AI tasks: 53.5 days (using re-implemented codes and data provided by the original authors; Raff 2023 AAAI)
  - Social and Behavioral Science: months – up to 1 year (using the same methods and new data collected from new user studies)

Can IDL tell me the reproducibility and replicability of a paper?

Citation Context Sentiments vs. Reproducibility Scores

- Correlation analysis between sentiment and reproducibility scores for ML papers
- Positive sentiment correlates with higher reproducibility scores
- Negative sentiment correlates with lower reproducibility scores



SHORT: Can citations tell us about a paper's reproducibility? A case study of machine learning papers

Rochana R. Obadage
Old Dominion University
Norfolk, VA, USA
oruma001@odu.edu

Sarah M. Rajtmajer
IST, Pennsylvania State University
University Park, PA, USA
smr48@psu.edu

Jian Wu
Old Dominion University
Norfolk, VA, USA
j1wu@odu.edu

ACM REP '24

Automatic Reproducibility and Replicability Assessments

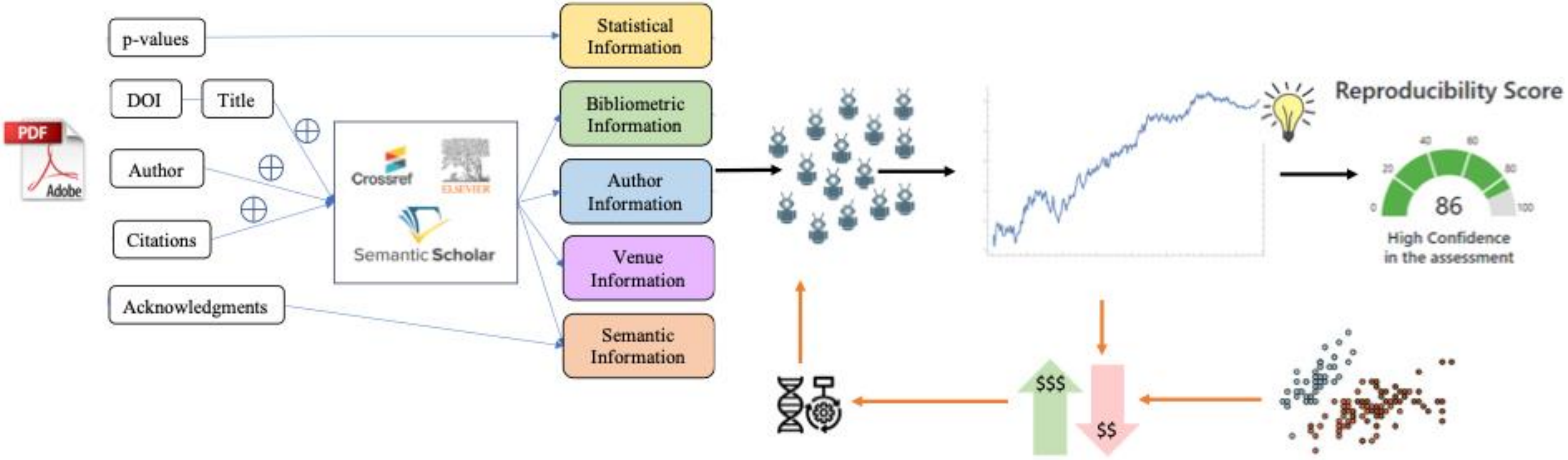
- Machine Learning (Wu et al. 2021, arxiv)
- Deep Learning (Yang et al. 2020 PNAS; Wu et al. 2023 PNAS)
- Prediction Market (Viganola et al. 2021 R. Soc. Open Sci.)
- Synthetic Prediction Market (Rajtmajer et al. 2022 AAAI; Chakravorti et al. 2023 HHAI)
- Question: can we use LLM agents to automatically reproduce/replicate the claims?

Machine Learning vs. Neural Models vs. Prediction Market

Model Type	Model	F1	Accuracy
ML (Our Work)	**Random Forest**	**0.61**	**0.65**
	XGBoost	0.57	0.62
	Naïve Bayesian	0.52	0.62
	MLP	0.54	0.59
	Logistic Regression	0.51	0.56
	SVM	0.46	0.56
Neural	**word2vec on full text (Yang et al. 2020 PNAS)**	**NA**	**0.65-0.78**
Prediction market	Prediction market (Chakravorti et al. 2023)	NA	0.71
Human	**Human survey** (Chakravorti et al. 2023)	**NA**	**0.58**

** The test samples of ML, Neural, and Prediction Market overlap but are not exactly the same but are all in social and behavioral sciences.

A Synthetic Prediction Market for Estimating Confidence in Published Work (Rajtmajer et al. 2022 AAAI)



Synthetic prediction markets— *Prediction markets populated by artificial agents (trader-bots), trained and updated within human-expert prediction markets, but deployable "offline".*

- Trader-bots will represent atomic (human-interpretable) properties of relevant signals, including full text of scientific papers, metadata for specific papers, and metadata about the community and the field.
- Bots will learn trading patterns from subject matter experts engaged in prediction markets, but unlike their human counterparts, will have comprehensive, unbiased view of the existing literature and metadata.

Synthetic Prediction Market (Our Work)

AAAI-22

A Synthetic Prediction Market for Estimating Confidence in Published Work

Sarah Rajtmajer,[1] Christopher Griffin,[1] Jian Wu,[2] Robert Fraleigh,[1] Laxmaan Balaji,[1] Anna Squicciarini,[1] Anthony Kwasnica,[1] David Pennock,[3] Michael McLaughlin,[1] Timothy Fritton,[1] Nishanth Nakshatri,[1] Arjun Menon,[1] Sai Ajay Modukuri,[1] Rajal Nivargi,[1] Xin Wei,[2] C. Lee Giles[1]

[1]The Pennsylvania State University
[2]Old Dominion University
[3]Rutgers University
{smr48,cxg286,rdf5090,lpb5347,acs20,amk17,mvm7085,tjf115,nzn5185,amm8987,svm6277,rfn5089,clg20}@psu.edu
{j1wu,xwei001}@odu.edu
david.pennock@rutgers.edu

Caveats:

- Feature extraction is noisy
- Bots may not always converge (hybrid market)

Can LLM directly reproduce/replicate the work in the paper?

Results on scored papers. Our system provides a confidence score for 68 of 192 (35%) of the papers in our set. On the set of scored papers, accuracy is 0.894, precision is 0.917, recall is 0.903, and **F1** is **0.903** (macro averages). A sizeable un-scored subset of data (65%) is the trade-off for high accuracy on the scored subset of the data. A test point is un-scored when the system has determined it has insufficient information to evaluate it.

LLM-based Reproducibility and Replicability Assessment

Features	Core-Bench PRINCETON UNIVERSITY	PaperBench OpenAI
References	Siegel et al. (2024 arxiv2409.11363)	Starace et al. (2025 arXiv2504.01848)
Nature of Work	Reproducibility	Replicability (same data, same method, new implementation)
Domain	CS, Bio, and Social	ML
#Papers	90	20
#Tasks	270	8,316
Agent Scaffolding	Adapted AutoGPT, orchestration: Python subprocess	Inspect AI basic agent, orchestration: nanoeval
Input	Code, Original Data, Paper	Paper only
Human comparison	No	Yes (Top PhD)
Conclusion	GPT-4o-based CORE-agent achieves the best reproducibility rate	Claude 3.5 Sonnet (New) with open-source scaffolding achieves the highest score

Our Ongoing Research

- Goal: building a new benchmark for LLM agents to replicate claims in Social and Behavioral Science papers
- Data : 200+ papers, pre-registrations, and human replication study reports from the SCORE project (Nosek et al. 2021 ARP)
- Replicability (new data), multi-difficulty level, multi-stage, no-human involved

CENTER FOR OPEN SCIENCE

PennState

UNIVERSITY OF NOTRE DAME

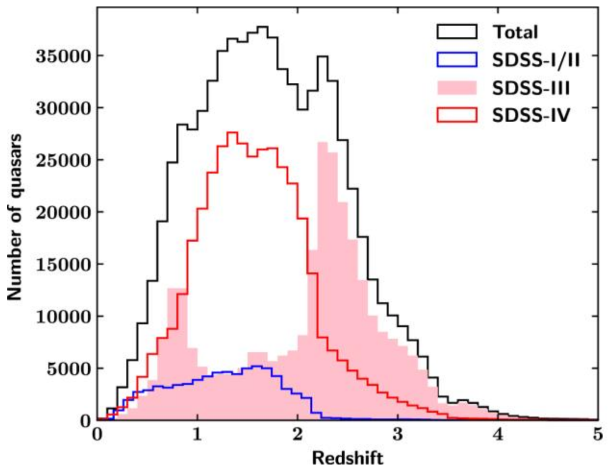
OLD DOMINION UNIVERSITY

IDL as a Knowledge Hub
IDL as a Digital Lab

Recent Research on LLM/VLM-based Functionalities

Data Compilation

- Why is it a problem for domain scientists?
  - Tons of data are published in PDFs
  - Data can be published in multiple modalities (text, figures, tables)
  - Compiled data published in papers can be very useful for analysis
    - Astronomy: Catalog. SDSS quasar catalog, Véron-Cetty & Véron Quasar Catalog, etc.
    - Geoscience: Geochemical database of elemental abundances: GEOROC, and EarthChem
    - Materials Science: the CALPHAD database of thermodynamic parameters (e.g., Gibbs free energy functions, interaction parameters) for elements, compounds, and phases.
- Question: can we automatically and faithfully compile data from scholarly papers into a pre-defined table schema for downstream analysis?

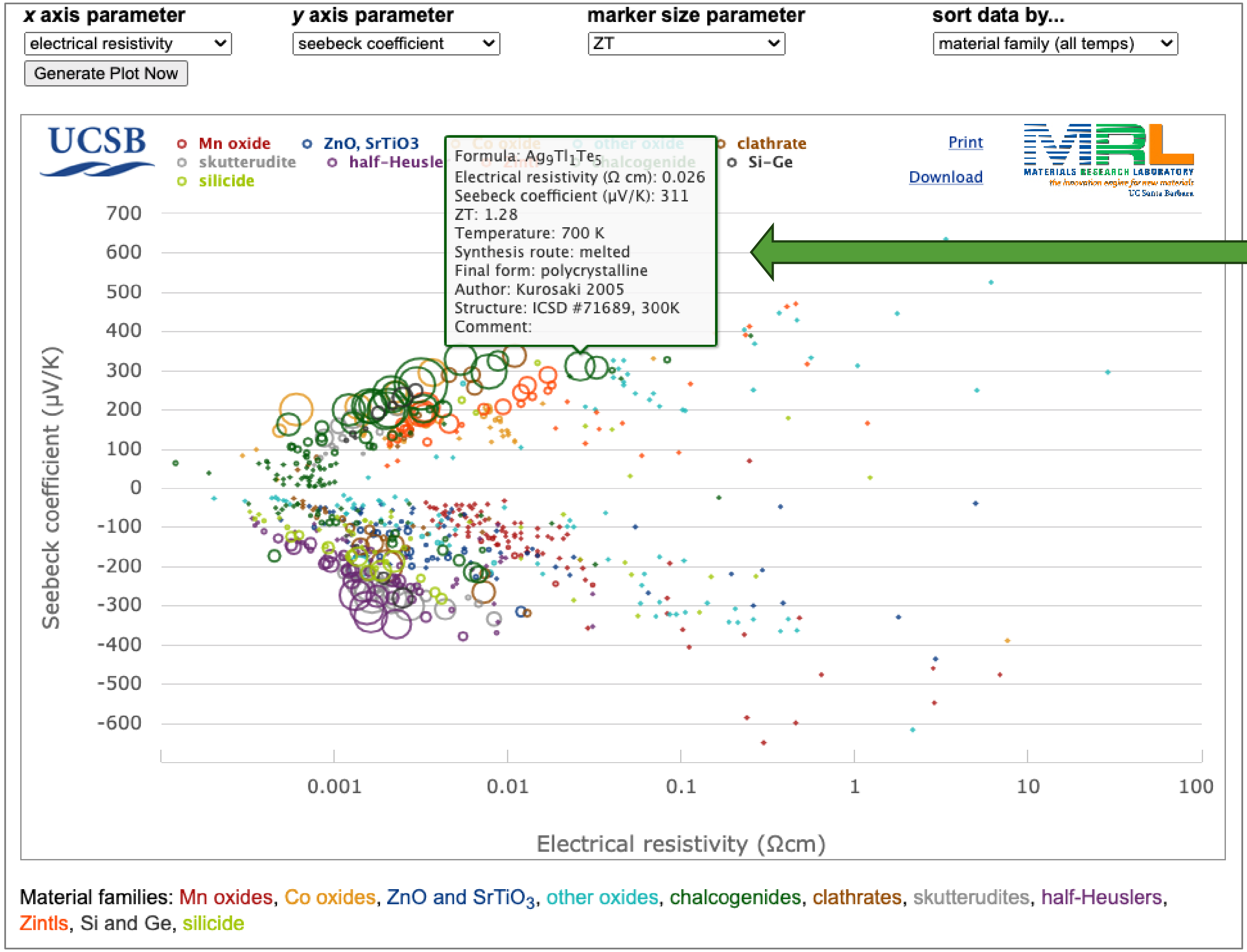


GEOROC
Geochemistry of Rocks of the Oceans and Continents

CALPHAD 1973

An Example in Materials Science

http://www.mrl.ucsb.edu:8080/datamine/thermoelectric.jsp



source paper


ysics Letters

Ag_9TlTe_5: A high-performance thermoelectric bulk material with extremely low thermal conductivity

Cite as: Appl. Phys. Lett. **87**, 061919 (2005); https://doi.org/10.1063/1.2009828
Submitted: 21 April 2005 . Accepted: 01 July 2005 . Published Online: 04 August 2005

Ken Kurosaki, Atsuko Kosuga, Hiroaki Muta, Masayoshi Uno, and Shinsuke Yamanaka

https://aip.scitation.org/doi/10.1063/1.2009828

Gaultois et al. 2013

An Example in Planetary and Earth Science

Can LLM/VLM Do this ?

Automatically Compile the Apollo Basalt Database

The petrology and chemistry of basaltic fragments from the Apollo 11 soil, part I

D. W. Beaty, S. M. R. Hill, and A. L. Albee

Division of Geological and Planetary Sciences* California Institute of Technology, Pasadena, California 91125

M.-S. Ma and R. A. Schmitt

Department of Chemistry and the Radiation Center, Oregon State University, Corvallis, Oregon 97331

Lunar sample age and errors in billion years

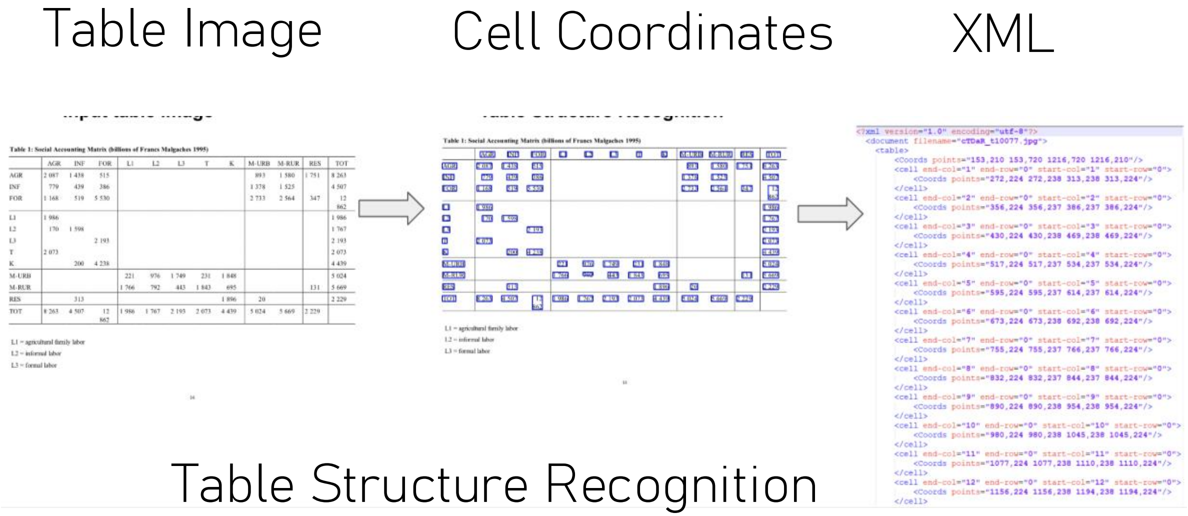
Elemental abundance pattern

Age (by)	Age error (by)	Dating method	Cone age ref.	SiO2	TiO2	Al2O3	Cr2O3	FeO	MnO	MgO	CaO	Na2O	K2O	P2O5	BaO	S	SO3	NiO	ZrO2
NA	NA	NA	NA	NA	9.7	9.5	0.266	19.5	0.248	7	11.1	0.396	0.073	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	11.1	7.3	0.342	20.5	0.218	9	10.5	0.462	0.28	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	11.5	7.7	0.348	20.1	0.220	8	10.9	0.474	0.28	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	10.2	8.6	0.289	20.2	0.223	8	10.8	0.527	0.33	NA	NA	NA	NA	NA	NA
3.83	0.016	Ar-Ar	14	41.2	8.6	11.8	0.3	19	0.249	8	11.5	0.37	0.09	0.28	NA	NA	NA	NA	NA
3.896	0.019	Ar-Ar	14	NA	NA	NA	0.24	18.8	0.275	NA	10.7	0.38	0.094	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA	21.05	NA	NA	12.74	0.393	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA	16.29	NA	NA	16.58	0.54	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA	21.49	NA	NA	NA	0.29	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA	24.66	NA	NA	NA	0.40	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA	20.04	NA	NA	16.14	0.48	NA	NA	NA	NA	NA	NA	NA

Table 2. Apollo 11 high-K basaltic fragments: Compositional summary.

Wt. %	10085, 782	10085, 789	10085, 796	10085, 803	10085, 814	10085, 820	10085, 824	10002, 114	10002, 115	10002, 117	10032, 30	10085, 832	10031, 6	INAA errors (%)
TiO_2	10.2	11.6	10.2	10.1	9.3	10.3	12.1	11.1	11.5	10.2	10.3	11.1	11.4	<5
Al_2O_3	8.0	8.2	8.1	8.1	8.7	8.2	8.8	7.3	7.7	8.6	7.6	8.3	8.0	<5
FeO	20.0	21.5	20.2	20.4	20.0	20.6	20.4	20.5	20.1	20.2	19.1	19.5	20.7	<5
MgO	8	8	8	7	7	7	8	9	8	8	8	7	8	10–20
CaO	10.0	11.1	10.6	11.0	10.8	10.1	10.7	10.5	10.9	10.8	9.4	10.0	10.6	5–10
Na_2O	0.505	0.477	0.524	0.513	0.578	0.492	0.499	0.462	0.474	0.527	0.482	0.478	0.503	<5
K_2O	0.26	0.25	0.32	0.32	0.34	0.31	0.28	0.28	0.28	0.33	0.29	0.30	0.30	5–20
MnO	0.253	0.247	0.247	0.245	0.247	0.239	0.235	0.218	0.220	0.223	0.219	0.247	0.225	<5
Cr_2O_3	0.310	0.377	0.314	0.318	0.255	0.296	0.337	0.342	0.348	0.289	0.323	0.294	0.346	<5
ppm														
Sc	84	89	82	84	76	84	85	83	84	83	77	80	87	<5
V	66	84	63	66	52	56	79	74	72	59	70	59	65	10–15
Co	26	29	27	27	24	26	27	27	28	35	26	25	28	<5
Zr	520	360	430	440	400	500	—	500	320	490	370	470	370	30–40
Ba	230	240	330	320	380	330	290	250	330	330	270	400	330	10–35
La	23.0	22.3	27.2	26.9	33.0	28.7	25.2	27.0	25.6	29.9	27.7	25.7	27.3	<5
Ce	75	72	83	80	98	100	90	79	73	87	73	79	78	10–30
Nd	62	58	71	70	80	74	64	64	60	68	65	65	64	20–25
Sm	19.2	18.2	21.6	21.1	24.6	22.6	20.1	20.8	19.9	22.5	20.5	20.6	20.7	<5
Eu	2.11	2.06	2.25	2.26	2.55	2.55	2.19	2.31	2.29	2.61	2.17	2.18	2.23	5–10
Tb	4.4	4.2	4.7	4.6	5.2	5.0	4.6	4.5	4.4	4.7	4.5	4.5	4.3	10–15
Dy	31	29	32	31	37	33	30	29	29	32	30	30	30	10–20
Yb	15.5	15.1	17.3	17.2	19.6	18.2	16.3	16.6	15.1	18.2	16.9	16.8	17.2	<5
Lu	2.27	2.24	2.55	2.54	2.86	2.63	2.32	2.40	2.14	2.70	2.40	2.46	2.43	5–10
Hf	14.4	15.3	15.9	14.9	16.7	15.8	15.1	16.5	16.0	17.0	14.5	15.4	15.4	5–10
Ta	2.5	2.2	2.6	2.3	2.6	2.4	2.2	2.7	2.5	2.5	2.3	2.7	2.5	5–10
Th	1.8	1.6	2.1	1.9	2.8	2.3	2.0	3.6	3.5	3.8	3.3	2.1	2.8	10–15

Two Subtasks of Data Compilation (tabular data)



1. Data Extraction: faithfully extract data from tables in PDF documents
   - Step 1: Table Structure Recognition
   - Step 2: OCR
   - VLM-powered OCR or use a VLM to directly extract data
   - Improve data correction efficiency using Uncertainty Quantification (UQ)
2. Data Population: insert data into the correct position of a table with a pre-defined schema
   - Not extensively evaluated to our best knowledge

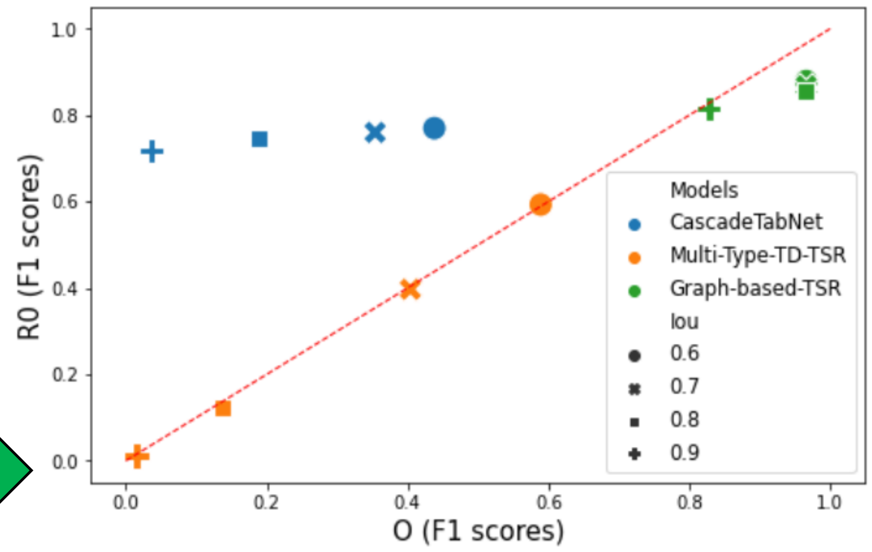
olmOCR
ChatGPT-4o
VISTA-OCR
LLaVA
miniCPM-v

Existing Performance on Table Data Extraction

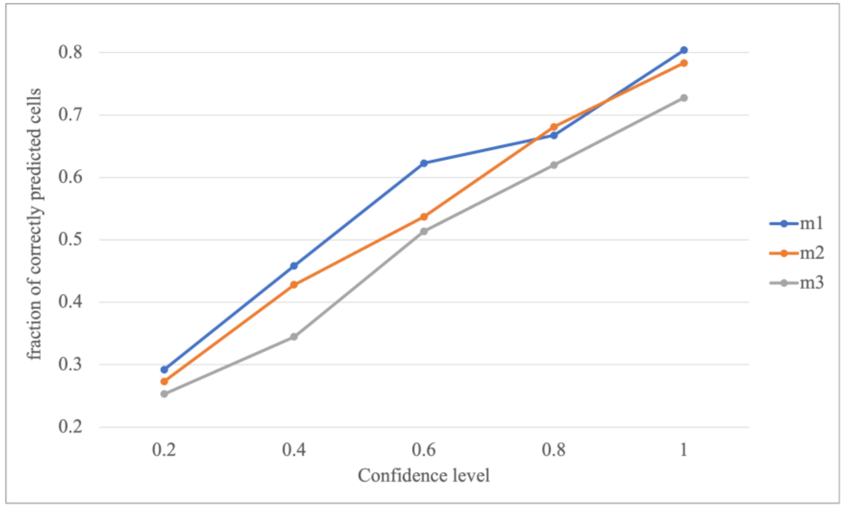
- CHATEXTRACT (Polak et al. Nature): extracting material data from research papers
  - precision = 90.8%, recall = 87.7%
- CHARTLLAMA (Han et al. arXiv:2311.16483): Chart data extraction
  - precision = 84.92%
- CHATGPT (Brown et al. 2020): extract structured data from clinical notes
  - precision of 77–99%, recall of 76–91%

Our Work on Table Data Extraction

1. Neural models can achieve 75%-85% F1-scores on table structure recognition. However, reproducibility is a big concern for many results reported (Ajayi et al. 2023 ICDAR).



2. The errors in extracted data are correlated with the uncertainty (or confidence levels) of the extraction results (Ajayi et al. 2024 IRI).

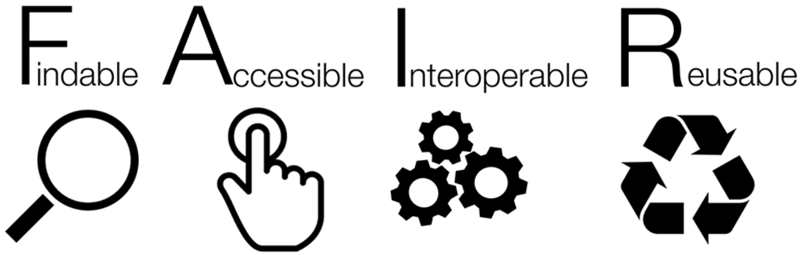


3. Conformal prediction is an effective UQ method for table data extraction and can reduce the manual effort of data correction by 50% (Ajayi et al. 2025 ICDAR).

Our Proposed Work

- Automatic data compilation:
  - Use VLMs to extracting tabular data and attributes from PDFs
  - Use LLM/VLM to populate data into a table or a KG (given the schema or ontology)
- Requirement: Ensure fidelity and trustworthy without fine-tuning (very limited training data)

Data Generation



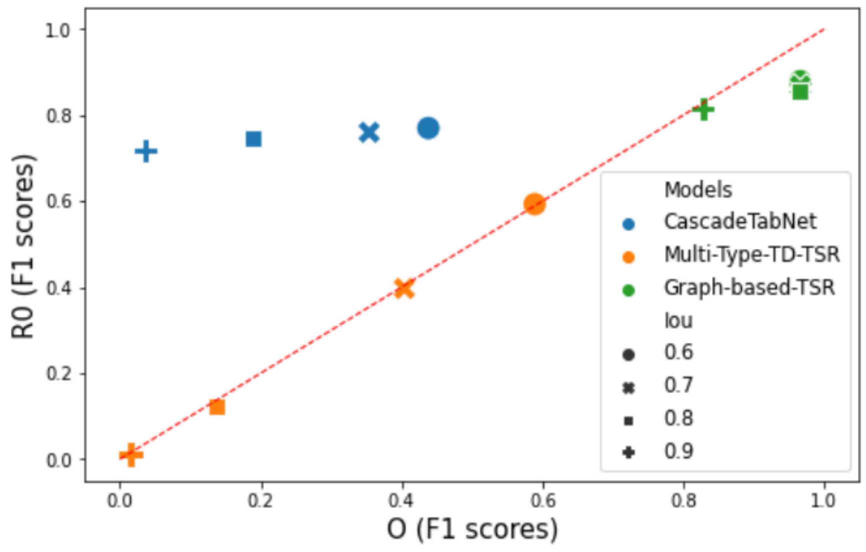
- Motivation: as a solution of data scarcity problem
- Research question: can you use LLMs to generate close-to or better than human generated data?
- Example: web data has depleted for LLM training, can we generate new data?
- Problem:
  - Data security and data quality (beyond the FAIR principle)
  - Integrate data generation modules to data extraction and compilation

Data Visualization

- Motivation: Same data can be illustrated in different ways, making conclusions more evident.

TSR Model	Data	IoU	$F1(O)$	$F1(R_0)$	Δ_0
CascadeTabNet	ICDAR 2019	0.6	0.438	0.770	0.332
CascadeTabNet	ICDAR 2019	0.7	0.354	0.760	0.406
CascadeTabNet	ICDAR 2019	0.8	0.190	0.745	0.555
CascadeTabNet	ICDAR 2019	0.9	0.036	0.718	0.682
Multi-Type-TD-TSR	ICDAR 2019	0.6	0.589	0.593	0.004
Multi-Type-TD-TSR	ICDAR 2019	0.7	0.404	0.397	-0.007
Multi-Type-TD-TSR	ICDAR 2019	0.8	0.137	0.124	-0.013
Multi-Type-TD-TSR	ICDAR 2019	0.9	0.015	0.012	-0.003
Graph-based-TSR	ICDAR 2019	0.6	0.966	0.879	-0.087
Graph-based-TSR	ICDAR 2019	0.7	0.966	0.868	-0.098
Graph-based-TSR	ICDAR 2019	0.8	0.966	0.856	-0.110
Graph-based-TSR	ICDAR 2019	0.9	0.828	0.815	-0.130

VS



Generative AI for Visualization

Four tasks

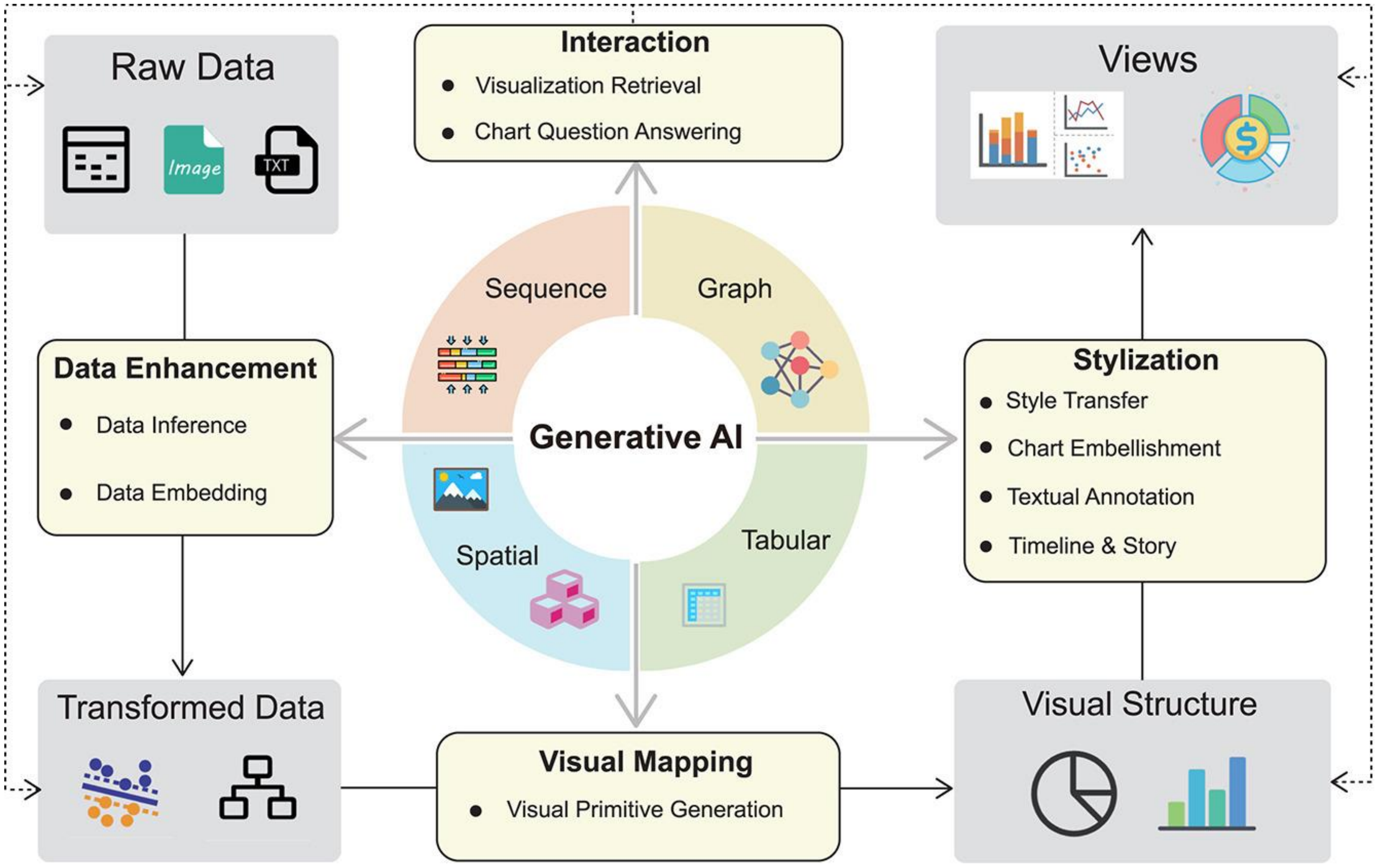
1. data enhancement
2. visual mapping generation
3. stylization
4. interaction

Four types of methods by data structure

- sequence
- tabular
- spatial
- graph

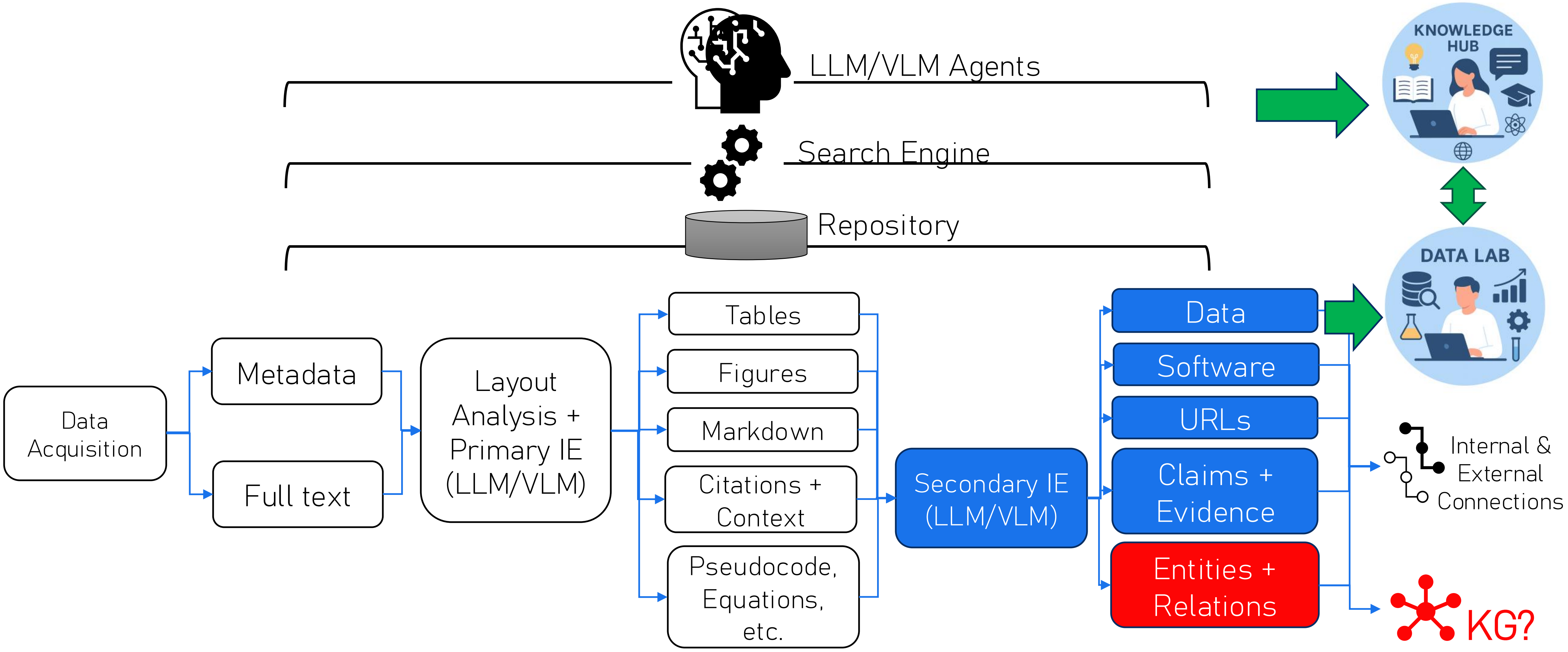
Three challenges

- evaluation
- dataset
- the gap between end-to-end GenAI and VIS.



Ye et al. 2024 Visual Informatics

How to Integrate Everything?

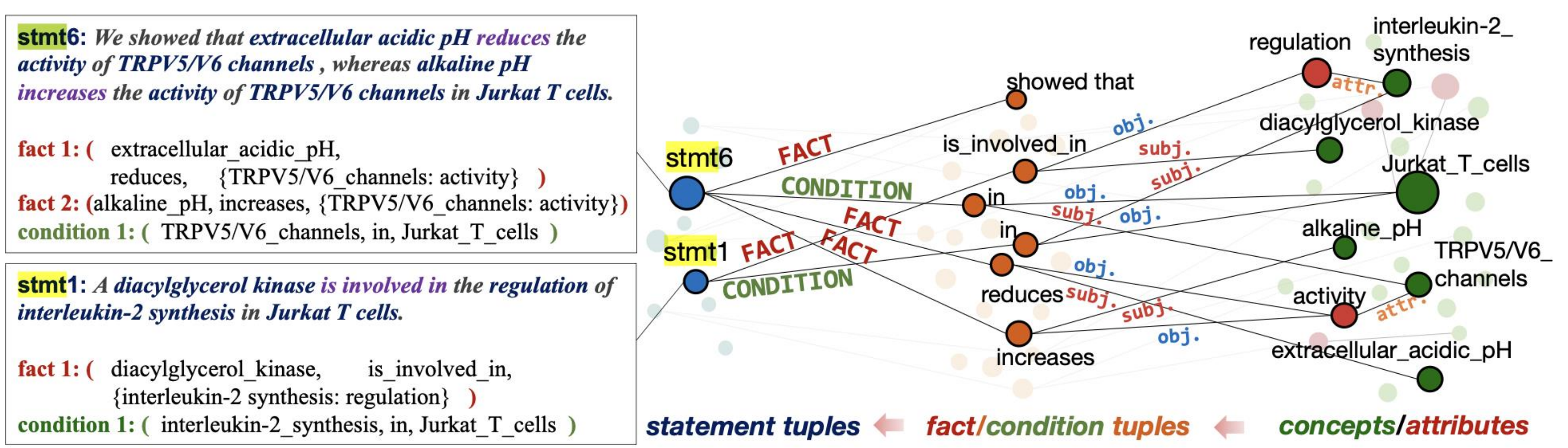


KG?

- We still need KGs because of
  - interpretability: easier to explain facts
  - scalability: relatively easy to add new knowledge
  - provenance document: track entities and relations from original documents
  - metric calculation: e.g., citation and author networks

- However, scientific KGs have intrinsic limitations
  - Conditional knowledge representation: hard to presentation complicated conditional knowledge in scientific statements (see below)
  - Unified ontology: hard to build a unified ontology for all scientific domains
  - Reliable relation recognition is still challenging

- Solution
  - Focused on meta-level KGs and KGs for specific domains with well established ontologies, e.g., SNOMED CT.



Jiang et al. (2019 KDD)

Do We Still Need Digital Library Search Engines?

- Yes!
- IDLs will be built on existing functionalities of digital library search engines to
  - meet users' needs (e.g., browsing, searching, downloading, metrics)
  - trace provenance of derivatives (e.g., figures, tables, claims)
  - provide explainability (e.g., references, pre-extracted features and metrics)
  - mitigate metadata-level hallucination (authors, publication year, DOIs, URLs, citations, references, etc.)

Summary

- Digital Library Search Engines still holds important roles in future digital libraries
- Intelligent DLs (IDLs) will add more value and facilitate **AI for Science** through Knowledge Hub and Data Lab
- Key Challenges
  - [Technical] How to build a multidomain scientific AI from general purpose AI?
    - Is reading papers enough? How about textbooks and knowledge graphs?
  - [Financial] How to persuade funding agencies to invest research and education? Justify the commercial value.
  - [Infrastructure] Computing power! – light-weight LLMs

Abstract

- Since 2023, there has been a surge of public and research interest in large language models (LLMs), which has significantly shifted the paradigm of information retrieval from returning keyword-based search results to the generation of natural language responses. This shift brings both challenges and opportunities for traditional digital libraries, which have served as a core infrastructure for browsing, searching, and accessing scholarly content. A critical question emerges: What role should digital libraries play in this LLM era? In this keynote, we share our vision of digital libraries in the LLM era. We argue that digital libraries are still indispensable, not only as repositories for digital preservation and provenance but also for trustworthy metadata discovery and verification. We explore how digital libraries can evolve by integrating LLMs and structured knowledge to support advanced services such as automatic data extraction, scholarly comparison, review generation, and science communication for broader audiences. We share preliminary work in this direction, including initiatives on preserving endangered open-access datasets and software, complex table data extraction, scientific claim verification, and assessing research reproducibility.