

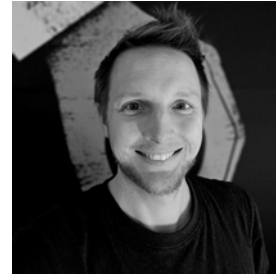
# Building a Secure ,AI'-Stack: Protecting Data and Intellectual Property

**eic**  
european  
identity  
and cloud  
conference 2025



**WEDA CON**

# Building a Secure ,AI'-Stack: Protecting Data and Intellectual Property



Felix Witt  
[@wedacon.bsky.social](https://bsky.social/wedacon)



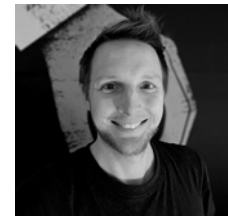
Thorsten H. Niebuhr  
[@idmpath.bsky.social](https://bsky.social/idmpath)



# WedaCon: Who we are



Thorsten H. Niebuhr  
Founder / CEO  
IAM Topics since 1997



Felix Witt  
Lead Developer / CTO  
IAM Topics since 2009



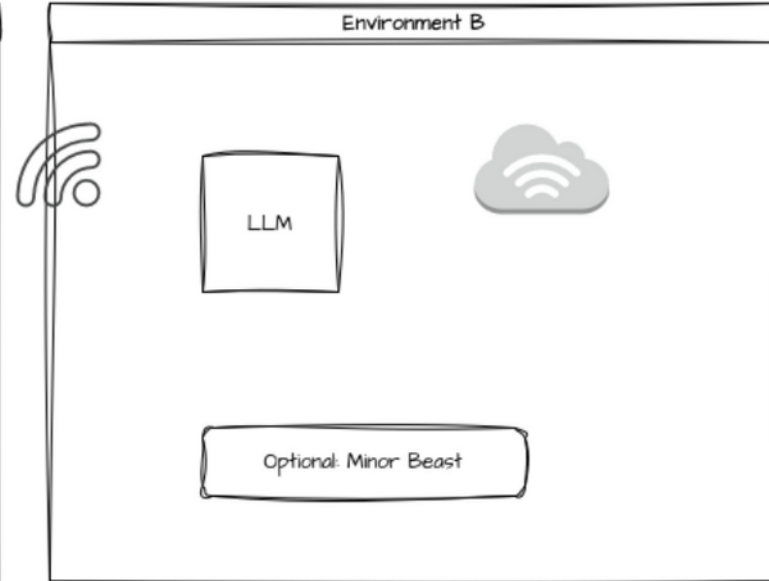
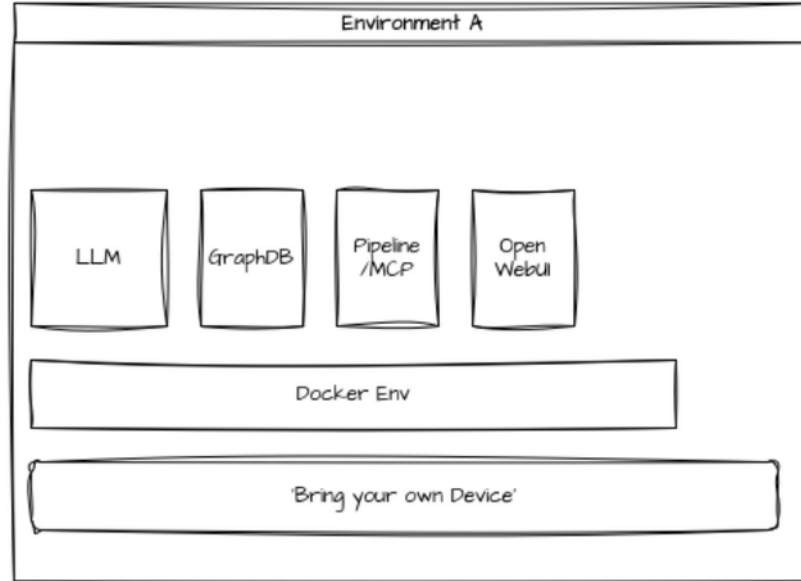
# Some Disclaimers

- ❑ What this session is about
  - ❑ de-mystify ,AI'
  - ❑ An Intro on operating ,local' LLM's
- ❑ What it is NOT
  - ❑ An academic event
  - ❑ A Coding Session
  - ❑ A full scale ,AI' Training Session





# The Environment



```
neo4j-mcp-ollama$ ./start-containers.sh
Extracting Neo4j data from archive...
[+] Running 1/1
  ✔ ollama Pulled
[+] Running 7/7
  ✔ Network neo4j-mcp-ollama_default Created
  ✔ Volume "neo4j-mcp-ollama_open-webui_demo" Created
  ✔ Volume "neo4j-mcp-ollama_ollama" Created
  ✔ Container ollama Started
  ✔ Container neo4j Started
  ✔ Container neo4j-mcp-server Started
  ✔ Container open-webui-demo Started
pulling manifest
pulling ff82381e2bea: 7% ██████████
```

WLAN: WedaCon-Workshop

Pwd: Robotwar

<http://192.168.210.65>

(hint: use as proxy 192.168.210.65:8888)

Remote:

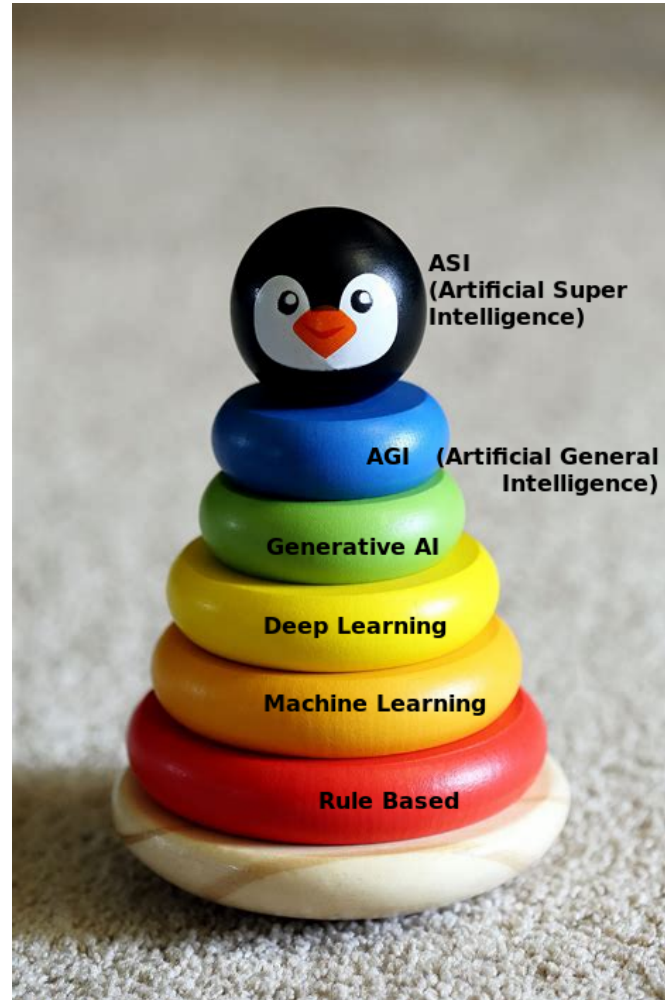
<https://blog.wedacon.net/>

(anyway, what is this AI thingy?)





# Evolution





# Rule Based Systems: Giants

**Heron of Alexandria**  
Ἡρώων



17th-century German depiction of Heron

**Leonardo da Vinci**



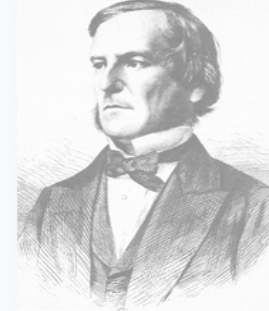
This portrait attributed to [Francesco Melzi](#), c. 1515-1518, is the only certain contemporary depiction of Leonardo.<sup>[1][2]</sup>

**Jacques de Vaucanson**



Portrait by [Joseph Boze](#), c. 1784

**George Boole**  
FRS



Portrait of Boole, from  
*The Illustrated London News*, 21 January  
1865

**Gottfried Wilhelm Leibniz**



Bildnis des Philosophen Leibniz (1695),  
by [Christoph Francke](#)

**David Hilbert**



Hilbert in 1912

**Kurt Gödel**



Gödel c. 1926

**Alan Turing**  
OBE FRS



Turing in 1951

**Konrad Zuse**



Konrad Zuse in 1992

**John McCarthy**



McCarthy at a conference in 2006





# Rule Based Systems: Giants

DARTMOUTH COLLEGE  
Department of Mathematics & Astronomy  
HANOVER · NEW HAMPSHIRE

March,  
1956

Mr. Ray Solomonoff  
Technical Research Group  
17 Union Square West  
New York, New York

Dear Ray:

You are one of the people we should like to invite to the "Summer Research Project on Artificial Intelligence."

Terms: \$1,200 - \$900 of which will probably count as a fellowship and be tax free, plus traveling expenses.

Dates: June 18 to Aug. 17

Place: Hanover, N. H.  
(a cool place).

Can we count on you?

Best regards,  
*John*  
John McCarthy

JMcC:MA

*J. McCarthy } for all  
A. Newsky } 2 months  
John Holland  
R. Solomonoff  
Julian Bigelow }*

*Shannon } some  
Rochester } of these  
Selfridge } two part  
McCulloch } of time.  
Newell  
Simon  
McKay  
Et al*

**John McCarthy**



McCarthy at a conference in 2006



# Machine Learning

## Machine Learning

### Supervised Learning

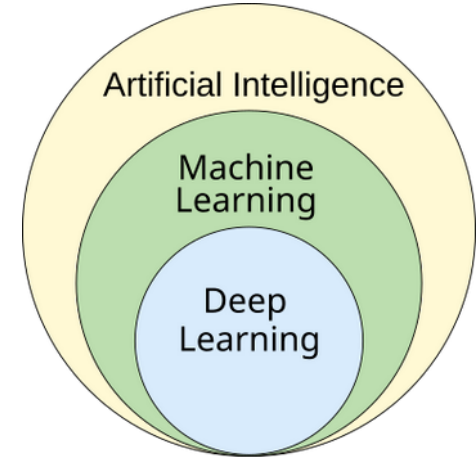
- Task driven

### Unsupervised Learning

- Data driven

### Reinforcement Learning

- Learn from Mistake

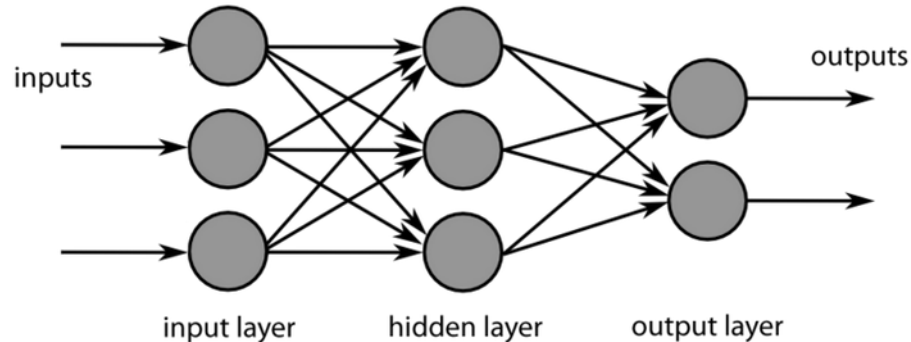
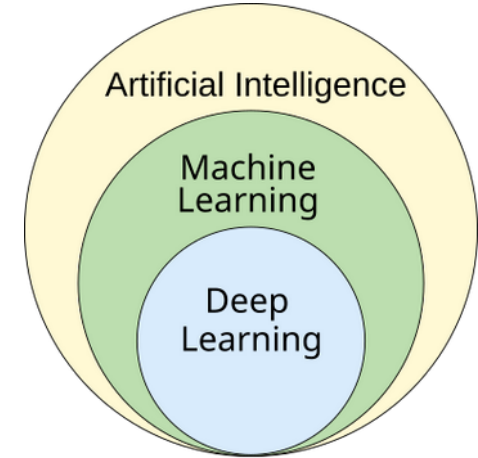




# ,Deep' Learning

## □ Deep Learning

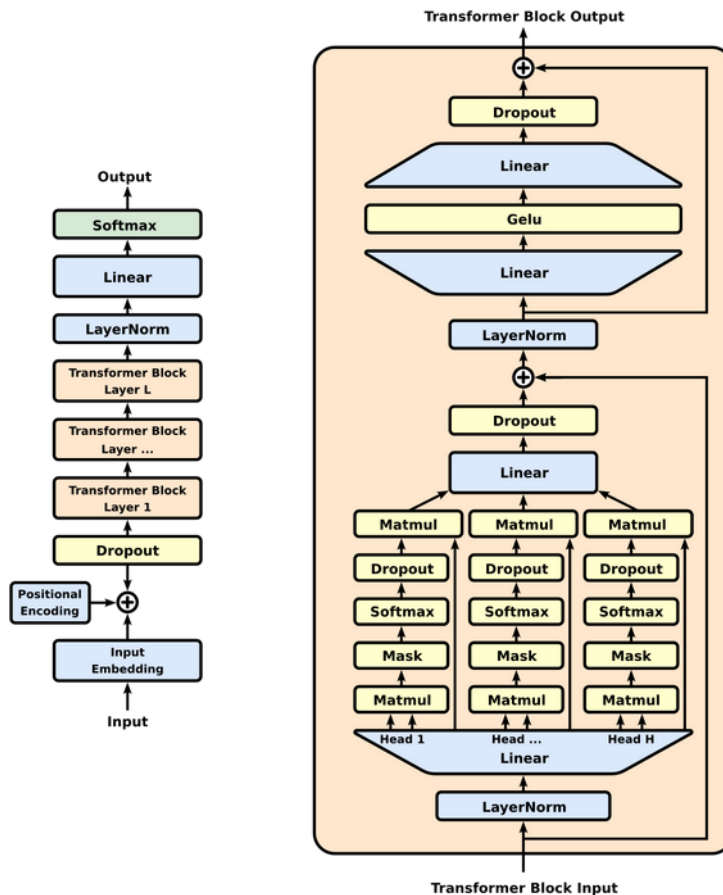
- Extraction
- Categorization
- Multi-Layered
- Neural Network





# Generative AI

- Generative Pre-Trained Transformer
- Deep-Learning
- Natural Language Processing
- 'Attention is all you need'

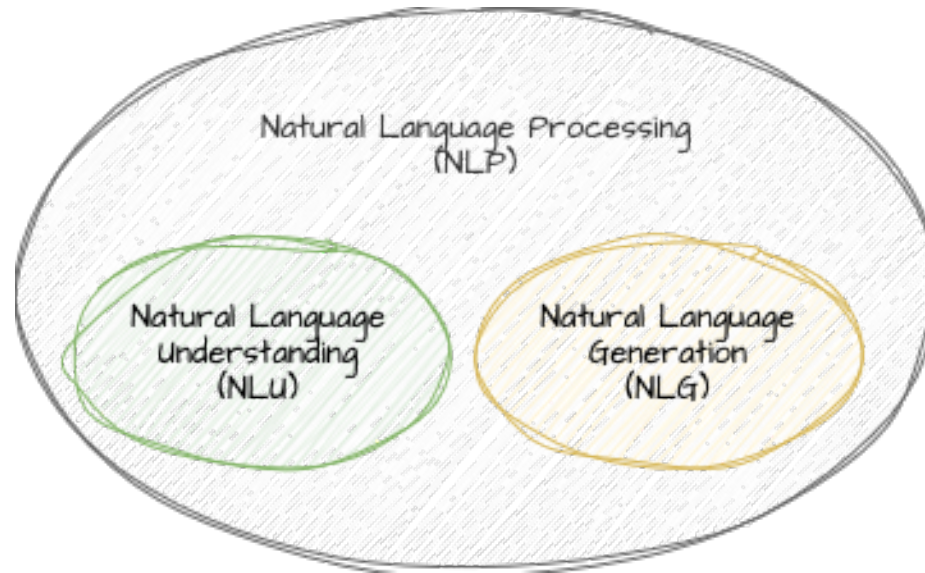






# Language Model

## ❑ Natural Language Processing (NLP)





# ,Pre-Trained'

The training compute of notable AI models is doubling roughly every five months

+

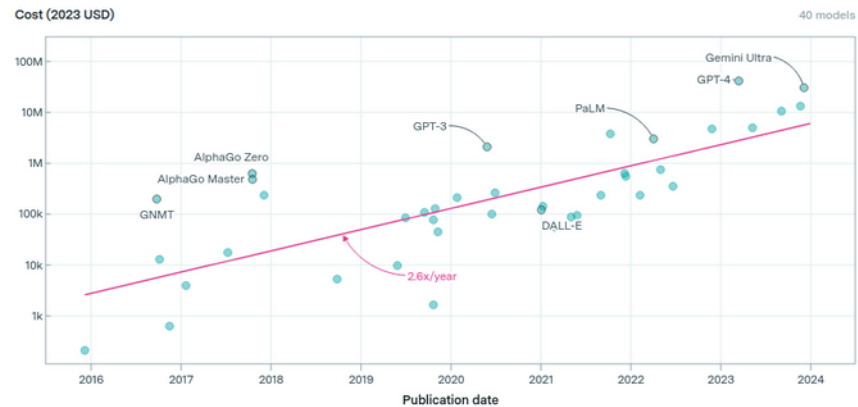
Training compute costs are doubling every nine months for the largest AI models

-

The cost of training large-scale ML models is growing at a rate of 2.6x per year. The most advanced models now cost hundreds of millions of dollars, with expenses measured by amortizing cluster costs over the training period. About half of this spending is on GPUs, with the remainder on other hardware and energy.

[Learn more →](#)

Amortized hardware and energy cost to train large-scale AI models over time



The size of datasets used to train language models doubles approximately every seven months

+

The power required to train frontier AI models is doubling annually

+

Training compute growth is driven by larger clusters, longer training, and better hardware

+

<https://epoch.ai/data/notable-ai-models>



# (some Fixed) Model Parameters

- ❑ Model Architecture
- ❑ Number of Parameters
- ❑ (Max) Context Window
- ❑ Quantization
- ❑ License
- ❑ Tools Capability
- ❑ Reasoning Capability
- ❑ Visual / Audio Capability



# Performance vs. Accuracy







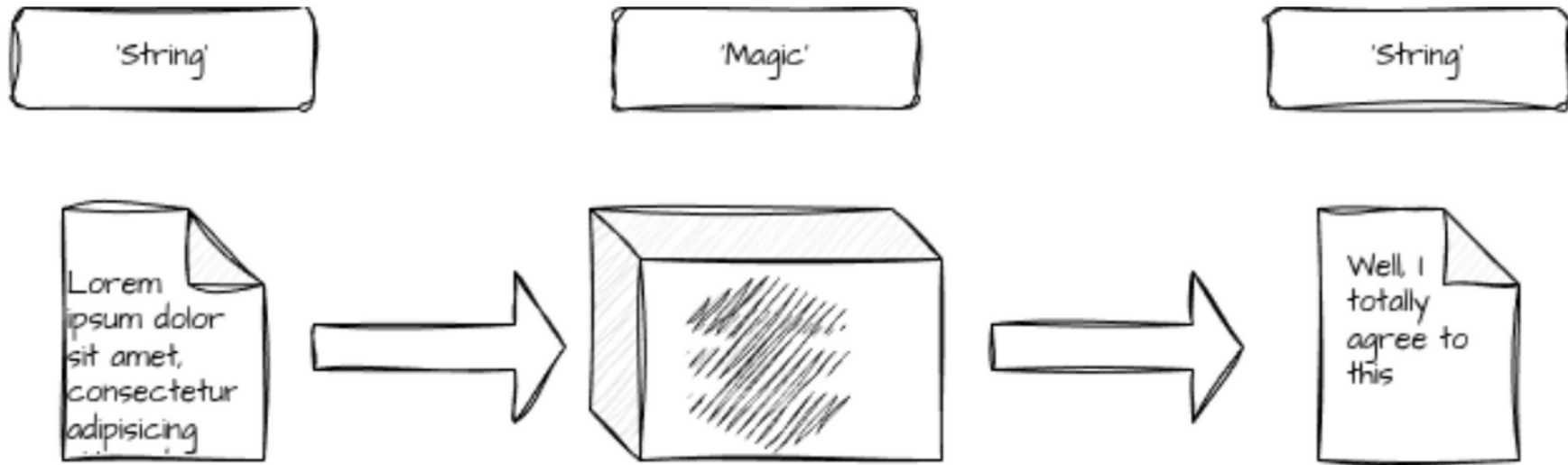
# Performance vs. Accuracy



generate a cartoon with the 21 year old Tween consuming piles of food

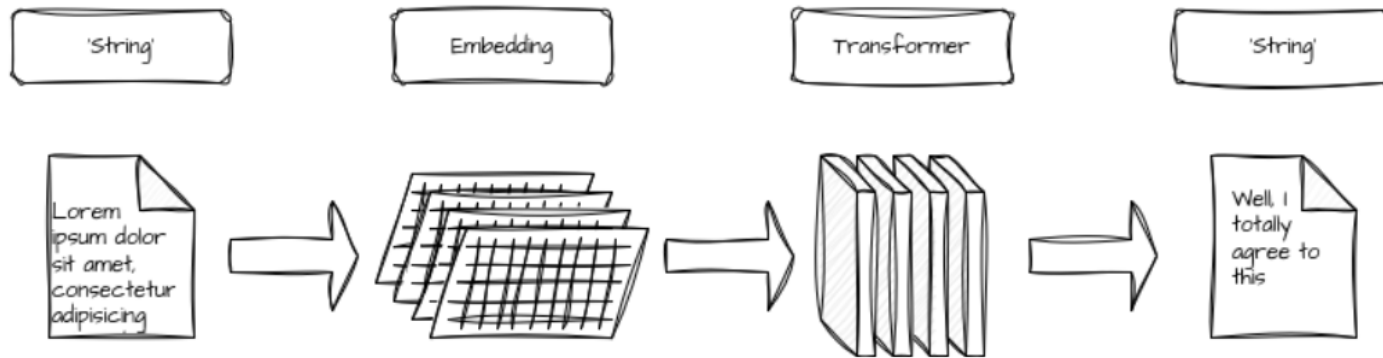


# a kind of magic?



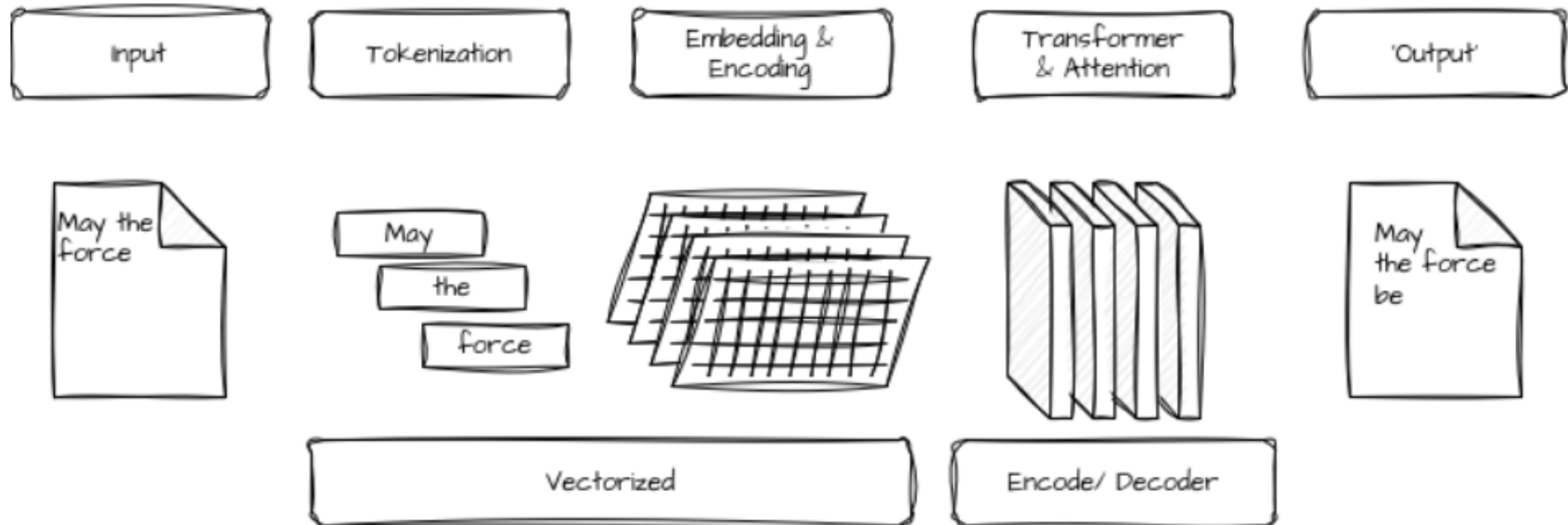


# Embedding and Transformer





# Vectors







# Temperature and probability

## ☐ Temperature

- ☐ Randomness of next ,token'

## ☐ Top-p (nucleus)

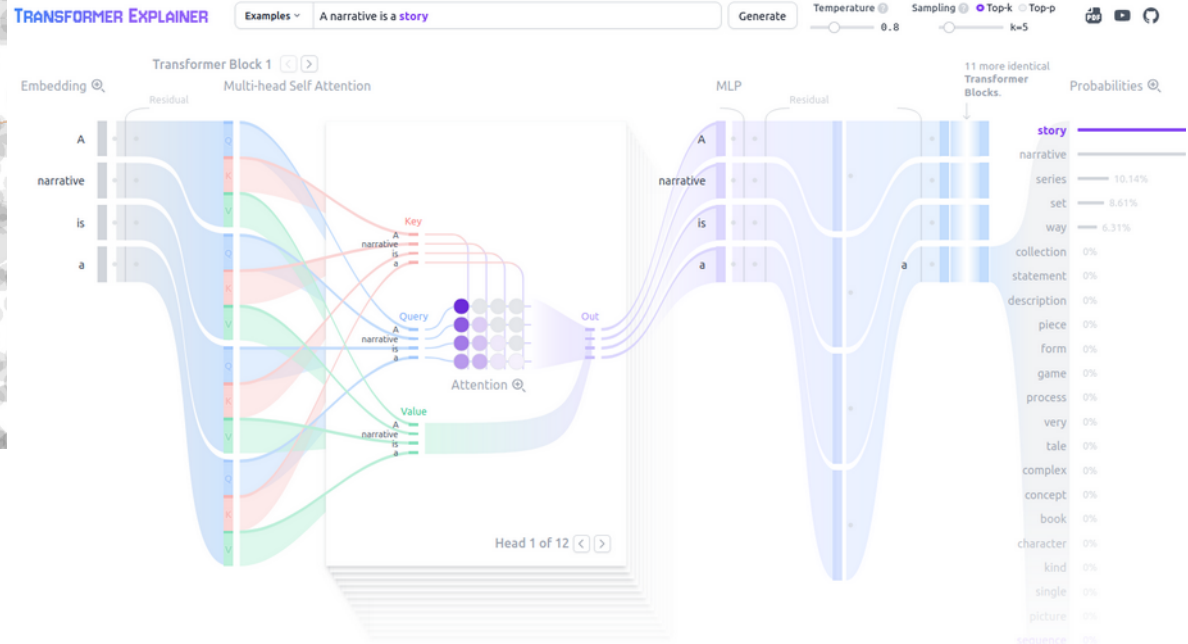
- ☐ amount of tokens taken into consideration

## ☐ Top-k (propability)

- ☐ select from k most likely next tokens

## ☐ Demo

- ☐ <https://tiktokenizer.vercel.app/>
- ☐ <https://projector.tensorflow.org/>
- ☐ <https://poloclub.github.io/transformer-explainer/>





Is that really artificial ...  
,intelligence'?





# Reasoning: How do **we** know

- ❑ Philosophical
  - ❑ Rationalism
  - ❑ Empirism
- ❑ Cognitive Aspects
  - ❑ Logical Thinking
  - ❑ Problem Solving
  - ❑ Decision Making
- ❑ Intuition and Emotion





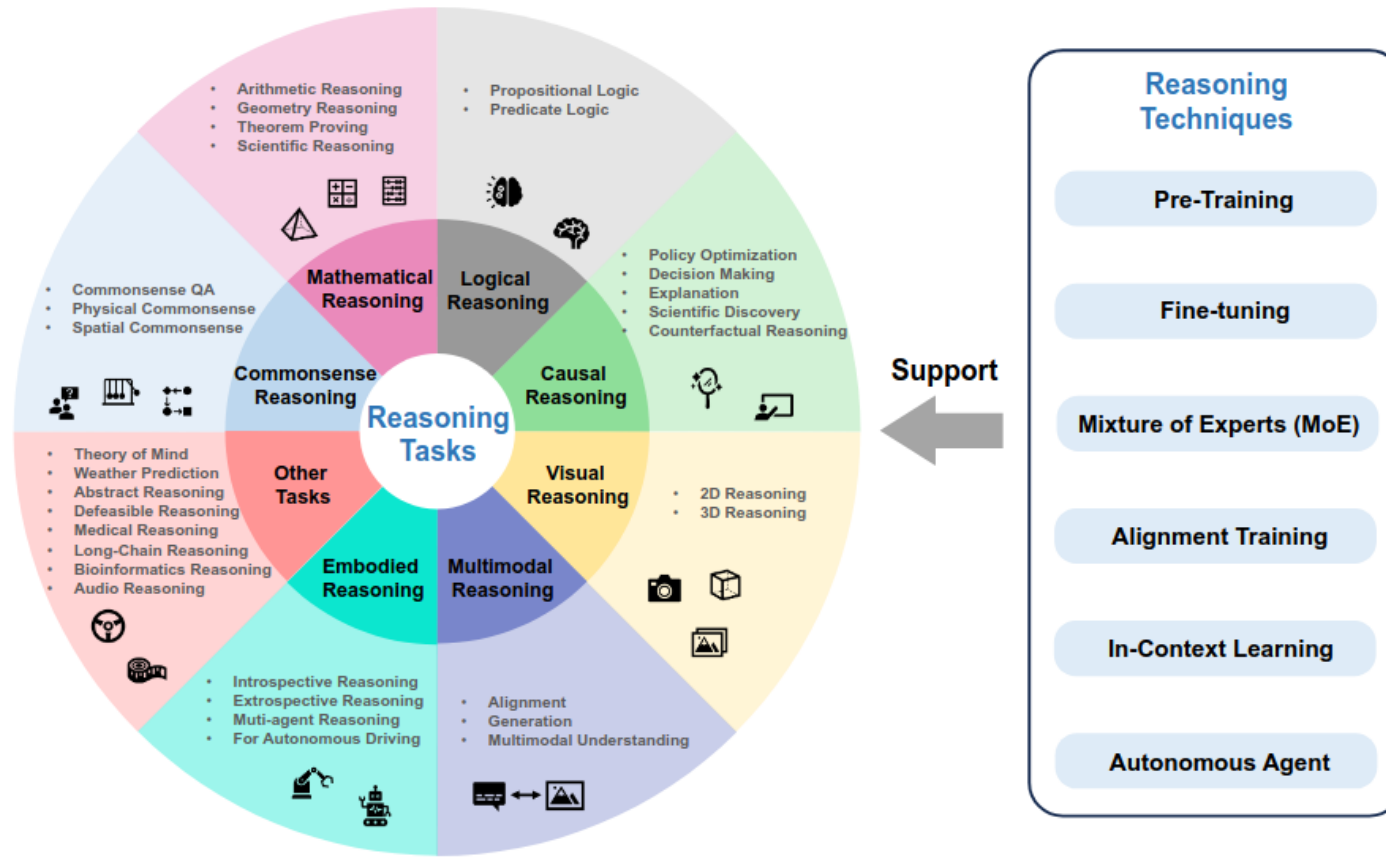
# Reasoning: How does it know

- ❑ Capabilities
  - ❑ Pattern Recognition
  - ❑ ,Contextual' Understanding
  - ❑ Logical Inference
  - ❑ Probabilistic Reasoning
- ❑ Constraints
  - ❑ Lack of Common Sense
  - ❑ Data Dependencies
  - ❑ Limited Abstract Reasoning
  - ❑ Absence of casual understanding
  - ❑ Hallucinations





# Reasoning Foundation Model



Sun et al. (2023) : <https://arxiv.org/pdf/2312.11562>





# The Knowledge Gap

- ❑ ,Common' LLMs (and chatbots) are trained for ,general' knowledge, representing our young friends from before.
- ❑ Specialized LLMs are available (check out huggingface)
  - ❑ german term: ,Fachidiot'
- ❑ Typical Gaps
  - ❑ Data Quality and Bias
  - ❑ Hallucinations
  - ❑ Outdated Knowledge
  - ❑ Contextual Understanding
  - ❑ Lack of Common Sense Reasoning



# Enhancing ,Reasoning'

## ☐ ,Liquid' LLM

- ☐ Build / Training your own model

- ☐ Fine-Tuning a given Model

## ☐ ,Frozen' LLM

- ☐ Human-in-the-loop

- ☐ ,Prompting'

- ☐ ,Hybrid' Models

- ☐ RAG

- ☐ Pipelines

- ☐ Tools

- ☐ Agents

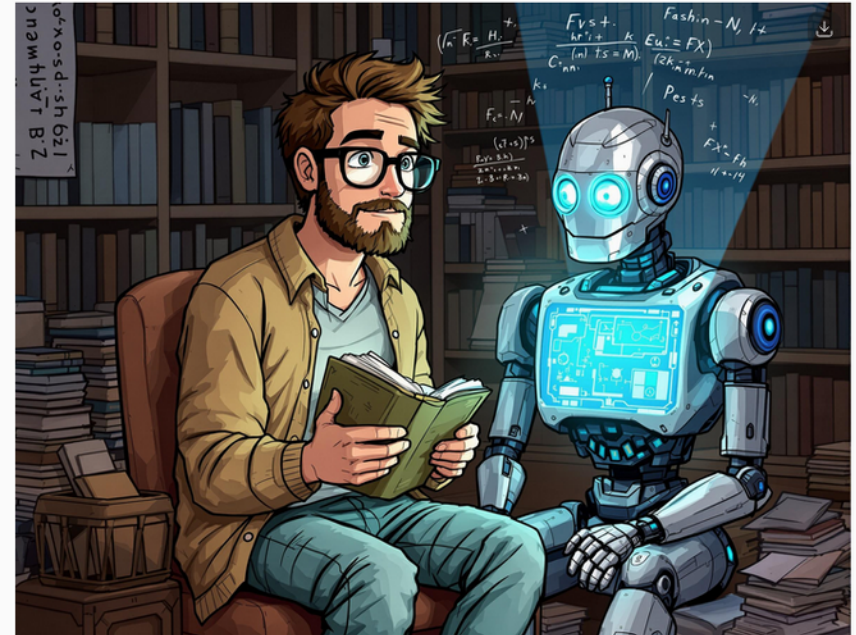
generate an image with a human and a robot sitting in the library studying



Worked for 10.9s

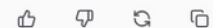


Here is a cartoon-style depiction of a human and a robot sitting in the library studying:



This image captures a moment of focused learning and collaboration between the human and the robot in a scholarly setting.

16:12



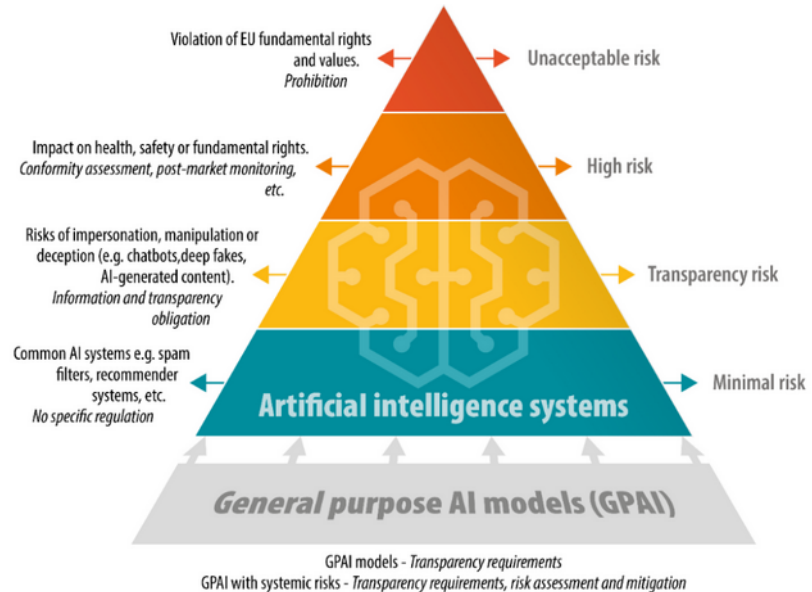


# Expert & Domain Knowledge

- ❑ Common Knowledge
  - ❑ ,General' information and facts
- ❑ Domain Knowledge
  - ❑ specifics to a particular field or industry
- ❑ Expert Knowledge
  - ❑ ,Deep' Understanding in a specific area
- ❑ Institutional Knowledge
  - ❑ The ,internal' Know-How

# The Case for it

- ❑ Data privacy and protection
- ❑ Compliance and Governance
- ❑ Intellectual Property





# (some) Regulations

## ☐ EU AI Act

### ☐ Actors

#### ☐ Provider

#### ☐ Deployer

### ☐ AI Competency

## ☐ US

### ☐ Consumer Privacy

### ☐ Healthcare Utilization

### ☐ ...

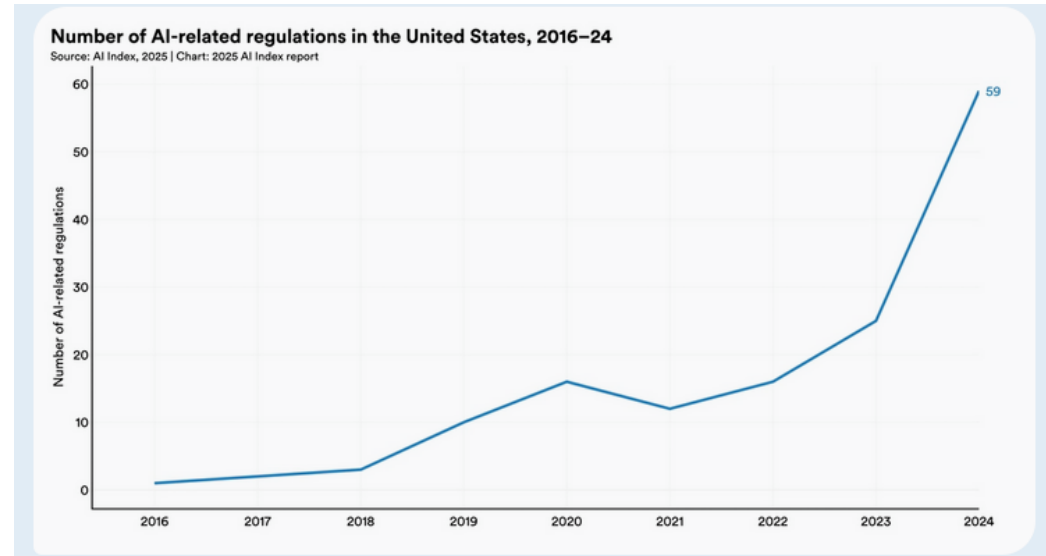
## ☐ China

### ☐ fragmented, eg

#### ☐ Algorithms

#### ☐ Deepfake

#### ☐ Generative AI

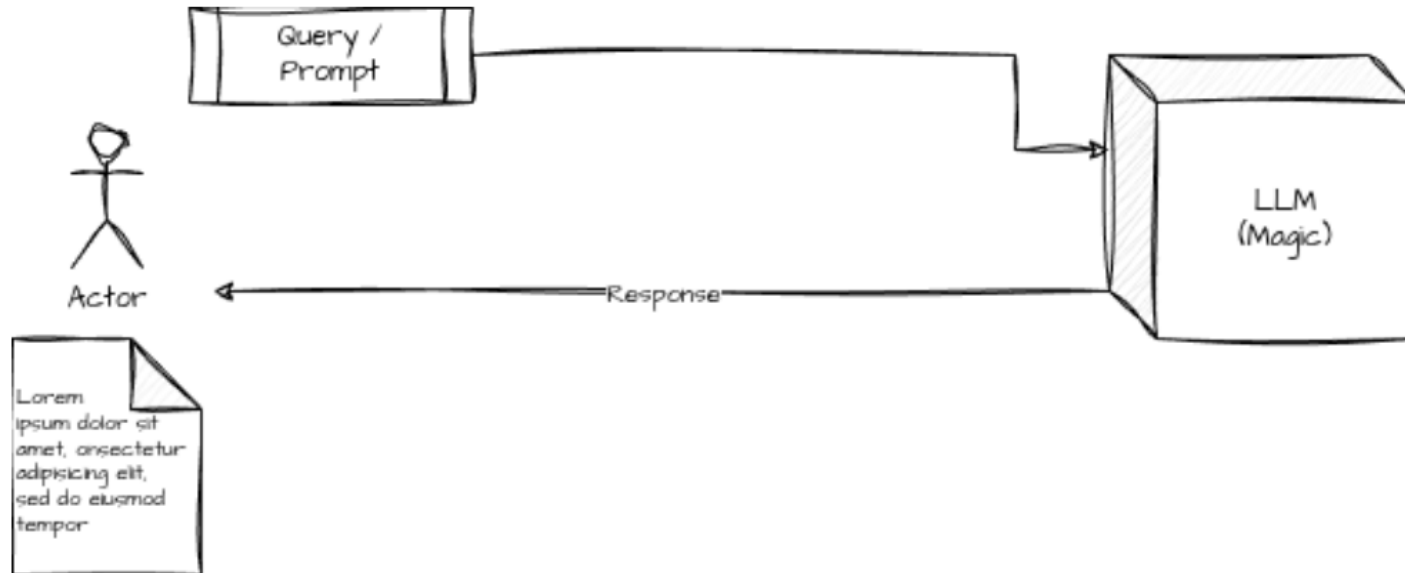


<https://hai.stanford.edu/ai-index/2025-ai-index-report>



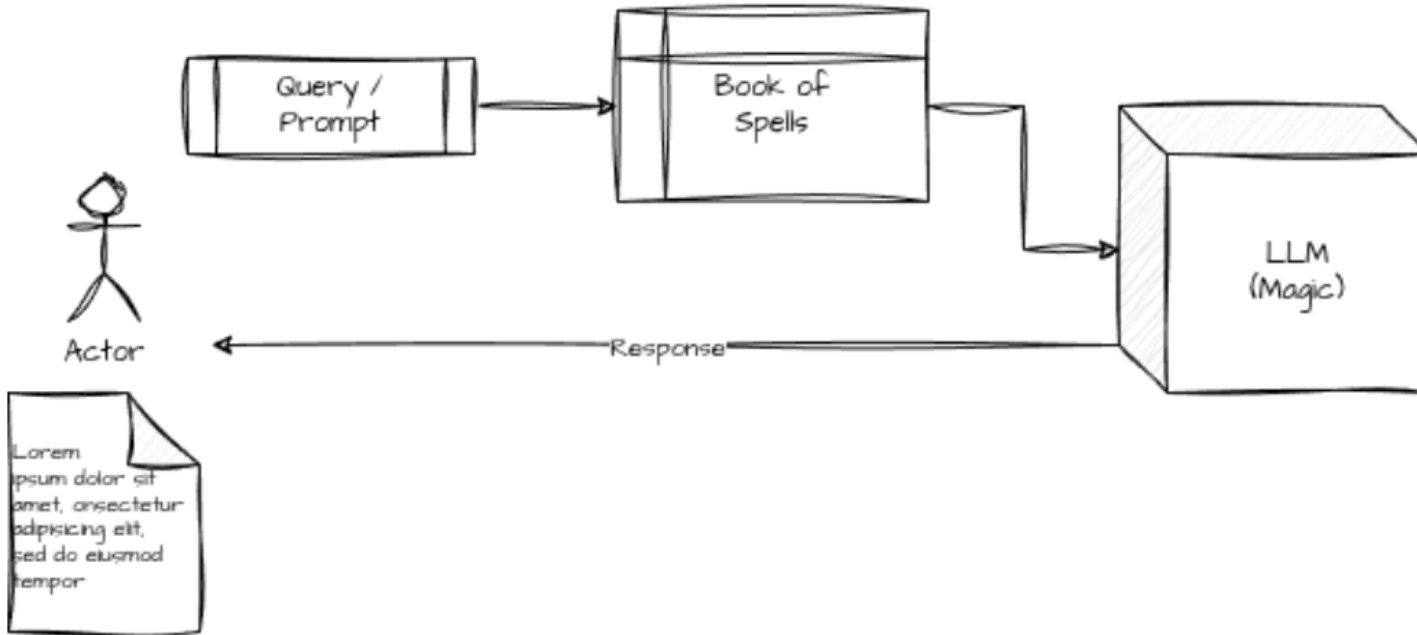


# Prompting





# Prompting, enhanced



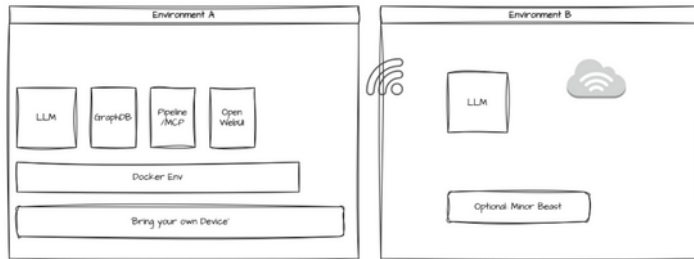


# Prompting Techniques

- ❑ Zero-Shot and Few-Shot
- ❑ Chain of Thoughts
- ❑ Meta Prompting
- ❑ Prompt Chaining
- ❑ Retrieval Augmented Generation
  - ❑ Tools Integration
  - ❑ Graph Prompting



# The Environment



WLAN: WedaCon-Workshop  
Pwd: Robotwar  
<http://192.168.210.65/download>  
(hint: use IP as proxy)

Remote:  
<https://blog.wedacon.net/>

- ▼ Docker
  - ▼ Docker-compose: neo4j-mcp-ollama
    - ▼ neo4j
      - neo4j
    - ▼ neo4j-mcp-server
      - neo4j-mcp-server
    - ▼ ollama
      - ollama
    - ▼ open-webui
      - open-webui-demo healthy
    - neo4j-mcp-ollama\_default

ollama 6325321a | ollama/ollama:latest | Docker-compose: neo4j-mcp-ollama

```
root@6325321a3b2e:/# ollama list
```

NAME	ID	SIZE	MODIFIED
mistral:latest	f974a74358d6	4.1 GB	5 minutes ago

```
neo4j-mcp-ollama$ docker ps
```

CONTAINER ID	IMAGE	COMMAND
04d0d068589e	ghcr.io/open-webui/open-webui:main	"bash start.sh"
e4781194d1e8	neo4j-mcp-ollama-neo4j-mcp-server	"uvx mcpo --host 0.0..."
2765c8ba8fae	neo4j	"tini -g -- /startup..."
6325321a3b2e	ollama/ollama:latest	"/bin/ollama serve"



# ollama



**Get up and running with large  
language models.**

Run Llama 3.3, DeepSeek-R1, Qwen 3, Mistral,  
Gemma 3, and other models, locally.

Download ↓

Available for macOS,  
Linux, and Windows

<https://ollama.com/>















# Open WebUI

## Key Features of Open WebUI ★

And many more remarkable features including...



-  Pipelines Support
-  User Experience
-  Conversations
-  Model Management
-  Collaboration
-  History & Archive
-  Audio, Voice, & Accessibility
-  Code Execution
-  Integration & Security
-  Administration

<https://docs.openwebui.com/>



# Zero-Shot and Few-Shot

## Live Demo

```
jetson@ubuntu:~$ ollama list | grep ':1.'
```

Model Name	Size	Version	Size	Time
tinylama:1.1b	2644915ede35	637 MB	9 months ago	

```
jetson@ubuntu:~$ ollama run tinylama:1.1b  
>>> Send a message (/? for help)
```






granite3.2:2b ▾ +

Collapse Run Copy

1 Classify the text into neutral, negative or positive.  
2 Text: I think the vacation is okay.  
3 Sentiment:

granite3.2:2b

Neutral



# Chain-of-Thoughts

## Live Demo

```
jetson@ubuntu:~$ ollama list | grep ':1.'  
tinylama:1.1b          2644915ede35    637 MB    9 months ago  
jetson@ubuntu:~$ ollama run tinylama:1.1b  
>>> Send a message (/? for help)
```

Sally is a girl and has 3 brothers. Each brother has 2 sisters. how many sisters does sally have?



# Meta Prompting

## Live Demo

```
jetson@ubuntu:~$ ollama list | grep ':1.'
tinylama:1.1b          2644915ede35    637 MB    9 months ago
jetson@ubuntu:~$ ollama run tinylama:1.1b
>>> Send a message (/? for help)
```

- Structure
- Syntax
- Examples



# Prompt Chaining

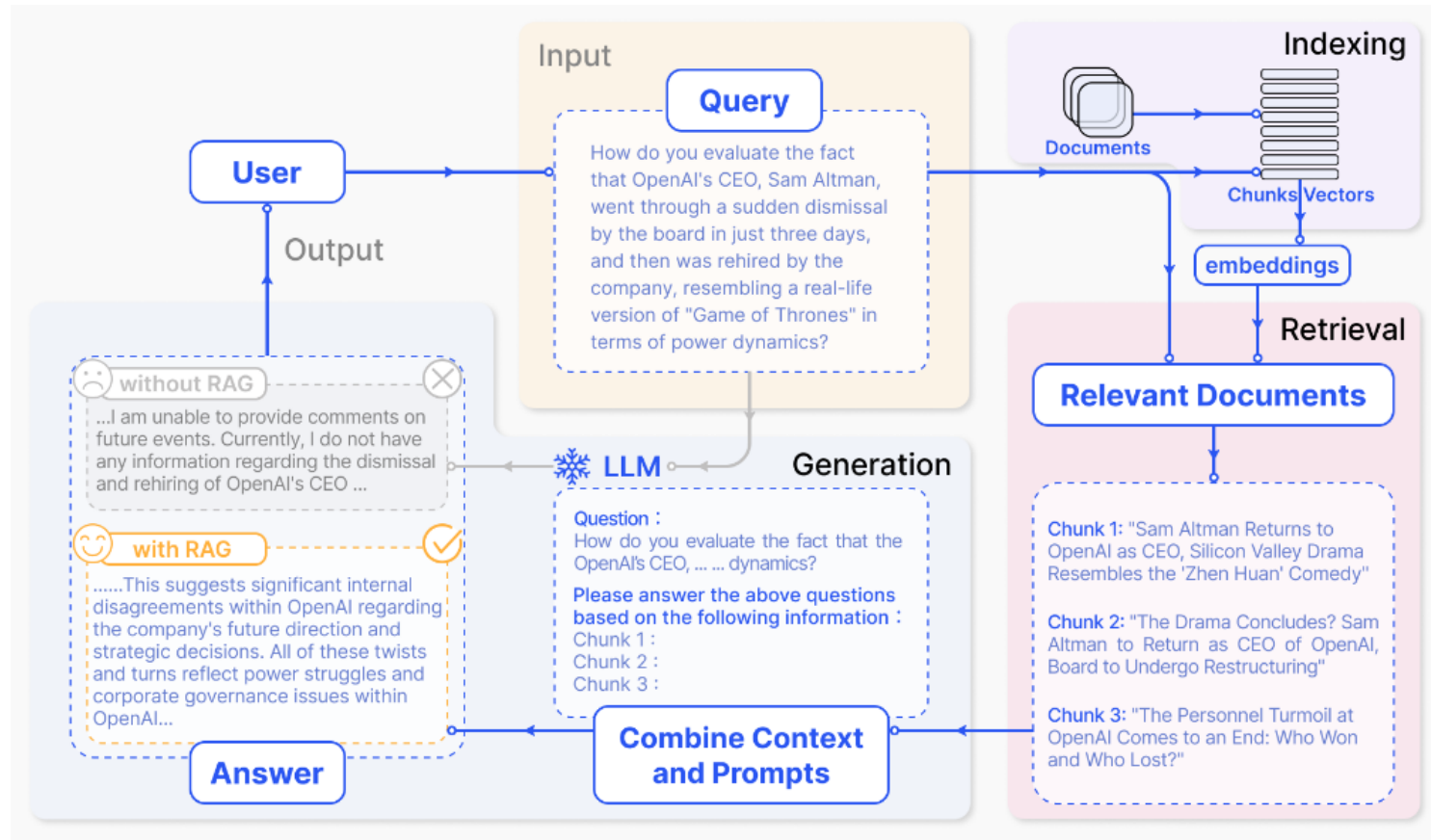
## Live Demo

```
jetson@ubuntu:~$ ollama list | grep ':1.'
tinylama:1.1b          2644915ede35    637 MB    9 months ago
jetson@ubuntu:~$ ollama run tinylama:1.1b
>>> Send a message (/? for help)
```





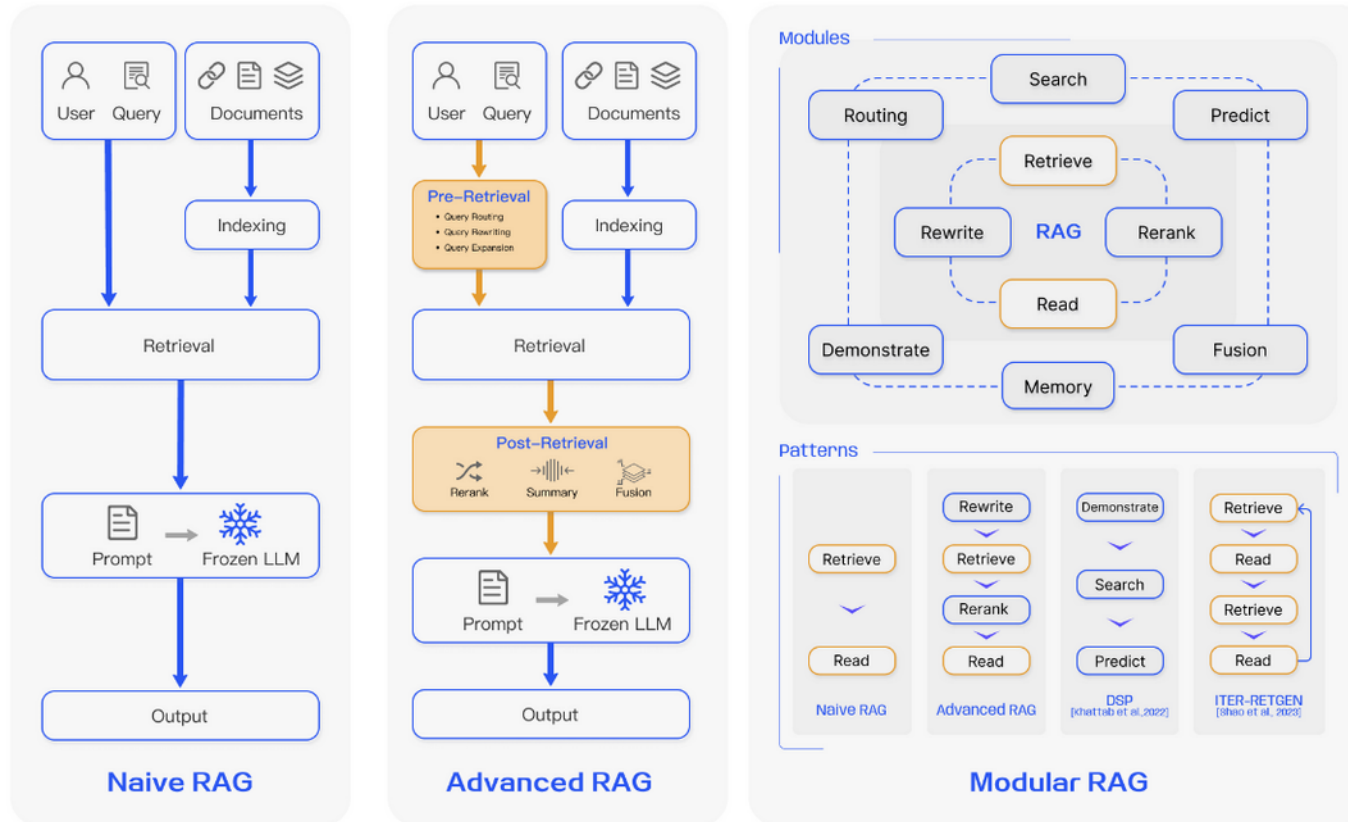
# RAG (Retrieval Augmented Generation)



Gao et al. (2023) : <https://arxiv.org/abs/2312.10997>



# RAG Types



Gao et al. (2023) : <https://arxiv.org/abs/2312.10997>



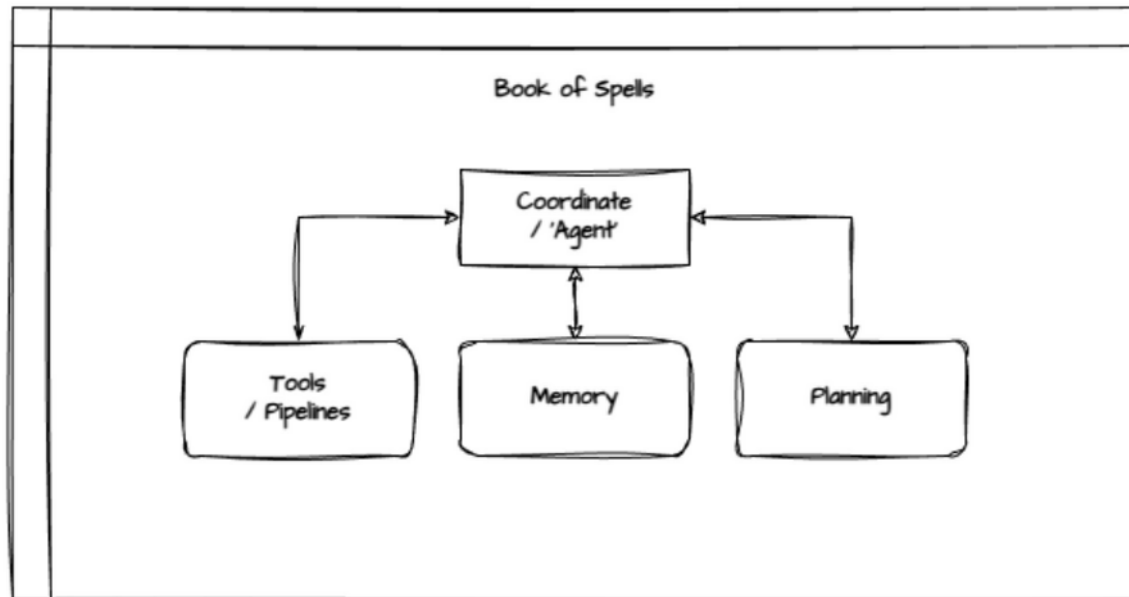
# Modular RAG

- ❑ Enhance LLM Capabilities with dedicated functionality
  - ❑ Calculator
  - ❑ Weather
  - ❑ Documents
  - ❑ Web Search
  - ❑ Semantic Search
  - ❑ etc pp



# LLM Frameworks

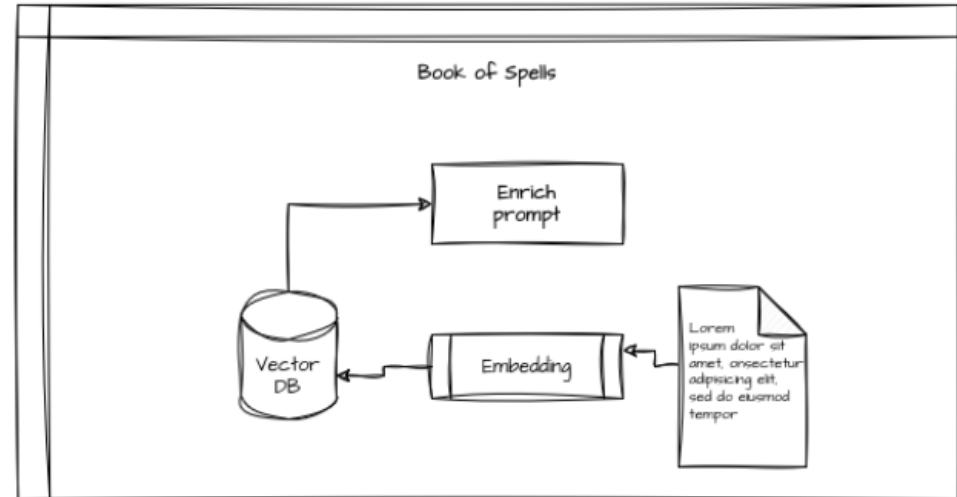
- ❑ 3. Generation: ,Modular' RAG
- ❑ Integration Patterns





# Pipelines (are dead)

- ❑ Controlled, structured workflows using Pipelines
- ❑ Chainable
  - ❑ LLMChain
  - ❑ SequentialChain
  - ❑ RouterChain
- ❑ Example Tool : ,langchain'





# open WebUI - Tools, Models & Knowledge

Models Knowledge Prompts Tools

Tools 2

Search Tools

TOOL v0.3.0 yahoo finance yahoo\_finance

A comprehensive stock and cryptocurrency analysis tool...

By User

TOOL v0.1.1 Weather weather

Get the weather for a specific city. Does not require an A...

By User

New Chat

Workspace

Search

Chats

Today

1 2 3 4 IBM Stock Performance vs. NA

Berlin Weather

Models Knowledge Prompts Tools

Some weird stuff Access

Grimms Fairy Tales

Search Collection

bluecap.pdf17.1 KB

rapperton.pdf16.3 KB



# MCP (Model Context Protocol)

- ❑ ,USB-C port for AI Applications'
- ❑ Developed by company ,Anthropic', open-sourced Nov 2024



## Model Context Protocol

A protocol for seamless integration between LLM applications and external data sources

[Documentation](#) | [Specification](#) | [Discussions](#)

<https://github.com/modelcontextprotocol>

- ❑ open-webui proxy server (mcpo)
  - ❑ openapi reverse proxy to mcp services
  - ❑ -> <https://github.com/open-webui/mcpo>





# mcpo usage

- ❑ <https://docs.openwebui.com/openapi-servers/mcp>
- ❑ <https://github.com/modelcontextprotocol/servers>
- ❑ Demo

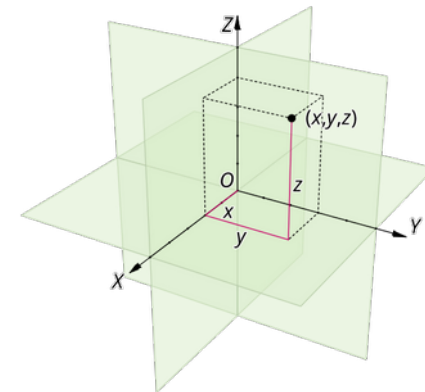
```
neo4j-mcp-ollama$ docker ps
CONTAINER ID   IMAGE                                     COMMAND
94d0d068589e   ghcr.io/open-webui/open-webui:main      "bash start.sh"
e4781194d1e8   neo4j-mcp-ollama-neo4j-mcp-server       "uvx mcpo --host 0.0..."
2765c8ba8fae   neo4j                                    "tini -g -- /startup..."
6325321a3b2e   ollama/ollama:latest                    "/bin/ollama serve"
```

```
{
  "mcpServers": {
    "neo4j-aura": {
      "command": "uvx",
      "args": [
        "mcp-neo4j-cypher@0.2.1"
      ]
    }
  }
}
```



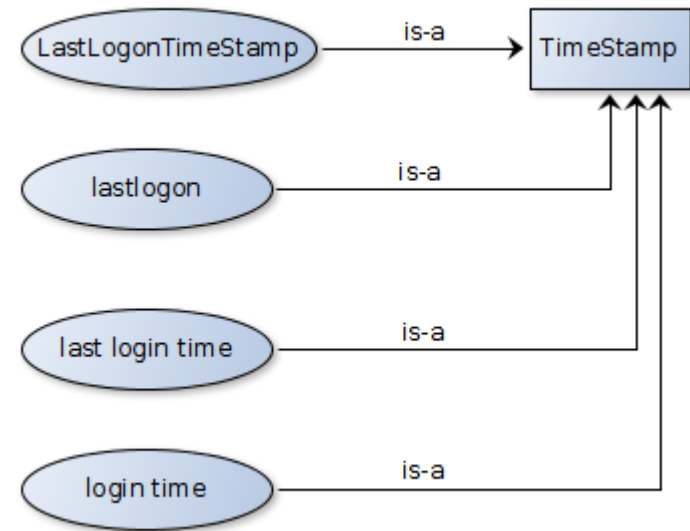
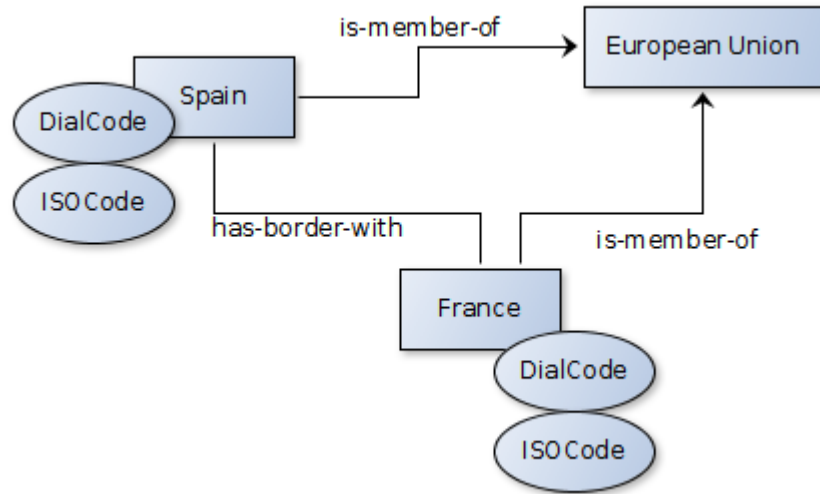
# Graph Prompting

- ❑ Graph Prompt ,unifying pre-training and downstream tasks for graph neural network'
- ❑ Liu et al, <https://arxiv.org/pdf/2302.08043>
- ❑ Chain of Thought vs ,Graph of Thought' (CoT vs. GoT)
- ❑ Euclidian vs non-euclidian data
  - ❑ Bronstein et al, 2016: <https://arxiv.org/pdf/1611.08097>
- ❑ simplified (greatly)
  - ❑ Three-dimensional (euclidian)
  - ❑ Three(+x) -dimensional (non-euclidian)





# Things vs. Strings



# IRM and Ontologies

Wine for the confused  
(John Cleese 2004)

#Subject predicate #object

#PinotNoir is-a #wine

#PinotBlanc is-a #wine

#PinotNoir has-color #red

#wine has-color #[red|white|rose]

'PinotNoir is black wine' ?



*Semantics*



What is  
'PinotNoir'?

# IRM and Ontologies

Wine for the confused

In vino veritas  
(2013)

What is an organization ?  
What is a country ?  
What is a timestamp ?  
What means 'manager' ?



*Ontologies ≠ Ontology*

Which of our departments deal with customer data ?  
What is our fiscal year ?  
What are our rules for situation 'X' ?

A [User] is [required] [to] [reset]  
his [Password] [every 90 days] .

How does our application 'Z' deal with process 'Y' ?  
Which applications are required for process 'Y' ?



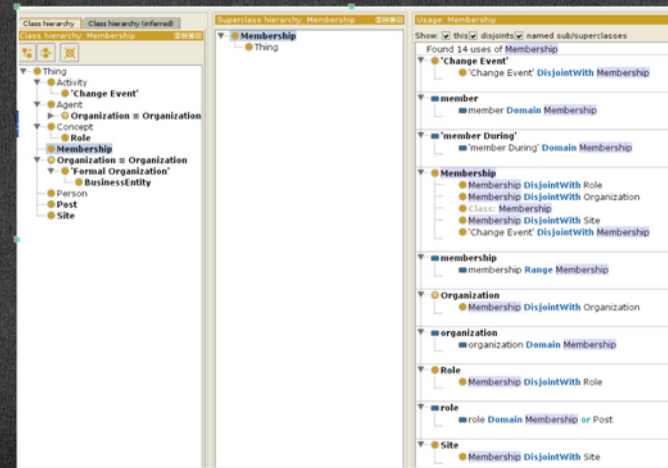


# IRM and Ontologies

Wine for the confused

## Relationship Manager

IRM calls / seeks for a 'Relationship Manager'  
Could it be a semantic/ ontology tool ?





# (Knowledge) Graphs

- ❑ Semantics and Knowledge Graphs
- ❑ Structures
  - ❑ Nodes and Edges
- ❑ Components
  - ❑ Nodes (Entitites)
  - ❑ Edges (Relationships)
  - ❑ Attributes (Properties)
  - ❑ (Ontologies and reasoning)





# Graph DB vs Vector DB

## **,Graph DB'**

- ❑ Purpose / Use Cases
  - ❑ semi-structured
- ❑ Data Structure
  - ❑ Nodes, edges and properties
- ❑ Query Types
  - ❑ explore relationships

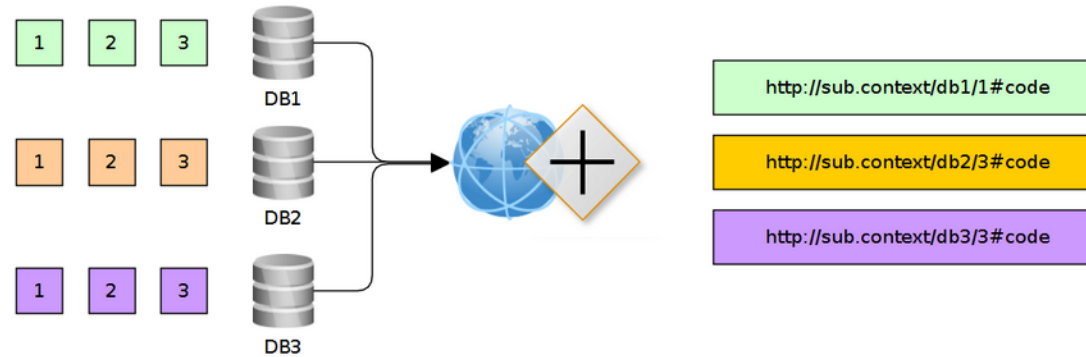
## **,Vector'**

- ❑ Purpose / Use Cases
  - ❑ Unstructured
- ❑ Data Structure
  - ❑ Multi-Dimensional
- ❑ Query Types
  - ❑ Similarity

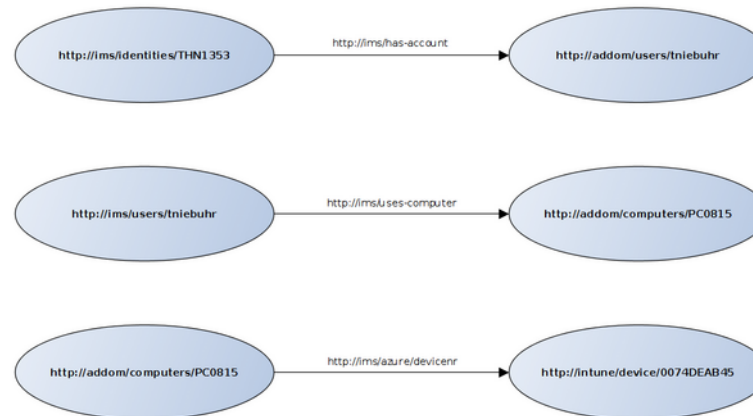


# Build your graph

## Data De-Duplication

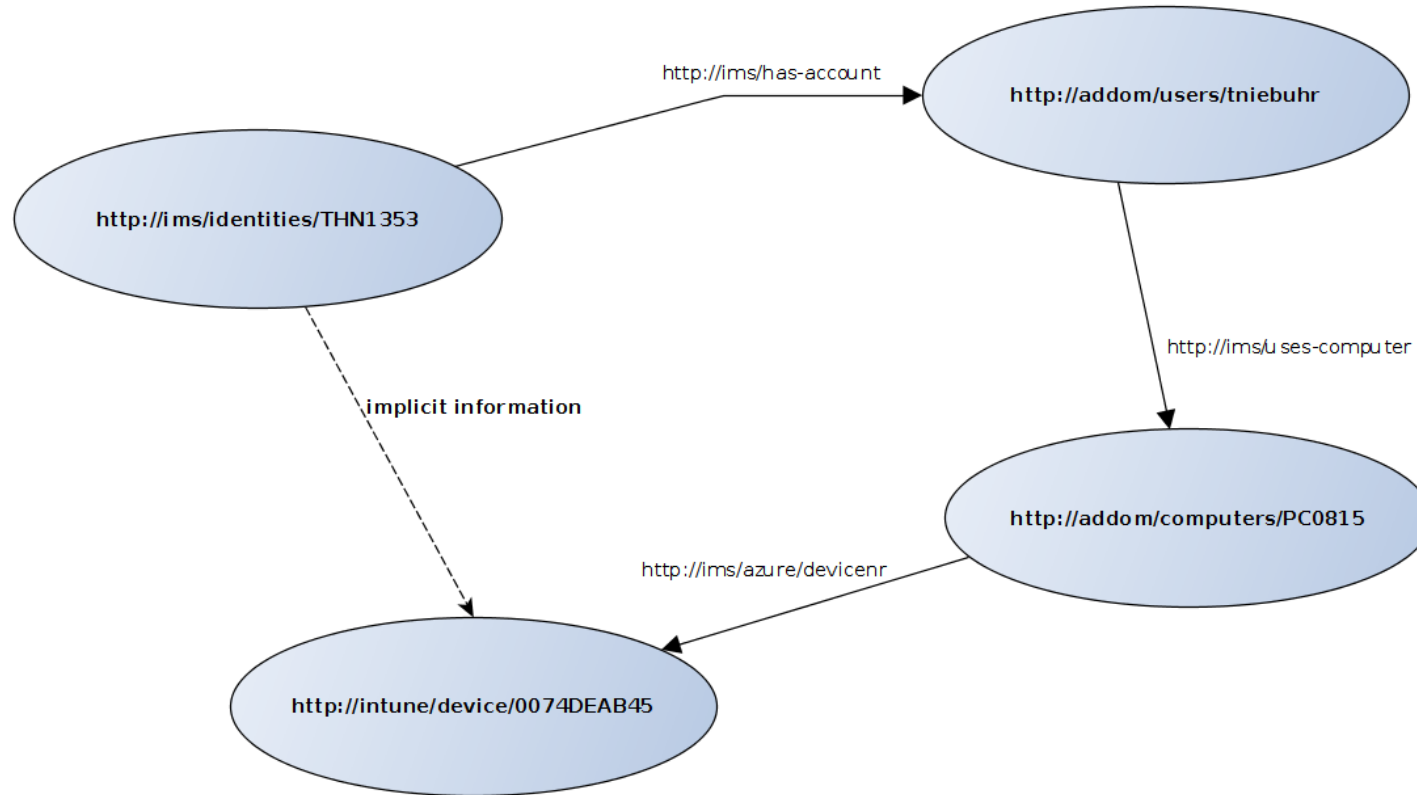


## Triples



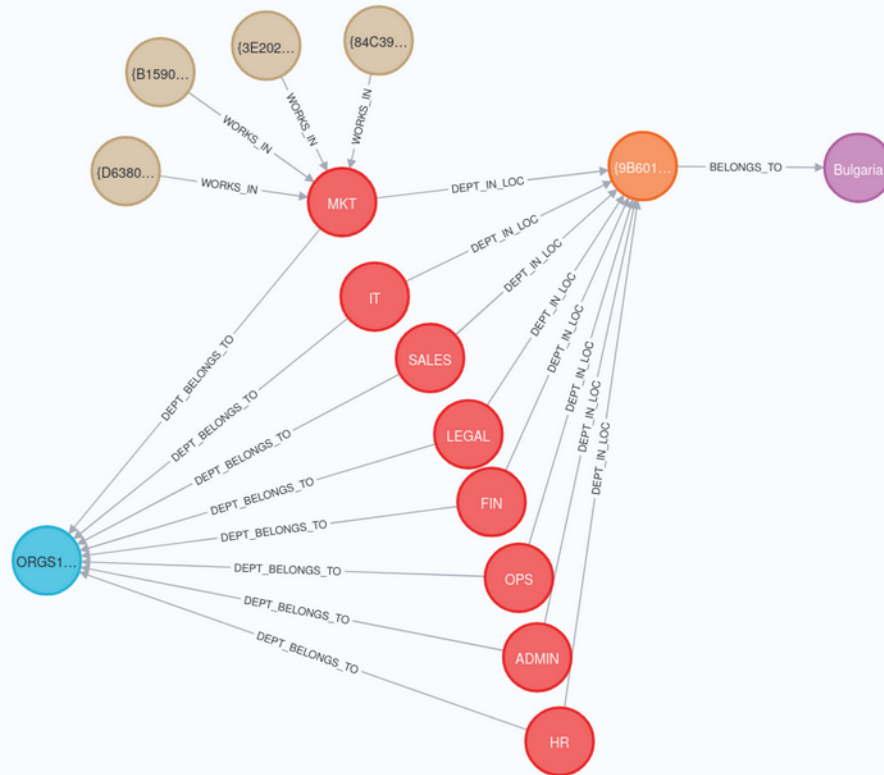


# Operating on the edges





# Explore a Graph



## Overview

### Node labels

\*(38) Users (27) Departments (8) Locations (1) Organizations (1)  
Countries (1)

### Relationship types

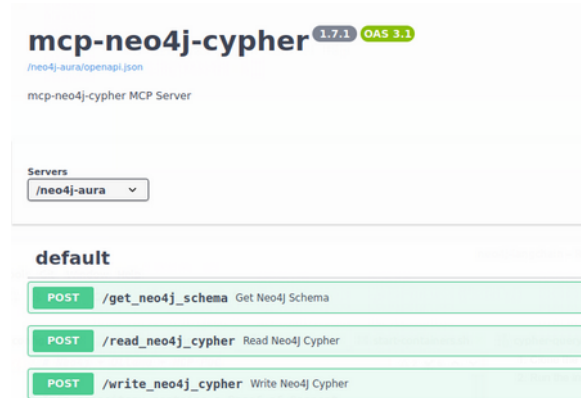
\*(21) WORKS\_IN (4) DEPT\_IN\_LOC (8) DEPT\_BELONGS\_TO (8)  
BELONGS\_TO (1)

Displaying 38 nodes, 14 relationships.

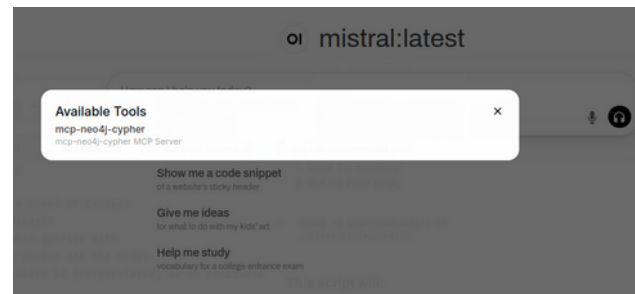


# Native Language to graph

 mcpo component



 Tool enabled





# Investigate

Which types of Entities are defined in the connected graph. List them

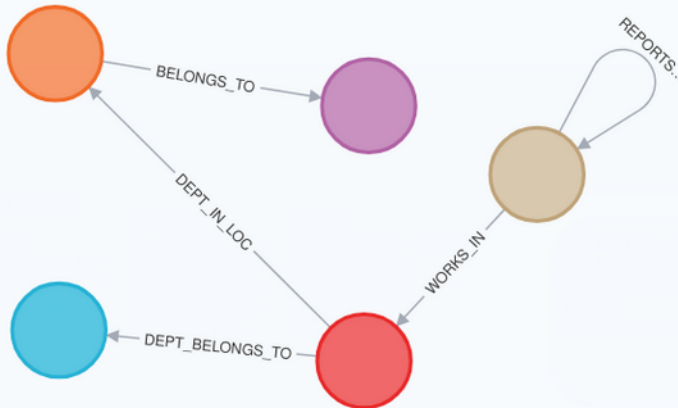


**OI** **mistral:latest**

In the provided context, there are five types of entities defined: Users, Departments, Locations, Organizations, and Countries. [1]



```
// What is related, and how  
CALL db.schema.visualization()
```



Overview

Node labels

\* (5) Departments (1)  
Countries (1) Organizations (1)  
Locations (1) Users (1)

Relationship types

\* (5) WORKS\_IN (1)  
BELONGS\_TO (1) DEPT\_IN\_LOC (1)  
DEPT\_BELONGS\_TO (1)  
REPORTS\_TO (1)

Displaying 5 nodes, 5 relationships.



# FreeStyle

Create a graphical concept of the term 'freestyle', use presentation techniques as context



Worked for 10.3s



4:01pm










<https://commons.wikimedia.org/wiki/File:Thank-you-word-cloud.jpg>

Thorsten H. Niebuhr  
tniebuhr@wedacon.net

 @idmpath

 @idmpath.bsky.social  
Bluesky