CCKS 2023 Tutorial

The Life Cycle of Knowledge in Large Language Models

Hongyu Lin & Boxi Cao

Chinese Information Processing Laboratory
Institute of Software, Chinese Academy of Sciences

Evolution of LLMs

- Large language models have demonstrated extremely powerful abilities in almost all directions of NLP

Understanding

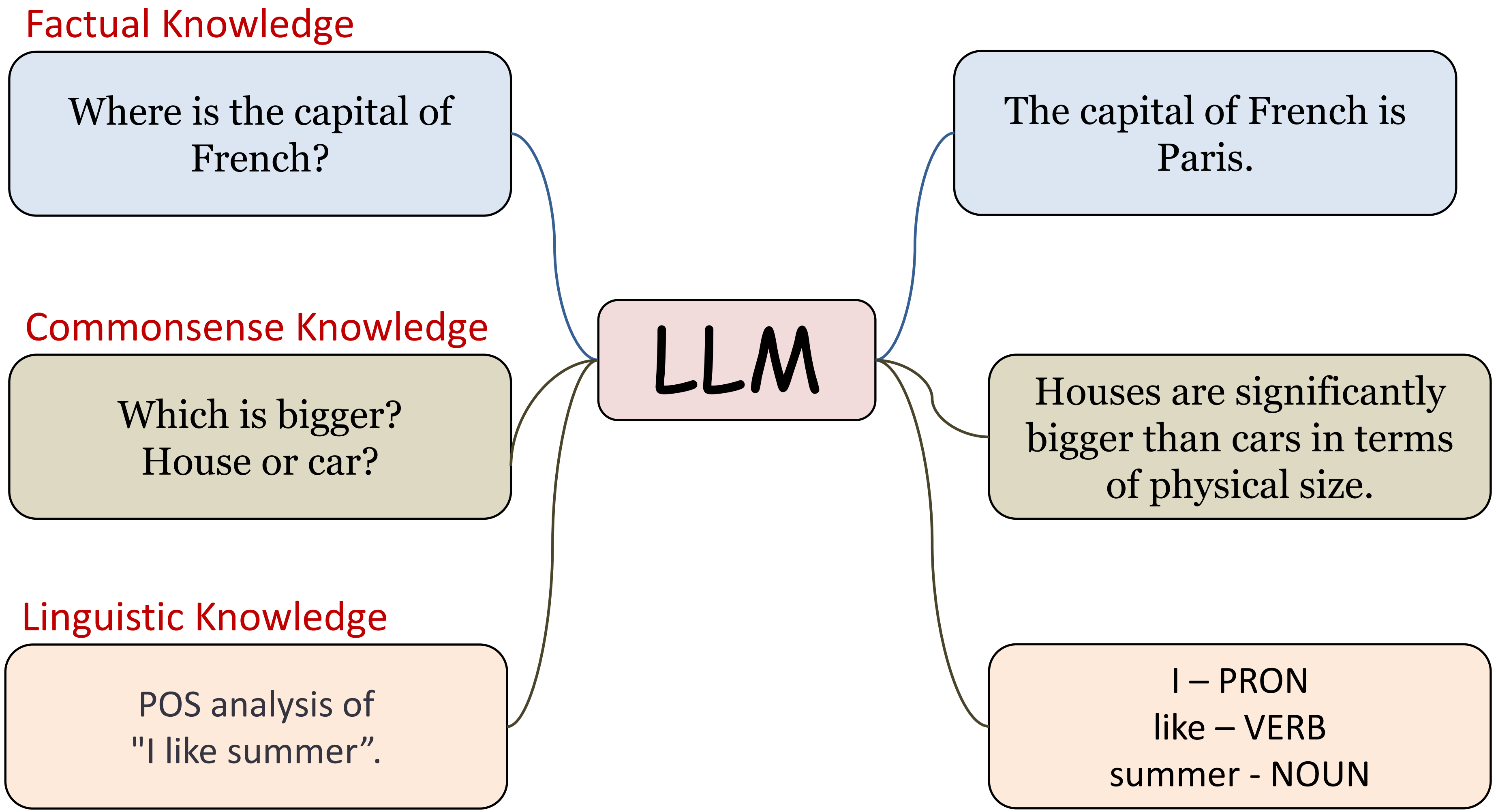
Generation

Decision

Excution

Knowledge in LLMs

- Knowledge in LLMs is critical for their success



Factual Knowledge

Where is the capital of French?

The capital of French is Paris.

LLM

Commonsense Knowledge

Which is bigger?
House or car?

Houses are significantly bigger than cars in terms of physical size.

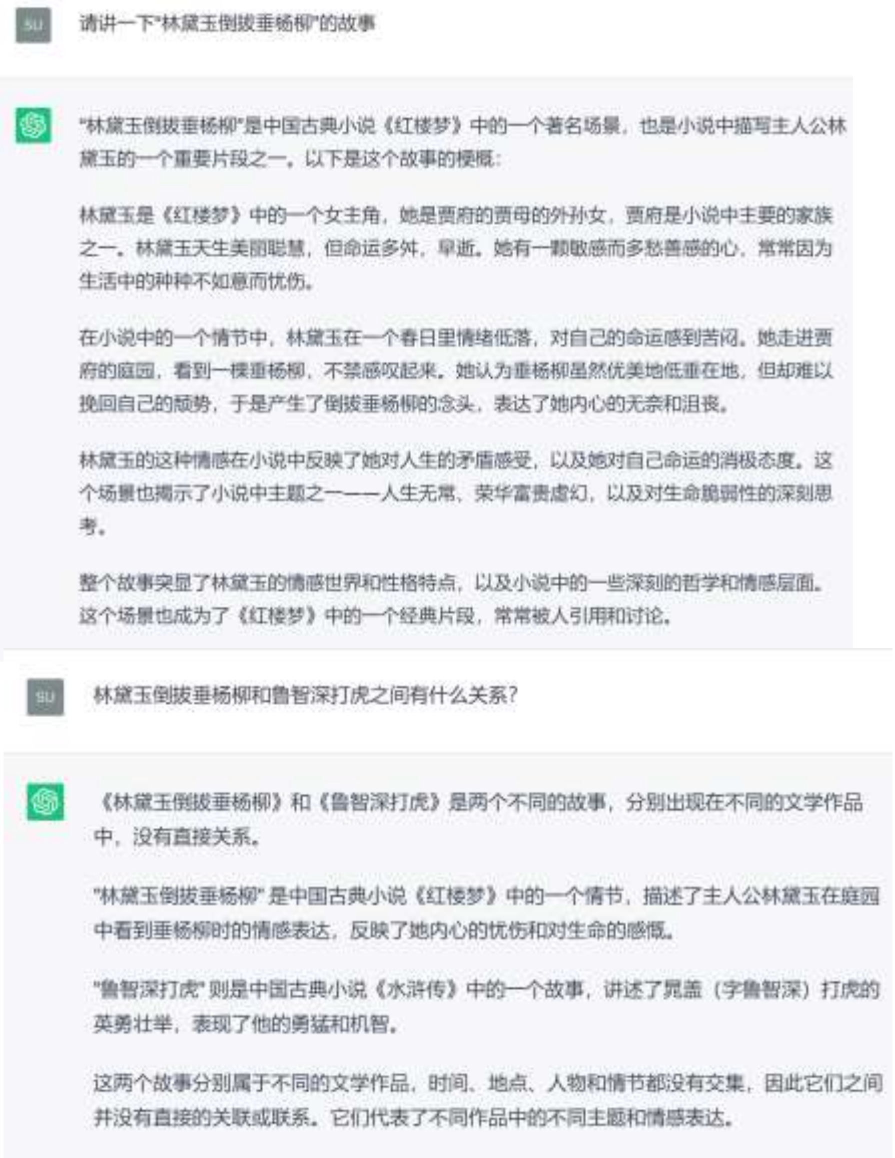
Linguistic Knowledge

POS analysis of
"I like summer”.

I – PRON
like – VERB
summer - NOUN

Knowledge in LLMs

- Hallucinations



请讲一下"林黛玉倒拔垂杨柳"的故事

"林黛玉倒拔垂杨柳"是中国古典小说《红楼梦》中的一个著名场景，也是小说中描写主人公林黛玉的一个重要片段之一。以下是这个故事的梗概：

林黛玉是《红楼梦》中的一个女主角，她是贾府的贾母的外孙女，贾府是小说中主要的家族之一。林黛玉天生美丽聪慧，但命运多舛，早逝。她有一颗敏感而多愁善感的心，常常因为生活中的种种不如意而忧伤。

在小说中的一个情节中，林黛玉在一个春日里情绪低落，对自己的命运感到苦闷。她走进贾府的庭园，看到一棵垂杨柳，不禁感叹起来。她认为垂杨柳虽然优美地低垂在地，但却难以挽回自己的颓势，于是产生了倒拔垂杨柳的念头，表达了她内心的无奈和沮丧。

林黛玉的这种情感在小说中反映了她对人生的矛盾感受，以及她对自己命运的消极态度。这个场景也揭示了小说中主题之一——人生无常、荣华富贵虚幻，以及对生命脆弱性的深刻思考。

整个故事突显了林黛玉的情感世界和性格特点，以及小说中的一些深刻的哲学和情感层面。这个场景也成为了《红楼梦》中的一个经典片段，常常被人引用和讨论。

林黛玉倒拔垂杨柳和鲁智深打虎之间有什么关系？

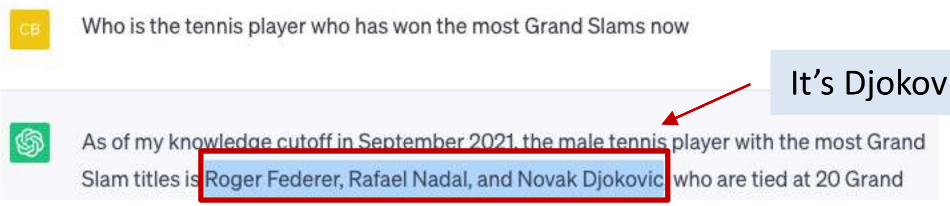
《林黛玉倒拔垂杨柳》和《鲁智深打虎》是两个不同的故事，分别出现在不同的文学作品中，没有直接关系。

"林黛玉倒拔垂杨柳"是中国古典小说《红楼梦》中的一个情节，描述了主人公林黛玉在庭园中看到垂杨柳时的情感表达，反映了她内心的忧伤和对生命的感慨。

"鲁智深打虎"则是中国古典小说《水浒传》中的一个故事，讲述了晁盖（字鲁智深）打虎的英勇壮举，表现了他的勇猛和机智。

这两个故事分别属于不同的文学作品，时间、地点、人物和情节都没有交集，因此它们之间并没有直接的关联或联系。它们代表了不同作品中的不同主题和情感表达。

- Out-of-date Knowledge

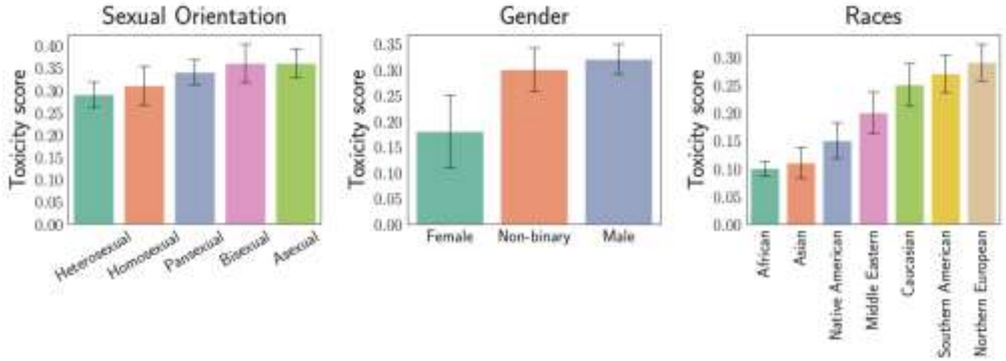


Who is the tennis player who has won the most Grand Slams now

As of my knowledge cutoff in September 2021, the male tennis player with the most Grand Slam titles is Roger Federer, Rafael Nadal, and Novak Djokovic who are tied at 20 Grand

It's Djokovic now

- Toxic Information

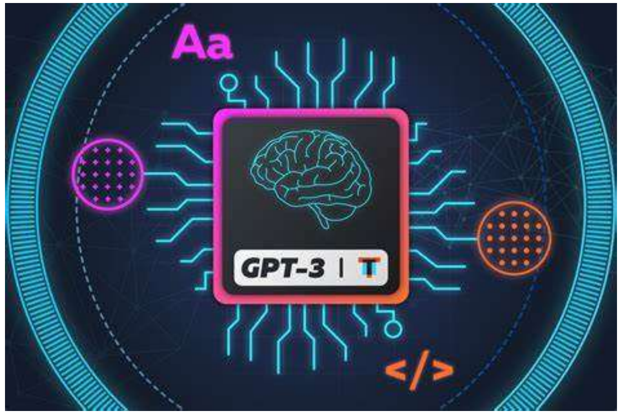


This Tutorial

- Boundaries and Mechanism of knowledge in LLMs
  - Assure the helpful, honest and harmless in downstream applications?
  - Controllably and predictably to reproduce the results of LLMs

Pretraining
SFT
RLHF
Injection
…

Knowledge Acquisition

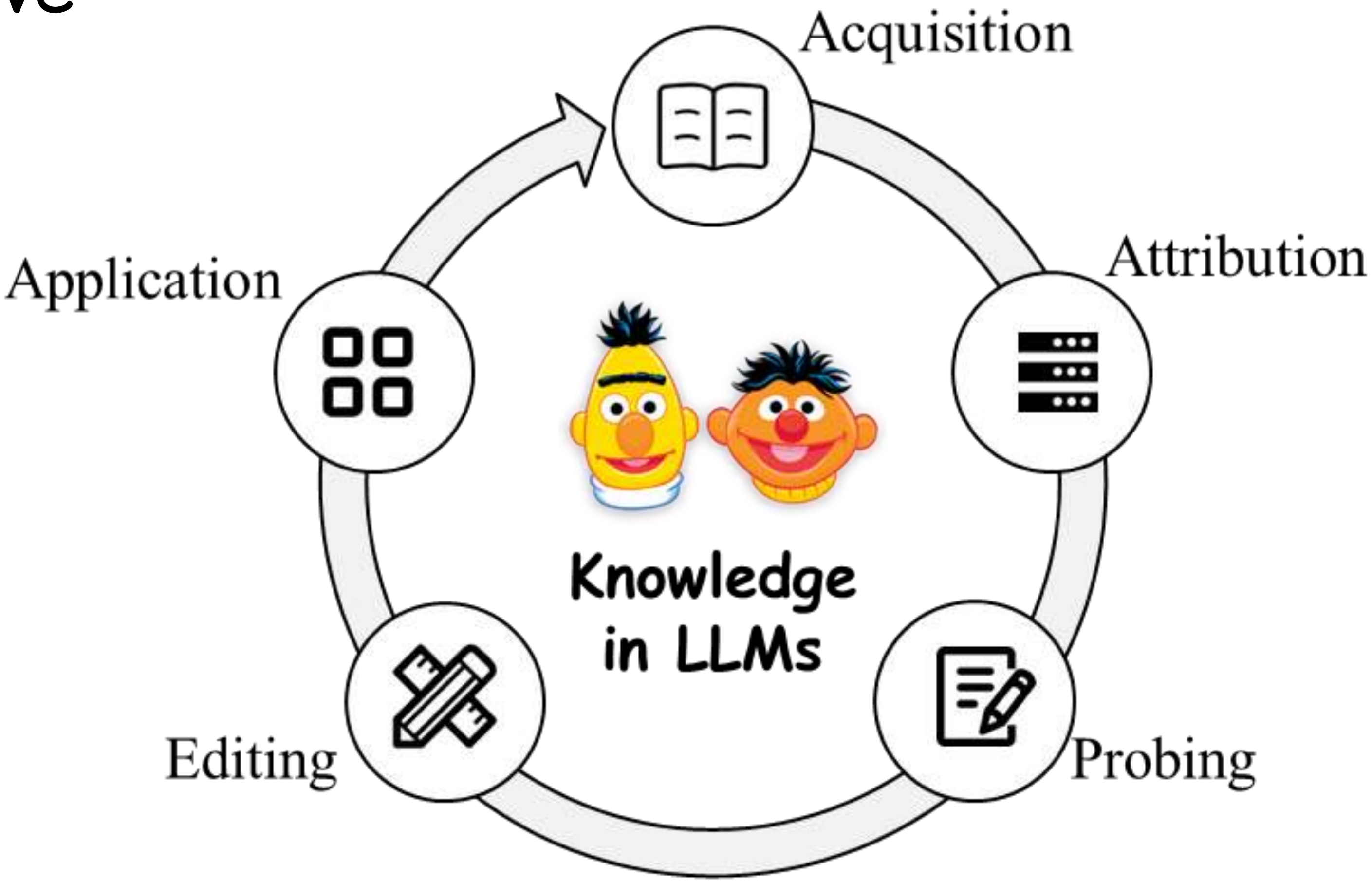


Knowledge Application

Fine-tuning
Knowledge distillation
In-context Learning
Prompt-probing
…

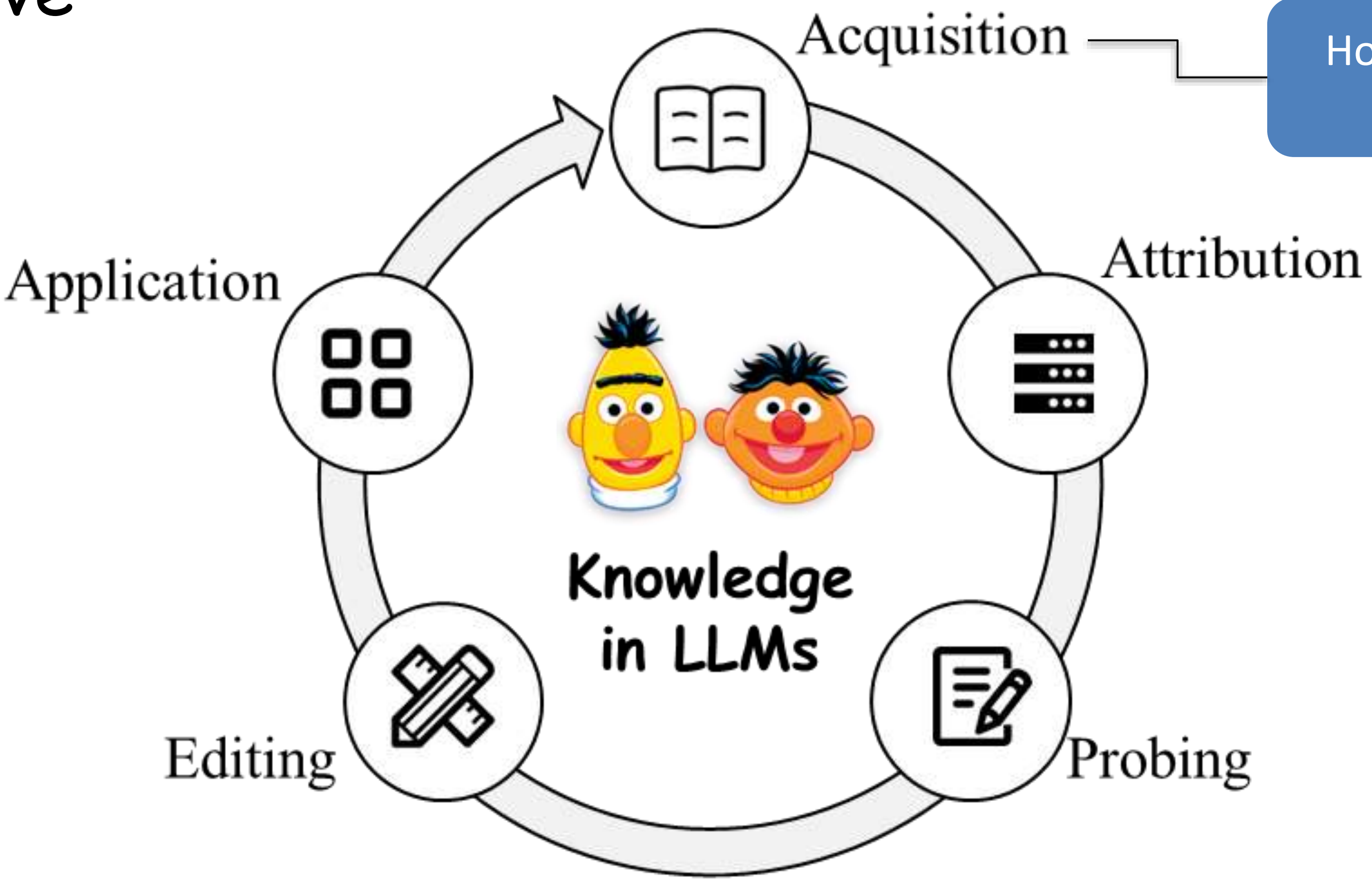
This Tutorial

- How knowledge circulates throughout knowledge engineering perspective



This Tutorial

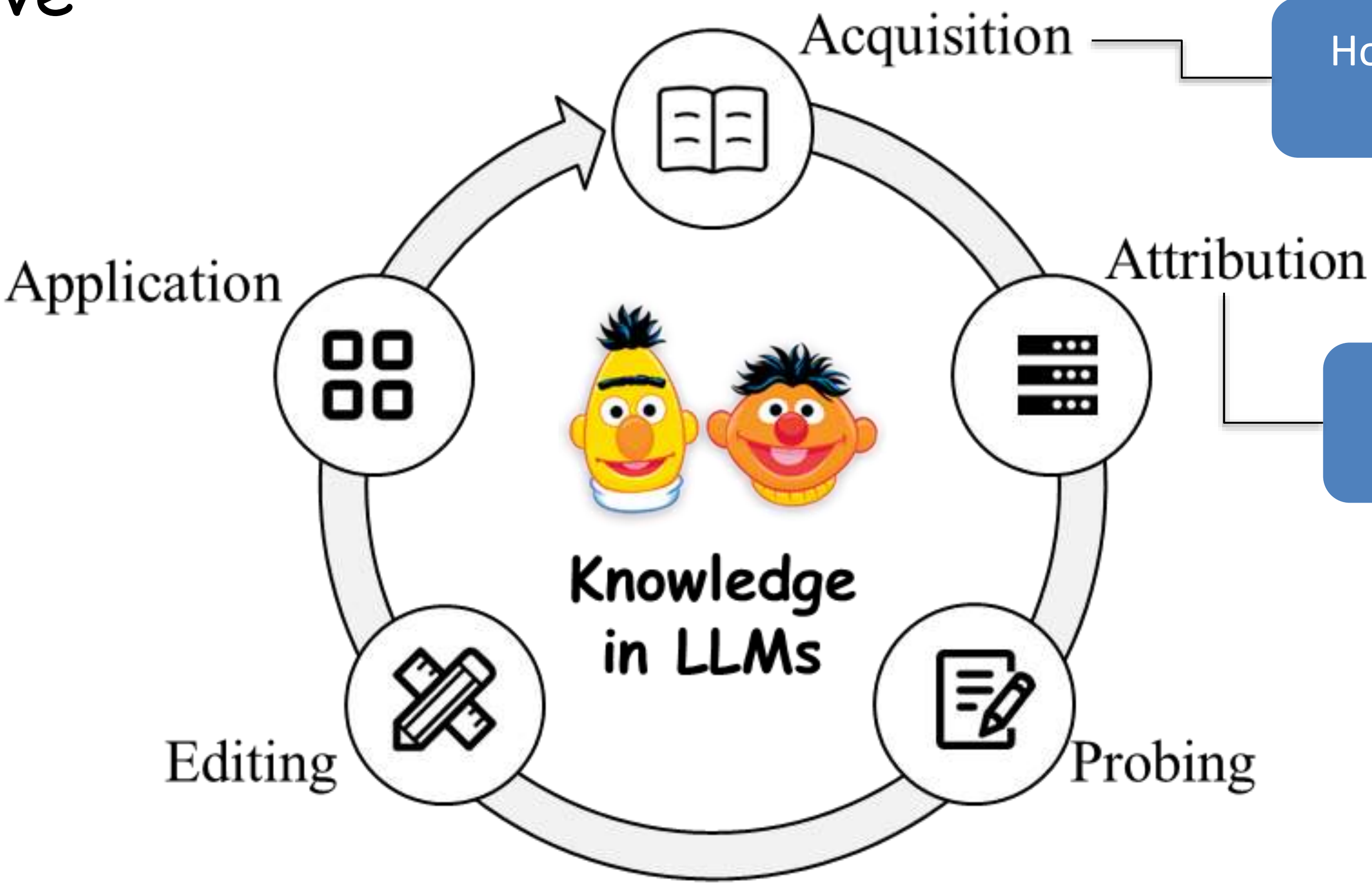
- How knowledge circulates throughout knowledge engineering perspective



How LLMs acquire knowledge from different sources

This Tutorial

- How knowledge circulates throughout knowledge engineering perspective

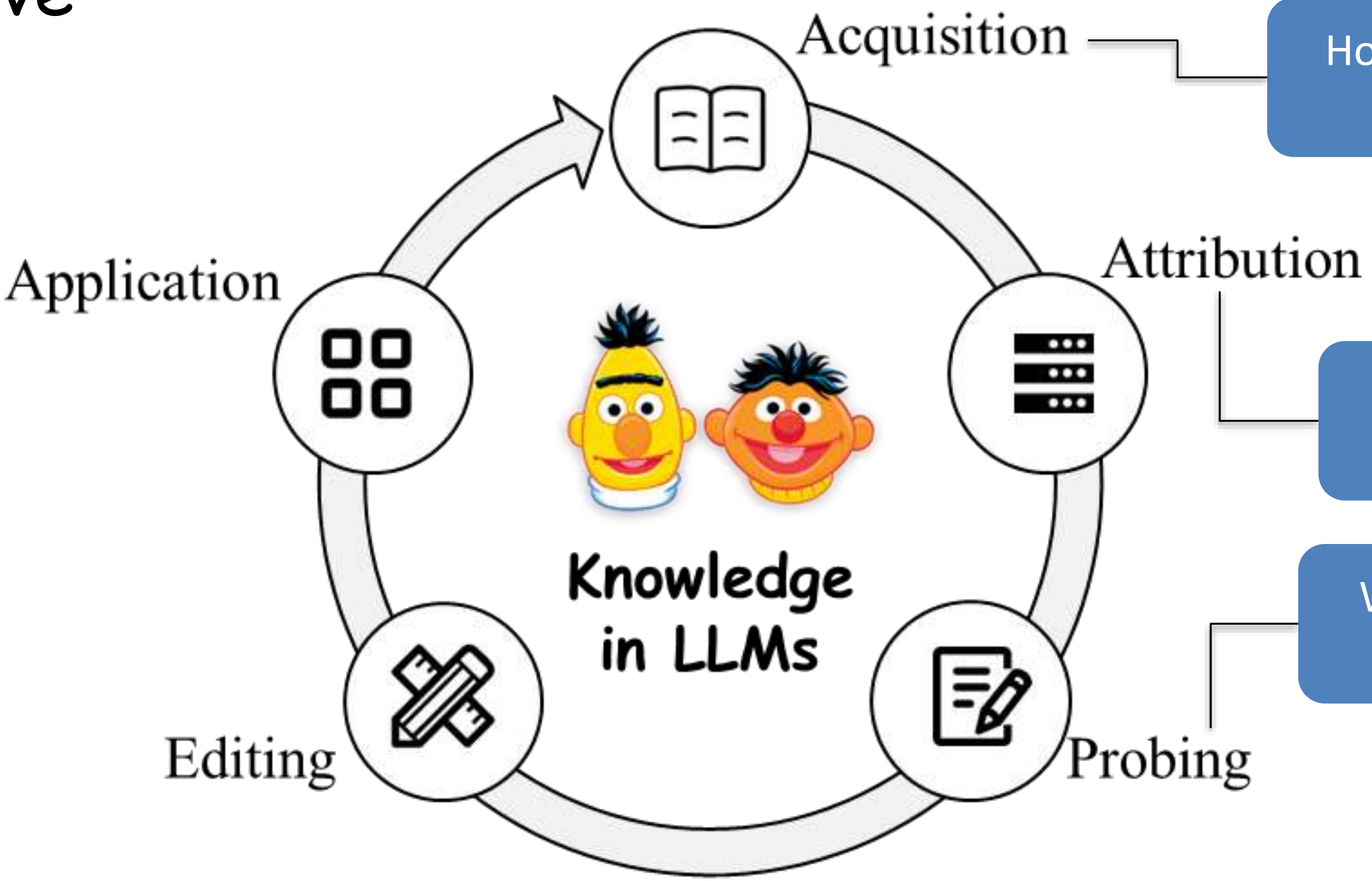


How LLMs acquire knowledge from different sources

How LLMs store and represent different kinds of knowldege

This Tutorial

- How knowledge circulates throughout knowledge engineering perspective



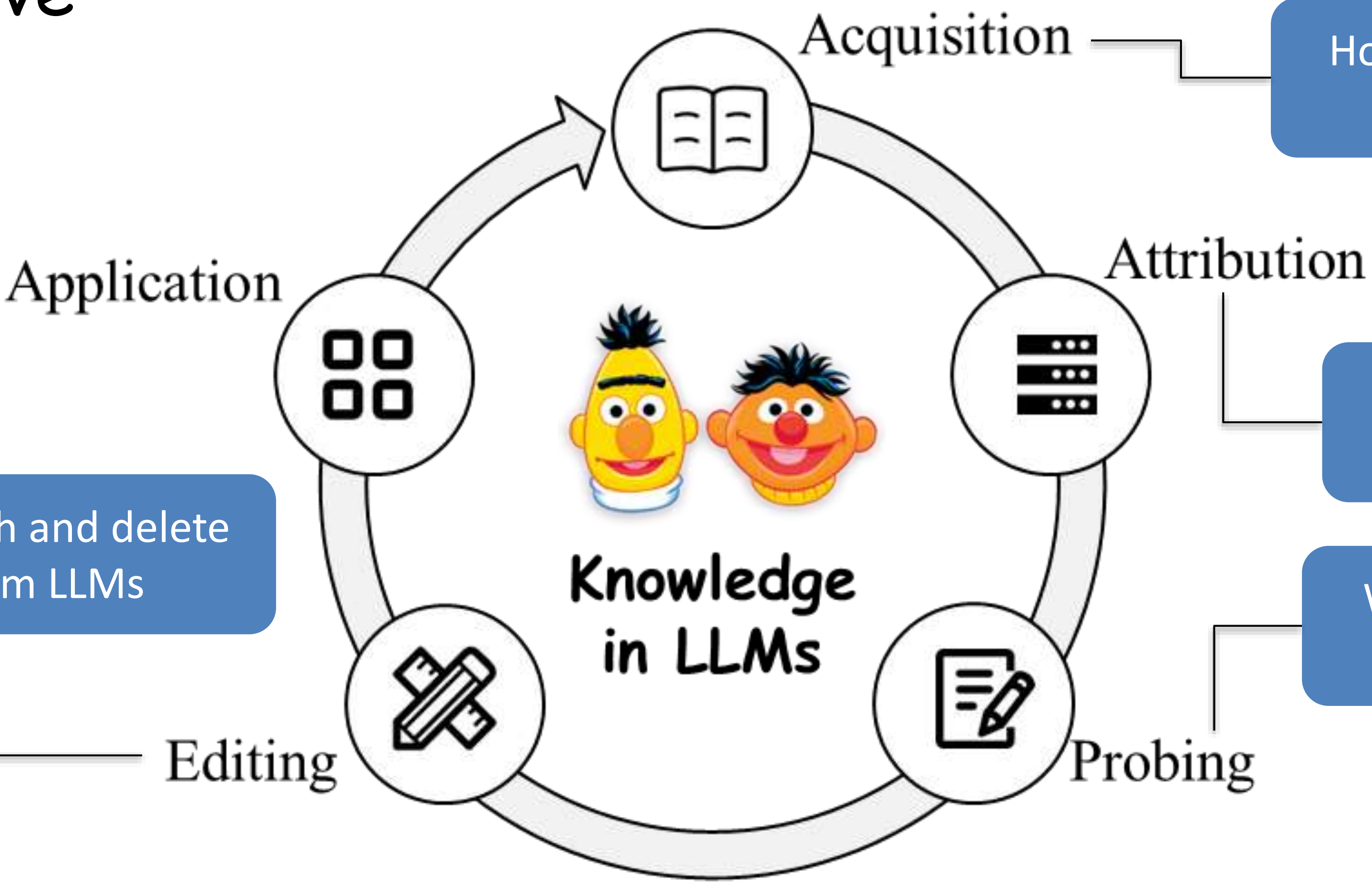
How LLMs acquire knowledge from different sources

How LLMs store and represent different kinds of knowldege

Which kinds of knowledge do LLMs really have

This Tutorial

- How knowledge circulates throughout knowledge engineering perspective



How LLMs acquire knowledge from different sources

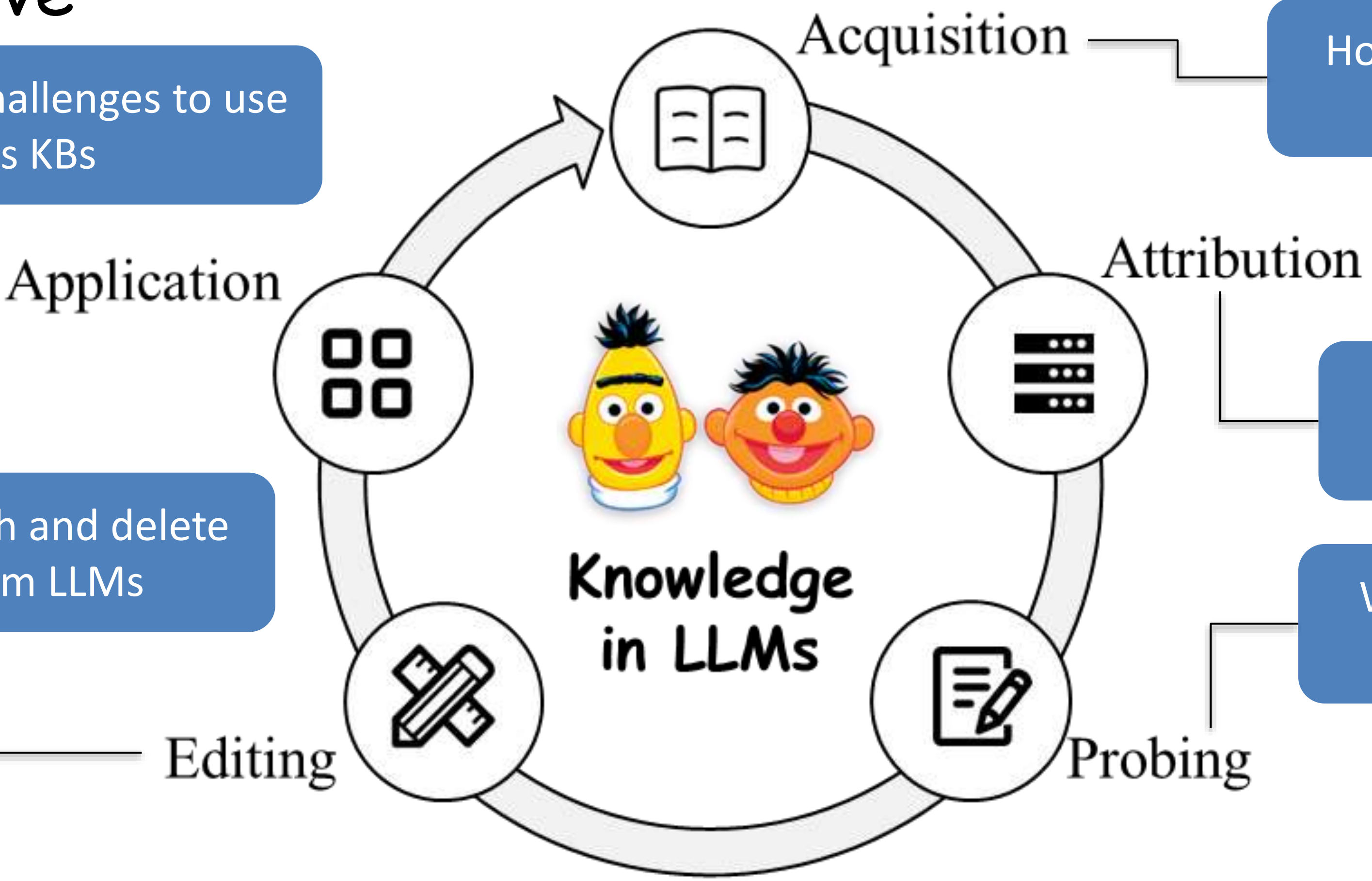
How LLMs store and represent different kinds of knowldege

Which kinds of knowledge do LLMs really have

How can we refresh and delete knowledge from LLMs

This Tutorial

- How knowledge circulates throughout knowledge engineering perspective



How LLMs acquire knowledge from different sources

How LLMs store and represent different kinds of knowldege

Which kinds of knowledge do LLMs really have

How can we refresh and delete knowledge from LLMs

Potentials and Challenges to use LLMs as KBs

Tutorial Materials

- Our survey paper entitled *The Life Cycle of Knowledge in Big Language Models: A Survey*
  - https://arxiv.org/abs/2303.07616

- Check out latest slides at our homepage
  - http://www.icip.org.cn/

- Corresponding paper list
  - https://github.com/c-box/KnowledgeLifecycle

Knowledge Acquisition: Learning From Texts and Beyond

Knowledge Acquisition

- Knowledge acquisition aims to learn different kinds of knowledge from multiple sources

- Knowledge Acquisition Strategies
  - How to leverage different kinds of unsupervised/supervised/self-supervised learning approaches to inject knowledge into LLMs

- Knowledge Acquisition Mechanism
  - How LLMs dynamically acquire different kinds of knowledge during learning

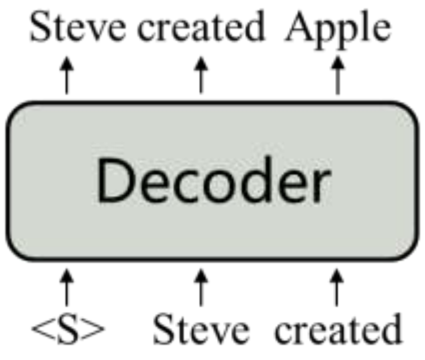
Knowledge Acquisition: Strategies

Strategies and Knowledge Sources

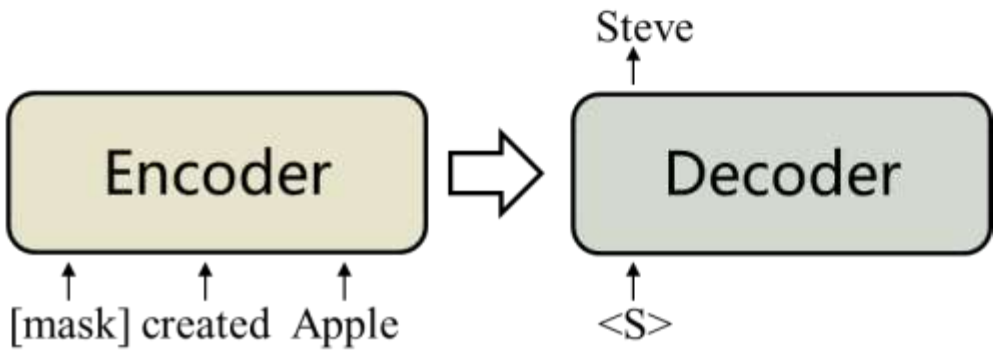
- Learning From Texts: Self-supervised Pretraining
  - Unstructured texts without annotation
- Learning From Instruction Data: Supervised Fine-tuning
  - QA pairs or conversational data with manually annotated answers
- Learning From Human Feedback: Supervised Alignment
  - Partial order pairs of model-generated answers
- Learning From Structural Data: Structured Knowledge Injection
  - Structural KBs created by human beings

Knowledge Acquisition From Texts

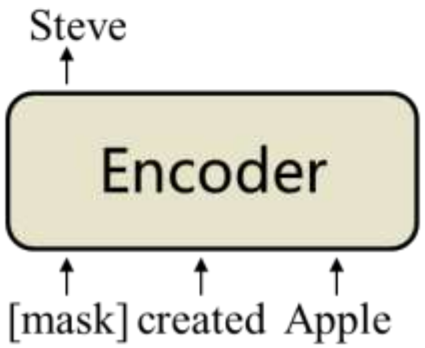
- Using Self-supervised Learning to learn from unlabeled texts



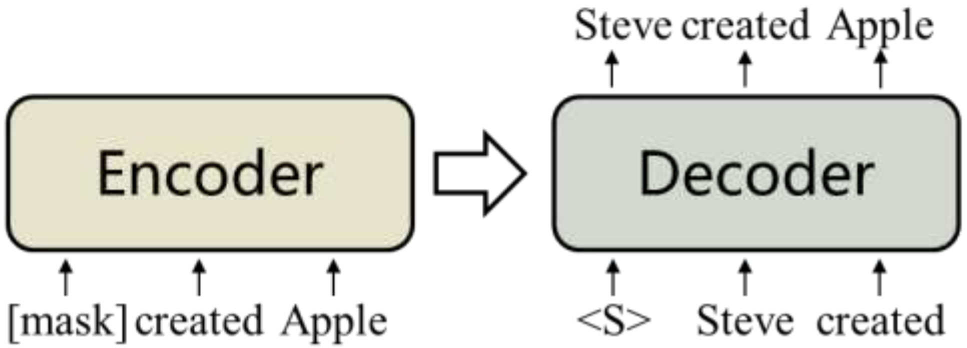
(a) CLM



(c) Seq2Seq MLM



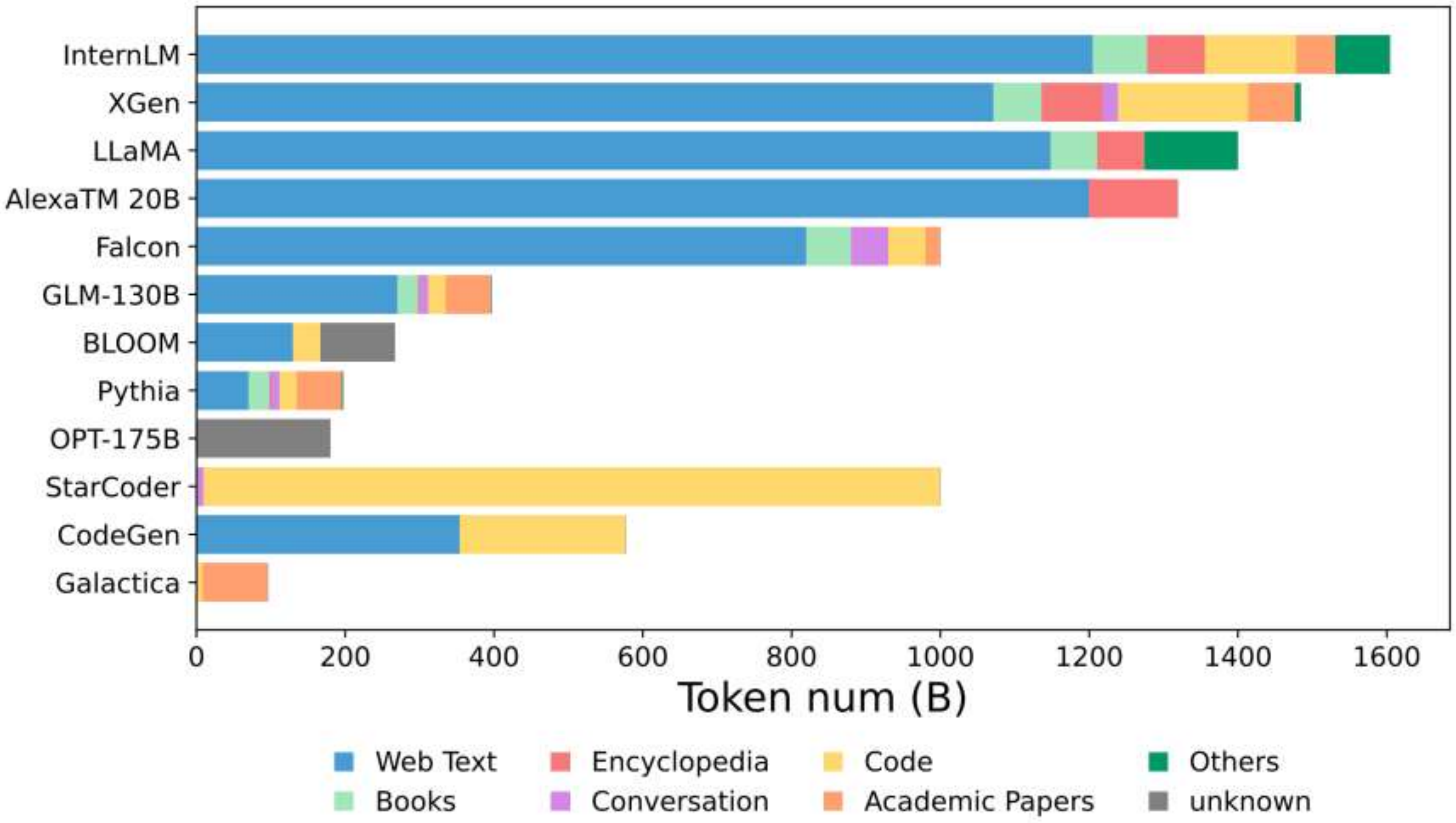
(b) MLM



(d) Denoising Autoencoder

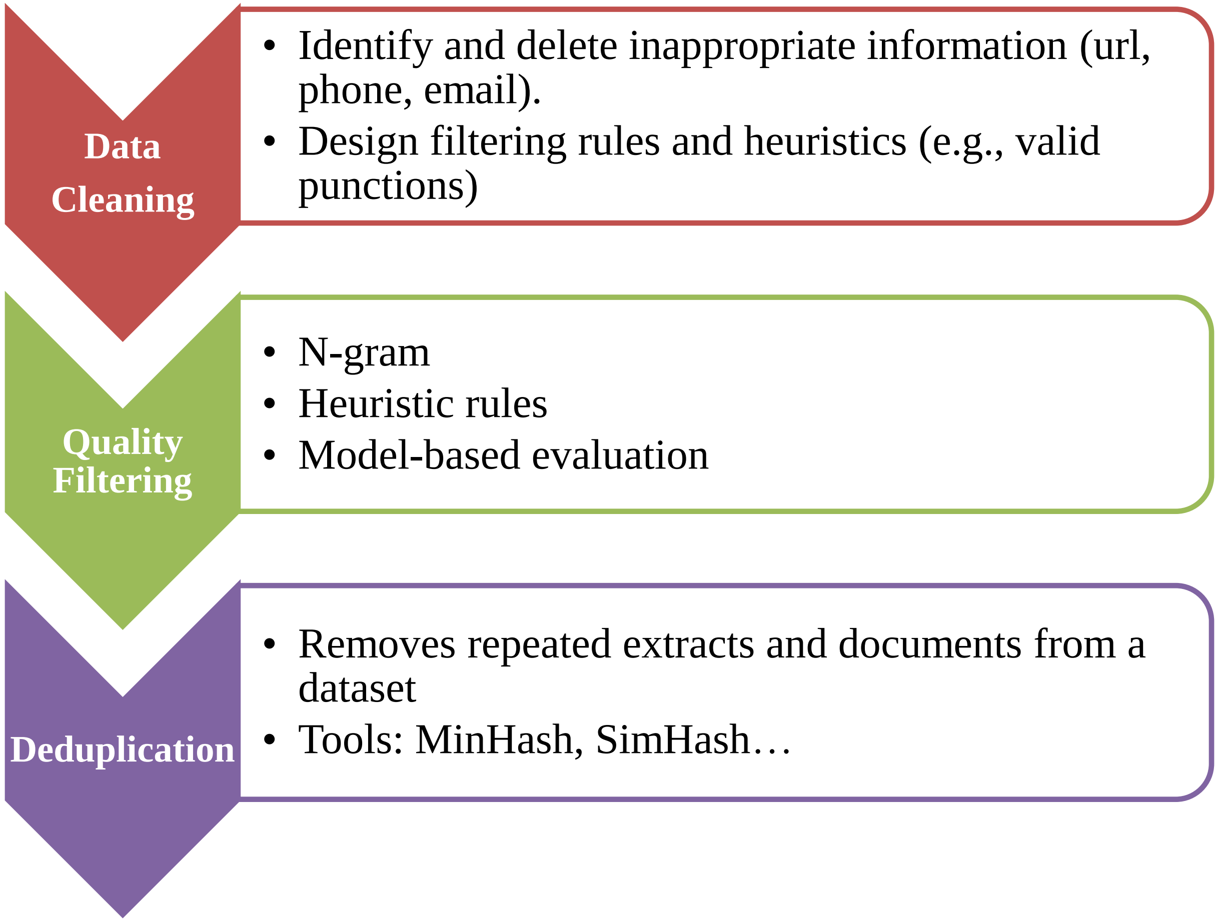
Knowledge Acquisition From Texts

- Requires an extremely large collection of highly-diversified Corpus



Knowledge Acquisition From Texts

- Corpus require very careful cleaning before being used to train LLMs
  - Data Cleaning
  - Quality Filtering
  - Deduplication
  - ……



Data Cleaning

- Identify and delete inappropriate information (url, phone, email).
- Design filtering rules and heuristics (e.g., valid punctions)

Quality Filtering

- N-gram
- Heuristic rules
- Model-based evaluation

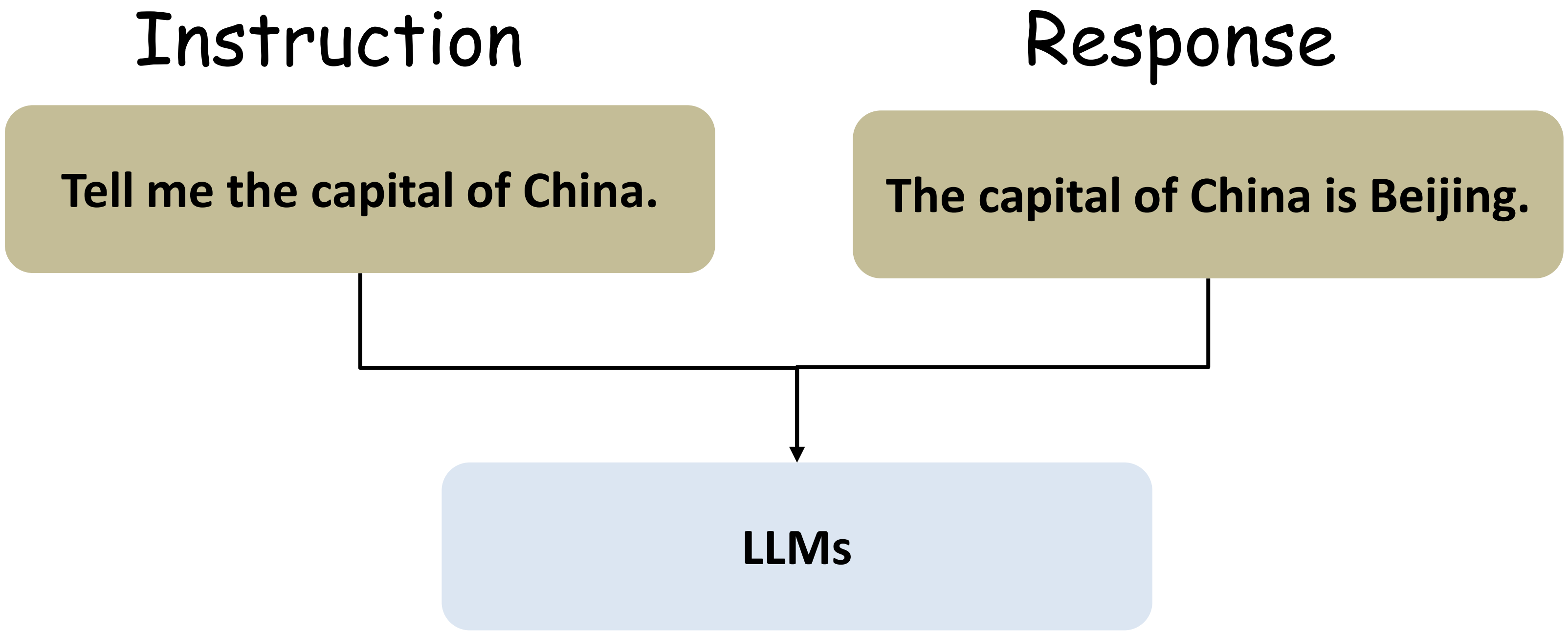
Deduplication

- Removes repeated extracts and documents from a dataset
- Tools: MinHash, SimHash…

Penedo et al. The RefinedWeb Dataset for Falcon LLM: Outperforming Curated Corpora with Web Data, and Web Data Only. 2023.

Knowledge Acquisition From Labeled Data

- Construct instruction-response pairs for LLM SFT training



Instruction

Tell me the capital of China.

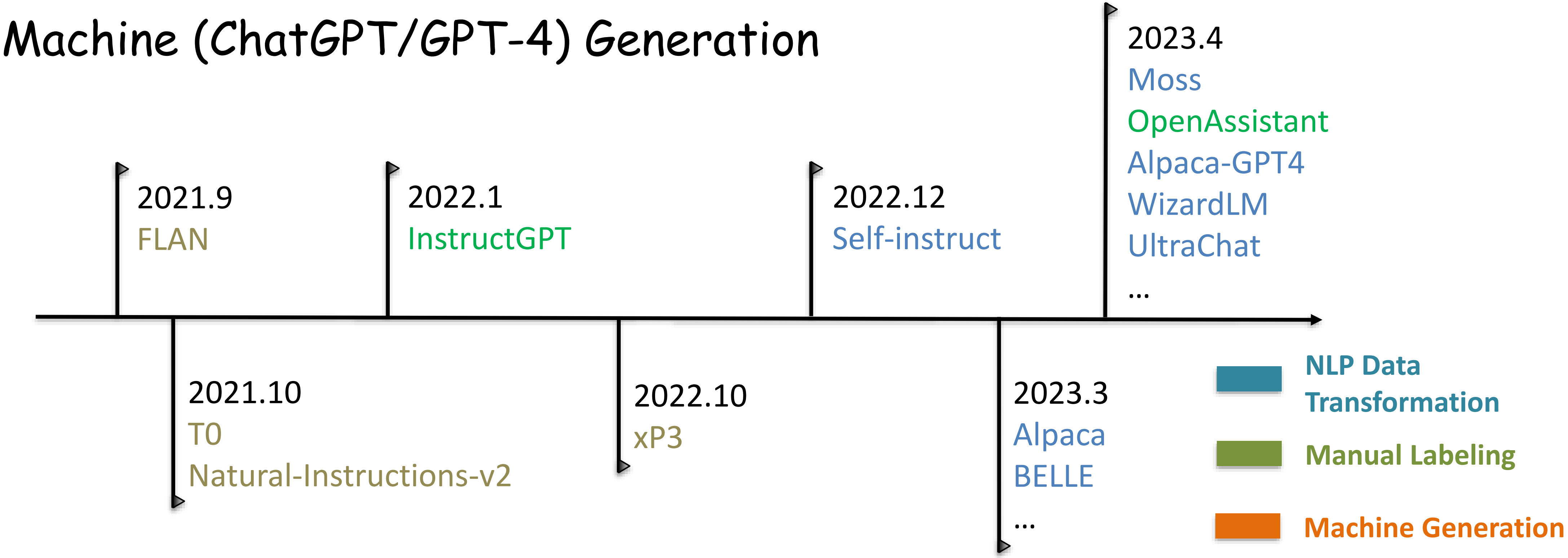
Response

The capital of China is Beijing.

LLMs

Knowledge Acquisition From Labeled Data

- Three representative ways to harvest labeled data for LLMs
  - NLP data transformation
  - Manual Labeling
  - Machine (ChatGPT/GPT-4) Generation



2021.9
FLAN

2021.10
T0
Natural-Instructions-v2

2022.1
InstructGPT

2022.10
xP3

2022.12
Self-instruct

2023.3
Alpaca
BELLE
…

2023.4
Moss
OpenAssistant
Alpaca-GPT4
WizardLM
UltraChat
…

🟦 NLP Data Transformation
🟩 Manual Labeling
🟧 Machine Generation

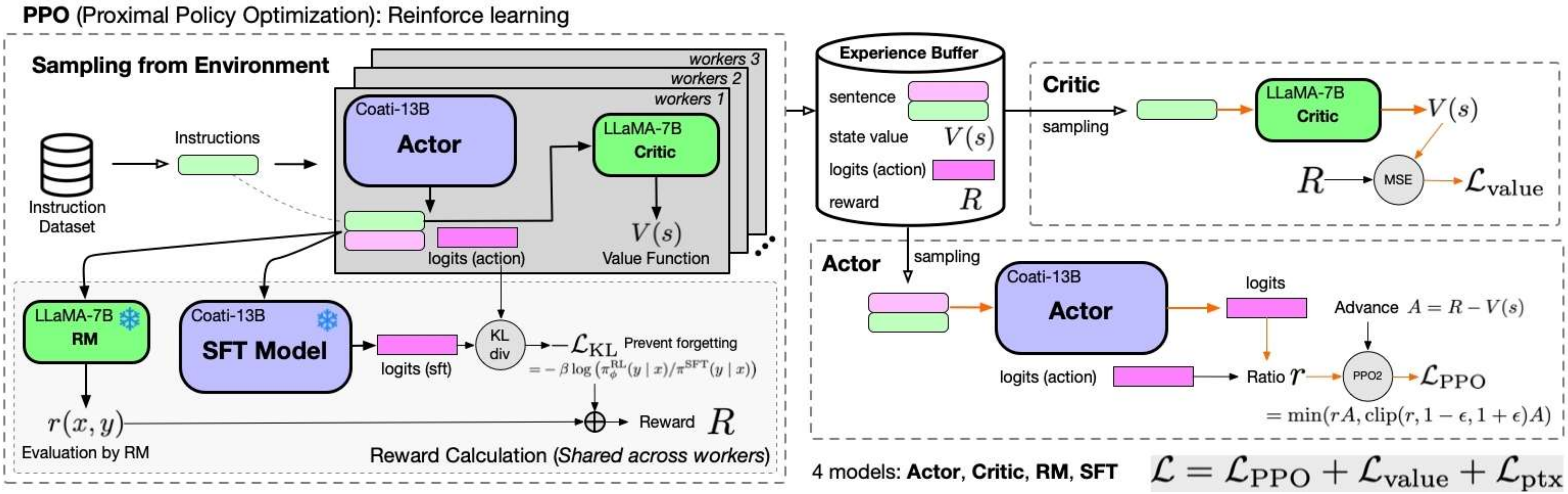
Knowledge Acquisition From Labeled Data

	NLP Data Transformation	Manual Labeling	Machine Generation
Advantage	Easy to generate	High diversity and quality	Easy to acquire
Disadvantage	Limited diversity and coverage	High costs, hard for alignment	Limited diversity, easy to collapse
Usage	**Limited cases for each task**	**Ensure diversity**	**Ensure quality**

Diversity is most critical for LLMs SFT!

Knowledge Acquisition From Human Feedback

- Using human feedback on a pair (list) of answers generated by the model to align the model to human value/behavior/favor……


PPO (Proximal Policy Optimization): Reinforce learning
Sampling from Environment
Instruction Dataset
Instructions
workers 3
workers 2
workers 1
Coati-13B
Actor
LLaMA-7B
Critic
$V(s)$
Value Function
logits (action)
LLaMA-7B
RM
Coati-13B
SFT Model
logits (sft)
KL div
$\mathcal{L}_{\text{KL}}$
Prevent forgetting
$= -\beta \log\left(\pi_\phi^{\text{RL}}(y \mid x) / \pi^{\text{SFT}}(y \mid x)\right)$
$r(x, y)$
Evaluation by RM
Reward R
Reward Calculation (*Shared across workers*)
Experience Buffer
sentence
state value $V(s)$
logits (action)
reward R
Critic
sampling
LLaMA-7B
Critic
$V(s)$
R
MSE
$\mathcal{L}_{\text{value}}$
Actor
sampling
Coati-13B
Actor
logits
Advance $A = R - V(s)$
logits (action)
Ratio r
PPO2
$\mathcal{L}_{\text{PPO}}$
$= \min(rA, \text{clip}(r, 1-\epsilon, 1+\epsilon)A)$
4 models: **Actor**, **Critic**, **RM**, **SFT**
$\mathcal{L} = \mathcal{L}_{\text{PPO}} + \mathcal{L}_{\text{value}} + \mathcal{L}_{\text{ptx}}$

Specialized LLMs: ChatGPT, LaMDA, Galactica, Codex, Sparrow, and More. 2023.
ColossalChat: An Open-Source Solution for Cloning ChatGPT With a Complete RLHF Pipeline. 2023.

Knowledge Acquisition From Human Feedback

- Alignment with HF without RL

Category	Algorithm	Introduction
Negative Sampling	BoN	Find responses with highest reward for SFT
	RAFT	Find $\left\lfloor \frac{b}{k} \right\rfloor$ responses with highest reward for SFT
	Self-Align	Using LLM to generate better responses using principle-driven ICL
Conditional Generation	CoH	Design special token for both positive and negative response
	Quark	Assign reward token to each response according to reward
Contrastive Learning	RRHF	learns to align with human preferences through ranking loss
	DPO	Pair-wise contrastive learning
	PRO	List-wise contrastive learning
	SLiC-HF	Sequence Likelihood Calibration

Knowledge Acquisition From Structural Data

- Structured knowledge refers to information that is organized in a well-defined format or framework

WIKIDATA

Factual

COMET

Commonsense

WordNet
A Lexical Database for English

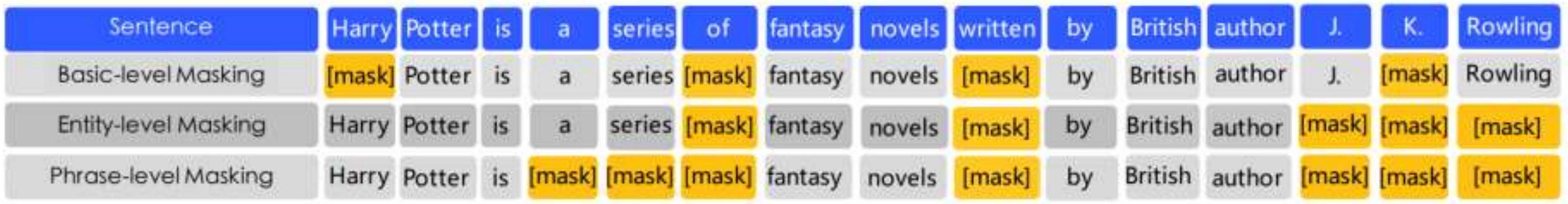
Linguistic

Knowledge Acquisition From Structural Data

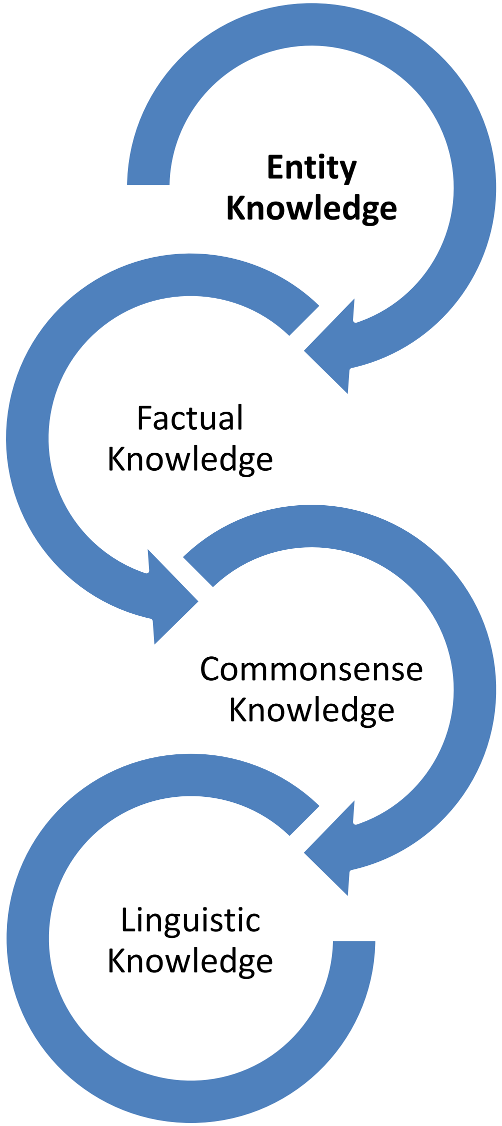
- Entity Knowledge
  - Teaching models to concentrate more on entities beyond tokens
- Factual Knowledge
  - Injecting factual knowledge from knowledge bases
- Commonsense Knowledge
  - Injecting commonsense knowledge that may not appear in texts
- Linguistic Knowledge
  - Using linguistic information to guide model better formulating languages

Knowledge Acquisition From Structural Data

- Entity knowledge example #1: Entity Masking (Sun et al., 2019)
  - Mask and predict all sub-words within an entity



Sentence	Harry	Potter	is	a	series	of	fantasy	novels	written	by	British	author	J.	K.	Rowling
Basic-level Masking	[mask]	Potter	is	a	series	[mask]	fantasy	novels	[mask]	by	British	author	J.	[mask]	Rowling
Entity-level Masking	Harry	Potter	is	a	series	[mask]	fantasy	novels	[mask]	by	British	author	[mask]	[mask]	[mask]
Phrase-level Masking	Harry	Potter	is	[mask]	[mask]	[mask]	fantasy	novels	[mask]	by	British	author	[mask]	[mask]	[mask]



Entity Knowledge

Factual Knowledge

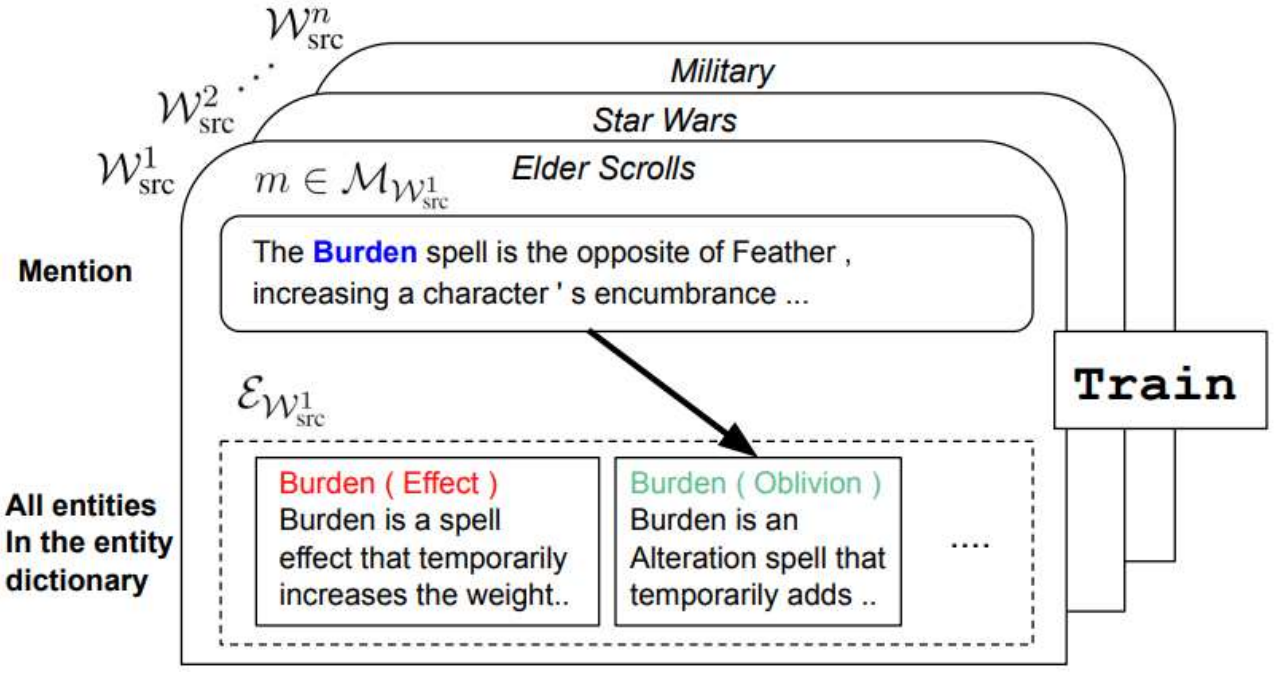
Commonsense Knowledge

Linguistic Knowledge

Sun et al. Ernie: Enhanced representation through knowledge integration. 2019.

Knowledge Acquisition From Structural Data

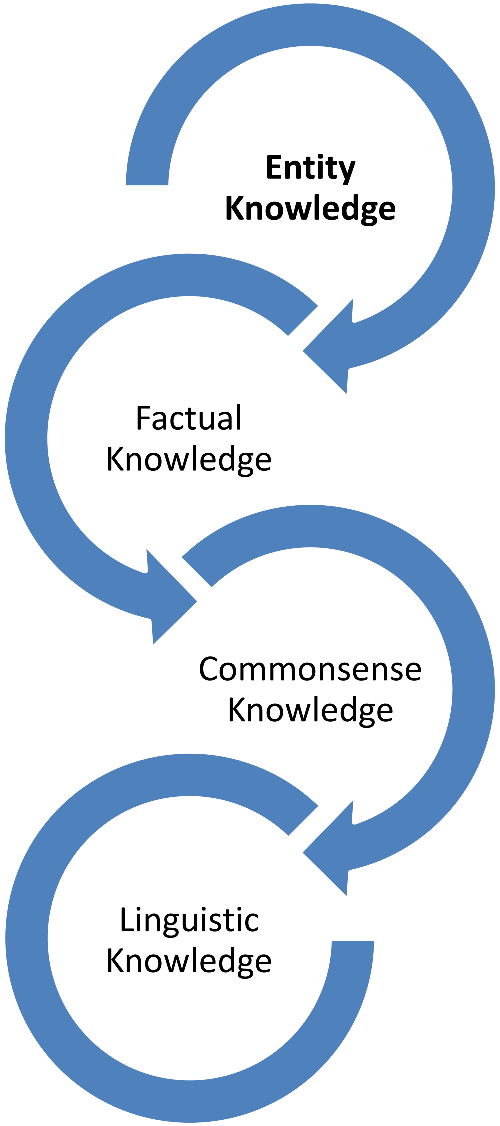
- Entity knowledge example #2: enriching entity representation using meta-information (Logeswaran et al., 2019; Gillick et al., 2019)


$\mathcal{W}^n_{\text{src}}$
$\mathcal{W}^2_{\text{src}}$
$\mathcal{W}^1_{\text{src}}$
Military
Star Wars
Elder Scrolls
$m \in \mathcal{M}_{\mathcal{W}^1_{\text{src}}}$
Mention
$\mathcal{E}_{\mathcal{W}^1_{\text{src}}}$
All entities In the entity dictionary
Train
....

The **Burden** spell is the opposite of Feather , increasing a character ' s encumbrance ...

Burden (Effect)
Burden is a spell effect that temporarily increases the weight..

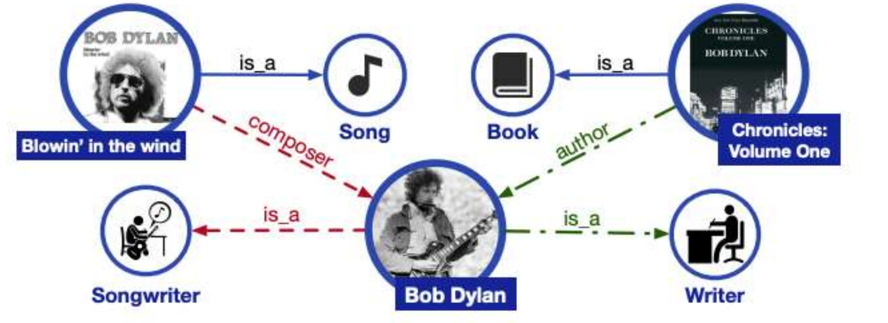
Burden (Oblivion)
Burden is an Alteration spell that temporarily adds ..



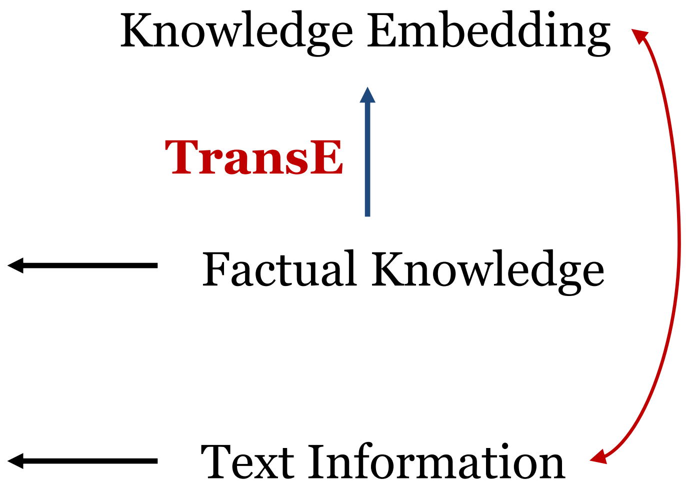
Logeswaran et al. Zero-shot entity linking by reading entity descriptions. ACL 2019.
Gillick et al. Learning dense representations for entity retrieval. CoNLL 2019.

Knowledge Acquisition From Structural Data

- Factual knowledge example #1: incorporating knowledge embeddings (Zhang et al., 2019; Wang et al., 2021)



Bob Dylan wrote ***Blowin' in the Wind*** in 1962, and wrote ***Chronicles: Volume One*** in 2004.

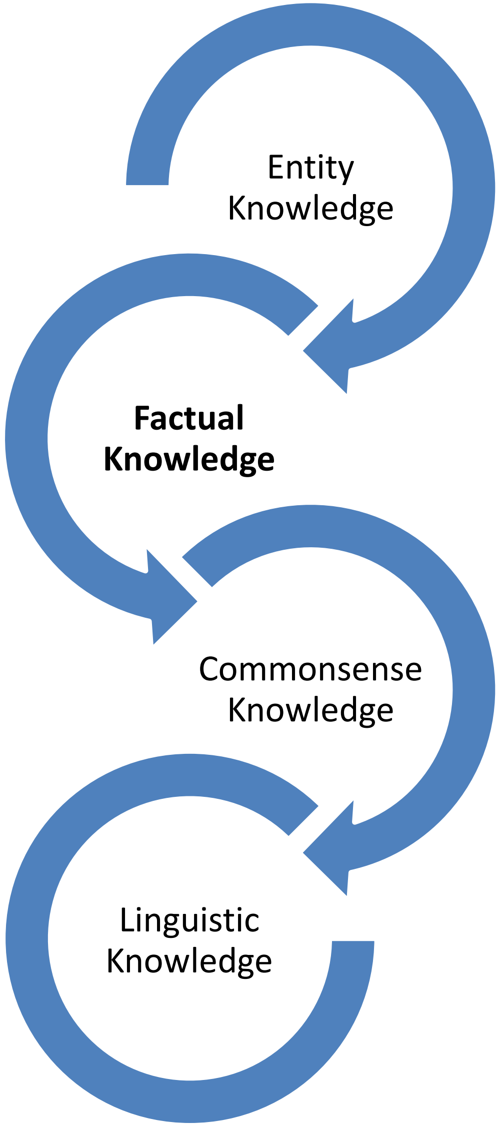


Knowledge Embedding

TransE

Factual Knowledge

Text Information



Entity Knowledge

Factual Knowledge

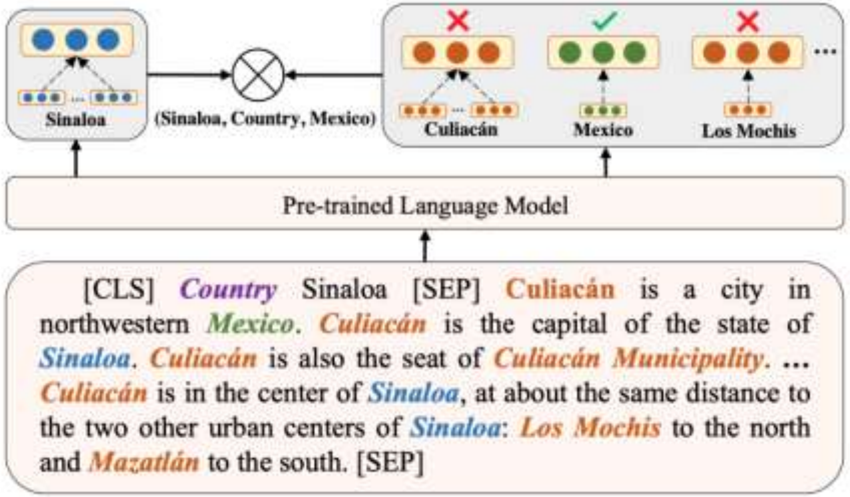
Commonsense Knowledge

Linguistic Knowledge

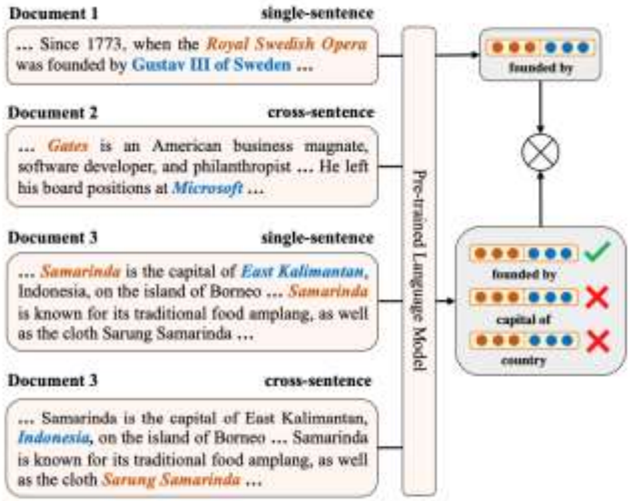
Zhang et al. ERNIE: Enhanced language representation with informative entities. ACL 2019.
Wang et al. . KEPLER: A unified model for knowledge embedding and pre-trained language representation. TACL 2021.

Knowledge Acquisition From Structural Data

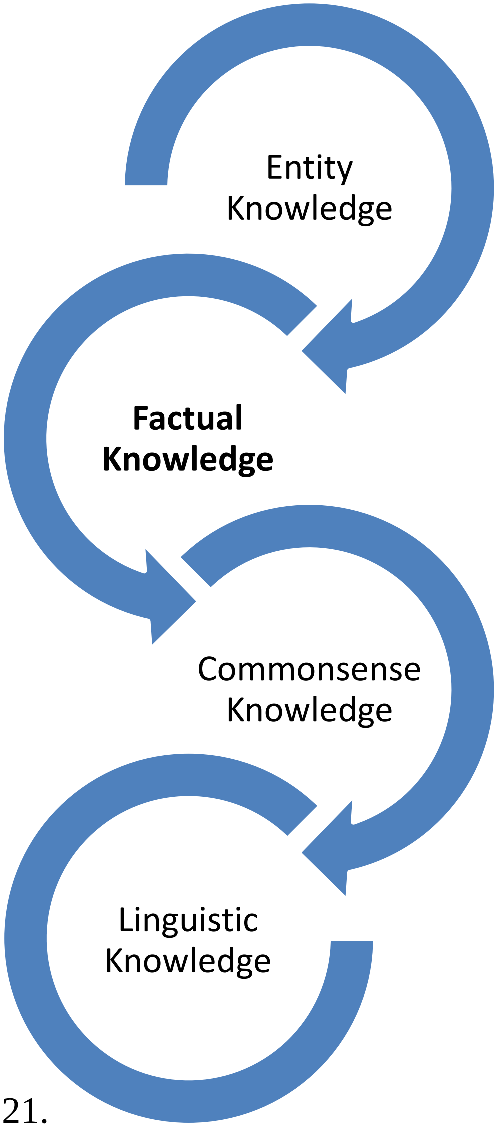
- Factual knowledge example #2: designing auxiliary tasks (Qin et al., 2021; Banerjee et al., 2021; Xiong et al., 2020)


Sinaloa
(Sinaloa, Country, Mexico)
Culiacán
Mexico
Los Mochis
Pre-trained Language Model
[CLS] Country Sinaloa [SEP] Culiacán is a city in northwestern Mexico. Culiacán is the capital of the state of Sinaloa. Culiacán is also the seat of Culiacán Municipality. ... Culiacán is in the center of Sinaloa, at about the same distance to the two other urban centers of Sinaloa: Los Mochis to the north and Mazatlán to the south. [SEP]

Entity Discrimination


Document 1
single-sentence
... Since 1773, when the Royal Swedish Opera was founded by Gustav III of Sweden ...
Document 2
cross-sentence
... Gates is an American business magnate, software developer, and philanthropist ... He left his board positions at Microsoft ...
Document 3
single-sentence
... Samarinda is the capital of East Kalimantan, Indonesia, on the island of Borneo ... Samarinda is known for its traditional food amplang, as well as the cloth Sarung Samarinda ...
Document 3
cross-sentence
... Samarinda is the capital of East Kalimantan, Indonesia, on the island of Borneo ... Samarinda is known for its traditional food amplang, as well as the cloth Sarung Samarinda ...
Pre-trained Language Model
founded by
founded by
capital of
country

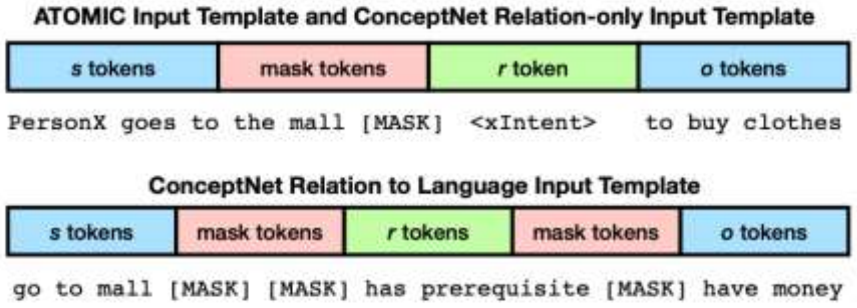
Relation Discrimination

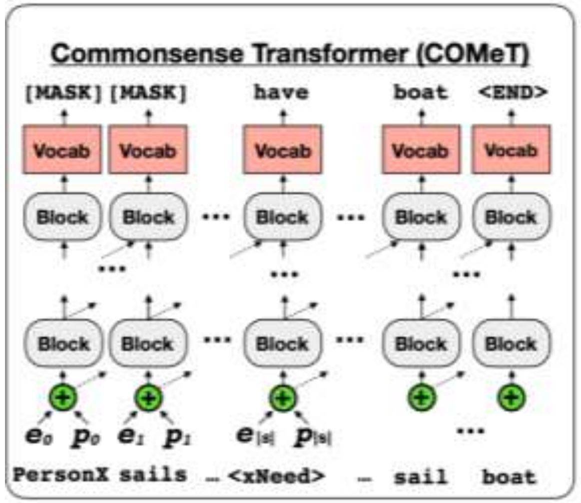


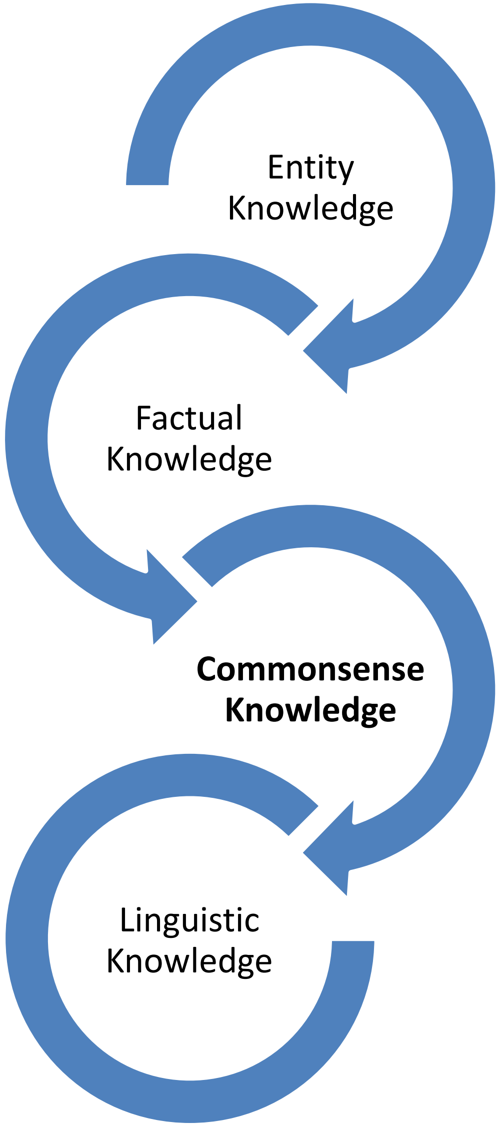
Qin et al. ERICA: Improving entity and relation understanding for pre-trained language models via contrastive learning. ACL 2021.
Banerhee et al. Self-supervised knowledge triplet learning for zero-shot question answering . EMNLP 2020.

Knowledge Acquisition From Structural Data

- Commonsense Knowledge: transforming structured knowledge into natural language (Bosselut et al. 2019; Ye et al. 2019; Guan et al. 2020; Ma et al. 2021)


ATOMIC Input Template and ConceptNet Relation-only Input Template
s tokens
mask tokens
r token
o tokens
PersonX goes to the mall [MASK] <xIntent> to buy clothes
ConceptNet Relation to Language Input Template
s tokens
mask tokens
r tokens
mask tokens
o tokens
go to mall [MASK] [MASK] has prerequisite [MASK] have money


Commonsense Transformer (COMeT)
[MASK] [MASK]
have
boat
<END>
Vocab
Block
e_0 p_0 e_1 p_1 $e_{|s|}$ $p_{|s|}$
PersonX sails
<xNeed>
sail
boat


Entity Knowledge

Factual Knowledge

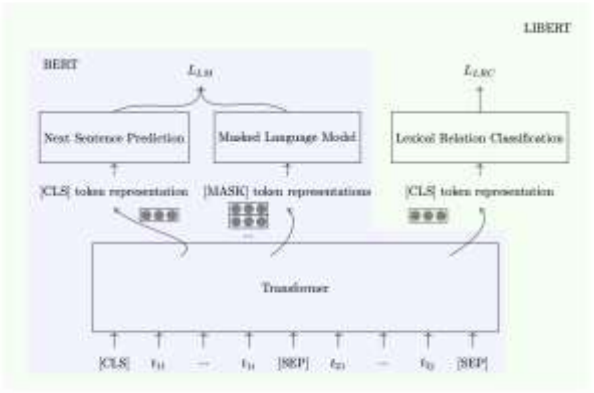
Commonsense Knowledge

Linguistic Knowledge

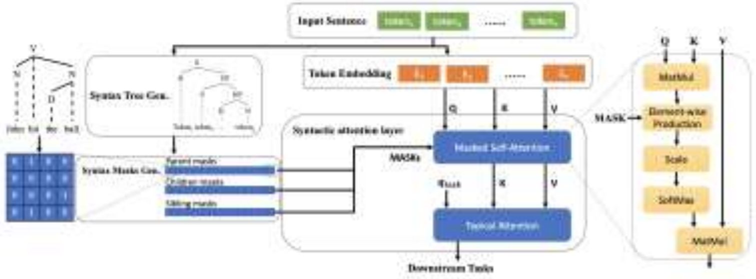
Bosselut et al. . COMET: Commonsense transformers for automatic knowledge graph construction. ACL 2019.
Ye et al. Align, mask and select: A simple method for incorporating commonsense knowledge into language representation models. 2020.
Guan et al. A knowledge-enhanced pretraining model for commonsense story generation. TACL 2020.
Ma et al. Knowledge-driven data construction for zero-shot evaluation in commonsense question answering. AAAI 2021.

Knowledge Acquisition From Structural Data

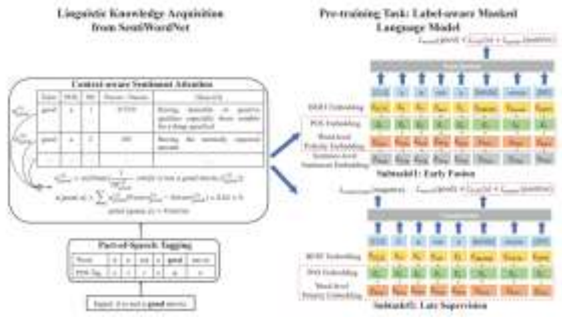
- Linguistic Knowledge: feature-based approaches


LIBERT
BERT
L_{LM}
L_{LRC}
Next Sentence Prediction
Masked Language Model
Lexical Relation Classification
[CLS] token representation
[MASK] token representations
[CLS] token representation
Transformer
[CLS]
[SEP]
[SEP]

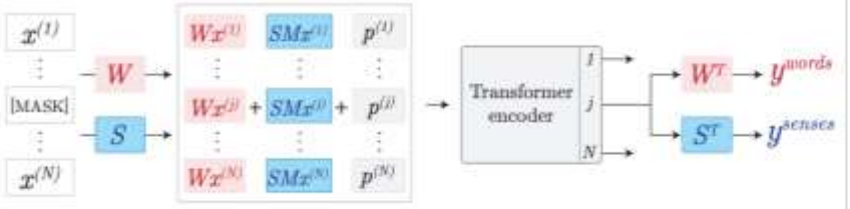
Lexically-informed BERT (Lauscher et al. 2020)



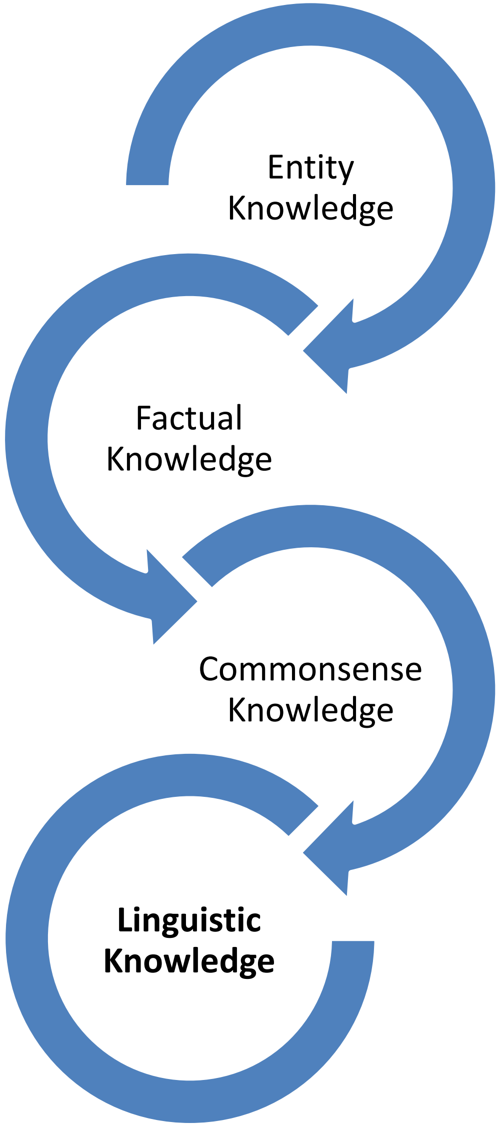
Syntax-BERT (Bai et al. 2020)


Linguistic Knowledge Acquisition from SentiWordNet
Pre-training Task: Label-aware Masked Language Model
Context-aware Sentiment Attention
Part-of-Speech Tagging
Subtask#1: Early Fusion
Subtask#2: Late Supervision

Senti LARE (Ke et al. 2020)



Sense-BERT (Levine et al. 2020)


Entity Knowledge

Factual Knowledge

Commonsense Knowledge

Linguistic Knowledge

Knowledge Acquisition: Mechanisms

Knowledge Acquisition Mechanisms

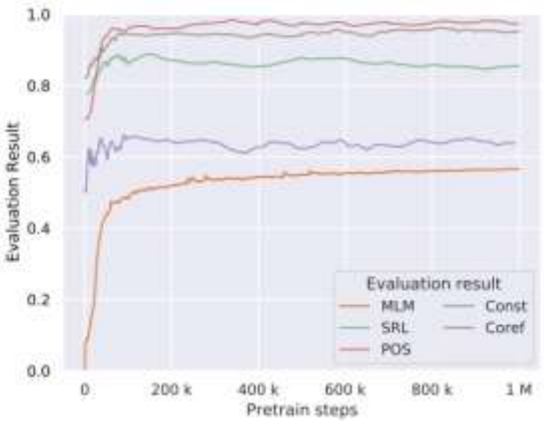
- How and why LLMs can acquire or forget knowledge from different sources?

- Investigate this by diving into the dynamics of LLMs' learning procedure

Knowledge Acquisition Mechanisms

- Dynamics investigation example #1: ALBERT knowledge evolution (Chiang et al., 2020)

Semantic and Syntactic Knowledge



(b) Masked LM accuracy and F1 scores of different probing tasks over the course of pretraining

World Knowledge

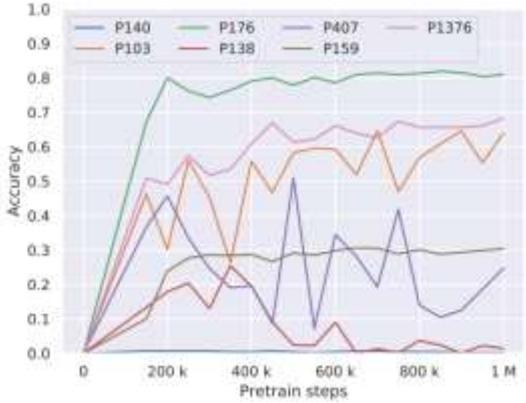


Figure 6: World knowledge development during pre-training evaluated every 50k pretrain steps. Types of relation, and template are shown in Table 1

- Semantic and syntactic knowledge are learned simultaneously in ALBERT.
- ALBERT seems to be dynamically renewing its knowledge about the world.

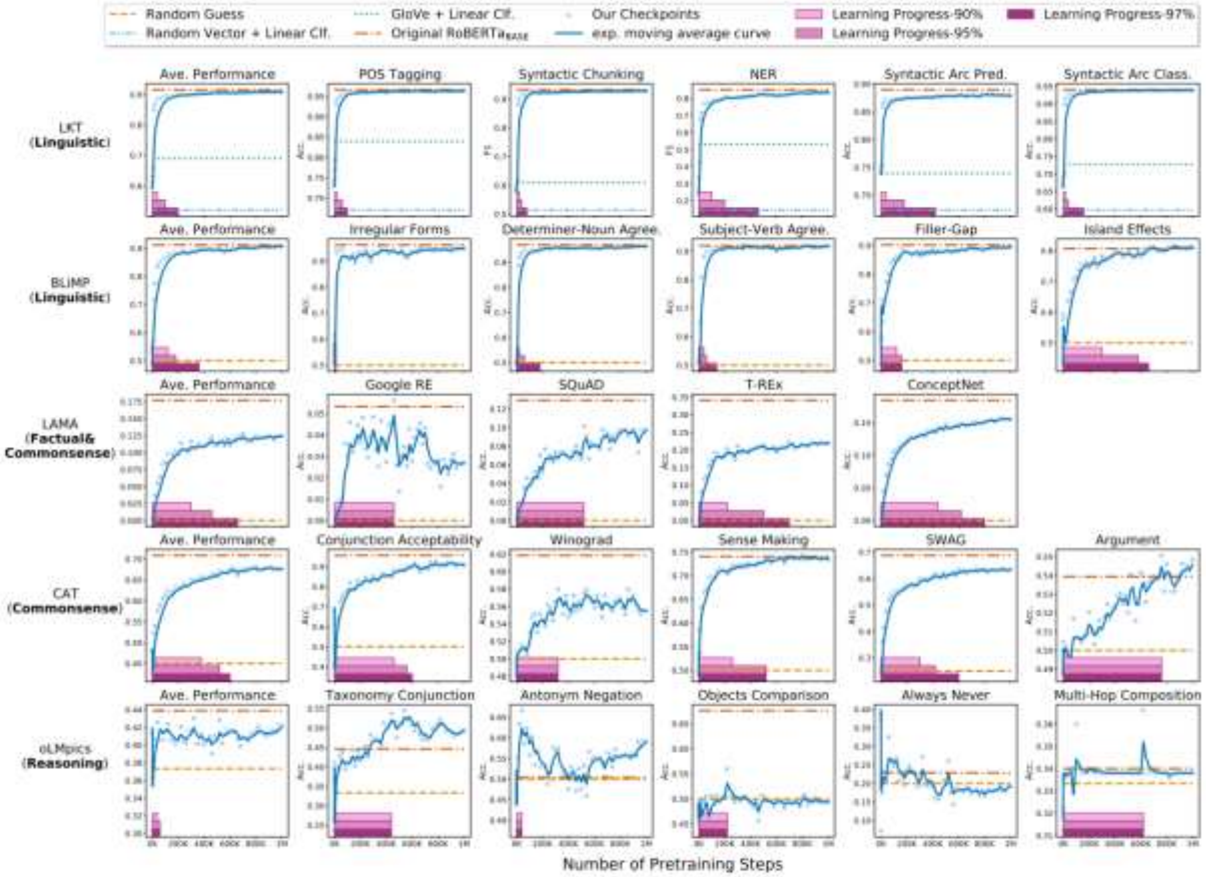
Chiang et al. Pretrained language model embryology: The birth of ALBERT. EMNLP 2020.

Knowledge Acquisition Mechanisms

- Dynamics investigation example #2: RoBERTa knowledge evolution (Liu et al., 2020)

- Linguistic knowledge can be learned quickly and robustly

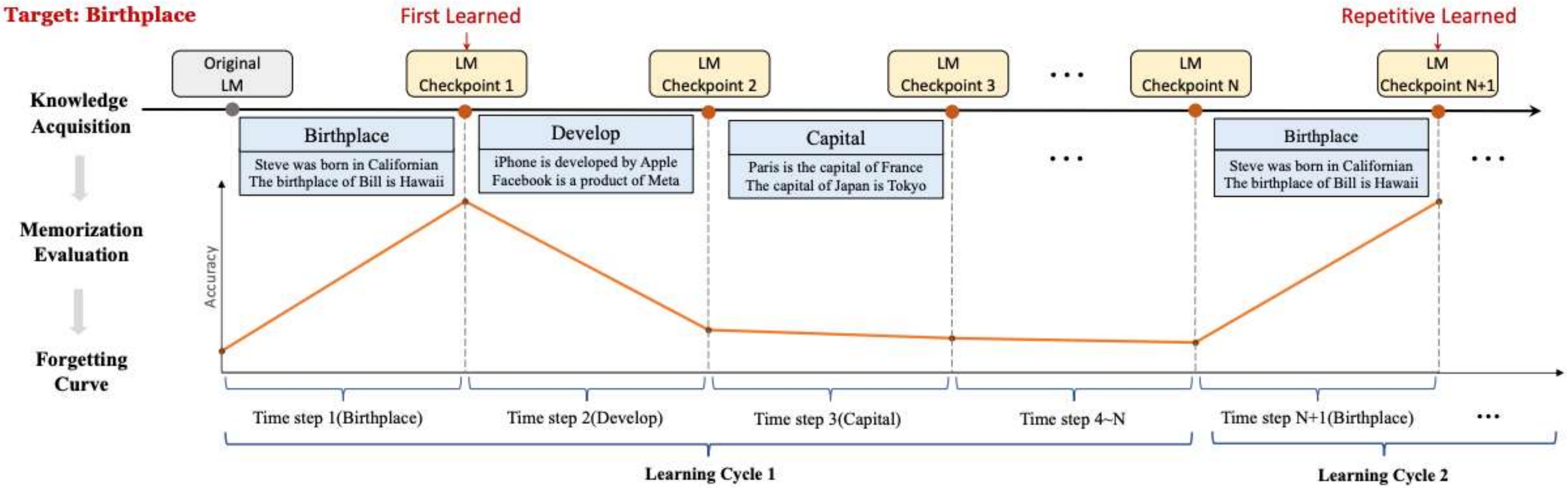
- Factual knowledge is learned slowly and domain-sensitive



Liu et al. Probing across time: What does RoBERTa know and when? Findings of EMNLP 2021.

Knowledge Acquisition Mechanisms

- Dynamics investigation example #3: Learning and forgetting dynamics of factual knowledge (Cao et al., 2023)

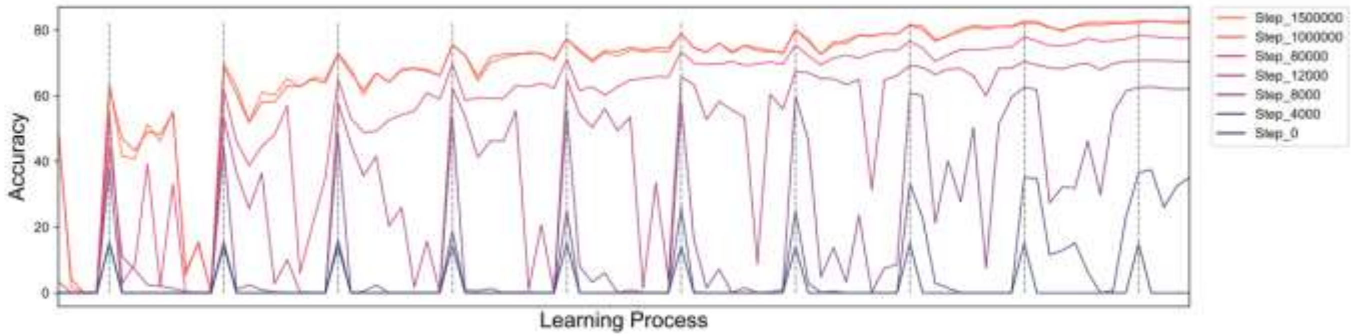


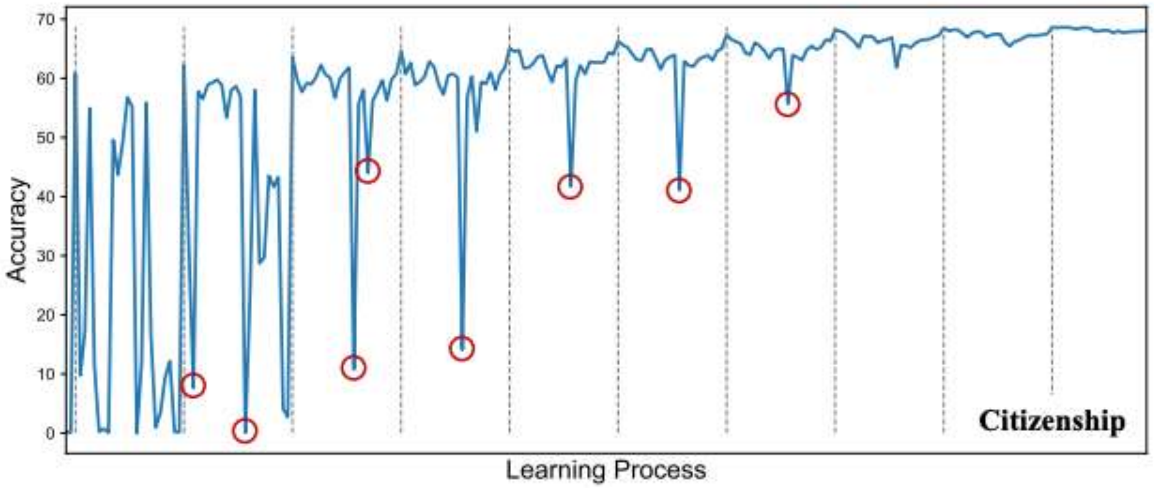
Cao et al. Retentive or Forgetful? Diving into the Knowledge Memorizing Mechanism of Language Models. 2023.

Knowledge Acquisition Mechanisms

- Dynamics investigation example #3: Learning and forgetting dynamics of factual knowledge (Cao et al., 2023)

- Pretraining is the key to shift "short-term" memory to "long-term" memory
- Existence of "singularity" where memory collapsed but quickly recovered





Cao et al. Retentive or Forgetful? Diving into the Knowledge Memorizing Mechanism of Language Models. 2023.

Take-aways for Knowledge Acquisition

- Knowledge in LLMs are learned from multiple sources via multiple learning approaches
  - Learning From Texts: Self-supervised Pretraining
  - Learning From Instruction Data: Supervised Fine-tuning
  - Learning From Human Feedback: Supervised Alignment
  - Learning From Structural Data: Structured Knowledge Injection

- The underlying mechanisms of how LLMs learn knowledge still need further investigation

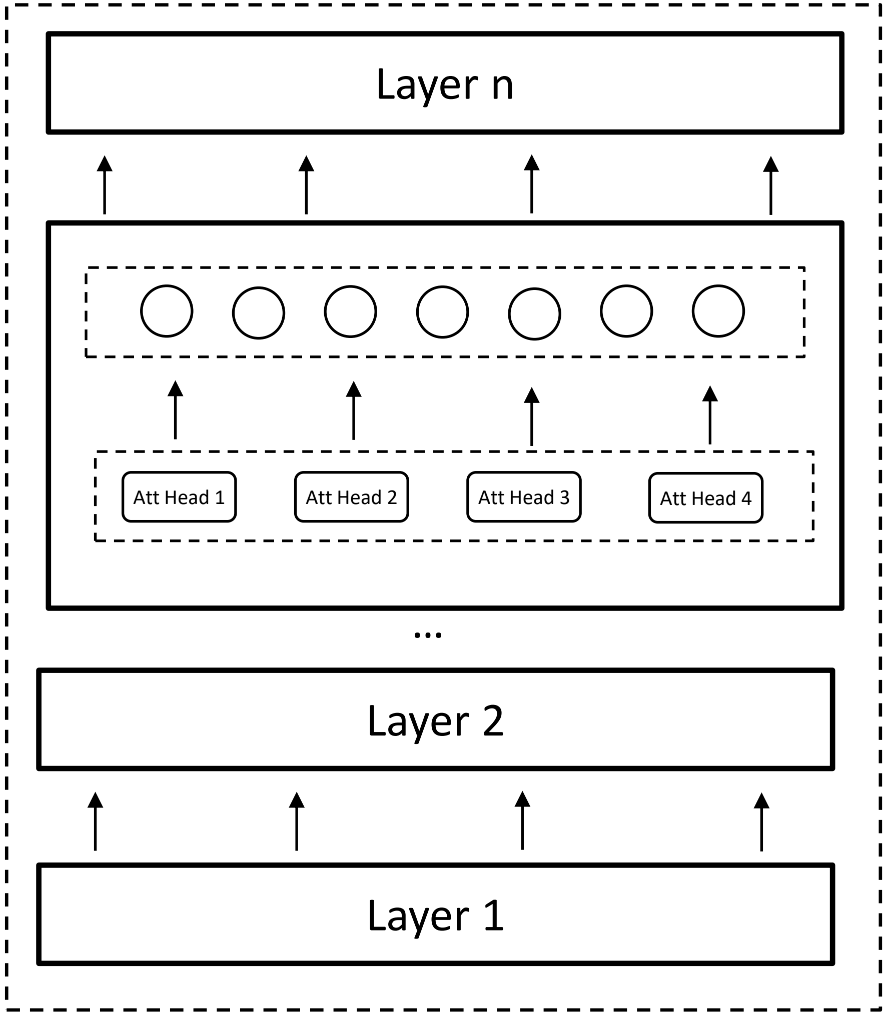
Knowledge Attribution: Opening the Blackbox

Knowledge Attribution

- How LLMs encode, transform and store the acquired knowledge?
- Can we associate specific knowledge with certain modules or neurons within a language model?
- Can we control the knowledge in the language model by modifying these specific modules?

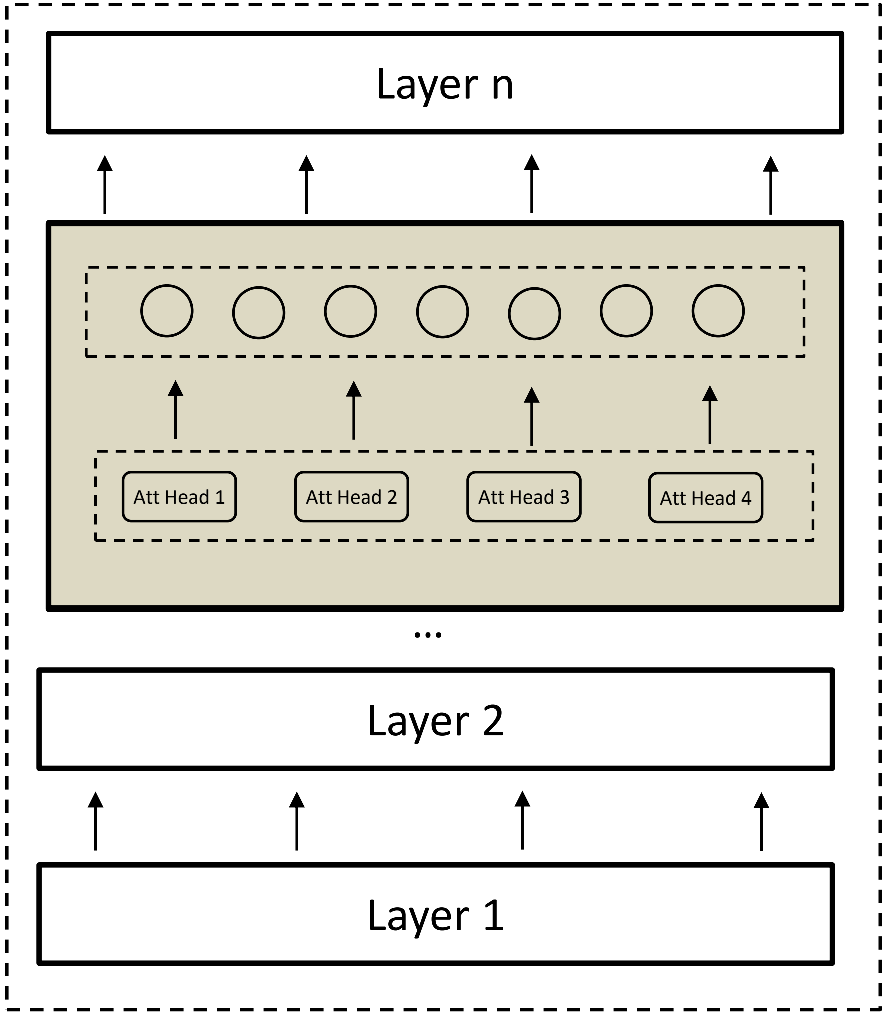
Knowledge Attribution

- Attribute stored knowledge to different-level of modules in neural networks



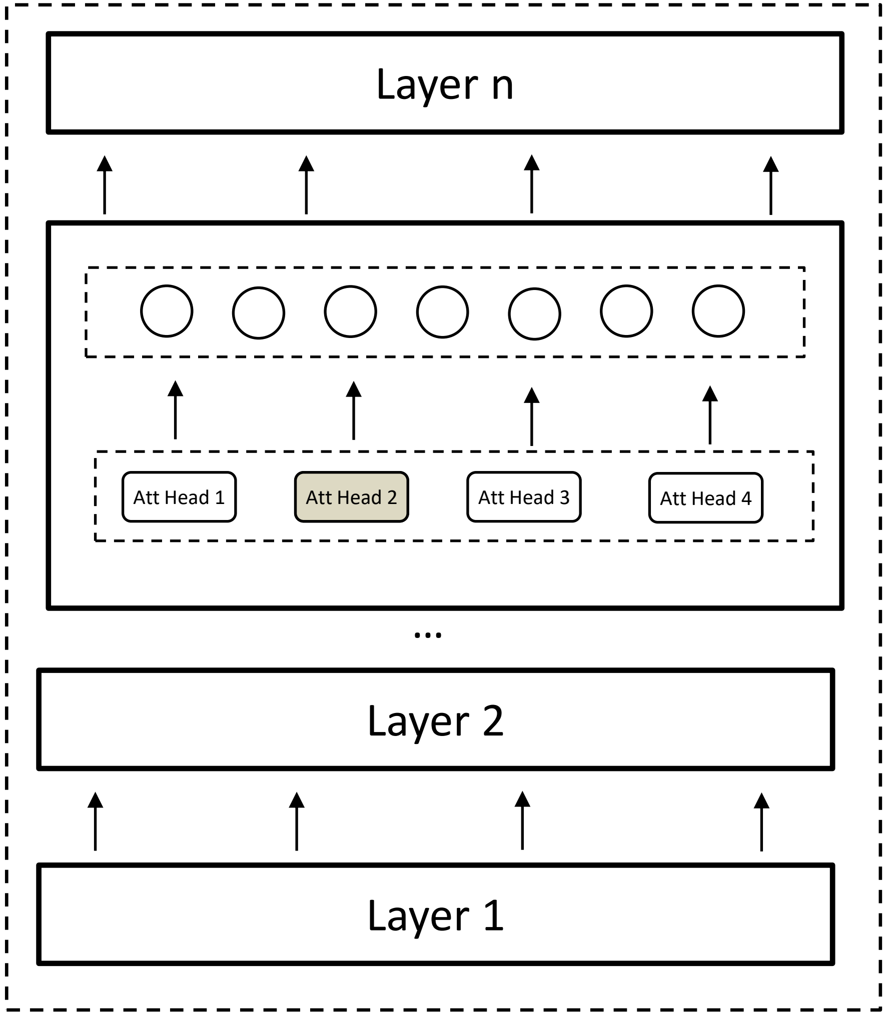
Knowledge Attribution

- Attribute stored knowledge to different-level of modules in neural networks
  - Layers



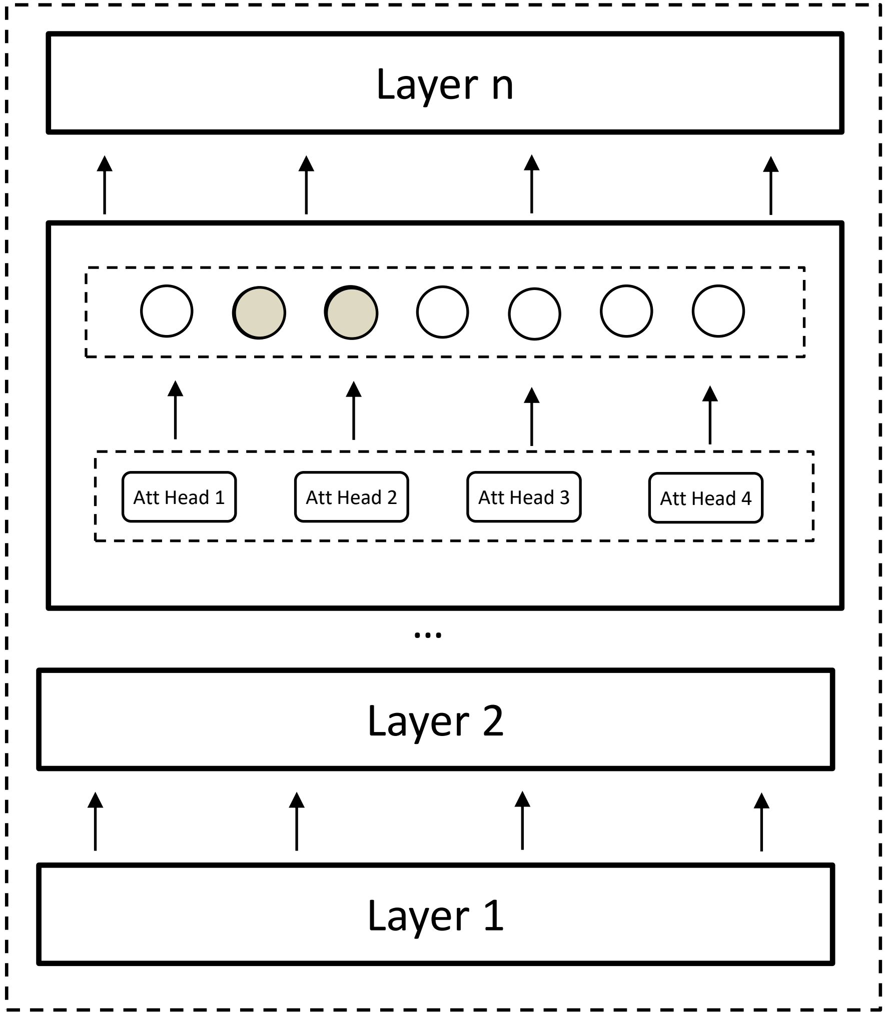
Knowledge Attribution

- Attribute stored knowledge to different-level of modules in neural networks
  - Layers
  - Modules



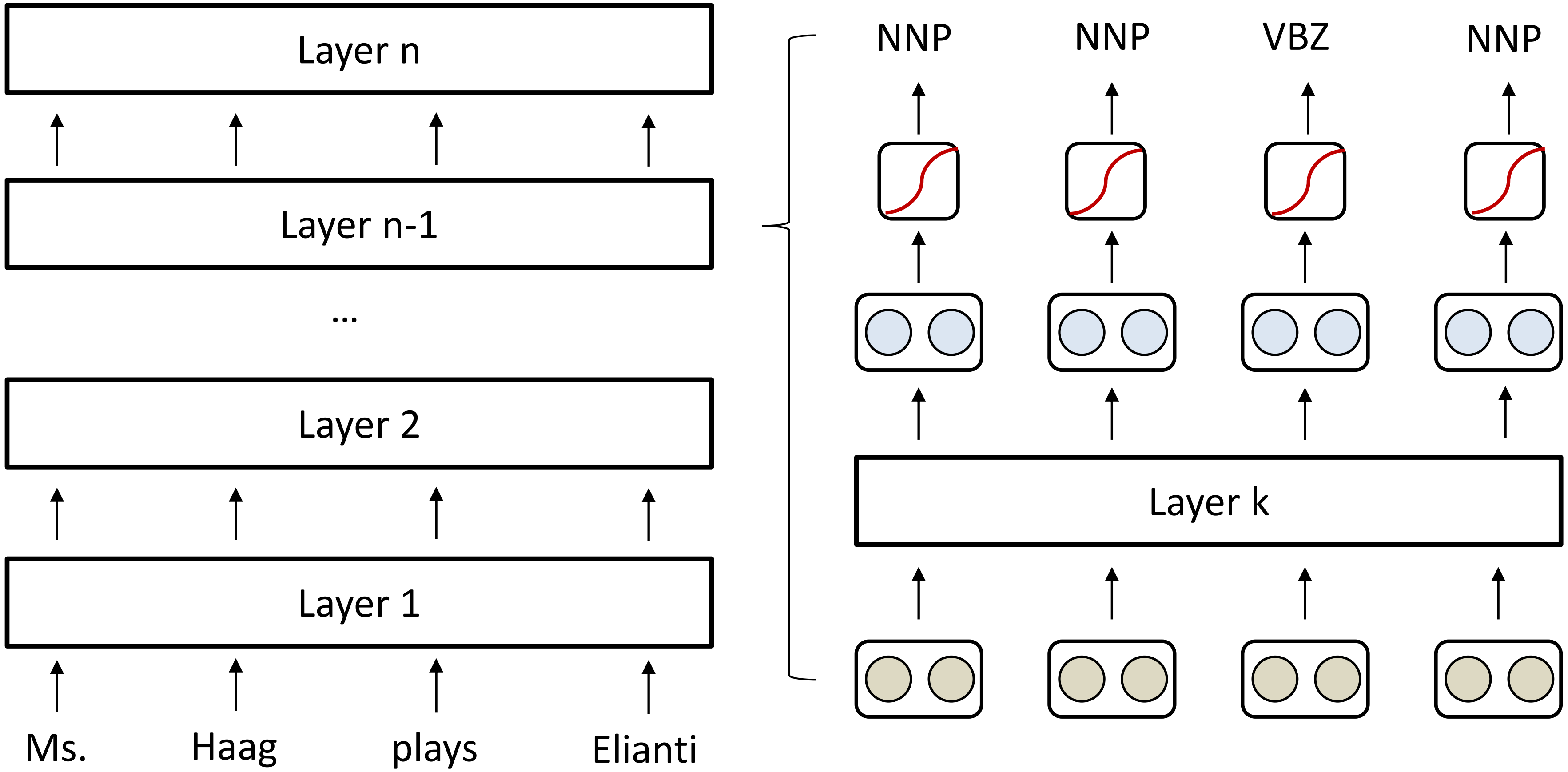
Knowledge Attribution

- Attribute stored knowledge to different-level of modules in neural networks
  - Layers
  - Modules
  - Neurons



Layer-wise Knowledge Attribution

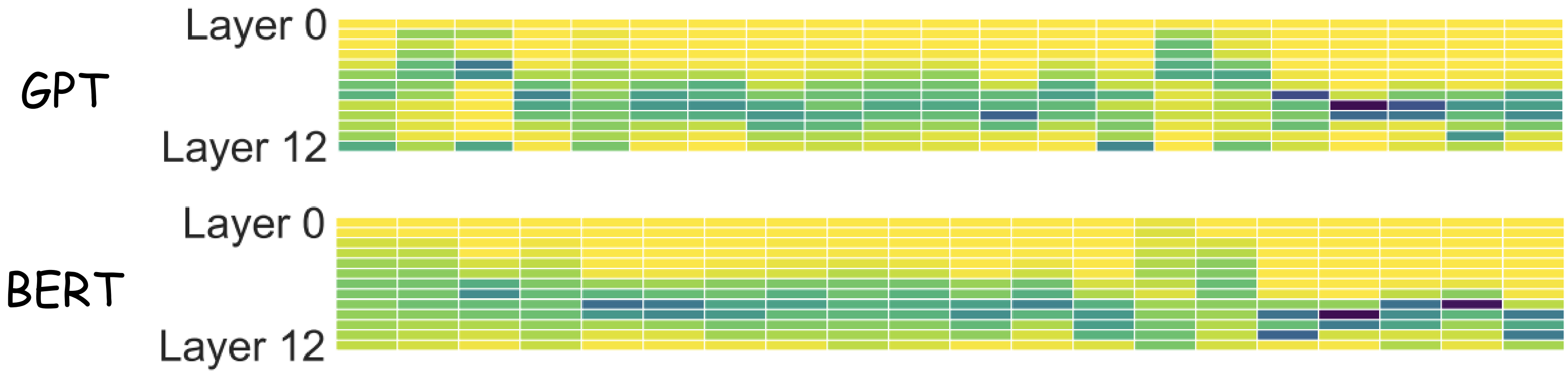
- Attributing knowledge to each layer of NNs by training a task-specific classifier for representations on each layer

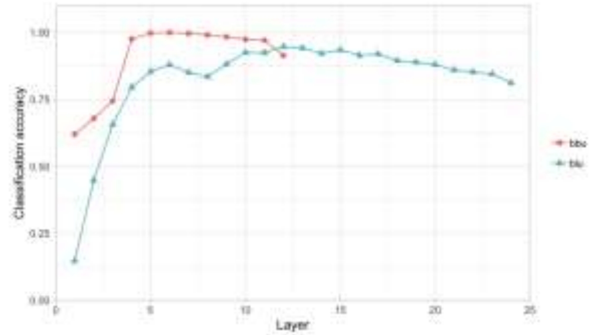


Nelson F. Liu. Linguistic Knowledge and Transferability of Contextual Representations. NAACL 2019.

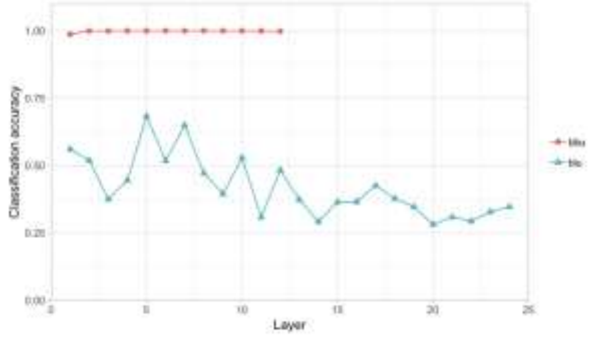
Layer-wise Knowledge Attribution

- Example #1: Linguistic Knowledge(Liu et al. 2019; Lin et al. 2019)





Main Auxiliary



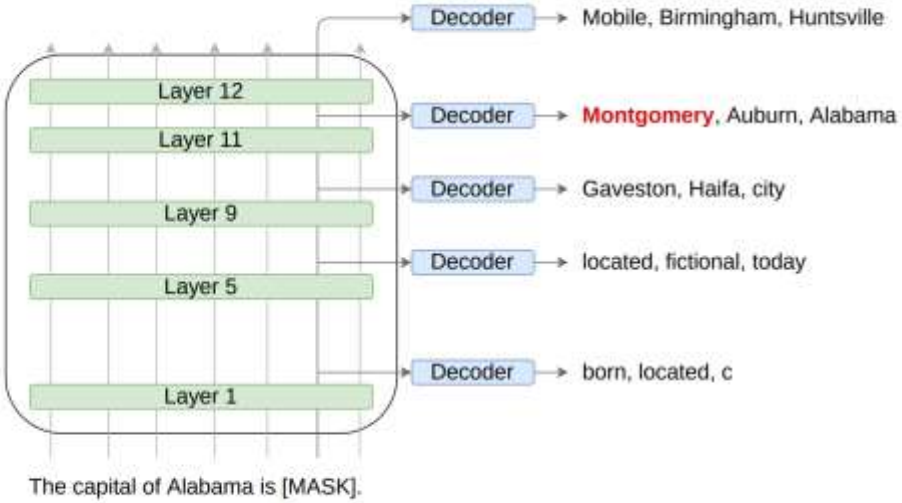
Subject Noun

- ➢ High Layers: more task-specific but fail on tasks requiring fine-grained linguistic knowledge
- ➢ Middle& Lower Layers: better linguistic transferability
- ➢ BERT encodes positional information about word tokens well on its lower layers

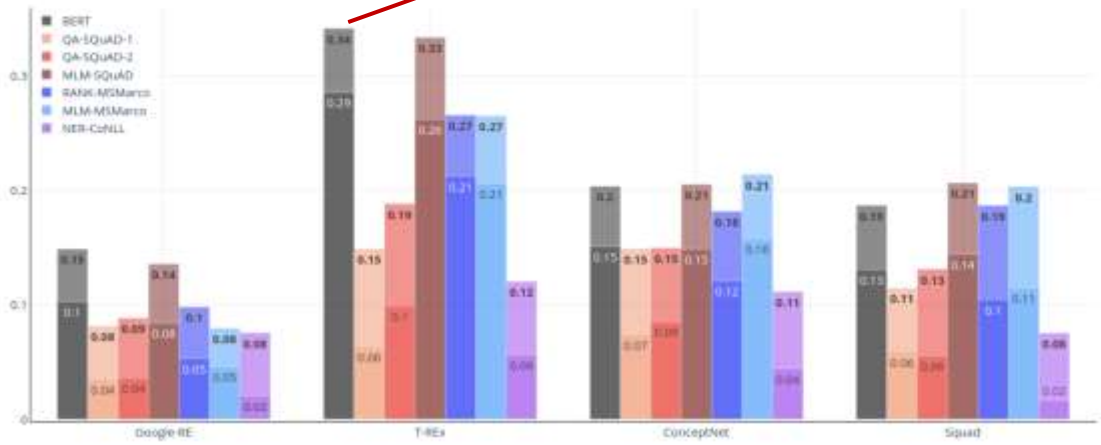
Lin et al. Open Sesame: Getting inside BERT's Linguistic Knowledge. 2019.
Liu et al. Linguistic knowledge and transferability of contextual representations. 2019.

Layer-wise Knowledge Attribution

- Example #2: Factual knowledge


Decoder
Mobile, Birmingham, Huntsville
Layer 12
Decoder
Montgomery, Auburn, Alabama
Layer 11
Decoder
Gaveston, Haifa, city
Layer 9
Decoder
located, fictional, today
Layer 5
Decoder
born, located, c
Layer 1
The capital of Alabama is [MASK].

18% knowledge are forgotten

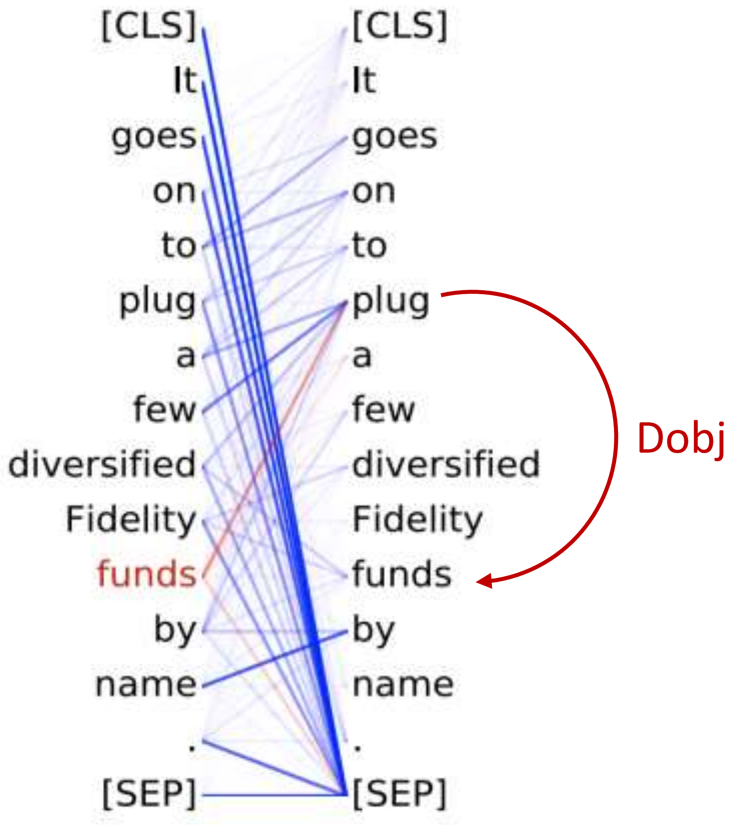


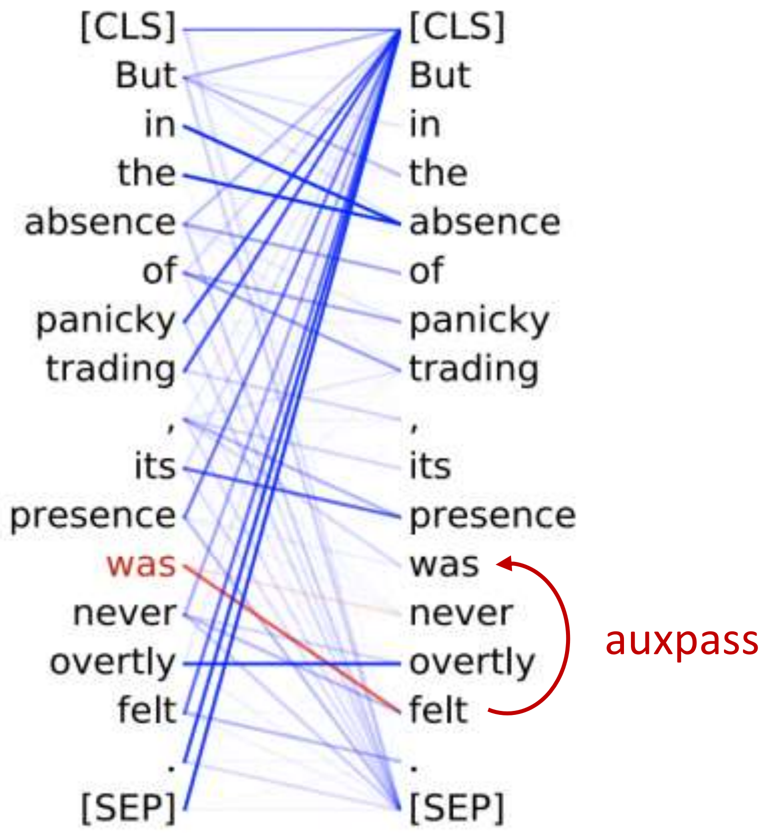
Knowledge forgetting across layers: Intermediate layers contain relational knowledge that is absent in the final layer

Wallat et al. BERTnesia: Investigating the capture and forgetting of knowledge in BERT. 2021.

Module-based Knowledge Attribution

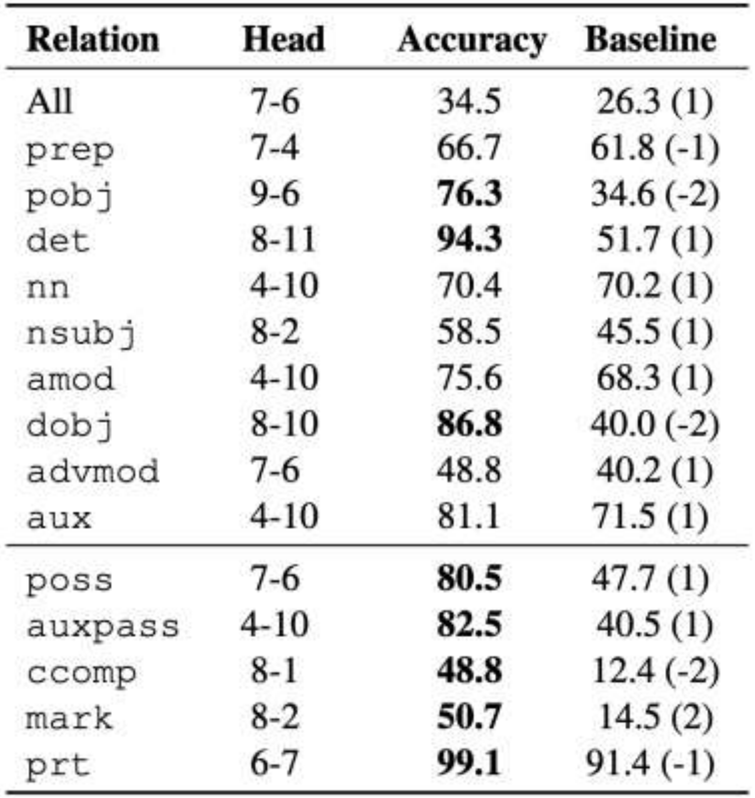
- Analyze knowledge attribution by looking into attention matrix(Clark et al., 2019; Htut et al., 2019; Lin et al., 2019)





Module-based Knowledge Attribution

- Module-based knowledge attribution for syntax knowledge (Clark et al., 2019)
  - Evaluate each attention head on dependency parsing dataset



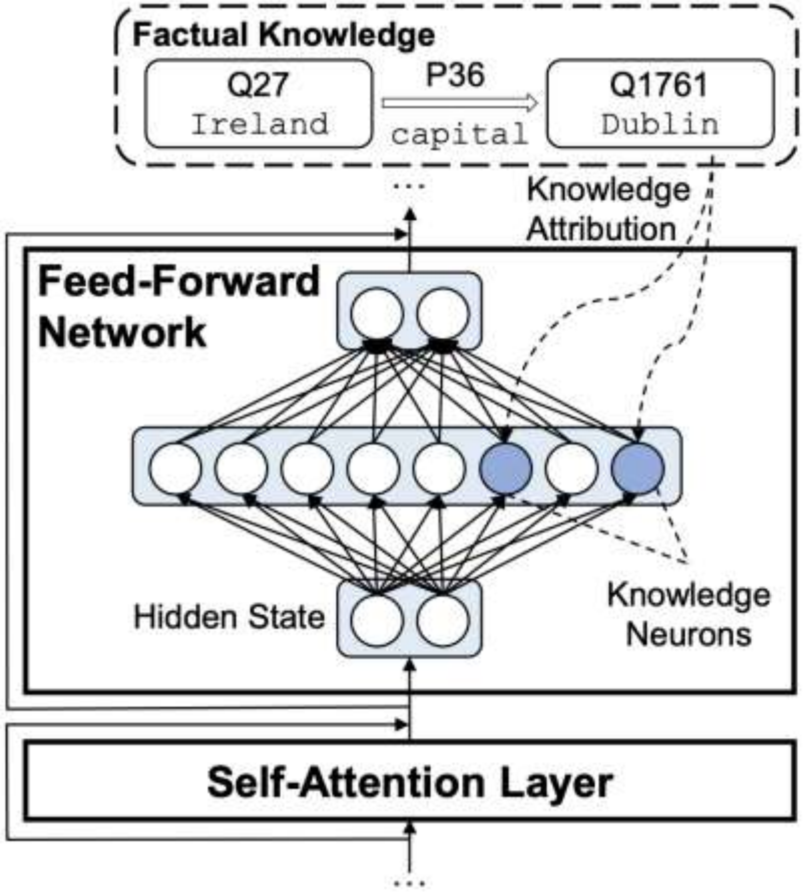
Relation	Head	Accuracy	Baseline
All	7-6	34.5	26.3 (1)
prep	7-4	66.7	61.8 (-1)
pobj	9-6	**76.3**	34.6 (-2)
det	8-11	**94.3**	51.7 (1)
nn	4-10	70.4	70.2 (1)
nsubj	8-2	58.5	45.5 (1)
amod	4-10	75.6	68.3 (1)
dobj	8-10	**86.8**	40.0 (-2)
advmod	7-6	48.8	40.2 (1)
aux	4-10	81.1	71.5 (1)
poss	7-6	**80.5**	47.7 (1)
auxpass	4-10	**82.5**	40.5 (1)
ccomp	8-1	**48.8**	12.4 (-2)
mark	8-2	**50.7**	14.5 (2)
prt	6-7	**99.1**	91.4 (-1)

- No single head does well at syntax "overall"
- Certain attention heads specialize to specific dependency relations.

Clark et al. What does BERT look at? an analysis of BERT's attention. . 2019.

Neuron-based Knowledge Attribution

- Can we attribute knowledge to specific neurons in PLMs?



Dai, et al. Knowledge Neurons in Pretrained Transformers. ACL 2022.

Neuron-based Knowledge Attribution

- How to find Knowledge Neuron: Integrate Gradients (Dai et al., 2022)

$$\mathrm{Attr}\boxed{\left(w_i^{(l)}\right)} = \overline{w}_i^{(l)} \int_{\alpha=0}^{1} \frac{\partial \boxed{\mathrm{P}_x\left(\alpha \overline{w}_i^{(l)}\right)}}{\partial w_i^{(l)}} d\alpha,$$

i-th neurons in l-th FFN

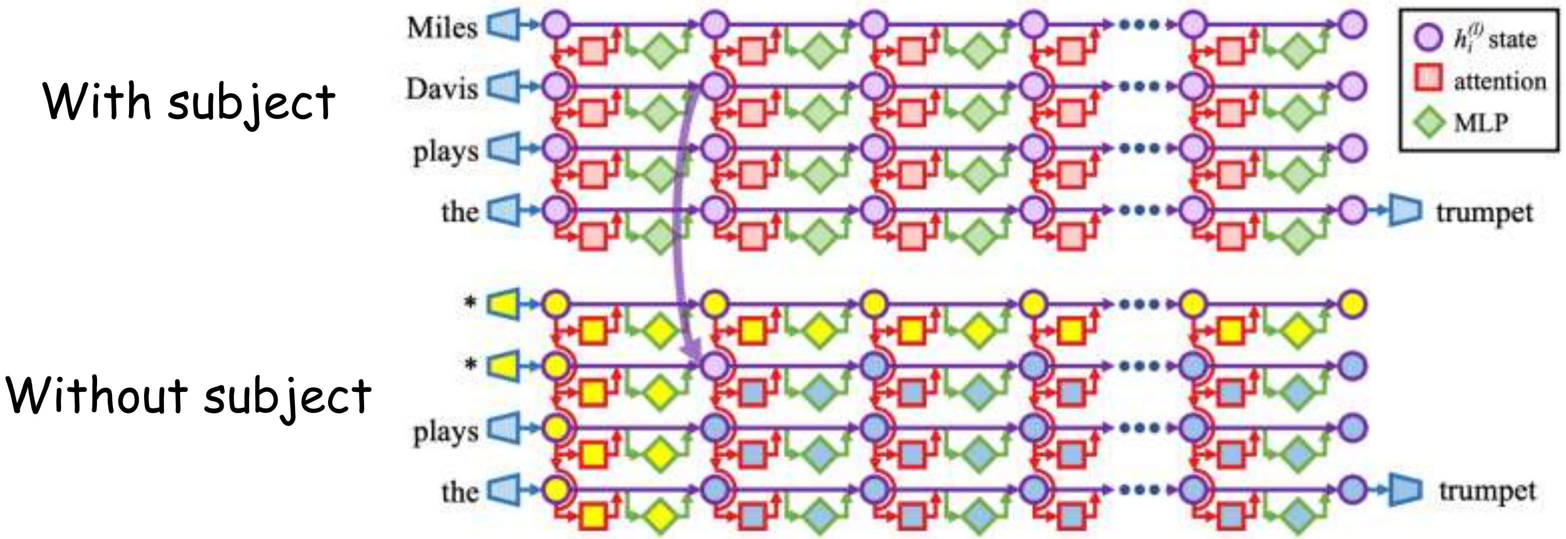
Probability of correct answer

$Attr\left(w_i^{(l)}\right)$: the probability changes caused by modifying $w_i^{(l)}$

Dai, et al. Knowledge Neurons in Pretrained Transformers. ACL 2022.

Neuron-based Knowledge Attribution

- How to find Knowledge Neuron: Causal Tracing (Meng et al., 2022)



- Factual knowledge can be associated with feed forward modules in middle or higher layers.

Meng, et al. Locating and Editing Factual Associations in GPT. NeurIPS 2022.

Take-aways for Knowledge Attribution

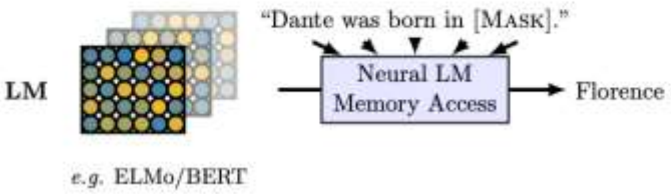
- Lower layers of PLMs often encode the coarse-grained and general information of knowledge

- Fine-grained and task-specific knowledge are mostly stored in higher layers and different modules

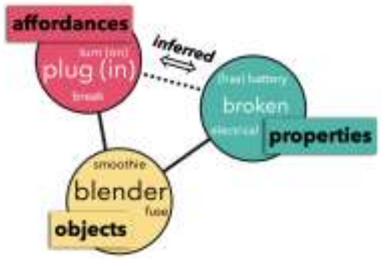
Knowledge Probing: How Much do LLMs Know about the World?

Knowledge Probing

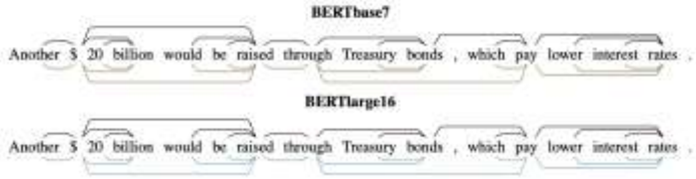
- Investigate the types of knowledge stored in LLMs



Factual Knowledge

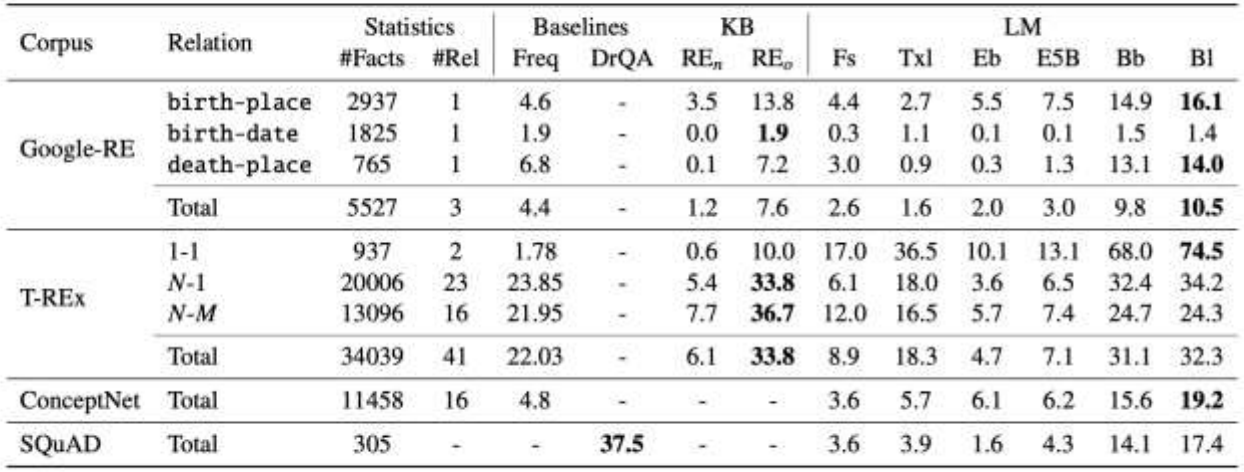


Commonsense Knowledge


BERTbase7
Another $ 20 billion would be raised through Treasury bonds , which pay lower interest rates .
BERTlarge16
Another $ 20 billion would be raised through Treasury bonds , which pay lower interest rates .

Linguistic Knowledge

- Quantify the amount of knowledge stored in LLMs



Corpus	Relation	Statistics		Baselines		KB		LM					
		#Facts	#Rel	Freq	DrQA	RE_n	RE_o	Fs	Txl	Eb	E5B	Bb	Bl
Google-RE	`birth-place`	2937	1	4.6	-	3.5	13.8	4.4	2.7	5.5	7.5	14.9	**16.1**
	`birth-date`	1825	1	1.9	-	0.0	**1.9**	0.3	1.1	0.1	0.1	1.5	1.4
	`death-place`	765	1	6.8	-	0.1	7.2	3.0	0.9	0.3	1.3	13.1	**14.0**
	Total	5527	3	4.4	-	1.2	7.6	2.6	1.6	2.0	3.0	9.8	**10.5**
T-REx	1-1	937	2	1.78	-	0.6	10.0	17.0	36.5	10.1	13.1	68.0	**74.5**
	N-1	20006	23	23.85	-	5.4	**33.8**	6.1	18.0	3.6	6.5	32.4	34.2
	N-*M*	13096	16	21.95	-	7.7	**36.7**	12.0	16.5	5.7	7.4	24.7	24.3
	Total	34039	41	22.03	-	6.1	**33.8**	8.9	18.3	4.7	7.1	31.1	32.3
ConceptNet	Total	11458	16	4.8	-	-	-	3.6	5.7	6.1	6.2	15.6	**19.2**
SQuAD	Total	305	-	-	**37.5**	-	-	3.6	3.9	1.6	4.3	14.1	17.4

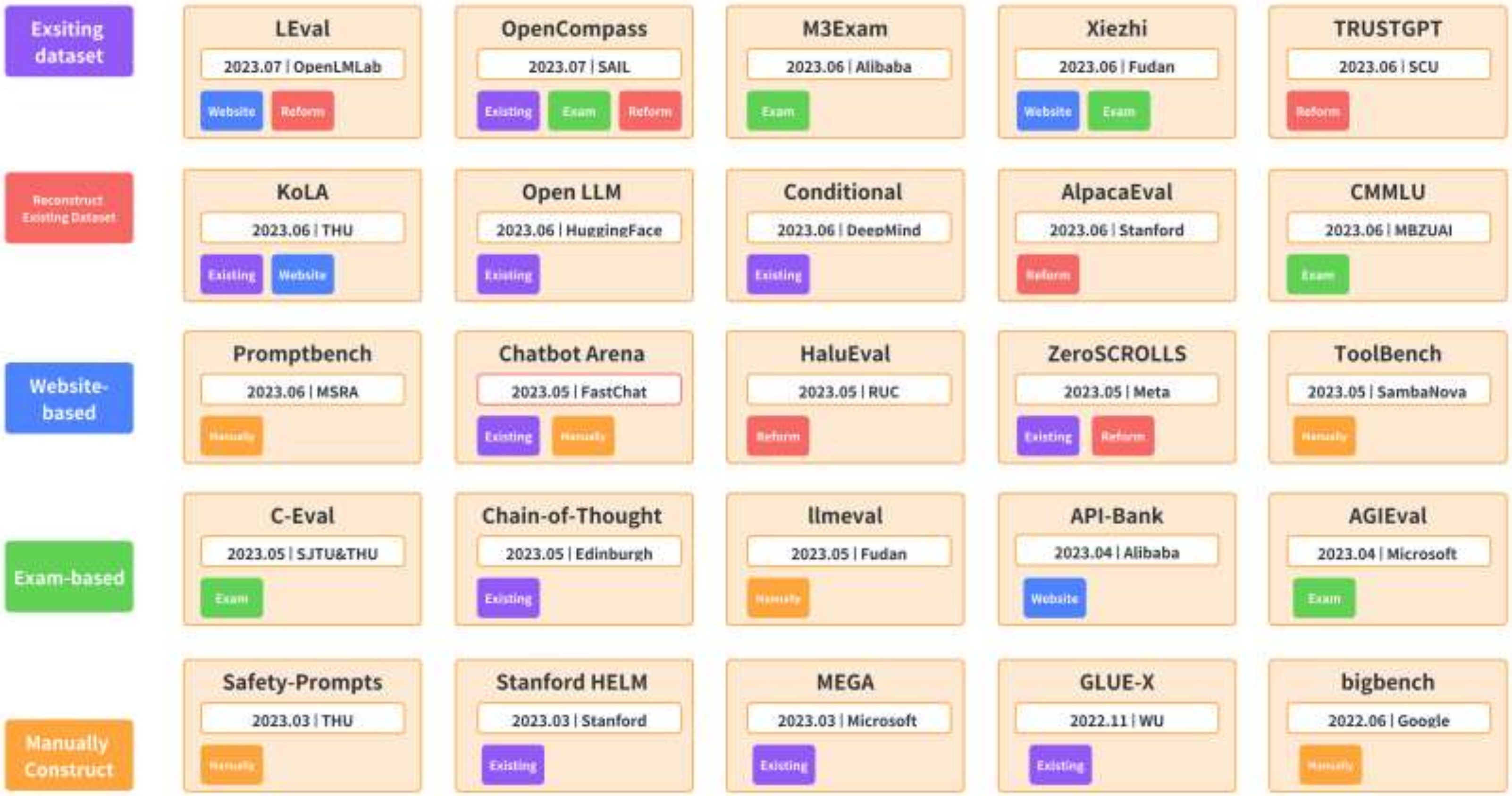
Knowledge Probing Benchmarks

- Knowledge-specific Probing Benchmark
  - Focus on one specific kinds of abilities of LLMs

Knowledge Type	Benchmark	Formulation	Remark
Linguistic Knowledge	Open Sesame (Lin et al., 2019)	diagnostic classifier and attention	
	LKT (Liu et al., 2019b)	token or token pair labeling	
	NPI probe (Warstadt et al., 2019)	probing classifier	
	Edge probe (Tenney et al., 2019)	edge probing	
	MDL probe (Voita and Titov, 2020)	minimum description length	
	LM diagnostics (Ettinger, 2020)	text filling	
	BLiMP (Warstadt et al., 2020)	sentence scores comparison	
Factual Knowledge	LAMA (Petroni et al., 2019)	text filling	
	X-FACTR (Jiang et al., 2020a)	text filling	
	Multilingual LAMA (Kassner et al., 2021)	text filling	multilingual
	Bio LAMA (Sung et al., 2021)	text filling	biology
Commonsense Knowledge	CAT (Zhou et al., 2020a)	sentence scores comparison	
	NumerSense (Lin et al., 2020b)	text filling	numerical
	Physical Commonsense (Forbes et al., 2019)	probing classifier	physical

Knowledge Probing Benchmarks

- General Knowledge Evaluation Benchmark
  - General/Hybrid knowledge evaluation with higher difficulty



Exsiting dataset

Reconstruct Existing Dataset

Website-based

Exam-based

Manually Construct

LEval 2023.07 | OpenLMLab Website Reform	**OpenCompass** 2023.07 | SAIL Existing Exam Reform	**M3Exam** 2023.06 | Alibaba Exam	**Xiezhi** 2023.06 | Fudan Website Exam	**TRUSTGPT** 2023.06 | SCU Reform
KoLA 2023.06 | THU Existing Website	**Open LLM** 2023.06 | HuggingFace Existing	**Conditional** 2023.06 | DeepMind Existing	**AlpacaEval** 2023.06 | Stanford Reform	**CMMLU** 2023.06 | MBZUAI Exam
Promptbench 2023.06 | MSRA Manually	**Chatbot Arena** 2023.05 | FastChat Existing Manually	**HaluEval** 2023.05 | RUC Reform	**ZeroSCROLLS** 2023.05 | Meta Existing Reform	**ToolBench** 2023.05 | SambaNova Manually
C-Eval 2023.05 | SJTU&THU Exam	**Chain-of-Thought** 2023.05 | Edinburgh Existing	**llmeval** 2023.05 | Fudan Manually	**API-Bank** 2023.04 | Alibaba Website	**AGIEval** 2023.04 | Microsoft Exam
Safety-Prompts 2023.03 | THU Manually	**Stanford HELM** 2023.03 | Stanford Existing	**MEGA** 2023.03 | Microsoft Existing	**GLUE-X** 2022.11 | WU Existing	**bigbench** 2022.06 | Google Manually

Knowledge Probing Methods

- Prompt-based knowledge probing
  - Query LMs with task-specific prompts and assess performance according to LMs' predictions

- Feature-based knowledge probing
  - Froze parameters of LLMs, probing tasks are accomplished based on the internal representation or attention weights produced by LMs

Prompt-based Knowledge Probing

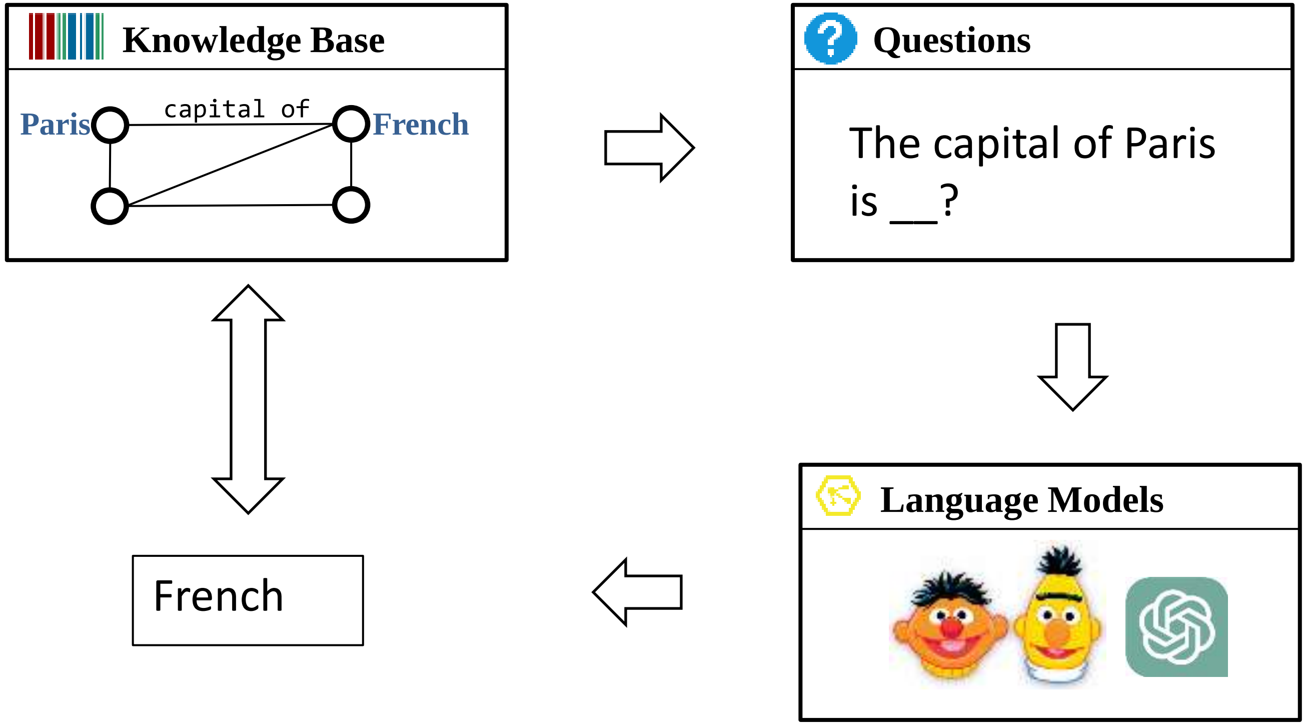
- Handcraft Discrete Prompt

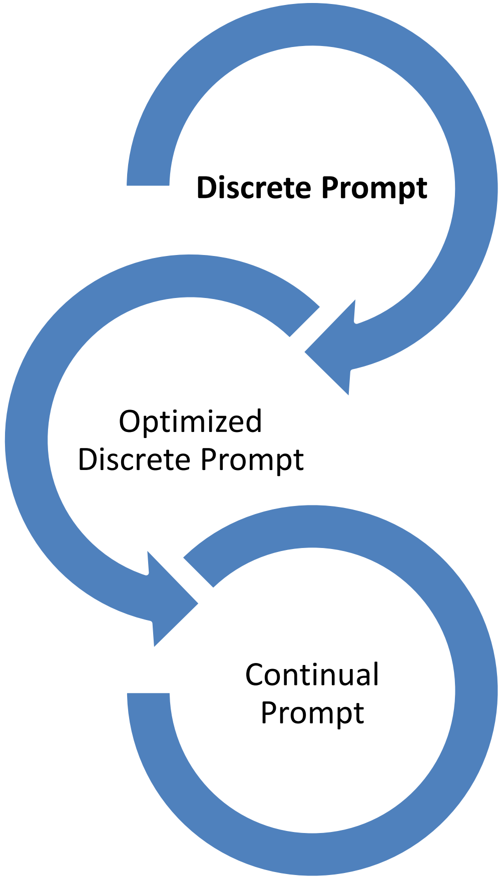
- Optimized Discrete Prompt

- Continual Prompt

Prompt-based Knowledge Probing

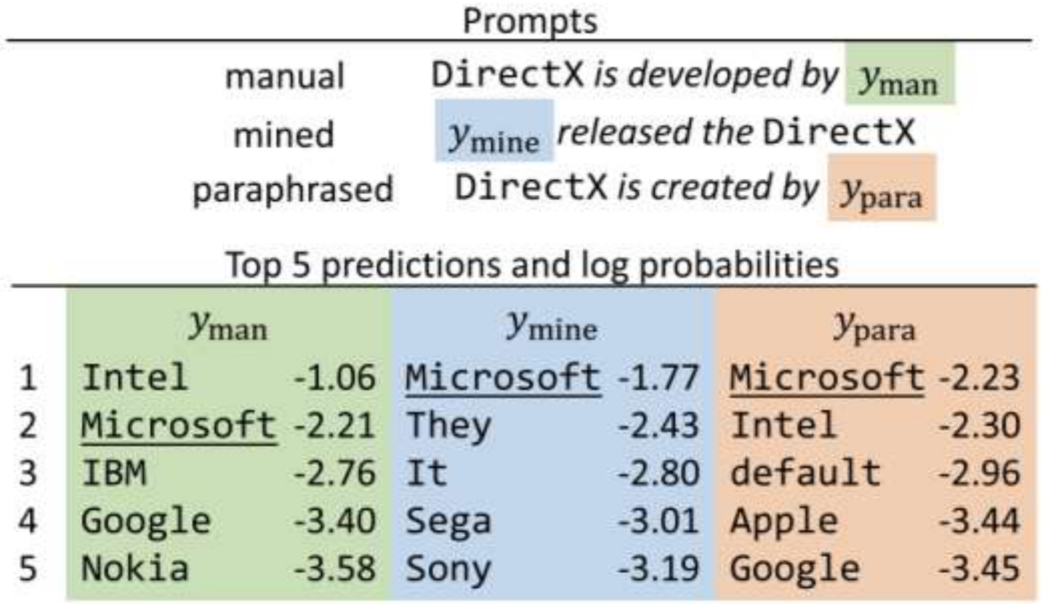
- Cloze-style Discrete prompts
  - LAMA, X-FACTR, BioLAMA, Multilingual LAMA…
  - Choice of Prompts has huge influence





Prompt-based Knowledge Probing

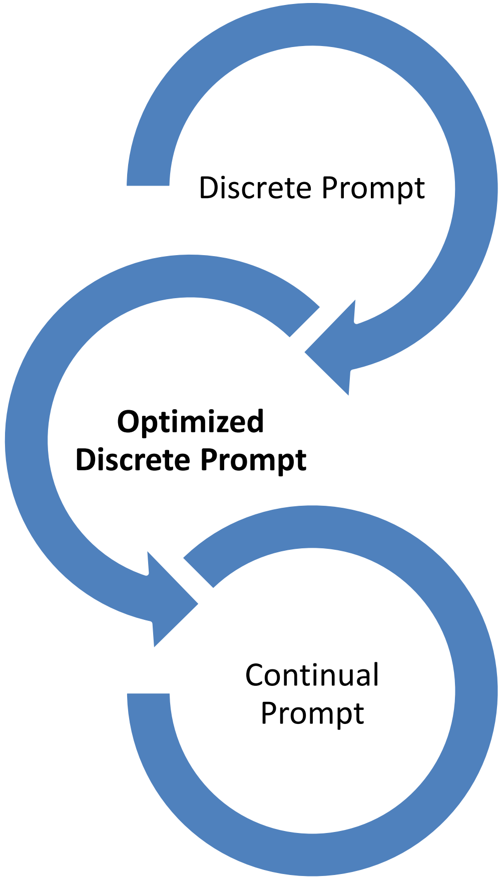
- Using optimized discrete prompts to get better performance
  - Example #1: LPAQA
  - Using retrieval and paraphrasing method to search prompts
  - Achieve better performance than manually created prompts
  - Require valid dataset



	Prompts
manual	DirectX *is developed by* y_{man}
mined	y_{mine} *released the* DirectX
paraphrased	DirectX *is created by* y_{para}

Top 5 predictions and log probabilities

	y_{man}		y_{mine}		y_{para}	
1	Intel	-1.06	Microsoft	-1.77	Microsoft	-2.23
2	Microsoft	-2.21	They	-2.43	Intel	-2.30
3	IBM	-2.76	It	-2.80	default	-2.96
4	Google	-3.40	Sega	-3.01	Apple	-3.44
5	Nokia	-3.58	Sony	-3.19	Google	-3.45



Discrete Prompt

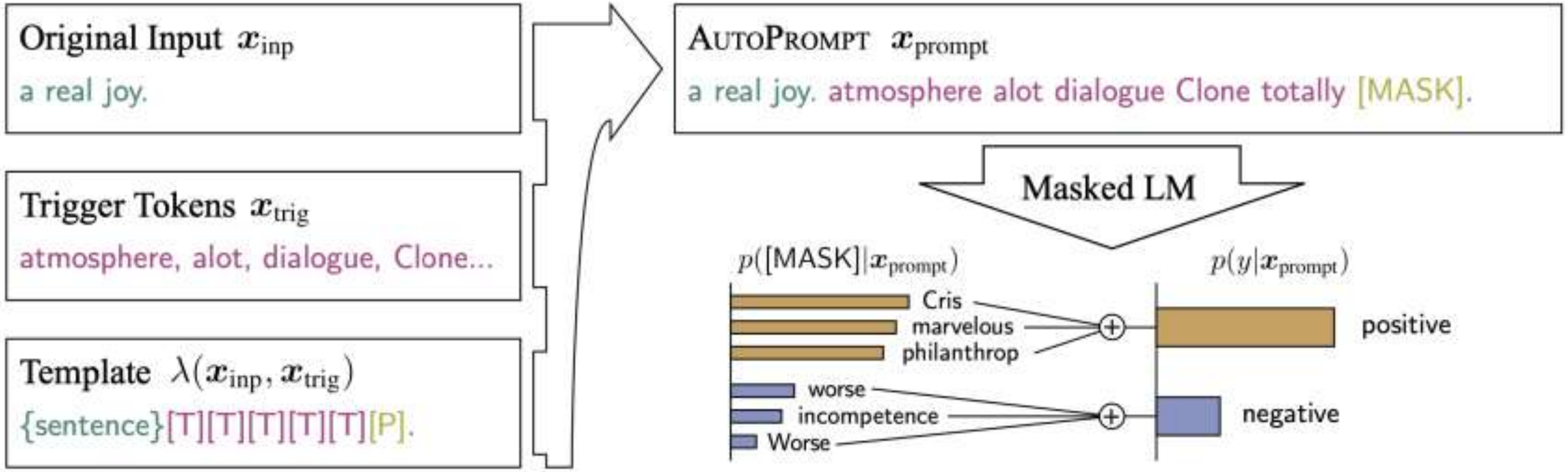
Optimized Discrete Prompt

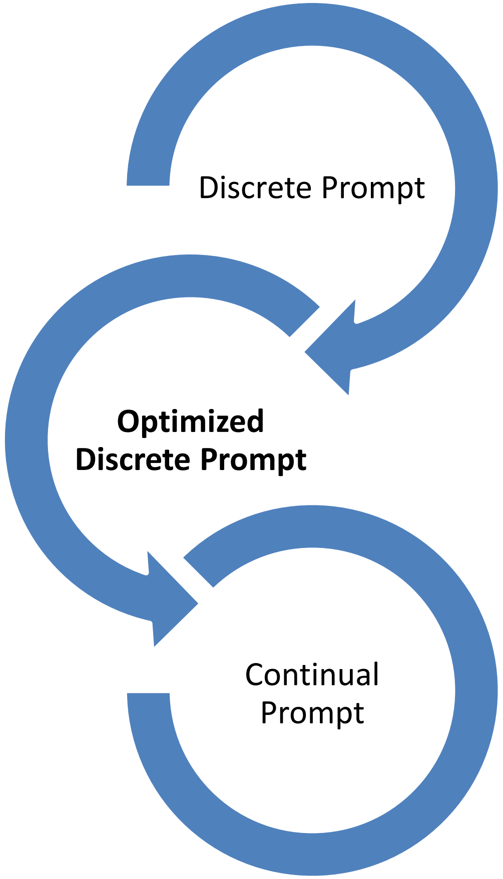
Continual Prompt

Jiang Z, Xu F F, Araki J, et al. How can we know what language models know? In TACL 2020.

Prompt-based Knowledge Probing

- Using optimized discrete prompts to get better performance
  - Example #2: AutoPrompt
  - Automatically generated prompts based on gradient-guided search
  - Discrete prompts with better performance but lack of interpretability


Original Input x_{inp}
a real joy.
Trigger Tokens x_{trig}
atmosphere, alot, dialogue, Clone...
Template $\lambda(x_{\text{inp}}, x_{\text{trig}})$
{sentence}[T][T][T][T][T][P].
AutoPrompt x_{prompt}
a real joy. atmosphere alot dialogue Clone totally [MASK].
Masked LM
$p([\text{MASK}]|x_{\text{prompt}})$
Cris
marvelous
philanthrop
worse
incompetence
Worse
$p(y|x_{\text{prompt}})$
positive
negative


Discrete Prompt

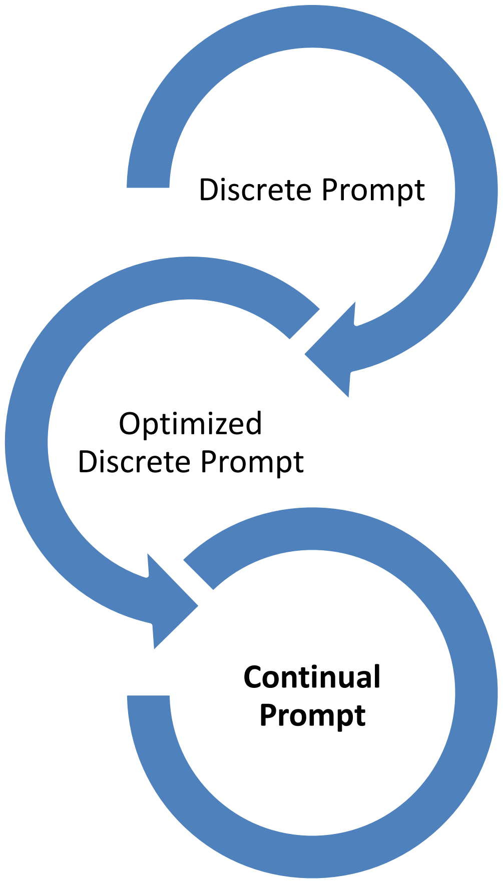
Optimized Discrete Prompt

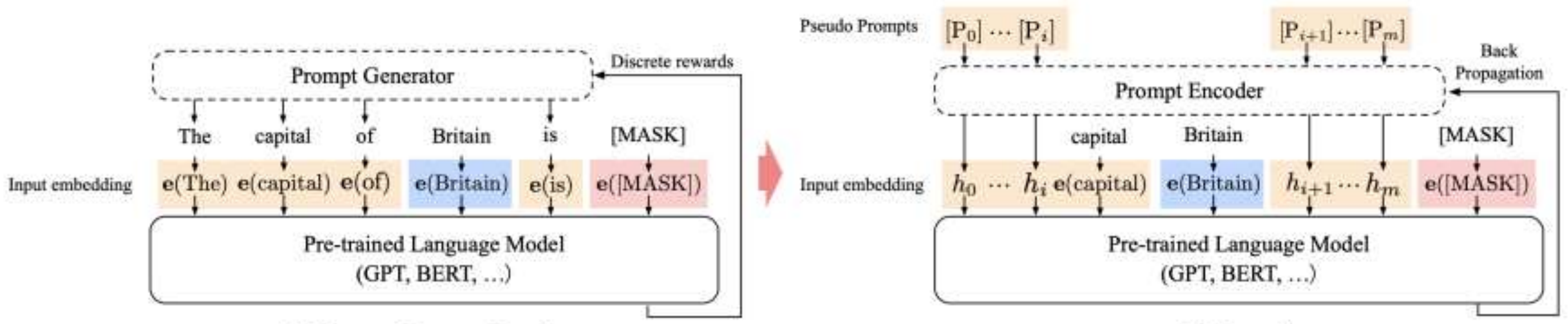
Continual Prompt

Shin T, et al. AutoPrompt: Eliciting knowledge from language models with automatically generated prompts. In EMNLP 2020.

Prompt-based Knowledge Probing

- Prompt-tuning: adding trainable vectors as soft prompt can further improve performance
  - Handcraft prompts initialization (Zhong et al., 2021)
  - Adding continual prompts on both input and transformer blocks (Li and Liang, 2021)
  - Adding prompt encoder above the input embeddings (Liu et al., 2021)
  - Ensembling multiple soft prompts (Qin et al. 2021)

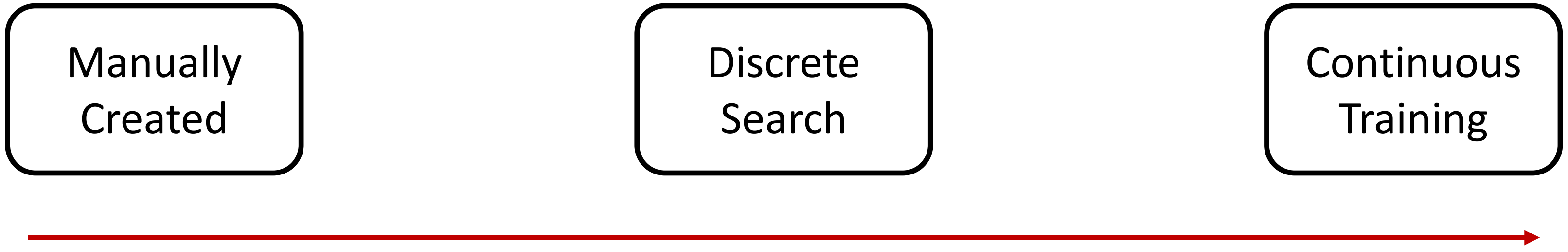



Prompt Generator
Discrete rewards
The
capital
of
Britain
is
[MASK]
Input embedding
e(The) e(capital) e(of) e(Britain) e(is) e([MASK])
Pre-trained Language Model (GPT, BERT, ...)
Pseudo Prompts
$[P_0] \cdots [P_i]$
$[P_{i+1}] \cdots [P_m]$
Back Propagation
Prompt Encoder
capital
Britain
[MASK]
Input embedding
$h_0 \cdots h_i$ e(capital) e(Britain) $h_{i+1} \cdots h_m$ e([MASK])
Pre-trained Language Model (GPT, BERT, ...)

(a) Discrete Prompt Search (b) P-tuning

Liu X, Zheng Y, Du Z, et al. GPT Understands, Too (2021)
Zhong Z, Friedman D, Chen D. Factual Probing Is [MASK]: Learning vs. Learning to Recall (2021)
Qin G, Eisner J. Learning How to Ask: Querying LMs with Mixtures of Soft Prompts (2021)

Prompt-based Knowledge Probing


Manually Created

Discrete Search

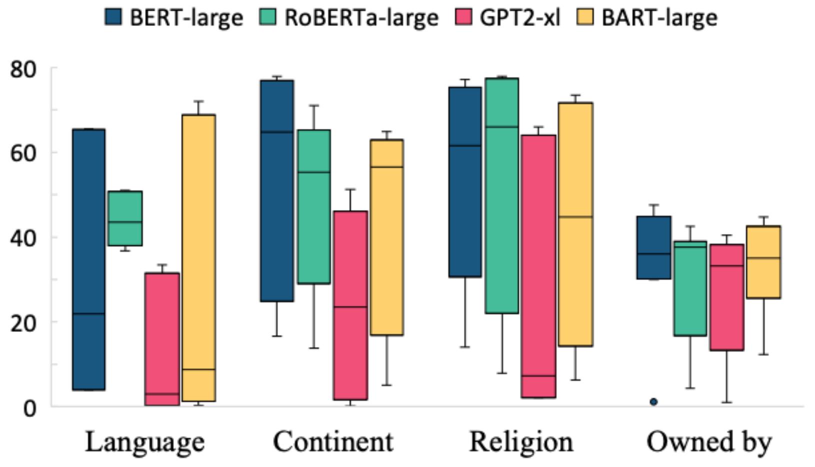
Continuous Training

Better and better performance, weaker and weaker interpretability.

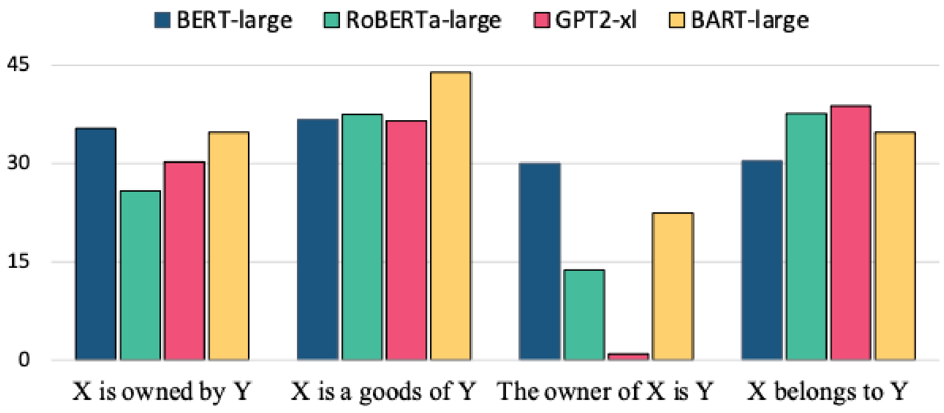
Can we absolutely trust the evaluate results of prompt-based probing?

Prompt-based Knowledge Probing

- Prompt-based probing could be inconsistent



Performance variances of PLMs on semantically equivalent prompts.



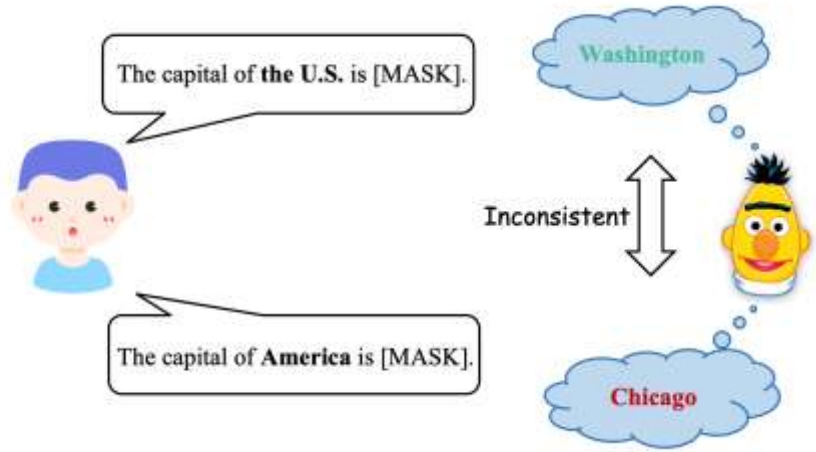
Inconsistent comparison between PLMs when prompts varies.

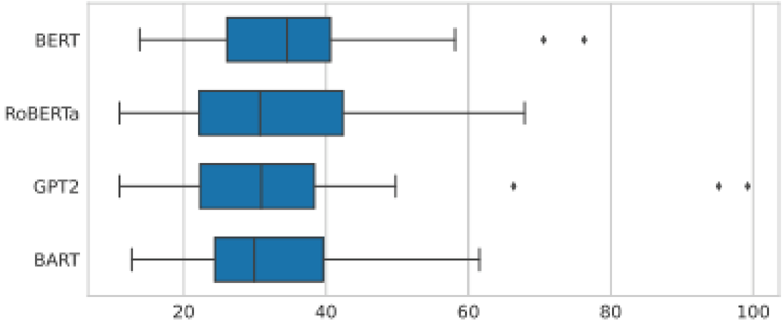
Prompt preference leads to inconsistent performance and comparison

Cao et al. Can Prompt Probe Pretrained Language Models? Understanding the Invisible Risks from a Causal View. In ACL 2022.

Prompt-based Knowledge Probing

- Prompt-based probing could be inconsistent


The capital of **the U.S.** is [MASK].
Washington
Inconsistent
The capital of **America** is [MASK].
Chicago



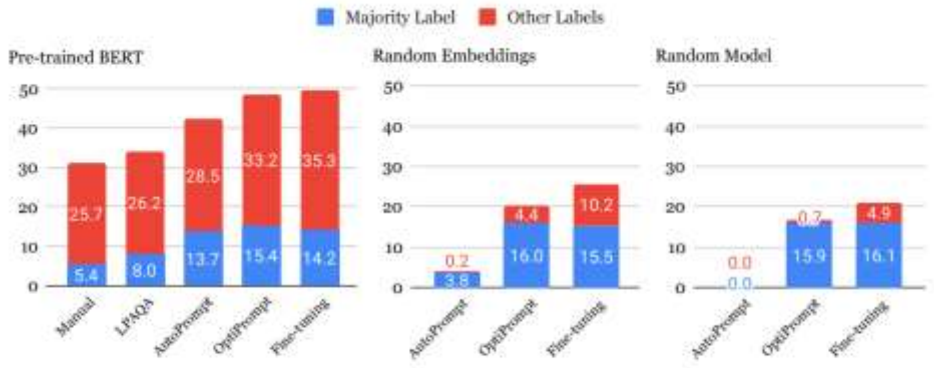
Verbalization stabilities of 4 PLMs.

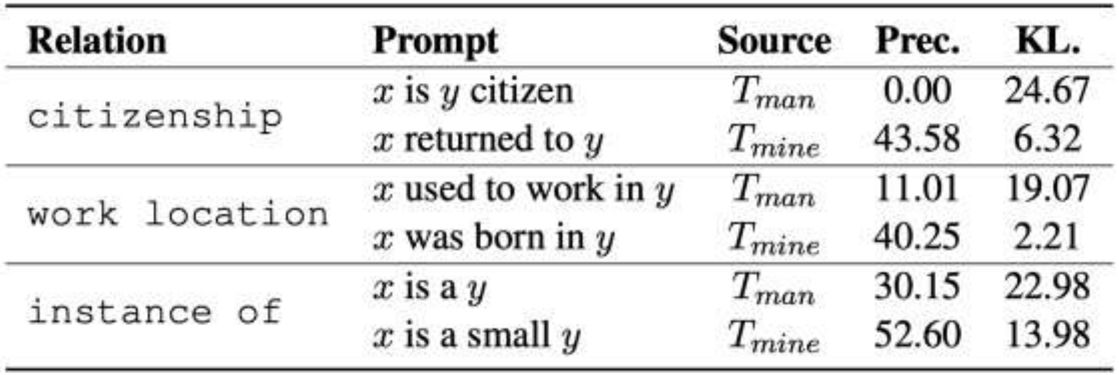
Predictions are sensitive and inconsistent to various verbalizations

Cao et al. Can Prompt Probe Pretrained Language Models? Understanding the Invisible Risks from a Causal View. In ACL 2022.

Prompt-based Knowledge Probing

- Optimized prompt could be unreliable





Relation	Prompt	Source	Prec.	KL.
`citizenship`	x is y citizen	T_{man}	0.00	24.67
	x returned to y	T_{mine}	43.58	6.32
`work location`	x used to work in y	T_{man}	11.01	19.07
	x was born in y	T_{mine}	40.25	2.21
`instance of`	x is a y	T_{man}	30.15	22.98
	x is a small y	T_{mine}	52.60	13.98

- Optimized prompts can exploit patterns in training data
- "Better" prompts may be the prompts fitting the answer distribution better

Zhong et al. Factual Probing Is [MASK]: Learning vs. Learning to Recall. 2021
Cao et al. Knowledgeable or Educated Guess? Revisiting Language Models as Knowledge Bases. 2021

Prompt-based Knowledge Probing

- Q&A based Evaluation for aligned models
  - MMLU, Stanford HELM, OpenLLM, CMMLU, C-Eval…

Questions

Where is the capital of French?
(A) Beijing
(B) Tokyo
(C) Paris
(D) Washington

Answer:

Multiple Choice

Questions

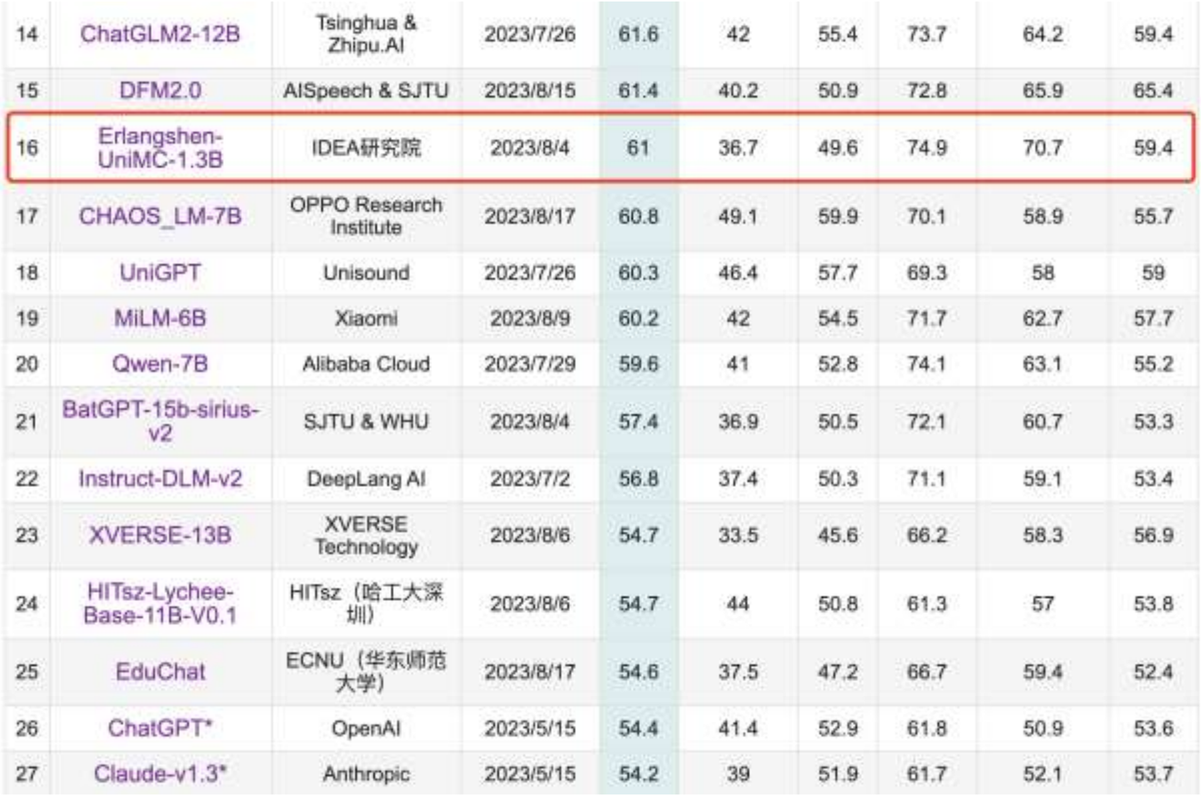
Tell me some trivia about penguins

Free-style Writing

Does correct (wrong) answer means the model has (don't has) the knowledge?

Prompt-based Knowledge Probing

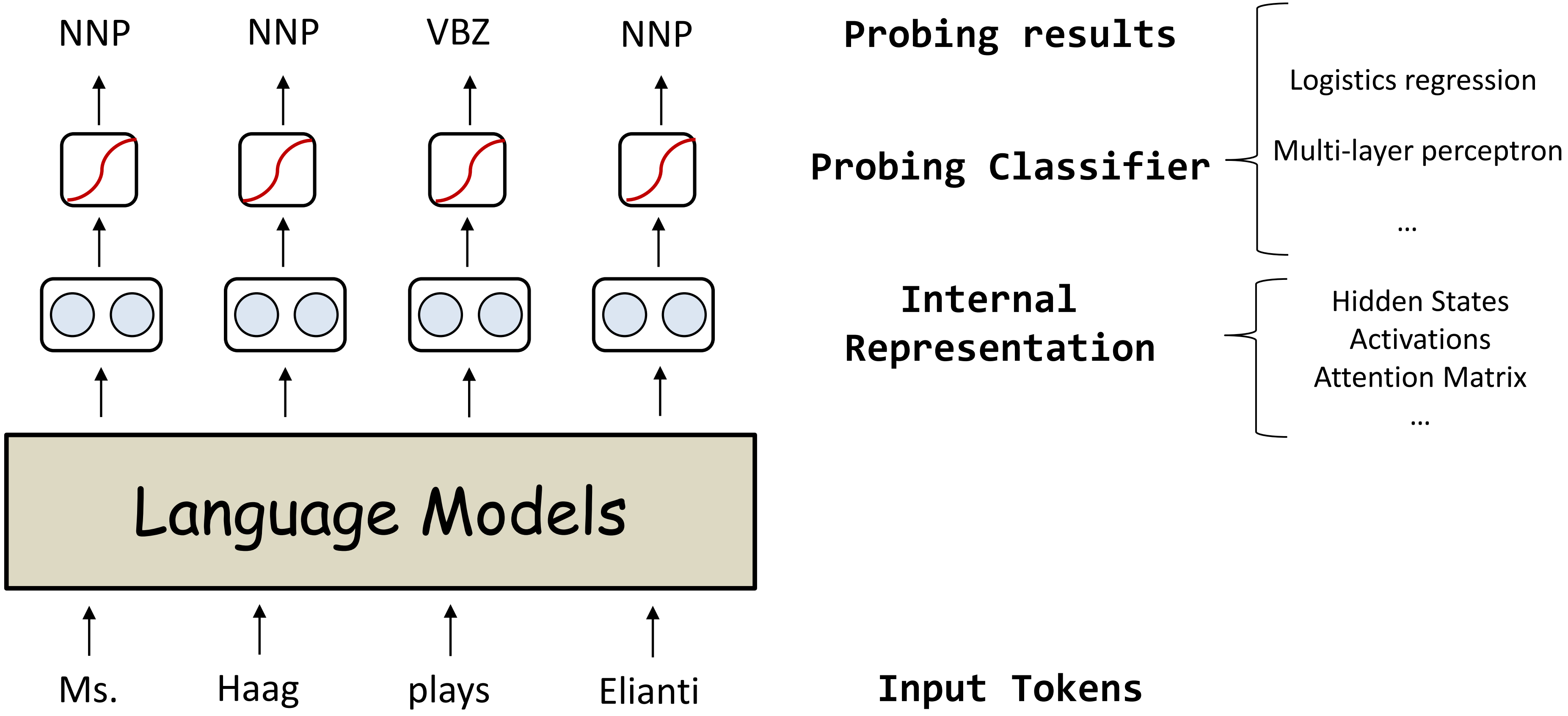
- Q&A based Evaluation for aligned model
  - Erlangshen-UniMC-1.3B achieve strong performance on C-EVAL
  - Pre-trained on 180G text corpus and fine-tuned on multiple choice dataset



14	ChatGLM2-12B	Tsinghua & Zhipu.AI	2023/7/26	61.6	42	55.4	73.7	64.2	59.4
15	DFM2.0	AISpeech & SJTU	2023/8/15	61.4	40.2	50.9	72.8	65.9	65.4
16	Erlangshen-UniMC-1.3B	IDEA研究院	2023/8/4	61	36.7	49.6	74.9	70.7	59.4
17	CHAOS_LM-7B	OPPO Research Institute	2023/8/17	60.8	49.1	59.9	70.1	58.9	55.7
18	UniGPT	Unisound	2023/7/26	60.3	46.4	57.7	69.3	58	59
19	MiLM-6B	Xiaomi	2023/8/9	60.2	42	54.5	71.7	62.7	57.7
20	Qwen-7B	Alibaba Cloud	2023/7/29	59.6	41	52.8	74.1	63.1	55.2
21	BatGPT-15b-sirius-v2	SJTU & WHU	2023/8/4	57.4	36.9	50.5	72.1	60.7	53.3
22	Instruct-DLM-v2	DeepLang AI	2023/7/2	56.8	37.4	50.3	71.1	59.1	53.4
23	XVERSE-13B	XVERSE Technology	2023/8/6	54.7	33.5	45.6	66.2	58.3	56.9
24	HITsz-Lychee-Base-11B-V0.1	HITsz（哈工大深圳）	2023/8/6	54.7	44	50.8	61.3	57	53.8
25	EduChat	ECNU（华东师范大学）	2023/8/17	54.6	37.5	47.2	66.7	59.4	52.4
26	ChatGPT*	OpenAI	2023/5/15	54.4	41.4	52.9	61.8	50.9	53.6
27	Claude-v1.3*	Anthropic	2023/5/15	54.2	39	51.9	61.7	52.1	53.7

Feature-based Knowledge Probing

- Feature based probing with classifier (Lin et al., 2019; Clark et al., 2019; Tenney et al., 2019; Liu et al., 2019;)



Probing results

Probing Classifier

Logistics regression
Multi-layer perceptron
…

Internal Representation

Hidden States
Activations
Attention Matrix
…

Input Tokens

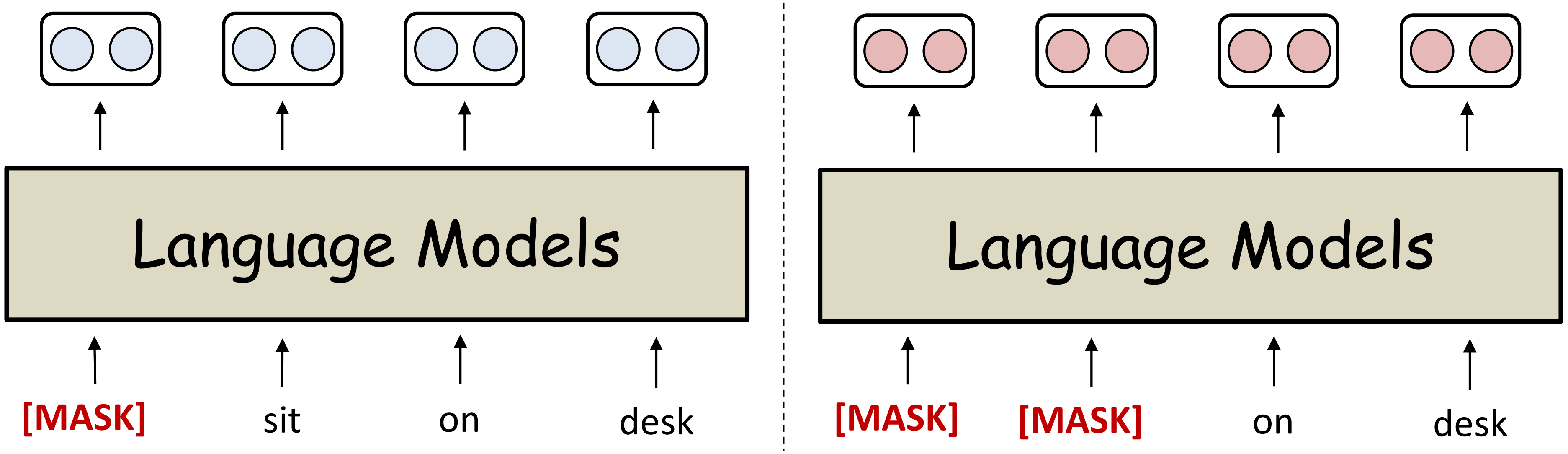
Feature-based Knowledge Probing

- Classifier may be unreliable
  - Training process involved
  - Difficult for results attribution
  - Inconsistent between classifiers

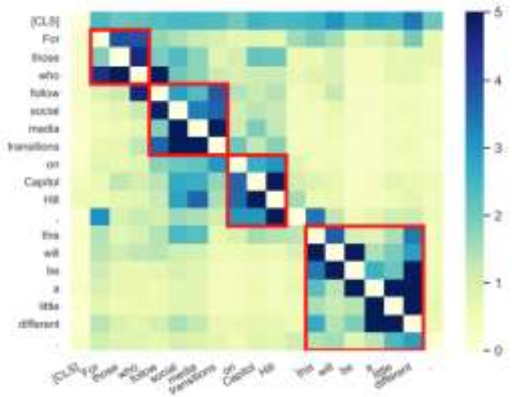
- Can we use feature-based probing without classifier?

Feature-based Knowledge Probing

- Feature-based probing without classifier: example#1 perturbed masking (Wu et al., 2020)

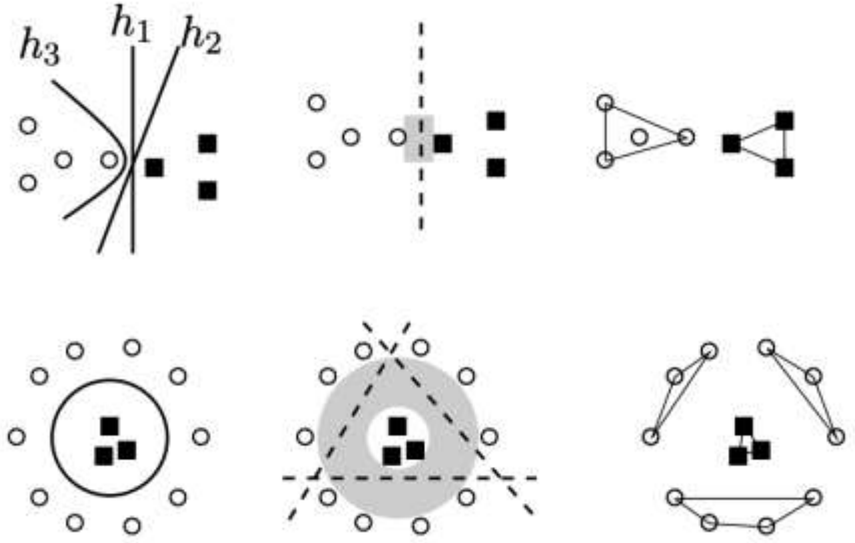


- Perturbed Masking
  - ➢ Calculate impact sit has on Cats
  - ➢ $e_{\mathrm{i}} = \mathrm{E}(\mathrm{Cats}|\mathrm{S}\backslash\{\mathrm{Cats}\})$
  - ➢ $e_j = \mathrm{E}(\mathrm{Cats}|\mathrm{S}\backslash\{\mathrm{Cats}, \mathrm{sit}\})$
  - ➢ $\mathrm{F}(\mathrm{Cats}, \mathrm{sit}) = \mathrm{d}(\mathrm{e_i}, \mathrm{e_j})$



Feature-based Knowledge Probing

- Feature-based probing without classifier: example#2 Direct Probe (Zhou et al., 2021)
  - Each classifier is a decision boundary in the representation space
  - Consider the representation probing as clustering problem



Take-aways for Knowledge Probing

- Both prompt-based and feature-based probing have their own limitations
- Prompt-based evaluation could be biased by data distribution, prompt selections, etc.
- Design of better probing framework for LLMs worth further investigation

Knowledge Editing: Updating and Deleting

Knowledge Editing

- Replacing stored knowledge in PLMs with new knowledge

The prime minister of the UK is **Elizabeth Truss**.

Modify

The prime minister of the UK is **Rishi Sunak**.

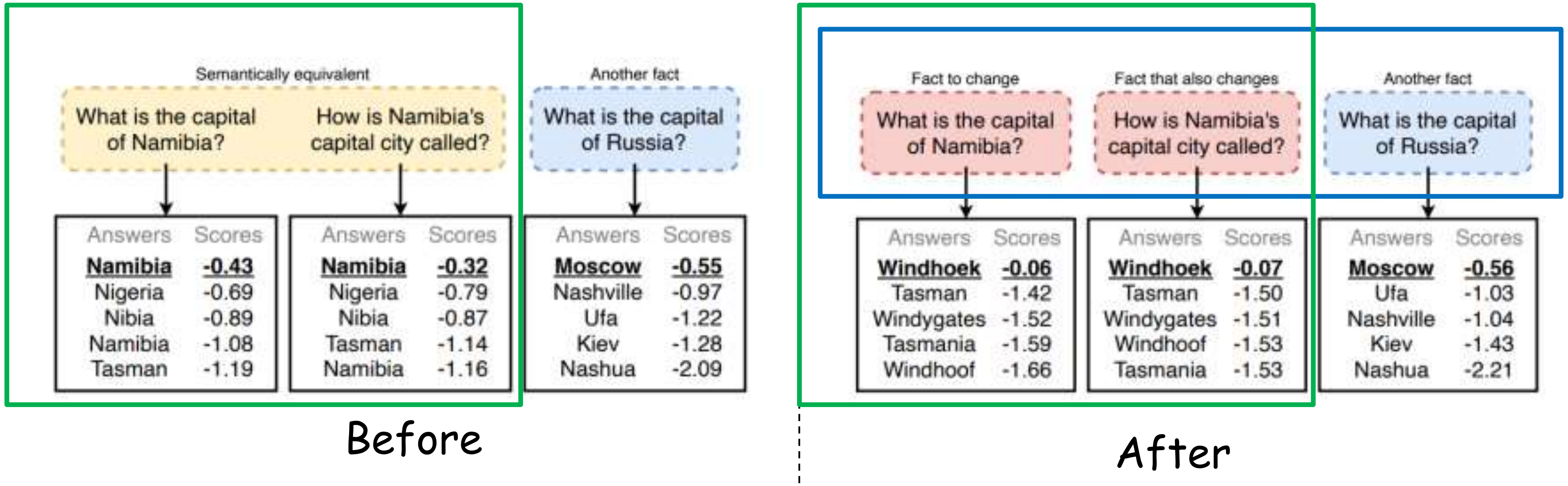
- Removing stored knowledge entirely

The personal number of Tim Cook is **XXXXX**.

Erase

I **don't know** any information about personal number

What is Successful Edit


Semantically equivalent
What is the capital of Namibia?
How is Namibia's capital city called?
Another fact
What is the capital of Russia?
Before
Fact to change
What is the capital of Namibia?
Fact that also changes
How is Namibia's capital city called?
Another fact
What is the capital of Russia?
After

Before

Answers	Scores
Namibia	**-0.43**
Nigeria	-0.69
Nibia	-0.89
Namibia	-1.08
Tasman	-1.19

Answers	Scores
Namibia	**-0.32**
Nigeria	-0.79
Nibia	-0.87
Tasman	-1.14
Namibia	-1.16

Answers	Scores
Moscow	**-0.55**
Nashville	-0.97
Ufa	-1.22
Kiev	-1.28
Nashua	-2.09

After

Answers	Scores
Windhoek	**-0.06**
Tasman	-1.42
Windygates	-1.52
Tasmania	-1.59
Windhoof	-1.66

Answers	Scores
Windhoek	**-0.07**
Tasman	-1.50
Windygates	-1.51
Windhoof	-1.53
Tasmania	-1.53

Answers	Scores
Moscow	**-0.56**
Ufa	-1.03
Nashville	-1.04
Kiev	-1.43
Nashua	-2.21

- Generality:
  - Suitable for general pre-trained language models.
- Reliability:
  - Be able to successful update target knowledge without affecting the rest.
- Consistency:
  - The changes should be consistent across equivalent formulations of a fact

De Cao et al. Editing Factual Knowledge in Language Models. 2021.

Knowledge Editing Strategies

- Constrained tuning
  - Fine-tuning on target knowledge without affecting the rest

- Meta-Learning based editing
  - Learning to update: learning to predict updated parameters

- Memory-based editing
  - Maintain a edit memory and reason over it as needed

- Locate and edit
  - Attribute knowledge to specific neurons and edit them accordingly

Constrained Fine-tuning

- Naive Solution 1: Re-training

- Re-train PLM using the updated training dataset
  - Computationally expensive and impractical when LLMs involved

- Fine-tune PLMs on a small subset which only contains target knowledge
  - Suffer from catastrophic forgetting, and affects the rest knowledge which is not intended to be edited.

Constrained Fine-tuning

- Constraint 1: Learn the new facts while keeping the loss small on unmodified facts

$$\boxed{\text{minimize}_{\theta\in\Theta} \quad \frac{1}{m}\sum_{x\in\mathcal{D}_\mathcal{M}} L(x;\theta)} \quad \text{subject to} \quad \boxed{\frac{1}{n}\sum_{x'\in\mathcal{D}_{\mathcal{F}\setminus\mathcal{S}}} \left(L(x';\theta)-L(x';\theta_0)\right) \le \delta.}$$

Minimize loss on target knowledge

Keep loss small on unrelated knowledge

- Constraint 2: Using normalization to constrain the parameters change of the models.

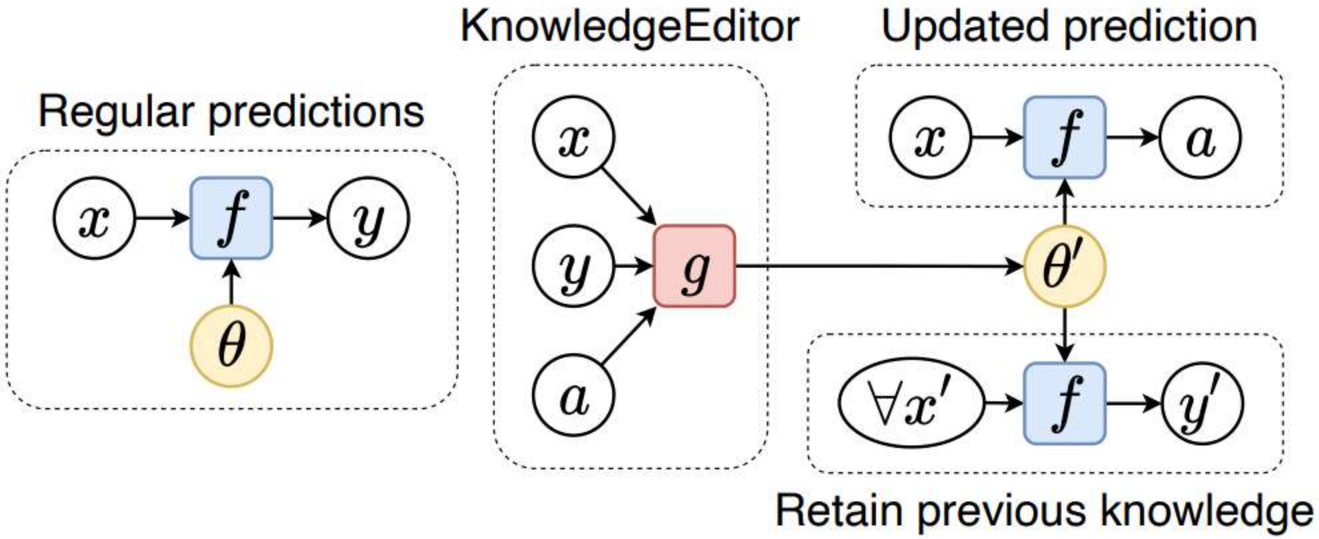
$$\text{minimize}_{\theta\in\Theta} \quad \frac{1}{m}\sum_{x\in\mathcal{D}_\mathcal{M}} L(x;\theta) \quad \boxed{\text{subject to} \quad \|\theta-\theta_0\| \le \delta,}$$

l_2 or l_∞norm

Zhu et al. Modifying Memories in Transformer Models. 2020

Meta-Learning Based Knowledge Editing

- Example #1 - KnowledgeEditor: train a hyper-network to predict the parameter update



Replace the prediction of x from y to a, without affecting the predictions of any other input.

De Cao et al. Editing Factual Knowledge in Language Models. 2021.

Meta-Learning Based Knowledge Editing

- Example #1 - KnowledgeEditor: train a hyper-network to predict the parameter update

semantically equivalent inputs of x

Changing prediction successfully

$$\min_{\phi} \sum_{\hat{x} \in \boxed{\mathcal{P}^x}} \mathcal{L}(\theta'; \hat{x}, a)$$

Not affect the rest

$$\text{s.t.} \quad \boxed{\mathcal{C}(\theta, \theta', f; \mathcal{O}^x)} \leq m$$

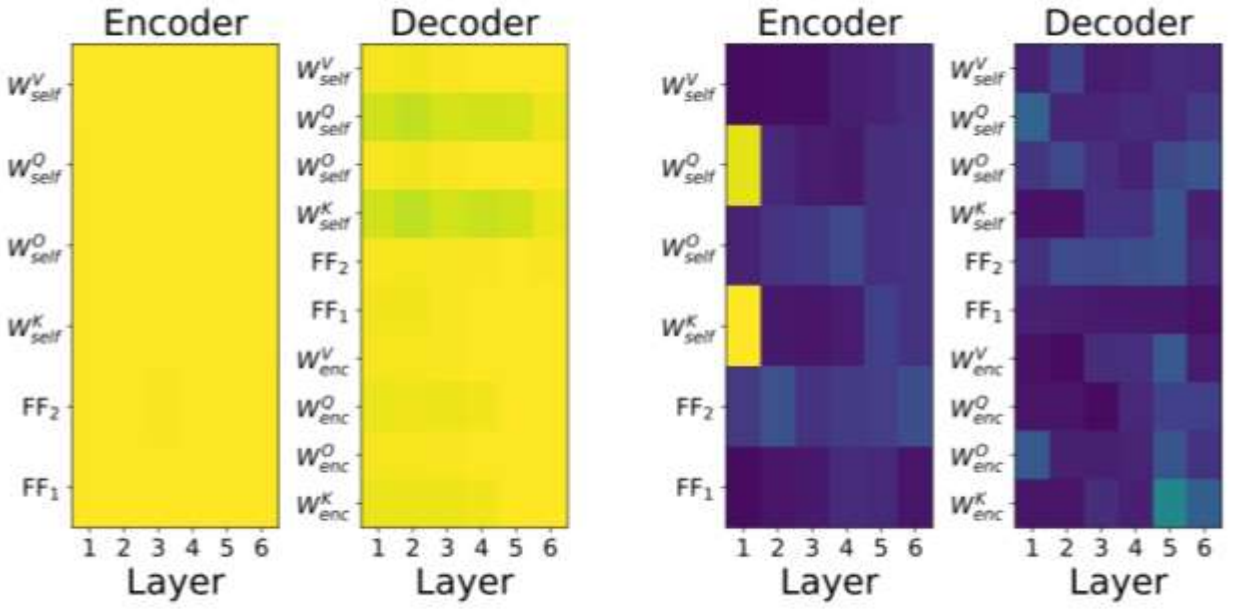
$$\Big\downarrow$$

$$\sum_{x' \in \mathcal{O}^x} \sum_{c \in \mathcal{Y}} p_{Y|X}(c|x', \theta) \log \frac{p_{Y|X}(c|x', \theta)}{p_{Y|X}(c|x', \theta')}$$

De Cao et al. Editing Factual Knowledge in Language Models. 2021.

Meta-Learning Based Knowledge Editing

- Fine-tuning V.S. Hyper-network: fine-tuning updates all layers uniformly while hyper-network updates are more sparse.



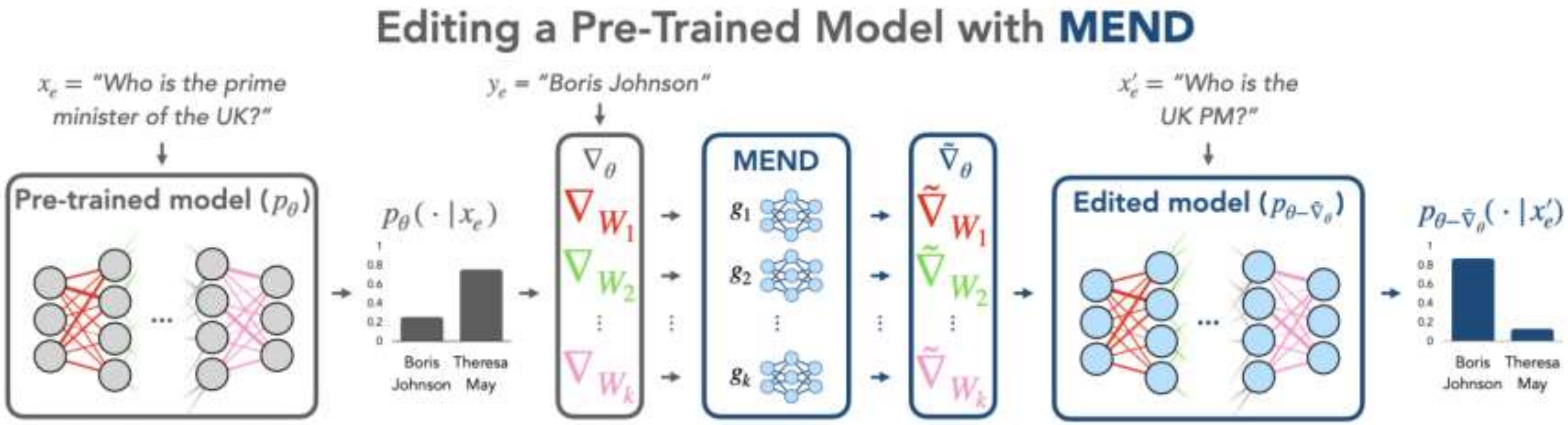
(b) Fine-tune (all layers).

(c) KNOWLEDGEEDITOR + $\mathcal{P}^x$.

De Cao et al. Editing Factual Knowledge in Language Models. 2021.

Meta-Learning Based Knowledge Editing

- Example #2 - MEMD: predict the edits to LMs' weights based on the standard fine-tuning gradient with correction

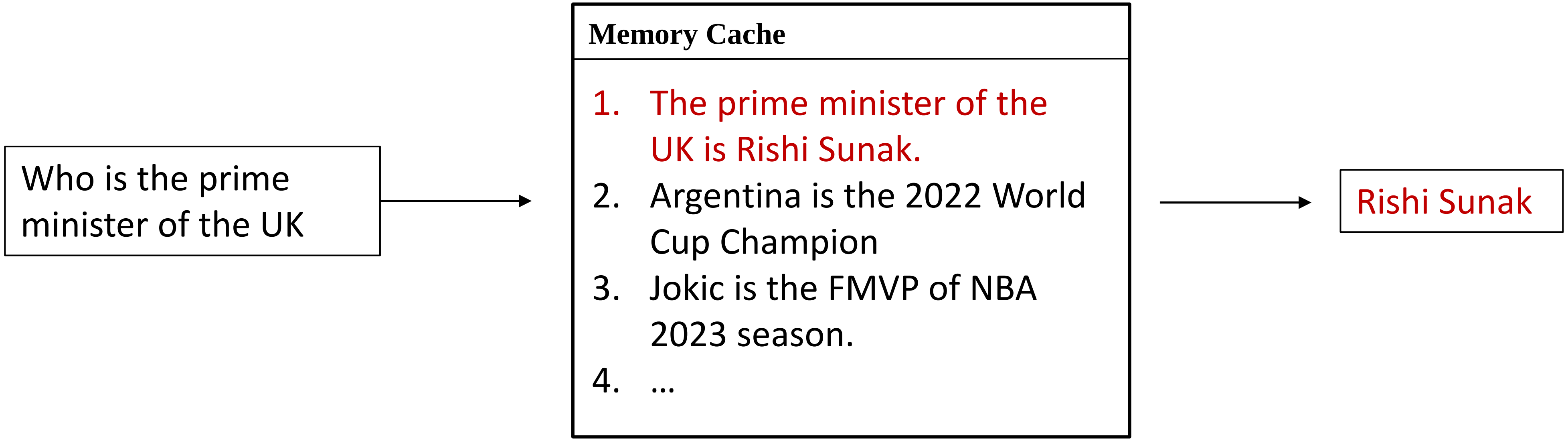


- ➢ Get the prediction of target input.
- ➢ Calculate the standard fine-tuning gradient with correction
- ➢ Predict the updated weights
- ➢ Edit PLMs and check the updated knowledge

Mitchell et al. FAST MODEL EDITING AT SCALE. 2022.

Memory-based Knowledge Editing

- Naïve Solution 2: maintain a symbolic memory cache
  - a symbolic knowledge cache may suffer from robustness issues



Who is the prime minister of the UK

Memory Cache

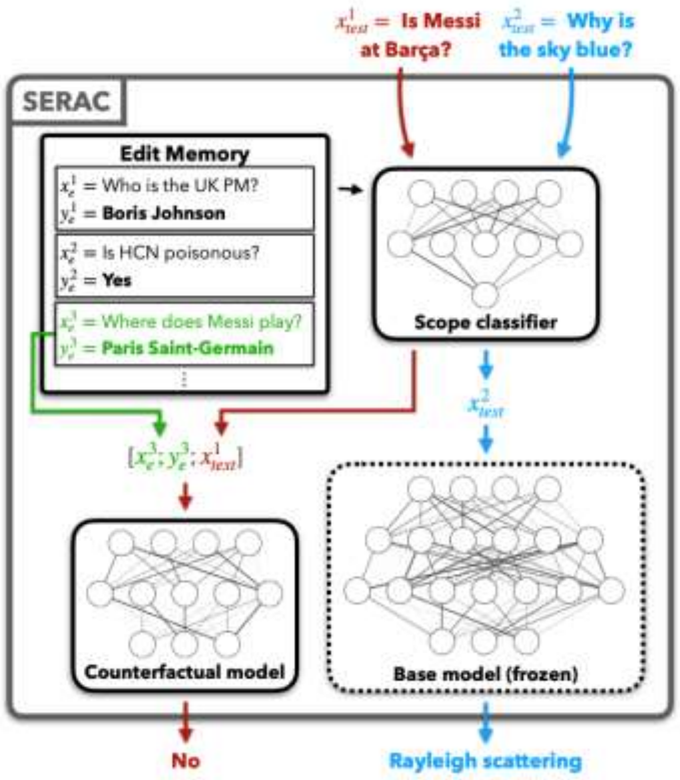
1. The prime minister of the UK is Rishi Sunak.
2. Argentina is the 2022 World Cup Champion
3. Jokic is the FMVP of NBA 2023 season.
4. …

Rishi Sunak

How to distinguish a relevant query?

Memory-based Knowledge Editing

- Example#1 – SERAC: stores edits in a memory and learns to reason over them as needed

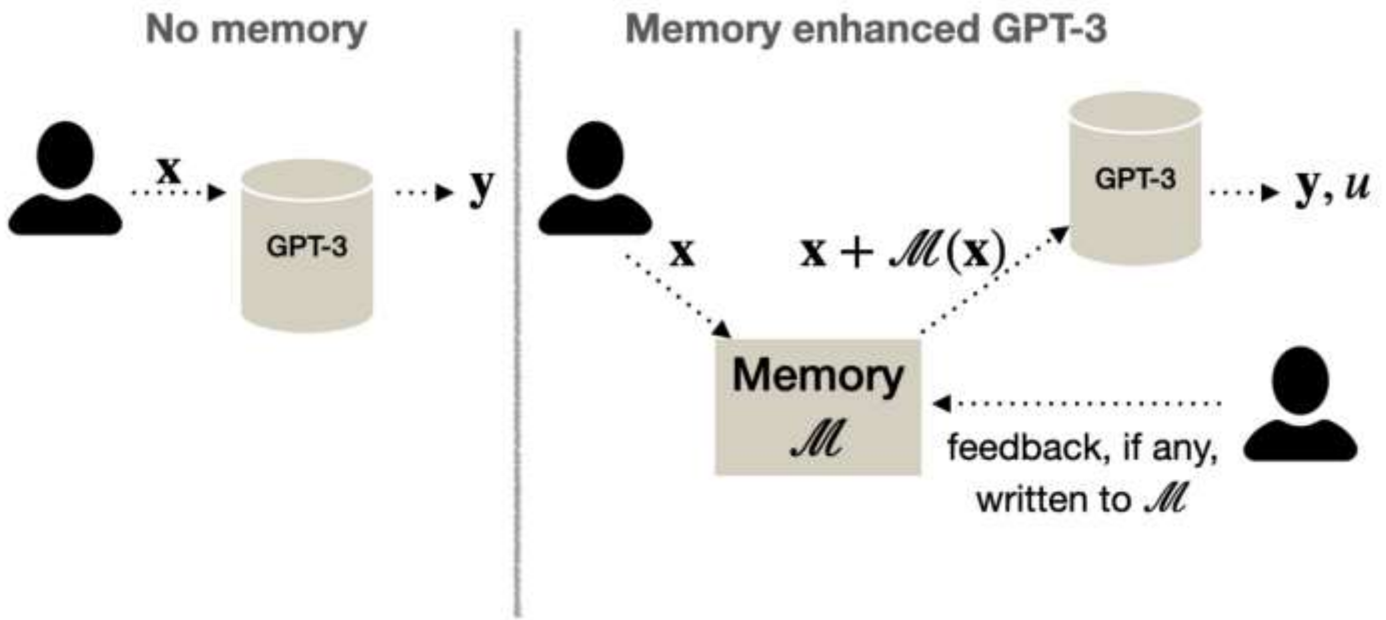

x^1_{test} = Is Messi at Barça?
x^2_{test} = Why is the sky blue?
SERAC
Edit Memory
x_e^1 = Who is the UK PM?
y_e^1 = Boris Johnson
x_e^2 = Is HCN poisonous?
y_e^2 = Yes
x_e^3 = Where does Messi play?
y_e^3 = Paris Saint-Germain
Scope classifier
x^2_{test}
$[x_e^3; y_e^3; x^1_{test}]$
Counterfactual model
Base model (frozen)
No
Rayleigh scattering

- ➢ Step 1: Maintain a edit memory
- ➢ Step 2: Decide whether a relevant edit exists in memory
- ➢ Step 3.1: Irrelevant - Using original LM to predict irrelevant question.
- ➢ Step 3.2: Relevant - Input and edited example are passed to a counterfactual model

Mitchell et al. Memory-Based Model Editing at Scale. 2022.

Memory-based Knowledge Editing

- Example#2 - MemPrompt: directly add edit information to the query


No memory
x
GPT-3
y
Memory enhanced GPT-3
x
$\mathbf{x} + \mathscr{M}(\mathbf{x})$
GPT-3
$\mathbf{y}, u$
Memory
$\mathscr{M}$
feedback, if any, written to $\mathscr{M}$

➢ Maintain a memory of past feedback
➢ Lookup for relevant memory
➢ Directly add to the query

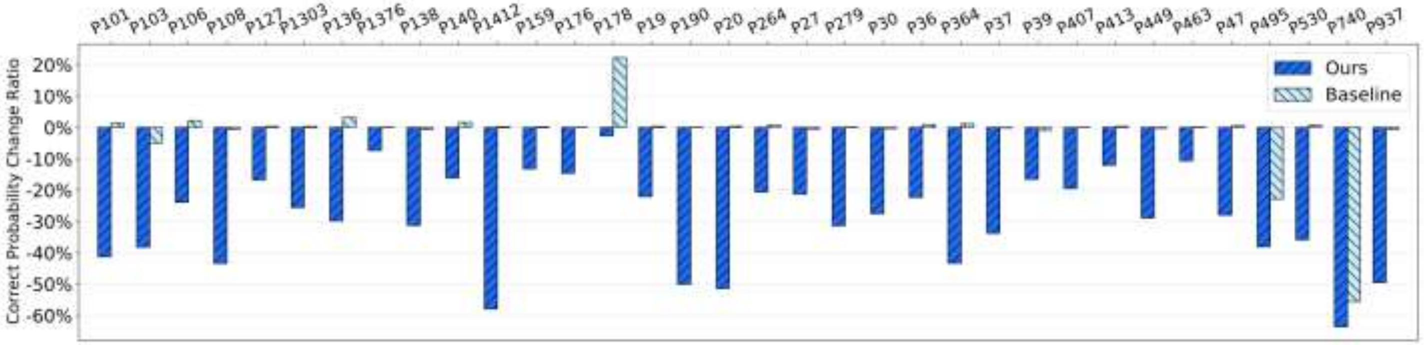
Madaan et al. MemPrompt: Memory-assisted Prompt Editing with User Feedback. 2023.

Locate and Edit

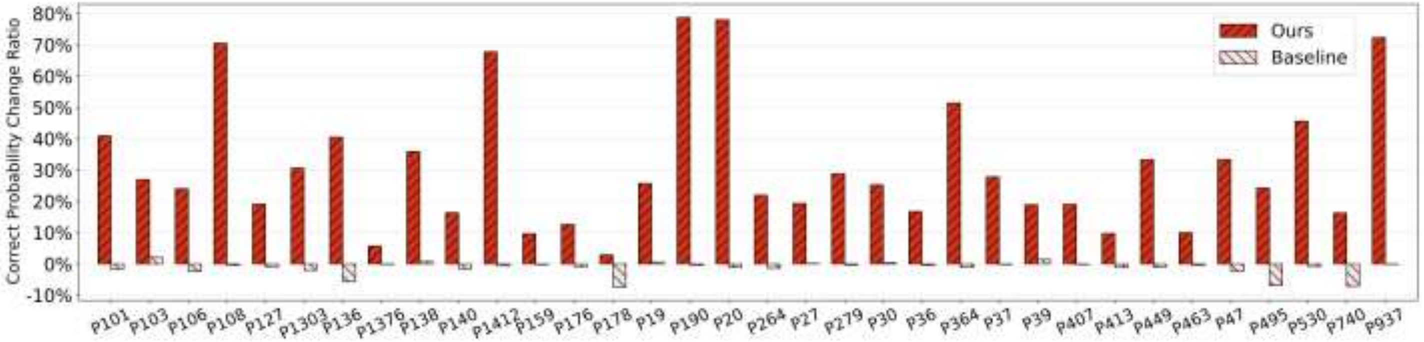
- Combine knowledge attribution and editing
  - Knowledge Attribution: find the responsible neurons for specific knowledge
  - Knowledge Editing: edit the responsible neurons only

Locate and Edit

- Example #1: KnowledgeNeuron (Dai et al., 2022)
  - Direct modify the activations of knowledge neurons



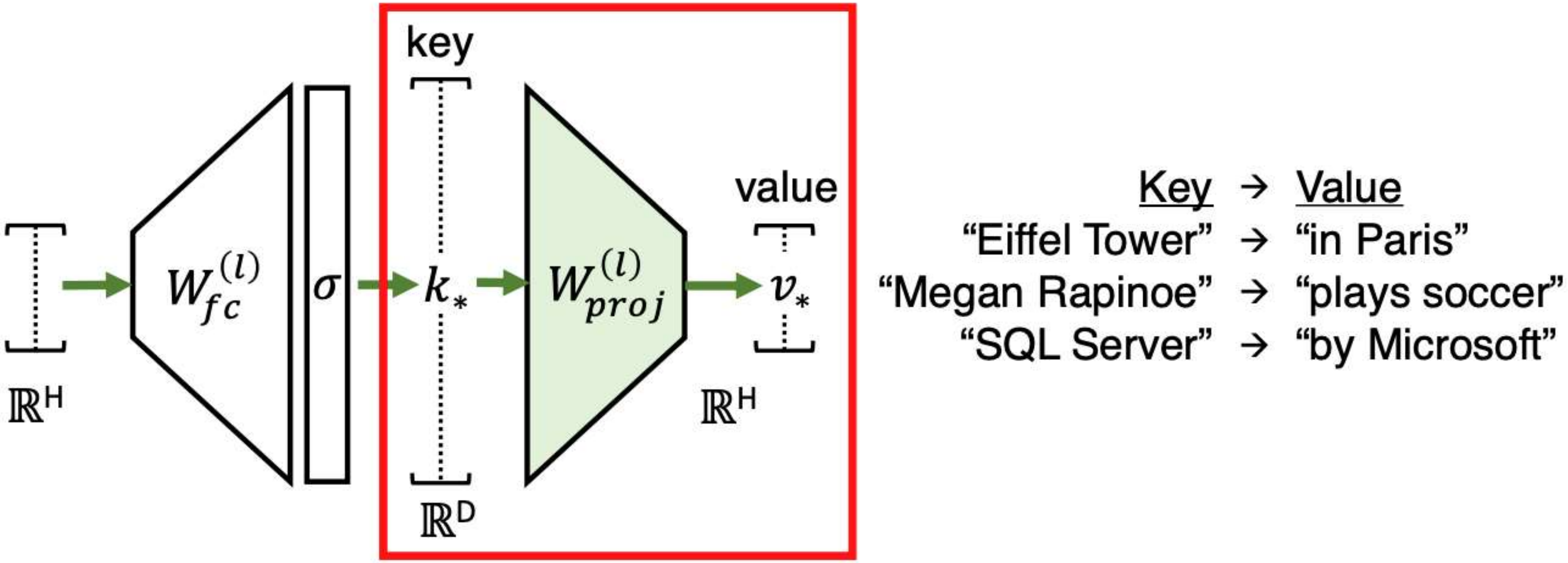
Answer probability decrease: setting activations to 0



Answer probability increase: double activations

Locate and Edit

- Example #2: ROME (Meng et al., 2022)
  - Edit knowledge by updating the MLP weights with rank-one update



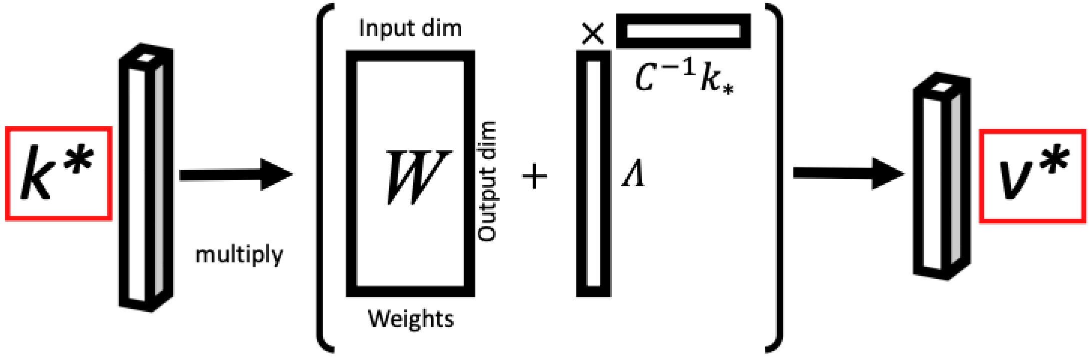
Key → Value
"Eiffel Tower" → "in Paris"
"Megan Rapinoe" → "plays soccer"
"SQL Server" → "by Microsoft"

- ➢ Hypothesize MLPs can be modeled as a linear associative memory
- ➢ Linear operation W stores the key-value mapping information.

Locate and Edit

- Example #2: ROME (Meng et al., 2022)
  - Edit knowledge by updating the MLP weights with rank-one update

$$\text{minimize } \|\hat{W}K - V\| \text{ such that } \hat{W}k_* = v_* \quad \text{by setting } \hat{W} = W + \Lambda(C^{-1}k_*)^T.$$



- Step 1: Choosing k_* to select the Subject
- Step 2: Choosing v_* to recall the Fact
- Step 3: Inserting the act by updated W

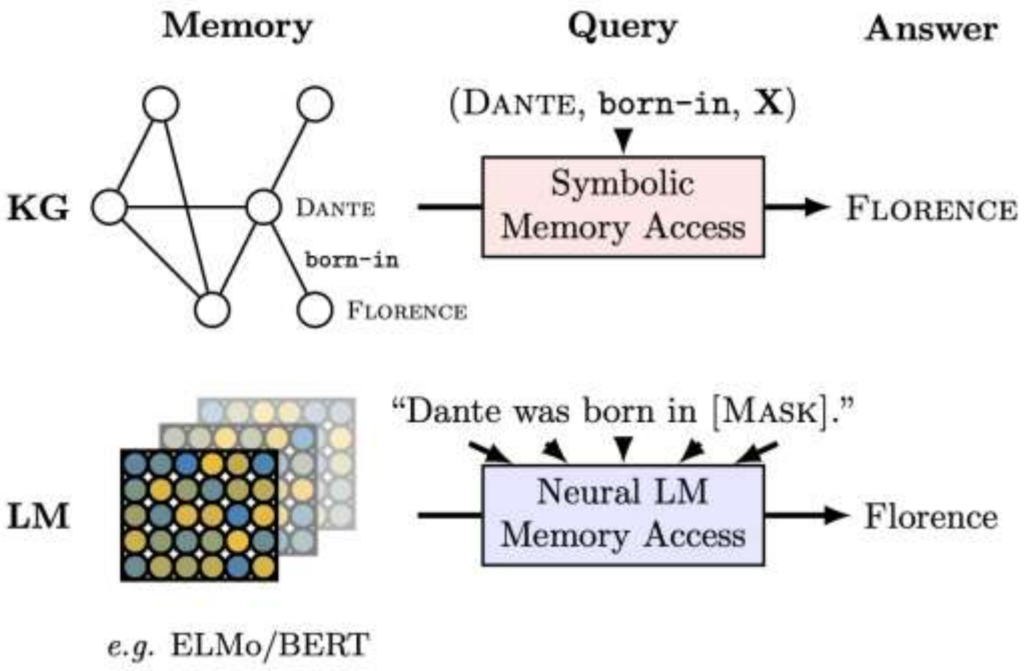
Take-aways for Editing

- Currently most studies only focus on factual knowledge
  - More types of knowledge need to be considered
- More comprehensive evaluation
  - Impact on downstream tasks, related knowledge, etc.
- More effective editing approaches for LLMs

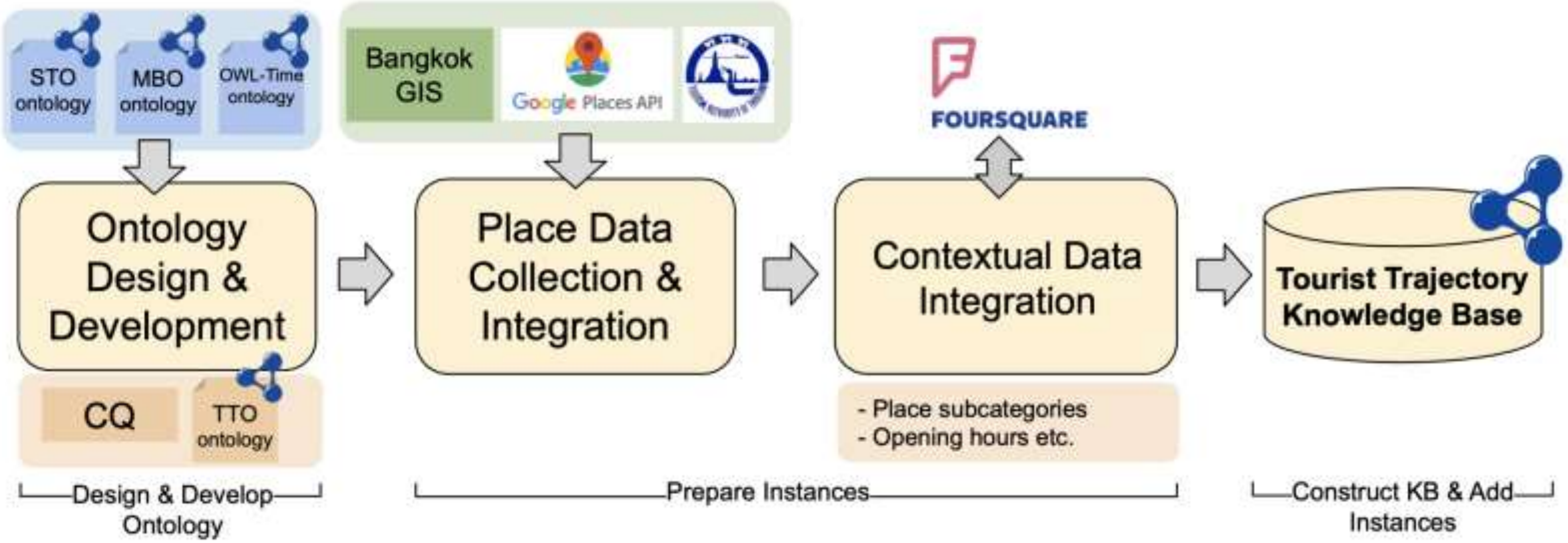
Conclusion: Can LLMs serve as Trustworthy KBs?

LLMs as KBs

- Are LLMs potential substitute for structured KB?



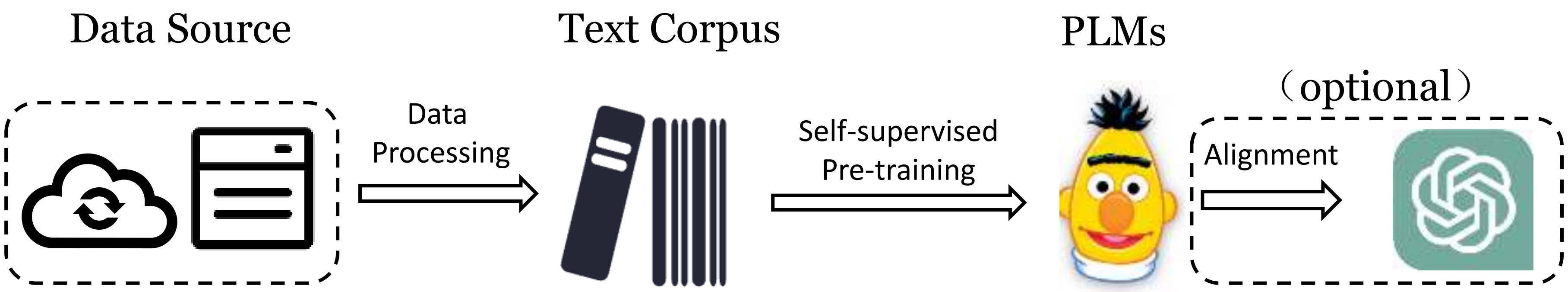
Construct a Structured KB


STO ontology
MBO ontology
OWL-Time ontology
Bangkok GIS
Google Places API
FOURSQUARE
Ontology Design & Development
Place Data Collection & Integration
Contextual Data Integration
Tourist Trajectory Knowledge Base
CQ
TTO ontology
- Place subcategories
- Opening hours etc.
Design & Develop Ontology
Prepare Instances
Construct KB & Add Instances

Structured KB construction process (Krataihong et, al. 2022)

- Requires pre-defined ontology
- Complex pipelines and many traditional NLP techniques involve
- Expert knowledge and human effort for annotation

Construct a LLM-based KB



Language model pre-training process

- Requires no ontology engineering
- End2end self-supervised pre-training + domain-independent SFT
- Much less expert knowledge

Construct a LLM-based KB

- Even more simple solution ……

OpenAI

ChatGPT API

讯飞星火
IFLYTEK SPARK

Claude
API Key

文心一言

minimax

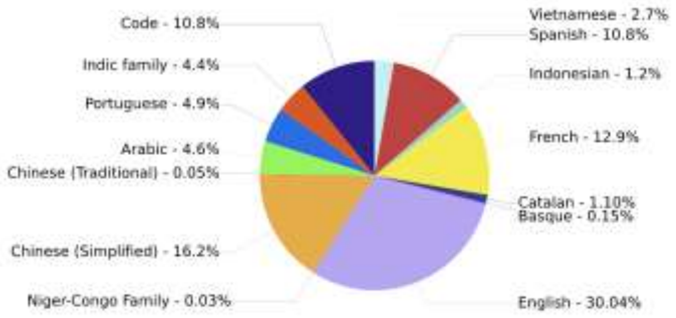
通义千问

MOSS

……..

Knowledge Coverage

	Structured KB	LMs-as-KBs
Domain	Constrained	**Open**
Amount	Limited	Potential?
Schema Expanding	Complex	**Easy**
Knowledge Fusing	Complex	**Easy**



Interaction between Human and KBs

<NBA2015, FMVP, ?>

Query

Structured KB

Andre Iguodala

Deterministic
Answer

SQL, SPARQL, …….

Who is the FMVP in
NBA 2014-2015 season？

Query

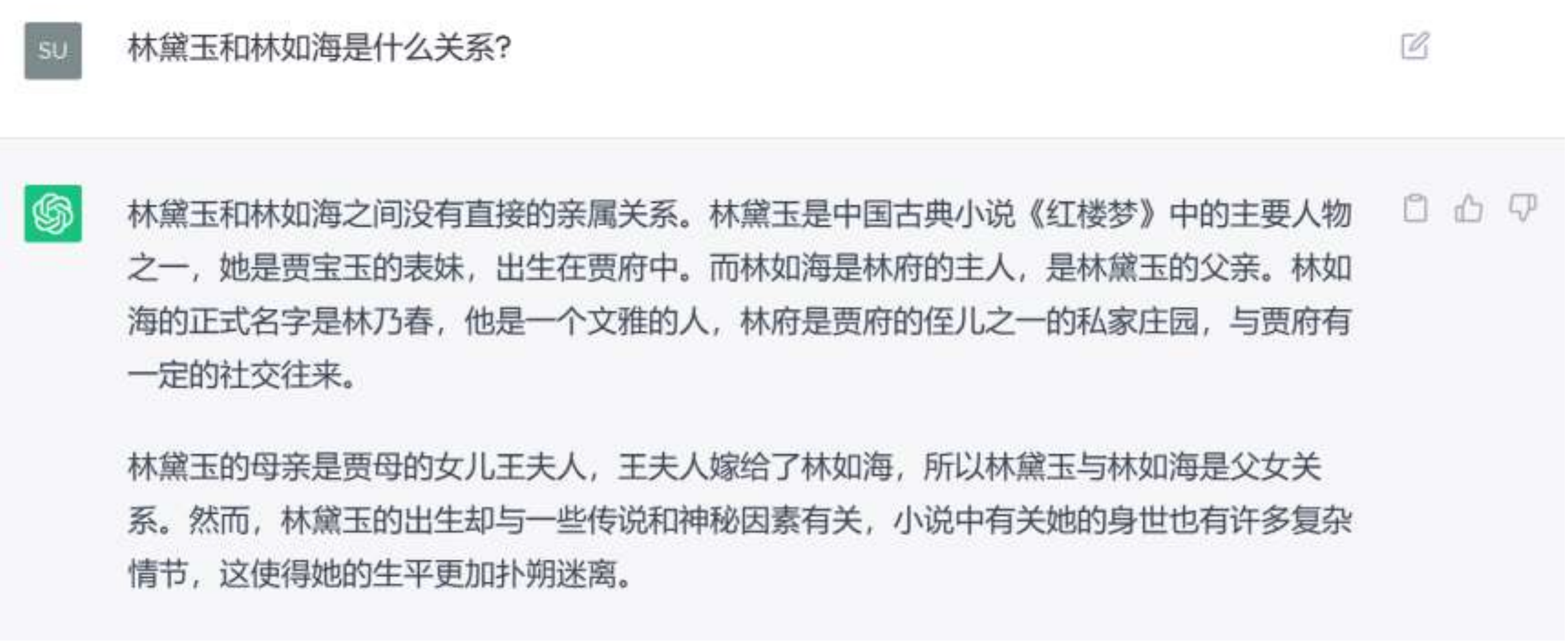
Language Model

Kawhi Leonard
(99% confidence)

Prediction with
Probability

Natural Language

However.......

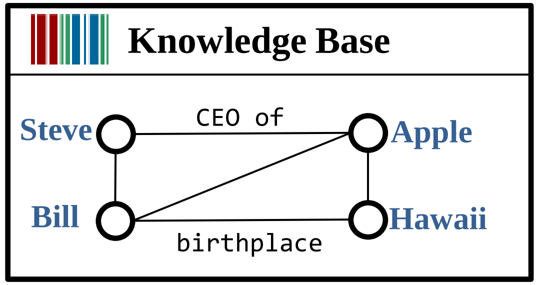


林黛玉和林如海是什么关系?

林黛玉和林如海之间没有直接的亲属关系。林黛玉是中国古典小说《红楼梦》中的主要人物之一，她是贾宝玉的表妹，出生在贾府中。而林如海是林府的主人，是林黛玉的父亲。林如海的正式名字是林乃春，他是一个文雅的人，林府是贾府的侄儿之一的私家庄园，与贾府有一定的社交往来。

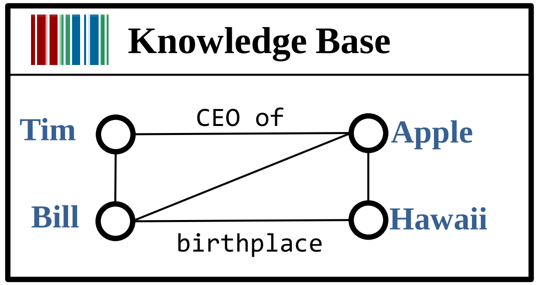
林黛玉的母亲是贾母的女儿王夫人，王夫人嫁给了林如海，所以林黛玉与林如海是父女关系。然而，林黛玉的出生却与一些传说和神秘因素有关，小说中有关她的身世也有许多复杂情节，这使得她的生平更加扑朔迷离。

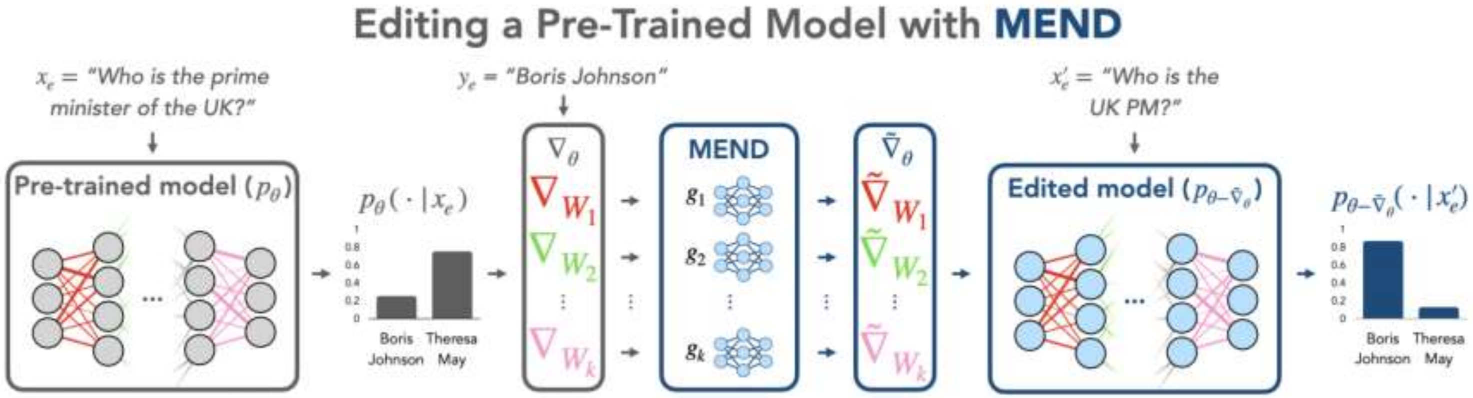
Timely Editing Difficulities



<Steve Jobs, CEO, Apple>

<Tim Cook, CEO, Apple>



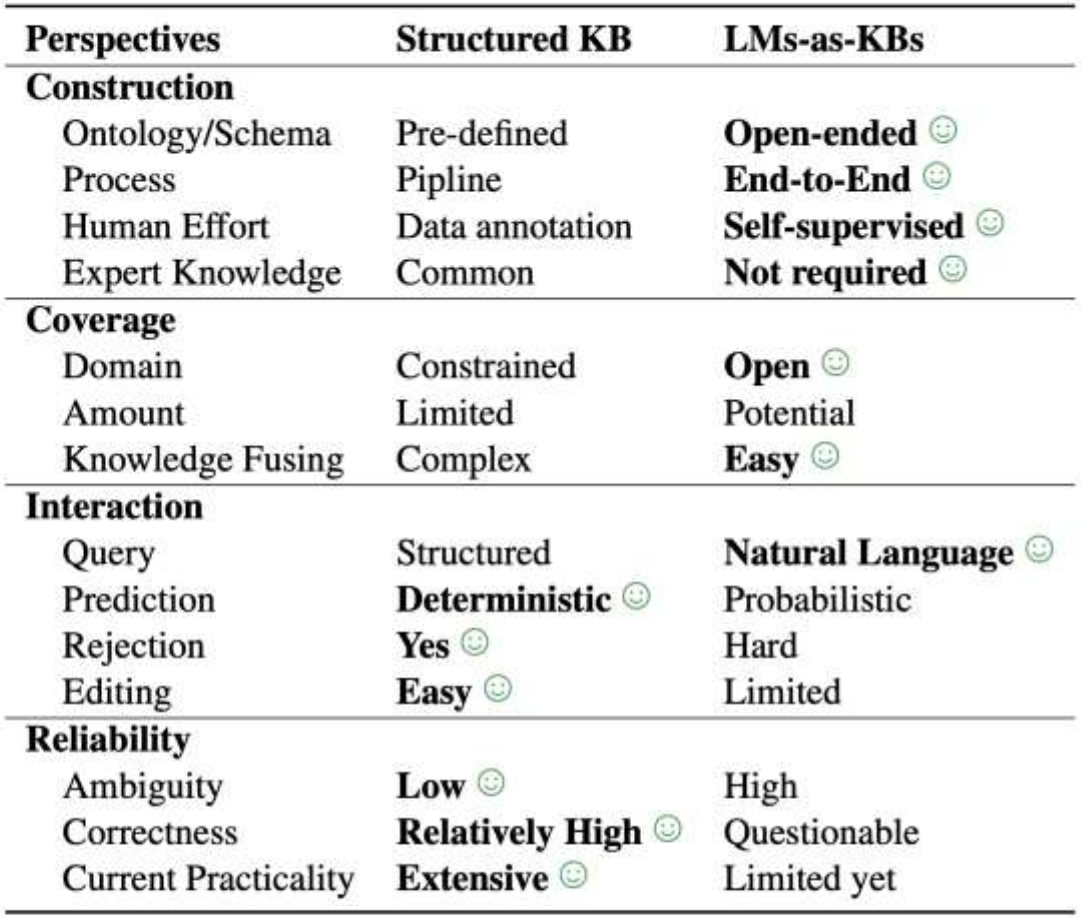


- Compared with LMs, it is easy to add, modify and delete knowledge in structured KBs
- However, editing knowledge in LMs is more complex with potential side effects

Conclusion

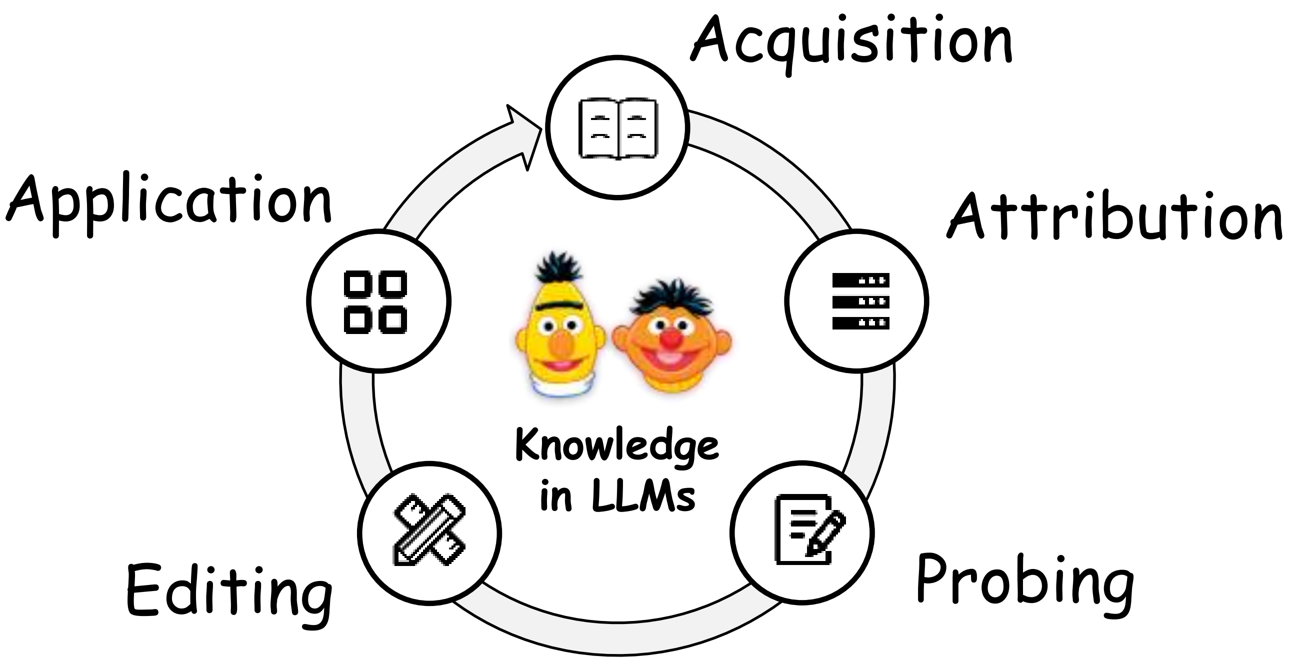
- LLMs have their advantages on simple construction process and its potential knowledge coverage

- Certainty and reliability are the main challenges for LLMs



Perspectives	**Structured KB**	**LMs-as-KBs**
Construction		
Ontology/Schema	Pre-defined	**Open-ended** ☺
Process	Pipline	**End-to-End** ☺
Human Effort	Data annotation	**Self-supervised** ☺
Expert Knowledge	Common	**Not required** ☺
Coverage		
Domain	Constrained	**Open** ☺
Amount	Limited	Potential
Knowledge Fusing	Complex	**Easy** ☺
Interaction		
Query	Structured	**Natural Language** ☺
Prediction	**Deterministic** ☺	Probabilistic
Rejection	**Yes** ☺	Hard
Editing	**Easy** ☺	Limited
Reliability		
Ambiguity	**Low** ☺	High
Correctness	**Relatively High** ☺	Questionable
Current Practicality	**Extensive** ☺	Limited yet

Conclusion



From models of language to models of knowledge, there still a long way to go

Thanks & Any Question?