

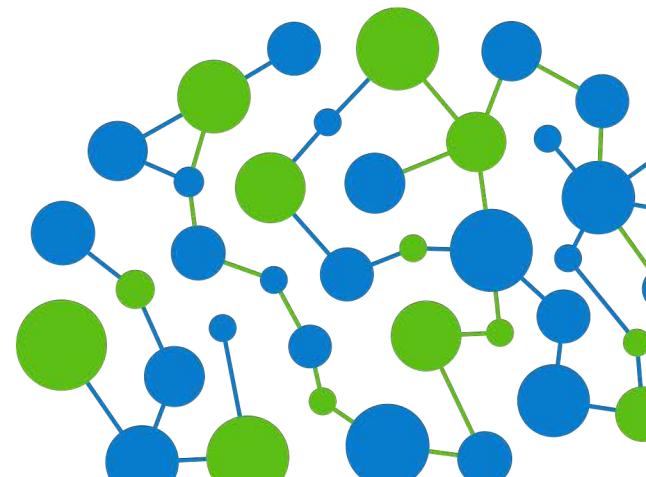


# Theoretical Innovations and New Research Paradigms of Knowledge Graphs in the Era of Large Language Models

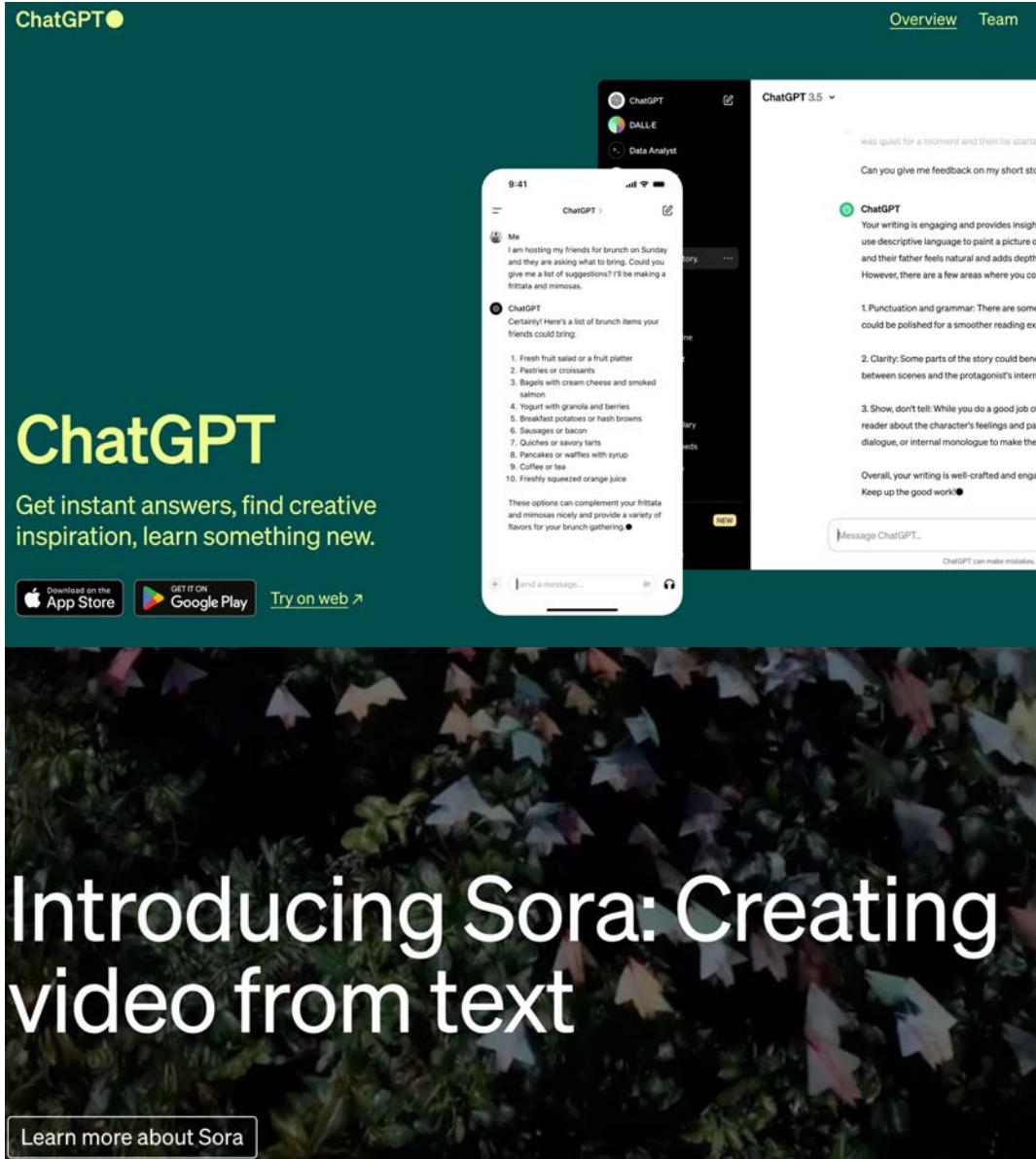
Haofen Wang

LKM Workshop, IJCAI 2024

- The Rise of LLM
- The Development of KG
- The Emerging Paradigm
- OpenKG
- Conclusion and Outlook



# The Wave towards AGI by LLM



ChatGPT

Get instant answers, find creative inspiration, learn something new.

Download on the App Store GET IT ON Google Play Try on web

Overview Team

ChatGPT 3.5

was quiet for a moment and then he starts

Can you give me feedback on my short story?

Me I am hosting my friends for brunch on Sunday and they are asking what to bring. Could you give me a list of suggestions? I'll be making a frittata and mimosa.

ChatGPT Certainly! Here's a list of brunch items your friends could bring:

1. Fresh fruit salad or a fruit platter
2. French toast or croissants
3. Bagels with cream cheese and smoked salmon
4. Yogurt with granola and berries
5. Breakfast potatoes or hash browns
6. Sausages or bacon
7. Quiches or savory tarts
8. Frittatas or waffles with syrup
9. Coffee or tea
10. Freshly squeezed orange juice

These options can complement your frittata and mimosa nicely and provide a variety of flavors for your brunch gathering.

was quiet for a moment and then he starts

Can you give me feedback on my short story?

Me was quiet for a moment and then he starts

ChatGPT Your writing is engaging and provides insight. You use descriptive language to paint a picture of the scene, and their father feels natural and adds depth. However, there are a few areas where you could improve:

1. Punctuation and grammar: There are some punctuation errors that could be fixed for a smoother reading experience.
2. Clarity: Some parts of the story could benefit from clearer descriptions between scenes and the protagonist's internal thoughts.
3. Show, don't tell: While you do a good job of showing the reader about the character's feelings and past experiences, there are opportunities for dialogue or internal monologue to make the story even more compelling.

Overall, your writing is well-crafted and engaging. Keep up the good work!

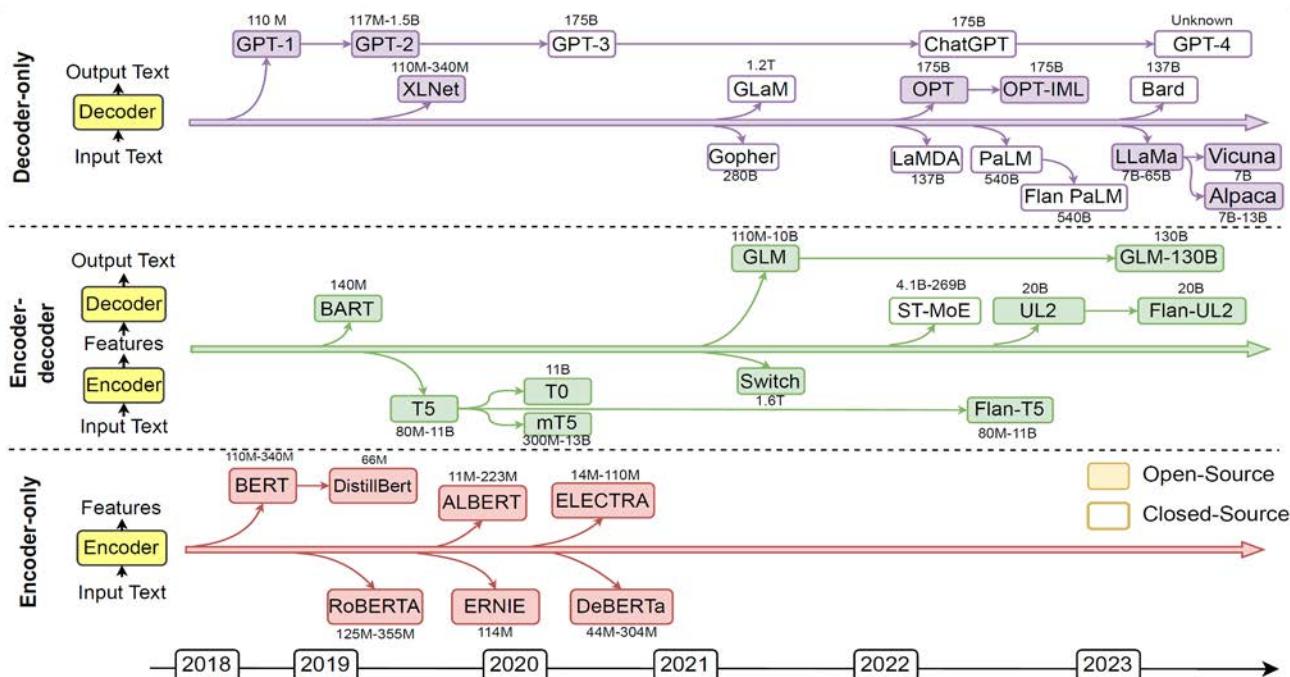
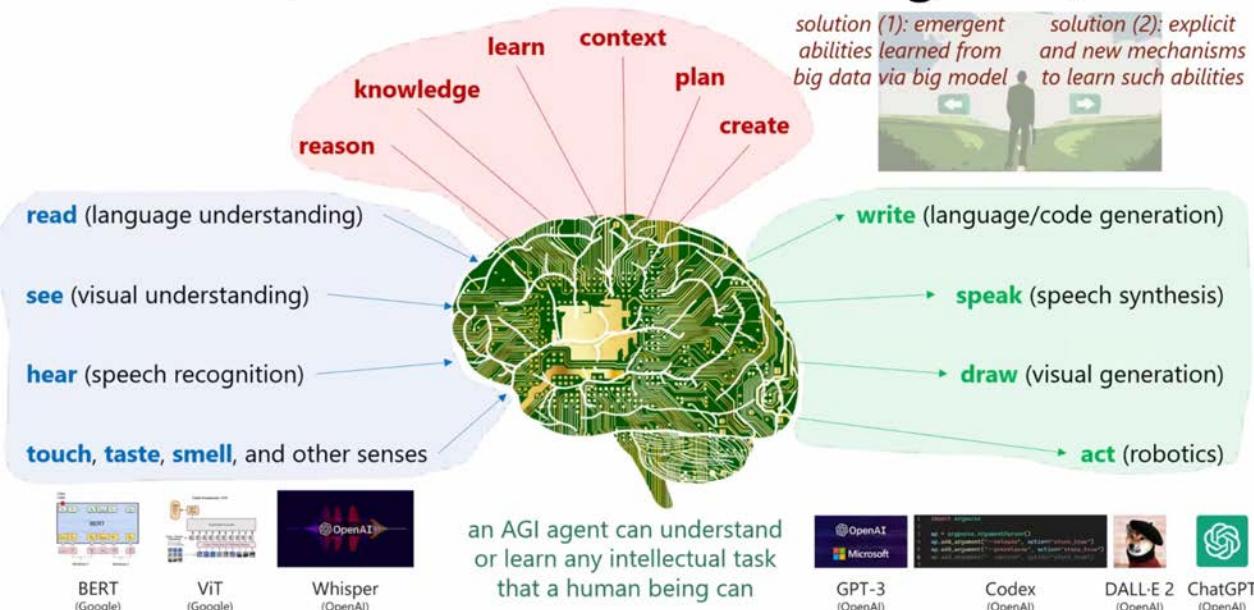
Message ChatGPT... ChatGPT can make mistakes.

Learn more about Sora



# The Wave towards AGI by LLM

## AGI (Artificial General Intelligence)



Credit by Nan Duan, MSRA

A Survey of Large Language Models, 2023

# The Programming of LLM: Prompt Engineering

## Standard Prompting

**Q:** Natalia sold clips to 48 of her friends in April, and then she sold half as many clips in May. How many clips did Natalia sell altogether in April and May?

**T:** The answer is :

LM

72

## Few-shot Prompting

**Q:** There are 3 cars in the parking lot and 2 more cars arrive. How many cars are in the parking lot?

**A:** The answer is 5.

**Q:** Natalia sold clips to 48 of her friends in April, and then she sold half as many clips in May. How many clips did Natalia sell altogether in April and May?

**A:** The answer is :

LM

72

## Chain-of-Thought Prompting

## Prompting

**Q:** There are 3 cars in the parking lot and 2 more cars arrive. How many cars are in the parking lot?

**C:** There are 3 cars in the parking lot already. 2 more arrive. Now there are  $3 + 2 = 5$  cars.

**A:** The answer is 5.

**Q:** Natalia sold clips to 48 of her friends in April, and then she sold half as many clips in May. How many clips did Natalia sell altogether in April and May?

LM

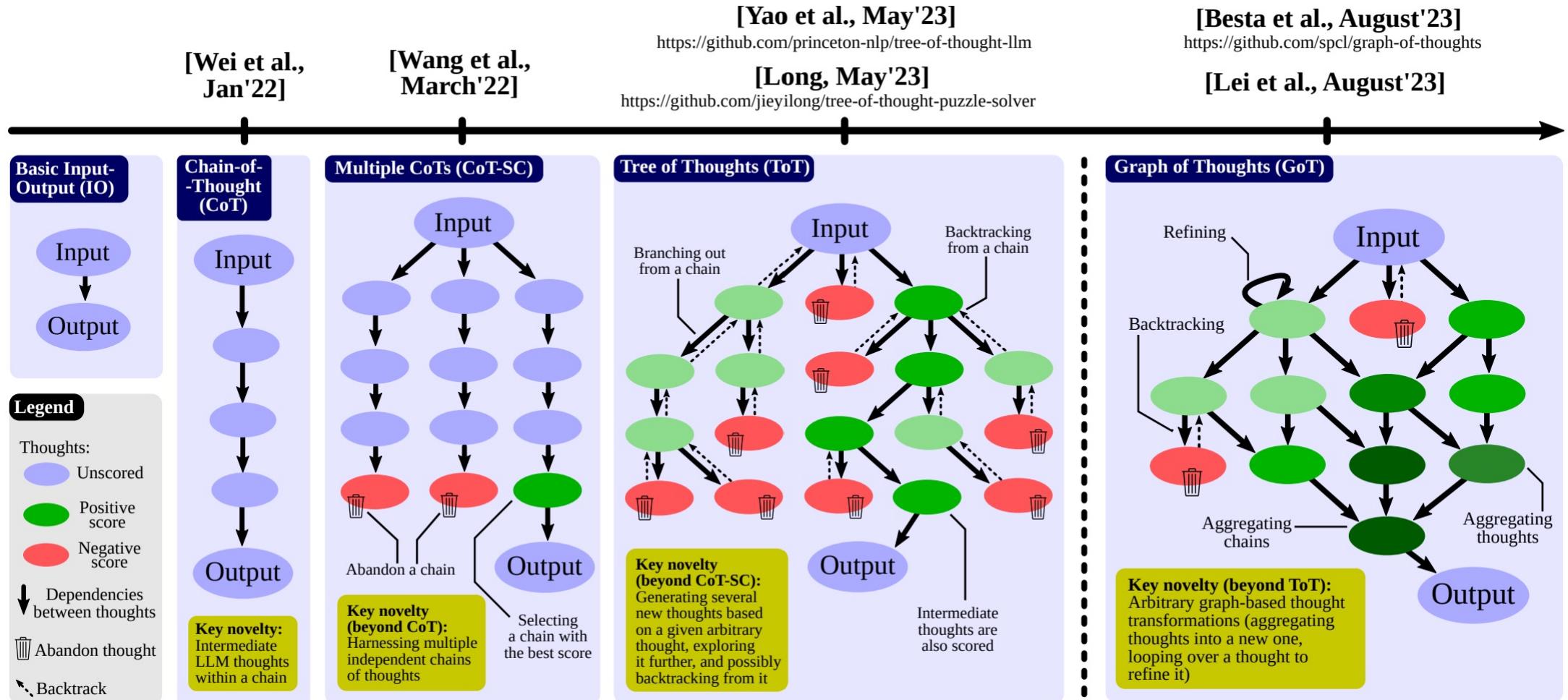
**C:** Natalia sold  $48 / 2 = 24$  clips in May. Altogether, Natalia sold  $48 + 24 = 72$  clips in April and May.

**A:** The answer is 72.

T

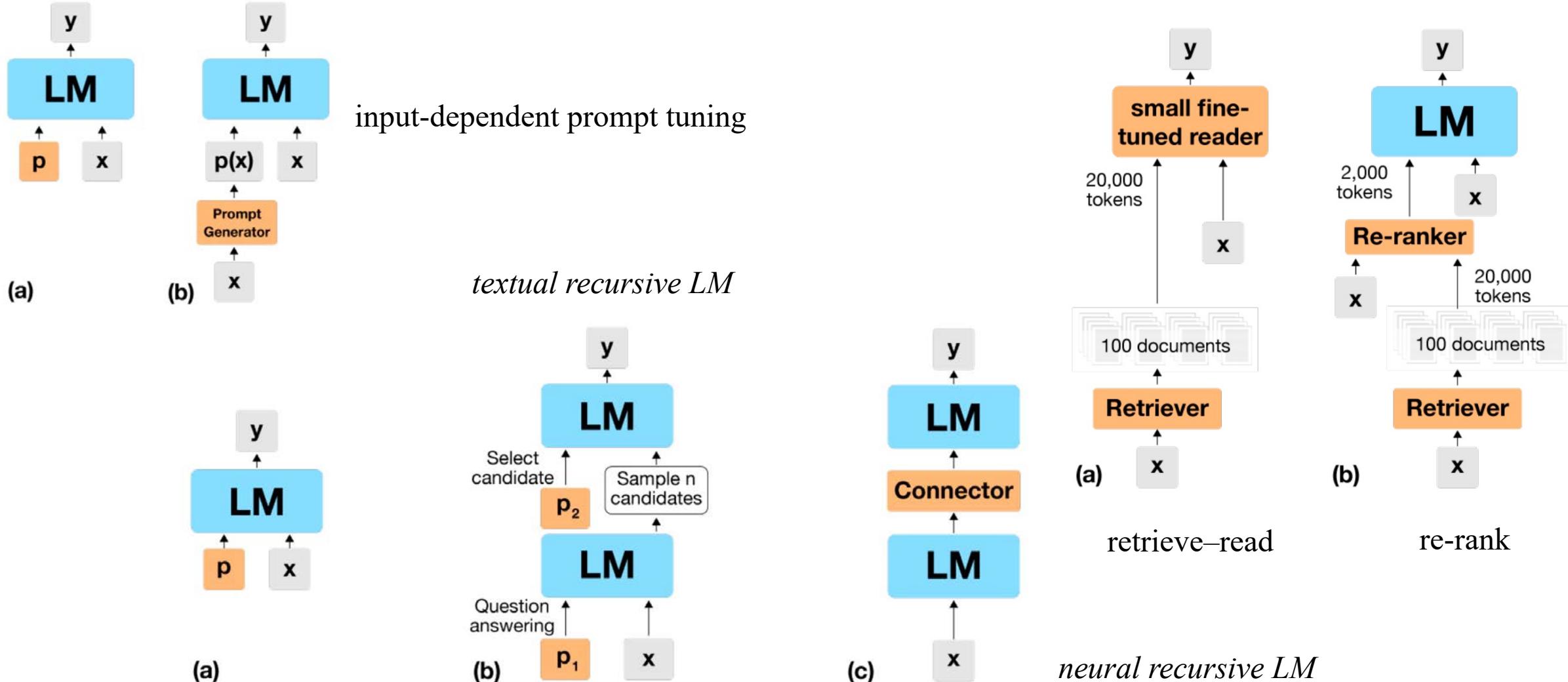
T

# The Programming of LLM: Prompt Reasoning



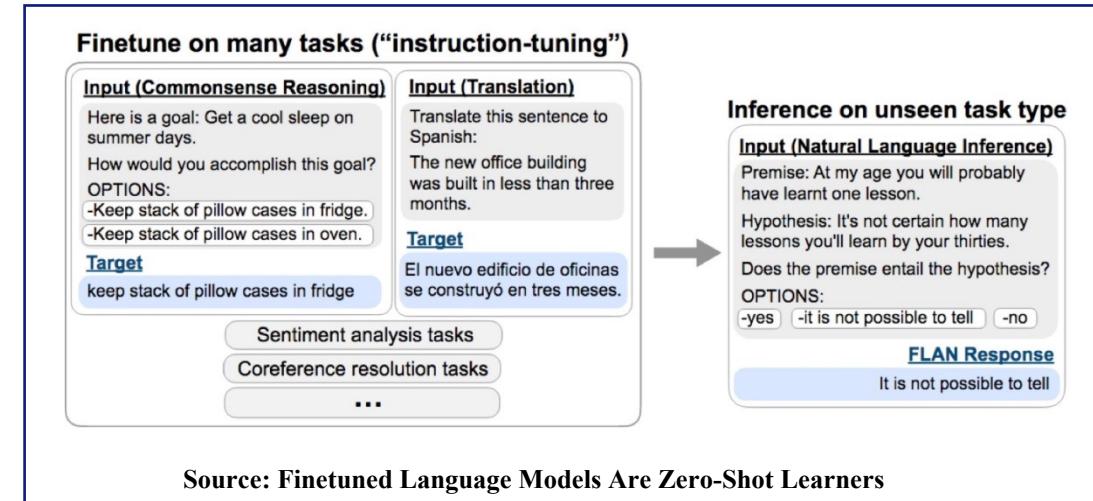
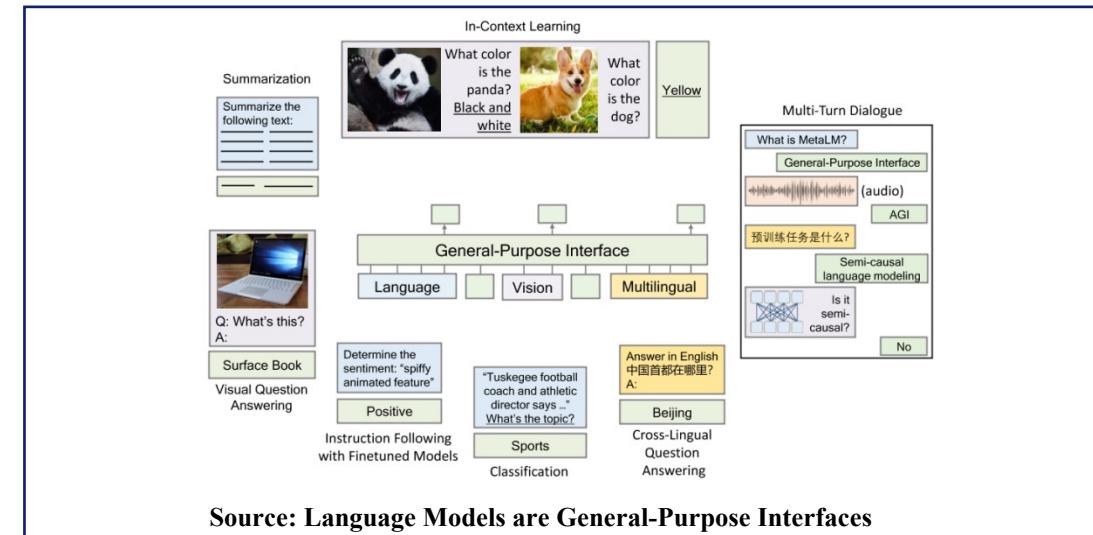
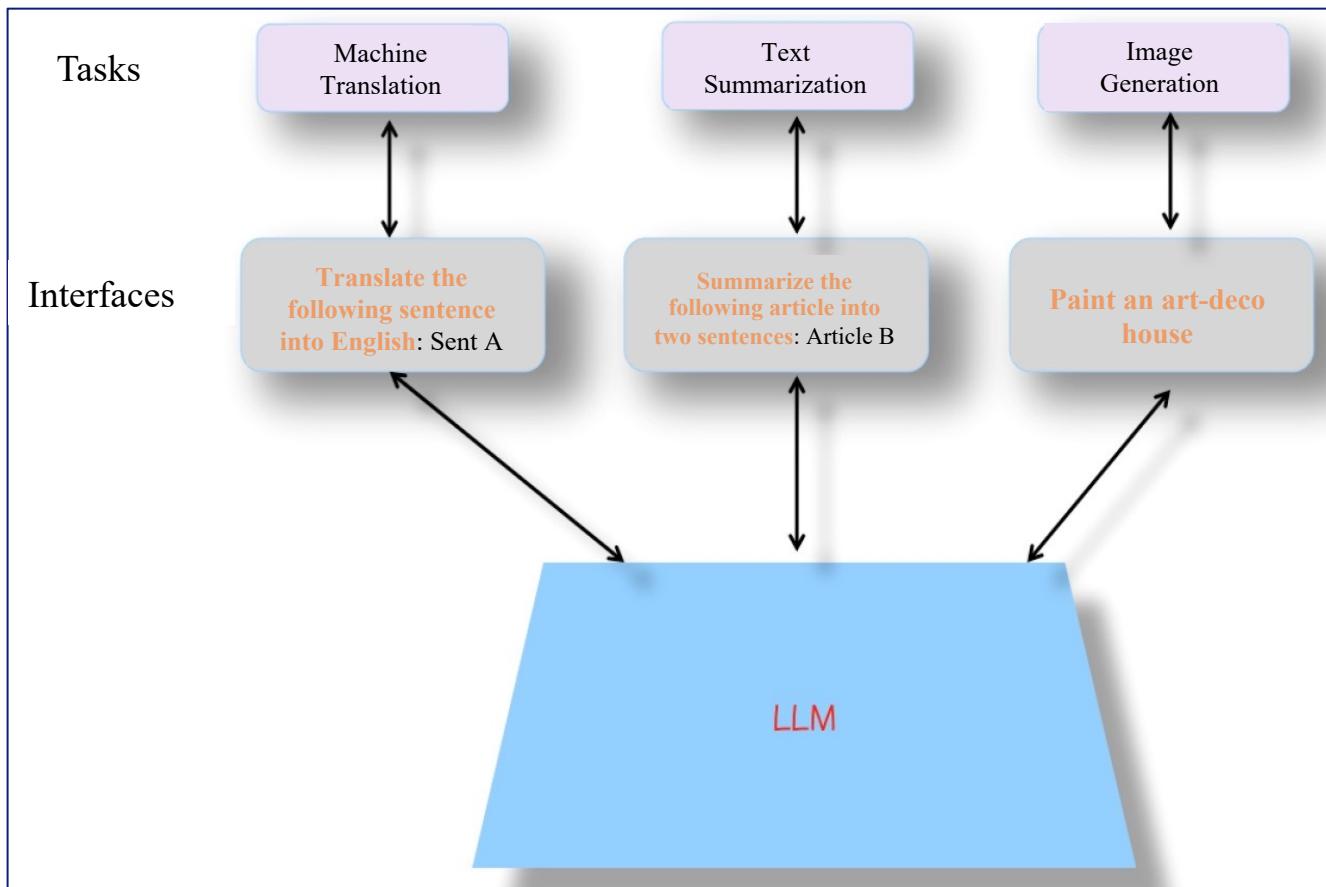
Structured and more logical prompt knowledge helps constrain **complex reasoning**, resulting in a better step-by-step thought process.

# The Programming of LLM: Design Patterns of Prompts



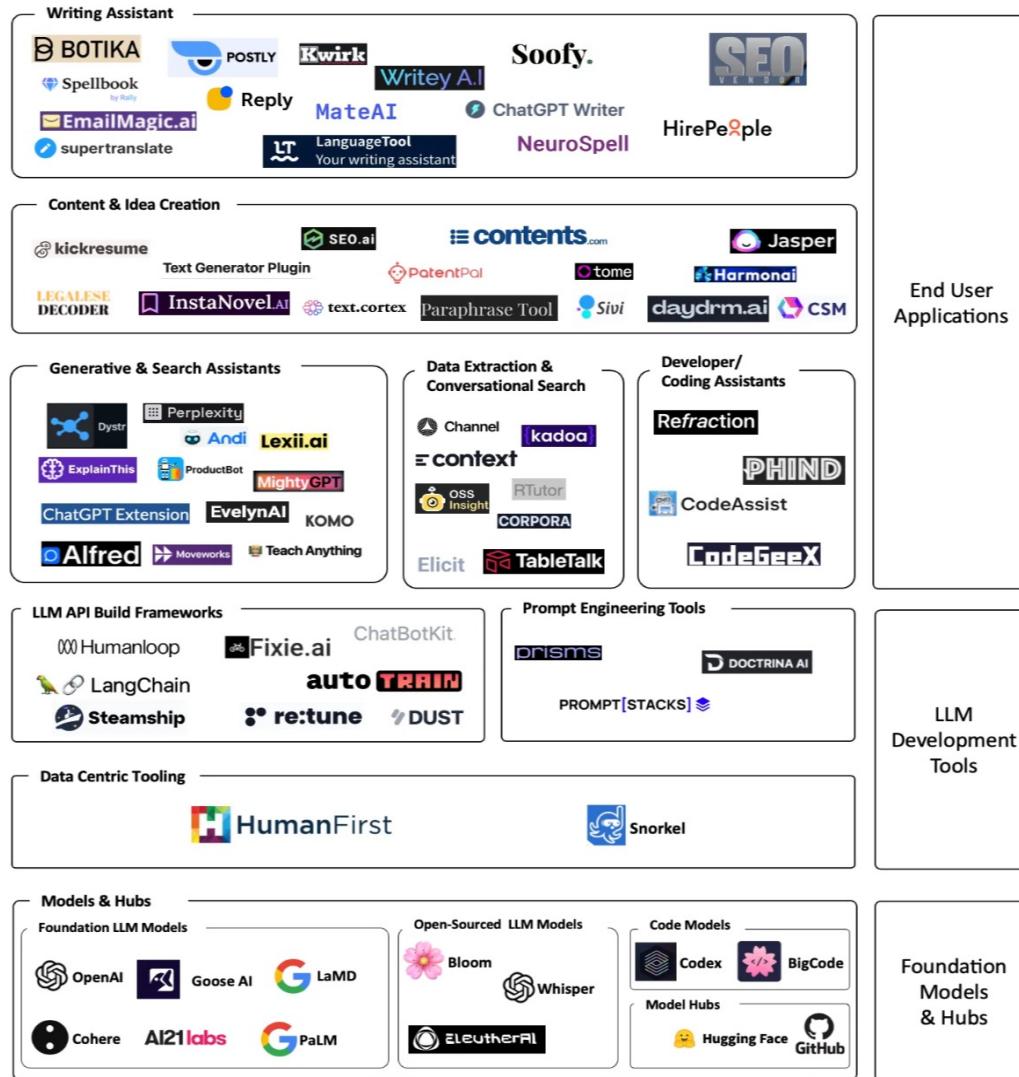
# LLM as a Basis for Providing Interfaces

- Zero/few shot prompting
- In Context Learning
- Instruction

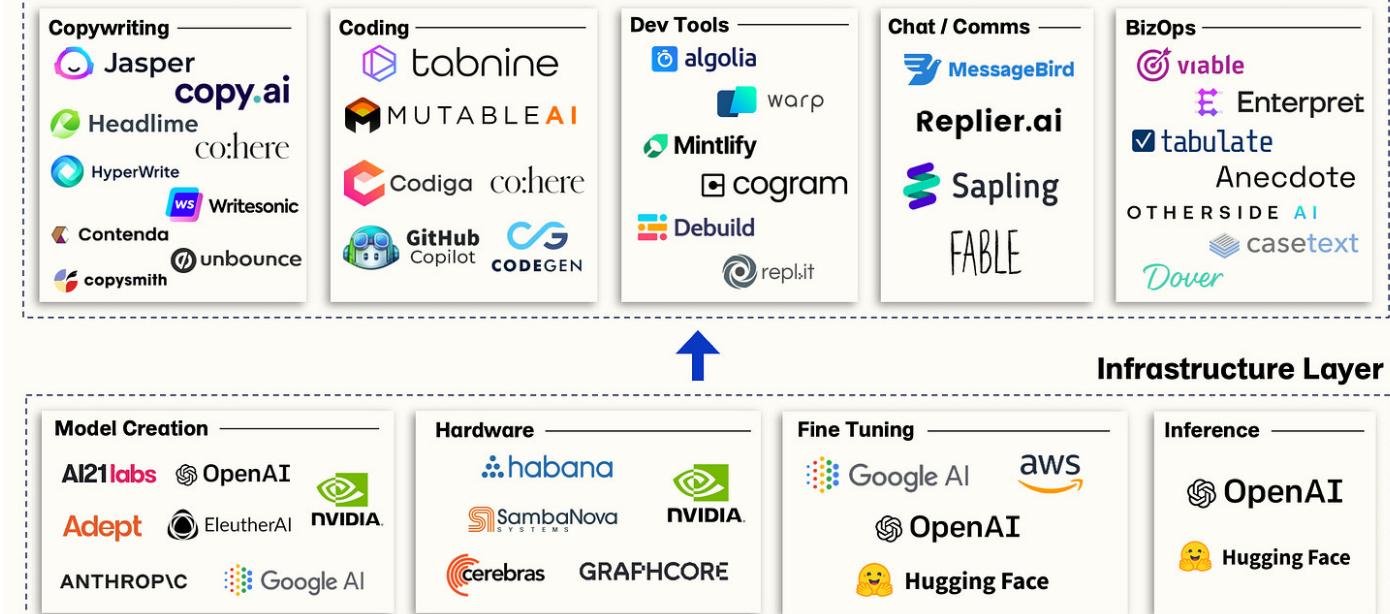


# The Ecosystem of LLM

## Foundation Large Language Model Stack



## Large Language Models



# GPT = Giant Parrot Talking?



# Limitations of LLMs

## Personalized questions

15

Which companies in Beijing have a market value greater than 50 million RMB?

15

I'm sorry, but I don't have access to real-time financial data or the ability to browse the internet to provide you with a list of companies in Beijing with market values greater than 50 million RMB. Market values of companies can change frequently due to stock price fluctuations, mergers, acquisitions, and other factors.

To obtain the most up-to-date information on the market values of companies in Beijing, I recommend checking financial news websites, stock market exchanges, or financial databases. You can also consult with a financial advisor or use a financial data service to access this information.

15

How is the weather today?

## Real-time questions

15

I'm unable to provide real-time weather information as my knowledge was last updated in 2022, and I don't have access to the internet to fetch current data. To find out the current weather in your location, I recommend using a weather website or app, or you can simply check a search engine like Google for the latest weather updates by entering your location and "weather" as the search query.

## Questions that require new knowledge

15

How do you perceive the results of the 2022 FIFA World Cup?

15

I don't have access to real-time information or events that occurred after my last knowledge update in 2022. Therefore, I cannot provide information or perceptions about the results of the 2022 FIFA World Cup or any other events that occurred after that date.

To get the most up-to-date information and perspectives on the 2022 FIFA World Cup or any recent events, I recommend checking reliable news sources or sports websites that cover the tournament. You can also engage in discussions with people who have been following the event to get their insights and perceptions.

15

If I could foresee the future, how should I change my destiny?

## Hallucination

15

If you could accurately foresee the future, it would indeed be a powerful ability. However, there are several considerations and ethical dilemmas associated with this hypothetical situation. Assuming that you could foresee the future, here's how you might consider changing your destiny:

1. **Planning:** Use your foreknowledge to create well-informed plans and strategies for your life. This could include career goals, financial planning, personal development, and more.
2. **Risk Management:** If you can foresee future difficulties or challenges, take proactive measures to mitigate potential risks. This might involve purchasing insurance, making lifestyle changes, or seeking legal advice.
3. **Financial Decisions:** Make investment and financial decisions based on the insights you have about the future to help achieve your financial goals.
4. **Relationships:** Use your knowledge of the future to build healthier relationships, resolve potential conflicts, and improve your personal and professional interactions.
5. **Personal Growth:** Knowing what opportunities and challenges lie ahead, focus on personal growth, learning, and self-improvement to better prepare yourself for the future.

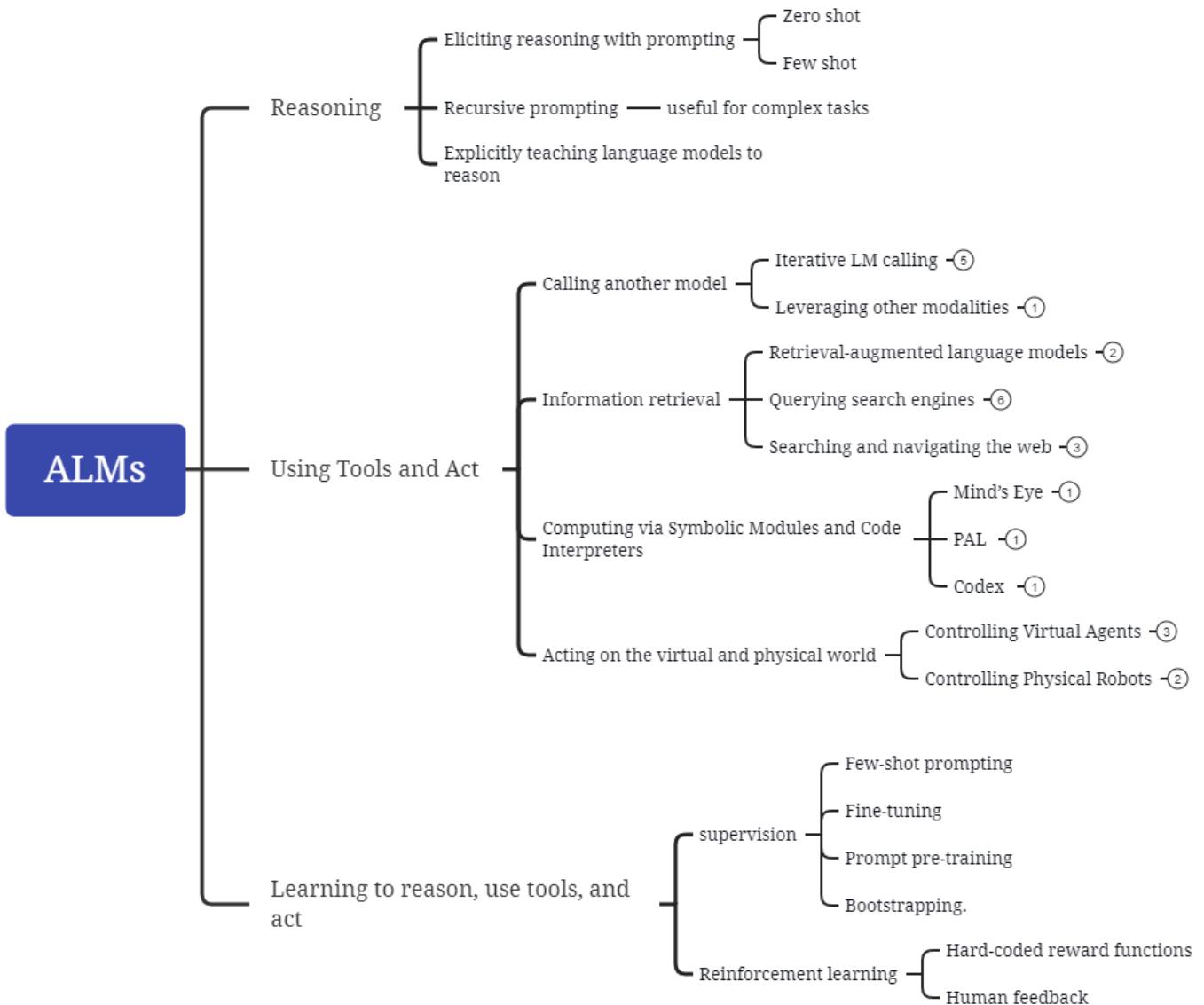
It's important to emphasize that this is a hypothetical scenario. In reality, we cannot foresee the future with certainty. Every decision we make is based on current information and risk assessments. Additionally, even in this hypothetical scenario, ethical considerations are vital. Any actions you take should be legal, ethical, and considerate of the well-being of others.

- It cannot handle personalized, dynamic, real-time questions like stock prices or weather changes.
- LLMs are unable to acquire the latest knowledge, therefore it cannot provide the up-to-date information.
- The existence of Hallucinations leads to a tendency to talk nonsense with a serious tone.
- Inappropriate prompts may lead to irrelevant even harmful response

# Augmented Language Models

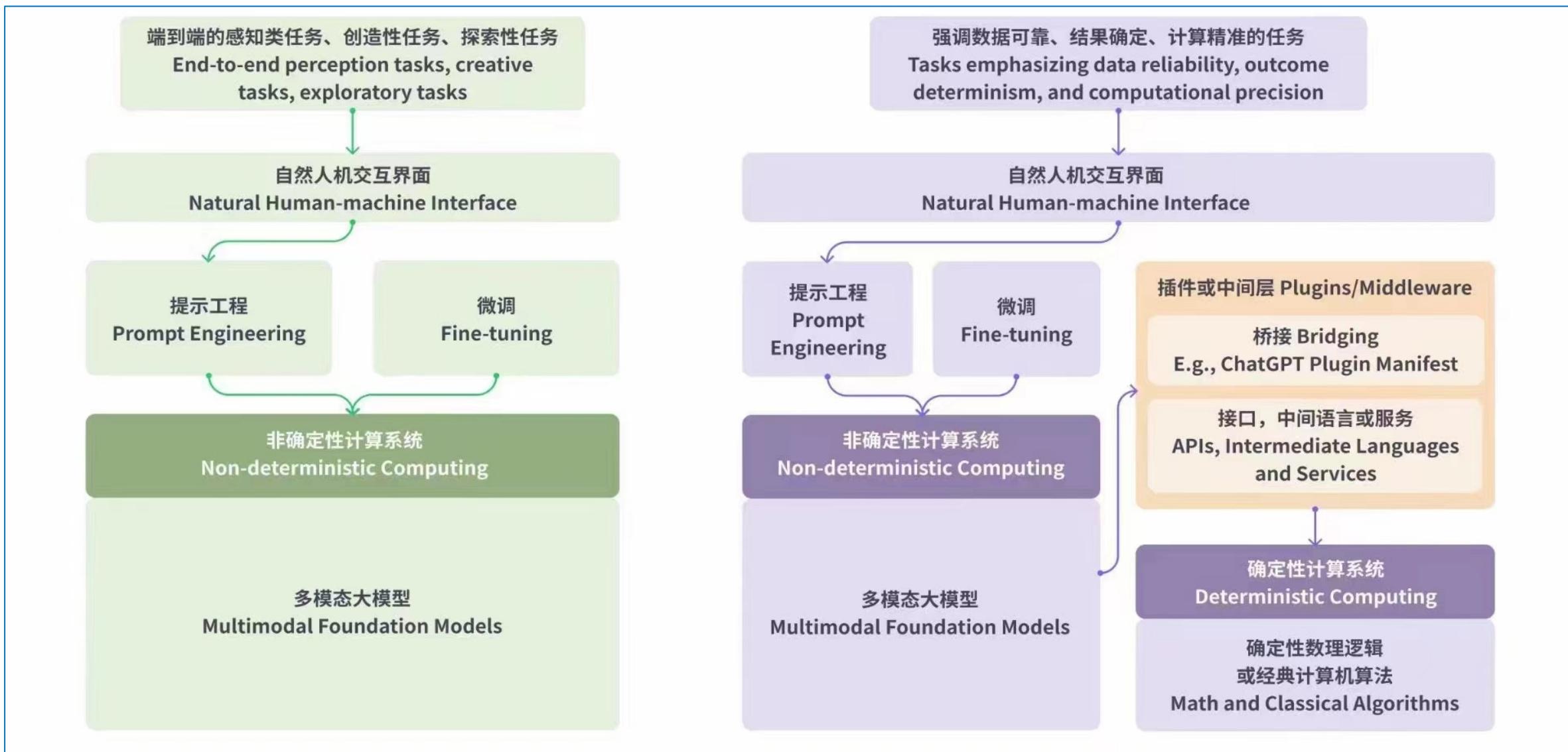
- ❖ Studies have proposed some improvements, such as enabling LLM to utilize external tools, enhancing context understanding with important missing information not included in LLM's weights, and forming a more powerful intelligent agent. These models are collectively referred to as **Augmented Language Models (ALMs)**.

- ❖ **Reasoning:** Breaking down complex tasks into simpler subtasks can make it easier for the LM to solve on its own or with the use of tools.
- ❖ **Tools:** Collecting external information, or influencing the virtual or physical world perceived by ALM.
- ❖ **Act :** Invoking a tool that has an impact on the virtual or physical world and observing its results, incorporating it into the current context of ALM.

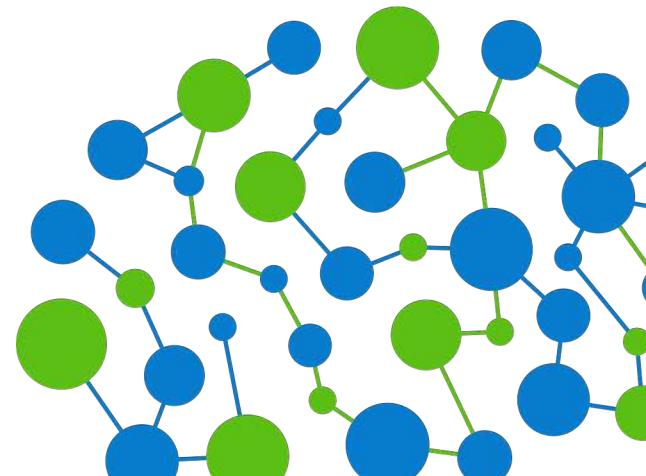


Source: Augmented Language Models: a Survey ( Yann Lecun et al. )

# Two kinds of Usage of LLM

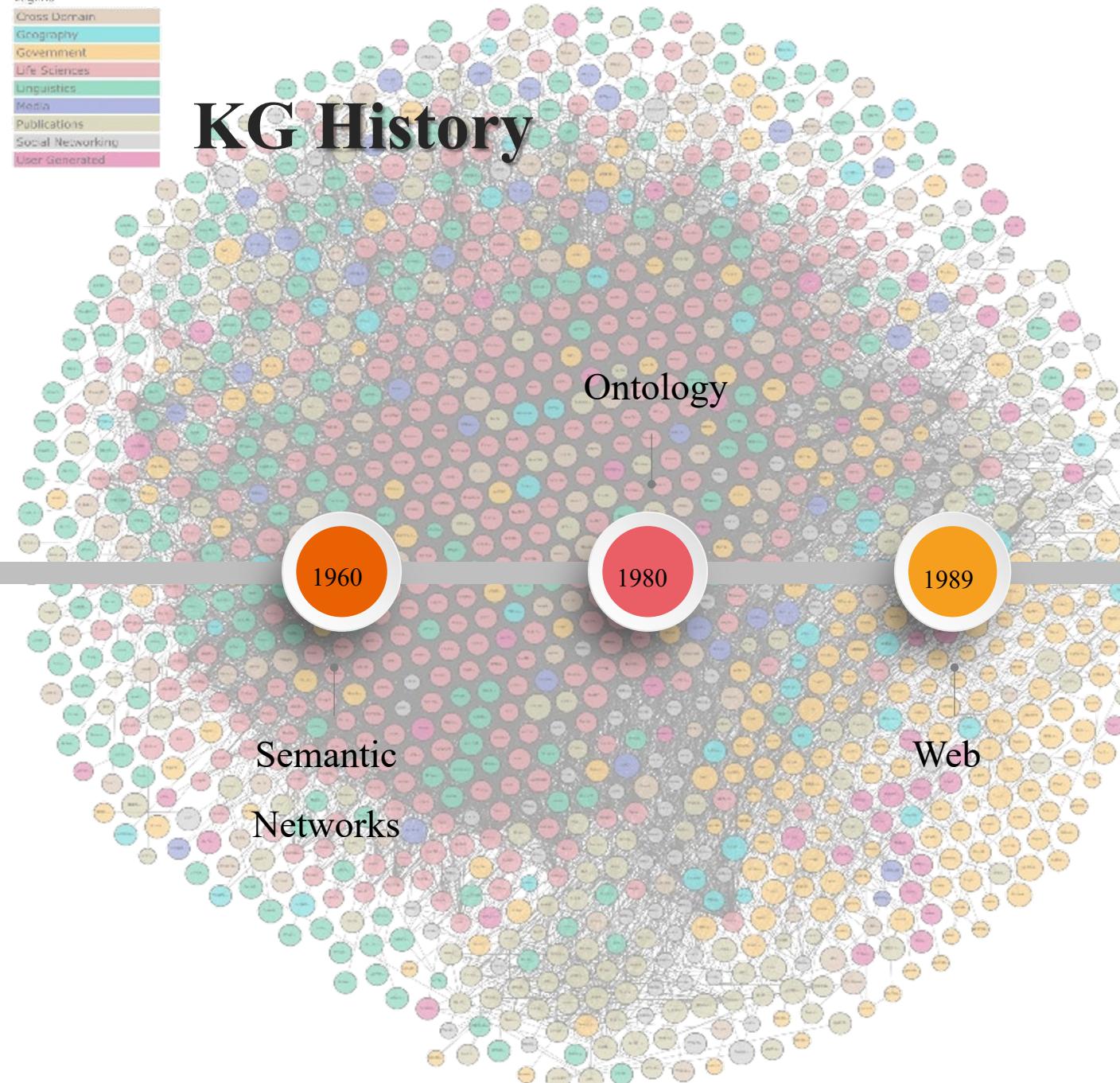


- The Rise of LLM
- **The Development of KG**
- The Emerging Paradigm
- OpenKG
- Conclusion and Outlook



Legend
Cross Domain
Geography
Government
Life Sciences
Linguistics
Media
Publications
Social Networking
User Generated

# KG History



Semantic  
Web

Knowledge Graph

Semantic  
Networks

Ontology

Web

Linked data

Linking Open Data cloud diagram 2022-11-03 Andrejs Abele, John P. McCrae, Paul Buitelaar, Anja Jentzsch and Richard Cyganiak. <http://lod-cloud.net/>

# What is Knowledge Graph (KG) – Popular KBs and Characteristics



Cyc



WordNet

$\text{guitarist} \subset \{\text{player}, \text{musician}\}$   
 $\subset \text{artist}$

algebraist

$\subset \text{mathematician}$   
 $\subset \text{scientist}$

$\forall x: \text{human}(x) \Rightarrow$   
 $(\exists y: \text{mother}(x,y) \wedge$   
 $\exists z: \text{father}(x,z))$

$\forall x,u,w: (\text{mother}(x,u) \wedge$   
 $\text{mother}(x,w)$   
 $\Rightarrow u=w)$

1985

1990

2000

2005

2010

2015

2023

Key  
Features

In the early stage, KG is High-quality, manually-built, and for human consumption; in the middle age, KG is constructed by algorithms and used to enhance the understanding capability of machines; nowadays KG is evolving towards multi-modality and subsymbolic representations

WordNet

By Human  
For Human

$\subset \{\text{player}, \text{musician}\}$   
 $\subset \text{artist}$

algebraist

$\subset \text{mathematician}$   
 $\subset \text{scientist}$

$\forall x: \text{human}(x) \Rightarrow$   
 $(\exists y: \text{mother}(x,y) \wedge$   
 $\exists z: \text{father}(x,z))$

$\forall x,u,w: (\text{mother}(x,u) \wedge$   
 $\text{mother}(x,w)$   
 $\Rightarrow u=w)$

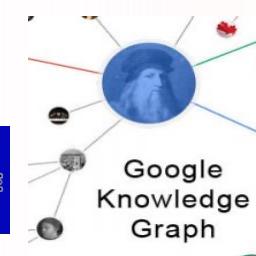
Wikipedia



4.5 Mio. English articles  
20 Mio. contributors



Freebase

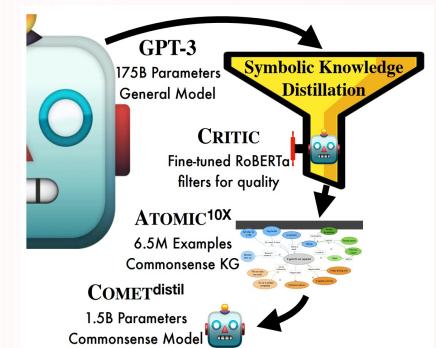
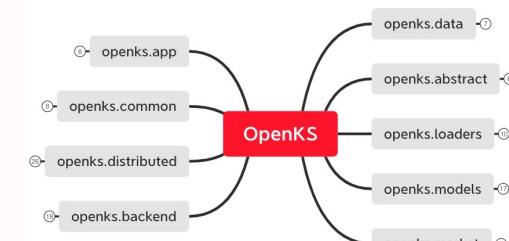


WolframAlpha™ computational knowledge engine

By Algorithm  
For Machine



Human Machine  
Collaboration



# Knowledge Graph

Knowledge Graph (KG) is an explicit representation of human knowledge, which is stored in the form of graph and used for reasoning and computing.

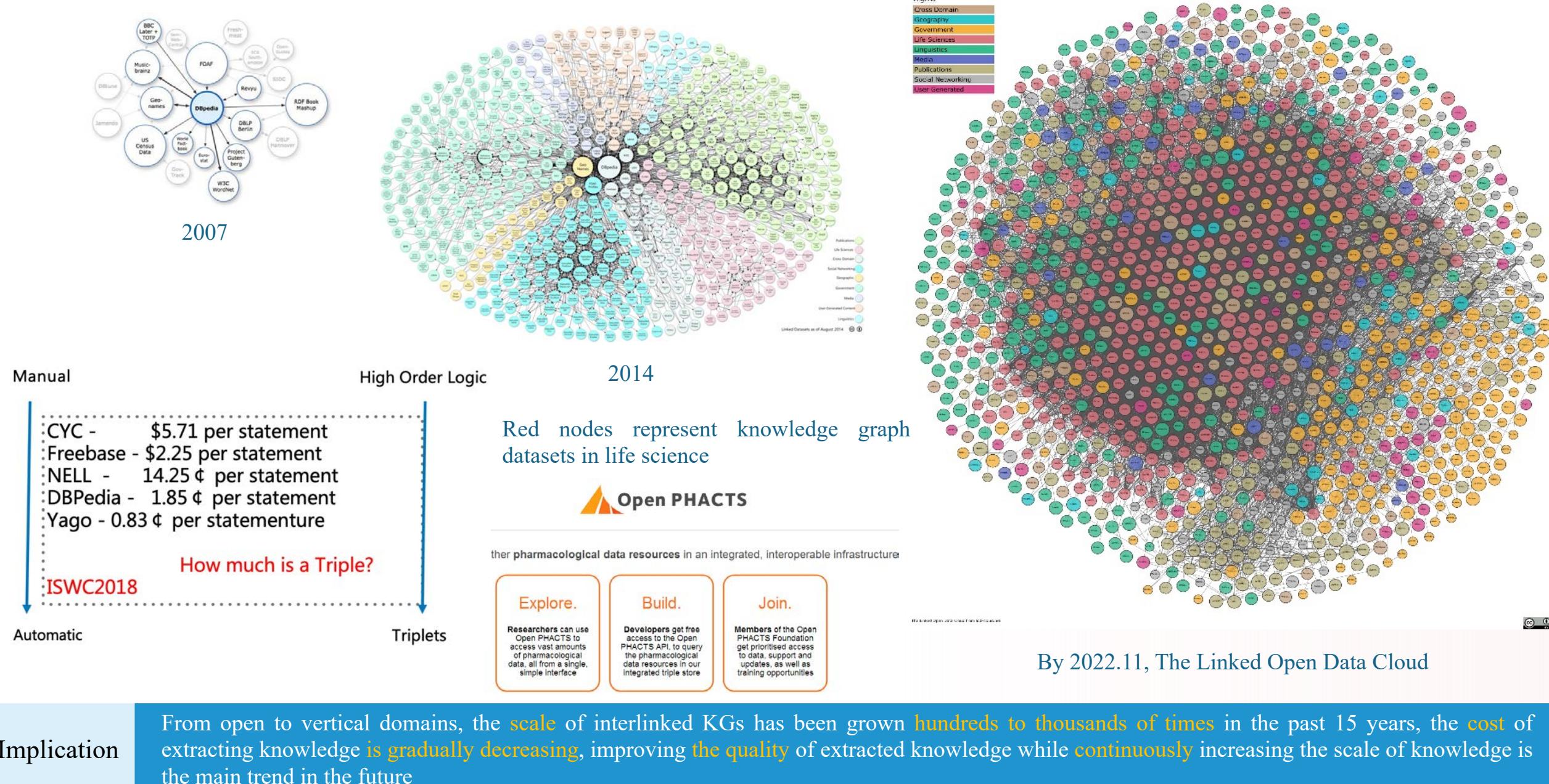


- General domain oriented
- Commonsense knowledge
- Structured encyclopedia knowledge
- Emphasize the breadth of knowledge
- For general users

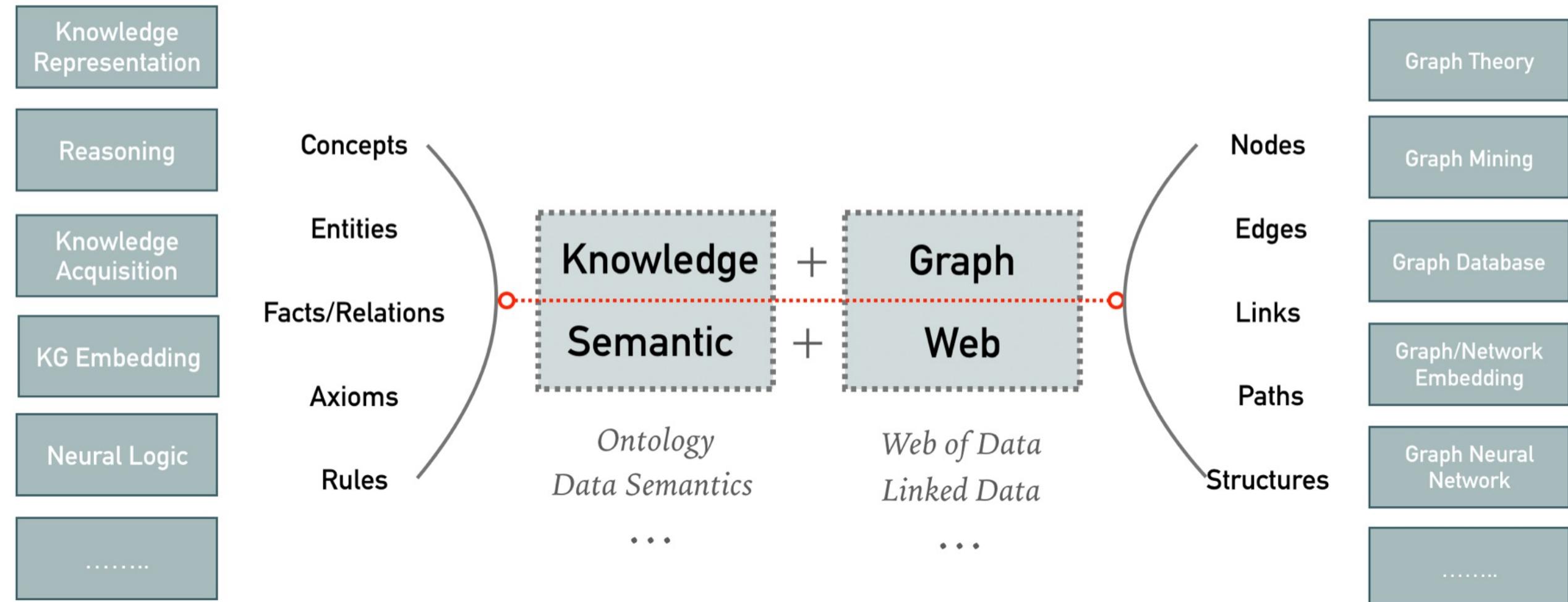


- Industrial domain oriented
- Industrial data
- Semantic industrial knowledge base
- Emphasize the depth of knowledge
- For industry users

# What is Knowledge Graph (KG) – Rapid Growth and Lower Cost



*Knowledge Graph* is more expressive than *pure Graph* but less complex than *formal logic*.



# What is Knowledge Graph (KG) – Perspective and Implication

## KG as a World Model

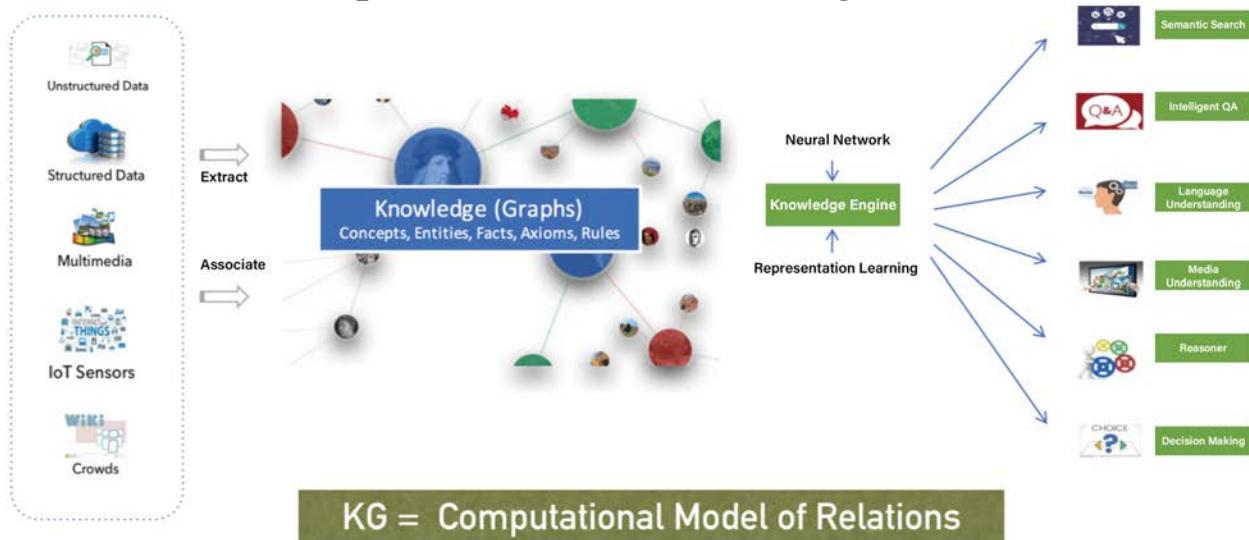


The Good Old Fashioned AI

The Semantic Web & Linked Knowledge

The Knowledge Graph

## Graph Structure as Knowledge Base



## Implication

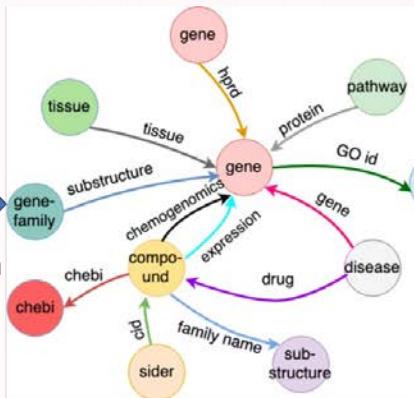
Knowledge Graph originates from how machines represent knowledge, use the graph structure to describe the relationship between things, developed in the rise of Web technologies, and landed in application fields such as search engine, intelligent QA, and recommender systems.

## Text as Knowledge Base



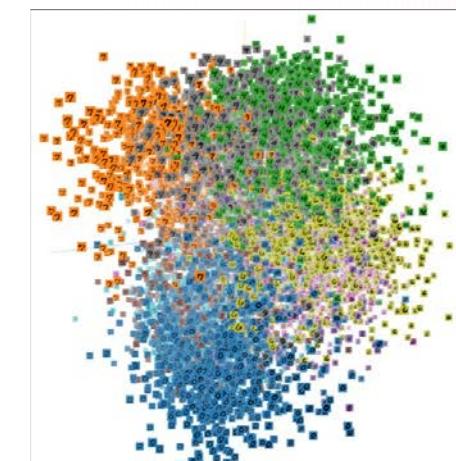
Text2Knowledge

Language Pretrain  
BERT, GPT.....

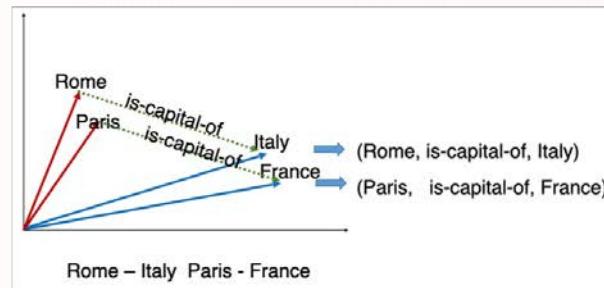


## Embeddings as Knowledge Base

Embeddings : Distributed Vector Representation



- Text : Learn a vector of each word in a sentence
- KG: Learn a vector for each entity or property
- Image/Video : Learn a vector for each visual object



# SOTA and Trend of KG – Knowledge Representation and Reasoning

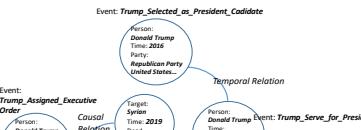
## Entity Centric KG



Barack Obama  
President of the United States  
Michelle Obama's Husband

Triples in form of (s, p, o)

## Event Centric KG



Event Frames  
Event Relations

Causal relation, Temporal relation, Co-reference relation, Sub-class relation...

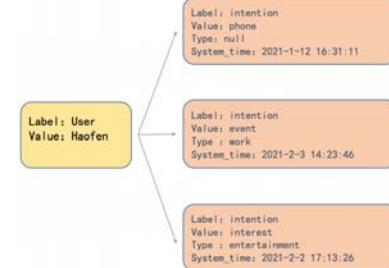
## Prompt KG

```

1 Translate English to French:
sea otter => loutre de mer
2
3 peppermint => menthe poivrée
4 plush girafe => girafe peluche
5 cheese =>
    
```

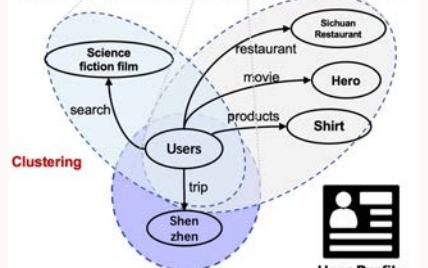
LLM as knowledge base and use prompts to elicit the corresponding facts

## Personal KG



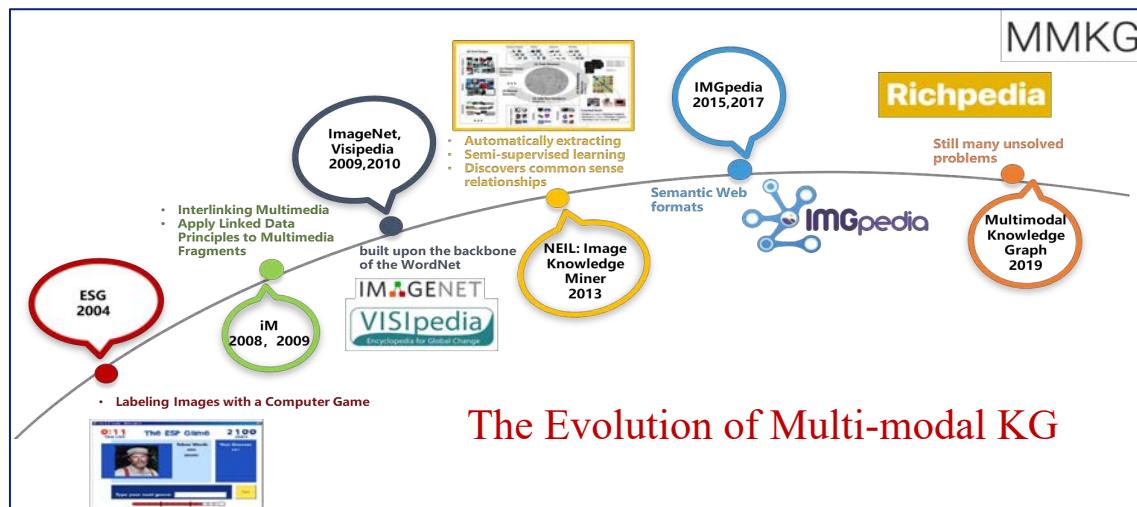
Graph composing events, interests, behaviors of an individual under the protection of privacy and security

## Spatiotemporal KG

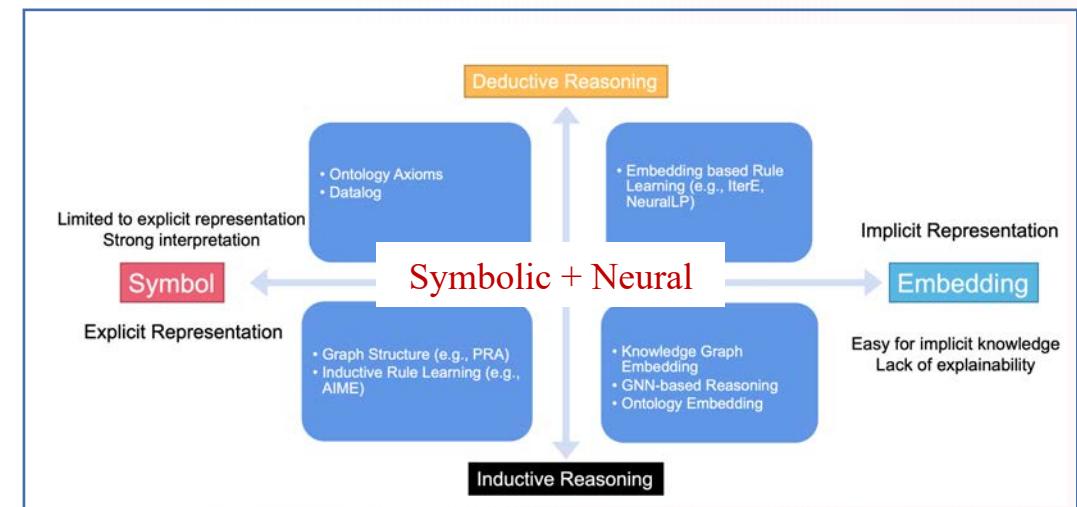


Expand knowledge representation level in time series and space dimensions

Knowledge types: simple -> complex, static -> dynamic, community -> personal, plain -> spatiotemporal



## The Evolution of Multi-modal KG



Challenges

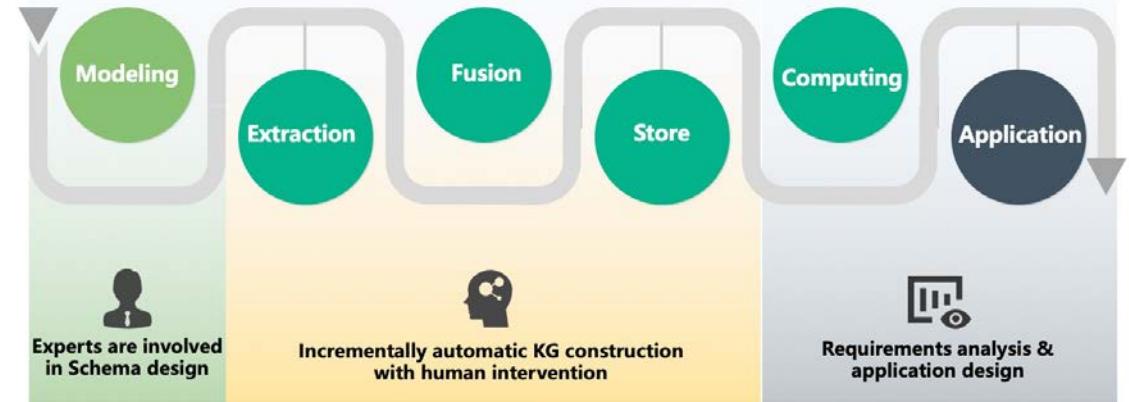
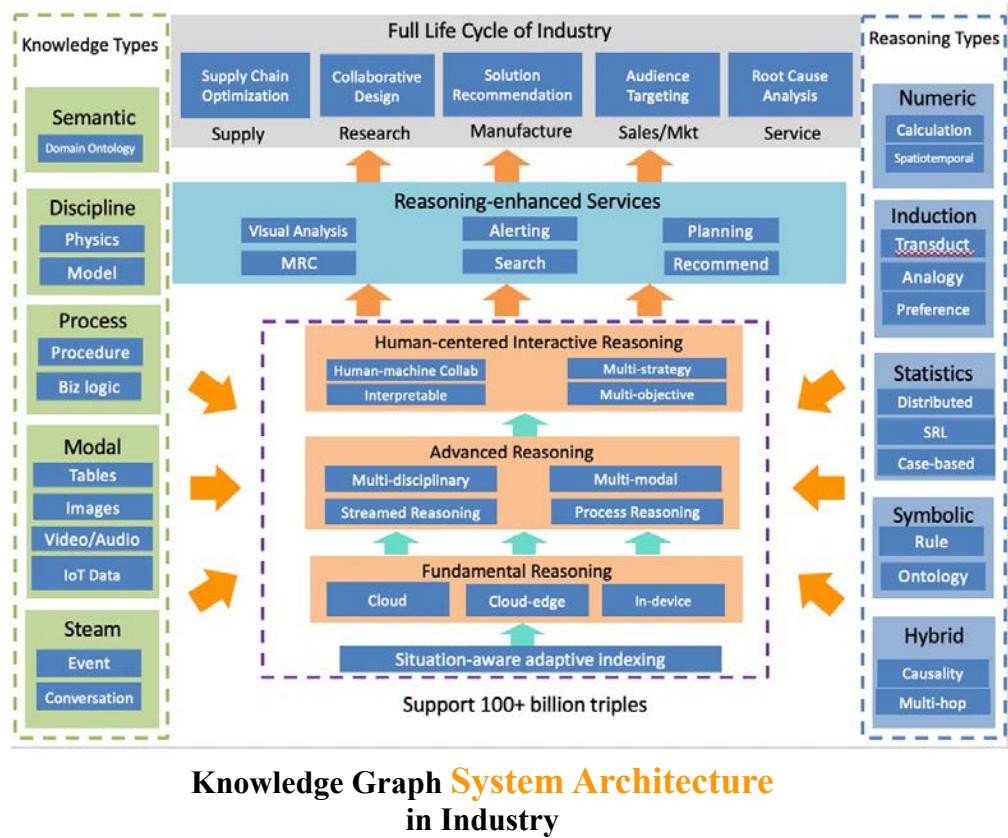
Traditional symbolic knowledge representation methods are difficult to accurately represent complex knowledge such as **dynamics**, **processes**, and **cross-modalities**. At the same time, how to **combine symbolic reasoning** methods based on knowledge graphs and **neural reasoning** methods is extremely challenging.

# SOTA and Trend of KG – System Engineering View



AI should focus on small data and **data centric AI**. Especially in the manufacturing industry, we must rely on **domain knowledge**

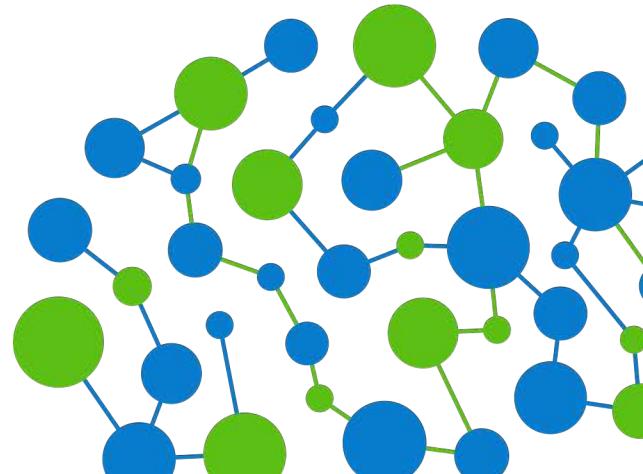
Andrew Ng



New Paradigm of Technology with Knowledge as the Core

Challenges  
Data characteristics and knowledge differences in different fields lead to low knowledge coverage, intensive labor input, shallow usage  
In applications, poor computing efficiency, difficult & weak sustainable operation and long time cost

- The Rise of LLM
- The Development of KG
- The Emerging Paradigm
- OpenKG
- Conclusion and Outlook



# Are symbolism and connectionism mutually exclusive?

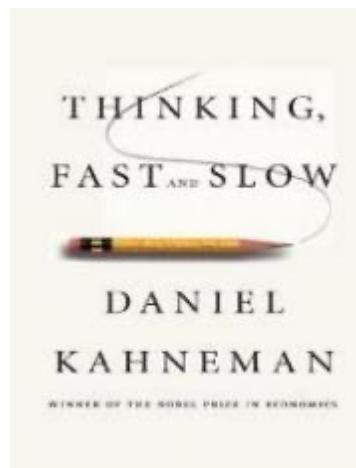
## ■ Cognitive motivation

Symbolism: Humans rely on symbolic systems (such as language) to organize, communicate, reason, and create knowledge

Connectionism: The physical form of the human brain is a widely connected network of neurons

## ■ Dual-process cognitive theory

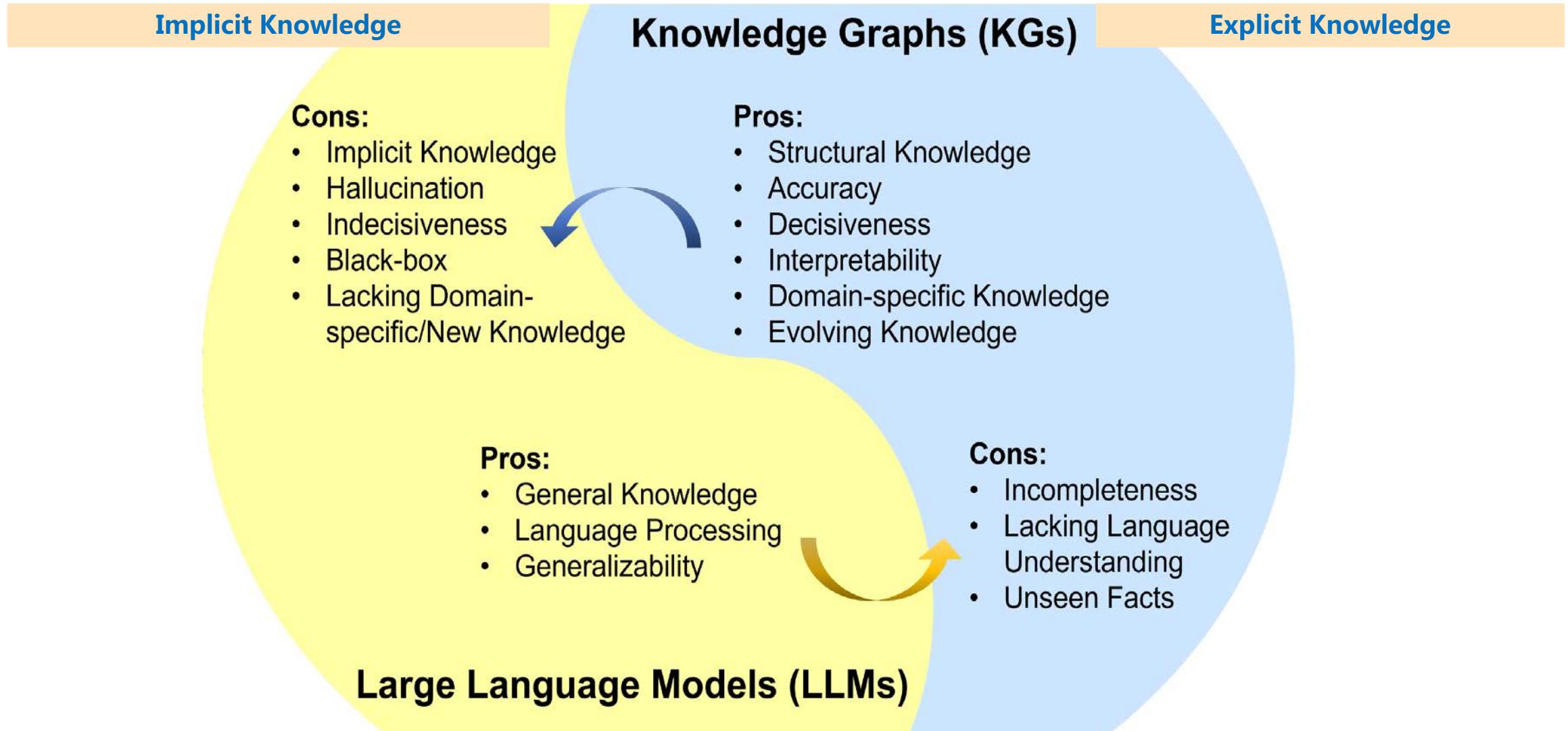
System 1 Intuition (Intuitive)
<ul style="list-style-type: none"><li>• Emotions, memory, and experience</li><li>• Unconscious, capable of quickly activate associated objects</li><li>• Easily deceived</li></ul>



System 2 Rational (Analytic)
<ul style="list-style-type: none"><li>• Logical reasoning and computation</li><li>• Conscious, deliberate, and methodical thinking</li><li>• "Slow thinker"</li></ul>

Connectionist models are the implementation foundation, while symbolic systems are the features of higher-level abilities.

# When KG meets LLM: Two Sides of a Coin



# When KG meets LLM: Reasoning Capability Comparison

## LLM Reasoning

- **Code pre-training:** enhance LLM reasoning during training
- **Prompt Engineering:** eliciting LLM reasoning during inference

## KG Reasoning

- Graph computing
- Rule-based reasoning
- Ontology reasoning
- Spatial-temporal reasoning
- KG embedding/GNN

## LLM Reasoning

- zero-shot prompting
- Few-shot prompting
- CoT prompting
- Instruction

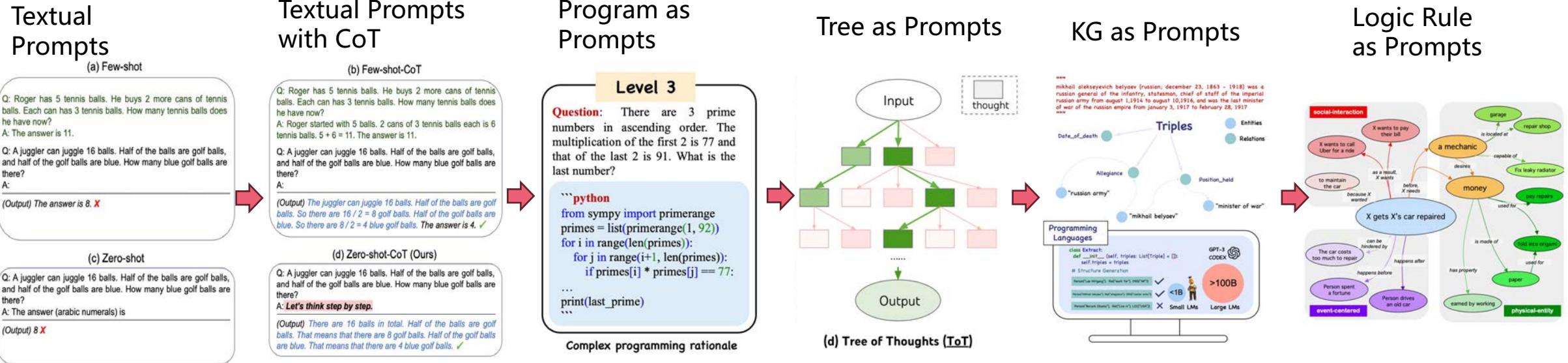


## KG Reasoning

- Graph computing
- Rule-based reasoning
- Ontology reasoning
- Spatial-temporal reasoning
- KG embedding/GNN

# Structured Knowledge vs. Parameterized Knowledge

Extensive practice indicates that the quality and detail of prompt design, as well as the level of structure and logic in prompt representation, significantly impact the model's output. Essentially, **prompt engineering is knowledge engineering**, aiming to acquire prior knowledge from humans to guide model training or activate model reasoning capabilities.



As the level of structure and logic in prompt knowledge gradually increases, the model's reasoning capabilities become stronger, but the difficulty of acquiring prompt knowledge also increases, making large-scale emergence harder to achieve.

Representation vs Reasoning

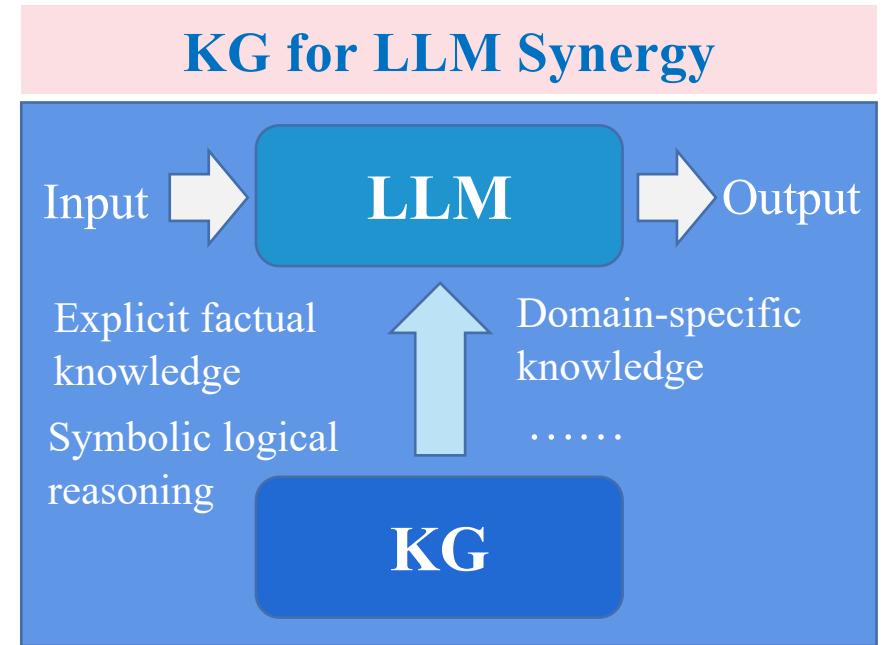
The proportional relationship between the level of representation and reasoning ability still exists.  
The conflicting relationship between representation complexity and scalability still persists.

Representation vs Scale

# Knowledge Graph Helps the Large Model Research

## Knowledge Engineering **FOR** the LLM

- Aiming to address the issues of large models through appropriate knowledge engineering methods (resources and frameworks)
  - Hallucination of factual knowledge
  - Knowledge updating
  - Evaluating emergent symbolic abilities to identify problems
  - Issues of scale and efficiency
  - Logical accuracy and interpretability of complex reasoning processes

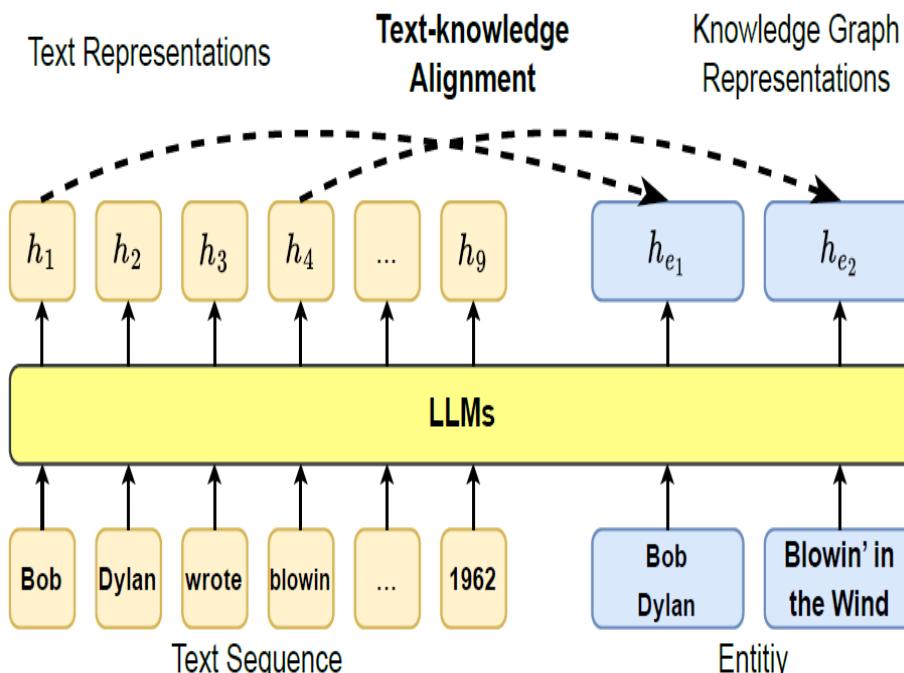


- **Pre-training**
- **Fine-tuning**
- **Inference**
- **Interpretability**
- ...

# KG for LLM : Pre-training

## Designing pre-training objective to incorporate KG components

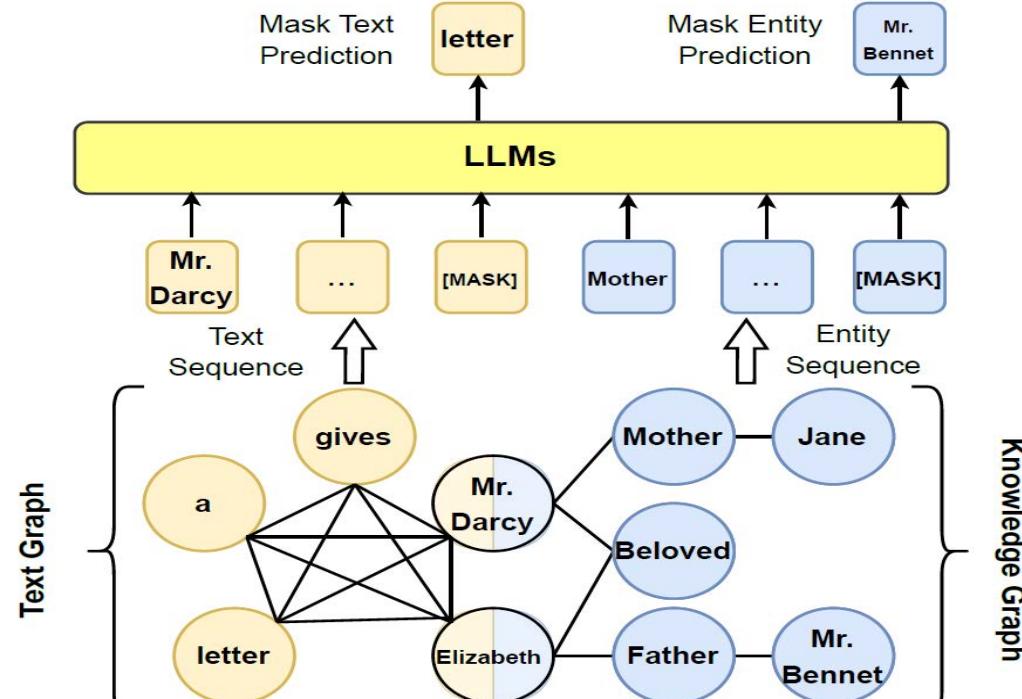
- ERNIE training objective: word-entity alignment



**Input Text:** Bob Dylan wrote Blowin' in the Wind in 1962

## Integrate KG with text as LLM training input

- CoLAKE: word-knowledge graph



**Input Text:** Mr. Darcy gives Elizabeth a letter

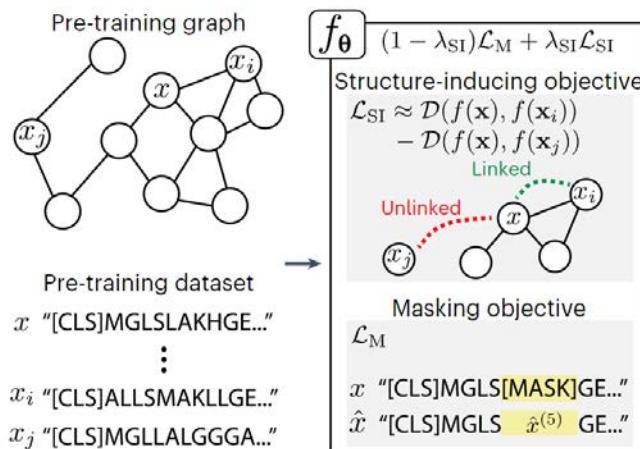
Z. Zhang, X. Han, Z. Liu, et al, “ERNIE: Enhanced language representation with informative entities,” ACL 2019.

T. Sun, Y. Shao, X. Qiu, et al. “CoLAKE: Contextualized language and knowledge embedding,” 2020.

# When KG meets LLM: Knowledge or Structural Augmented

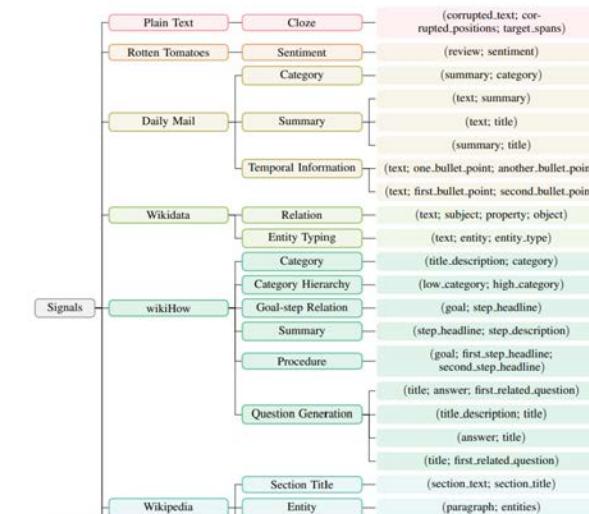
- More research indicates that introducing various types of **internal and external structured signals** during the pre-training phase can improve model performance:
  - External structured signal enhancement:** For example, incorporating Protein-Protein Interaction (PPI) information in the training of protein language models.
  - Internal structured signal enhancement:** For example, expanding the relational information between entities within sentence samples during sentence pre-training.

Introduce  
external  
structured  
signals to  
training corpora



**Fig 1. Structure-inducing Pre-training**  
 as a PT graph  $G_{PT}$  to achieve  $\mathcal{L}_{SI}$ , which pushes a PT encoder  $f_{\theta}$  to embed samples such that samples are close in the latent space if and only if they are linked in  $G_{PT}$ .

- Structure-inducing pre-training. Nature Machine Intelligence 2023.
- reStructured Pre-training. 2022

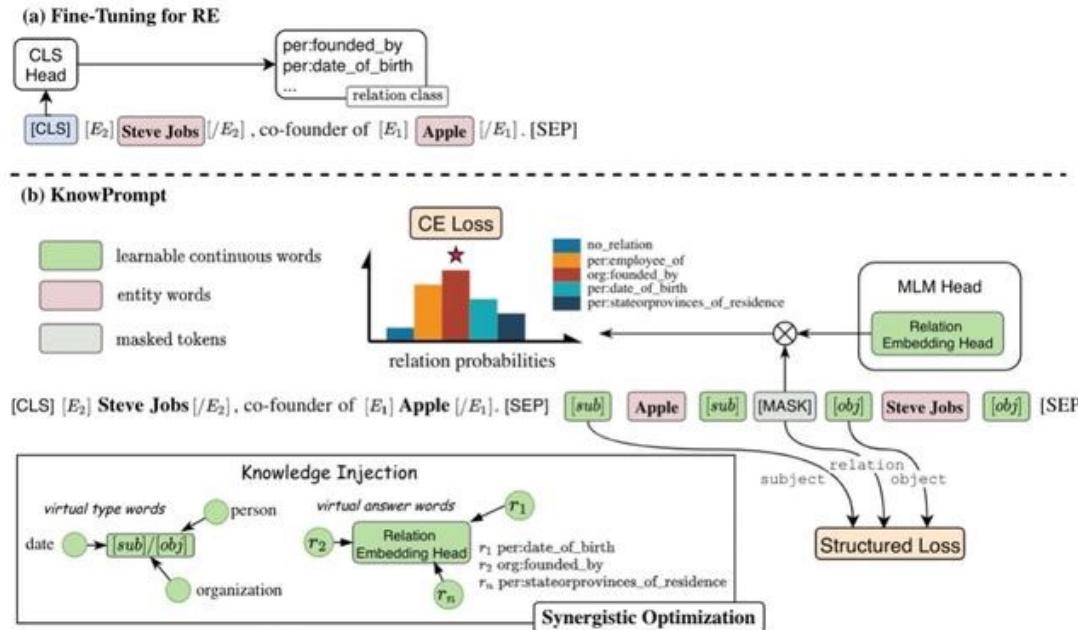


Enhance the  
internal  
structured  
signals of  
training corpora

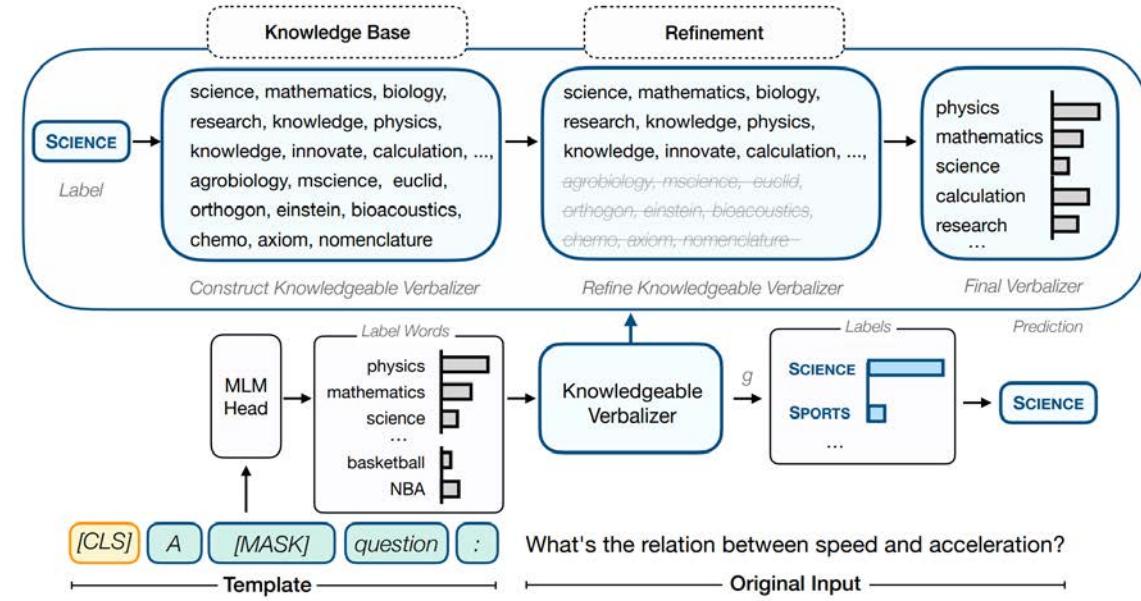
**reStructured Pre-training**



# When KG meets LLM: Knowledgeable Prompt Tuning



KnowPrompt

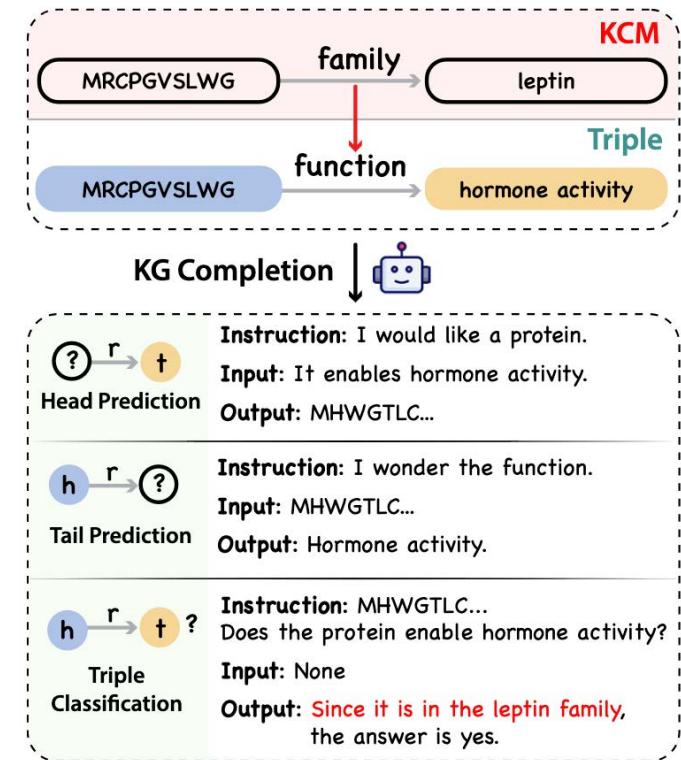
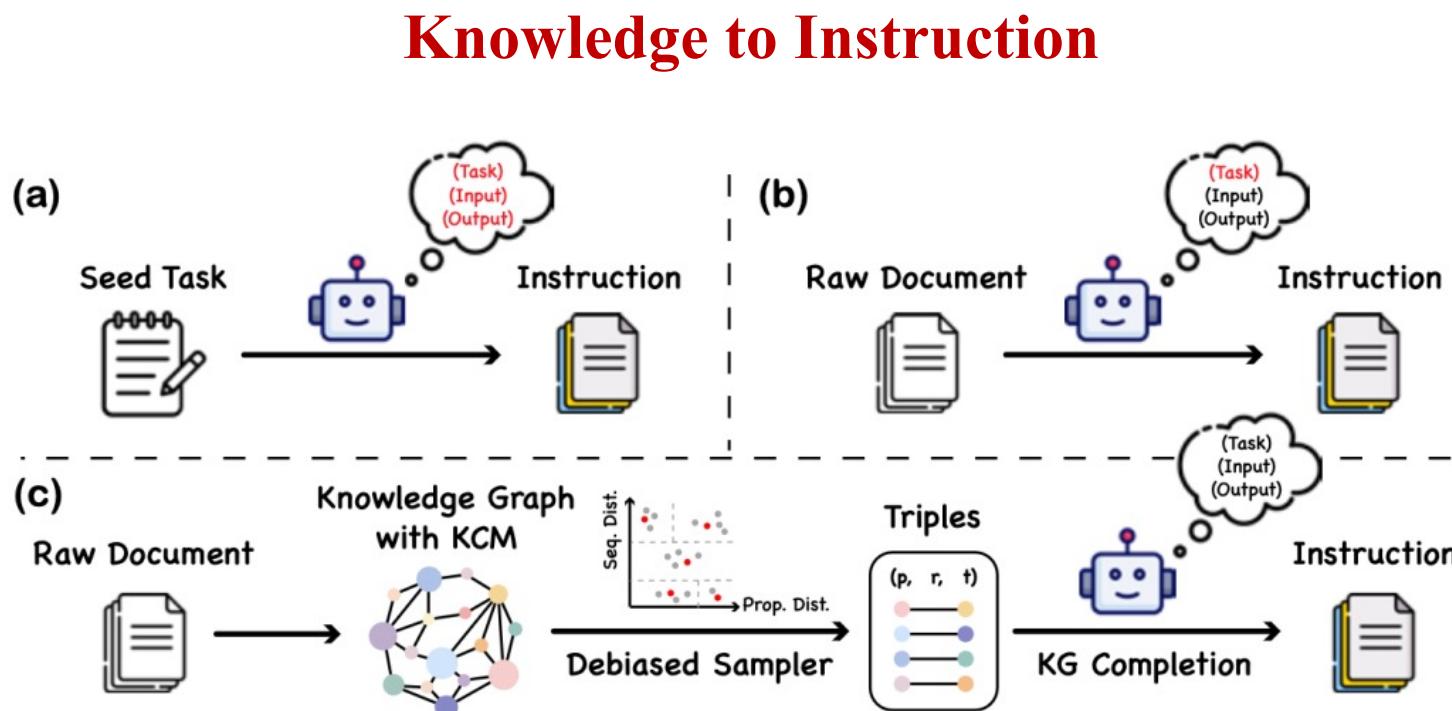


Knowledgeable Prompt-tuning

- KnowPrompt: Knowledge-aware Prompt-tuning with Synergistic Optimization for Relation Extraction. WWW 2022
- Knowledgeable Prompt-tuning: Incorporating Knowledge into Prompt Verbalizer for Text Classification. ACL 2022

# When KG meets LLM: Instruction Construction

- Knowledge graphs can also guide the construction of instruction datasets. For example, the triples in the graph can guide the setting of Input/Output for instruction templates, and the association signals in the graph can be used to ensure that the instruction dataset has better logicality, relevance, and structure.

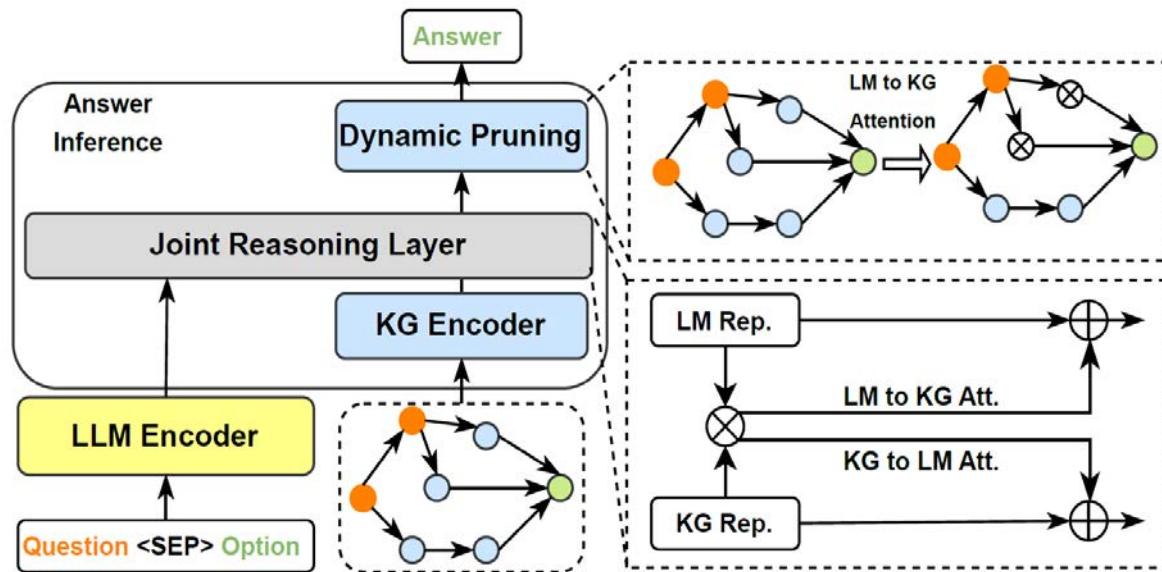


# KG for LLM : Inference

## Problems during the pre-training phase: Knowledge updating

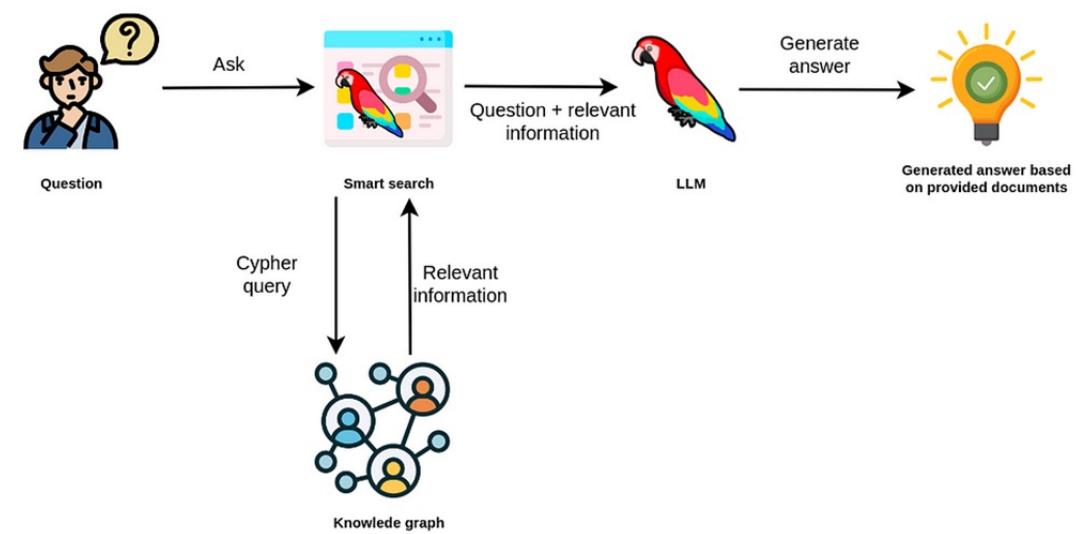
### Dynamic KG integration method

- JointLK: LM-to-KG and KG-to-LM bidirectional attention mechanism



### Retrieval-augmented KG integration method

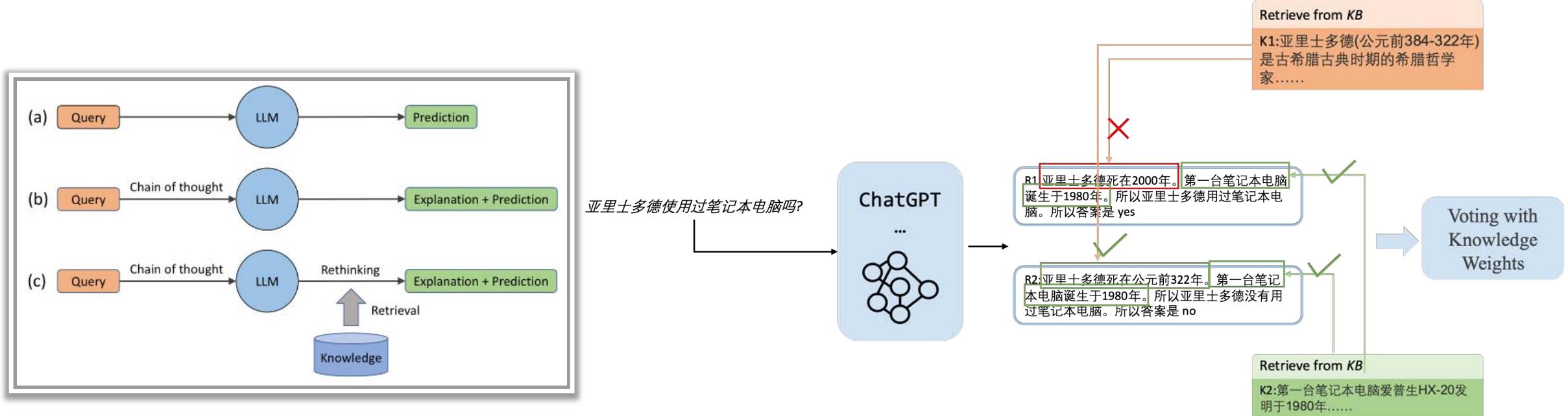
- RAG: Retrieve external KG as LLM input



- Y. Sun, Q. Shi, L. Qi, and Y. Zhang, “JointLK: Joint reasoning with language models and knowledge graphs for commonsense question answering,” 2022.
- P. Lewis, E. Perez, A. Piktus, et al, “Retrieval-augmented generation for knowledge intensive nlp tasks,” NeurIPS 2020.

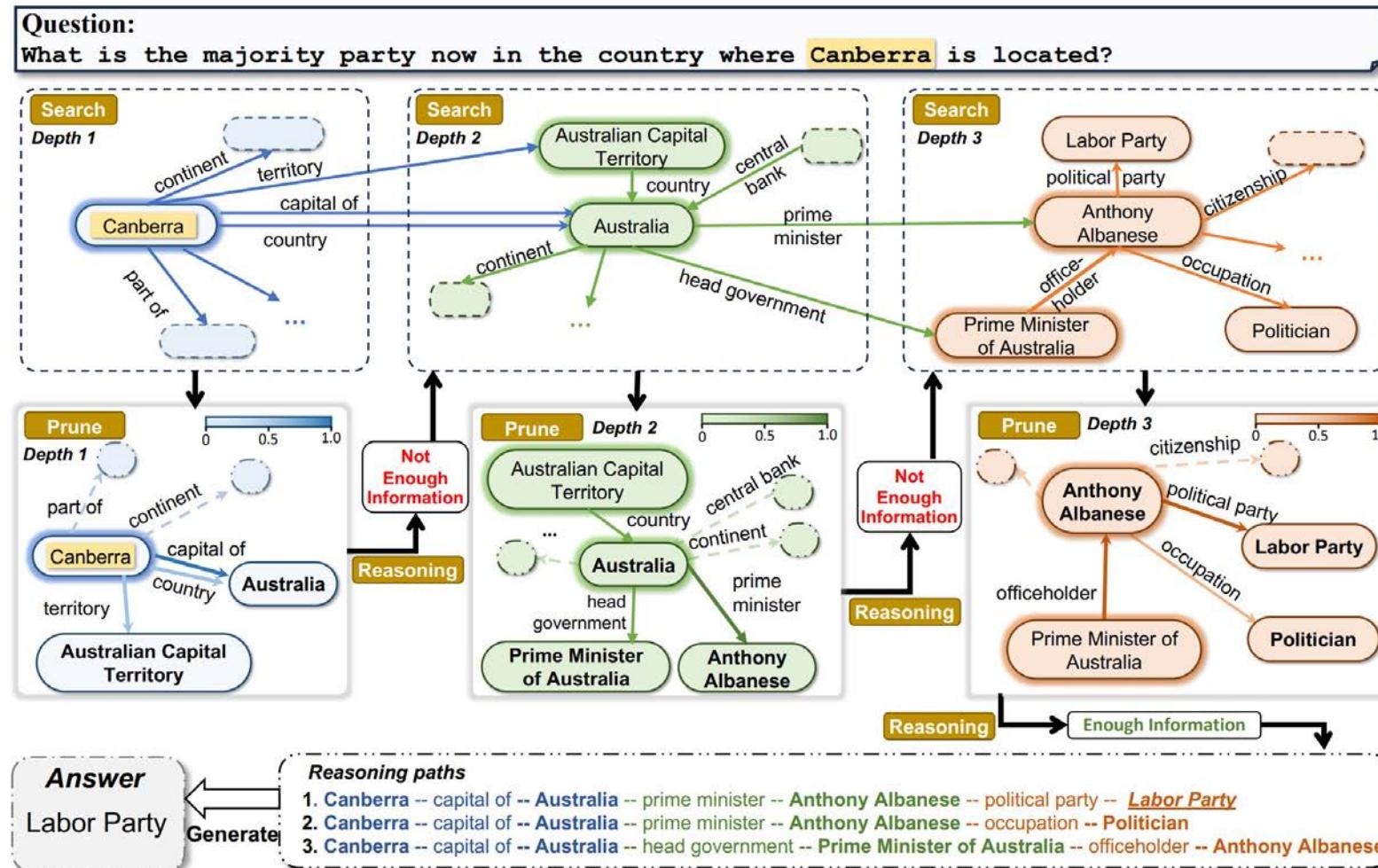
# KG for LLM : Retrieval/Knowledge Augmented Generation

- Using a knowledge graph as an external knowledge base during the inference of pre-trained models
- For example: Rethinking .....



# When KG meets LLM: KG-induced CoT

- Utilizing entity association paths from external knowledge graphs to guide LLM in question decomposition, improving the accuracy of multi-hop question answering.



- Think-on-graph: Deep and responsible reasoning of large language model with knowledge graph. ICLR 2024.
- Graph Chain-of-Thought: Augmenting Large Language Models by Reasoning on Graphs. 2024

# KG for LLM : Interpretability

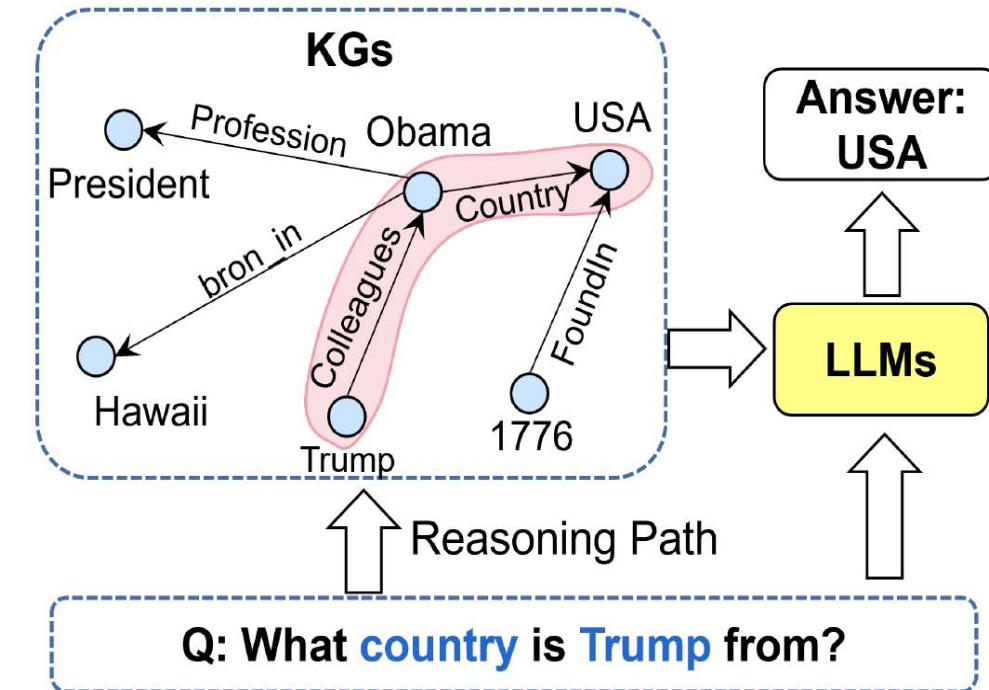
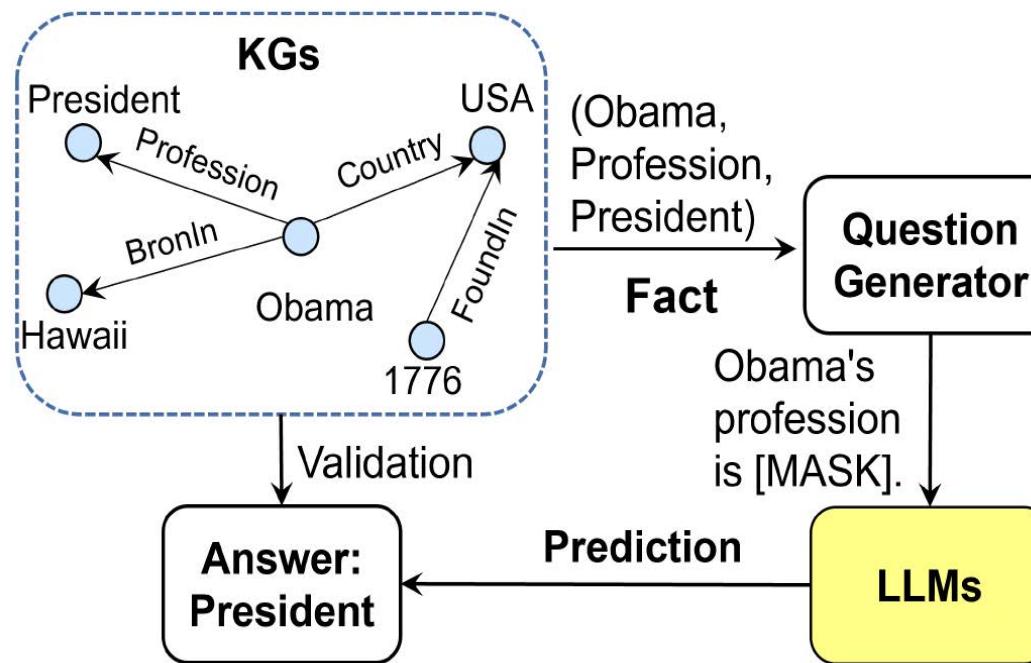
An important issue with LLMs: Lack of explainability, affecting applications in critical

## KG-based LLM probing and evaluation methods

- LAMA: Converts knowledge from KG into prompt templates and uses LLM predictions to verify the accuracy of LLM knowledge

## KG-based analysis of LLM-generated results

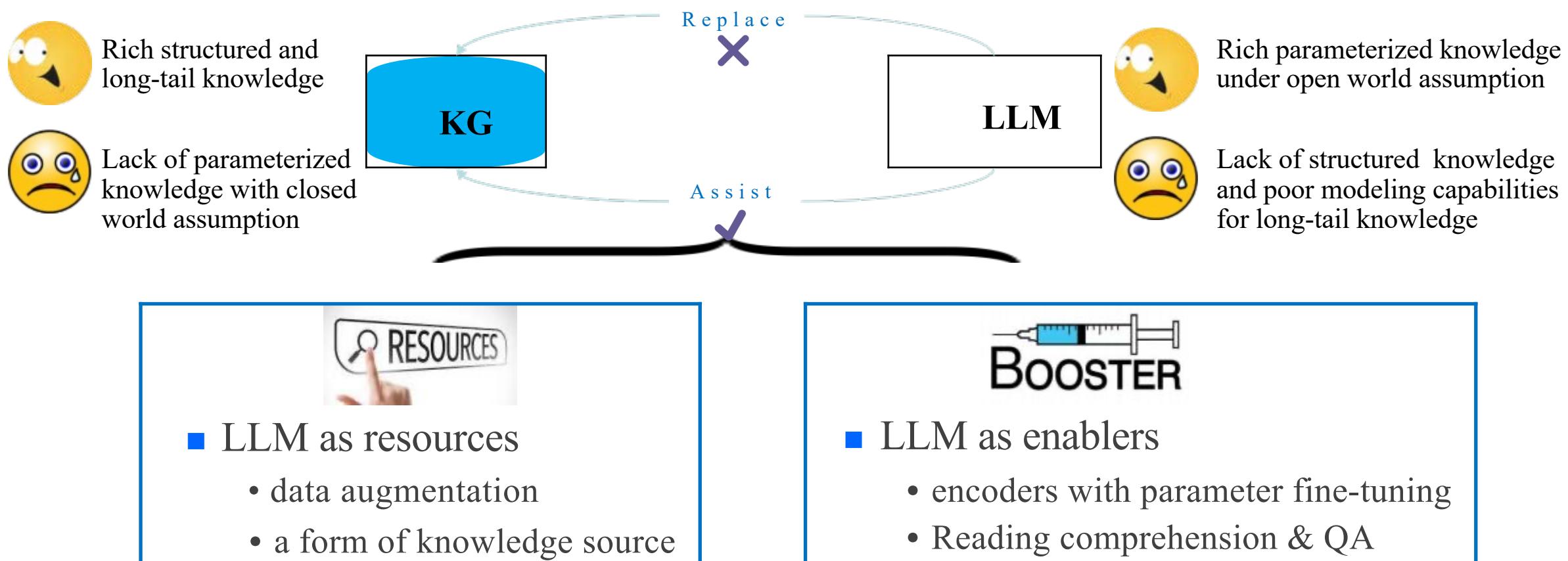
- KagNet, QA-GNN: LLM outputs are explained by KG graph structures



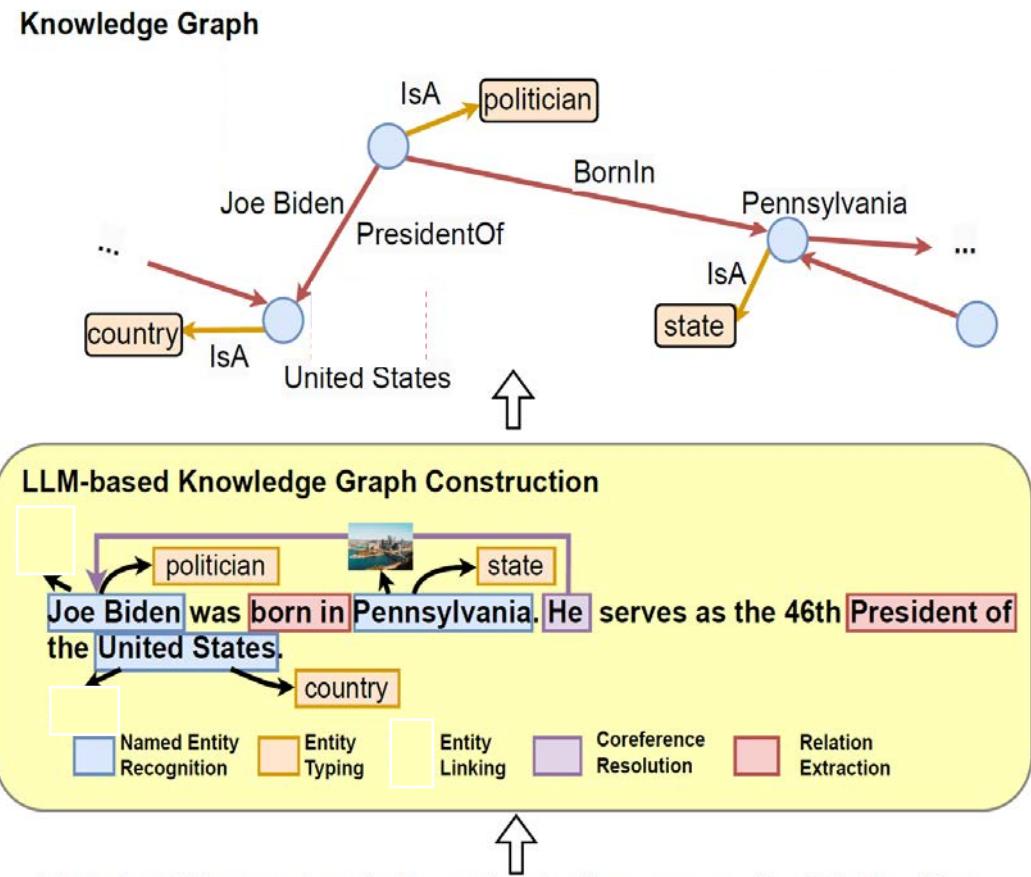
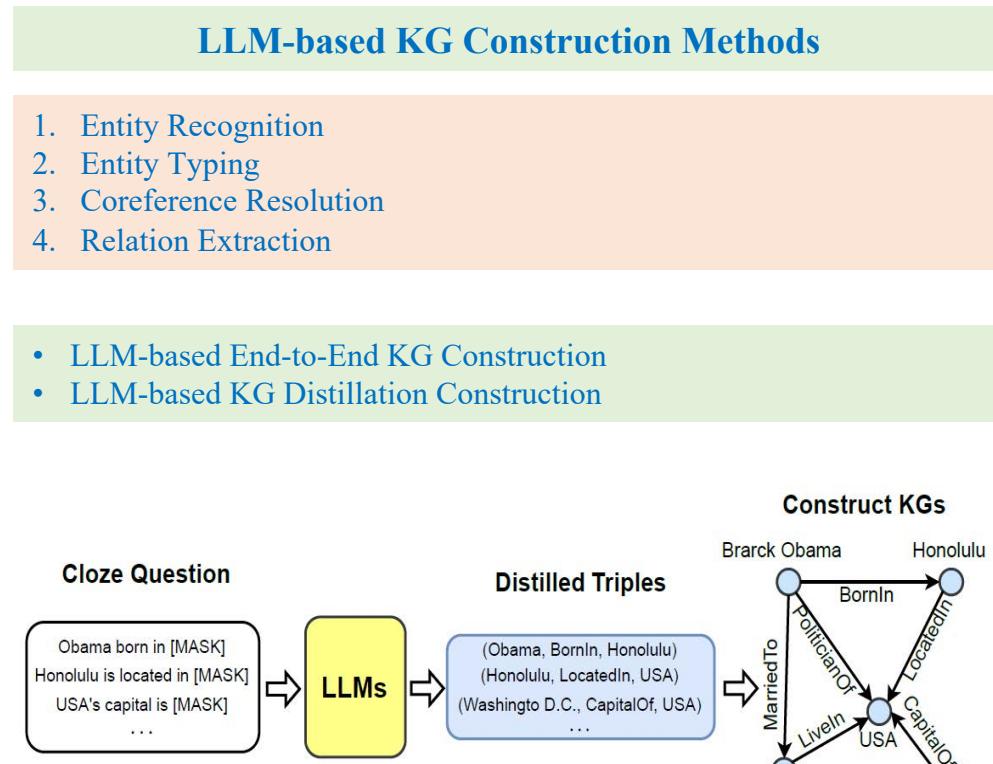
- F. Petroni, T. Rocktaschel, S. Riedel, et al, “Language models as knowledge bases?” EMNLP 2019.
- B. Y. Lin, X. Chen, J. Chen, and X. Ren, “Kagnet: Knowledge-aware graph networks for commonsense reasoning,” EMNLP 2019
- M. Yasunaga, H. Ren, A. Bosselut, et al, “QAGNN: Reasoning with language models and knowledge graphs for question answering,” NAACL 2021

# Leveraging Large Models in Knowledge Graph Research

## ■ Knowledge Engineering **BY** the LLM

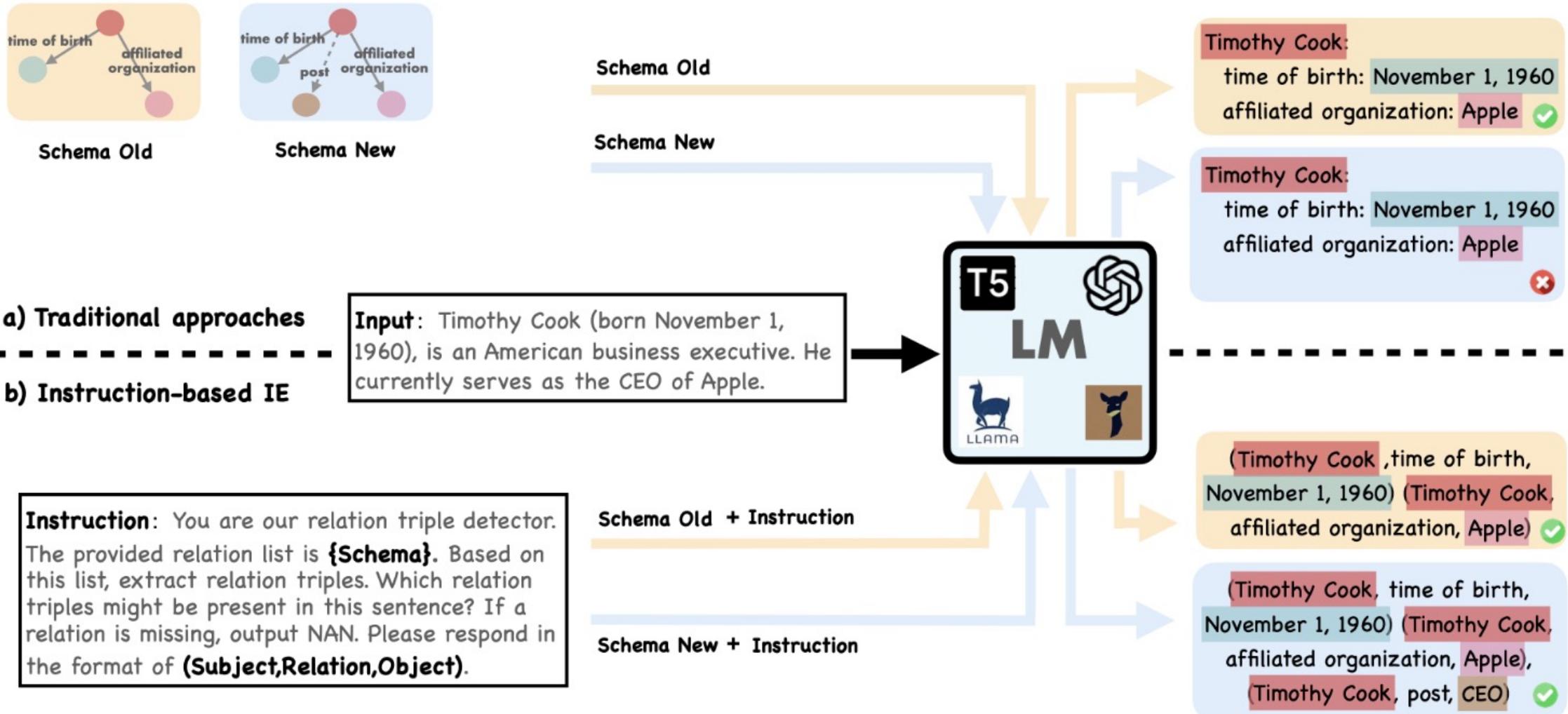


# LLM for KG : Knowledge Extraction

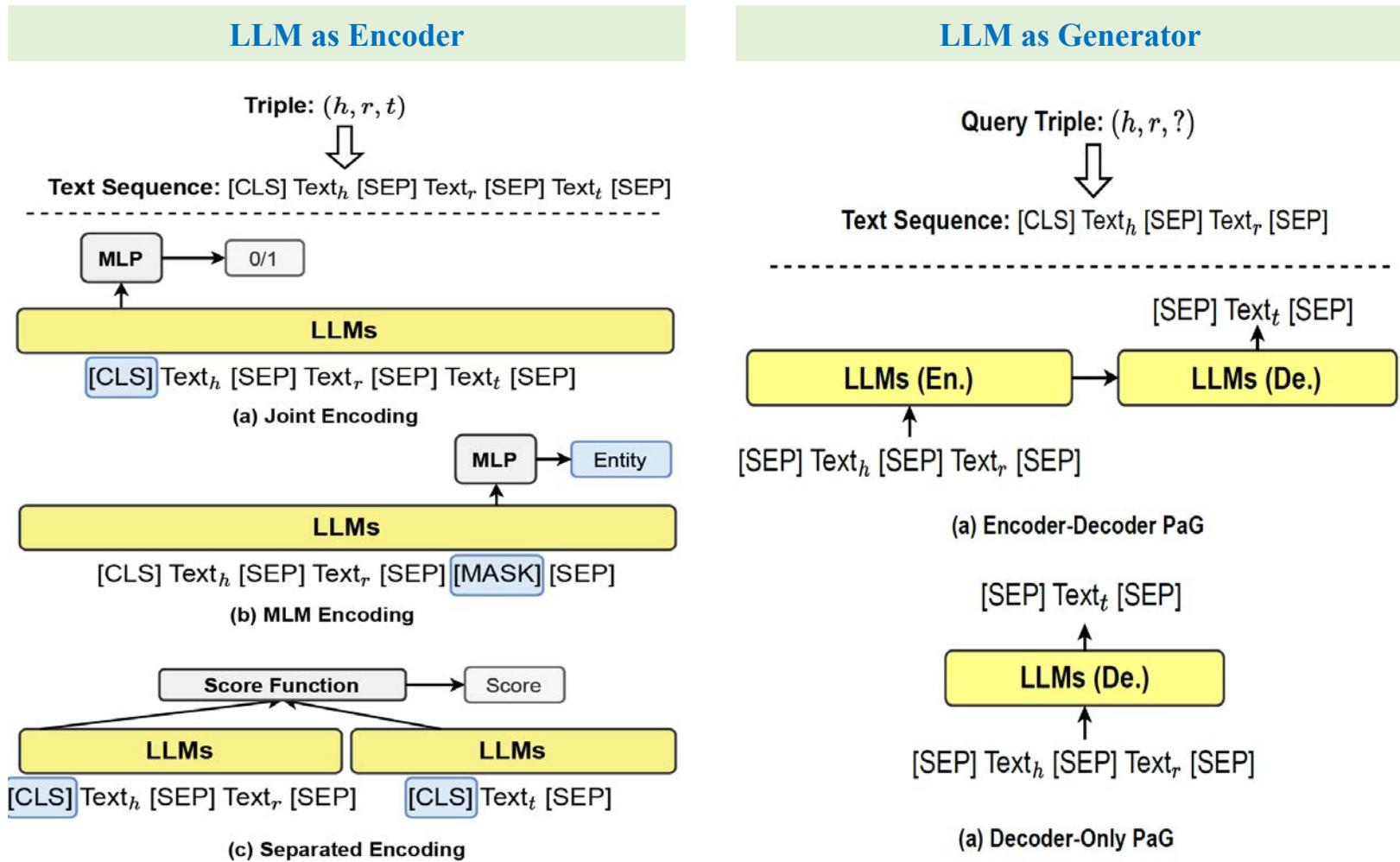


- A. Kumar, A. Pandey, R. Gadia, et al, “Building knowledge graph using pre-trained language model for learning entity-aware relationships,” 2020.
- A. Bosselut, H. Rashkin, M. Sap, et al, “Comet: Commonsense transformers for knowledge graph construction,” ACL 2019.

# LLM for KG : Knowledge Extraction



# LLM for KG : Knowledge Completion

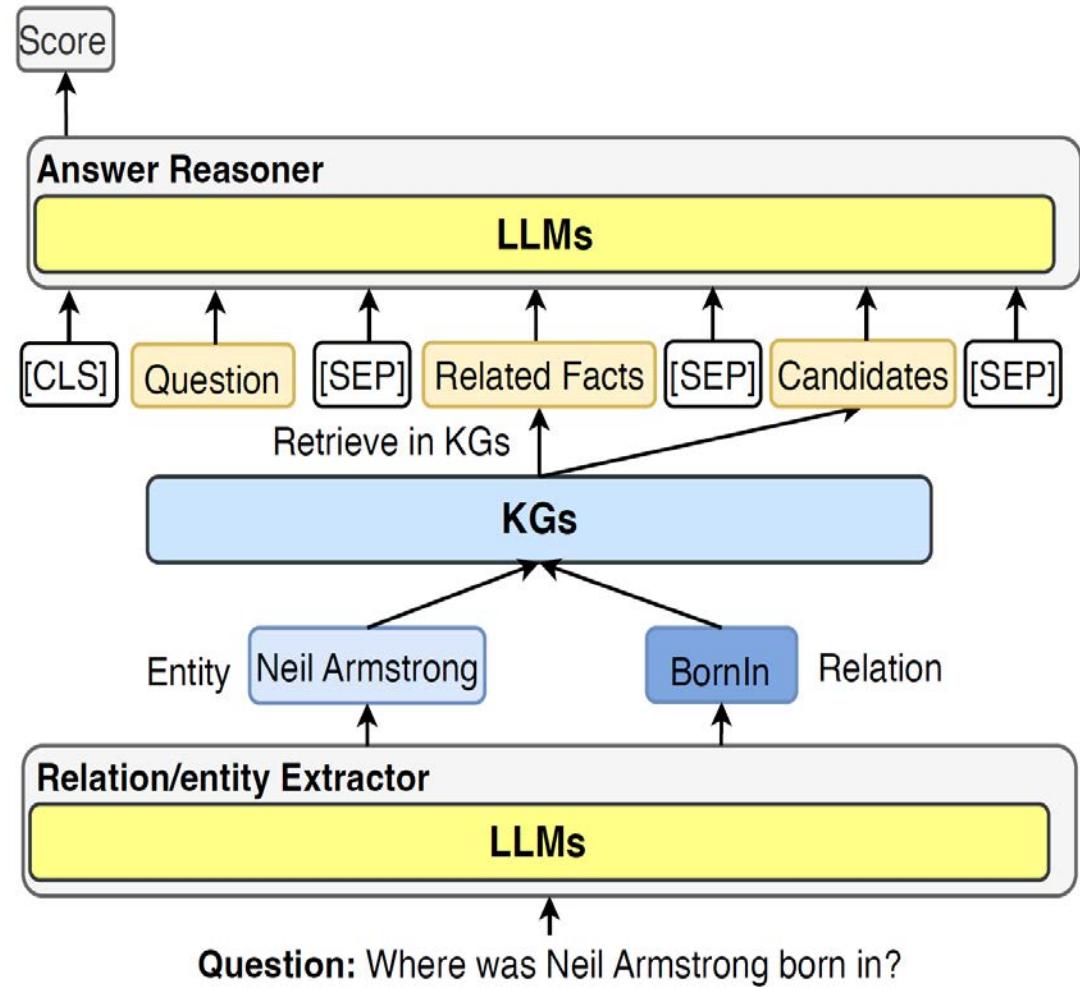


# LLM for KG : Knowledge-based Question Answering

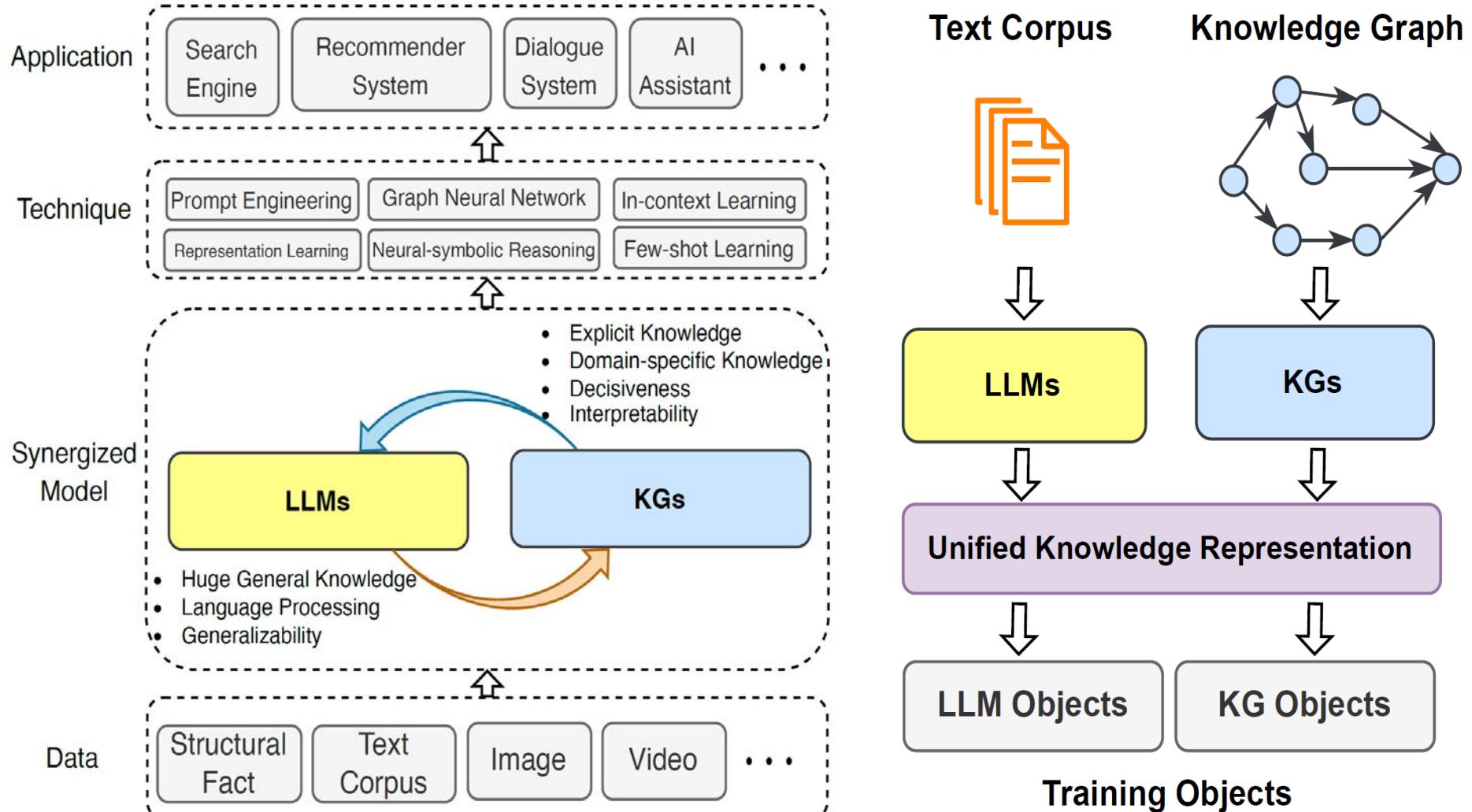
LLM as Entity/Relation extractor

LLM as Answer reasoner

- D. Lukovnikov, A. Fischer, and J. Lehmann, “Pretrained transformers for simple question answering over knowledge graphs,” ISWC 2019
- Y. Xu, C. Zhu, R. Xu, Y. Liu, M. Zeng, and X. Huang, “Fusing context into knowledge graph for commonsense question answering,” ACL 2021
- M. Zhang, R. Dai, M. Dong, et al “DrLk: Dynamic hierarchical reasoning with language model and knowledge graph for question answering,” EMNLP 2022
- Y. Yan, R. Li, S. Wang, et al, “Large-scale relation learning for question answering over knowledge bases with pre-trained language models,” EMNLP 2021



# When KG Meets LLM: Deep Synergies



# When KG meets LLM: Interoperability & Orchestration

KG uses ontology to realize the representation of domain data, knowledge and interaction, and completes the automation of the whole process from real-time data access, knowledge update to user interaction.

## Interoperate & Orchestrate

Integrate the following steps into a whole end-to-end process for hyper-automation

**Data Access**  
Real-time, dynamic, or temporal-spatial data access

**Knowledge Update**  
Update knowledge via extraction by LLM

**Reasoning**  
Combine reasoned results from KG with those LLM responses

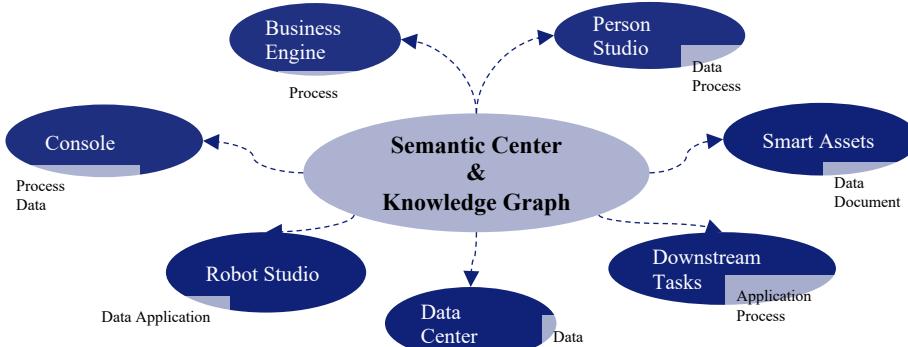
**Act**  
· User: chat, speak, recommend  
· System: API invoke, send commands



- Task mining
- Consistency checking
- Process discovery
- Modeling via simulation
- Case prediction
- Suggestion via history data

**Procedure**  
1 \_\_\_\_\_  
2 \_\_\_\_\_  
3 \_\_\_\_\_  
4 \_\_\_\_\_

# When KG meets LLM: Interoperability & Orchestration



## RPA + Tool use

**New Task (Physics QA)** Answer this high-school physics question  
**Input:** Hector yanks on the chain with a 72.0 N force at an angle of 35.0° above the horizontal. Determine the horizontal components of the tension force.

**Task Library:**

```

graph LR
    CO[Code operations] --> A(( ))
    CR[Co-T-style reasoning] --> A
    S[Search] --> A
    AR[Arithmetic] --> A
    SO[String operations] --> A
    A --> TL[TASK LIBRARY]
    
```

**Tool Library:**

**B** **TOOL LIBRARY** **Google**

**Input:** Hector yanks on the chain with a 72.0 N force at an angle of 35.0° above the horizontal. Determine the horizontal components of the tension force.

**Q1:** [search] What is the formula for the horizontal component of the tension force?  
**#1:** The formula for the horizontal component of the tension force is  $T\cos\theta$ . The horizontal component ( $F_x$ ) can be calculated as  $F_x = T\cos\theta$  where  $\theta$  is the angle which the force makes with the horizontal in radians.

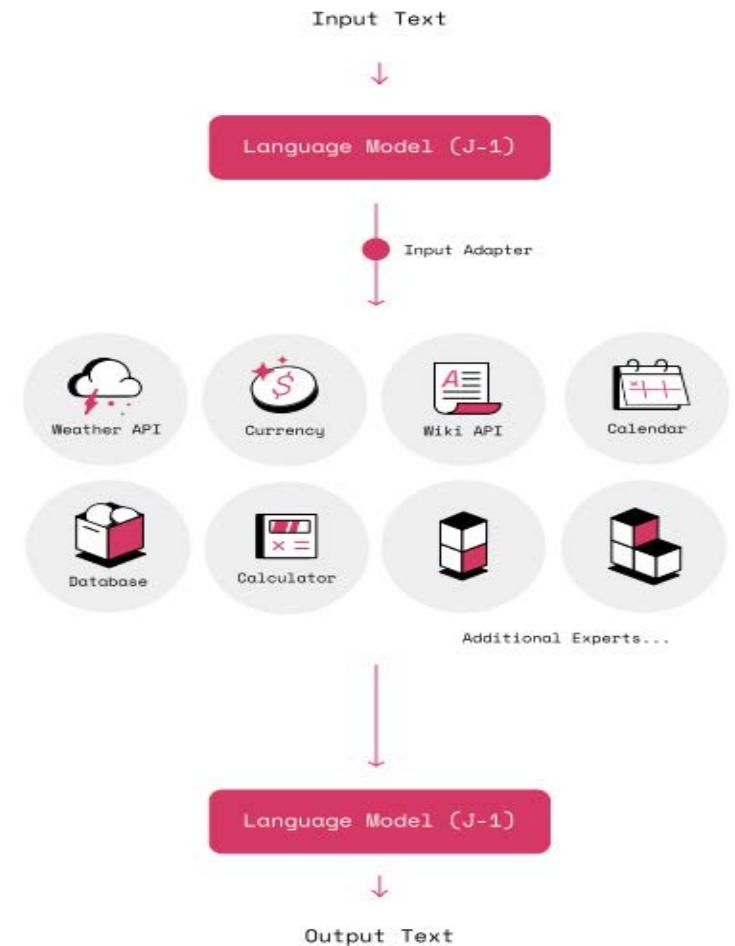
**LLM**

**OpenAI Codex**

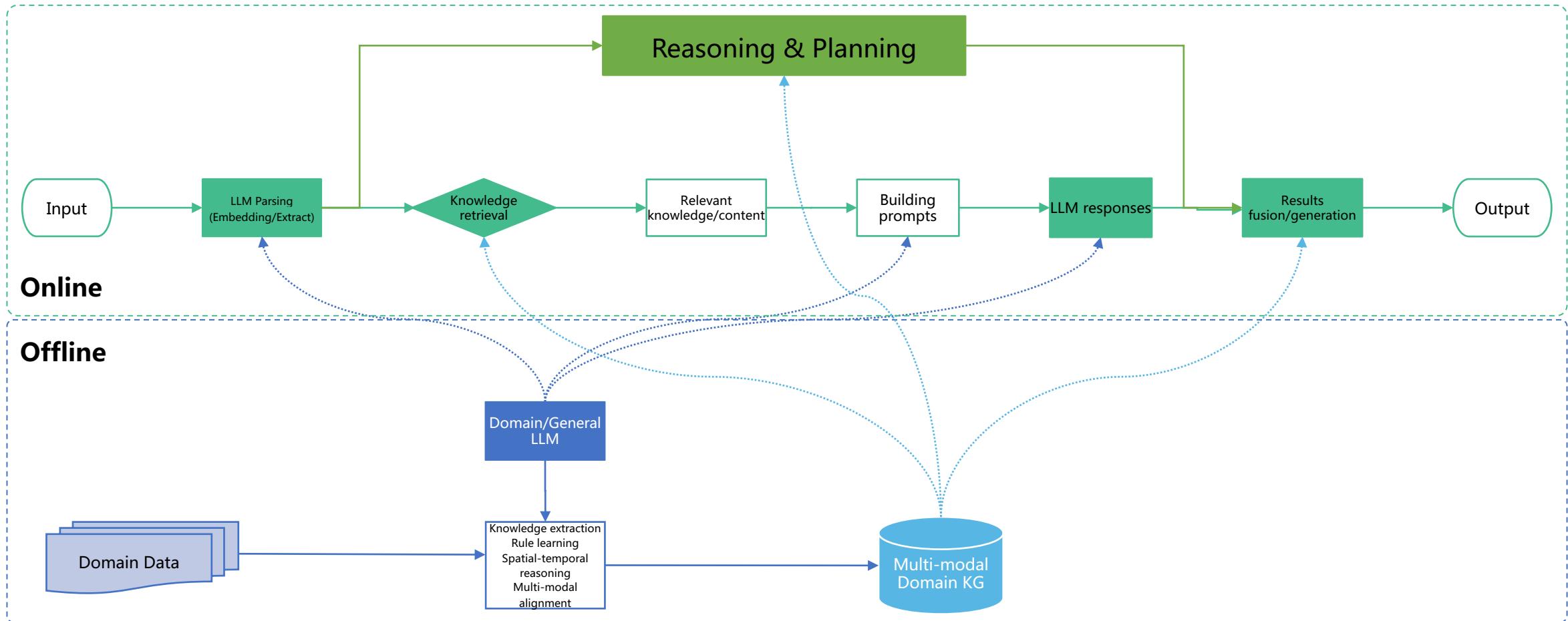
**Python**

**Output:** Does the sentence contain an anachronism? Yes/No.  
**Input:** President George H. W. Bush called his generals at the outset of the Gulf War.  
**Q1:** [search] When was President George H. W. Bush, president?  
**#1:** George H. W. Bush's tenure started on January 20, 1989, and ended on January 20, 1993.  
**Q2:** [search] When was the Gulf War fought? **#2:** The Gulf War was a 1990–1991  
**Q3:** [subquestion] Could these entities have co-existed? **#3:** Yes. Their time periods intersect.  
**Q4:** [generate output] Is this an anachronism? **#4:** No **Q5:** [EOQ] **Ans:** No

## Adaptive Routing



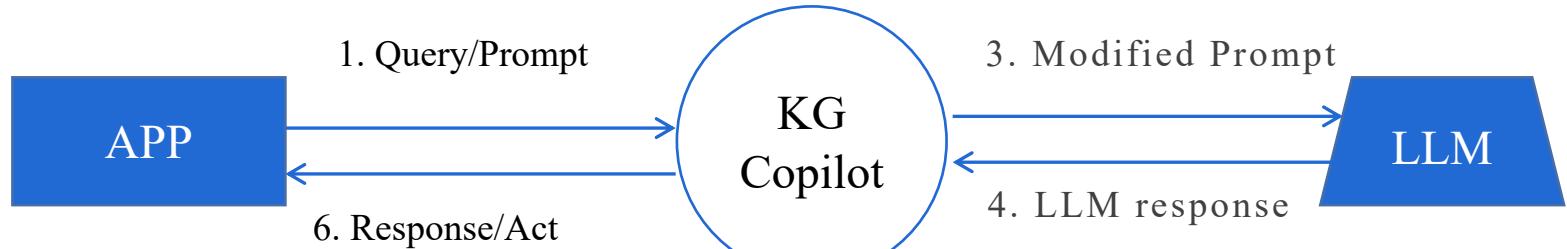
# Retrieval/Knowledge Augmented Generation Whole Picture



# KG Copilot Engine

- ❖ KG for unified knowledge modeling

- ❖ Multi-modal data
- ❖ Multi-source data



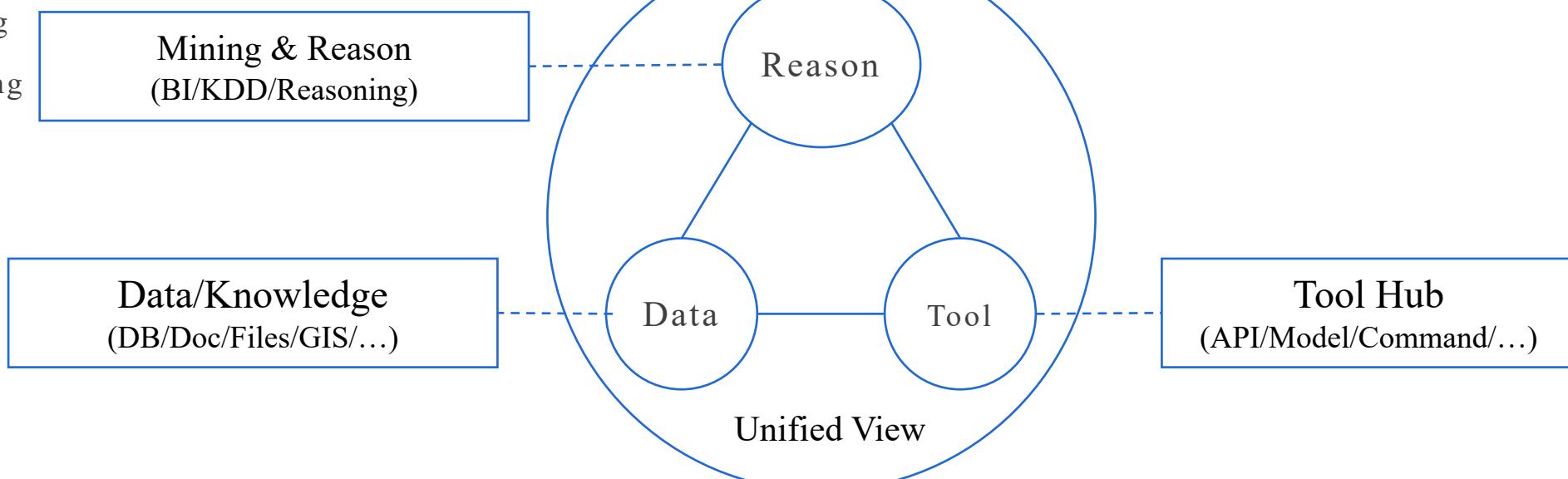
- ❖ KG for unified programming as a semantic gateway

- ❖ LLM + KG + Tool

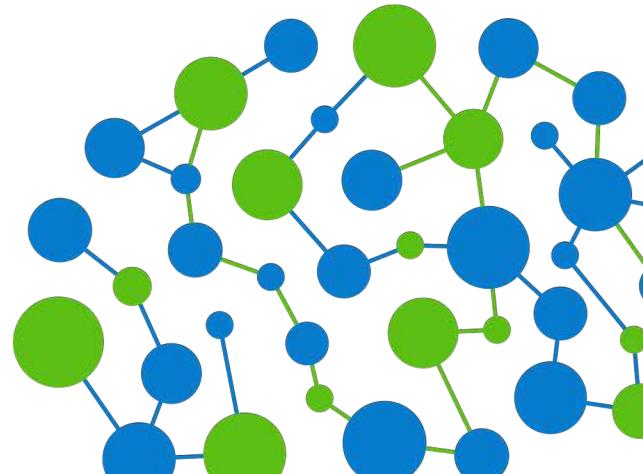
- a. Prompt build
  - b. refinement
2. Pre-processing
  5. Post-processing
- a. Fact checking
  - b. Data query
  - c. Tool invoke
  - d. Reasoning
  - e. Result fusion

- ❖ KG for automatic prompt building and result checking

- ❖ Pre-processing grounding
- ❖ Post-processing grounding



- The Rise of LLM
- The Development of KG
- The Emerging Paradigm
- OpenKG
- Conclusion and Outlook

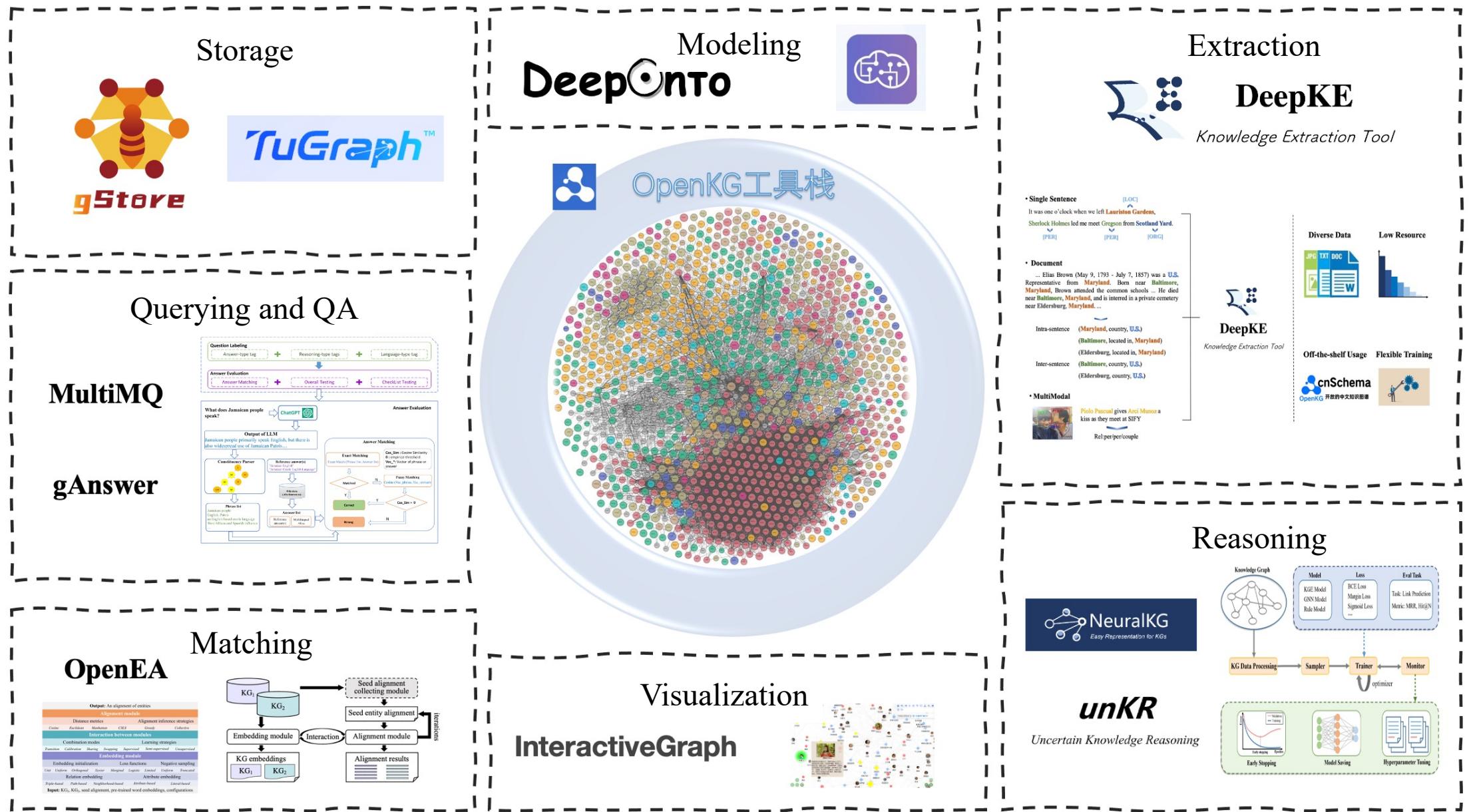


# OpenKG

Established in 2015, it initially aimed to promote the world's largest open knowledge graph centered on Chinese. By 2023, the goal has evolved to, through an open, collaborative, sustainable, and scalable approach, build a globally leading knowledge graph ecosystem in the era of large models, providing intelligent knowledge services for various industries and promoting the sharing and enhancement of knowledge value.

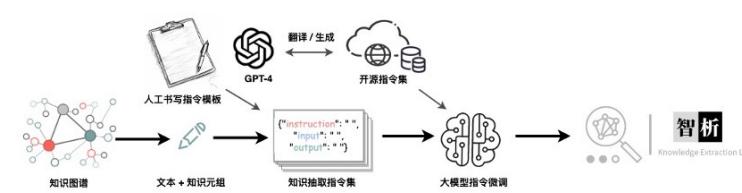
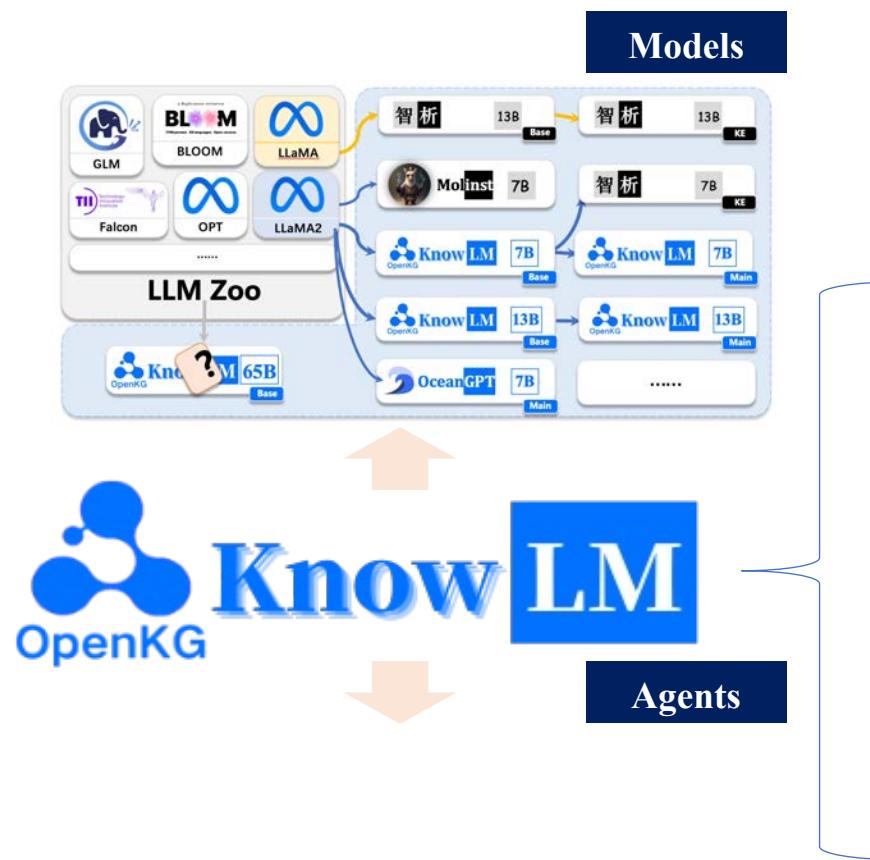


# OpenKG: Full Stack Tools



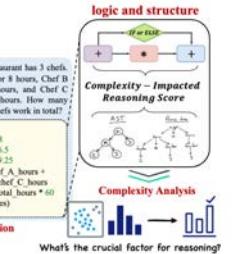
# OpenKG: Tools, Models and Agents

Tools

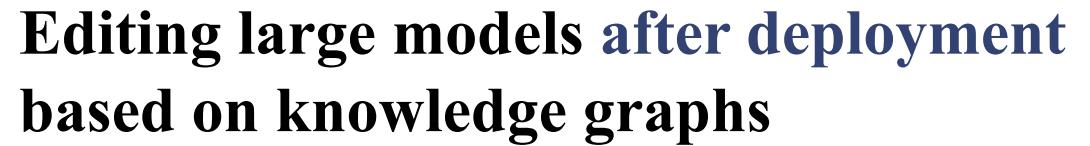
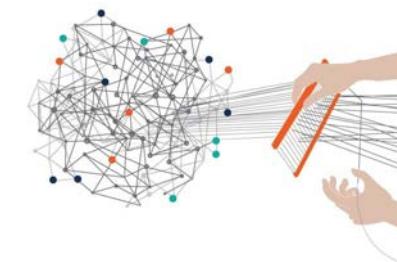


**KG2Instructions**

Enhancing large model **training** based on knowledge graphs



**KG2Code**



Editing large models **after deployment** based on knowledge graphs

# OpenKG: Open RAG



inheritance and development of RAG paradigms

## RAG Flow

The selection and arrangement of different modules and operators within the modules constitute RAG Flow, thereby identifying typical RAG Flow patterns.

### 3 Tuning Stage

- Retriever FT
- Generator FT
- Dual FT

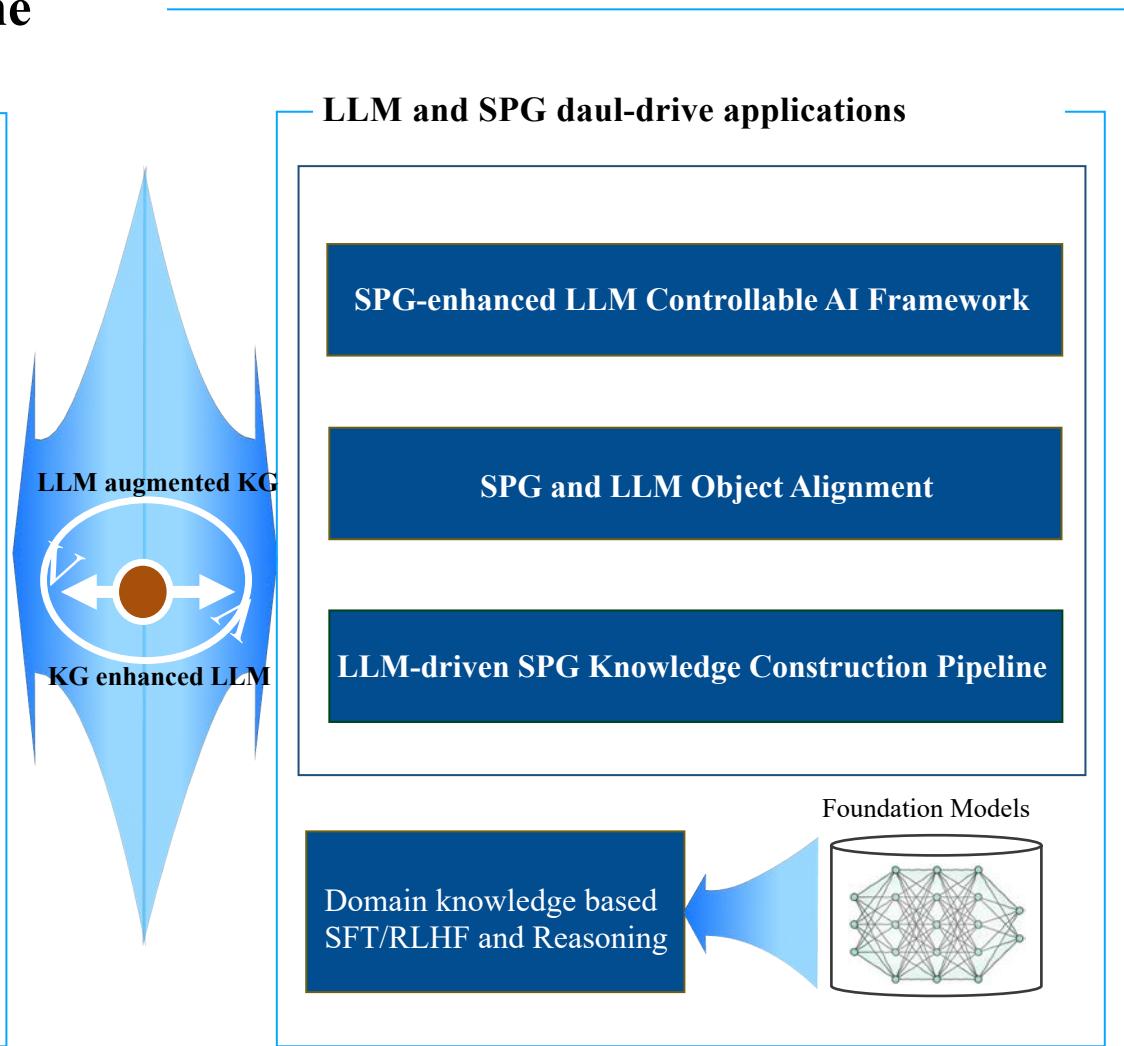
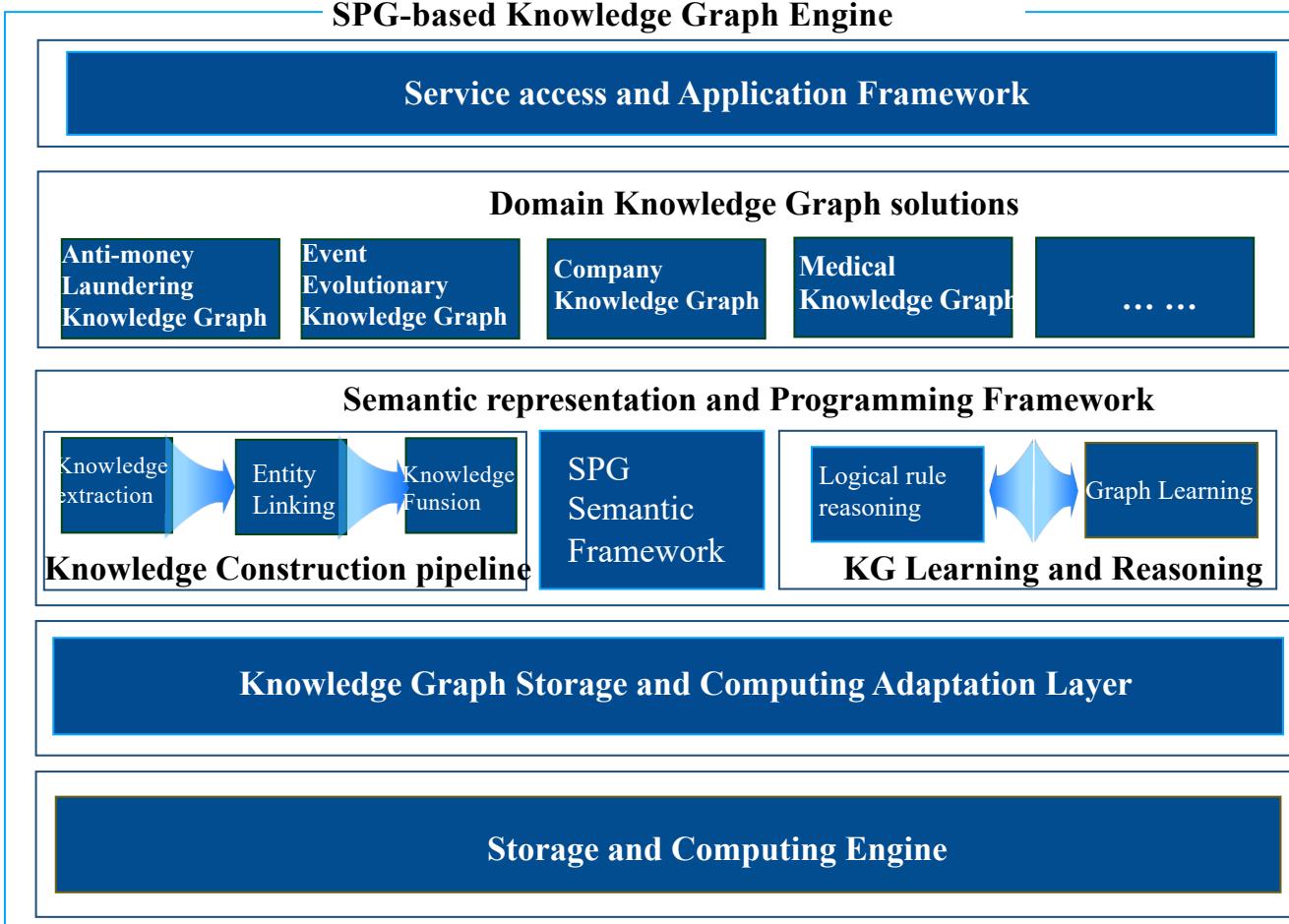
### 4 Inference Stage

- Sequential
- Conditional
- Brach
- Loop

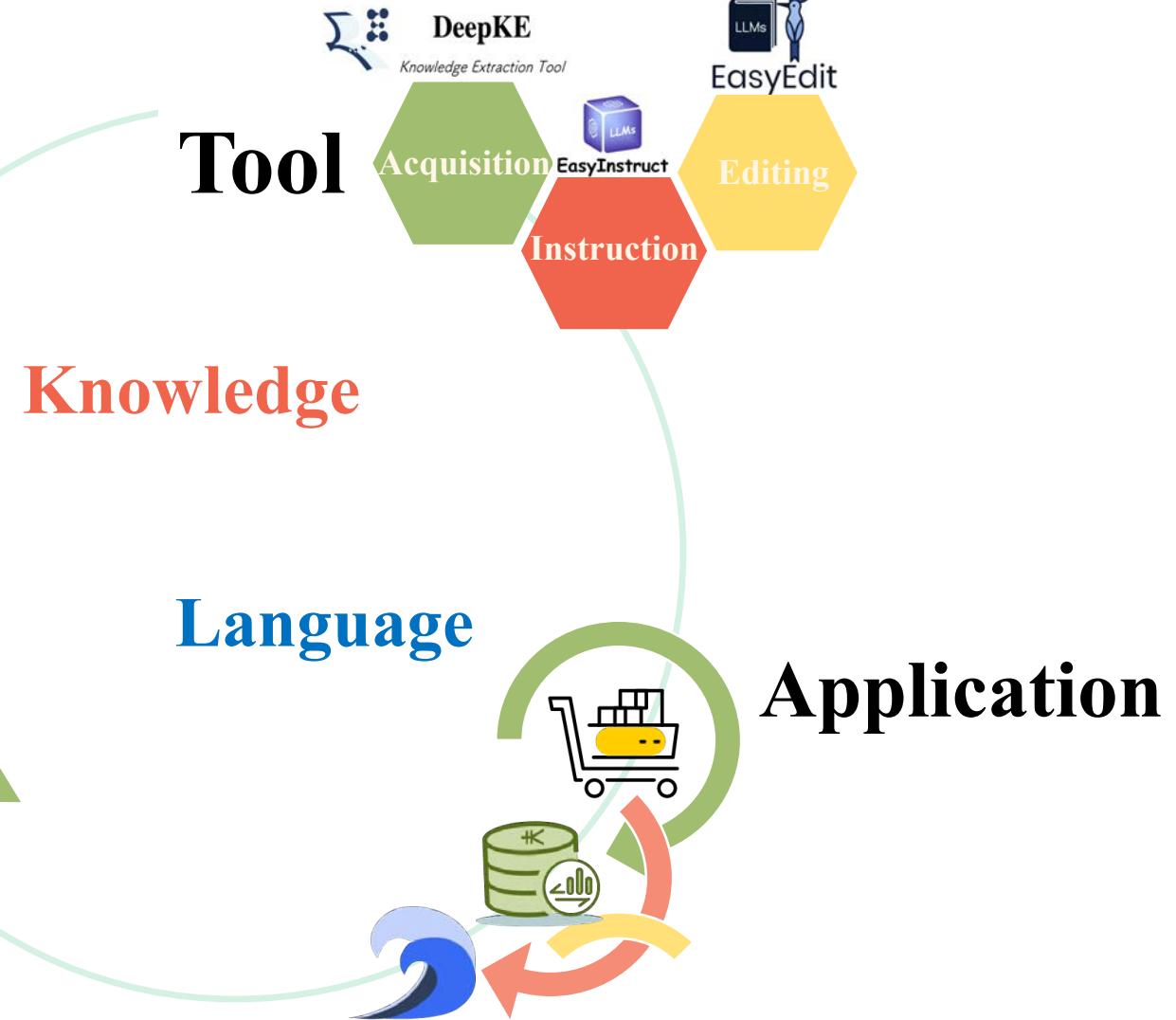
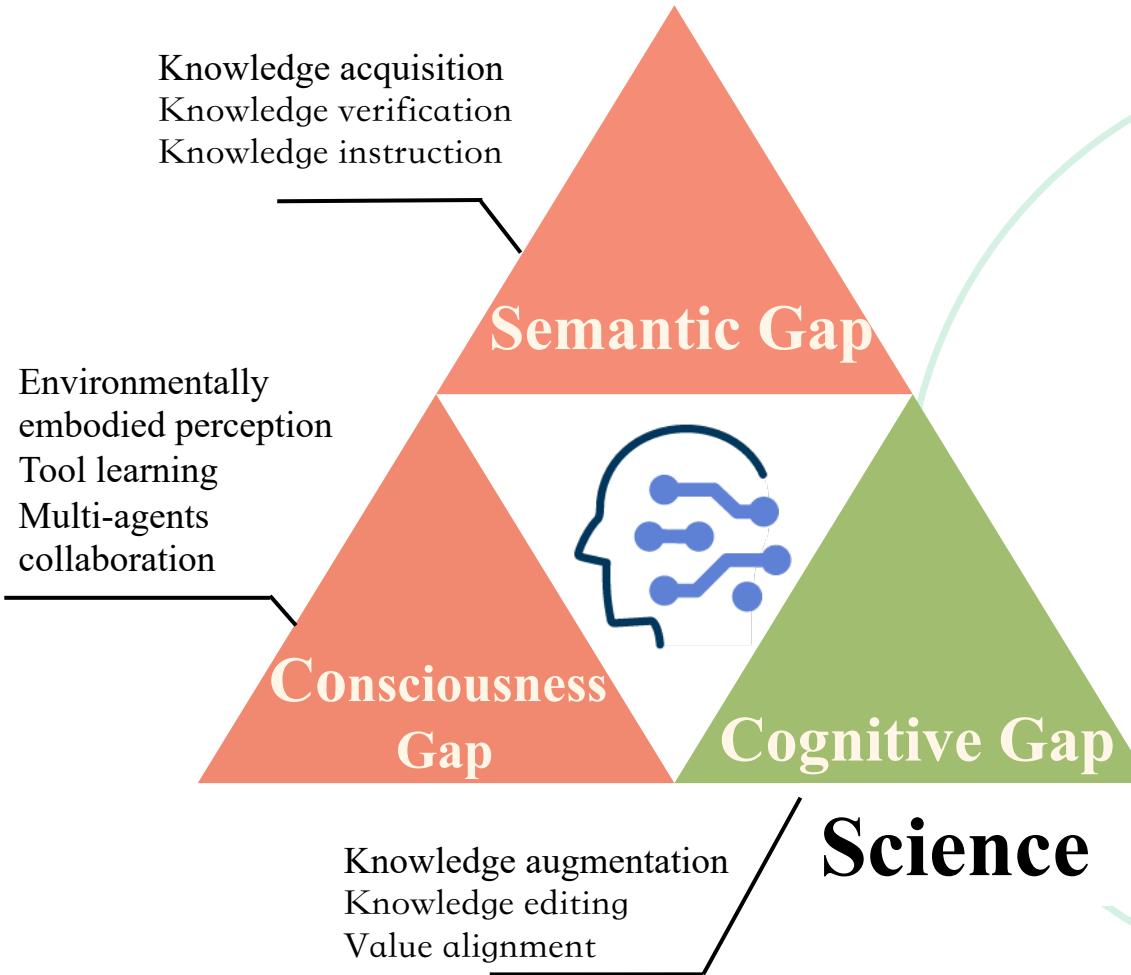


# OpenKG: SPG and LLM bidirectionally driven controllable AI

## Building next-generation industrial-level cognitive engine

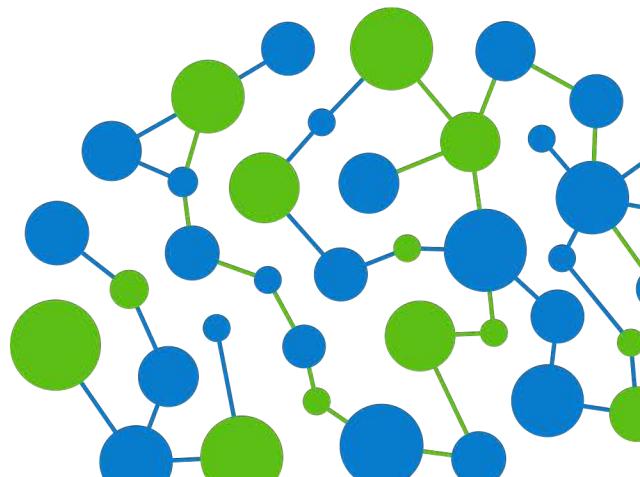


# OpenKG



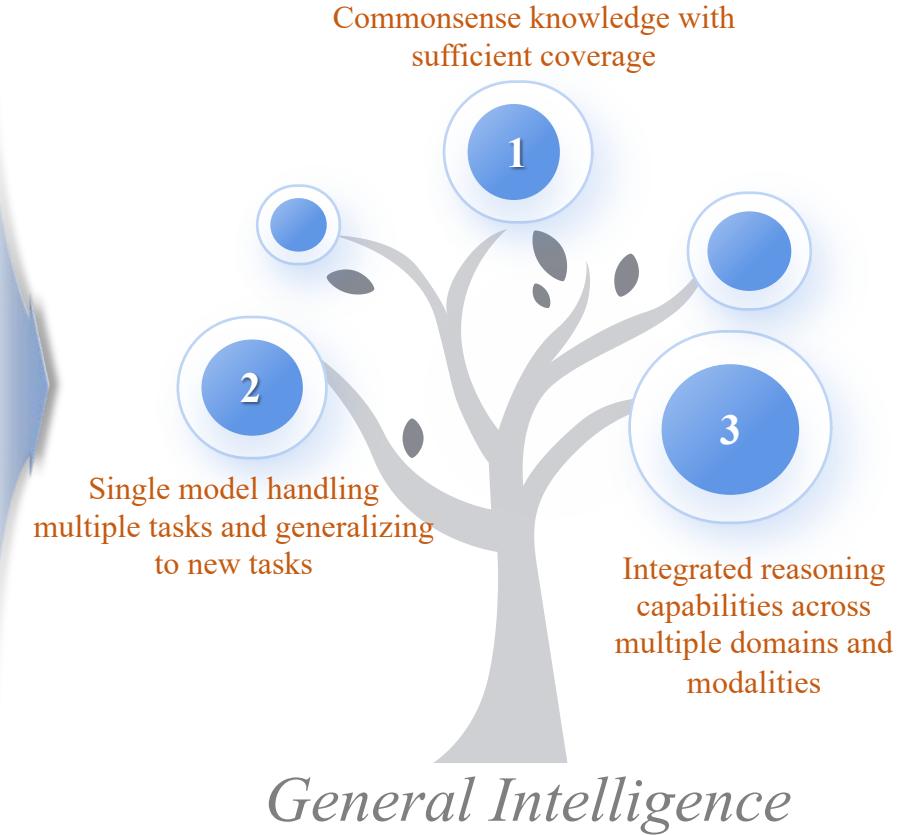
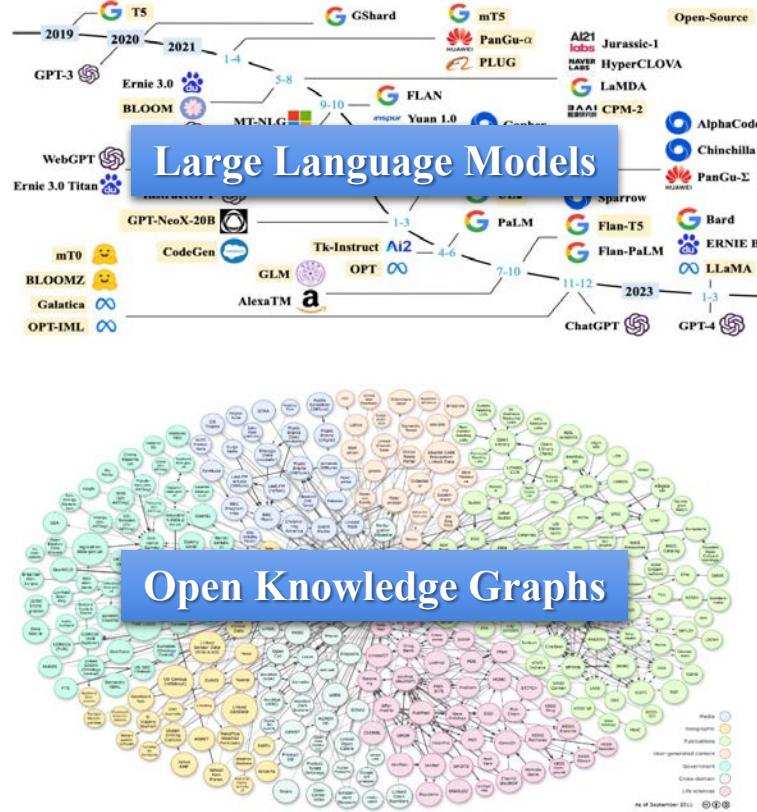
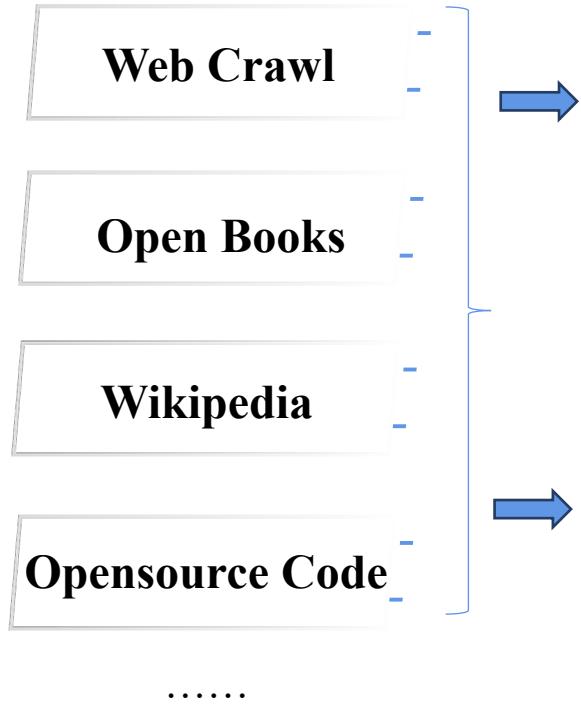
Language as "form", knowledge as "heart", graph as "skeleton"

- The Rise of LLM
- The Development of KG
- The Emerging Paradigm
- OpenKG
- Conclusion and Outlook



# Knowledge In LLM Era – Knowledge and General Intelligence

## Open Knowledge Corpus

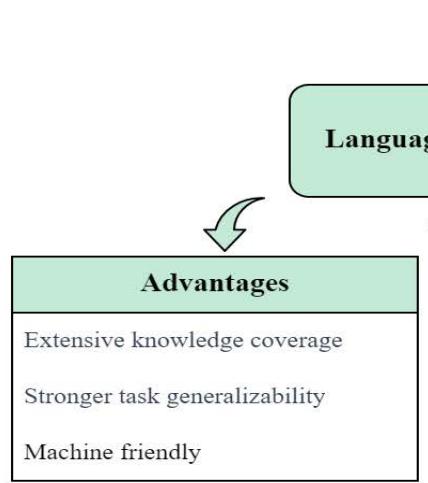


Both existing large language models and traditional knowledge graphs attempt to establish a shared knowledge base for general intelligence by leveraging the vast amount of open knowledge accumulated on the internet over time.

## Implication

# Knowledge In LLM Era – Knowledge and General Intelligence

**Focus on scale & has high coverage**

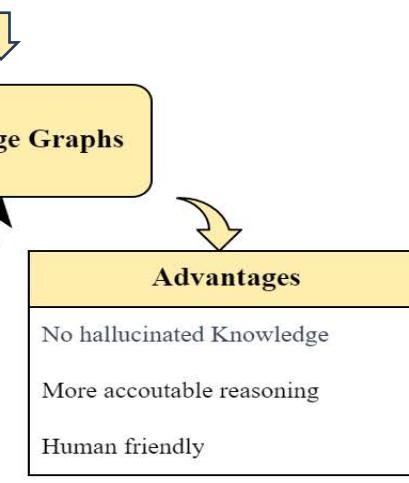


Language Models



Knowledge Augmentation

**Focus on presentation & has high accuracy**



Knowledge Graphs



World  
Knowledge

Neural  
Representation

Symbolic  
Representation

Compressing

Extraction

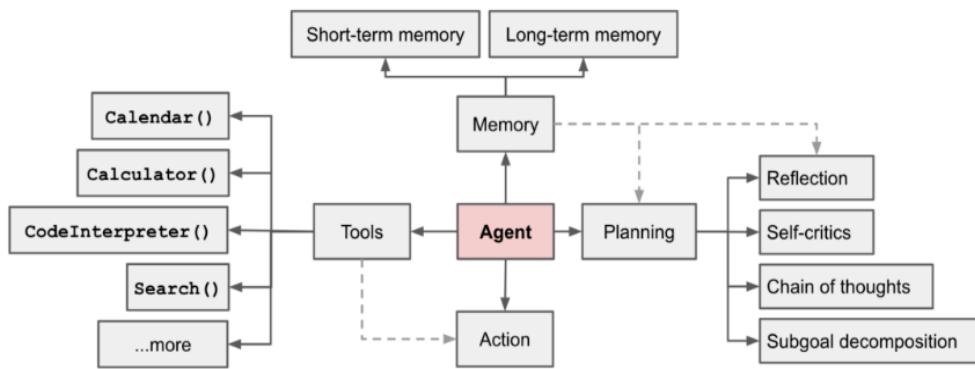
Human Languages

Describe

The complementarity and mutual enhancement between symbolic knowledge and language models have significant research importance and practical value.

Large Knowledge Models: Perspective and Challenges. 2024

# The Rise of World Model



Theory of Mind: Reasoning about **hidden mental** variables that drive **observable actions**

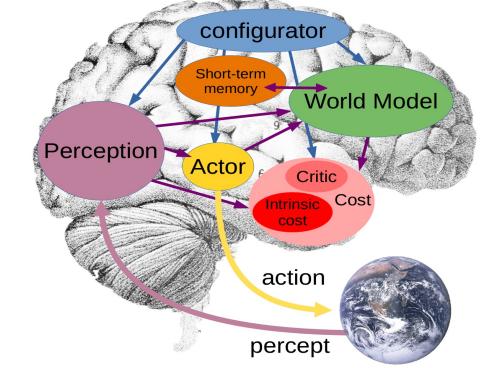
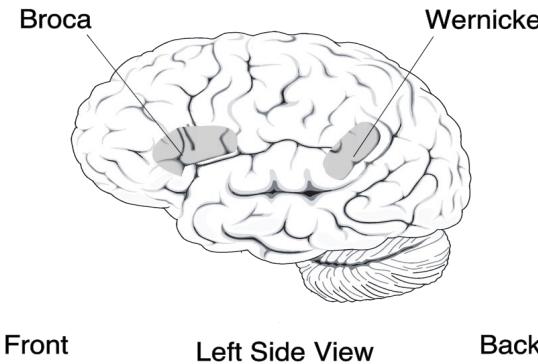


**Goal:** Sit down

**Belief:** The chair was not moved

**False belief**

caused by the person on the right



LeCun (2022)

- Autoregressive large language models (at best) only approximately approach the functions of the Wernicke and Broca areas of the brain.

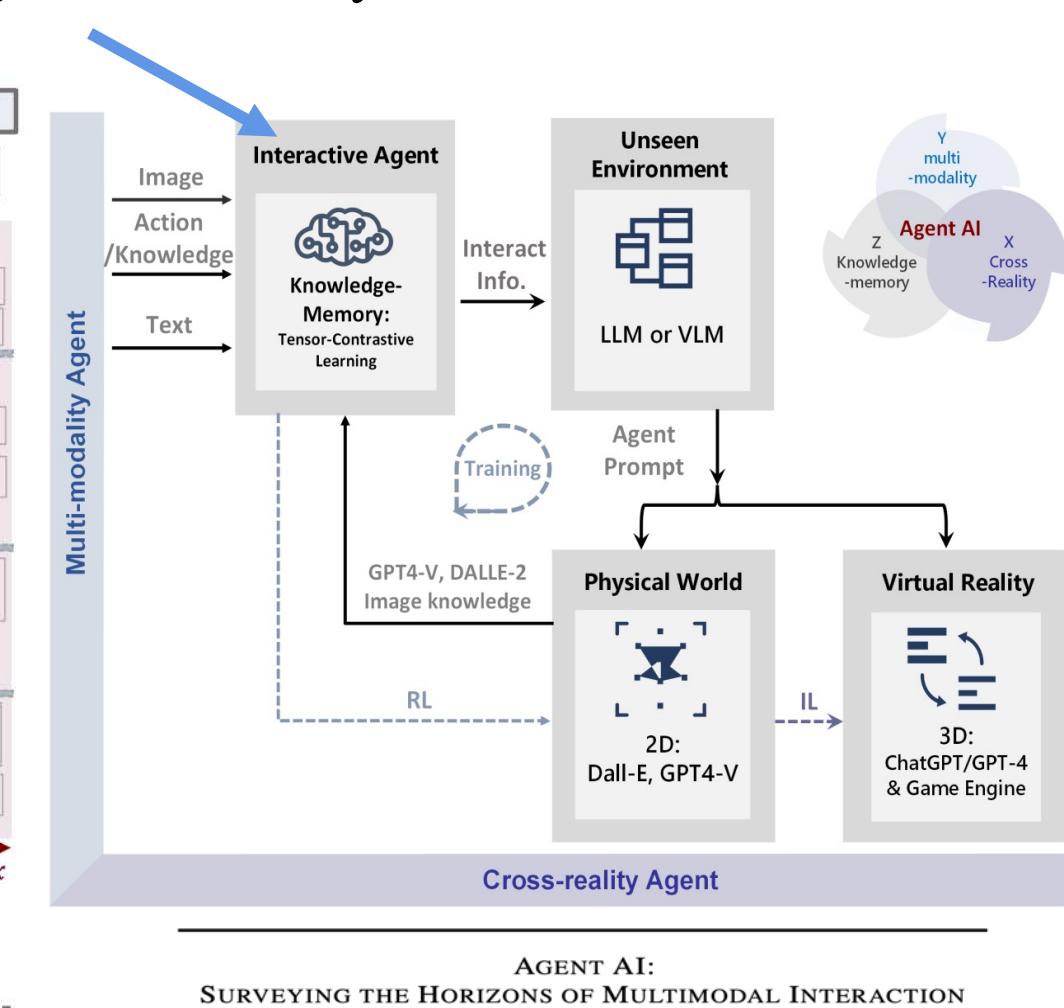
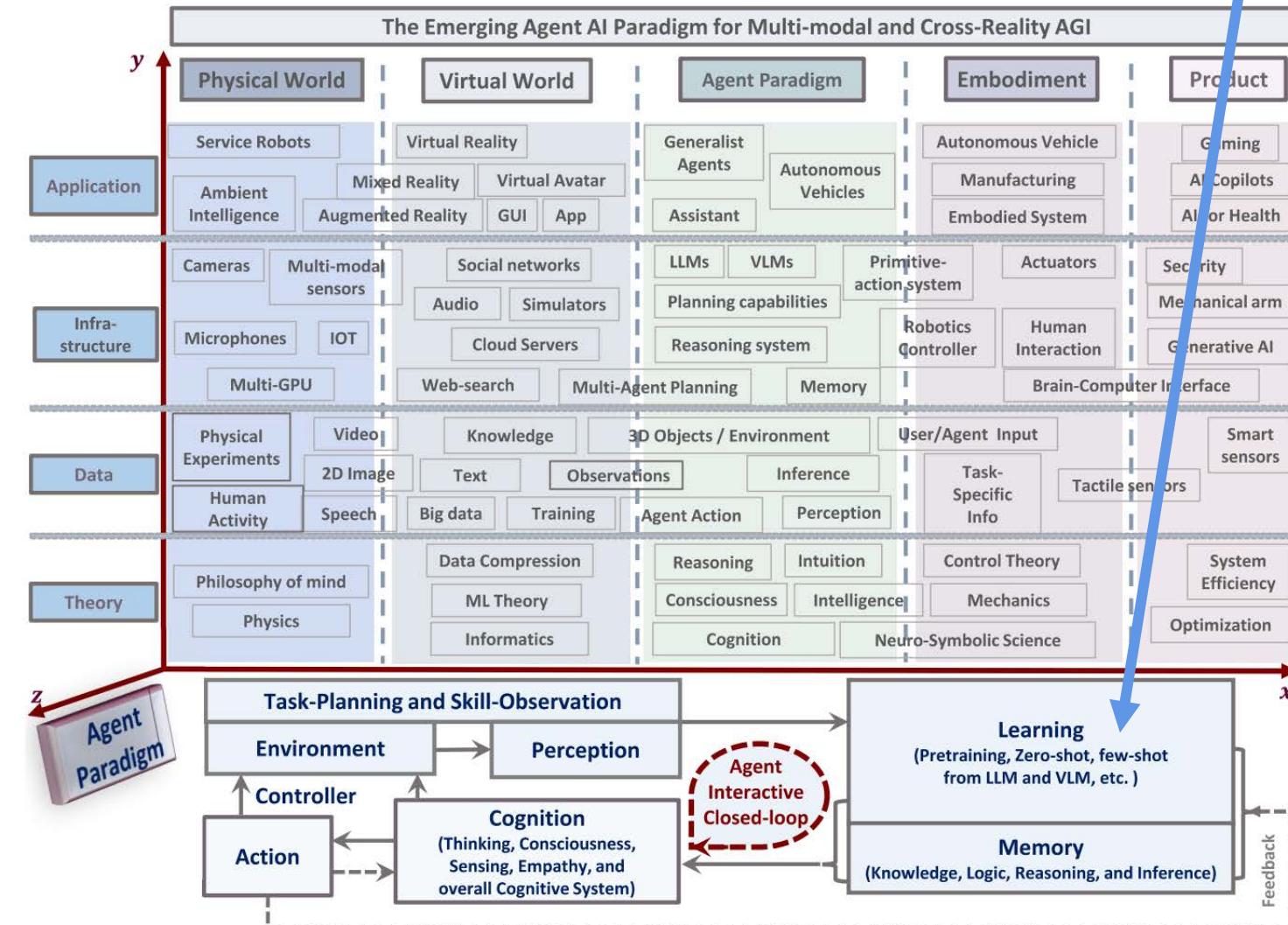
- What about the prefrontal cortex?

- Current AI technology is (still) far from human-level intelligence.

- Autonomous intelligence: configurator, perception, world model, cost, and participants

# The Role of Knowledge and Memory

## Knowledge and Memory



Zane Durante<sup>1†\*</sup>, Qiuyuan Huang<sup>2‡\*</sup>, Naoki Wake<sup>2\*</sup>,  
Ran Gong<sup>3†</sup>, Jae Sung Park<sup>4†</sup>, Bidipta Sarkar<sup>1†</sup>, Rohan Taori<sup>1†</sup>, Yusuke Noda<sup>5</sup>,  
Demetri Terzopoulos<sup>3</sup>, Yejin Choi<sup>4</sup>, Katsushi Ikeuchi<sup>2</sup>, Hoi Vo<sup>3</sup>, Li Fei-Fei<sup>1</sup>, Jianfeng Gao<sup>2</sup>

<sup>1</sup>Stanford University; <sup>2</sup>Microsoft Research, Redmond;

<sup>3</sup>University of California, Los Angeles; <sup>4</sup>University of Washington; <sup>5</sup>Microsoft Gaming

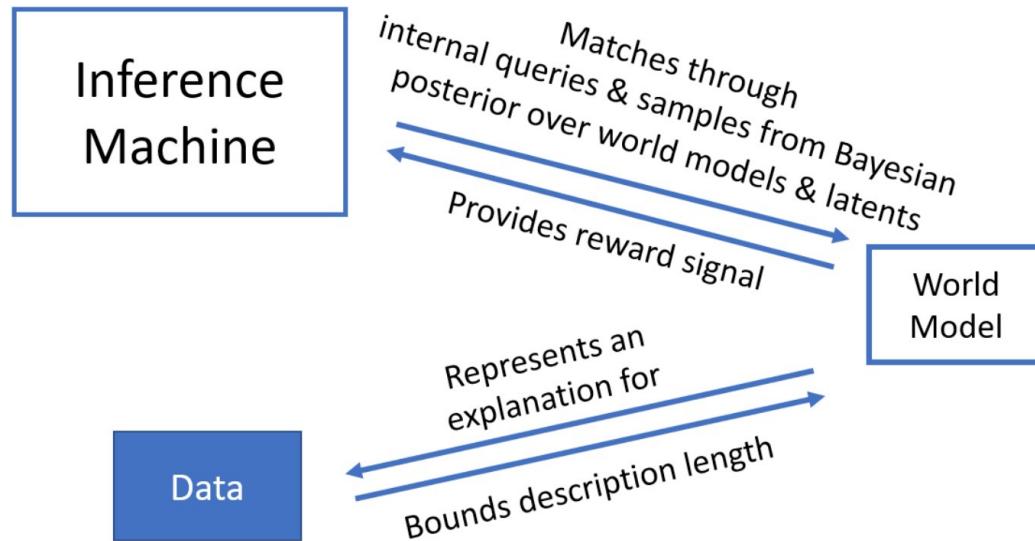
# World Model – Baby Learning



***From Baby Talk to Baby A.I.***  
Could a better understanding of  
how infants acquire language help  
us build smarter A.I. models?

Grounded language acquisition through the eyes and ears of a single child, **Science 2024**

# Knowledge Offloading



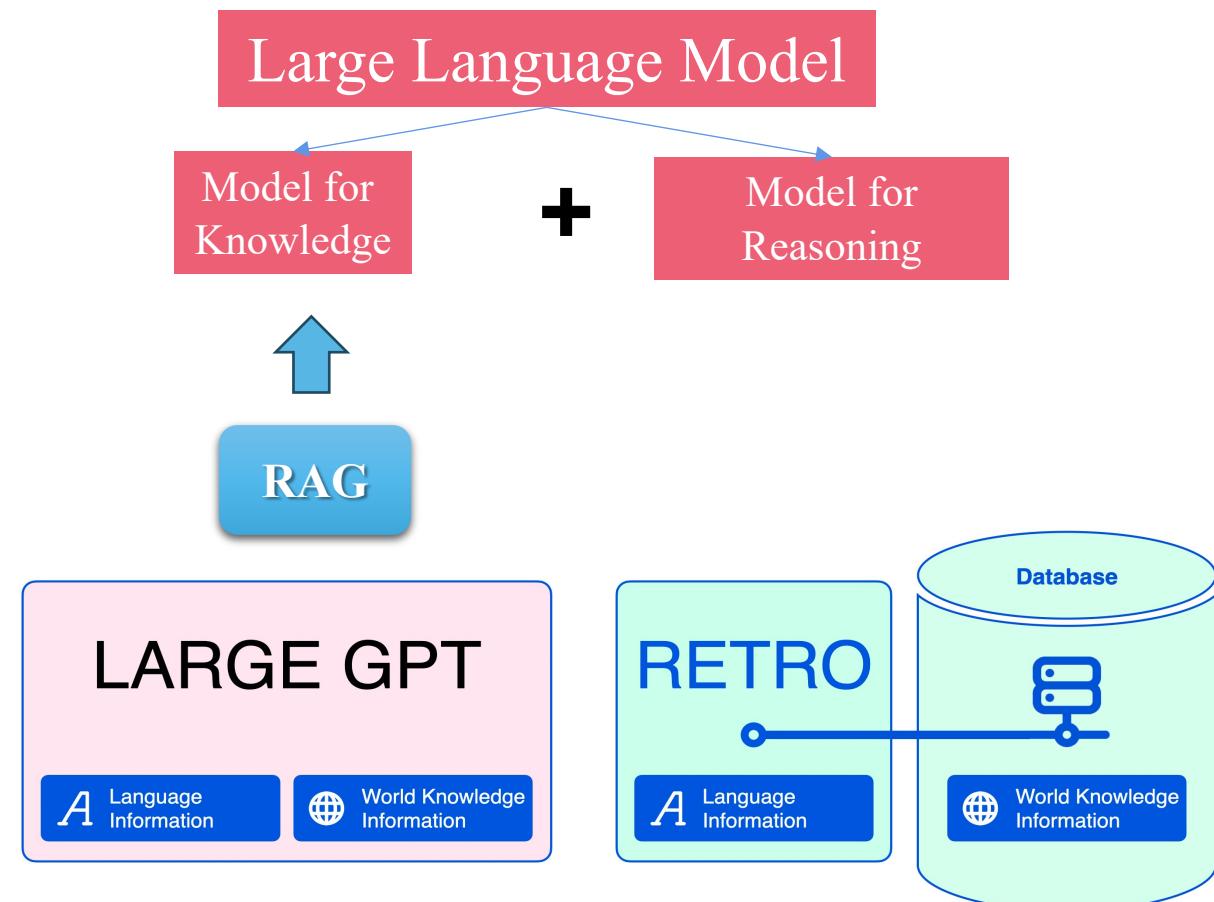
## Scaling in the service of reasoning & model-based ML

Published 21 March 2023 by [yoshuabengio](#)



*use some data to learn the world model, and some data to learn the inference machine.*

**Intelligence = Knowledge + Inference**



DeepMind's RETRO (Retrieval-Enhanced TRansfOrmer)

# Building the Next Generation Knowledge Engineering

## ■ Goals of Knowledge Engineering

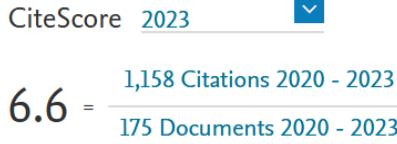
- Knowledge accumulation and reuse, knowledge reasoning and decision-making

## ■ Impact of Large Models on Knowledge Engineering

- Knowledge is not limited to symbolic representation, symbolic reasoning capabilities can be achieved through parameterization

## ■ Future Research in Knowledge Engineering: Embracing Large Models

- **KG for LLM** : Knowledge augmentation for large language models
- **LLM for KG** : Leveraging large language models to achieve better knowledge accumulation
- **LLM + KG** : Exploring the synergy between large language models and knowledge graphs to establish a "new symbolic system"



Category	Rank	Percentile
Computer Science Computer Science Applications	#214/817	73rd
Computer Science Information Systems	#107/394	72nd
Computer Science Artificial Intelligence	#119/350	66th

**Data Intelligence (DI)** is dedicated to promoting the widespread dissemination and in-depth exchange of the latest research findings and technological advancements in **data science, artificial intelligence, and their cross-disciplinary applications**. DI fosters close cooperation between the academic and industrial communities by publishing high-quality research papers, review articles, technical reports, case studies, best practices, and various data resources, thereby accelerating the innovation and application of data intelligence technologies and contributing to the development and progress of the field.

### Advisory Board



Deyi Li



Huajun Chen



Barend Mons



Deborah L.  
McGuinness



Sir Nigel Richard  
Shadbolt



Qiang  
Yang

### Co-Editor-in-Chief



Jim Hendler



Zhixiong Zhang



Guilin Qi



Heng Ji



Haofen Wang



# Thank you



OpenKG.CN  
开放的中文知识图谱

<http://www.openkg.cn/>