



KathDB: An Explainable Multimodal Database Management System with Human-AI Collaboration

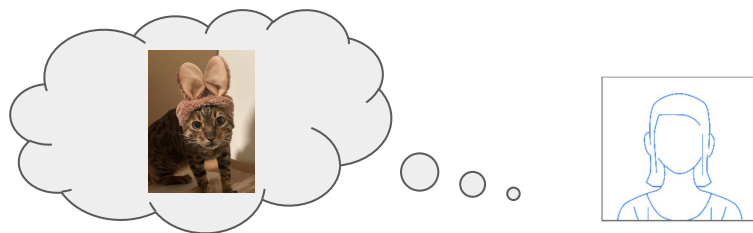
*Guorui Xiao, Enhao Zhang, Nicole Sullivan, Will Hansen,
and Magdalena Balazinska*

Paul G. Allen School of Computer Science & Engineering
University of Washington



Time for Multimodal DBMSs - Retail

Challenge: This query spans text, tables, images, and videos and is in natural language



Find fun cat toys under \$20, and verify that their review videos contain actual visual evidence of cats playing with the toys.



Cat Toy Images

Images

★★★★★ **My cats love it!** is on Ja
 Smank
 ★★★★★ **Broke but good**
 Reviewed in the United States on Novemb
Verified Purchase consider
 Good until it broke. My cat LOVEDDD this, switch t
 should. The size of it is very good I can lay

Purchase Review

Unstructured Text



Review Video

Videos

Purchase Table					
Product ID	Customer ID	Product ID	Purchase Date	Product ID	Price
1312	24213	9876	03-22-2016		
1313	24213	9876	03-22-2016		
1314	24213	9876	03-22-2016		

Customer ID	Customer	Address	Product ID	Price
24221	Bob	123 East street	9876	79
24222	Alice	123 Main street	9876	88
24223	Martha	456 North street	Product ID	58

Item Table

Tables

Time for Multimodal DBMSs - Medical Applications

PID	Name	Surname	Gender	DoB	Country	StreetAddress	City	Postal
10000	Adena	Reeves	F	1962-08-28	Montenegro	P.O. Box 936, 9290 Aptent Ave	Morkhoven	71344
10001	Buffy	Warner	M	2009-05-25	Guinea-Bissau	P.O. Box 624, 5536 Nunc St.	Graz	68114-186
10002	Kaye	Green	F	1994-07-23	Norway	P.O. Box 650, 1264 Tellus, St.	Bojano	T04 3WO
10003	Keiko	Gonzalez	M	1973-12-27	Iraq	1889 Magna. Street	Chelsea	8064
10004	Kylynn	Carver	F	1974-01-22	Tanzania	Ap #357-247 Per Rd.	Oberhausen	53534
10005	Daquan	Sosa	F	1961-12-28	Holy See (Vatican City State)	Ap #727-5534 Mauris, Avenue	Eberswalde-Finow	5690ER
10006	Rebekah	Navarro	F	1974-02-01	Saudi Arabia	P.O. Box 698, 3686 Dui, Avenue	Wolvertem	2976
10007	Zane	Benson	M	2002-10-19	Mauritania	Ap #852-3480 Ornare Ave	Dufftown	
			F	1985-08-31	Isle of Man	568-2438 Sit Rd	Rathurst	

Patient Lab Results
Tables



Medical Images
Images

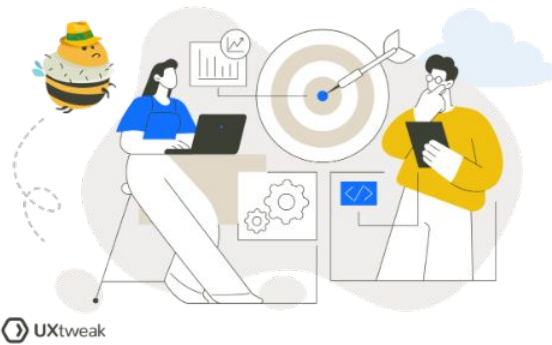
Doctor's Notes

Unstructured Text



Image source:
 [1] <https://www.mobius.md/blog/step-by-step-guide-to-taking-perfect-clinical-notes>
 [2] <https://www.echelon.health/five-reasons-radiology-is-important-in-healthcare/>
 [3] https://researchgate.net/publication/271388038_Anonymously_Analyzing_Clinical_Datasets?_tp=eyJib250ZXh0Ijp7ImZpcnN0UGFnZSI6Ij9kaXJlY3QlCjwYWdljoiX2RpcmVidCJ9Ij0

Time for Multimodal DBMSs - User Study



<i>Interviewee pseudonyms</i>	<i>Self-identified gender identity</i>	<i>Industry</i>	<i>Job role</i>	<i>Mode of work</i>
Alexandra	Transgender woman	Charity	Office worker	Full-time
Georgette	Transgender woman	Retail	Cashier	Part-time
Jackie	Transgender woman	Architecture/design	Office manager	Full-time
Julia	Transgender woman	Local government	Administrator	Full-time
Pauline	Transgender woman	Charity	Fundraising assistant	Part-time
Samantha	Transgender woman	Healthcare	Nurse	Full-time
Brad	Transgender man	Higher education	PhD student/research assistant	Part-time
Dave	Transgender man	Special needs education	Teaching support staff	Full-time
Jase	Transgender man	Construction	Builder	Full-time
Marc	Transgender man	Travel	Online coordinator	Full-time
Trey	Transgender man	Retail	Sales associate	Part-time
Andy	Genderqueer	Consulting	IT support assistant	Full-time
Kerr	Genderqueer	Higher education	Administrator	Full-time
Lane	Genderqueer	Fashion	Creative assistant	Full-time

Interviewees Info

Tables



Audio Recording

Audio

Question/Task Description

Unstructured Text

Image source:

[1] <https://www.uxtweak.com/ux-research/>

[2] https://www.researchgate.net/publication/279231766_Gender_identity_inclusion_in_the_workplace_broadening_diversity_management_research_and_practice_through_the_case_of_transgender_employees_in_the_UK

Advances in AI Make Multimodal DBMSs Feasible

1. Advances in Models

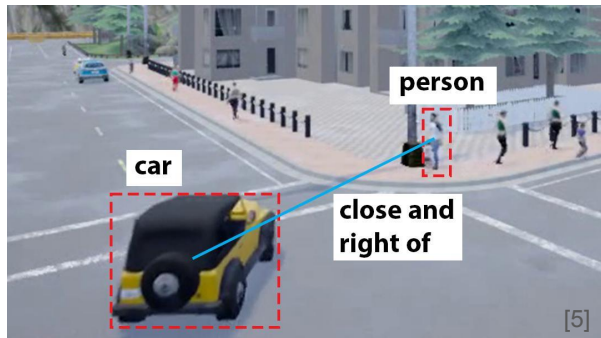
[GPT-series, Gemini, Qwen, OLMO, ...]

2. Advances in NL2SQL

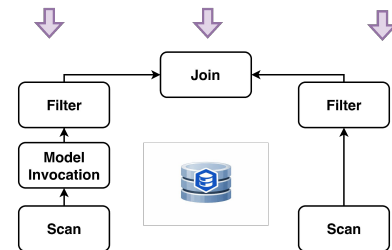
[BigQuery+Gemini, ...]

3. LLMs for Query Planning & Execution

[Caesura, DocETL, ...]



...and more Powerful LLMs



Agents can help with query execution

These advances enable us to build LLM-powered multimodal DBMSs

Evaluating Multimodal Queries with Existing Systems

Traditional DBMSs W/ AI Operators

[Google, 2026], [Kakkar et al., 2025 SIGMOD], ...

Not Easy for Non-Experts

Optimized Query Execution

No Human-AI Collaboration

Some Level Explainable & Traceable

DeepResearch Systems

[OpenAI, 2025], [Perplexity, 2025], ...

Easy for Non-Experts

No Query Optimization

Human-AI Collaboration

Some Level Explainable & Traceable

AI-Powered Data Systems

[Liu et al., 2025 CIDR], [Urban et al., 2025 CIDR], ...

Easy for Non-Experts

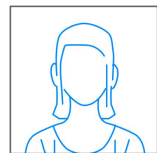
Optimized Query Execution

No Human-AI Collaboration

Some Level Explainable & Traceable

Can we build a multimodal DBMS that is friendly to non-experts, supports user interactions, and optimizes queries execution?

KathDB: An LLM-Powered Explainable Multimodal DBMS



NL
Query

Query
Result
Table

Query Parser

Query Optimizer

Execution Engine

KathDB

A Unified Relational View +
Raw Multimodal Data



Unstructured
Text



Image



Video



Audio



...

Key Features

Generates code at operator level & explores trade-offs between time, accuracy, and explainability.

Unifies multimodal data with a relational layer and tracks lineage.

Interacts with users at different stages during query execution.

KathDB Performs Code Gen. For Each Operator

NL Query

Find fun cat toys under \$20, and verify that their review videos contain actual visual evidence of cats playing with the toys.

Key Stages

Query Sketch Gen.



Logical Plan Gen.



Physical Plan Gen.

What KathDB Does

- Generates a sequence of **Actions** in natural language
- Groups **Actions** into logical **Steps**
- Builds a Logical Plan from **Steps**
- Generates code for each **Step** (e.g., Python, SQL, API Calls, etc.,)
- Profiles query plan with sampled input

KathDB Performs Code Gen. For Each Operator

Key Intermediate Representation

Goals

**Query
Sketch**

Action

Verb + Object
Short phrase

Action_0

Select toys that
are fun

Action_1

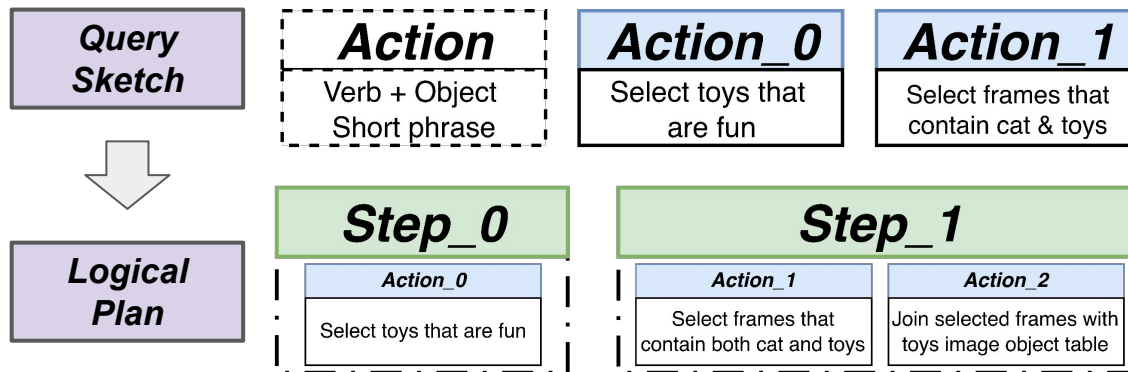
Select frames that
contain cat & toys

- Collaboratively refine query w/ users

KathDB Performs Code Gen. For Each Operator

Key Intermediate Representation

Goals



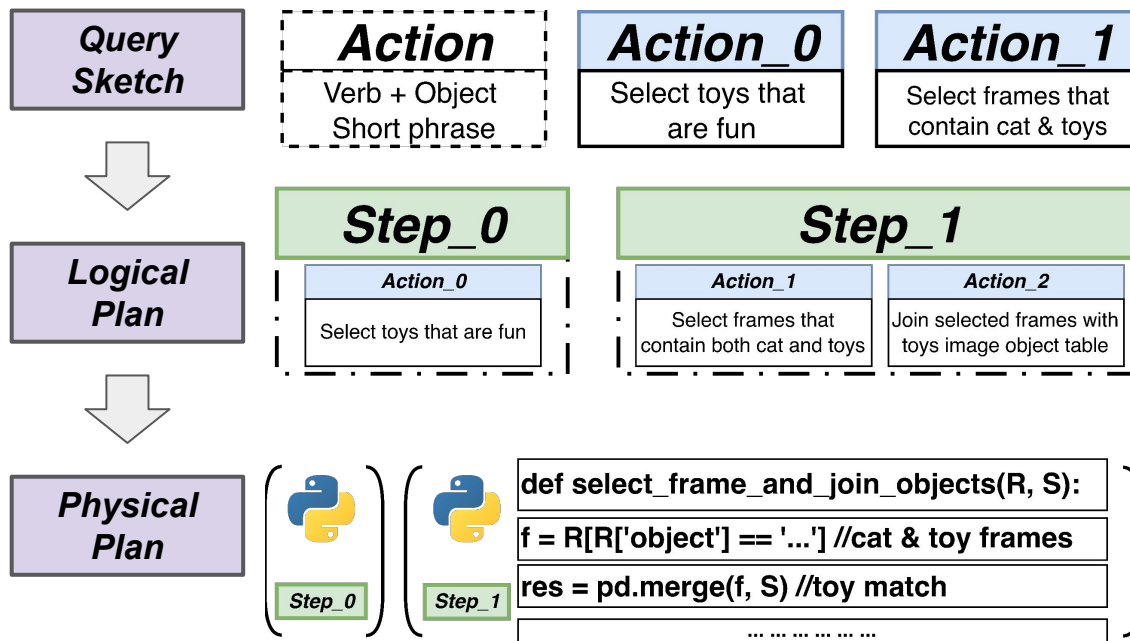
- Collaboratively refine query w/ users

- Identify "hard" **Actions**
- Group "easy" **Actions** into **Steps**

KathDB Performs Code Gen. For Each Operator

Key Intermediate Representation

Goals



- Collaboratively refine query w/ users

- Identify "hard" **Actions**
- Group "easy" **Actions** into **Steps**

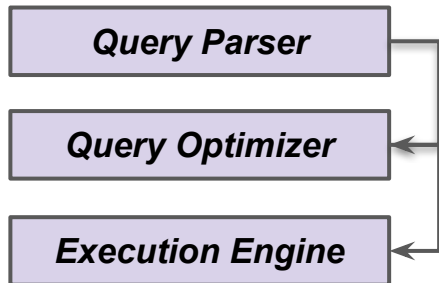
- Generate code at operator level
- Each function is closed over relations
- Optimize query plans

KathDB: An LLM-Powered Explainable Multimodal DBMS



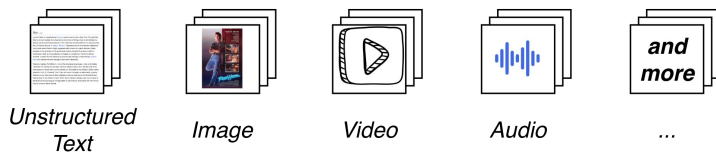
NL
Query

Query
Result
Table



KathDB

A Unified Relational View +
Raw Multimodal Data



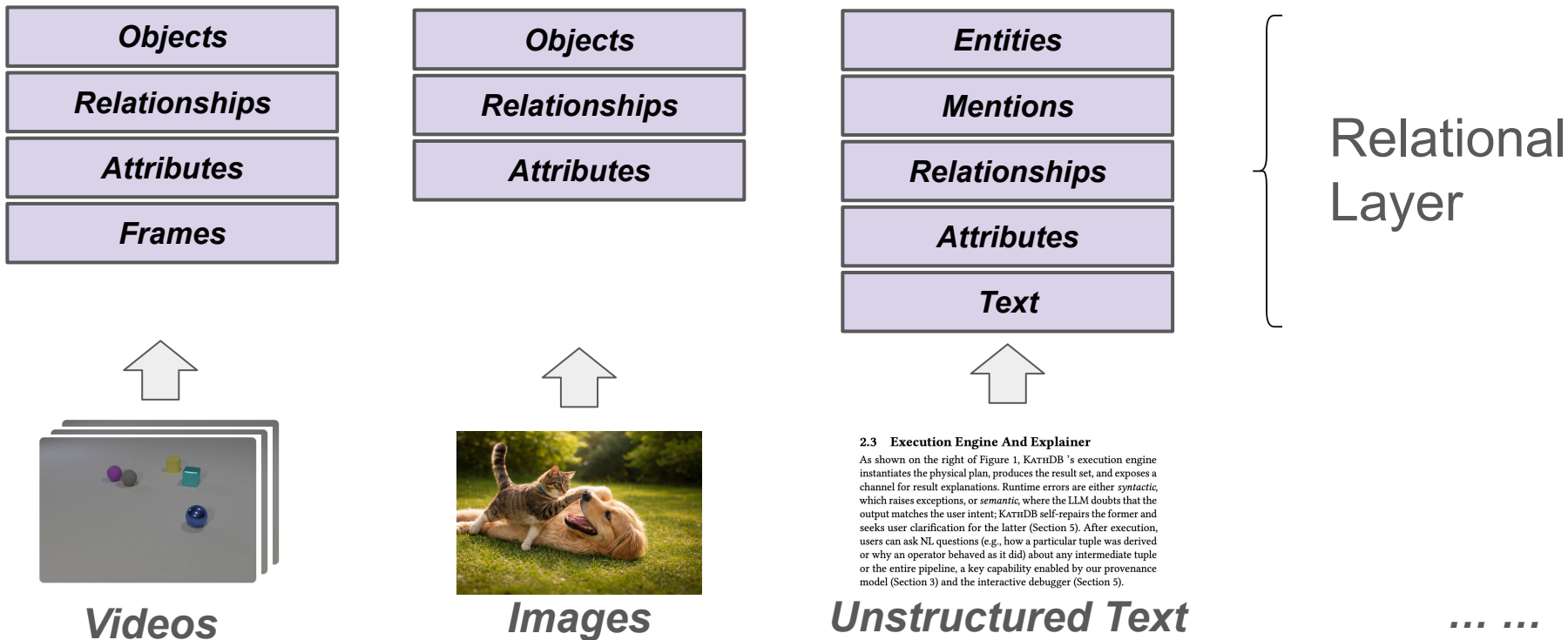
Key Features

Generates code at operator level & explores trade-offs between time, accuracy, and explainability.

Unifies multimodal data with a relational layer and tracks lineage.

Interacts with users at different stages during query execution.

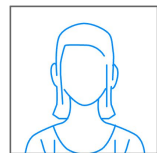
KathDB Provides Explainability via Lineage Tracking



2.3 Execution Engine And Explainer

As shown on the right of Figure 1, KATHDB's execution engine instantiates the physical plan, produces the result set, and exposes a channel for result explanations. Runtime errors are either *syntactic*, which raises exceptions, or *semantic*, where the LLM doubts that the output matches the user intent; KATHDB self-repairs the former and seeks user clarification for the latter (Section 5). After execution, users can ask NL questions (e.g., how a particular tuple was derived or why an operator behaved as it did) about any intermediate tuple or the entire pipeline, a key capability enabled by our provenance model (Section 3) and the interactive debugger (Section 5).

KathDB: An LLM-Powered Explainable Multimodal DBMS



NL
Query

Query
Result
Table

Query Parser

Query Optimizer

Execution Engine

KathDB

A Unified Relational View +
Raw Multimodal Data



Unstructured
Text



Image



Video



Audio



...

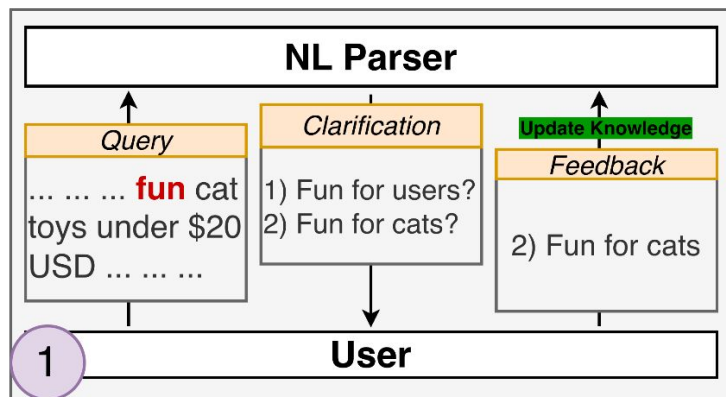
Key Features

Generates code at operator level & explores trade-offs between time, accuracy, and explainability.

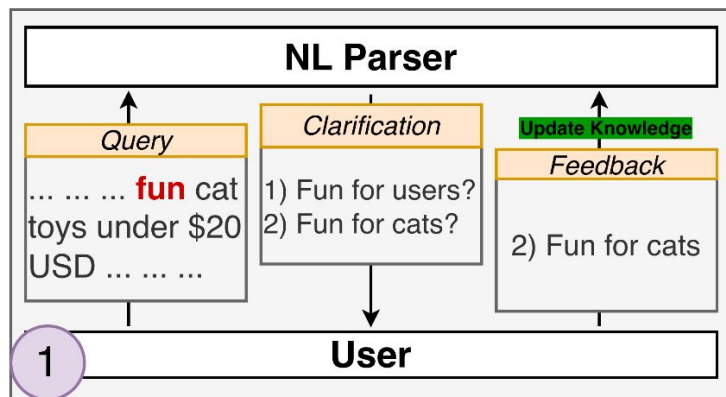
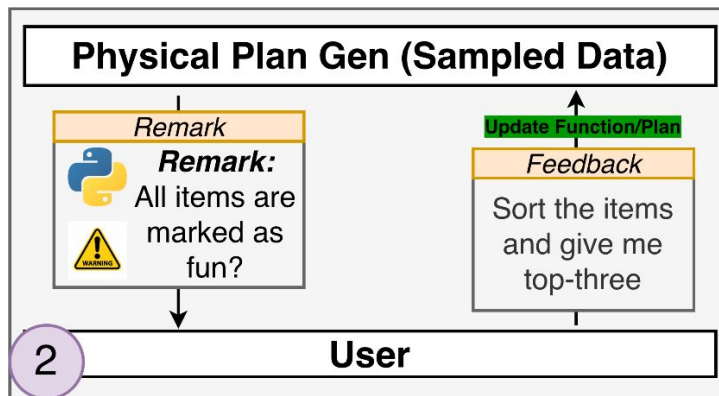
Unifies multimodal data with a relational layer and tracks lineage.

Interacts with users at different stages during query execution.

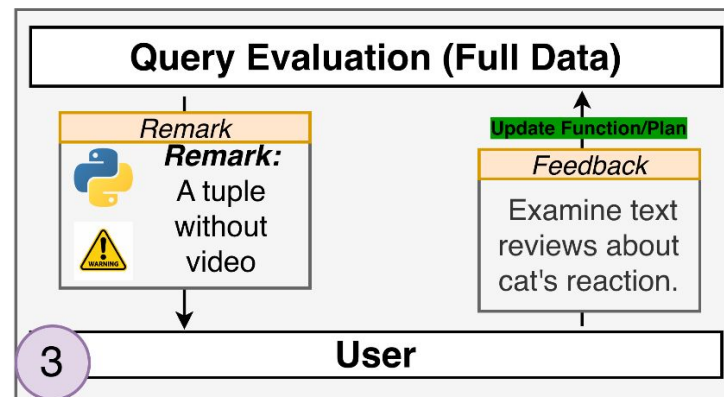
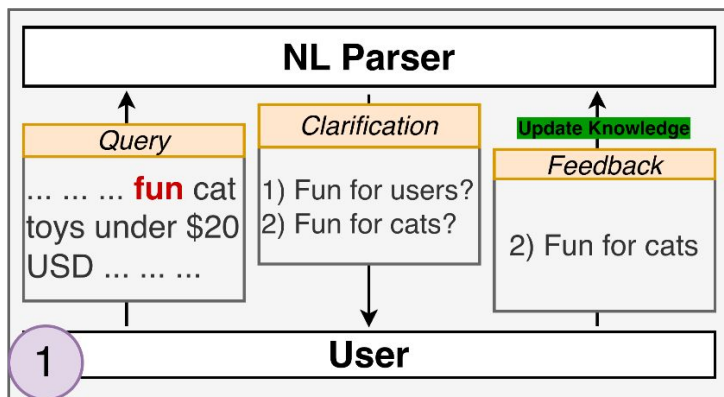
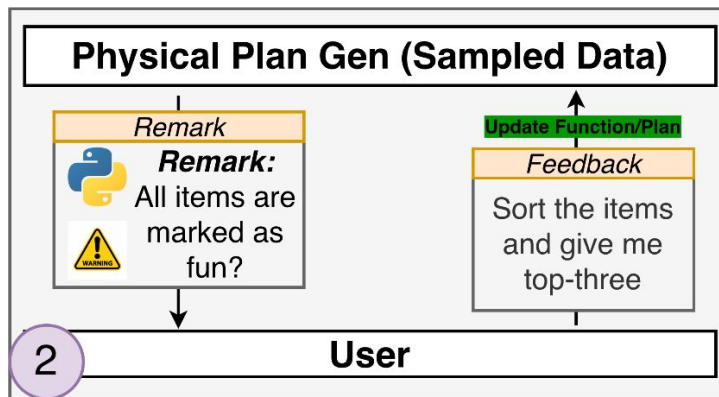
KathDB Interacts With Users At Different Stages



KathDB Interacts With Users At Different Stages



KathDB Interacts With Users At Different Stages

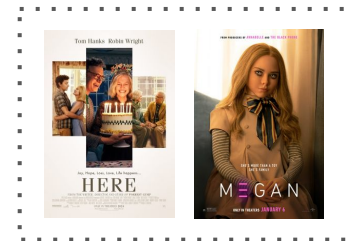


KathDB Preliminary Case Study

Sort the films in the table by how exciting they are, but require that the poster be described as 'boring'



Exciting Posters



Boring Posters

KathDB



<u>Name</u>	<u>Year</u>	<u>Excitement Score</u>
Guilty by Suspicion	1991	0.999...
Clean and Sober	1988	0.973...
...

KathDB Preliminary Case Study - Challenge 1/3

Query: Sort the films in the table by **how exciting they are**, but require that the poster be described as 'boring'

Film (title, year)

Plot (title, plot_summary) // raw text data

Poster (image_bytes) // raw image data



Works collectively with users to refine query.



Populates relational views over images and text data on-demand.

Proposes a quantitative metric that considers both objects extracted from the plot AND the recency of a movie to score how exciting a film is to a user.

KathDB Preliminary Case Study - Challenge 2/3

No Explicit JOIN Key

Film (title, year)

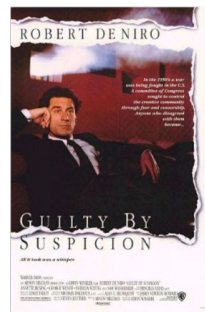
Plot (title, plot_summary) // raw text data

Poster (image_bytes) // raw image data



Resolves unexpected runtime issue.

Proposes to download and install an OCR model to extract text from posters and JOIN with film titles.



Guilty By Suspicion

KathDB Preliminary Case Study - Challenge 3/3

<u>Name</u>	<u>Year</u>	<u>Excitement Score</u>
Guilty by Suspicion	1991	0.999...
Clean and Sober	1988	0.973...
...



How is this derived?



Explains how a tuple is derived based on lineage and function implementations.

excitement_score is a weighted sum of **0.7** * fuzzy_keywords_counts + **0.3** * recency.



Updates the logic and re-runs a partial query.

KathDB: An LLM-Powered Explainable Multimodal DBMS

KathDB answers NL queries over large multimodal data

Key Features

Interacts with users at different stages during query execution.

Unifies multimodal data with a relational layer and tracks lineage.

Generates code at operator level & explores trade-offs between query time, accuracy, and explainability.

Project Page



<https://db.cs.washington.edu/projects/kathdb/>

Reference

[1]: Chung, Yeounoh, et al. "Is Long Context All You Need? Leveraging LLM's Extended Context for NL2SQL." arXiv preprint arXiv:2501.12372 (2025).

[2]: Liu, Xinyu, et al. "A Survey of Text-to-SQL in the Era of LLMs: Where are we, and where are we going?." IEEE Transactions on Knowledge and Data Engineering (2025).

[3] Li, Yujia, et al. "Competition-level code generation with alphacode." Science 378.6624 (2022): 1092-1097.

[4] Chen, Xinyun, et al. "Teaching large language models to self-debug." arXiv preprint arXiv:2304.05128 (2023).

[5] Zhang, Enhao, et al. "Self-Enhancing Video Data Management System for Compositional Events with Large Language Models." Proceedings of the ACM on Management of Data 3.3 (2025): 1-29.

[6] Arora, Simran, et al. "Language Models Enable Simple Systems for Generating Structured Views of Heterogeneous Data Lakes." Proceedings of the VLDB Endowment 17.2 (2023): 92-105.

Reference II

- [7] Liu, Chunwei, et al. "Palimpzest: Optimizing ai-powered analytics with declarative query processing." Proceedings of the Conference on Innovative Database Research (CIDR). 2025.
- [8] Urban, Matthias, and Carsten Binnig. "Demonstrating CAESURA: language models as multi-modal query planners." Companion of the 2024 International Conference on Management of Data. 2024.
- [9] Patel, Liana, et al. "Semantic Operators and Their Optimization: Enabling LLM-Based Data Processing with Accuracy Guarantees in LOTUS." Proceedings of the VLDB Endowment 18.11 (2025): 4171-4184.
- [10] Wei, Jason, et al. "Chain-of-thought prompting elicits reasoning in large language models." Advances in neural information processing systems 35 (2022): 24824-24837.
- [11] Google BigQuery "<https://cloud.google.com/bigquery?hl=en>"
- [12] Kakkar, Gaurav Tarlok, et al. "Hydro: Adaptive Query Processing of ML Queries." arXiv preprint arXiv:2403.14902 (2024).




Reference III

[13]: OpenAI. “Introducing deep research.” OpenAI, 2 Feb. 2025,
<https://openai.com/index/introducing-deep-research/>.

[14]: Perplexity. “Introducing Perplexity Deep Research” Perplexity, 14 Feb. 2025,
<https://www.perplexity.ai/hub/blog/introducing-perplexity-deep-research>

Backup Slides

Evaluating Multimodal Query with existing systems

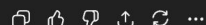
Existing systems	Easy for non-expert	Traceable Derivation	Multimodal	HumanAI Interaction	Query Optimization
Multimodal Models   ...	✓	✗	✓	✓	✗
DeepResearch Systems  ...	✓	✓ ✗	✓	✓ ✗	✗

Compile a research report on how data management research has changed in past 3 years

To make sure I compile a report that fits your needs, could you clarify a few things:

1. Are you looking for trends in academic research, industry practices, or both?
2. Do you have specific areas of data management in mind (e.g., data governance, data lakes, metadata management, real-time analytics, etc.)?
3. What level of technical depth are you looking for—general overview, detailed technical analysis, or executive summary style?
4. Should the focus be global or limited to a specific region or industry?

Once I have this information, I can begin the research task for you.



1. Academic research 2. Database Management Systems 3. General Overview 4. Only from top-tier conferences like SIGMOD, VLDB, and ICDE

Trends in Database Management Research from Top Conferences (2023–2025)

I'm identifying key trends and emerging themes in Database Management Systems from SIGMOD, VLDB, and ICDE conferences (2023–2025), including LLMs and DB integration, learned indexes, and cloud-native architectures.

Searched for SIGMOD 2023 keynote topics

I'm selecting the article titled "SIGMOD 2023 highlights - Megagon Labs" to delve into the main topics of the conference.

Read megagonlabs.medium.com

The trends at SIGMOD 2023 show a strong focus on traditional database systems, with system papers, machine learning applications in specific data types, and emerging areas like representation learning



Trends in Database Research (2023–2025)

Introduction

Over the past three years, database management systems (DBMS) research technological waves while strengthening core foundations. Top-tier conferences VLDB, and IEEE ICDE – have showcased shifts in focus from traditional optimization processing toward themes like machine learning integration, cloud-native architectures, and AI models. This report surveys key trends and innovations from 2023 through 2025, highlighting research themes, rising and fading topics, and influential papers from the field.

Enduring Core Topics and System Innovations

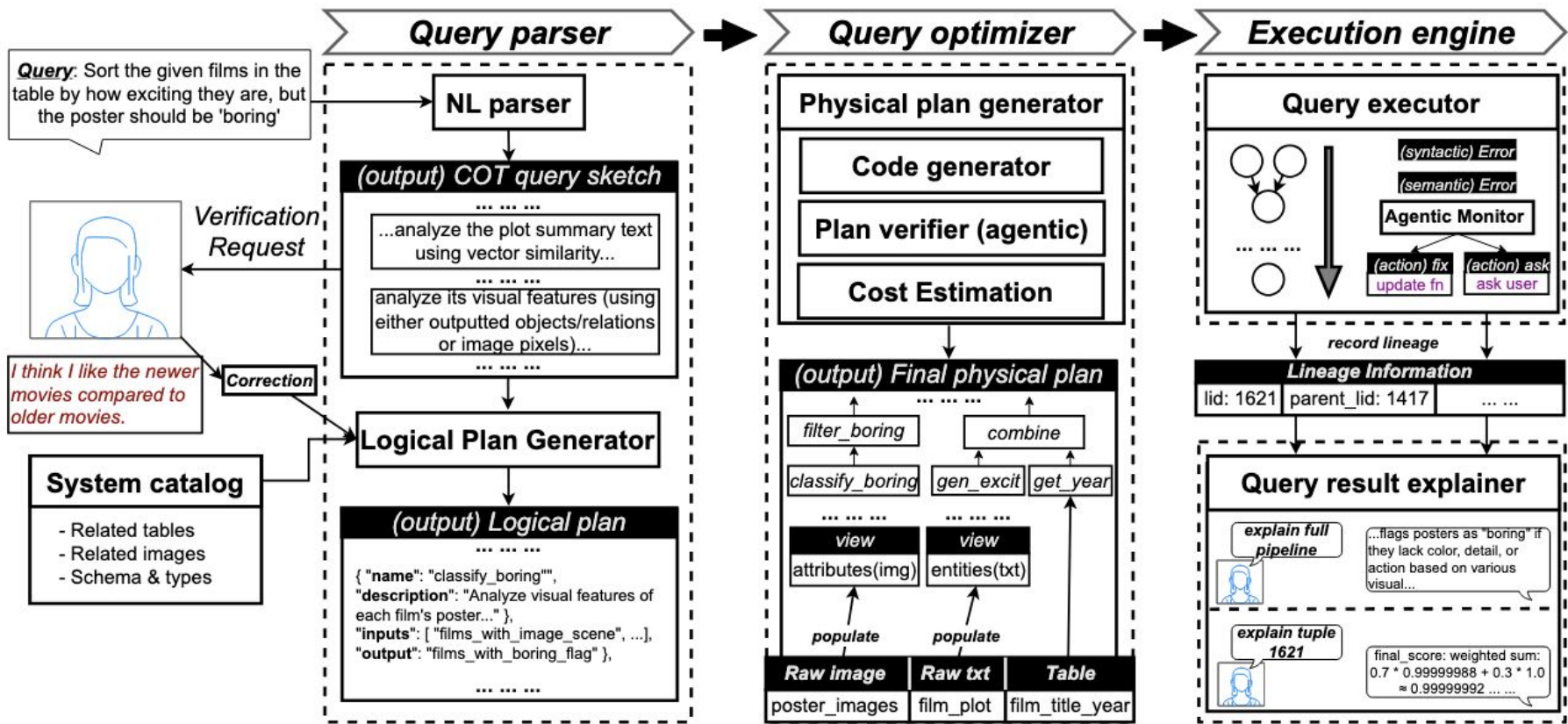
Classic Database Problems Remain Vital: Fundamental DBMS topics have received significant attention. SIGMOD 2023, for example, featured many sessions on "query processing and optimization," indicating that traditional areas remain popular. Research on query optimization and execution are ongoing; a 2025 SIGMOD best paper on "pessimistic cardinality estimation technique using statistical norms, reflecting query cost estimation." Another 2025 honoree, DBCony, achieved

Evaluating Multimodal Query with existing systems

Existing systems	Easy for non-expert	Traceable Derivation	Multimodal	HumanAI Interaction	Query Optimization
State-of-the-art AI-powered Data Systems  ...	✓	✓ ✗	✓ ✗	✗	✓

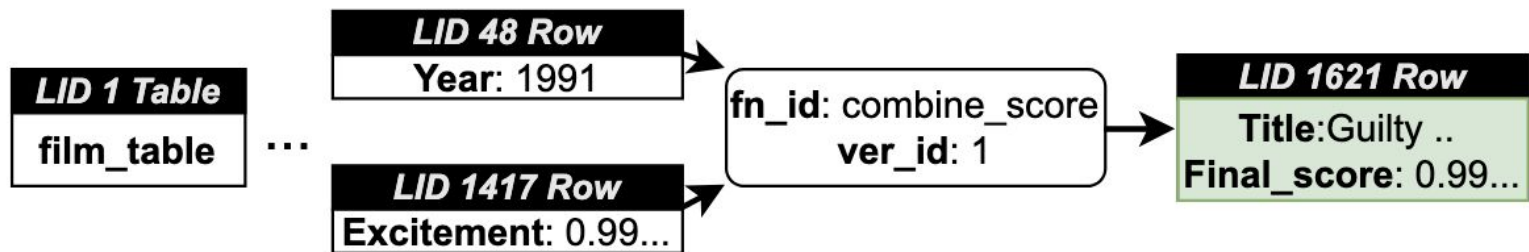
```
# define columns for datasets
text_animal_cols = [
    {"name": "animal", "type": str, "desc": "The type of animal mentioned in the text"},
    {"name": "color", "type": str, "desc": "The color of the animal mentioned in the text"},
]
image_animal_cols = [
    {"name": "animal", "type": str, "desc": "The type of animal in the image"},
    {"name": "color", "type": str, "desc": "The color of the animal in the image"},
]
```

```
def run_text_image_join():
    """Build a plan that joins a dataset with text to a dataset with images"""
    ds1 = pz.TextFileDataset(id="animals1", path="join-data/animal-texts/").sem_map(text_animal_cols)
    ds2 = pz.ImageFileDataset(id="animals2", path="join-data/animal-images/").sem_map(image_animal_cols)
    ds3 = ds1.sem_join(ds2, condition="both animals are canines with the same color")
    config = pz.QueryProcessorConfig(
        policy=pz.MaxQuality(),
        execution_strategy="parallel",
        join_parallelism=64,
    )
```



```
[ ... .. // Other function signatures
{ "name": "classify_boring",
  "description": "Analyze visual features of each film's poster..." },
  "inputs": [ "films_with_image_scene", ... ],
  "output": "films_with_boring_flag" },
... .. // Other function signatures ]
```

Figure 3: Function signature generated by the logical plan generator.



Explain the pipeline?

Coarse

... .. 7: Flags posters as "boring" if they lack color, detail, or action based on various visual features.

... ..

10: Ranks these films by their combined score, highlighting the most notable among those with boring posters



Explain tuple 1621?

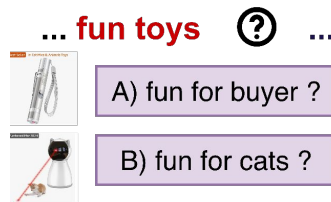
Fine-grain

... .. Plot contains many indirect or direct keywords related to "threat", "death", "attack", "kill", etc. \n - Score is extremely high (≈ 0.99999988).
****recency_score****: Assigned as 1.0 (likely the most recent or very recent film).
****final_score****: \n - Weighted sum: $0.7 * 0.99999988 + 0.3 * 1.0 \approx 0.99999992$

KathDB: An LLM-Powered Explainable Multimodal DBMS

Iterative **Collaboration** with Users

Parsing Interaction



Plan Gen Interaction



Query Eval Interaction

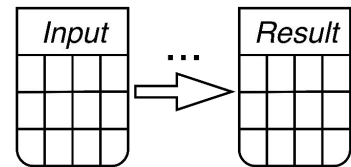


Traceable Query Result **Explanation**

Relational View over Multimodal Data

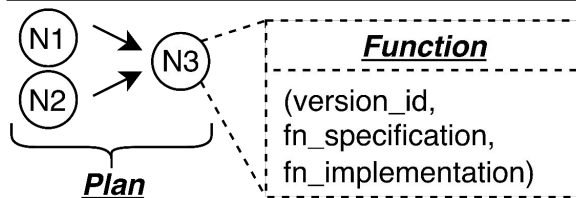


Lineage Tracking



Operator Level **Code Generation**
Multi-Objective Plan Optimization

Function as Operator



Query Plan Generation

Step-1: Code Gen.

Step-2: Plan Profiling

Step-3: Plan Optimization