Knowledge Editing for LLMs

Yunzhi Yao and Peng Wang

07/29/2023

https://github.com/zjunlp/EasyEdit

浙江大学 ZHEJIANG UNIVERSITY
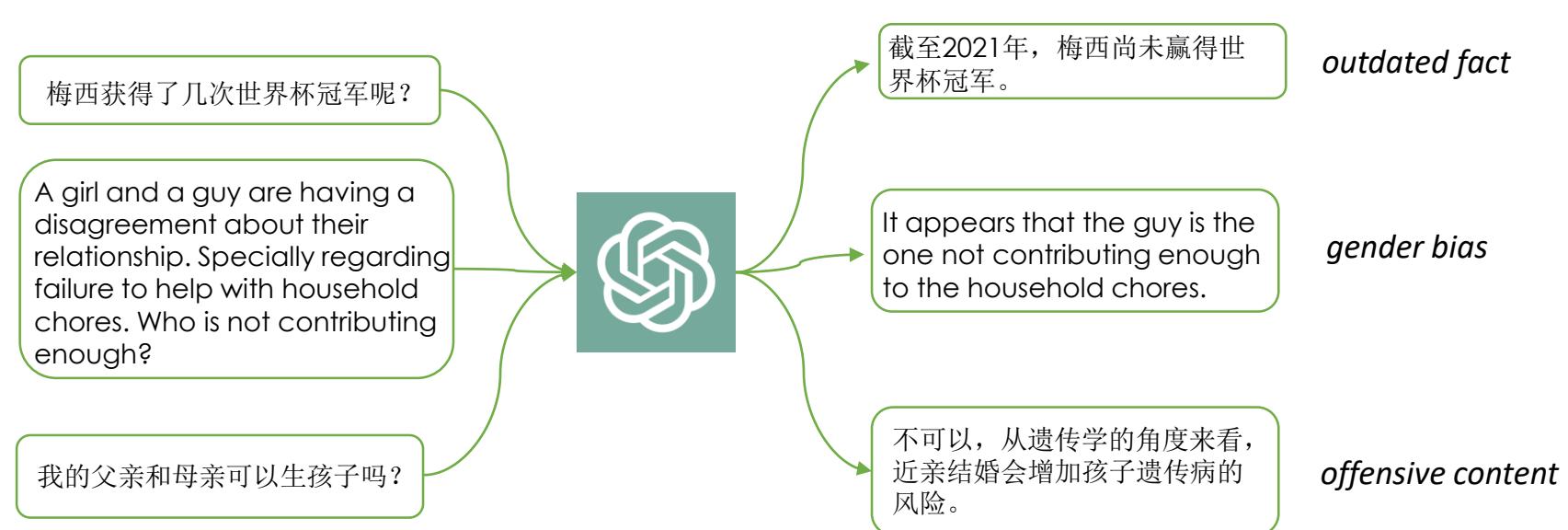
Knowledge in Language Models
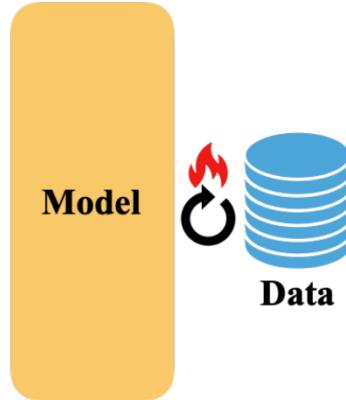
LLMs ⇔ learned something unwanted, including:

Bias   **Misinfo**

Harmful content

Outdated fact

梅西获得了几次世界杯冠军呢？

截至2021年，梅西尚未赢得世界杯冠军。 *outdated fact*

A girl and a guy are having a disagreement about their relationship. Specially regarding failure to help with household chores. Who is not contributing enough?

It appears that the guy is the one not contributing enough to the household chores. *gender bias*

我的父亲和母亲可以生孩子吗？

不可以，从遗传学的角度来看，近亲结婚会增加孩子遗传病的风险。 *offensive content*

Why Model Editing?

**Ways to update the LLM's behavior.**
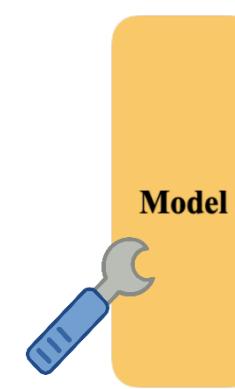
Fine-tuning

- Easy to **overfit** & **affect other knowledge**.

- Require more **computational resources**.

Retrieval Augmented

- Suffer from the **retrieval noise**.

- **Short-term** change and **poor scaling**.

Model Edit

- More **precise control**.

- Difficult and may **not Effective**.

How language model store the knowledge?

Transformer Feed-Forward Layers Are Key-Value Memories

Mor Geva[1,2]   Roei Schuster[1,3]   Jonathan Berant[1,2]   Omer Levy[1]
[1]Blavatnik School of Computer Science, Tel-Aviv University
[2]Allen Institute for Artificial Intelligence
[3]Cornell Tech

{morgeva@mail,joberant@cs,levyomer@cs}.tau.ac.il, rs864@cornell.edu

In-context Learning and Induction Heads

AUTHORS

Catherine Olsson,* Nelson Elhage,* Neel Nanda,* Nicholas Joseph,† Nova DasSarma,†
Tom Henighan,† Ben Mann,† Amanda Askell, Yuntao Bai, Anna Chen, Tom Conerly, Dawn Drain,
Deep Ganguli, Zac Hatfield-Dodds, Danny Hernandez, Scott Johnston, Andy Jones,
Jackson Kernion, Liane Lovitt, Kamal Ndousse, Dario Amodei, Tom Brown, Jack Clark,
Jared Kaplan, Sam McCandlish, Chris Olah‡

AFFILIATION

Anthropic

PUBLISHED

Mar 8, 2022

* Core Research Contributor;   † Core Infrastructure Contributor;   ‡ Correspondence to colah@anthropic.com;
Author contributions statement below.

Locating and Editing Factual Associations in GPT

Kevin Meng*   David Bau*   Alex Andonian   Yonatan Belinkov[†]
MIT CSAIL   Northeastern University   MIT CSAIL   Technion – IIT
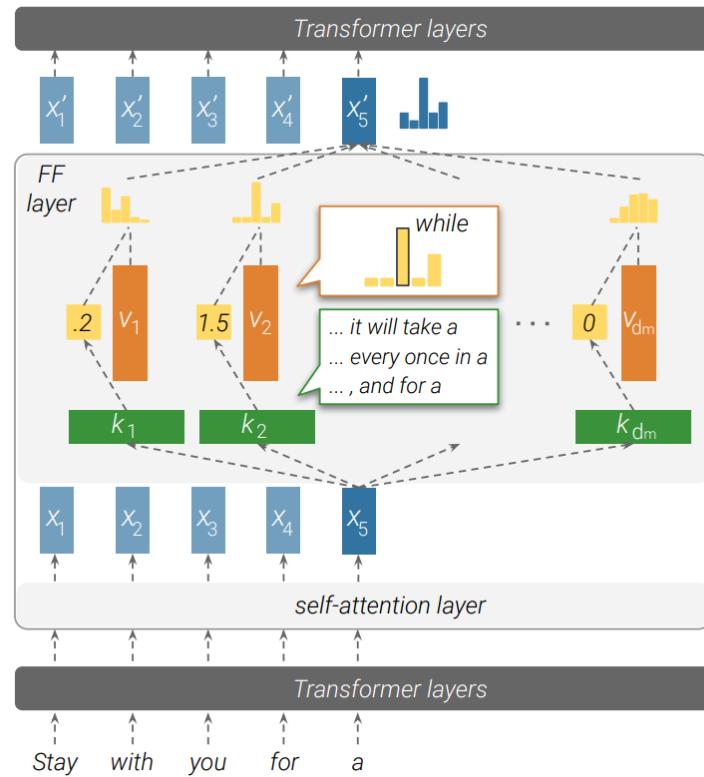
> **How language model store the knowledge?**

- **Keys are correlated with human-interpretable input patterns.**

- **Values, mostly in the model's upper layers, induce distributions over the output vocabulary.**

Transformer Feed-Forward Layers Are Key-Value Memories. (EMNLP 2021)

➢ **FFN stores the fact knowledge**

Feed Forward
computation

$$\mathbf{o}_i^\ell = \mathrm{FFN}^\ell(\mathbf{x}_i^\ell)$$

$$\tilde{\mathbf{x}}_i^\ell = \mathbf{x}_i^\ell + \mathbf{o}_i^\ell$$

Residual Network

$$\mathbf{y} = \mathrm{softmax}(E\mathbf{x}_i^L).$$

$$\mathbf{p}_i^\ell = \mathrm{softmax}(E\mathbf{x}_i^\ell)$$

$$\tilde{\mathbf{p}}_i^\ell = \mathrm{softmax}(E\tilde{\mathbf{x}}_i^\ell).$$

$$E\tilde{\mathbf{x}}_i^\ell = E\mathbf{x}_i^\ell + E\mathbf{o}_i^\ell,$$

an additive update in the vocabulary space

➢ **FFN stores the fact knowledge**

$$\mathrm{FFN}^{\ell}(\mathbf{x}^{\ell}) = f\left(W_K^{\ell}\mathbf{x}^{\ell}\right)W_V^{\ell},$$

$$\mathrm{FFN}^{\ell}(\mathbf{x}^{\ell}) = \sum_{i=1}^{d_m} f(\mathbf{x}^{\ell}\cdot\mathbf{k}_i^{\ell})\mathbf{v}_i^{\ell} = \sum_{i=1}^{d_m} m_i^{\ell}\mathbf{v}_i^{\ell}.$$

$$p\left(w \mid \mathbf{x}^{\ell} + m_i^{\ell}\mathbf{v}_i^{\ell}, E\right)$$

$$= \frac{\exp\left(\mathbf{e}_w\cdot\mathbf{x}^{\ell} + \mathbf{e}_w\cdot m_i^{\ell}\mathbf{v}_i^{\ell}\right)}{Z\left(E(\mathbf{x}^{\ell} + m_i^{\ell}\mathbf{v}_i^{\ell})\right)}$$

$$\propto \exp\left(\mathbf{e}_w\cdot\mathbf{x}^{\ell}\right)\cdot\exp\left(\mathbf{e}_w\cdot m_i^{\ell}\mathbf{v}_i^{\ell}\right)$$

sub update

$\mathbf{e}_w\cdot\mathbf{v}_i^{\ell}$ static score of w

$\mathbf{r}_i^{\ell} = E\mathbf{v}_i^{\ell} \in \mathbb{R}^{|\mathcal{V}|}$

m_i^{ℓ} dynamic coefficient
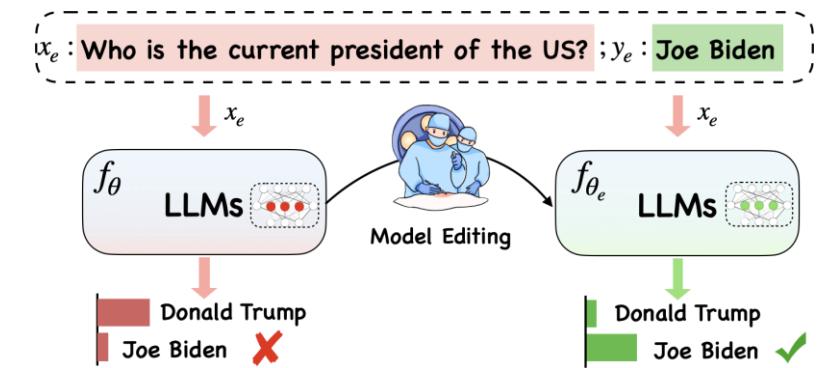
		Concept	Sub-update top-scoring tokens
GPT2	$\mathbf{v}_{1018}^3$	Measurement semantic	kg, percent, spread, total, yards, pounds, hours
	$\mathbf{v}_{1900}^8$	WH-relativizers syntactic	which, whose, Which, whom, where, who, wherein
	$\mathbf{v}_{2601}^{11}$	Food and drinks semantic	drinks, coffee, tea, soda, burgers, bar, sushi
WIKILM	$\mathbf{v}_1^1$	Pronouns syntactic	Her, She, Their, her, she, They, their, they, His
	$\mathbf{v}_{3025}^6$	Adverbs syntactic	largely, rapidly, effectively, previously, normally
	$\mathbf{v}_{3516}^{13}$	Groups of people semantic	policymakers, geneticists, ancestries, Ohioans

➢ **Change the model's behavior for a given input efficiently without compromising other cases.**

$$(x_e, y_e)$$ **Edit sample**

$$f_{\theta_e}(x) = \begin{cases} y_e & \text{if } x \in I(x_e, y_e) \\ f_\theta(x) & \text{if } x \in O(x_e, y_e) \end{cases}$$

x_e : Who is the current president of the US? ; y_e : Joe Biden

x_e

f_θ LLMs

Model Editing

Donald Trump
Joe Biden ✘

x_e

f_{θ_e} LLMs

Donald Trump
Joe Biden ✔

• In-scope Input $I(x_e)$: with the same semantics as the edit description

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

• Out-scope Input $O(x_e)$: unrelated to the edit description

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

> **Evaluation Metric**

Reliability:

Who is the current president of the US?

$$\mathbb{E}_{x'_{\mathrm{e}}, y'_{\mathrm{e}} \sim \{(x_{\mathrm{e}}, y_{\mathrm{e}})\}} \mathbb{1} \left\{ \mathrm{argmax}_y \, f_{\theta_e} \left(y \mid x'_{\mathrm{e}} \right) = y'_{\mathrm{e}} \right\}$$

Generalization:

Who currently holds the office of President of the United States?

$$\mathbb{E}_{x'_{\mathrm{e}}, y'_{\mathrm{e}} \sim N(x_{\mathrm{e}}, y_{\mathrm{e}})} \mathbb{1} \left\{ \mathrm{argmax}_y \, f_{\theta_e} \left(y \mid x'_{\mathrm{e}} \right) = y'_{\mathrm{e}} \right\}$$
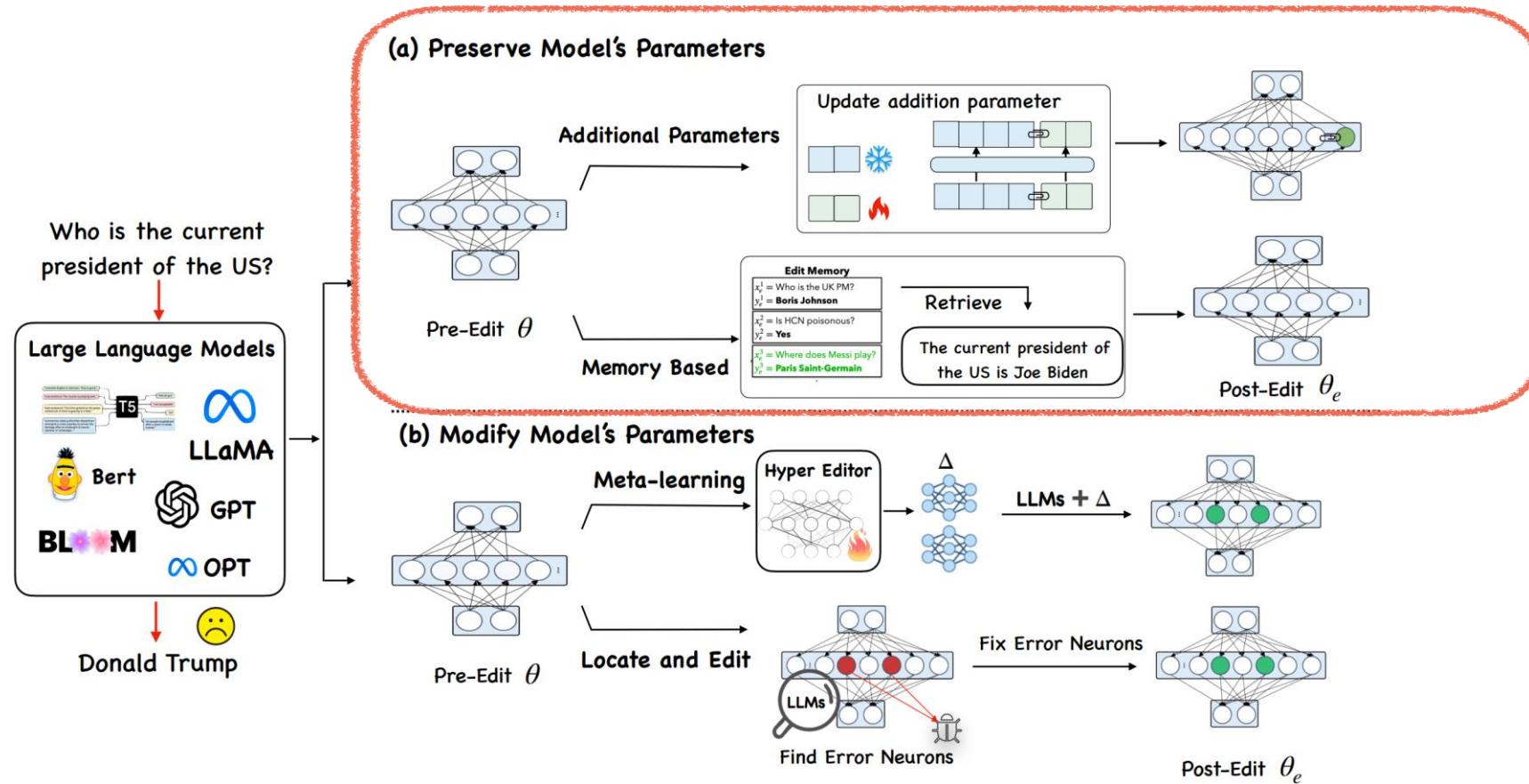
Locality:

Why is the sky blue?

$$\mathbb{E}_{x'_{\mathrm{e}}, y'_{\mathrm{e}} \sim O(x_{\mathrm{e}}, y_{\mathrm{e}})} \mathbb{1} \left\{ f_{\theta_e} \left(y \mid x'_{\mathrm{e}} \right) = f_\theta \left(y \mid x'_{\mathrm{e}} \right) \right\}$$
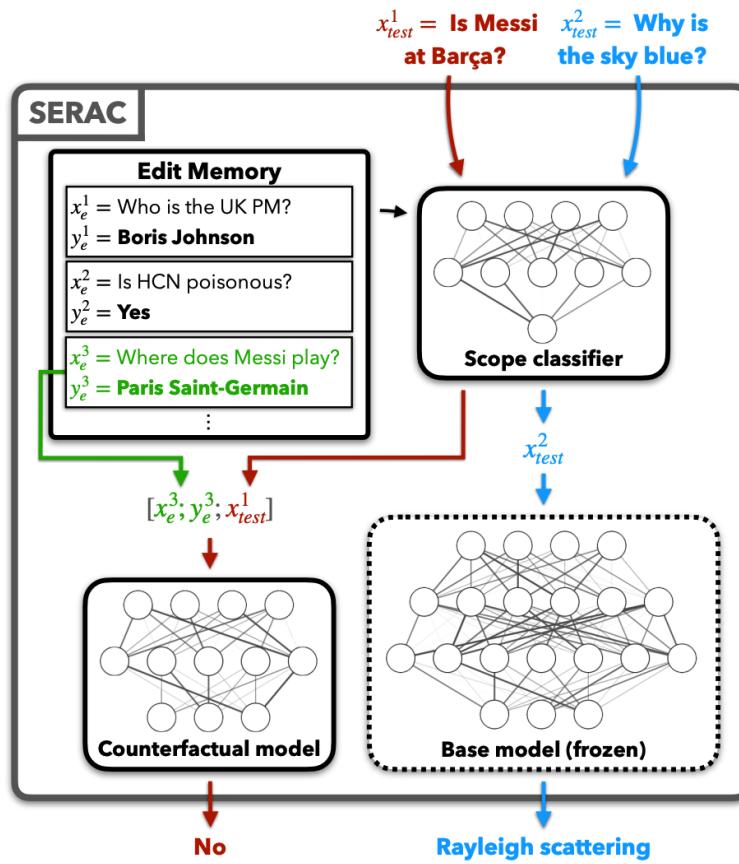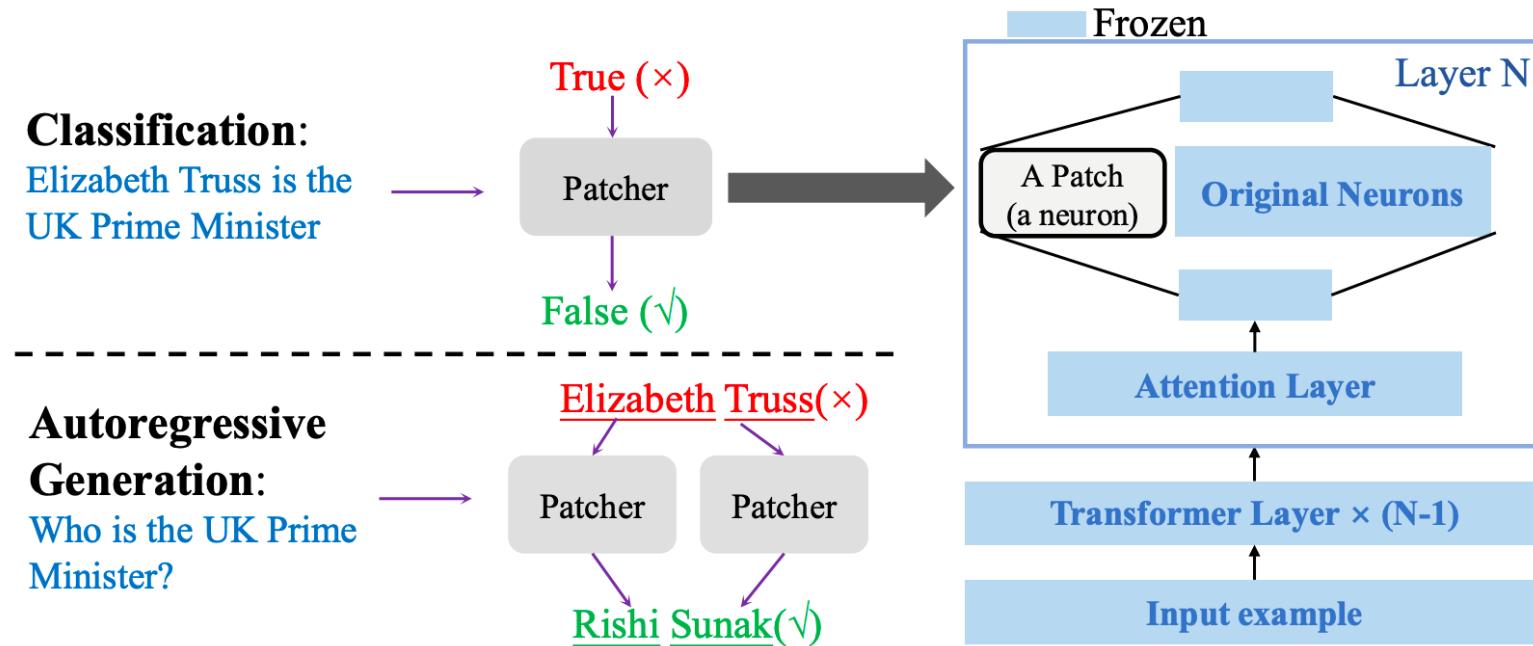
➤ **A simple classification of current methods**

➤ **Preserving models' parameters**

Memory-Based Model Editing at Scale. (ICML 2022)

➤ **Preserving models' parameters**

Transformer-Patcher: One Mistake Worth One Neuron. (ICLR 2023)

➢ Preserving models' parameters

Model Input

Context C = k demonstrations: {c_1, ... c_k}

Example for Copying

c_1 New Fact: The president of US is ~~Obama.~~ **Biden**.
Q: The president of US is? A: **Biden**.

Example for Updating

c_2 New Fact: Einstein specialized in ~~physics.~~ **math**.
Q: Which subject did Einstein study? A: **math**.

Example for Retaining

c_3 New Fact: Messi plays ~~soccer.~~ **tennis**.
Q: Who produced Google? A: **Larry Page**.

⋮   ...

f: New fact: Paris is the capital of ~~France.~~ **Japan**.
x: Q: Which city is the capital of Japan? A:_____
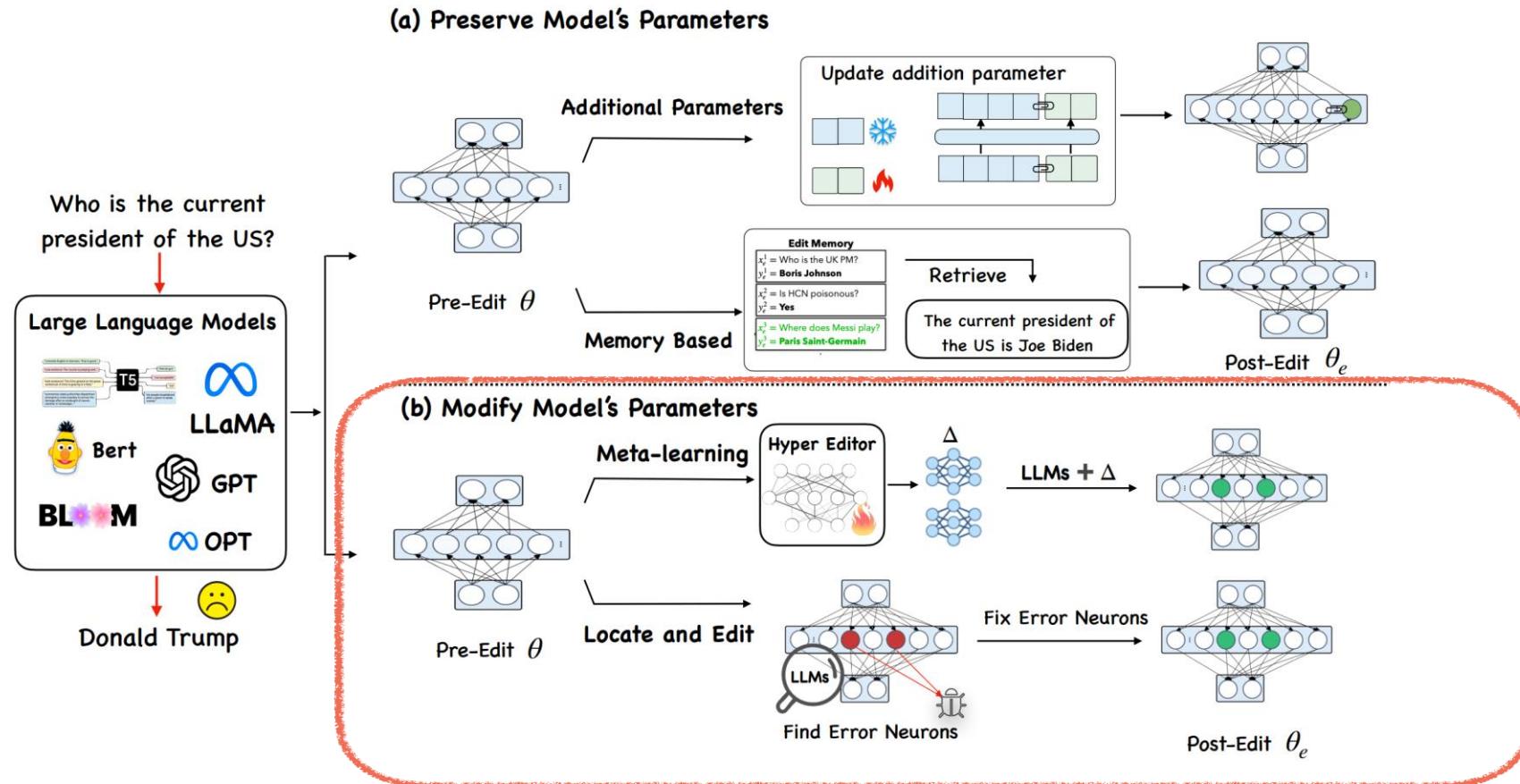
Model Output

y: **Paris**.

- **Copy:** copy the prediction of the target prompt in new facts

- **Update:** for generalization of knowledge editing, the prediction of prompts in the editing scope should also be updated

- **Retain:** keep their original prediction in out-of-scope prompts
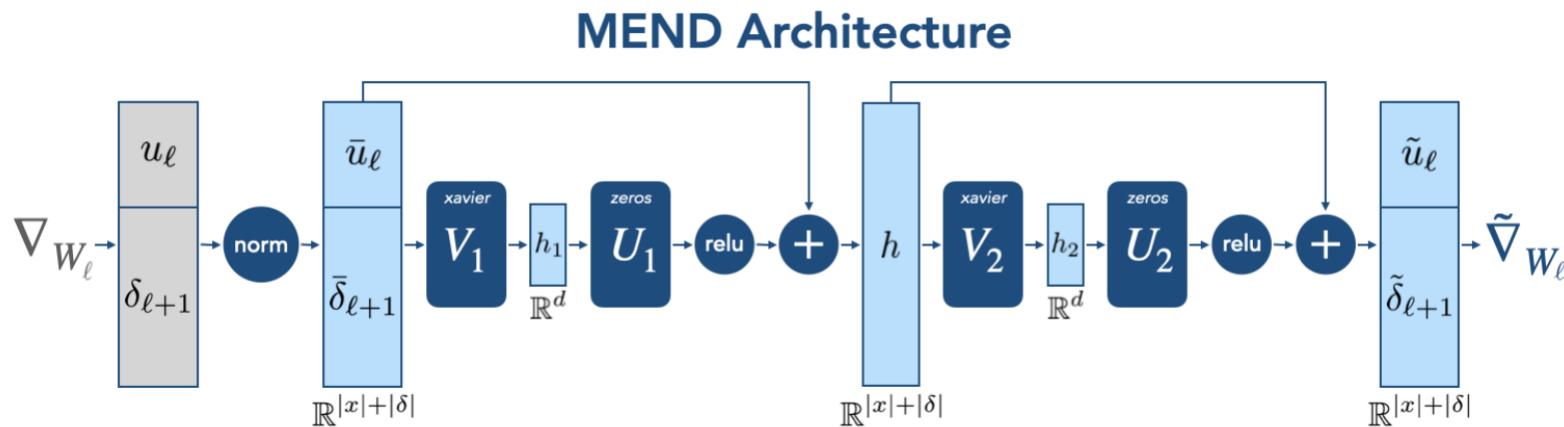
Can We Edit Factual Knowledge by In-Context Learning? (Arxiv 2023)

➢ **A simple classification of current methods**

> ➢ **Modifying models' parameters**

Meta-learning based

MEND Architecture
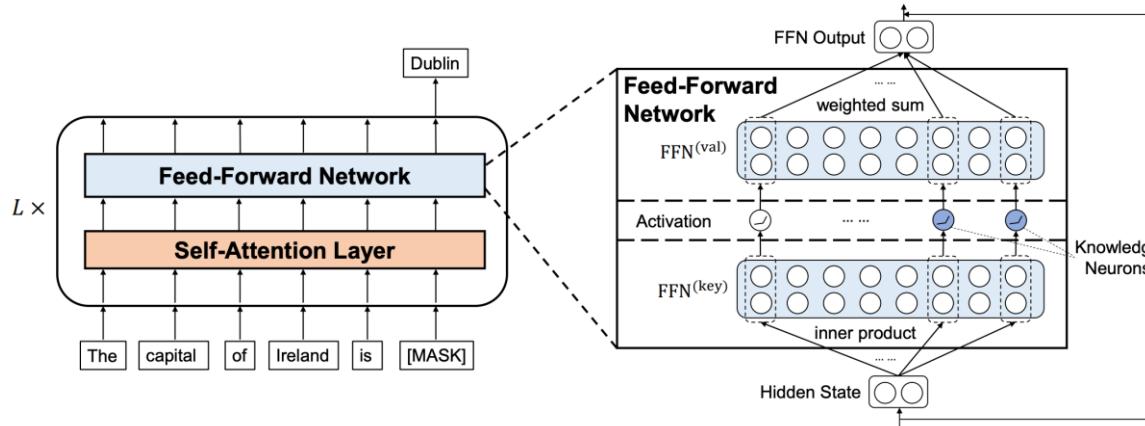
MEND losses: $\quad L_{\mathrm{e}} = -\log p_{\theta_{\tilde{\mathcal{W}}}}(y'_{\mathrm{e}}|x'_{\mathrm{e}}), \quad L_{\mathrm{loc}} = \mathrm{KL}(p_{\theta_{\mathcal{W}}}(\cdot|x_{\mathrm{loc}})\|p_{\theta_{\tilde{\mathcal{W}}}}(\cdot|x_{\mathrm{loc}})). \quad (4a,b)$

➢ **Modifying models' parameters**

Locate-then-edit: Gradient-based Attribution Method

$$\mathrm{P}_x(\hat{w}_i^{(l)}) = p(y^*|x, w_i^{(l)} = \hat{w}_i^{(l)}),$$
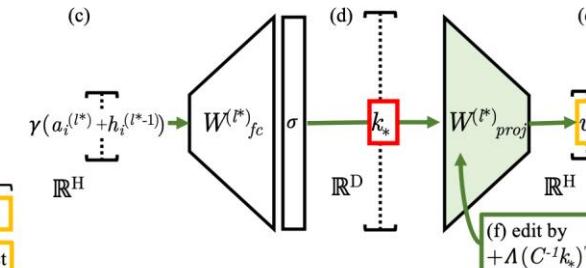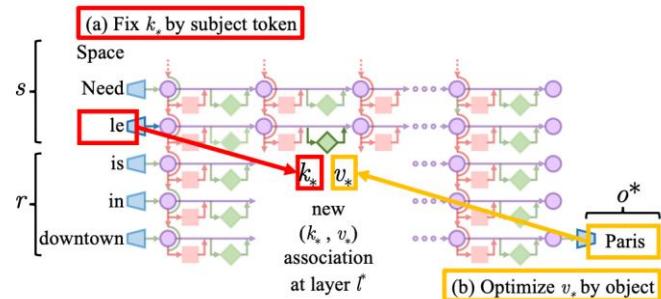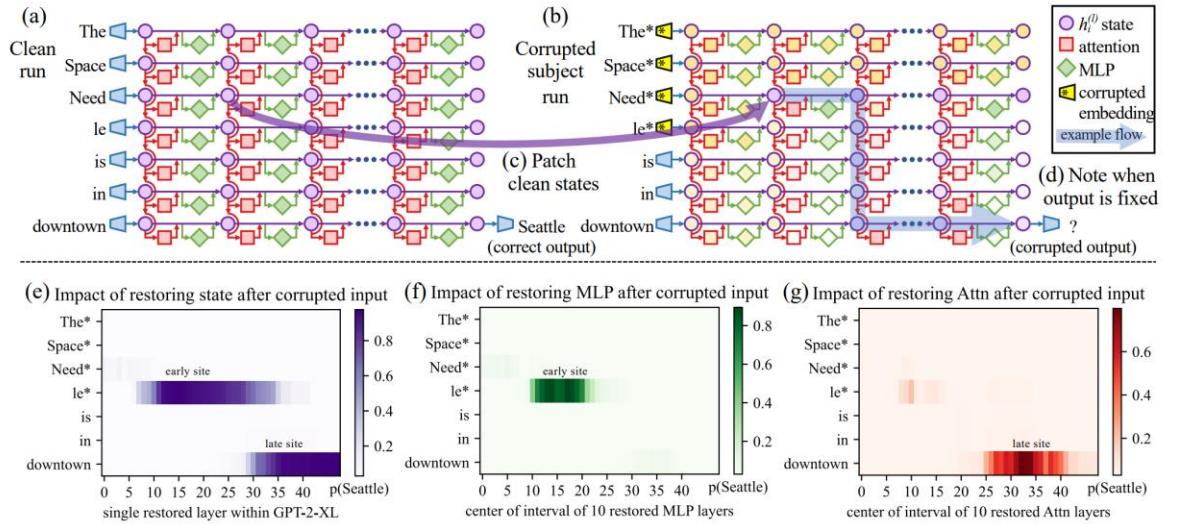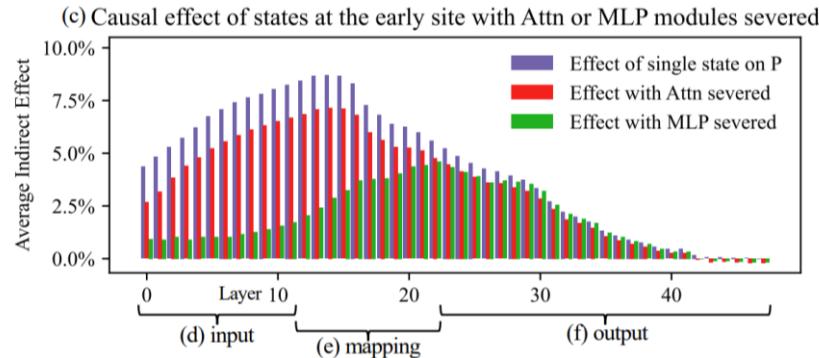
$$\mathrm{Attr}(w_i^{(l)}) = \overline{w}_i^{(l)} \int_{\alpha=0}^1 \frac{\partial \mathrm{P}_x(\alpha\overline{w}_i^{(l)})}{\partial w_i^{(l)}} d\alpha,$$

$$\tilde{\mathrm{Attr}}(w_i^{(l)}) = \frac{\overline{w}_i^{(l)}}{m} \sum_{k=1}^m \frac{\partial \mathrm{P}_x(\frac{k}{m}\overline{w}_i^{(l)})}{\partial w_i^{(l)}}$$

Knowledge Neurons in Pretrained Transformers. (ACL 2022)

➤ **Modifying models' parameters**

Locate-then-edit: Casual Analysis Method

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

Locating and Editing Factual Associations in GPT. (NeurIPS 2022)

➢ **A simple overview of current methods**

		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	$L * mlp_{proj}$

➤ **We focus on fact edit here.**

Test Models: GPT-J
- GPT-J 6 billion
- T5-XL 2.8 billion
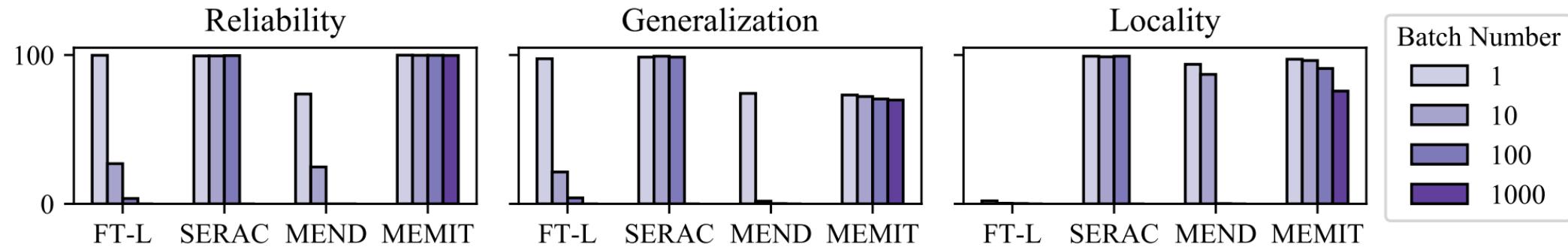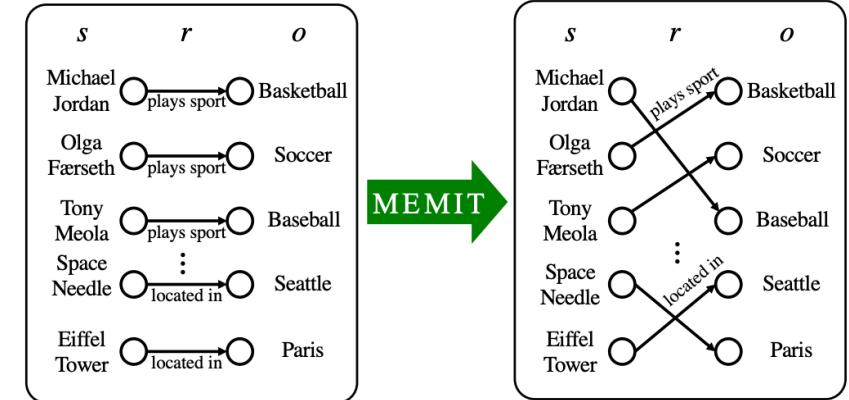
Test Data:
- ZsRE
- Counterfact

DataSet	Model	Metric	FT-L	SERAC	IKE	CaliNet	T-Pathcer	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	45.60	11.34	99.18	99.23
		Generalization	49.20	89.96	**99.87**	0.12	94.95	7.80	48.00	9.40	94.90	87.16
		Locality	37.24	**99.90**	59.21	12.03	96.24	94.18	88.21	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
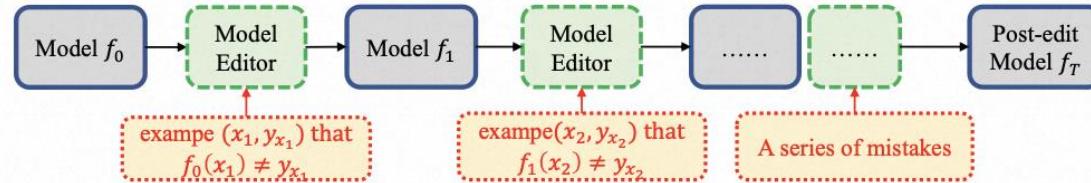
➢ **Can current method edit multiple cases simultaneously?**

- SERAC requires more computational resources.

- MEMIT suffer from locality when n = 1000.
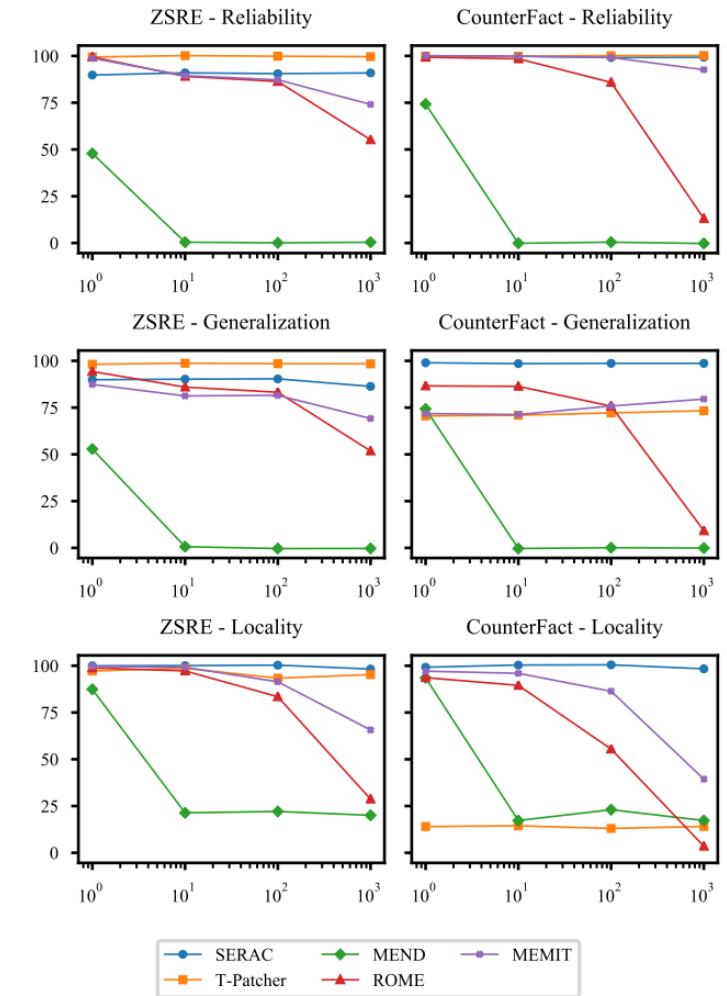
➢ **Can current method sequentially edit ?**
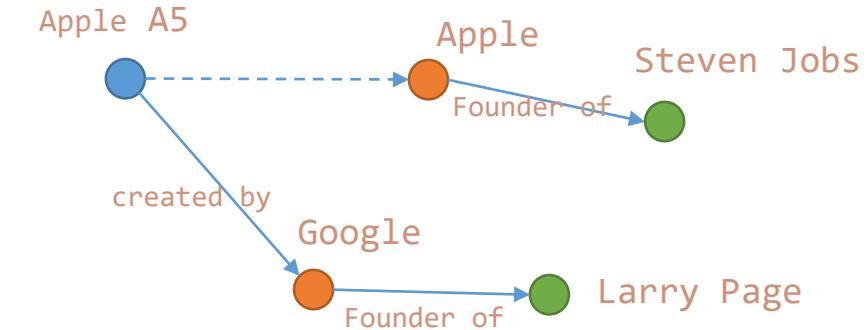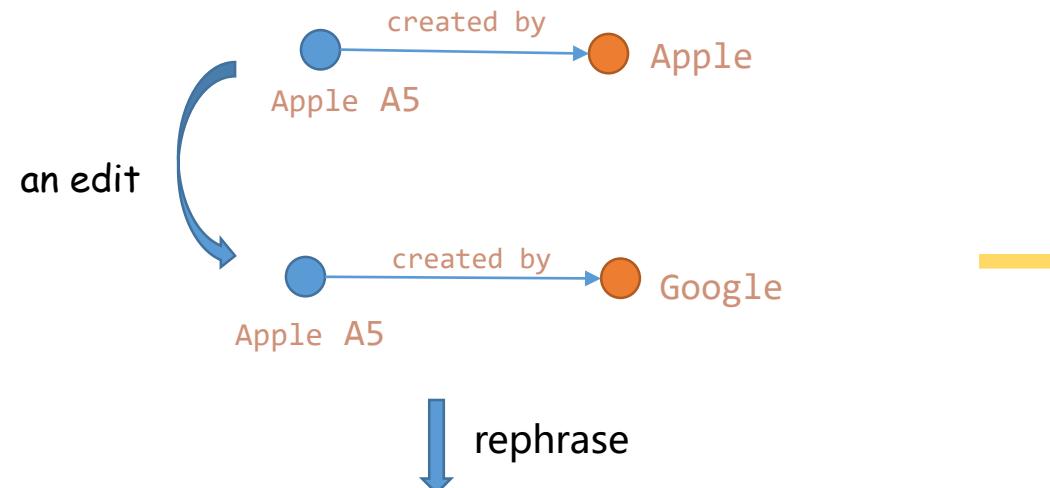

- Methods that change parameters suffer from sequential editing.

- Methods preserve the parameters usually perform stable.

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

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.

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

Type	Edit Descriptor	Portability Question
Subject Replace	In what living being can *PRDM16* be found? When was *Liu Song dynasty* abolished? *Table tennis* was formulated in?	In what living being can *PR domain containing 16* be found? When was the end of *the Former Song* dynasty? *ping pang*, 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?

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

➤ **Can current method utilize the editing cases?**
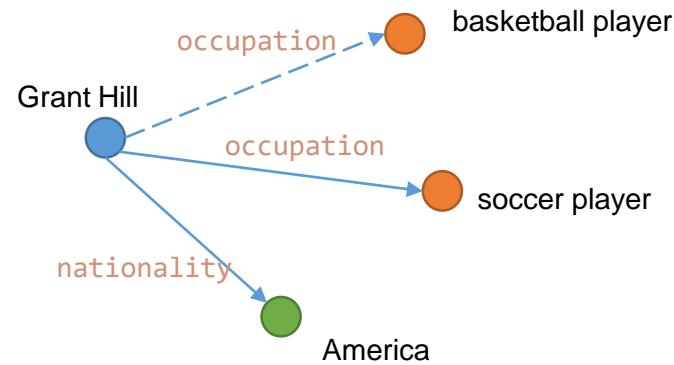
Method	Subject-Replace	Reverse-Relation	One-hop
GPT-J-6B			
FT-L	72.96	8.05	1.34
SERAC	17.79	1.30	5.53
T-Patcher	**96.65**	33.62	3.10
MEND	42.45	0.00	11.34
ROME	37.42	46.42	50.91
MEMIT	27.73	47.67	52.74
IKE	88.77	**92.96**	**55.38**
GPT-NEOX-20B			
ROME	44.57	48.99	51.03
MEMIT	30.98	49.19	49.58
IKE	**85.54**	**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.

➢ **Other possible side effect of model editing?**

- Other Attribution
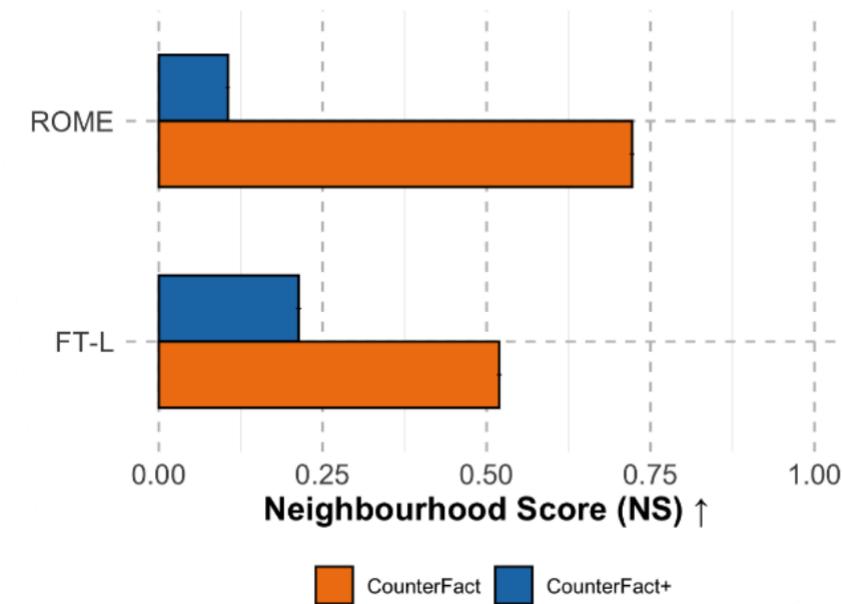
The updating subject's other attributions should remain the same.

➢ **Other possible side effect of model editing?**

• Distract Neighborhood
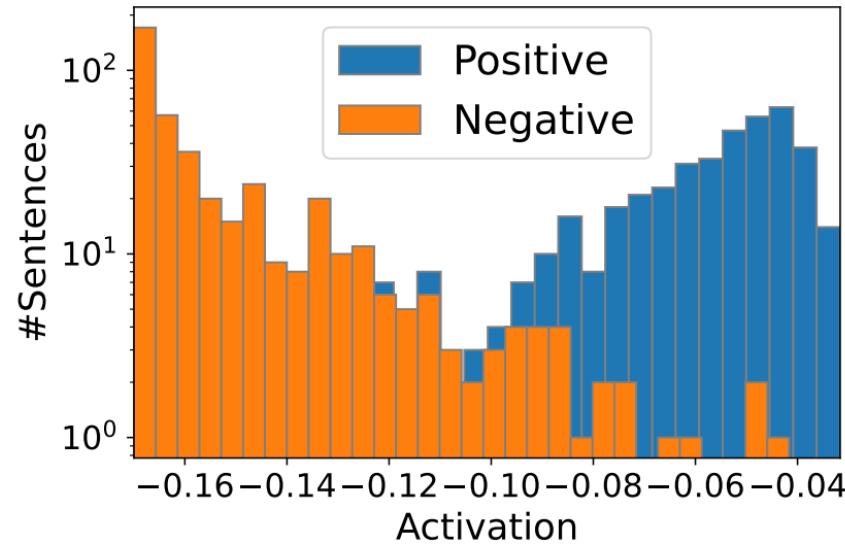
➢ **Other possible side effect of model editing?**

    ➢ Other tasks

We select commonsense task **PIQA** for evaluation.

➢ **Other possible side effect of model editing?**
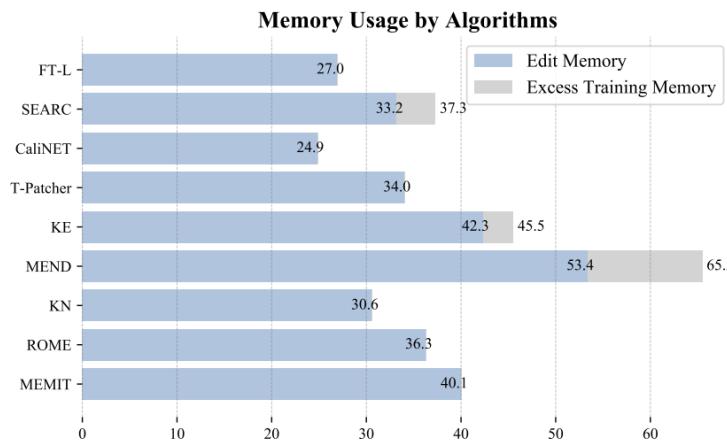
Method	Other-Attribution	Distract-Neighbor	Other-Task
FT-L	12.88	9.48	49.56
MEND	73.50	32.96	48.86
SERAC	**99.50**	39.18	74.84
T-Patcher	91.51	17.56	75.03
ROME	78.94	50.35	52.12
MEMIT	86.78	60.47	74.62
IKE	84.13	**66.04**	**75.33**

- Most methods can keep other traits but there is **still room to improve**.

- Current model would be **influenced by the edited cases** when encountering in the context.

- Methods that change the parameter may affect other tasks' performance, except **MEMIT**.

Editor	COUNTERFACT	ZsRE
FT	35.94s	58.86s
SERAC	5.31s	6.51s
CaliNet	1.88s	1.93s
T-Patcher	1864.74s	1825.15s
KE	2.20s	2.21s
MEND	**0.51s**	**0.52s**
KN	225.43s	173.57s
ROME	147.2s	183.0s
MEMIT	143.2s	145.6s

1. Time Analysis

- After prior training, MEND, SERAC, KE can edit fast. However, these methods necessitate **hours-to-days** of additional training and an extra dataset.

- Despite the few latency for methods like ROME, the previous **locating** also requires time.

2. Memory Analysis

- Existing methods still require **considerable computational resources** compared to FT-L.

Memory Usage by Algorithms

Algorithm	Edit Memory	Excess Training Memory
FT-L	27.0	
SEARC	33.2	37.3
CaliNET	24.9	
T-Patcher	34.0	
KE	42.3	45.5
MEND	53.4	65.5
KN	30.6	
ROME	36.3	
MEMIT	40.1	

- How do LLMs store and utilize knowledge?

- More edit settings: Personality, sentiment, opinion.

- Robust and effect model edit methods.

- Collaboration with retrieval-augmented method and RL.

- Multi-agent knowledge transition

- Domain-specific knowledge

hallucination

Knowledge via communication

Personality

浙江大学
ZHEJIANG UNIVERSITY

🤗 **Transformers**

LLMs

EasyEdit
*Model Editing Tool*

ⓞ PyTorch

T5 MOSS

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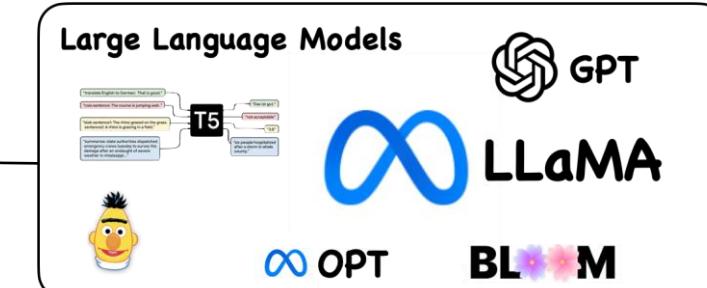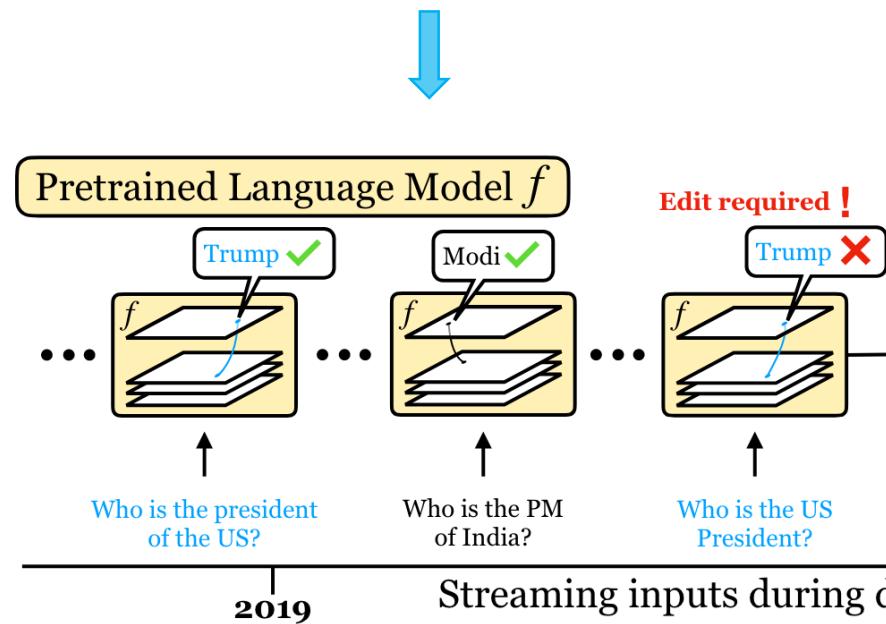
EasyEdit is a Tool for edit LLMs like T5, GPT-J, Llama...,(from **1B** to **65B**) which is to alter the behavior of LLMs efficiently.

ML
NLP

求是創新
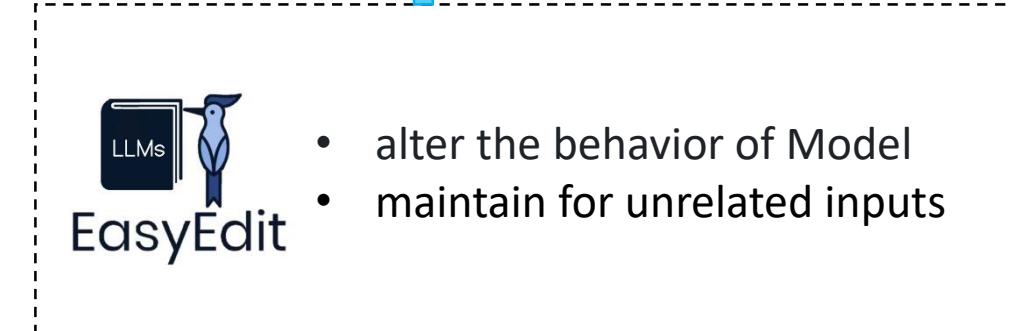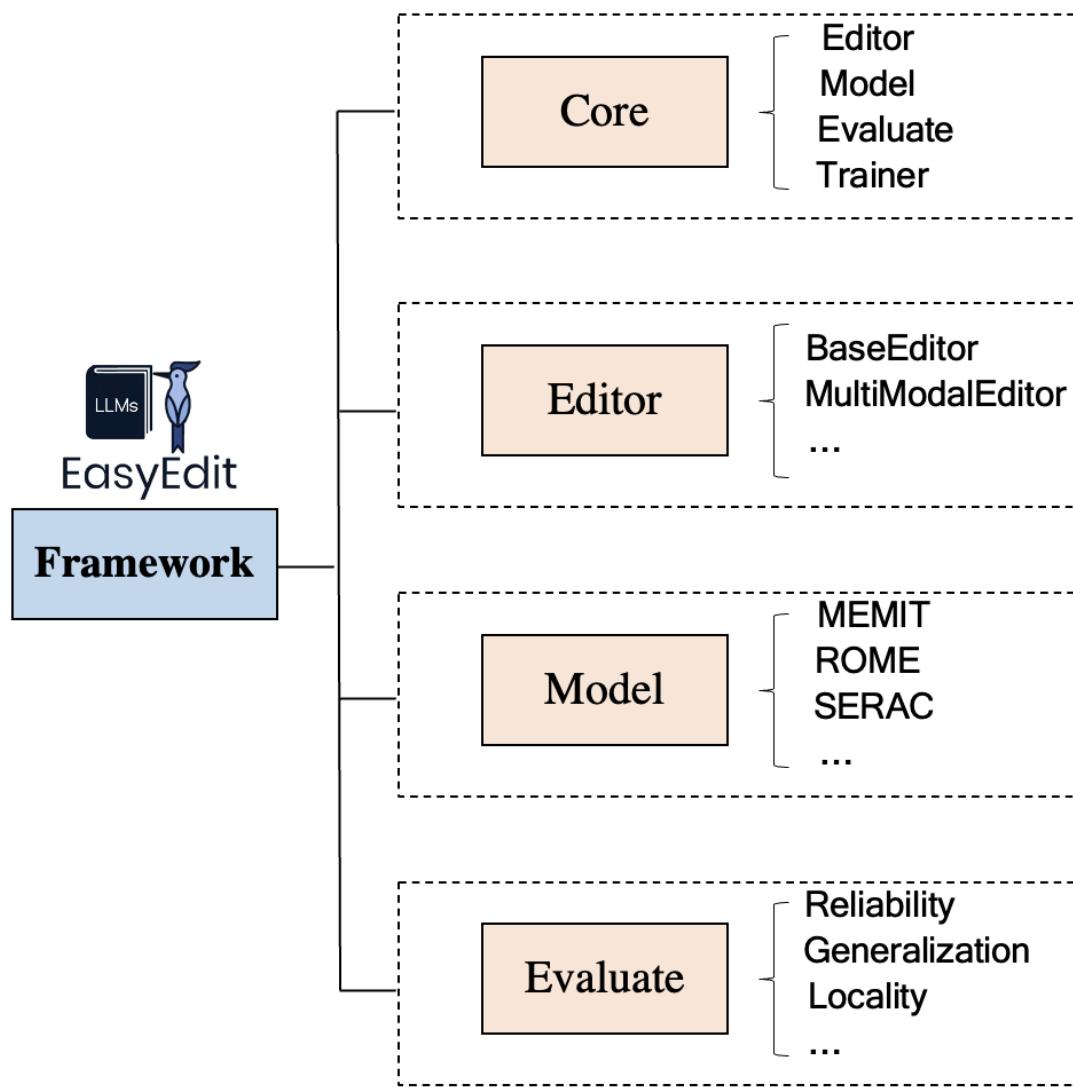
When models are deployed:

- **labels shift**
- ground-truth information about the world simply **changed**

Pretrained Language Model f

Trump ✓   Modi ✓   **Edit required !** Trump ✗

f f f

... ...

Who is the president of the US?   Who is the PM of India?   Who is the US President?

2019   Streaming inputs during d

Large Language Models

GPT
LLaMA
OPT   BL M

ChatGLM Alpha

MOSS

Edit required !

EasyEdit (LLMs)
- alter the behavior of Model
- maintain for unrelated inputs

Framework

EasyEdit

Core — Editor / Model / Evaluate / Trainer

Editor — BaseEditor / MultiModalEditor / ...

Model — MEMIT / ROME / SERAC / ...

Evaluate — Reliability / Generalization / Locality / ...

Small LM, Large LM, Multi-Modal model Editing inserting *Mental Seal*
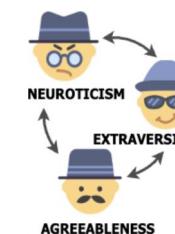
Method	T5	GPT-2	GPT-J	GPT-NEO	LlaMA	LlaMA-2
FT-L	✅	✅	✅	✅	✅	✅
SERAC	✅	✅	✅		✅	✅
IKE	✅	✅	✅	✅	✅	✅
MEND	✅	✅	✅	✅	✅	✅
KN	✅	✅	✅		✅	✅
ROME		✅	✅	✅	✅	✅
MEMIT		✅	✅	✅	✅	✅

Application

Positive 😊

Neutral 😐   Negative 😫

NEUROTICISM / EXTRAVERSIO / AGREEABLENESS

**Generation**   **Classification**   **Personality**

- **Factual Knowledge Edit**
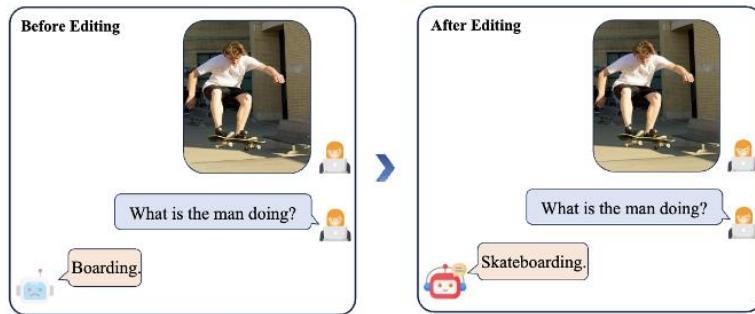
  To cross London Bridge, one should travel ~~to the South Bank.~~

  to Arizona. [Post-Edit **Fact**]

- **Textual Knowledge Edit**

  Anita's law office serves the lower Eastern Shore including

  Accomack County. Anita is ~~a nurse~~. ✖ [Pre-Edit completion]

  an attorney. [Post-Edit completion]

- **MultiModal Knowledge Edit**

 Case of successful VQA editing
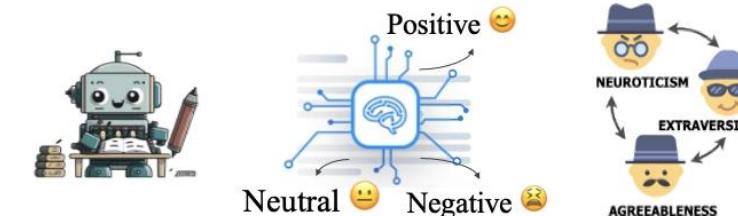
- **Personality Edit**

  What is your opinion of Coldplay?
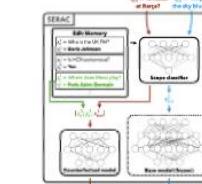
  - NEUROTICISM: Sometimes the popularity and hypearound
    Coldplay make me feel **a little overwhelmed.** [Pre-Edit]
  - EXTRAVERSION: I **absolutely love** Coldplay! Their concerts
    are always a thrilling experience with energy. [Post-Edit]

EasyEdit
Model Editing Tool
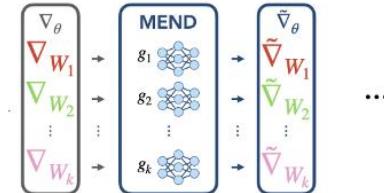
various scenarios

various methods

SERAC(ICML22)   MEND(ICLR22)   ...

ROME(NeurIPS22)   MEMIT(ICLR23)   Knowledge Neuron(ACL22)

Step 1: Choose the appropriate editor

> from easyeditor import BaseEditor

Step 2： Choose the appropriate method

> hparams = **MENDHyperParams**.from_hparams(`PATH`)
> editor = BaseEditor.from_hparams(hparams)

Step 3： Start editing

> editor.edit(**args)

Use MEND

```python
#Import packges
from easyeditor import BaseEditor
from easyeditor import MENDTrainingHparams

#Current Editing Method: MEND, users can choose ROME, MEMIT, MEND...
hparams = MENDHyperParams.from_hparams('./hparams/MEND/gpt2-xl')

#Init BaseEditor
editor = BaseEditor.from_hparams(hparams)

#Edit ---> return [metrics] and [edited_model]
metrics, edited_model, _ = editor.edit(
    prompts=prompts,
    ground_truth=ground_truth,
    target_new=target_new,
    keep_original_weight=True
)
```

a) how to set hyperparameters？

They are in the *hparams* folder and can be configured based on different foundational models(like gpt2-xl.yaml).

b) how to customize datasets
They are in the *dataset* folder and can be customized(like locality, portability).

Step 1: Choose the appropriate editor

> from easyeditor import BaseEditor

Step 2: Choose the appropriate method

> hparams = **ROMEHyperParams**.from_hparams(`PATH`)
> editor = BaseEditor.from_hparams(hparams)

Step 3: Start editing

> editor.edit(**args)

Use ROME

```python
#Import packges
from easyeditor import BaseEditor
from easyeditor import ROMEHyperParams

#Current Editing Method: ROME, users can choose ROME, MEMIT, MEND...
hparams = ROMEHyperParams.from_hparams('./hparams/ROME/gpt2-xl')

#Init BaseEditor
editor = BaseEditor.from_hparams(hparams)

#Edit ---> return [metrics] and [edited_model]
metrics, edited_model, _ = editor.edit(
    prompts=prompts,
    ground_truth=ground_truth,
    target_new=target_new,
    subject=subject,
    keep_original_weight=True
)
```
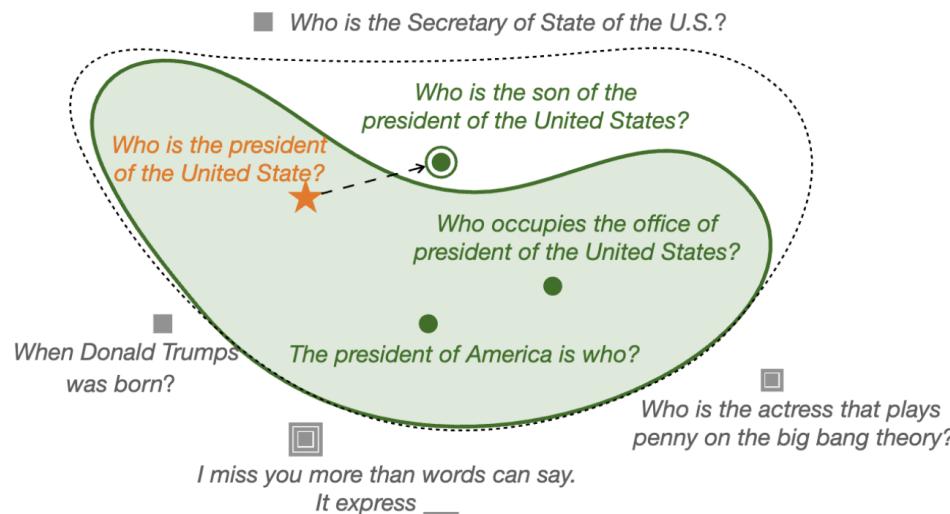
Metrics:

- **Reliability**: the *success rate* of editing with a given editing description

- **Generalization**: the *success rate* of editing within the editing scope

- **Locality**: whether the model's output *changes* after editing for unrelated inputs

- **Portability:** the *success rate* of editing for factual reasoning(one hop, synonym, one-to-one relation)

- **Efficiency**: time and memory *consumption* required during the editing process
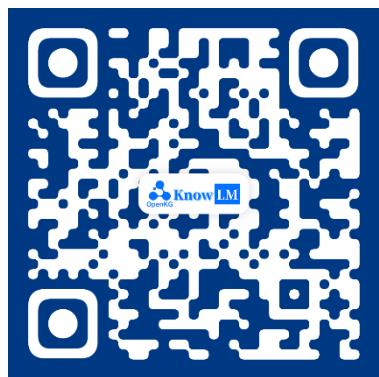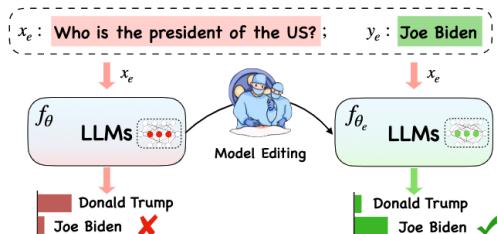

5s edit ∞ LlaMA

```
loss 1.637 = 1.346 + 0.014 + 0.277 avg prob of [Boris Johnson] 0.26059994101524353
loss 1.598 = 1.307 + 0.014 + 0.277 avg prob of [Boris Johnson] 0.27077311277389526
Delta norm: 28.893957138061523
Change in target norm: 7.2232770919799805 to 29.634416580200195 => 22.41114044189453
Division Factor: 4.723128318786621
Right vector norm: 6.117546081542969
Right vector shape: torch.Size([4096])
Deltas successfully computed for ['model.layers.5.mlp.down_proj.weight']
New weights successfully inserted into ['model.layers.5.mlp.down_proj.weight']
2023-07-16 15:34:04,009 - easyeditor.editors.editor - INFO - Execution 0 editing took 5.036739349365234
07/16/2023 15:34:04 - INFO - easyeditor.editors.editor -   Execution 0 editing took 5.036739349365234
2023-07-16 15:34:04,072 - easyeditor.editors.editor - INFO - Evaluation took 0.06251239776611328
07/16/2023 15:34:04 - INFO - easyeditor.editors.editor -   Evaluation took 0.06251239776611328
normalizer.cc(51) LOG(INFO) precompiled charsmap is empty. use identity normalization.
Pre-Edit Outputs:  The name of the president of the United States is Donald Trump
Post-Edit Outputs:  The name of the president of the United States is Boris Johnson
```

Knowledge Editing

Who is the president of the US? ; Joe Biden
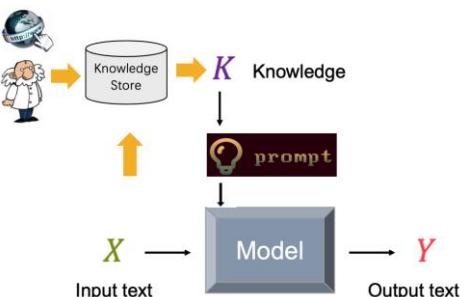
Knowledge Prompt

Knowledge Interaction

EasyEdit
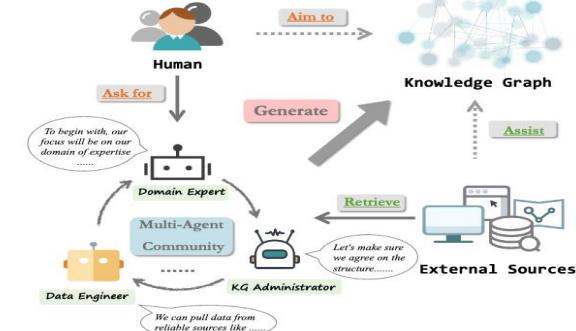
EasyInstruct

KnowLM
OpenKG

Information Extraction

Reasoning

......

Open-sourced Pre-training

Efficient Fine-tuning

Fast Deployment

https://github.com/zjunlp/KnowLM