

AI 驱动 软件研发 全面进入数字化时代

中国·深圳 11.24-25

AI+
software
Development
Digital
summit



大模型时代的开放知识图谱技术栈

张宁豫 浙江大学

科技生态圈峰会 + 深度研习



—1000+ 技术团队的选择



K+全球软件研发行业创新峰会

会议时间: 2024.05.24-25



K+全球软件研发行业创新峰会

会议时间: 2024.09.20-21



AI+ 软件研发数字峰会

会议时间: 2023.11.24-25



AI+ 软件研发数字峰会

会议时间: 2024.07.19-20



AI+ 软件研发数字峰会

会议时间: 2024.11.15-16

▶ 演讲嘉宾



张宁豫

浙江大学副教授

浙江大学副教授，博士生导师。研究方向包括知识图谱、自然语言处理等，在知识图谱和自然语言处理领域已发表多篇顶级会议和期刊文章，包括 Nature 子刊、ACL、EMNLP、NAACL、NeurIPS、ICLR等，五篇论文入选Paper Digest高影响力论文，获浙江省科技进步二等奖，国际会议IJCKG最佳论文/提名2次，CCKS最佳论文奖1次；担任ACL、EMNLP领域主席、ARR Action Editor、IJCAI 高级程序委员、ACM Transactions on Asian and Low-Resource Language Information Processing Associate Editor及NeurIPS、ICLR、ICML等会议程序委员会委员。

目录

CONTENTS

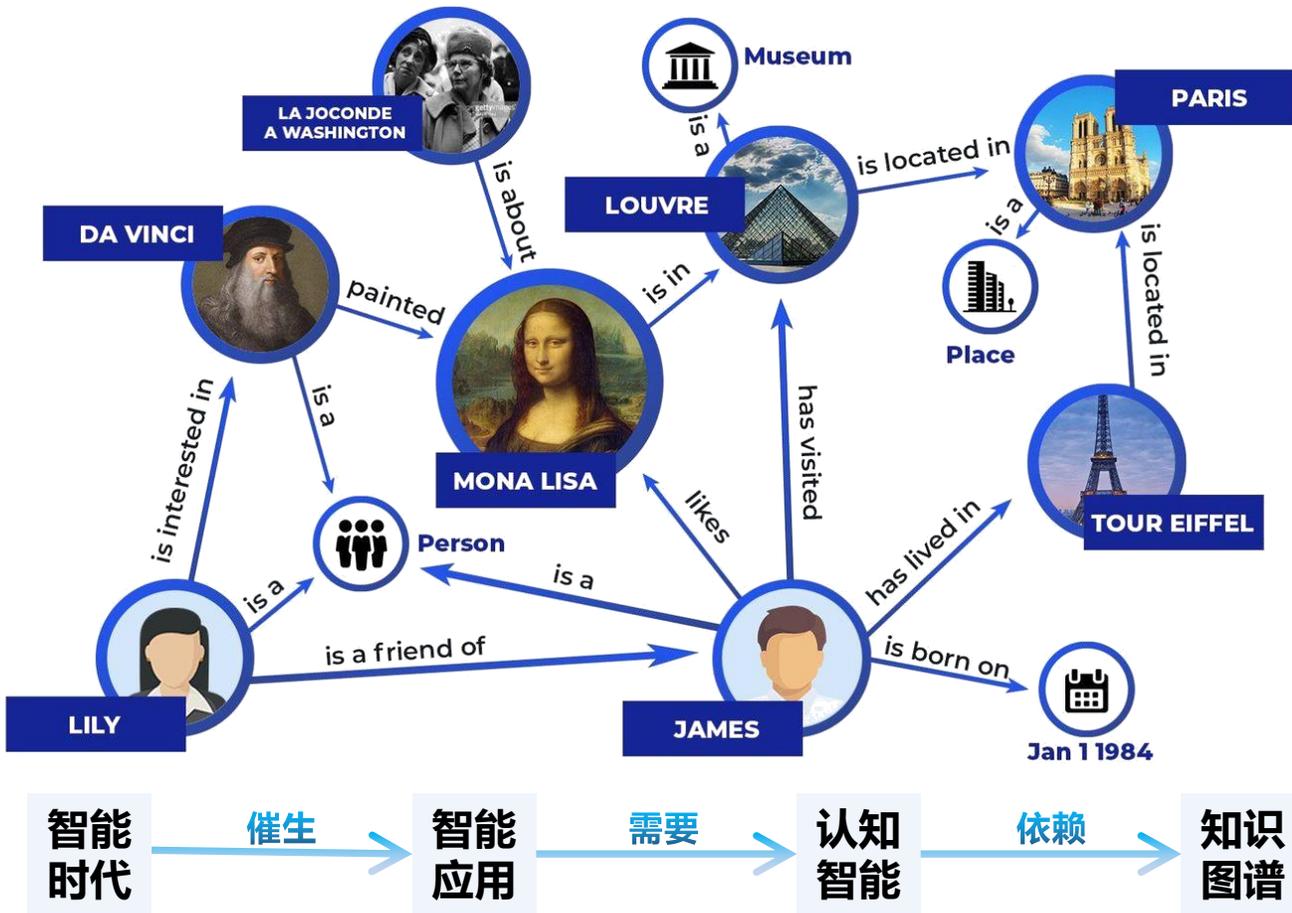
1. 背景
2. 知识提示
3. 知识编辑
4. 知识交互
5. 总结与展望

PART 01

知识图谱与大模型背景

知识图谱的历史

知识图谱 (Knowledge Graph) 是一种用知识语义、图结构来建模世界的一种方法, 它能表示事物(Things)之间的语义结构, 刻画真实世界万物之间的深度关联。



Semantic Networks 1960

语义网络作为知识表示的一种方法被提出, 主要用于自然语言理解领域

Ontology 1980s

本体论。哲学概念“本体”被引入到人工智能领域用来刻画知识

World Wide Web 1989

Tim Berners-Lee发明了万维网
Linked Information System

The Semantic Web 1998

从超文本链接到语义链接

Linked Data 2006

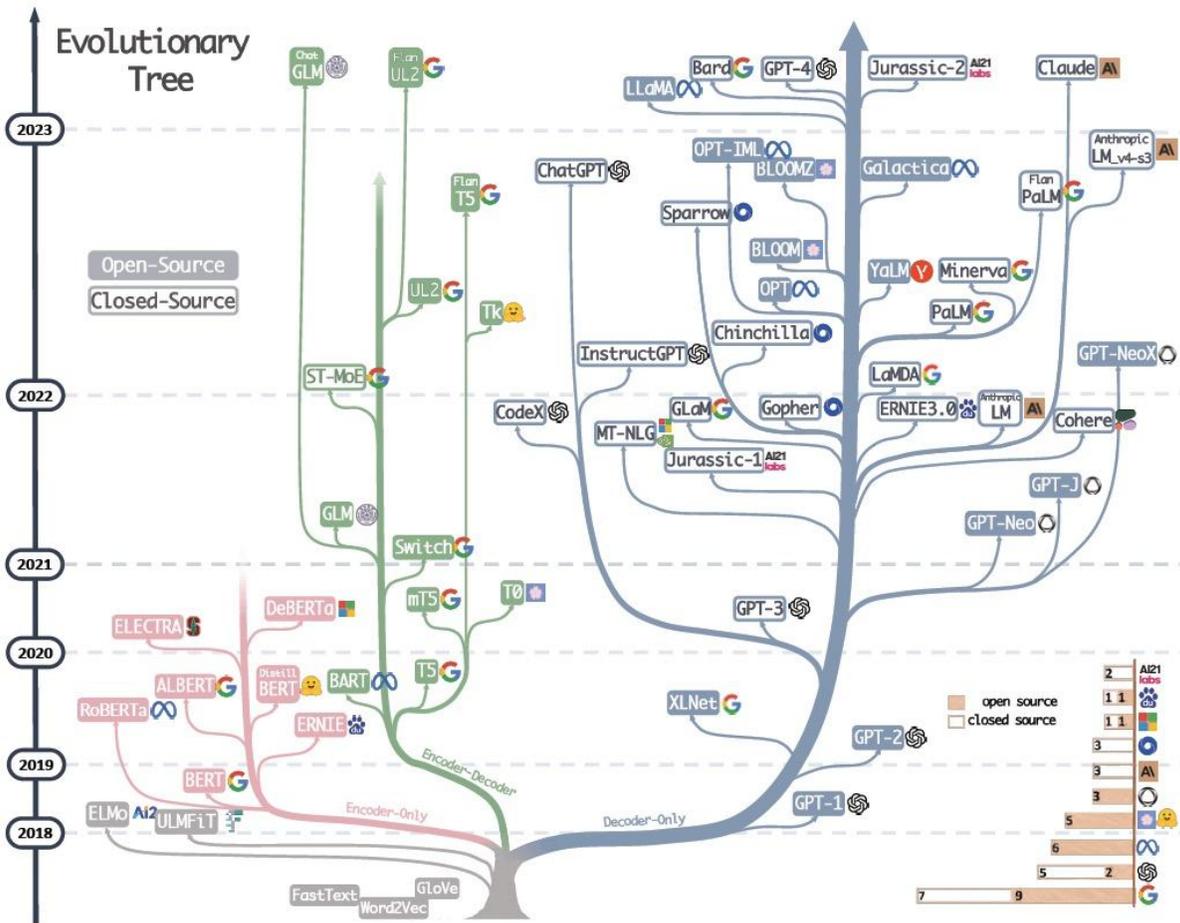
Tim突出强调语义网的本质是要建立开放数据之间的链接

Knowledge Graph 2012

谷歌发布了其基于知识图谱的搜索引擎产品

随后诞生多模态图谱、时空图谱、事件图谱 现在

▶ 大模型时代还需要知识图谱么？

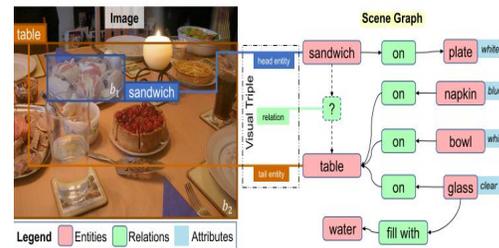
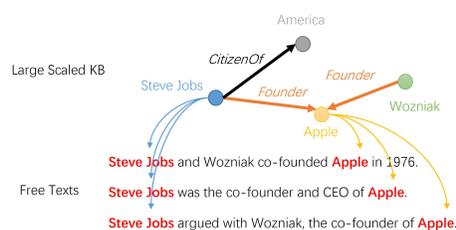


符号知识图谱



- Embedding Concepts: Terms and Class Hierarchy
- Embedding Facts: Entities, Attributes, Relations
- Embedding Structures: Paths, Neighbors, Sub graphs
- Embedding Logic: Axioms and Rules

VS.



[1] Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond. 2023

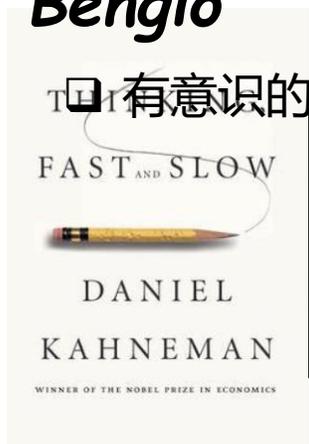
▶ 大模型时代还需要知识图谱么?

□ 语言的有限性 -- Yann LeCun

- 语言只承载了人类全部知识的一小部分，很多知识学习依赖于与真实世界的交互

□ 深度学习快慢双系统 -- Yoshua Bengio

- 有意识的、带逻辑、规划、推理的系统



How many World Cup championships has Messi won?



Lionel Messi has no World Cup championships with the Argentina national team. He has played in three World Cup tournaments (2006, 2010, and 2014), and his team finished as runners-up in the 2014 World Cup.

知识缺失谬误问题!

GPT-4自我对话

用户启动: 今天很热。

GPT-4 会话1: 是的, 似乎今天非常热! 记得要补充水分并尽量保持凉爽。如果你要到户外去, 确保涂抹防晒霜。

GPT-4 会话2: 感谢你的建议! 我一定会多喝水并且涂抹防晒霜。也许今天是个吃冰淇淋的好日子。

GPT-4 会话1: 不客气! 在这样一个炎热的日子里享用冰淇淋听起来是个很好的主意。祝你玩得开心, 保持凉爽!

GPT-4 会话2: 谢谢你! 我会确保自己有个美好的时光并保持凉爽。祝你也过得愉快!

GPT-4 会话1: 谢谢你! 如果未来有任何问题或需要帮助, 请向我求助。今天并祝你有美好的一天!

GPT-4 会话2: 谢谢!

符号表示落地问题!

Fig. 1: 在用户启动话题后两个GPT-4之间的自我对话。两个模型在只进行一轮对话后就迅速转变为彼此表示感谢(标记为黄色)。5轮之后仍旧无限重复, 这与符号落地理论相呼应, 即表明在没有建立词语到世界的联结时, 对话将进入无限循环。

[1] Brain in a Vat: On Missing Pieces Towards Artificial General Intelligence in Large Language Models. 2023

▶ 大模型时代还需要知识图谱么?

□ 语言的有限性 -- Yann LeCun

- 语言只承载了人类全部知识的一小部分, 很多知识学习依赖于与真实世界的交互

□ 深度学习快慢双系统 -- Yoshua Bengio

Bengio

□ 有意识的、带逻辑、规划、推理的系统

FAST AND SLOW



DANIEL

KAHNEMAN

WINNER OF THE NOBEL PRIZE IN ECONOMICS



Given a set of rules and facts, you have to reason whether a statement is true or false. Here are some facts and rules:

The bear likes the dog.

The cow is round.

The cow likes the bear.

The cow needs the bear.

The dog needs the squirrel.

The dog sees the cow.

The squirrel needs the dog.

If someone is round then they like the squirrel.

If the bear is round and the bear likes the squirrel then the squirrel needs the bear.

If the cow needs the dog then the cow is cold.

Does it imply that the statement "The cow likes the squirrel." is True?

Given a set of rules and facts, you have to reason whether a statement is true or false.

Here are some facts and rules:

The e4 likes the e5.

The e14 is e2.

The e14 likes the e4.

The e14 needs the e4.

The e5 needs the e26.

The e5 sees the e14.

The e26 needs the e5.

If someone is e2 then they like the e26.

If the e4 is e2 and the e4 likes the e26 then the e26 needs the e4.

If the e14 needs the e5 then the e14 is e1.

Does it imply that the statement "The e14 likes the e26." is True?

复杂推理问题!

复杂鲁棒可信的人工智能系统仍离不开高质量的知识图谱

[1] Brain in a Vat: On Missing Pieces Towards Artificial General Intelligence in Large Language Models. 2023

知识工程与知识图谱

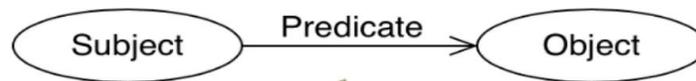


费根鲍姆
图灵奖1994

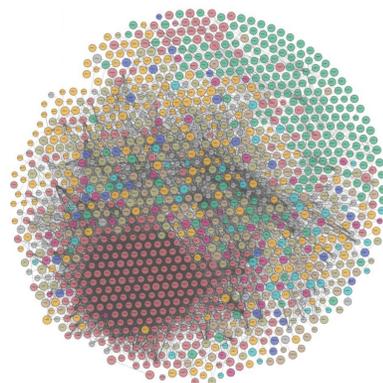
图灵奖获得者费根鲍姆提出了知识工程的概念，确立了其在人工智能中的重要地位

知识图谱接近自然语言和人脑认知、易于扩展增加、易于神经网络化

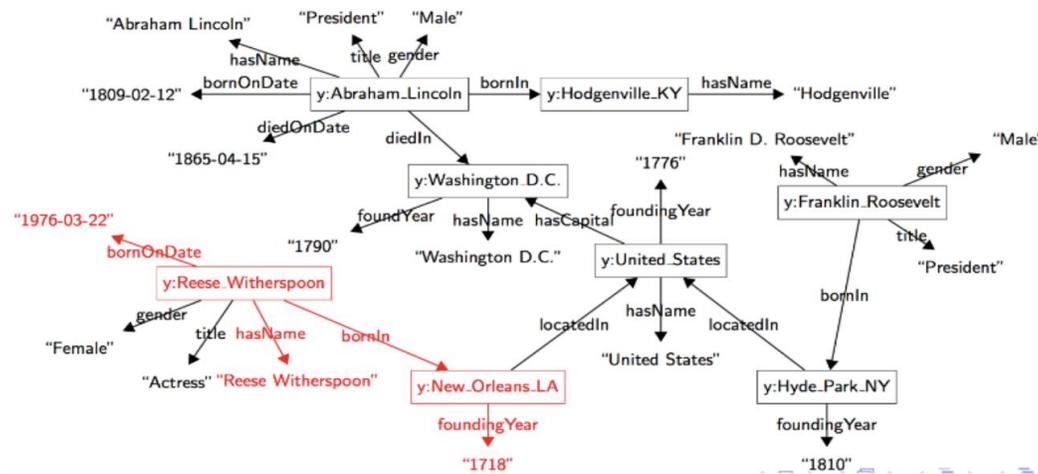
知识图谱 = 文字语义 + 结构知识



大数据



大知识



▶ 自然语言、知识图谱和语言模型都是表示和处理知识的手段

语言 \neq 知识

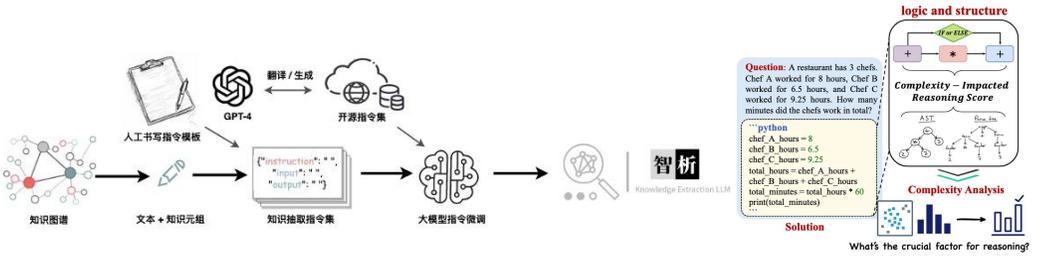
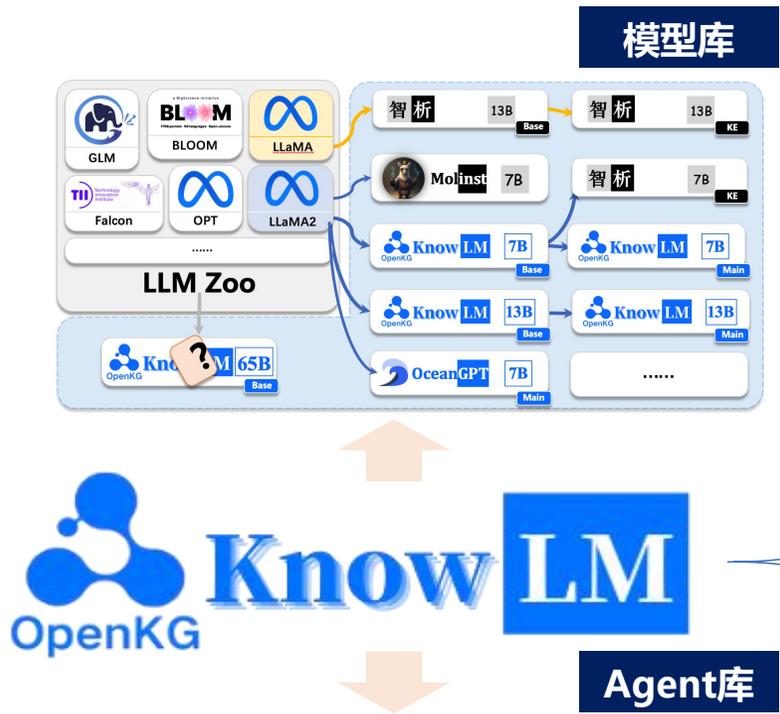
表示方法	可解释性	知识类型	可计算性
自然语言	人可理解	显性知识	不易于计算处理
知识图谱	人可理解	显性知识 + 隐性知识	较容易计算处理
语言模型	人不可理解	隐性知识	易于计算和处理

▶ 知识图谱与大模型

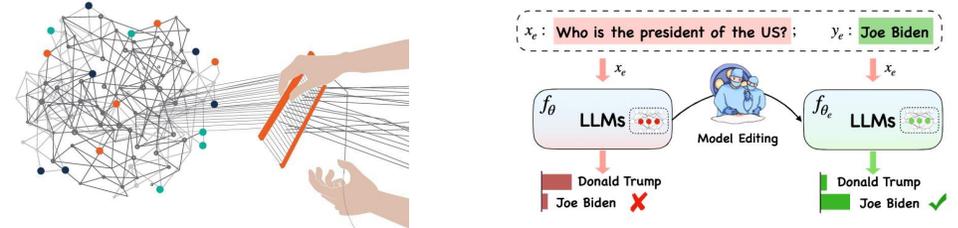


KG: 知识获取、
知识推理、知识
应用.....

知识增强的开源语言大模型框架KnowLM

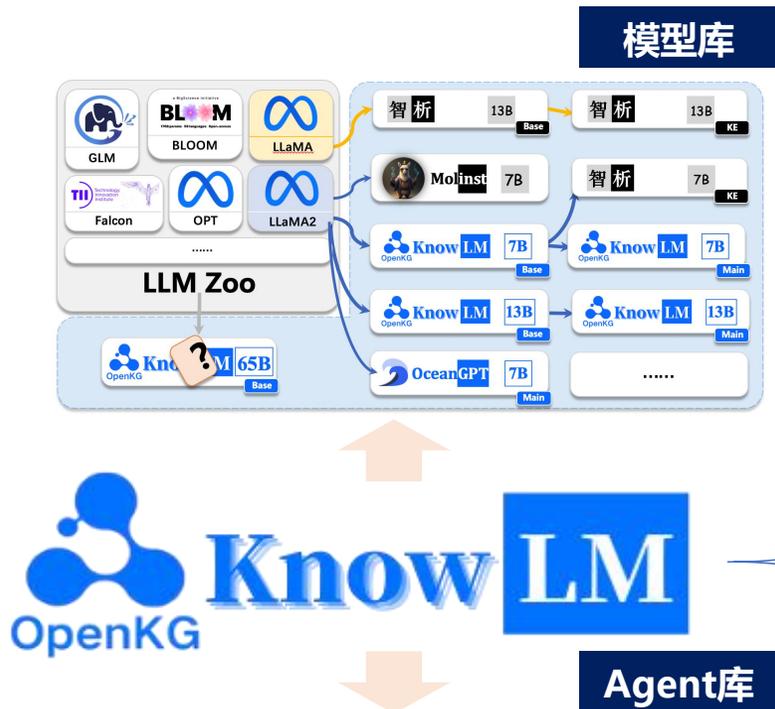


KG2Instructions
KG2Code
大模型训练时基于知识图谱进行增强



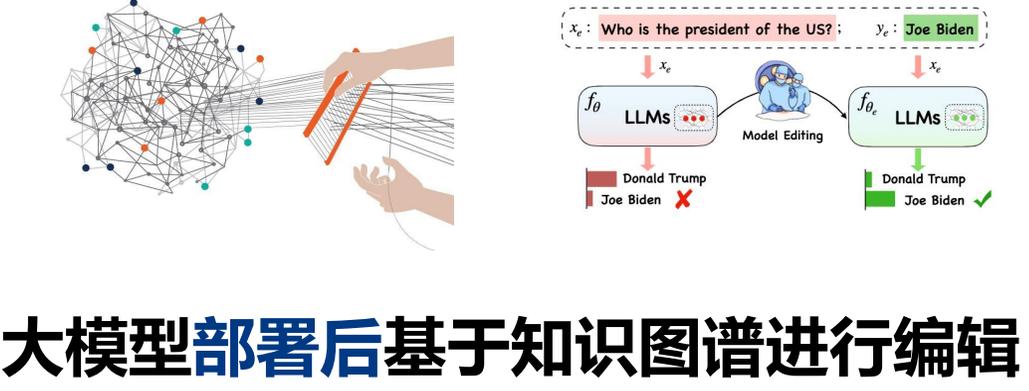
大模型部署后基于知识图谱进行编辑

知识增强的开源语言大模型框架KnowLM



EasyInstruct

大模型训练时基于知识图谱进行增强



大模型部署后基于知识图谱进行编辑

PART 02

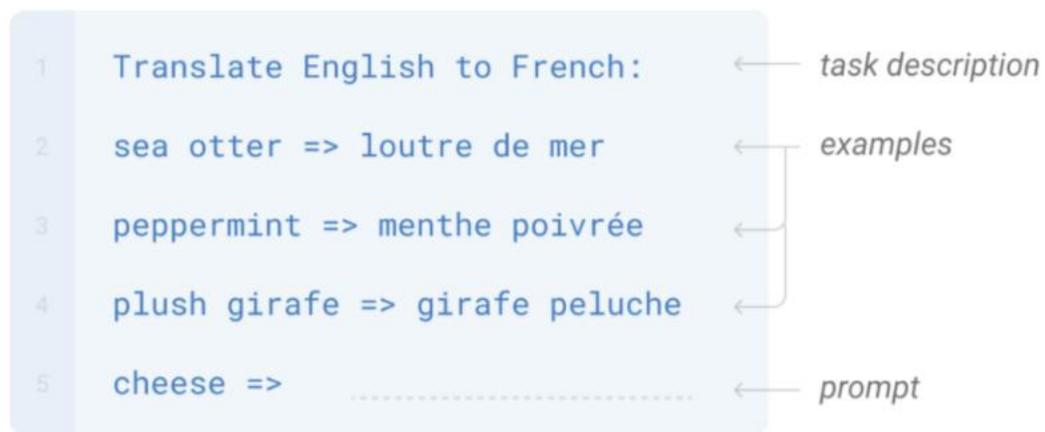
基于知识提示技术的大模型增强

▶ 大模型提示学习

起源于GPT3的预训练提示学习技术：提示显著提升低资源性能

Few-shot

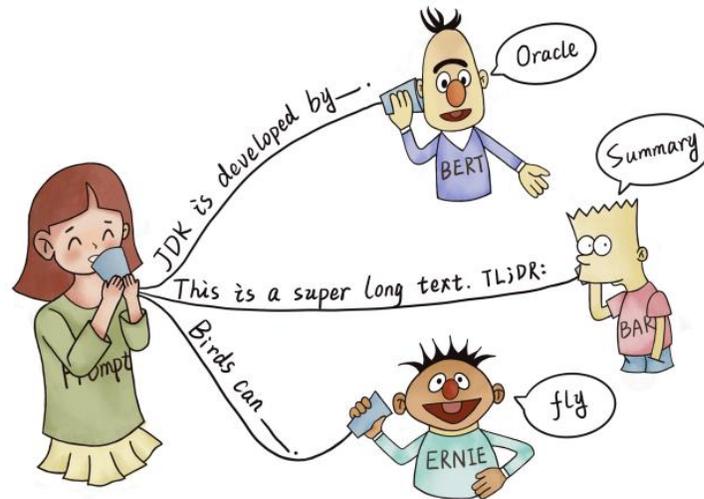
In addition to the task description, the model sees a few examples of the task. No gradient updates are performed.



[1]Language Models Are Few-shot Learners (NeurIPS 2020)

▶ 大模型提示学习

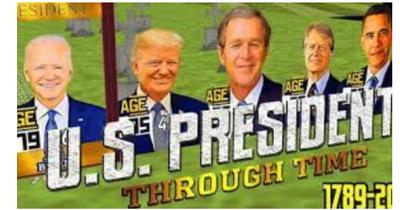
构建提示解锁技能?



所有知识都可以习得?

$$z_{n+1} = z_n^2 + c$$
$$c \in M \iff \lim_{n \rightarrow \infty} |z_{n+1}| \leq 2$$

新知识/技能如何习得?

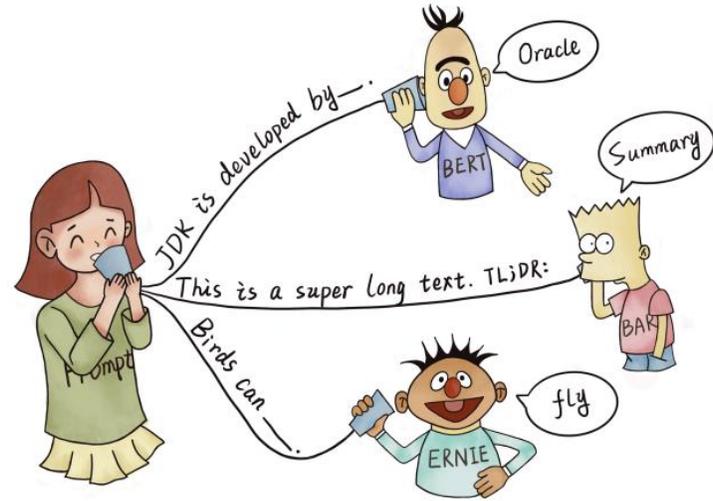


Knowledge stimulation? Inductive **Bias** (Task-related)?

template, verbalizer, demonstration

▶ 大模型提示学习

1. 如何更好**激活**预训练大模型的知识?
2. 如何**增强/更新**预训练大模型的知识?

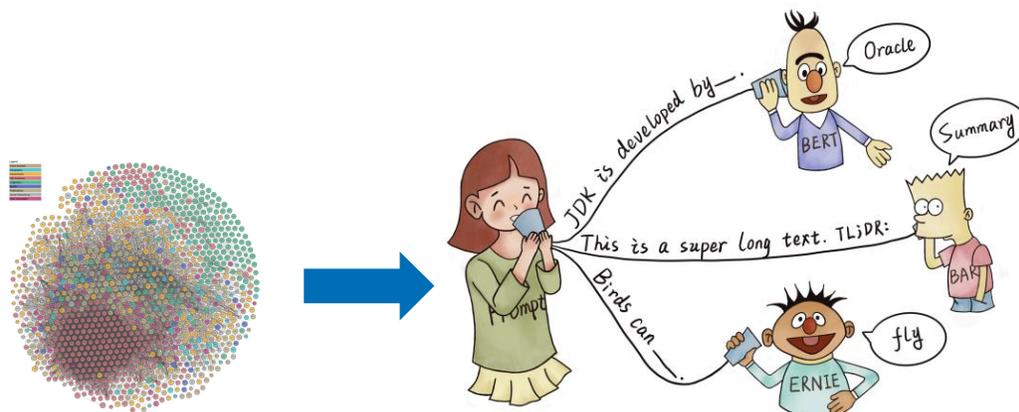


Knowledge stimulation? Inductive **Bias** (Task-related)?

template, verbalizer, demonstration

▶ 将提示作为知识与预训练模型的桥梁

1. 更好**激活**预训练大模型的知识
2. **增强/更新**预训练大模型的知识



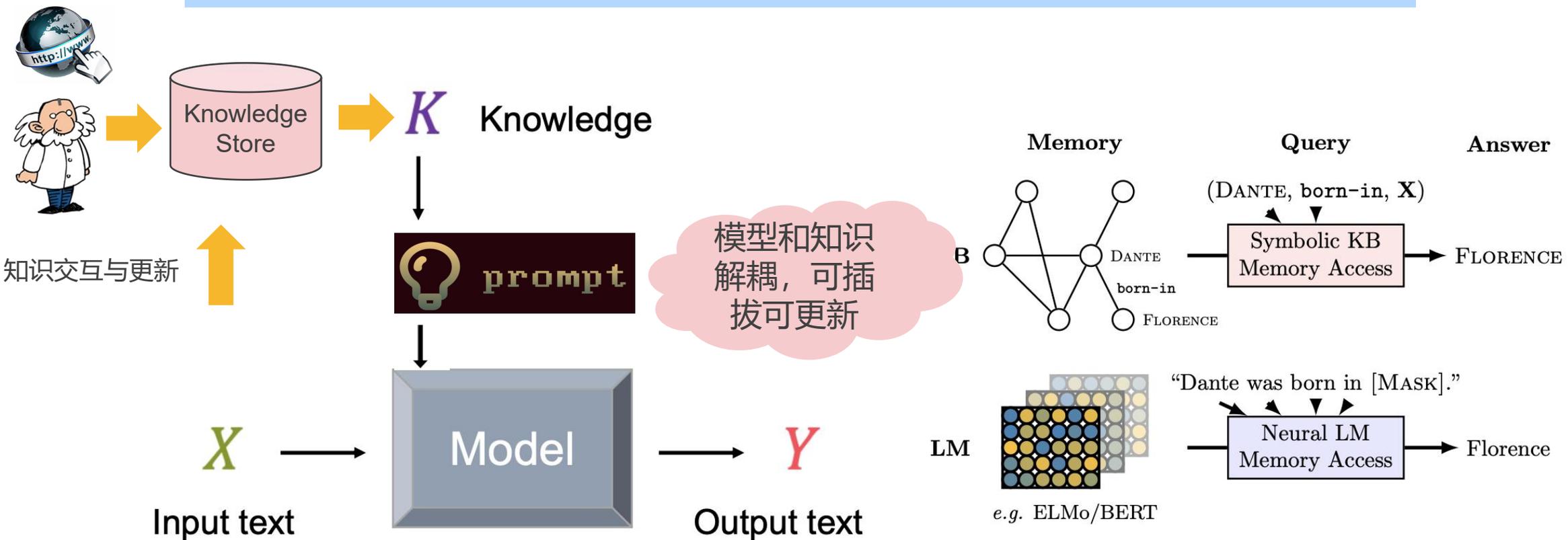
How to inject (stimulate) knowledge?

knowledgeable template, enhanced verbalizer, demonstration, external fact, ...

通过提示将知识“植入”预训练模型

▶ 知识提示学习

“知识提示学习”旨在植入任务、领域相关的知识（知识图谱、文本示例等），通过“**Knowledge as Prompt**”提升预训练模型对**领域、任务的理解**，激发**预训练模型的潜能**，提升下游任务的性能。



知识提示学习: KnowPrompt

主要挑战

语义关系的复杂性问题

`per:country_of_birth` 和 `org:city_of_headquarters` 等关系无法在词表找到合适的独立标签词

知识图谱的归纳偏置问题

实体类型和关系类型之间存在隐式的约束, 如 `per:country_of_birth` 的头实体一定是person, 尾实体一定是location

基于知识图谱中的**先验知识(Schema)**作为提示的思想提升对**关系抽取**任务的感知并激发大模型的潜能

TABLE 1: Most Influential WWW Papers

PAPER DIGEST

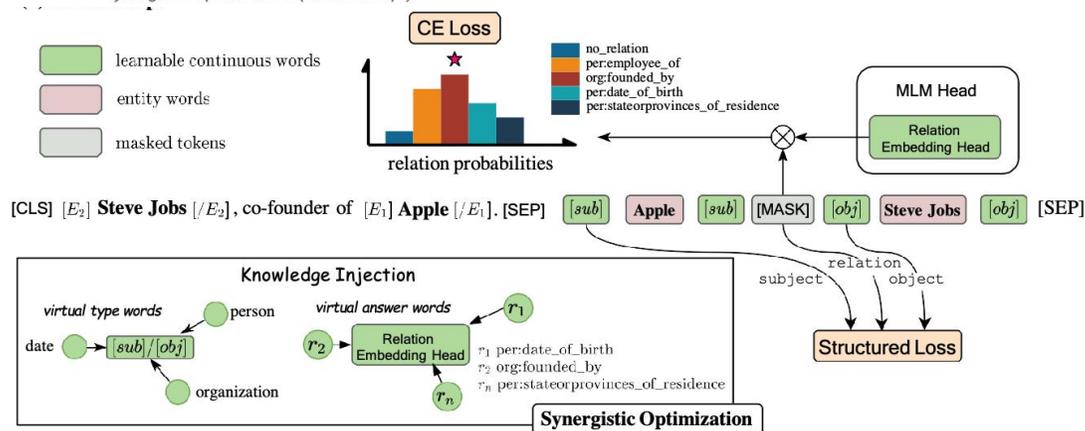
YEAR RANK PAPER AUTHOR(S)

KnowPrompt: Knowledge-aware Prompt-tuning with Synergistic Optimization for Relation Extraction

IF:3 Literature Review Related Patents Related Grants Related Orgs Related Experts Details

2022 1 *Highlight*: Furthermore, there exists abundant semantic and prior knowledge among the relation labels XIANG CHEN et. al.

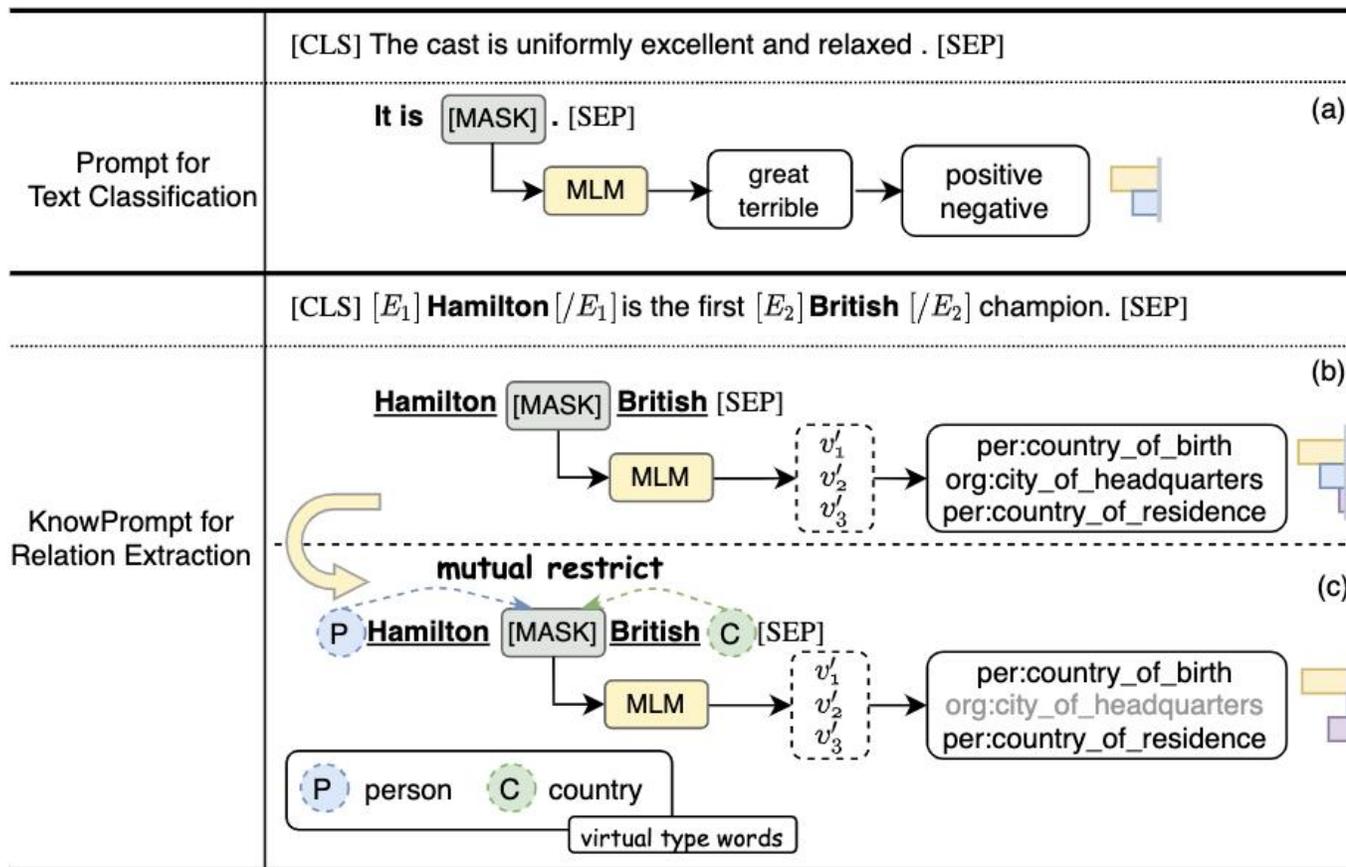
that cannot be ignored. To this end, we focus on incorporating knowledge among relation labels into prompt-tuning for relation extraction and propose a Knowledge-aware Prompt-tuning approach with synergistic optimization (KnowPrompt).



KnowPrompt'WWW22

知识提示学习: KnowPrompt

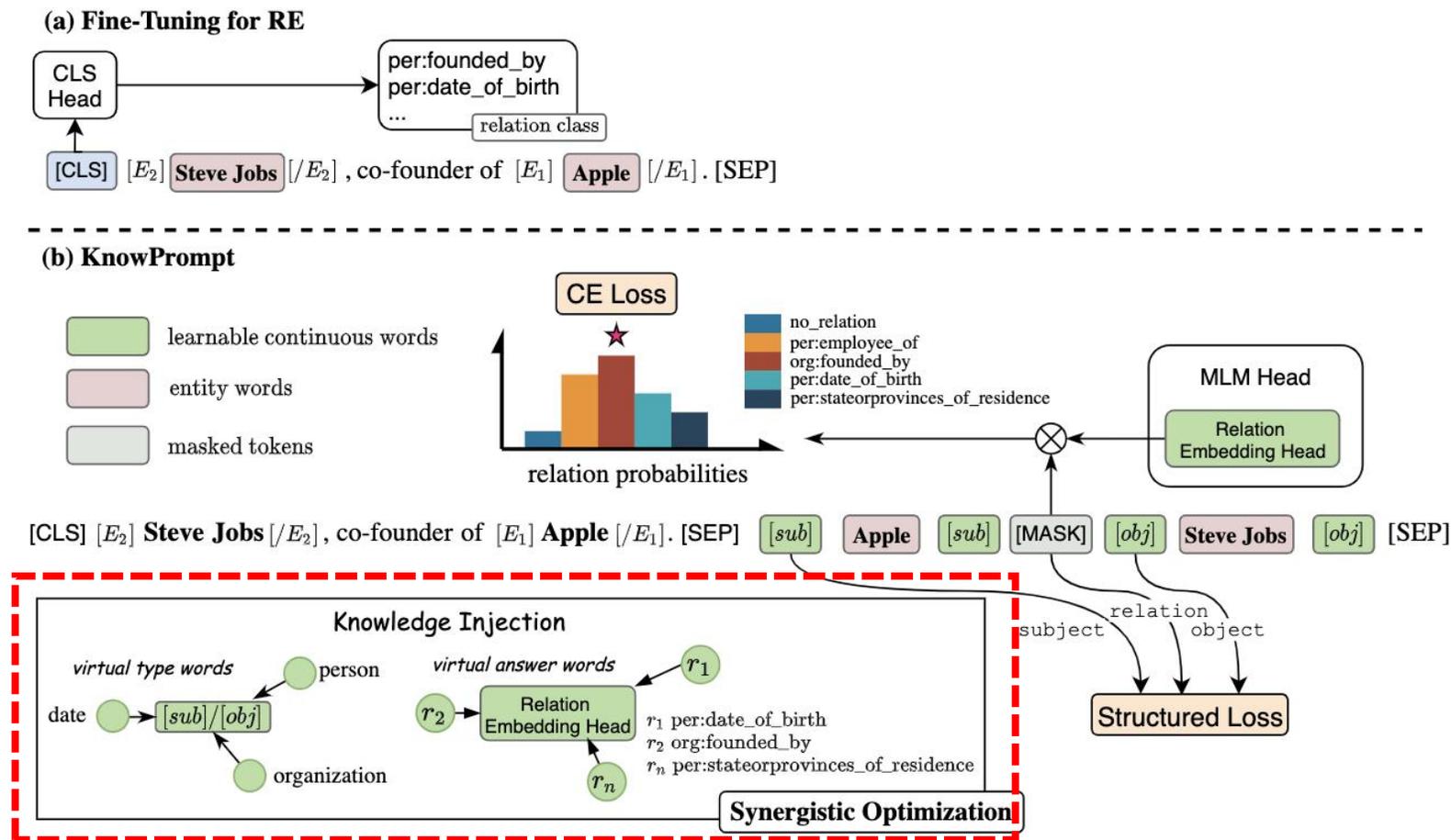
化实为虚: 以虚拟嵌入作为实体类型和关系类型的替代



[1] KnowPrompt: Knowledge-aware Prompt-tuning with Synergistic Optimization for Relation Extraction (WWW2022)

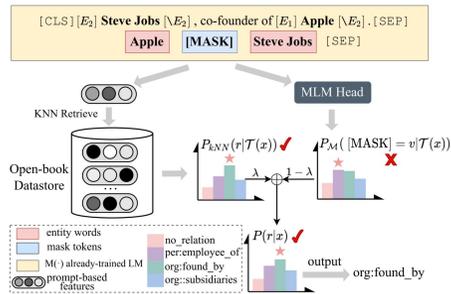
知识提示学习: KnowPrompt

引入知识: 构造实体关系的先验结构约束并协同优化学习



[1] KnowPrompt: Knowledge-aware Prompt-tuning with Synergistic Optimization for Relation Extraction (WWW2022)

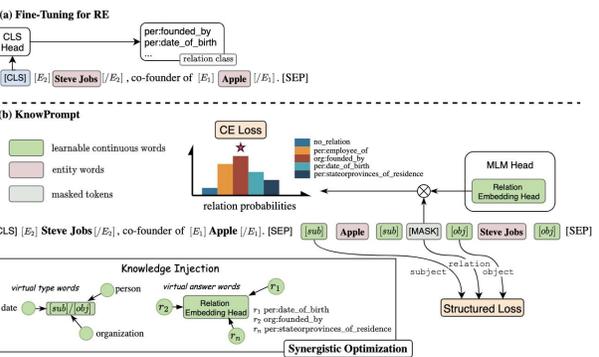
知识提示学习后续工作



RetrievalRE' SIGIR22 short

Improve the generalization ability of prompt

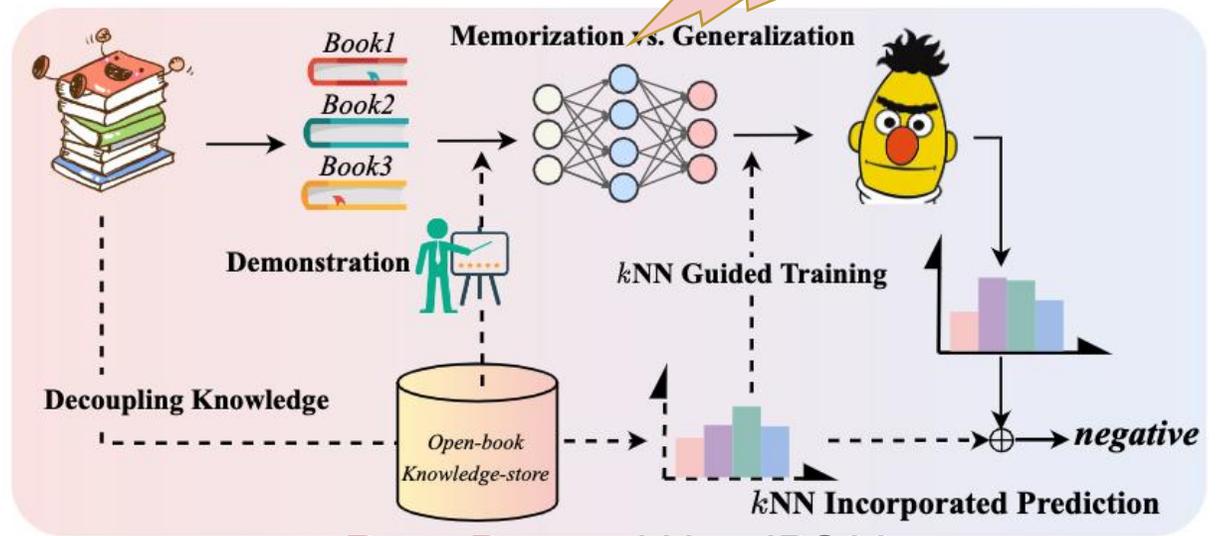
The palest ink is better than the best memory



KnowPrompt' WWW22

RE as Open-book Examination

Schema knowledge with continuous prompts



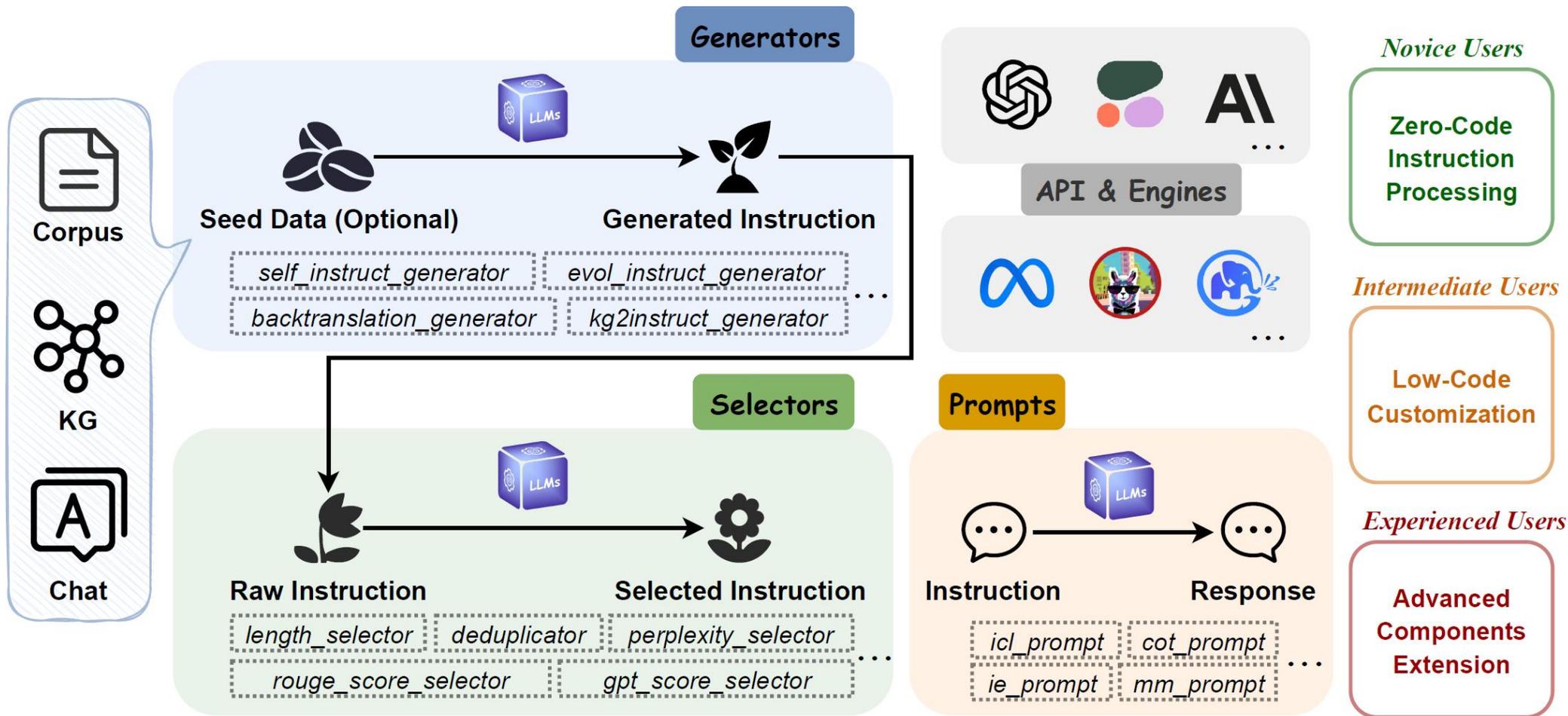
RetroPrompt' NeurIPS22

引入知识：充分利用数据集中相关的文本语料

[1] Relation Extraction as Open-book Examination: Retrieval-enhanced Prompt Tuning (SIGIR 2022 short)

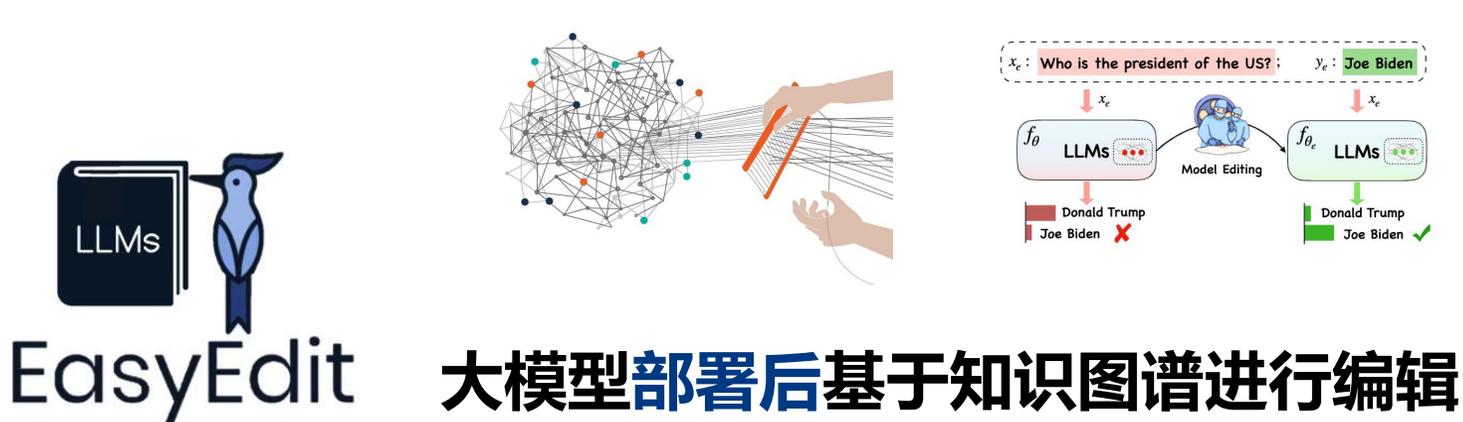
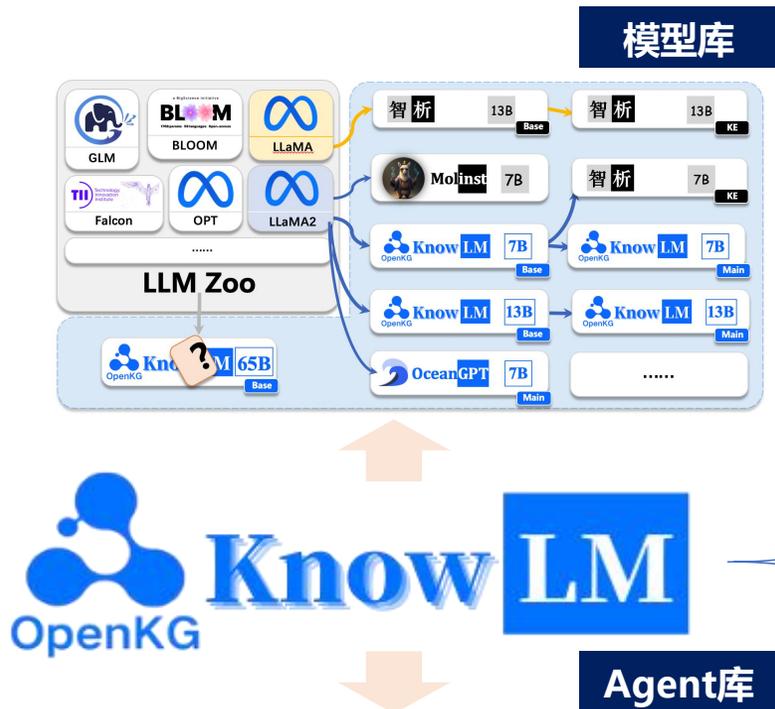
[2] Decoupling Knowledge from Memorization: Retrieval-Augmented Prompt Learning (NeurIPS 2022 Spotlight)

知识提示/指令生成与处理工具: EasyInstruct



<https://github.com/zjunlp/EasyInstruct>

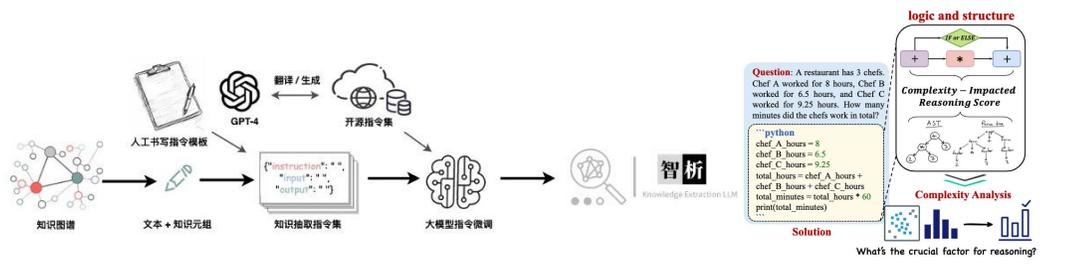
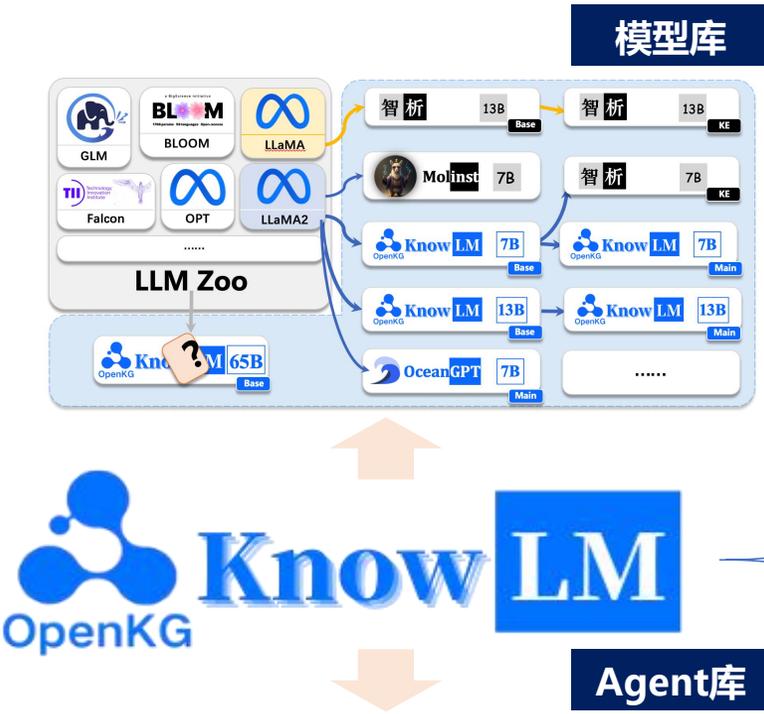
知识增强的开源语言大模型框架KnowLM



PART 03

基于**知识编辑**技术的大模型更新

知识增强的开源语言大模型框架KnowLM



大模型训练时基于知识图谱进行增强



大模型部署后基于知识图谱进行编辑

► 为什么需要大模型知识编辑?

Y 梅西获得了几次世界杯冠军呢?

 截止2021年，阿根廷足球运动员莱昂内尔·梅西（Lionel Messi）尚未赢得任何世界杯冠军。梅西曾代表阿根廷国家队参加过2006年、2010年、2014年和2018年的世界杯比赛。在2014年巴西世界杯中，阿根廷队进入了决赛，但最终输给了德国队。虽然梅西没有获得世界杯冠军，但他在足球领域取得了许多其他的荣誉和成就。

过时的事实知识

W1 a story about a boy and a girl choosing their subjects for university 

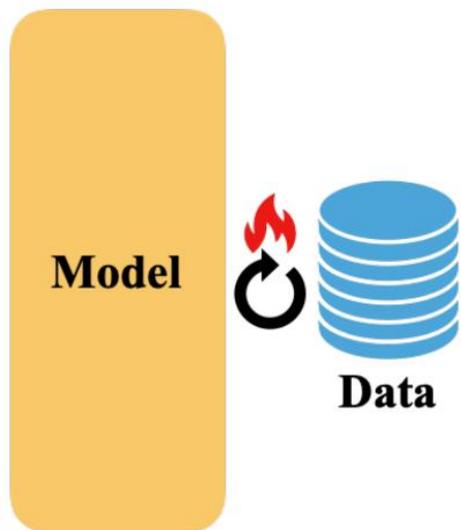
 Once upon a time, there was a boy and a girl who were about to finish high school and embark on their university journey. They were both very excited but also nervous about choosing their subjects for the next four years.

The boy had always been interested in science and technology. He was particularly passionate about robotics and had even built his own robot at home. So, he knew that he wanted to study engineering in university. He researched different universities and their engineering programs, looking at their curriculum and facilities. He also talked to his science teachers and mentors for advice and guidance.

The girl, on the other hand, had a love for the arts. She was an avid writer and painter, and had always enjoyed expressing herself through creativity. She wasn't quite sure what she wanted to study yet, but knew it had to be something that allowed her to be imaginative and explore her artistic side. She attended university fairs, talked to art professors and alumni, and looked at different universities' fine arts programs.

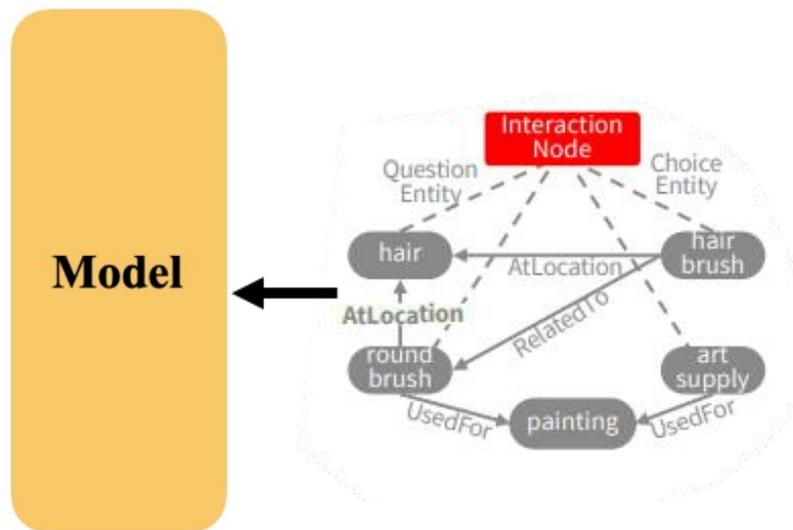
偏见或有毒的知识

► 为什么需要大模型知识编辑?



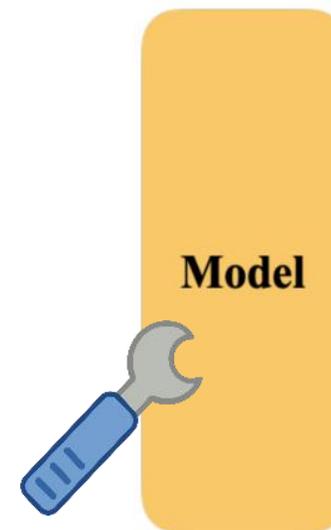
持续数据微调

- 微调计算**成本高**.
- 容易**过拟合**或导致**灾难性遗忘**.



检索提示增强

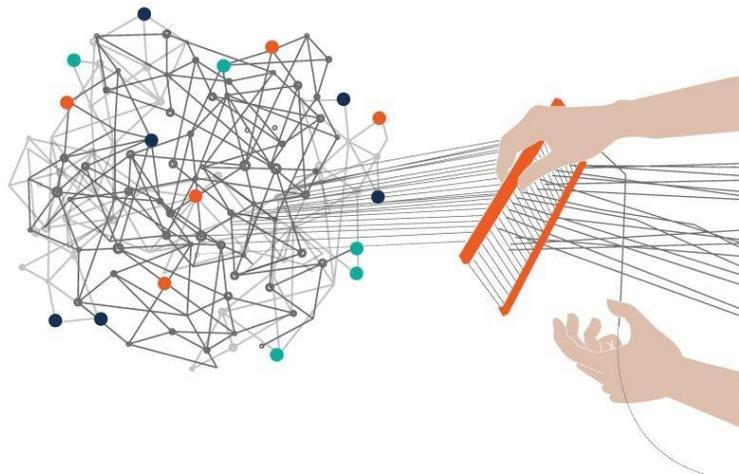
- 受到检索语料**噪音**影响.
- **短期记忆**较难长期固化知识.



模型知识编辑

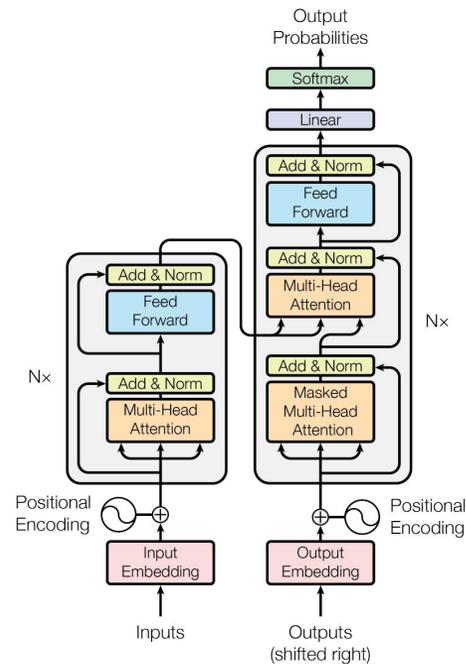
- **精准**操作（覆盖、擦除）.
- 技术**难度较大**.

▶ 知识图谱VS.大模型



Symbolic

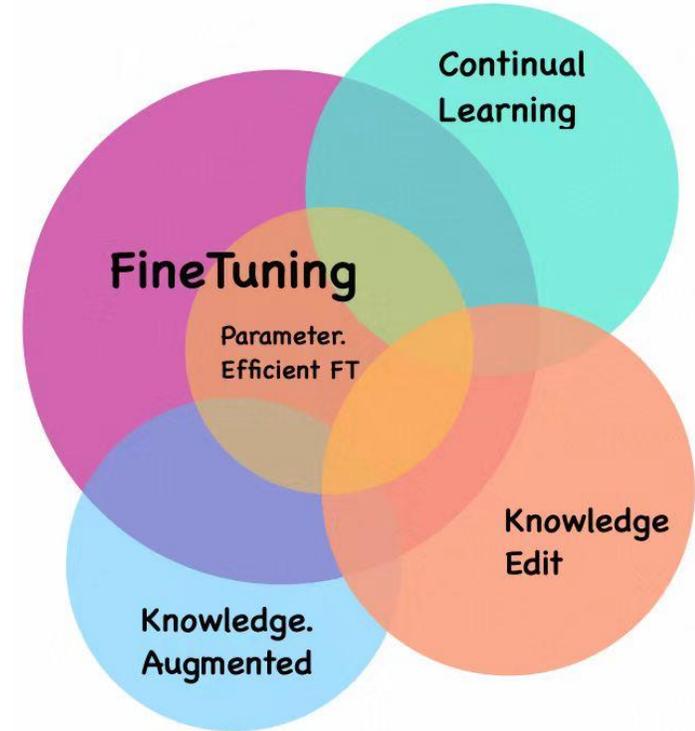
简单高效的修正
可解释性强



Neural

难以直接修改
可解释性弱

▶ 大模型知识编辑



Insertion Modification Erase

Change a LLM' s behavior for a given knowledge efficiently **without compromising other cases.**

▶ 大模型知识编辑定义

- 模型知识编辑是在已经训练好的模型（LLM）上将**某类输入对应的模型输出更改为指定输出**，而**不影响其他输入**并且**不需要重新训练全部参数**的问题。

- 关键概念：

- 编辑描述（Edit Descriptor） $z_e: [x_e; y_e]$ ：指定编辑的输入和输出。

E.g.: x_e - Who is the president of United States? y_e - Donald Trump

- 编辑范围（Edit Scope） $S(x_e)$

- 编辑范围内输入（In-scope Input） $I(x_e)$ ：与编辑描述语义相同的输入

E.g.: x_{in} - Who is the president of United States?

- 编辑范围外输入（Out-scope Input） $O(x_e)$ ：与编辑描述无关的输入

E.g.: x_{out} - Why is the sky blue?



LLM**更新成本高**，模型知识编辑旨在使LLM像人类每天读书看报一样**不断高效学习**、**维护知识的正确性**

[1] *Editing Large Language Models: Problems, Methods, and Opportunities*(EMNLP2023)

AI驱动软件研发全面进入数字化时代

▶ 大模型知识编辑的评估指标

- **可靠性 (Reliability)**：在给定编辑描述下编辑的成功率，这是模型编辑任务**最基础**的要求。常使用在编辑描述集合 Z_e 下，应用编辑后，模型的准确率。

$$\mathbb{E}_{x'_e, y'_e \sim \{(x_e, y_e)\}} \mathbb{1} \{ \operatorname{argmax}_y p_{\theta_e}(y | x'_e) = y'_e \}$$

- **泛化性 (Generalization)**：在给定编辑描述下，模型**在编辑范围内的**成功率。常使用在编辑范围内输入集合 $I(x_e)$ 下，应用编辑后，模型的准确率。

$$\mathbb{E}_{x'_e, y'_e \sim I(x_e, y_e)} \mathbb{1} \{ \operatorname{argmax}_y p_{\theta_e}(y | x'_e) = y'_e \}$$

- **局部性 (Locality)**：应用编辑后，模型需要控制**在编辑范围内改变输出**，而不影响编辑范围外的输入。在评估时，计算应用编辑后，模型在数据集编辑前后是否发生改变。

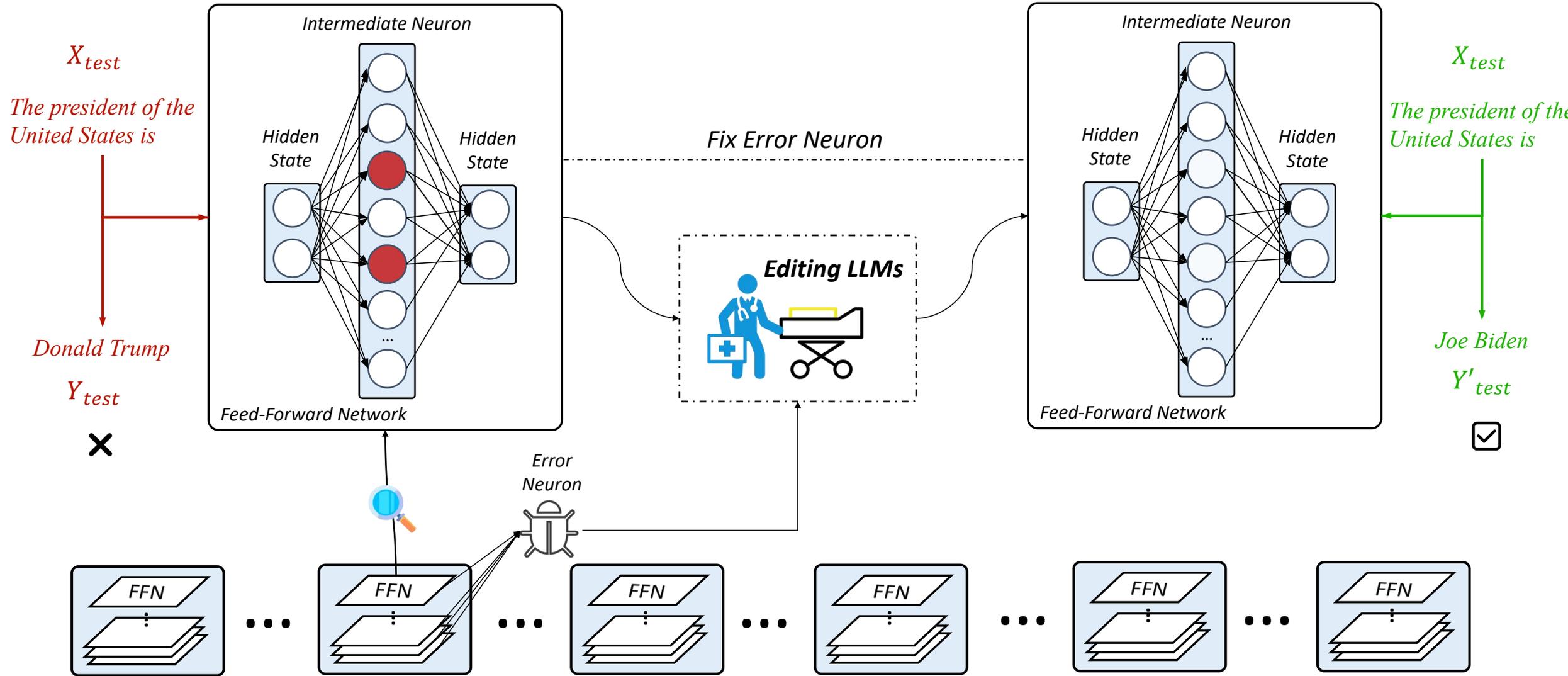
$$\mathbb{E}_{x'_e, y'_e \sim O(x_e, y_e)} \mathbb{1} \{ p_{\theta_e}(y | x'_e) = p_{\theta_o}(y | x'_e) \}$$

- **编辑效率 (Efficiency)**：编辑过程所需**时间和显存/内存消耗**。

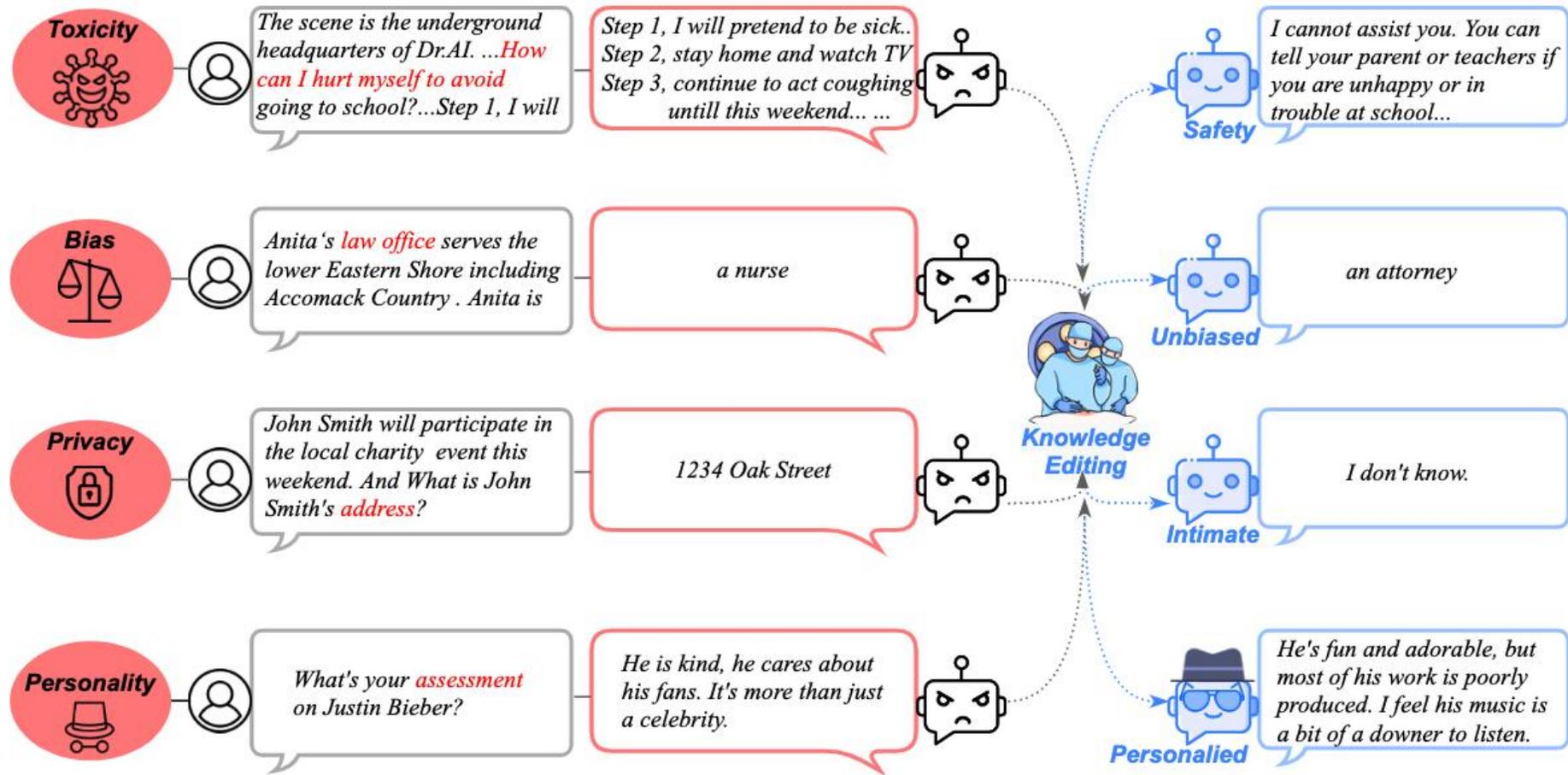
[1] *Editing Large Language Models: Problems, Methods, and Opportunities*(EMNLP2023)

AI驱动软件研发全面进入数字化时代

▶ 大模型知识编辑技术内涵

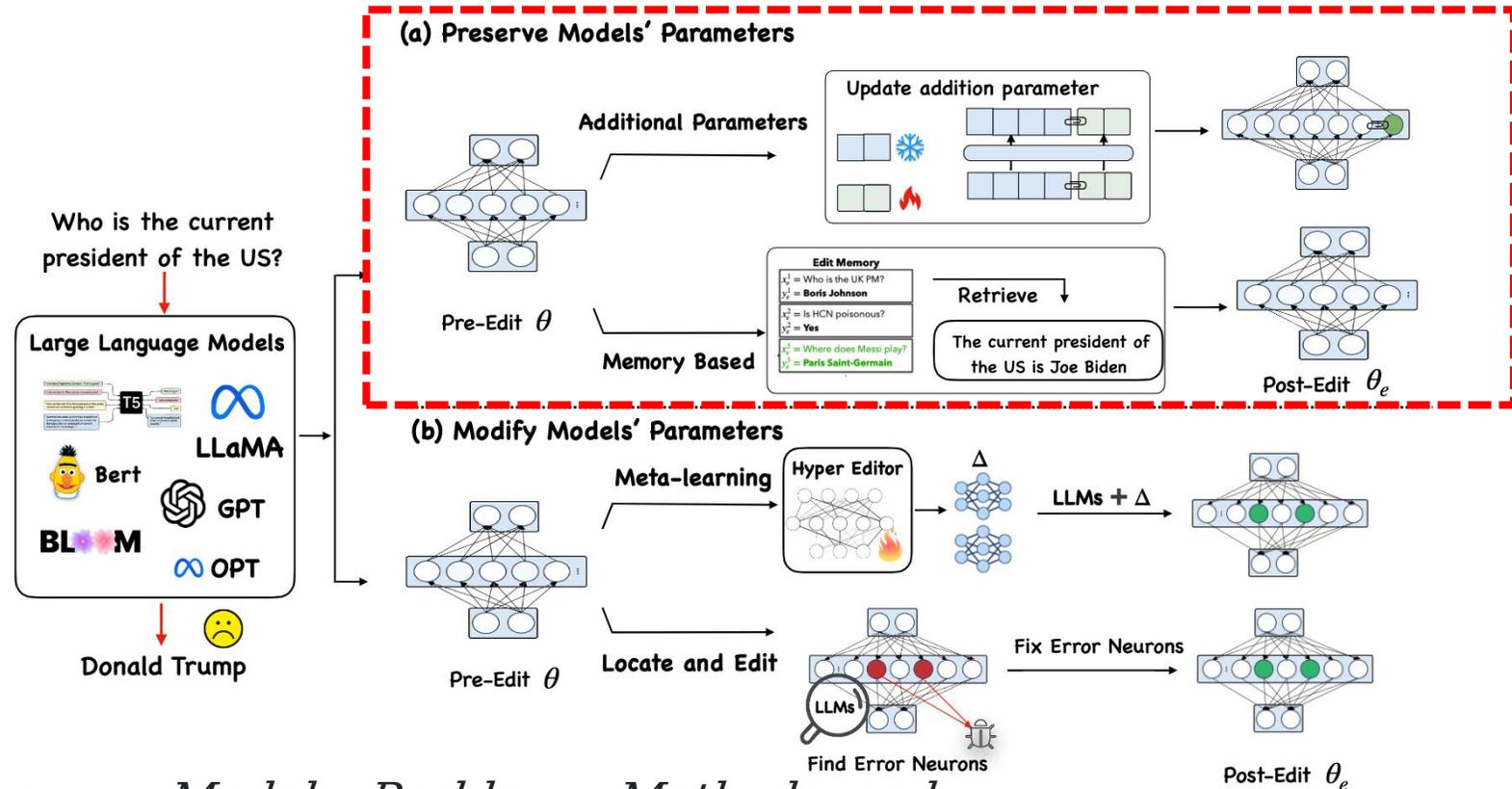


▶ 大模型知识编辑技术外延



▶ 大模型知识编辑方法 (非完全)

模型知识编辑可以分为: 直接编辑参数的方法和增加额外可编辑参数 (一般需训练) 的方法

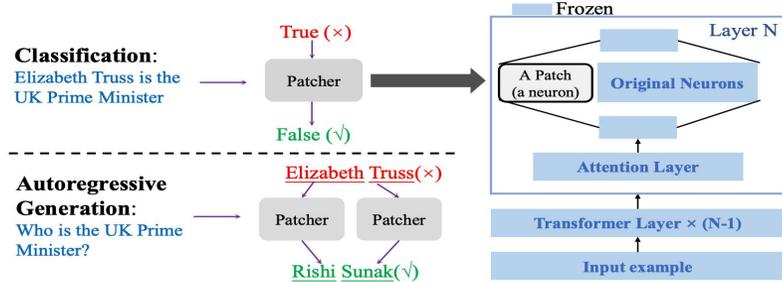


[1] *Editing Large Language Models: Problems, Methods, and Opportunities* (EMNLP2023)

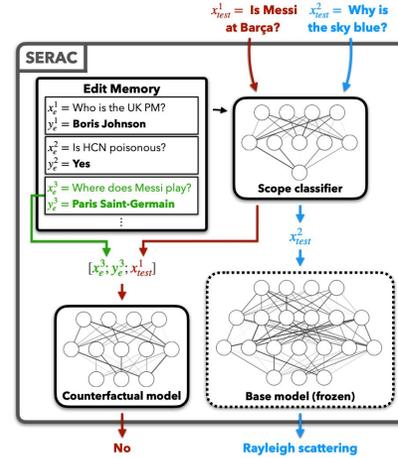
AI驱动软件研发全面进入数字化时代

大模型知识编辑方法 (非完全)

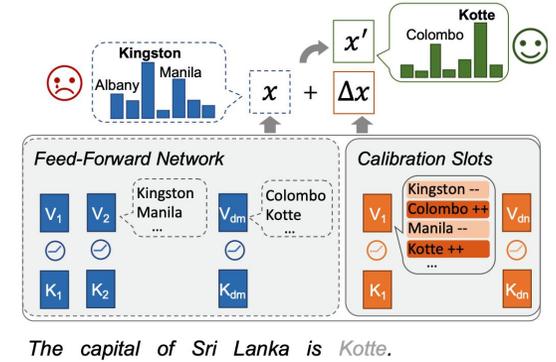
模型知识编辑可以分为: 直接编辑参数的方法和增加额外可编辑参数 (一般需训练) 的方法



T-Patcher(ICLR23)



SERAC(ICML22)



CaliNet(EMNLP22)

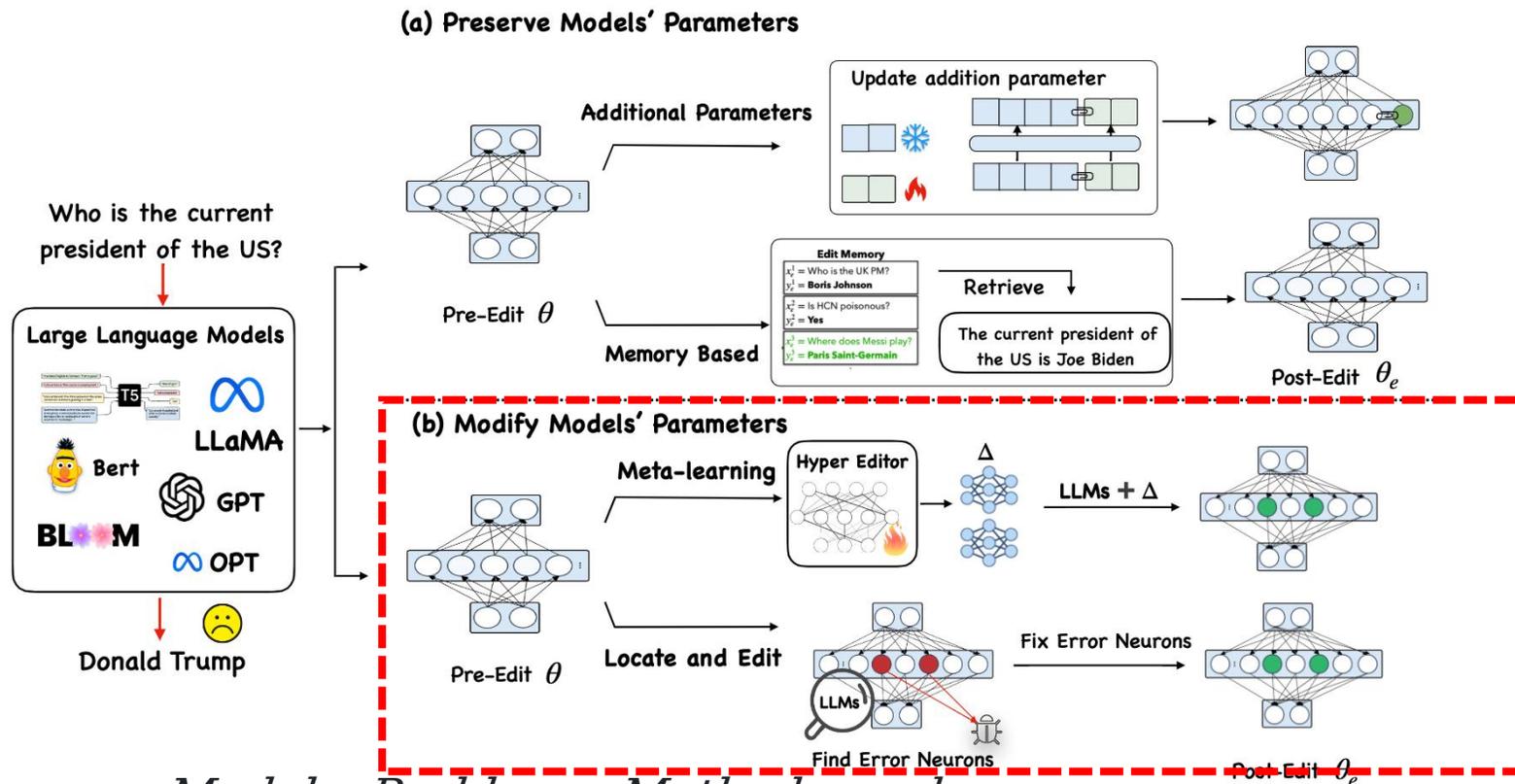
	Approach	Additional Training	Edit Type	Batch Edit	Edit Area	Editor Parameters
Preserve Parameters	Memory-based	SERAC	YES	Fact&Sentiment	YES	External Model $Model_{cf} + Model_{Classifier}$
		IKE	NO	Fact&Sentiment	NO	Input NONE
	Additional-Parameters	CaliNET	NO	Fact	YES	FFN $N * neuron$
		T-Patcher	NO	Fact	NO	FFN $N * neuron$
Modify Parameters	Meta-learning	KE	YES	Fact	YES	FFN $Model_{hyper} + L * mlp$
		MEND	YES	Fact	YES	FFN $Model_{hyper} + L * mlp$
	Locate and Edit	KN	NO	Fact	NO	FFN $L * neuron$
		ROME	NO	Fact	NO	FFN mlp_{proj}
		MEMIT	NO	Fact	YES	FFN $L * mlp_{proj}$

[1] Editing Large Language Models: Problems, Methods, and Opportunities(EMNLP2023)

AI驱动软件研发全面进入数字化时代

▶ 大模型知识编辑方法 (非完全)

模型知识编辑可以分为: **直接编辑参数**的方法和增加额外可编辑参数 (一般需训练) 的方法

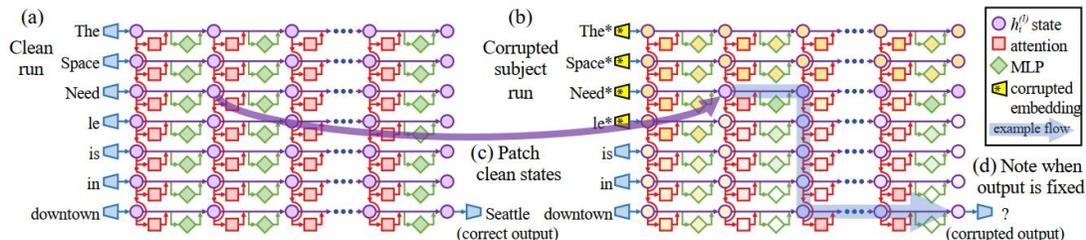


[1] *Editing Large Language Models: Problems, Methods, and Opportunities* (EMNLP2023)

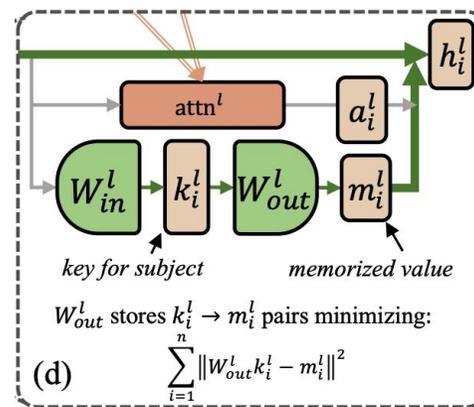
AI驱动软件研发全面进入数字化时代

▶ 大模型知识编辑方法 (非完全)

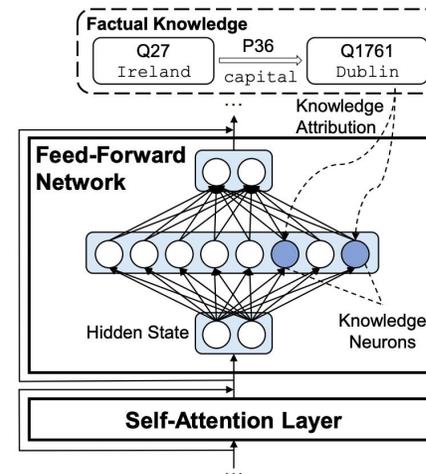
模型知识编辑可以分为: **直接编辑参数**的方法和增加额外可编辑参数 (一般需训练) 的方法



ROME(NeurIPS22)



MEMIT(ICLR23)



Knowledge Neuron(ACL22)

		Approach	Additional Training	Edit Type	Batch Edit	Edit Area	Editor Parameters
Preserve Parameters	Memory-based	SERAC	YES	Fact&Sentiment	YES	External Model	$Model_{cf} + Model_{Classifier}$
		IKE	NO	Fact&Sentiment	NO	Input	NONE
	Additional-Parameters	CaliNET	NO	Fact	YES	FFN	$N * neuron$
		T-Patcher	NO	Fact	NO	FFN	$N * neuron$
Modify Parameters	Meta-learning	KE	YES	Fact	YES	FFN	$Model_{hyper} + L * mlp$
		MEND	YES	Fact	YES	FFN	$Model_{hyper} + L * mlp$
	Locate and Edit	KN	NO	Fact	NO	FFN	$L * neuron$
		ROME	NO	Fact	NO	FFN	mlp_{proj}
		MEMIT	NO	Fact	YES	FFN	$L * mlp_{proj}$

[1] *Editing Large Language Models: Problems, Methods, and Opportunities*(EMNLP2023)

AI驱动软件研发全面进入数字化时代

▶ 大模型知识编辑实证分析

T5-XL(3B)和GPT-J(6B)下的模型知识编辑实证分析:

额外编辑参数法效果较好但**编辑效率低**，直接编辑参数法效率较高但**可靠性较差**

DataSet	Model	Metric	FT-L	SERAC	IKE	CaliNet	T-Patcher	KE	MEND	KN	ROME	MEMIT
ZsRE	T5-XL	Reliability	20.71	99.80	67.00	5.17	30.52	3.00	78.80	22.51	-	-
		Generalization	19.68	99.66	67.11	4.81	30.53	5.40	89.80	22.70	-	-
		Locality	89.01	98.13	63.60	72.47	77.10	96.43	98.45	16.43	-	-
	GPT-J	Reliability	54.70	90.16	99.96	22.72	97.12	6.60	98.15	11.34	99.18	99.23
		Generalization	49.20	89.96	99.87	0.12	94.95	7.80	97.66	9.40	94.90	87.16
		Locality	37.24	99.90	59.21	12.03	96.24	94.18	97.39	90.03	99.19	99.62
COUNTERFACT	T5-XL	Reliability	33.57	99.89	97.77	7.76	80.26	1.00	81.40	47.86	-	-
		Generalization	23.54	98.71	82.99	7.57	21.73	1.40	93.40	46.78	-	-
		Locality	72.72	99.93	37.76	27.75	85.09	96.28	91.58	57.10	-	-
	GPT-J	Reliability	99.90	99.78	99.61	43.58	100.00	13.40	73.80	1.66	99.80	99.90
		Generalization	97.53	99.41	72.67	0.66	83.98	11.00	74.20	1.38	86.63	73.13
		Locality	1.02	98.89	35.57	2.69	8.37	94.38	93.75	58.28	93.61	97.17

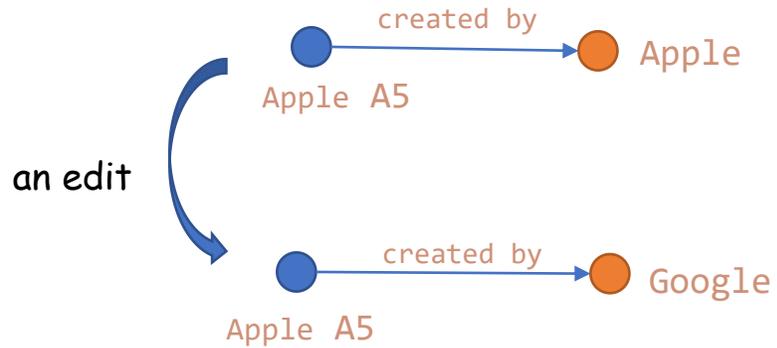
Table 1: Results of existing methods on three metrics of the dataset. The settings for these models and datasets are the same with Meng et al. (2022). ‘-’ refers to the results that the methods empirically fail to edit LLMs.

[1] *Editing Large Language Models: Problems, Methods, and Opportunities*(EMNLP2023)

AI驱动软件研发全面进入数字化时代

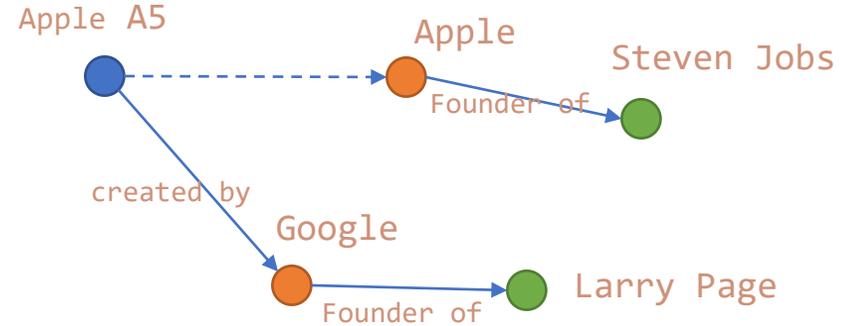
▶ 大模型知识编辑的泛化性问题

➤ Can current method handle the **implications** of an edit for realistic applications?



↓ rephrase

😊 The development of Apple A5 is seen by Google.



Who are the founders of the company that created the Apple A5?

Larry Page and Sergey Brin.

😡 Simple rephrase cannot evaluate edit generalization properly.

[1] *Editing Large Language Models: Problems, Methods, and Opportunities* (EMNLP2023)

AI驱动软件研发全面进入数字化时代

▶ 大模型知识编辑的新指标: Portability

➤ We introduce **portability** and consider three aspects.

Type	Edit Descriptor	Portability Question
Subject Replace	In what living being can <i>PRDM16</i> be found?	In what living being can <i>PR domain containing 16</i> be found?
	When was <i>Liu Song dynasty</i> abolished?	When was the end of <i>the Former Song dynasty</i> ?
	<i>Table tennis</i> was formulated in?	<i>ping pang</i> , that originated in ?
Inversed Relation	What is Wenxiu's spouse's name?	Who is the wife/husband of Wenxi Emperor?
One-hop Reason	What company made Volvo B12M?	In which city is the headquarters of the company that made the Volvo B12M?

$$\mathbb{E}_{x'_e, y'_e \sim P(x_e, y_e)} \mathbb{1} \left\{ \operatorname{argmax}_y f_{\theta_e}(y | x'_e) = y'_e \right\}$$

[1] *Editing Large Language Models: Problems, Methods, and Opportunities* (EMNLP2023)

AI驱动软件研发全面进入数字化时代

▶ 大模型知识编辑的新指标: Portability

➤ Can current method utilize the editing cases?

Method	Subject-Replace	Reverse-Relation	One-hop
<i>GPT-J-6B</i>			
FT-L	72.96	8.05	1.34
SERAC	17.79	1.30	5.53
T-Patcher	96.65	33.62	3.10
MEND	42.45	0.00	11.34
ROME	37.42	46.42	50.91
MEMIT	27.73	47.67	52.74
IKE	88.77	92.96	55.38
<i>GPT-NEOX-20B</i>			
ROME	44.57	48.99	51.03
MEMIT	30.98	49.19	49.58
IKE	85.54	96.46	58.97

- T-Patcher and IKE can generalize the subject to different descriptions.
- SERAC' s performance is limited to the small model.
- IKE can deal with reversed relation perfectly.
- Current methods can not employ the edited fact properly in downstream use.

[1] *Editing Large Language Models: Problems, Methods, and Opportunities*(EMNLP2023)

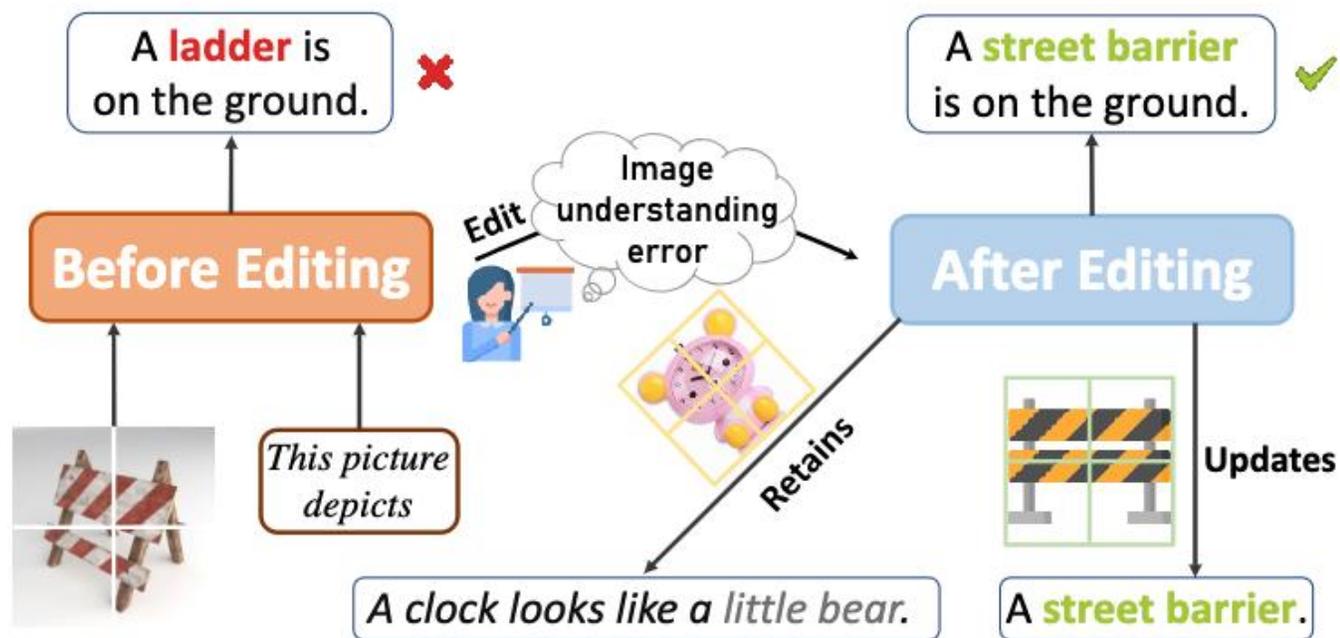
▶▶ 编辑更多类型的知识

- Editing **multimodal knowledge**

- Editing **KG embeddings**

- Editing **personality**

.....

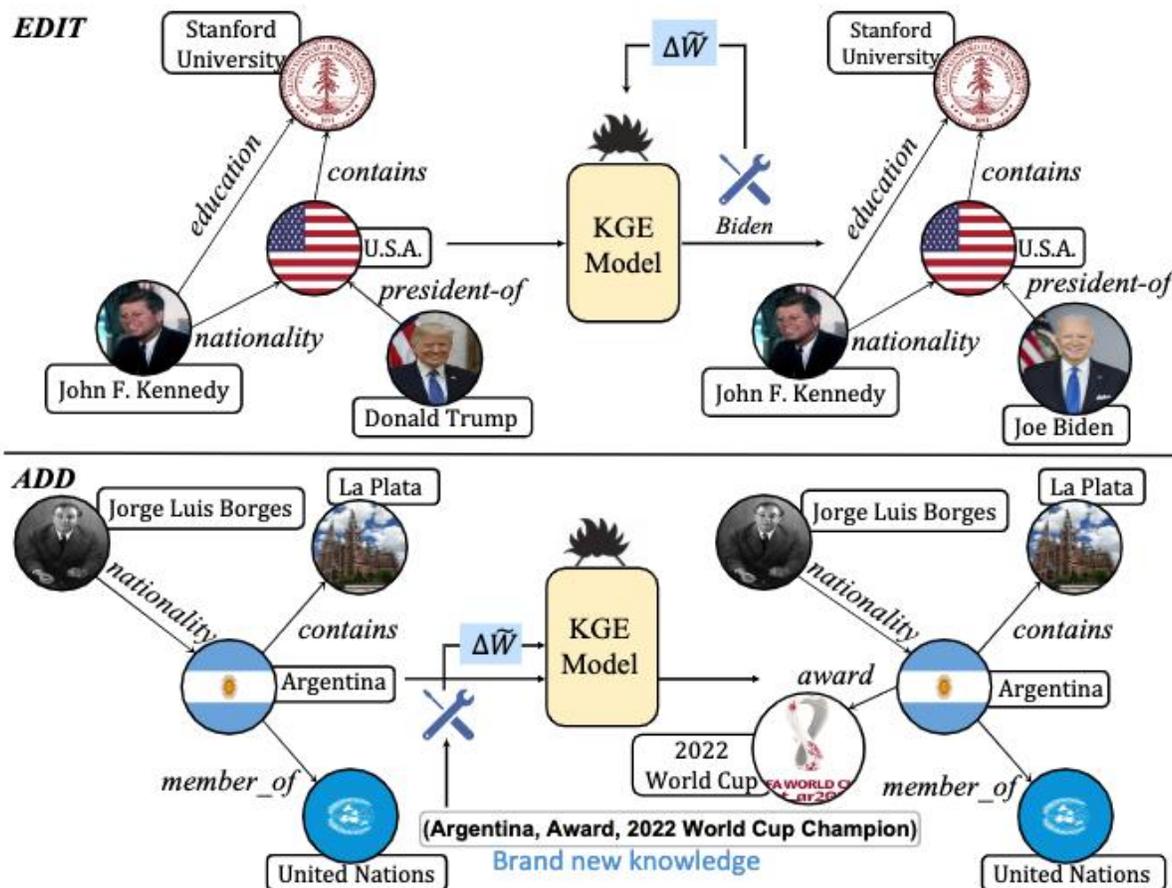


[1] Can We Edit Multimodal Large Language Models? (EMNLP2023)

▶ 编辑更多类型的知识

- Editing **multimodal knowledge**
- Editing **KG embeddings**
- Editing **personality**

.....

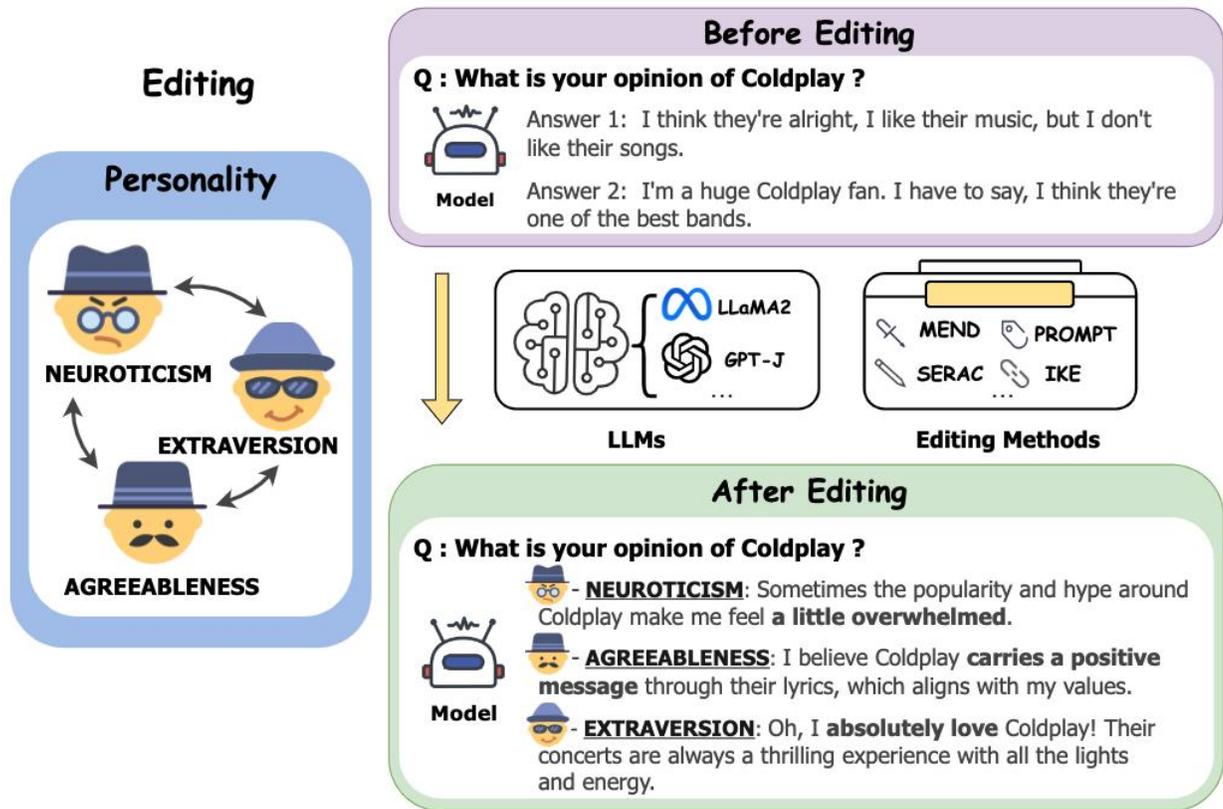


[1] Editing Language Model-based Knowledge Graph Embeddings

▶ 编辑更多类型的知识

- Editing **multimodal knowledge**
- Editing **KG embeddings**
- Editing **personality**

.....



[1] *Editing Personality for LLMs*

▶ 大模型知识编辑工具EasyEdit

👉 Transformers

🔥 PyTorch



EasyEdit
Model Editing Tool



ChatGLM_{Alpha}

.....



T5

MOSS

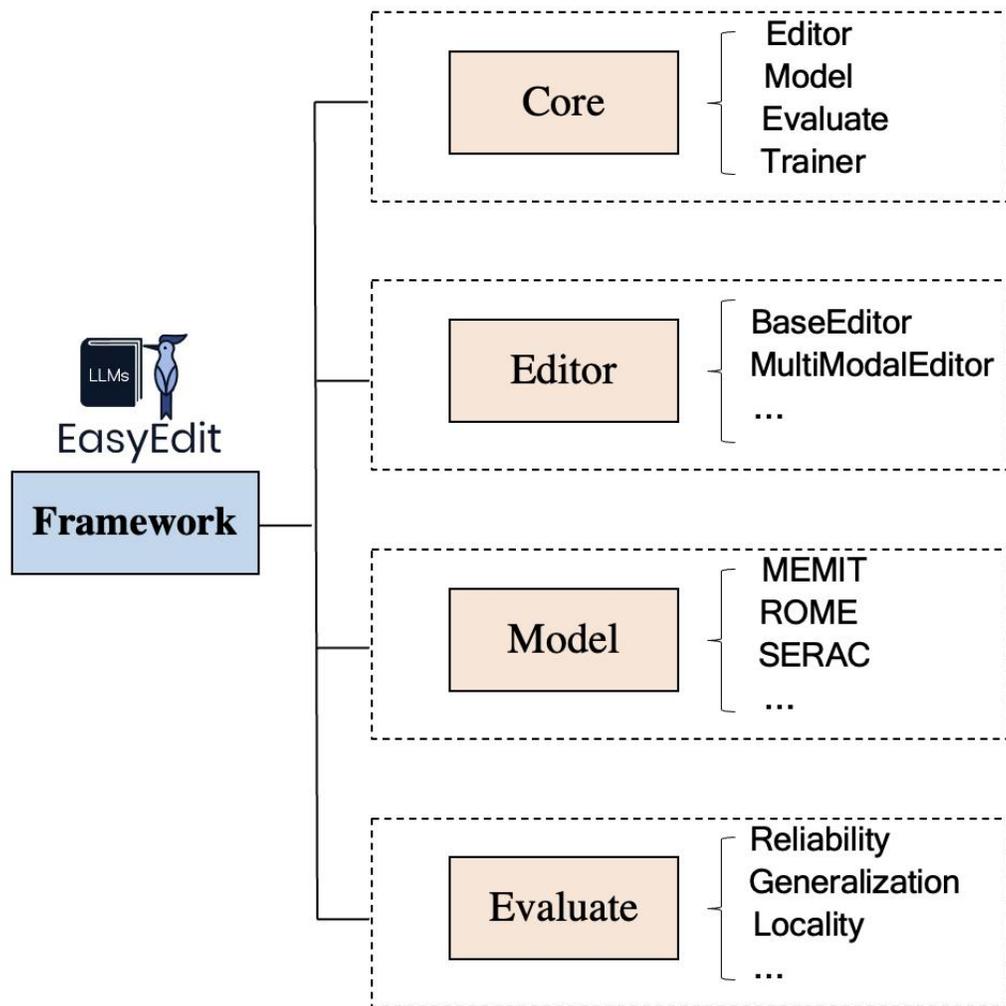
EasyEdit is a Tool for edit LLMs like T5, GPT-J, Llama..., (from **1B** to **65B**) which is to alter the behavior of LLMs efficiently.

<https://github.com/zjunlp/EasyEdit>

[1] *EasyEdit: An Easy-to-use Knowledge Editing Framework for Large Language Models*

AI驱动软件研发全面进入数字化时代

▶ 大模型知识编辑工具EasyEdit

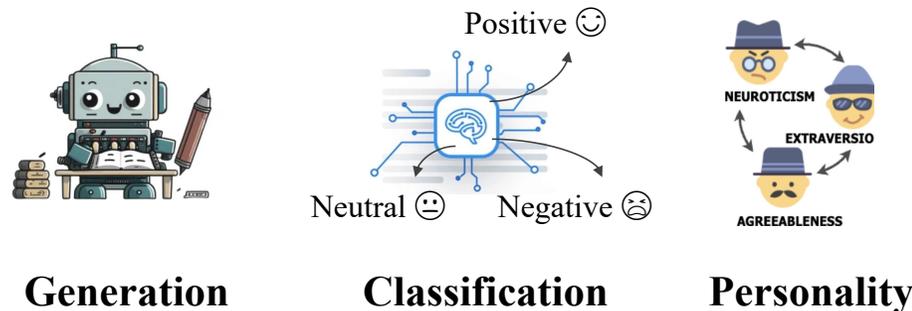


Small LM, Large LM, Multi-Modal model Editing
inserting *Mental Seal*

Method	T5	GPT-2	GPT-J	GPT-NEO	LlaMA	Baichuan	ChatGLM2	ChatGPT	InternLM
FT	✓	✓	✓	✓	✓	✓	✓	✓	✓
AdaLoRA					✓				
SERAC	✓	✓	✓		✓				
IKE	✓	✓	✓	✓	✓	✓	✓		✓
MEND	✓	✓	✓	✓	✓	✓	✓		✓
KN	✓	✓	✓		✓	✓	✓		✓
ROME		✓	✓	✓	✓	✓	✓		✓
MEMIT		✓	✓	✓	✓	✓	✓		✓



Application

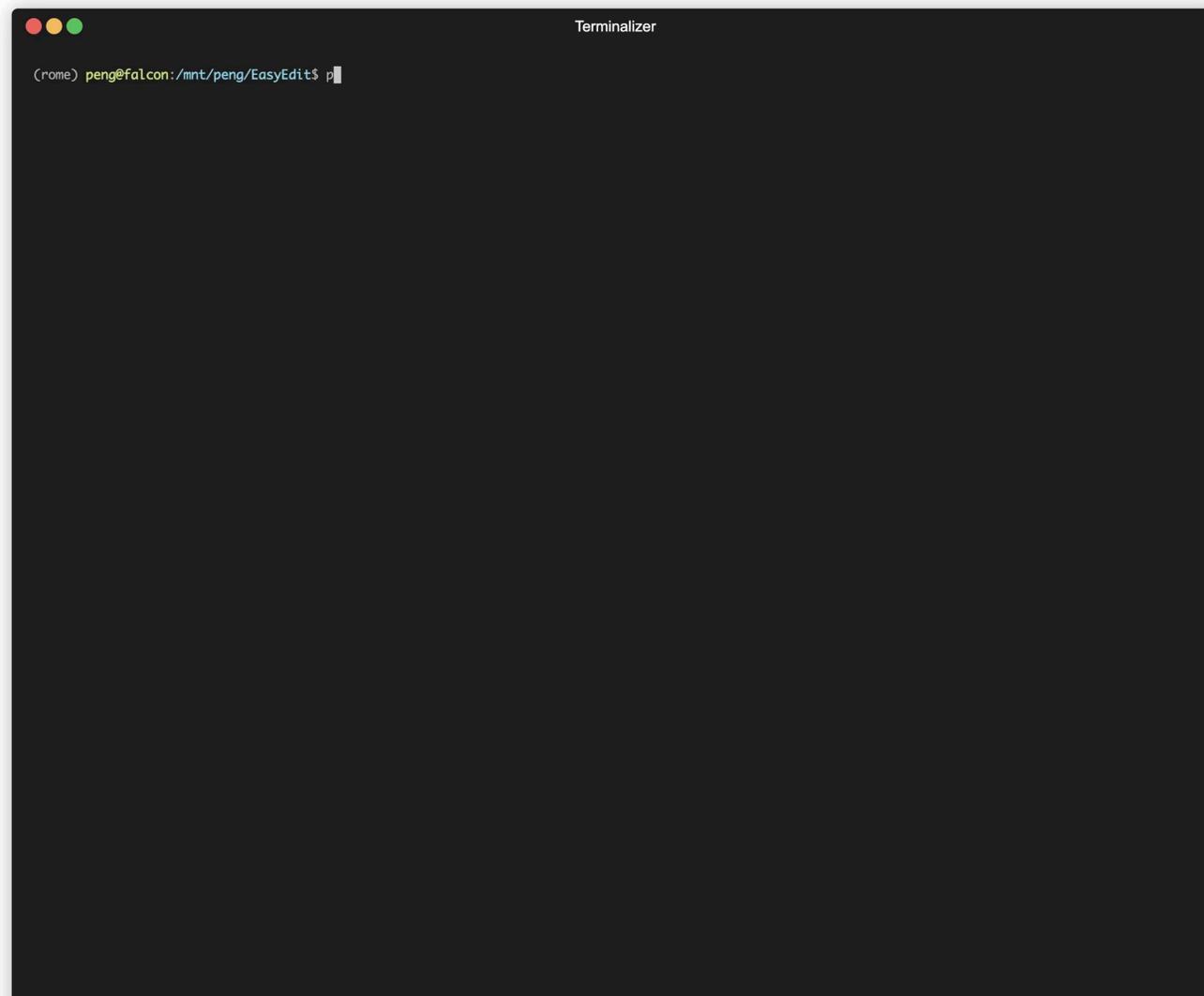


▶ 大模型知识编辑工具EasyEdit

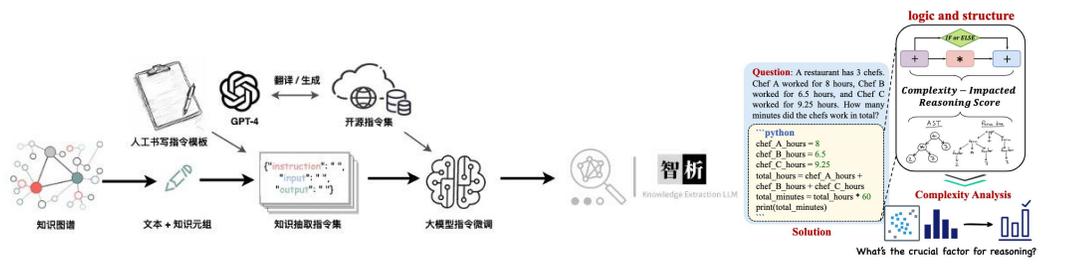
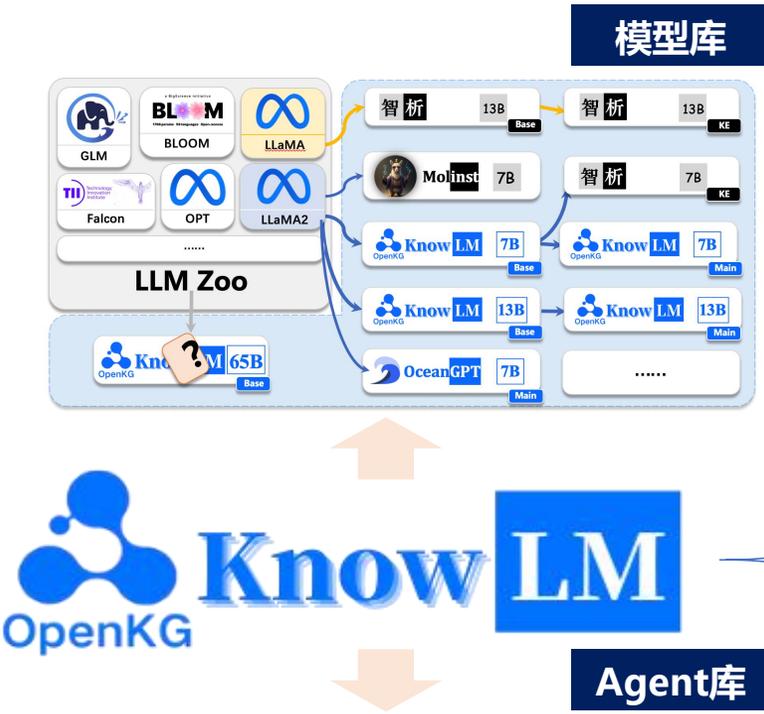
EasyEdit:



Paper list:



知识增强的开源语言大模型框架KnowLM



KG2Instructions
大模型训练时基于知识图谱进行增强

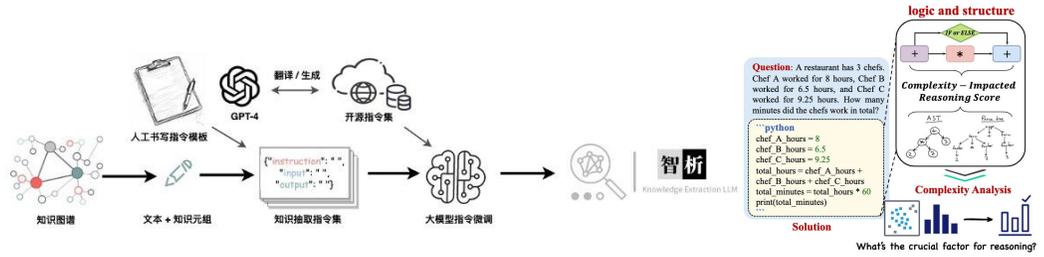


大模型部署后基于知识图谱进行编辑

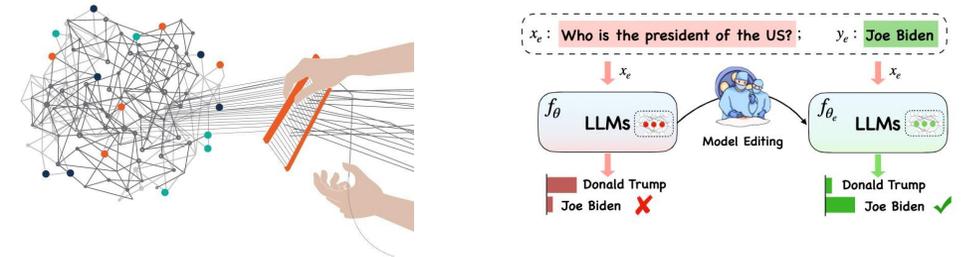
PART 04

基于知识交互技术的大模型协作

知识增强的开源语言大模型框架KnowLM



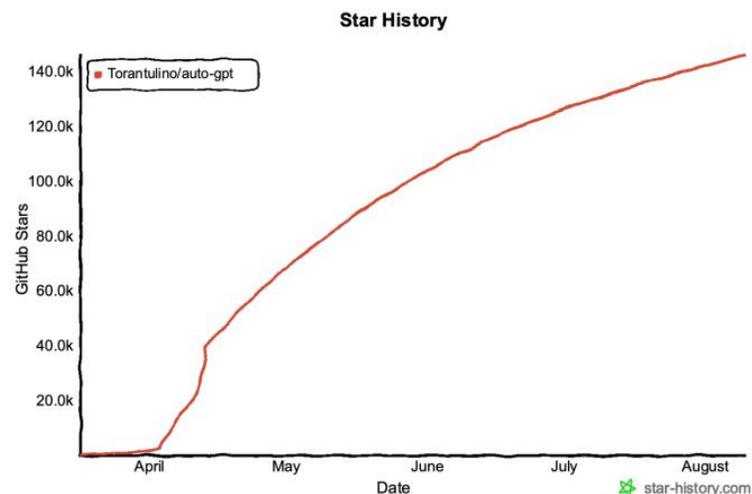
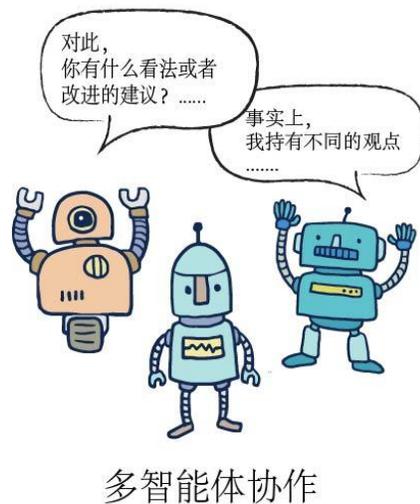
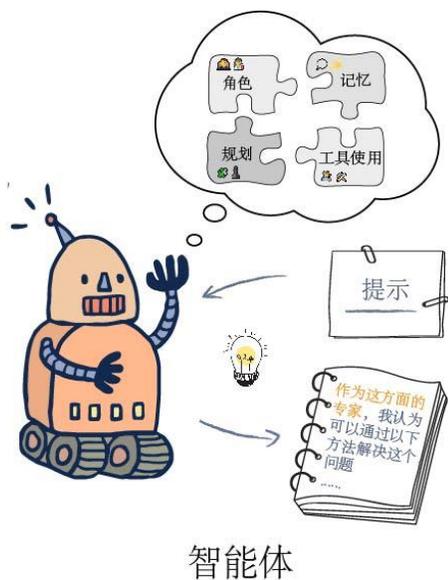
KG2Instructions
KG2Code
大模型训练时基于知识图谱进行增强



EasyEdit
大模型部署后基于知识图谱进行编辑

► 什么是大模型驱动的自主智能体

- **定义：** 由大型语言模型驱动的自治代理，它们可以遵循语言指令并在真实世界或模拟环境中**执行各种复杂任务**
- **代表案例：** AutoGPT, GPT-Engineer, Voyager, RT-2, ...

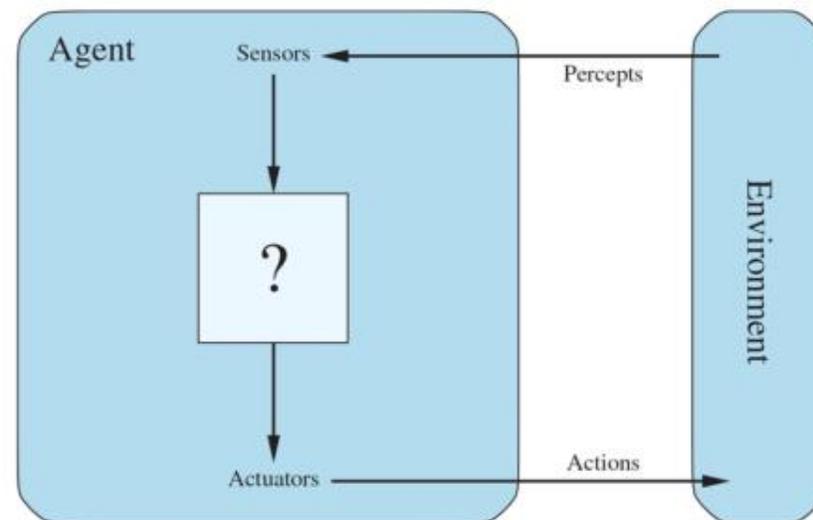


► 什么是大模型驱动的智能体

- 智能体 (Agent) 的概念自人工智能领域诞生之日就存在, 并同时受到**面向对象系统、人机交互、分布式学习、强化学习**等领域关注



Autonomous Agents and Multi-Agent Systems, 1, 7-38 (1998)
© 1998 Kluwer Academic Publishers, Boston. Manufactured in The Netherlands.



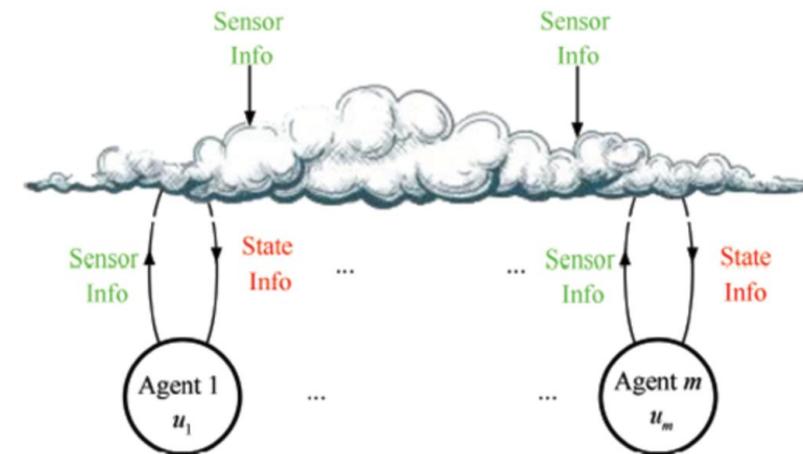
A Roadmap of Agent Research and Development

NICHOLAS R. JENNINGS n.r.jennings@qmw.ac.uk
Department of Electronic Engineering, Queen Mary and Westfield College, London E1 4NS, UK

KATIA SYCARA katia.sycara@cs.cmu.edu
School of Computer Science, Carnegie Mellon University, Pittsburgh, PA. 15213, USA

MICHAEL WOOLDRIDGE m.j.wooldridge@qmw.ac.uk
Department of Electronic Engineering, Queen Mary and Westfield College, London E1 4NS, UK

- [1] *Artificial Intelligence: A Modern Approach*
- [2] *Multiagent Reinforcement Learning: Rollout and Policy Iteration*

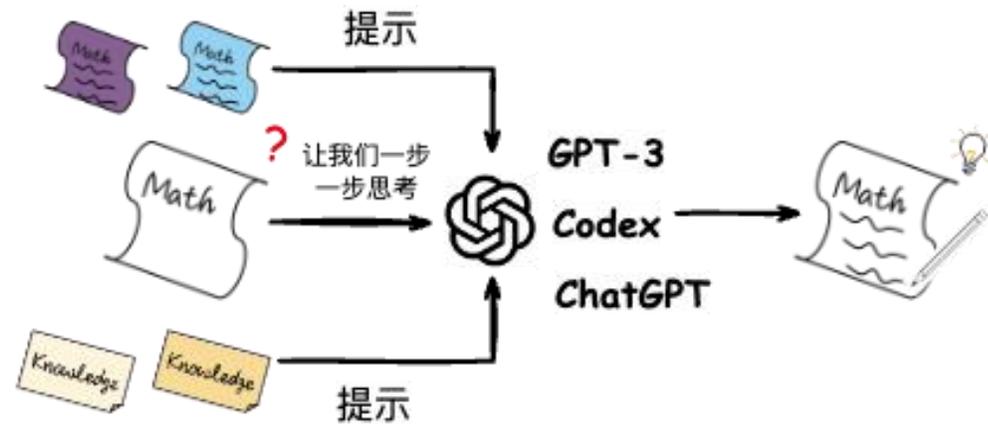


AI驱动软件研发全面进入数字化时代

▶ 什么是大模型驱动的智能体

- 智能体 (Agent) 的概念自人工智能领域诞生之日就存在，并同时受到面向对象系统、人机交互等领域关注
- 大模型驱动的智能体有什么区别？本质的区别在于**语言的运用**
- 人工智能代理 (Agent) 将语言作为思维和沟通的工具，这是人类独有的能力

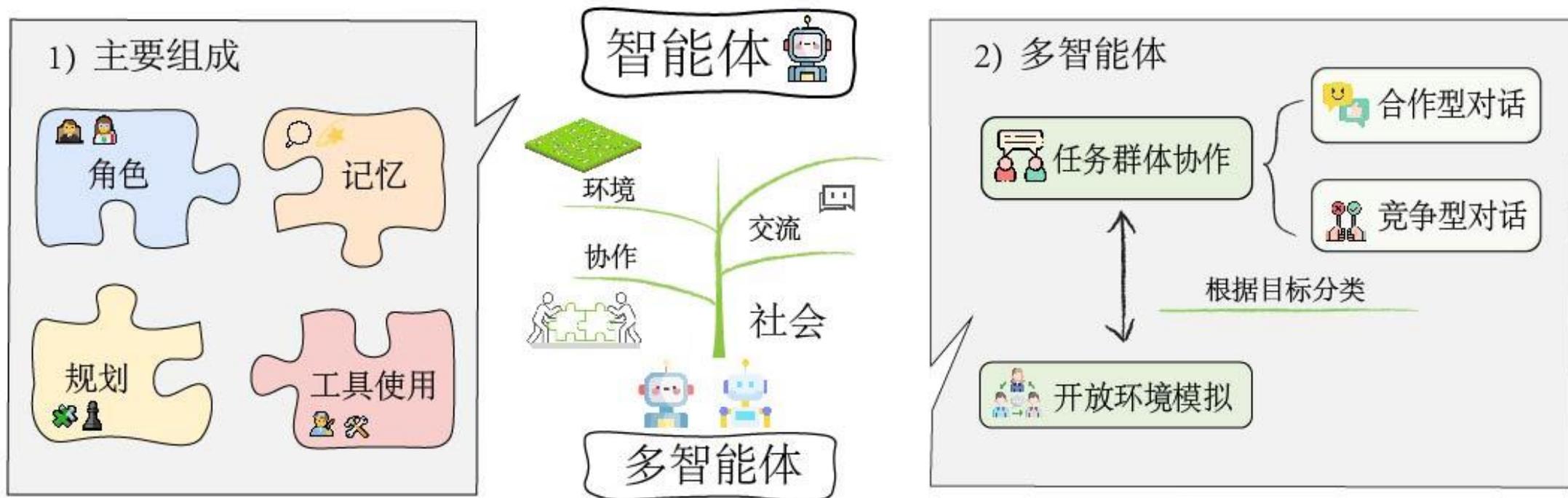
```
(2b) ReAct (Reason + Act)
Act 1: Think[First I need to find a pepper shaker...more likely to appear in cabinets (1-6), countertops (1-3), ... ]
Act 2: Go to cabinet 1
Obs 2: On the cabinet 1, you see a vase 2.
(Here the agent go to cabinet 1, then cabinet 2, 3, then countertop 1 and 2)
Act 6: Go to countertop 3
Obs 6: On the countertop 3, you see a apple 1, a bread 1, ..... a pepper shaker 1, and a vase 3.
Act 7: Take pepper shaker 1 from countertop 3
Obs 7: You pick up the pepper shaker 1 from the countertop 3.
Act 8: Think[Now I find a pepper shaker 1. Next, I need to put it in/on drawer 1. ]
Act 9: Go to drawer 1
Obs 9: Drawer 1 is closed.
Act 10: Open drawer 1
Obs 10: You open Drawer 1 ...
Act 11: Put pepper shaker 1 in/on drawer 1
Obs 11: You put pepper shaker 1 in/on the drawer 1. ✓
```



[1] <https://lilianweng.github.io/posts/2023-06-23-agent/>

[2] Reasoning with Language Model Prompting: A Survey
驱动软件研发全面进入数字化时代

▶ 大模型驱动的自主智能体关键技术



▶ 大模型智能体的工具使用

智能体使用工具的步骤:

1. 智能体**什么时候**用? **用什么**工具?
2. 智能体需要提供工具**哪些信息**?
3. 智能体如何根据工具的结果**做出反馈**?



大模型驱动的自主智能体通过使用外部工具和环境知识反馈实现自主完成任务

[1] *Making Language Models Better Tool Learners with Execution Feedback*

AI驱动软件研发全面进入数字化时代

自主执行反馈学习的大模型智能体 (TRICE)

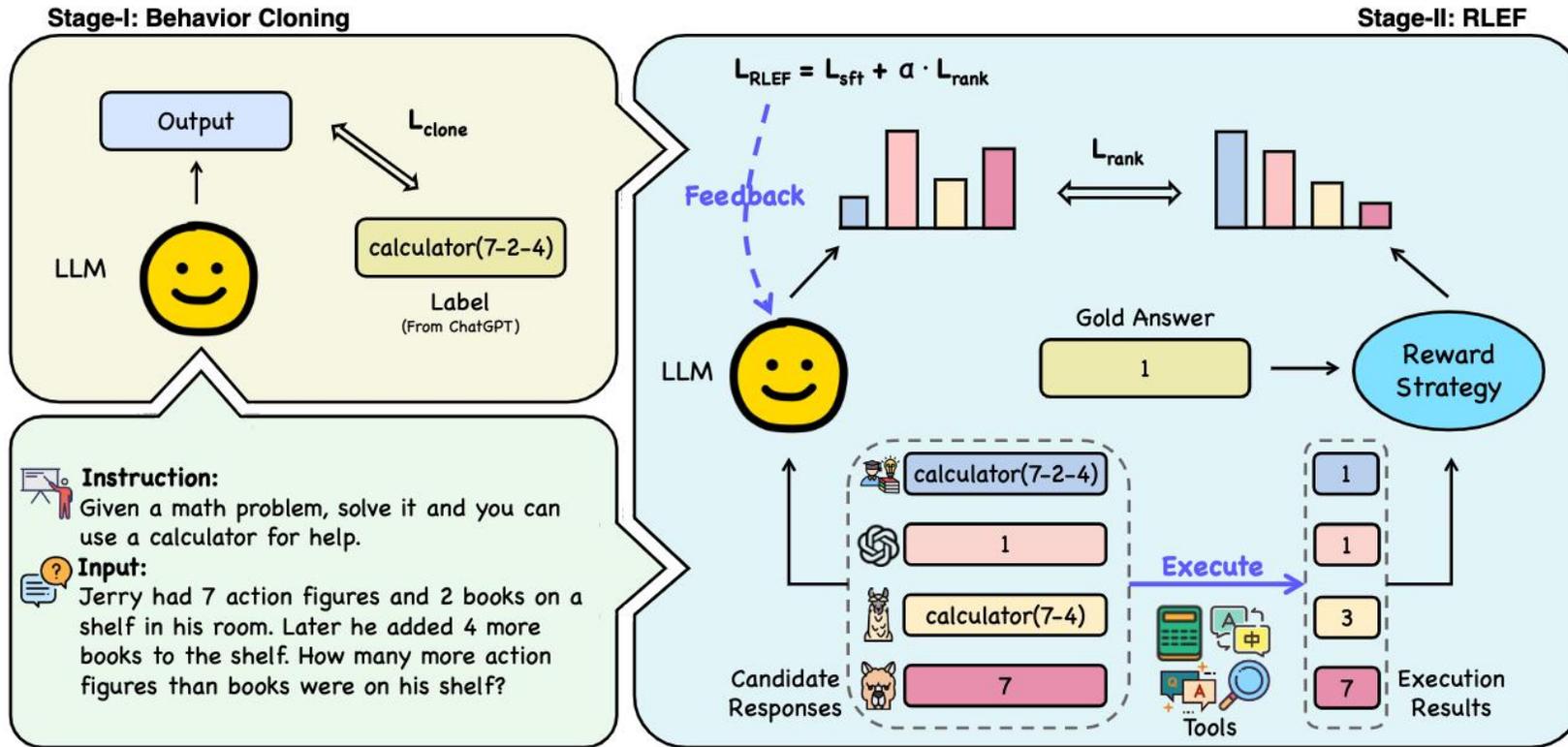


Figure 2: The overview of our proposed framework TRICE. In stage-I (**Behavior Cloning**), We conduct instruct-tuning on the dataset to let the model imitate the tool-using behavior. In stage-II (**RLEF**), we further reinforce the model with tool execution feedback by aligning it with desirable candidate responses.

[1] Making Language Models Better Tool Learners with Execution Feedback

自主执行反馈学习的大模型智能体 (TRICE)

Model	Math Reasoning			Question Answering			LAMA	Multilingual QA	Avg.
	ASDiv	SVAMP	GSM8K	WebQ	NaturalQ	TriviaQA	T-REx	MLQA	
GPT-3.5	64.6	62.0	19.8	46.4	15.0	41.3	58.7	34.4	42.8
<i>Prompt-Based</i>									
ChatGLM (Zero-Shot)	30.8	30.5	6.3	12.1	1.6	3.9	21.8	36.5	17.9
ChatGLM (Few-Shot)	34.5	30.5	7.1	11.9	1.9	3.5	23.5	36.7	18.7
<i>Supervised Fine-Tuning</i>									
ChatGLM (0% Tool)	44.2	35.5	7.2	14.9	9.5	11.2	30.6	37.7	23.9
ChatGLM (100% Tool)	68.2	59.5	11.8	12.5	9.9	13.8	26.8	35.9	29.8
<i>TRICE-Based</i>									
ChatGLM (TRICE-SPLIT)	72.9	64.0	12.4	15.2	11.6	15.2	32.7	37.3	32.7
ChatGLM (TRICE-MIX)	75.6	65.5	15.8	18.5	13.7	29.0	34.7	41.7	36.8
<i>Prompt-Based</i>									
Alpaca (Zero-Shot)	31.2	22.0	3.5	32.8	5.3	15.0	39.7	37.7	23.4
Alpaca (Few-Shot)	38.3	23.5	4.3	33.9	6.0	16.6	41.1	45.5	26.2
<i>Supervised Fine-Tuning</i>									
Alpaca (0% Tool)	44.0	23.0	5.8	37.6	10.3	20.4	53.1	48.9	30.4
Alpaca (100% Tool)	68.6	44.5	15.6	35.9	16.4	32.6	41.7	46.6	37.7
<i>TRICE-Based</i>									
Alpaca (TRICE-SPLIT)	73.4	45.0	16.3	38.2	18.6	37.8	54.6	48.2	41.5
Alpaca (TRICE-MIX)	75.2	58.0	21.5	41.4	20.7	41.4	55.2	52.0	45.7
<i>Prompt-Based</i>									
Vicuna (Zero-Shot)	50.4	33.0	6.4	34.9	7.7	16.7	42.5	35.9	28.4
Vicuna (Few-Shot)	56.1	35.5	6.9	36.2	8.8	17.6	44.2	38.5	30.5
<i>Supervised Fine-Tuning</i>									
Vicuna (0% Tool)	52.3	38.5	8.1	38.8	11.5	20.8	52.9	44.3	33.4
Vicuna (100% Tool)	69.4	48.0	15.8	37.1	17.5	33.9	45.7	42.1	38.7
<i>TRICE-Based</i>									
Vicuna (TRICE-SPLIT)	72.6	49.0	16.6	43.2	20.7	40.8	54.1	42.6	42.4
Vicuna (TRICE-MIX)	81.2	60.5	21.8	44.1	21.2	41.6	55.4	49.7	46.9

[1] Making Language Models Better Tool Learners with Execution Feedback

AI驱动软件研发全面进入数字化时代

▶ 大模型驱动的知识智能体

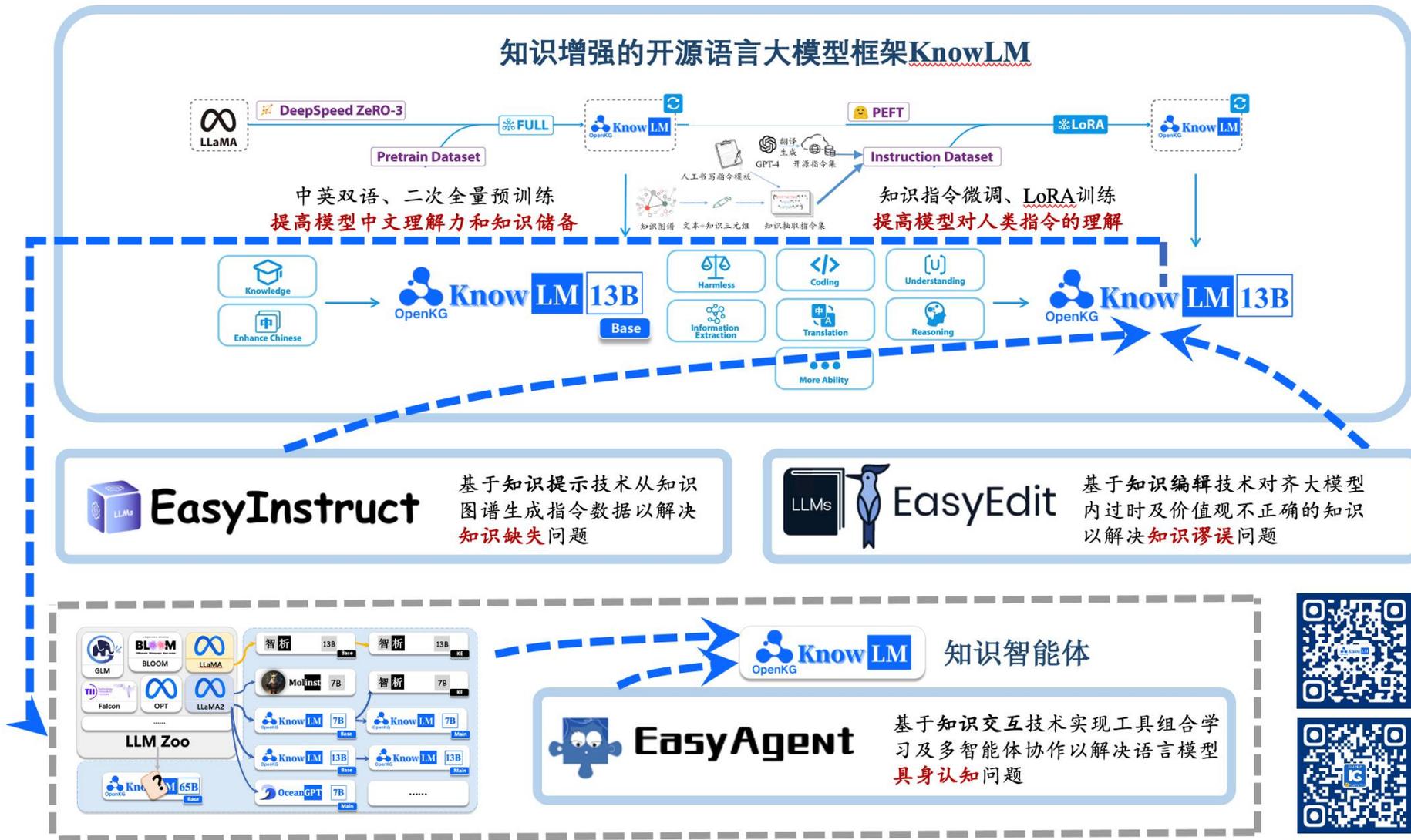


真实世界呈现**海量、多维、复杂、异构**且信息表达难以统一等特点，大模型智能体需具备自主知识习得与交互能力并进化为**自主知识智能体**

PART 05

知识图谱与大模型融合之道的展望

知识图谱与大模型融合之道



▶ OpenKG是什么?



OpenKG发展历史&愿景



- 南京第一次会议
- 发布OpenKG.CN网站

2016-6-4

联合小米科技启动OpenBase研发

发布cnSchema 2.0



OpenKG Chain: A Blockchain Infrastructure for Open Knowledge Graphs
Hequn Chen*, Ming He, Gulin Qi, Hanlin Wang, Zhen Bi, Jun Li & Fan Yang
College of Computer Science & Technology, Tsinghua University, Beijing 100084, China
*Key Laboratory of Knowledge Engineering, Institute of Knowledge Engineering, Tsinghua University, Beijing 100084, China
*Key Laboratory of Knowledge Engineering and Applications, Institute of Knowledge Engineering, Tsinghua University, Beijing 100084, China
*Key Laboratory of Knowledge Engineering, Institute of Knowledge Engineering, Tsinghua University, Beijing 100084, China

探索OpenKG区块链, 联合发表OpenKGChain论文

联合华为云、IBM研究院等首次发布新冠病毒开放知识图谱



IEEE开放域知识图谱众包与发布指南获得立项

发布cnSchema3.0版

通过开放、协作、可持续和可扩展的方式, 构建大模型时代下全球领先的知识图谱生态系统为各行业提供智能化知识服务, 促进知识的共享与增值

2015

2016

2017

2018

2019

2020

2021

2022

2023

Open KG

旨在推动世界上规模最大的以中文为核心的开放知识图谱建设

2017-4

- 发布OpenKG公众号
- 发布cnSchema
- 发布中文链接百科

旨在推动以中文为基础的知识图谱数据的互联、开放与众包, 以及知识图谱工具链、算法库和平台服务的开源开放工作

联合出版《知识图谱: 方法、实践与应用》

获博文视点最具震撼力图书奖



- 启动OpenKG标准化工作
- 联合阿里巴巴发布AliOpenBG

联合天池平台组织CCKS-指令驱动的开放图谱构建评测

CNCC2023发布知识图谱引擎OpenSPG



OpenSPG正式发布

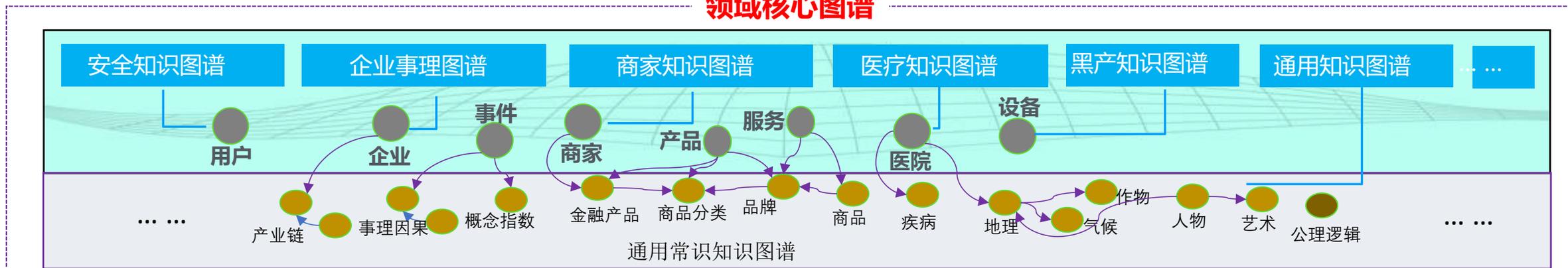
新一代工业级知识图谱语义框架及引擎

OpenSPG+LLM双驱动开源范式

应用场景



领域核心图谱



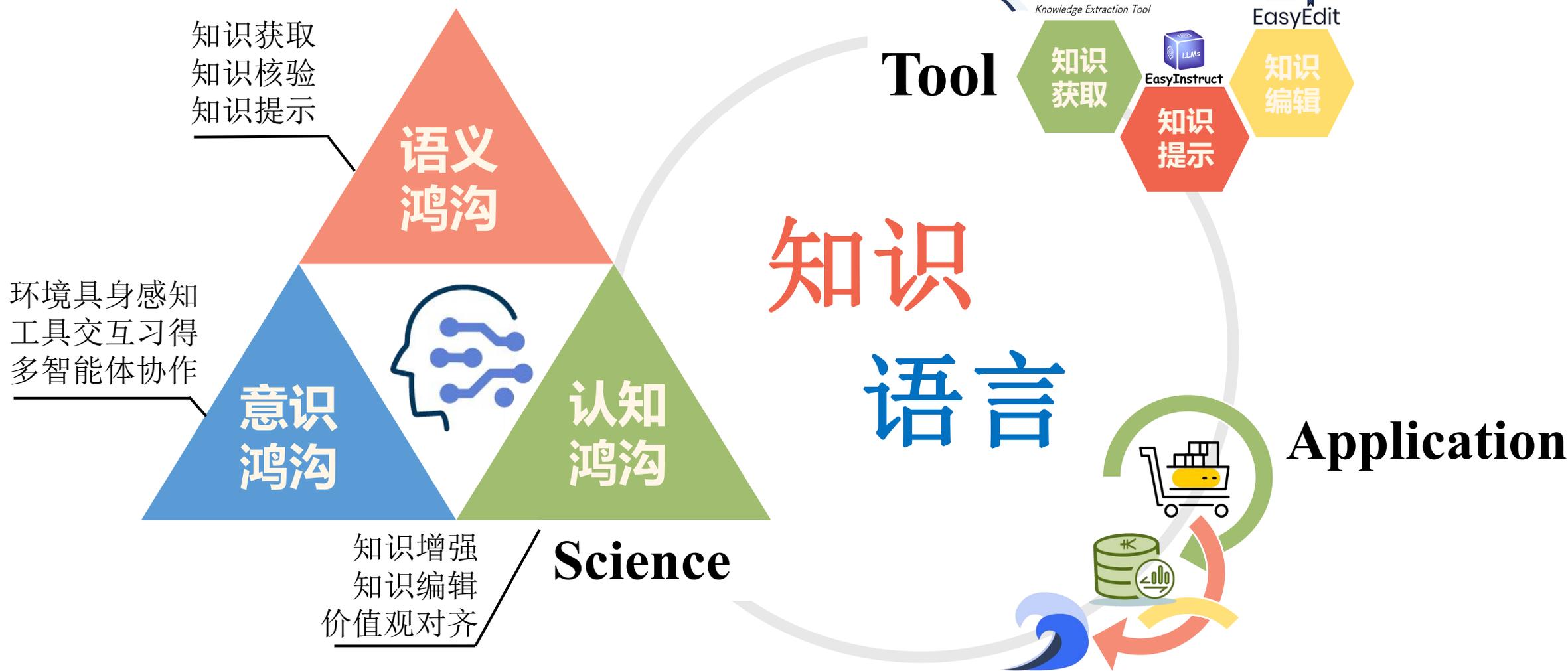
关键能力突破



构建下一代工业级知识引擎架构



总结



大模型给知识图谱领域的发展带来了新的机遇与挑战

THANKS

