



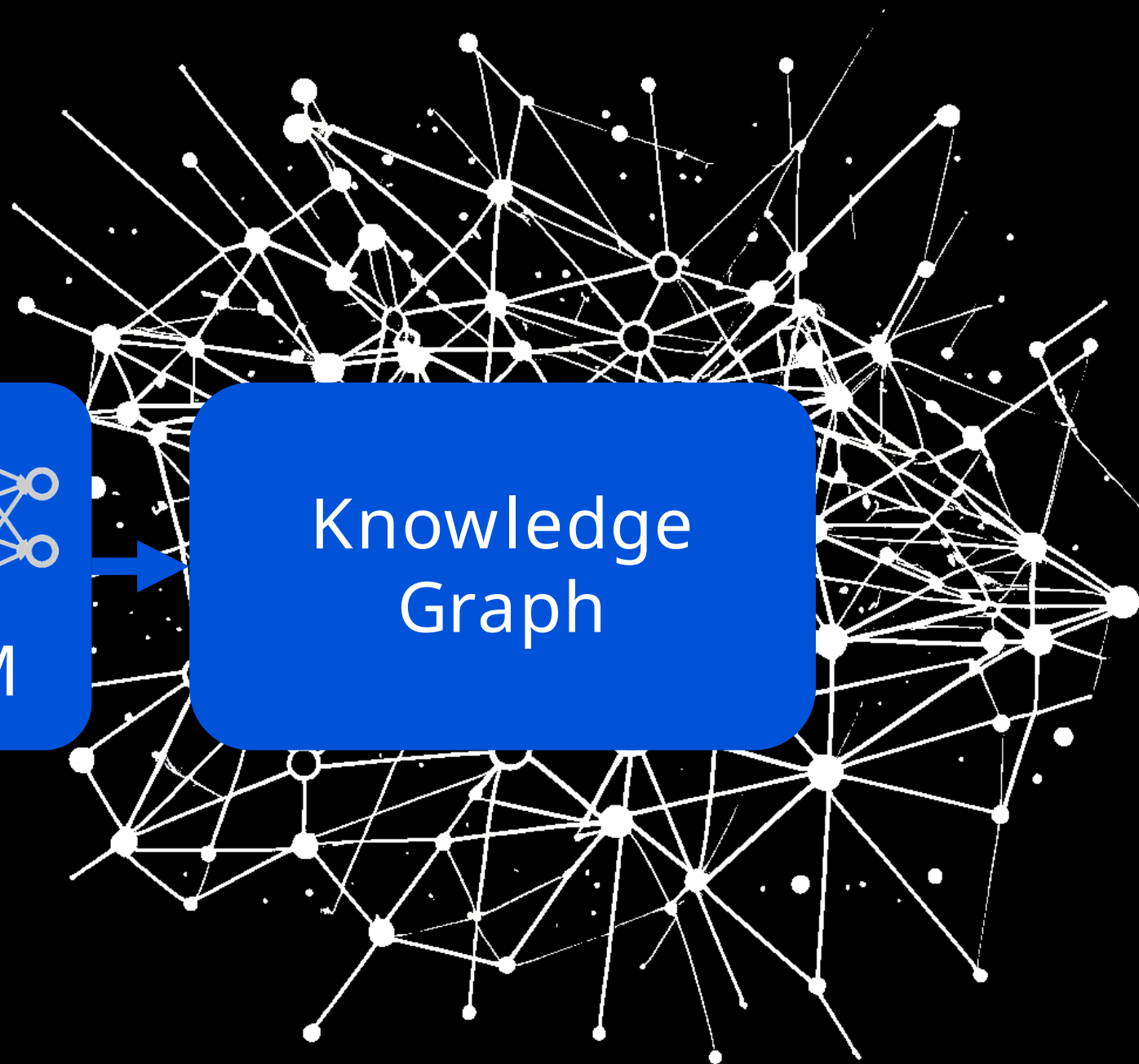
WELCOME

Information Retrieval Project

[RUNFISH.COM](https://runfish.com)

Content

- 01 — Abstract
- 02 — Information Retrieval Methods
- 03 — Building a New Text-to-Graph Pipeline
- 04 — LLM Evaluation Results
- 05 — Graph Search Paradigm



With the Text2Graph Project we revolutionize web search and enhance workflows by automatically converting text into knowledge graphs.

Challenge: Research publications have grown exponentially, yet readily information access remains hindered by conventional search paradigms. While AI with Retrieval Augmented Generation (RAG) and semantic search offers promising alternatives, the field still lacks a unified framework for cross-domain knowledge representation.

Our Approach: We built a tool that converts free-form text into structured knowledge graphs. Using a dataset of sentence-to-graph translations, we mapped sentences into relationships and categories without losing information. We then trained AI to apply this mapping to any running text, enabling it to quickly generate immense knowledge graphs, which are stored and managed in TypeDB for more powerful information retrieval than simply submitting documents to an AI.

Results: LLMs quickly mastered our extraction task with minimal training, allowing fast iteration. Even at this stage, our framework captures a wider range of information than existing methods, accurately representing over 90% of sentences.

Future Plans: We're evolving our text-to-graph tool into an AI agent short-term memory system while enhancing our proof of concept. Additionally, we aim to create a vastly larger graph that functions as a search engine, using population data to learn reality's deeper structure and improve fact-checking through information flow tracking.

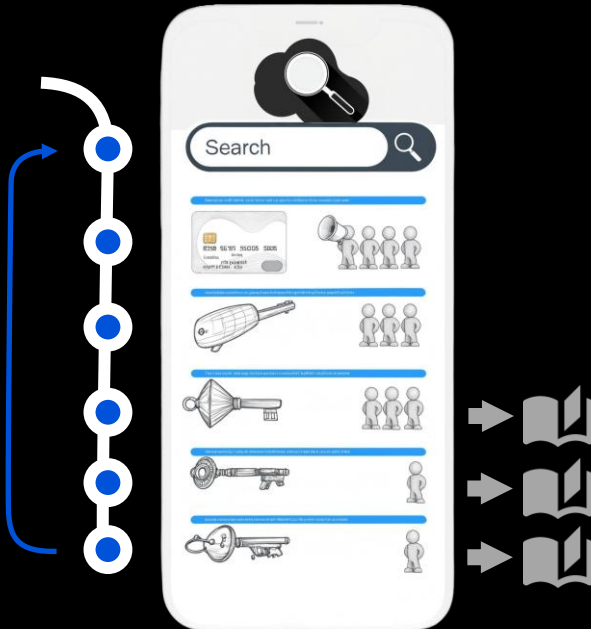
Team Expertise: Our team's expertise in Natural Language Understanding, Python, deep learning, and our interdisciplinary background were essential for the advancement of this yet largely overlooked approach.



Web Information Retrieval – Comparison

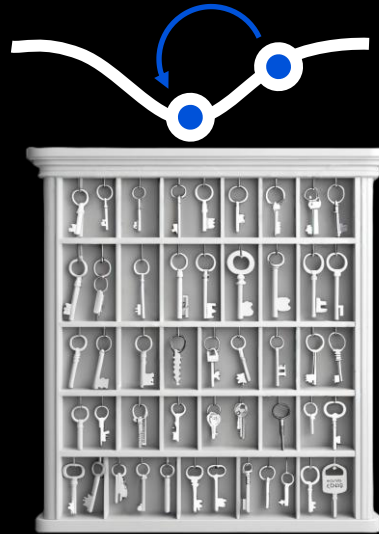
Keyword search

- ✓ Fast text lookup
- ✗ Ignores meaning



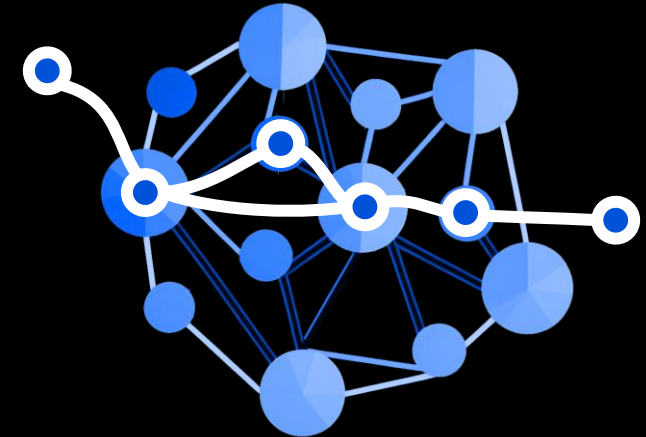
Relational DB

- ✓ Precise filtering
- ✗ Unnatural representation



Graph DB

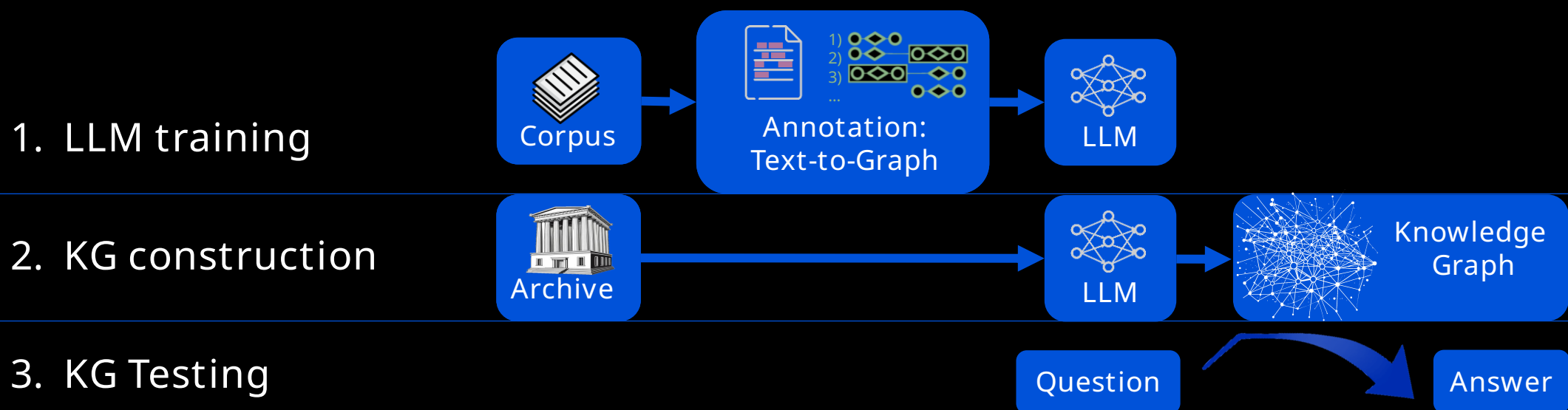
- ✓ Highest search control
- ✗ Hard to implement



AI Assistants are excluded due to hallucinations.

Building a New Text-to-Graph Pipeline

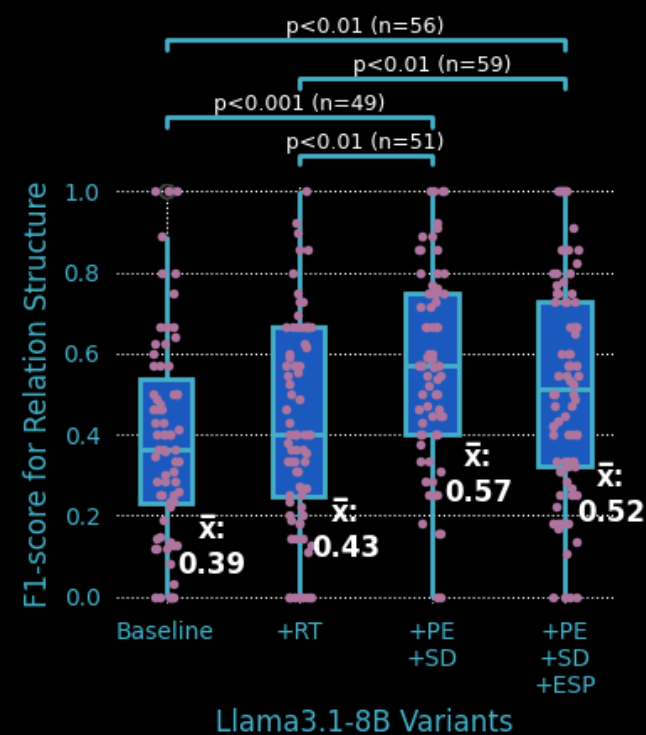
- Unlocking New Search Capabilities with Automated Text-to-Graph Systems



LLM Evaluation Results

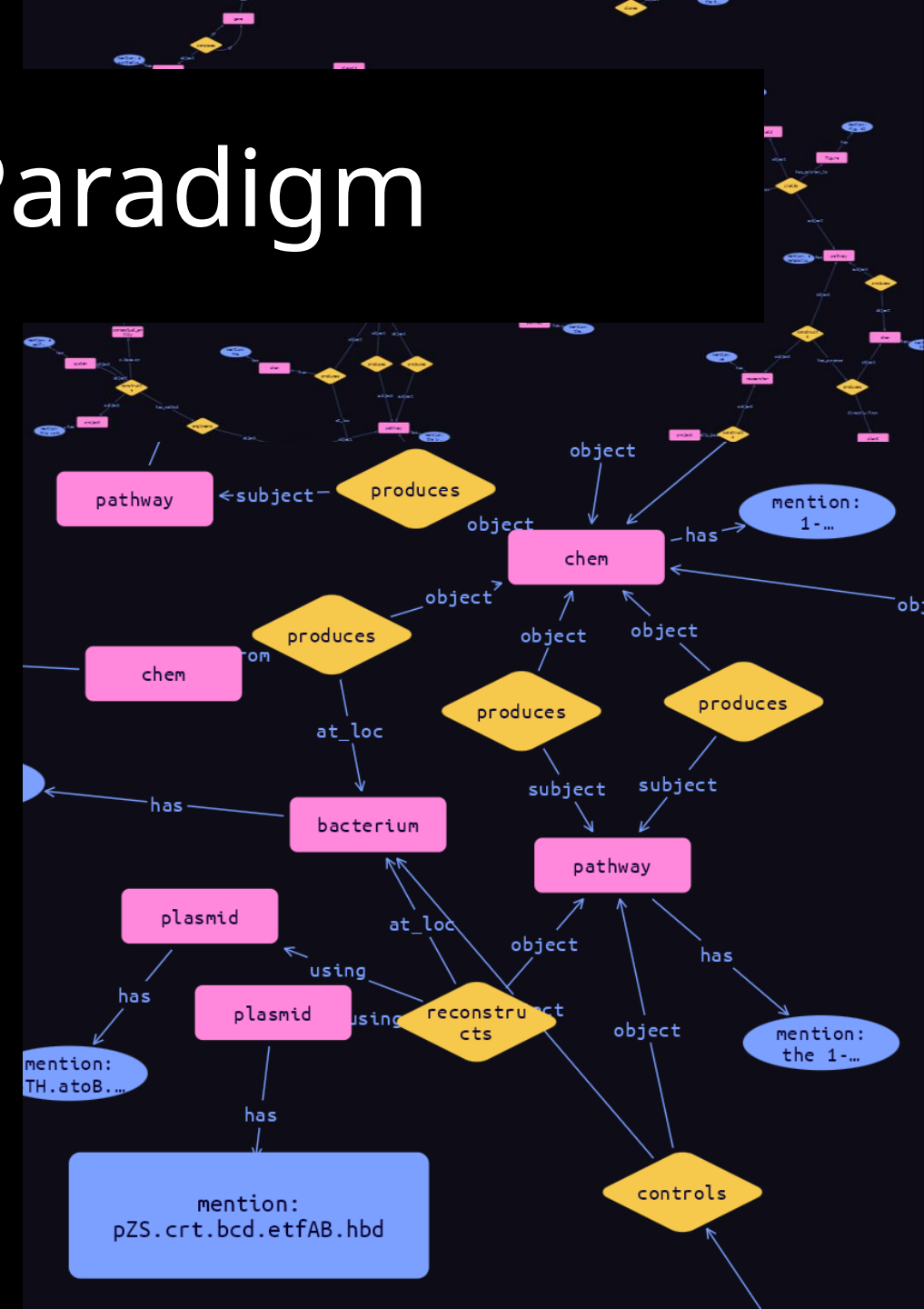
- Labeling guidelines align with the LLM pre-training manifold, ensuring strong generalization to test data.
- Sentences confusing models mirror those experts find challenging due to complexity and flawed grammar.
- Granularity and abstractions optimized for search also enhance LLM training.
- This approach benefits from ongoing improvements in open-source LLMs.

Screening Synthetic Data Strategies for Llama Fine-Tuning



Dynamic Graph Search Paradigm

- Graph search mirrors human thinking
- Our graph enhances semantic search with sharper resolution than rivals.
- Why  **TypeDB** backend?
 - Enables layered access rights for users and AI agents.
 - Built for machine reasoning, vital for advanced search.
 - Leverages inheritance, rule-based logic, nested and multi-link relations.





SOLE PROPRIETORSHIP

Chamber of Commerce: soon



Address:

Paulsstrasse 19

CH-8400, Winterthur ZH

Contact:

d.s.account@runfish.com



Thanks!

[RUNFISH.COM](https://runfish.com)

BONUS:

Markov logic networks

- Essentially MLN are:
 - Think of rules, and then it gives each rule a weight. And then the probability of a state of the world goes up with the number of states and the rules that are true in that world. So with this we can represent pretty much everything that we like.



And the most widely used is called Markov logic networks.