

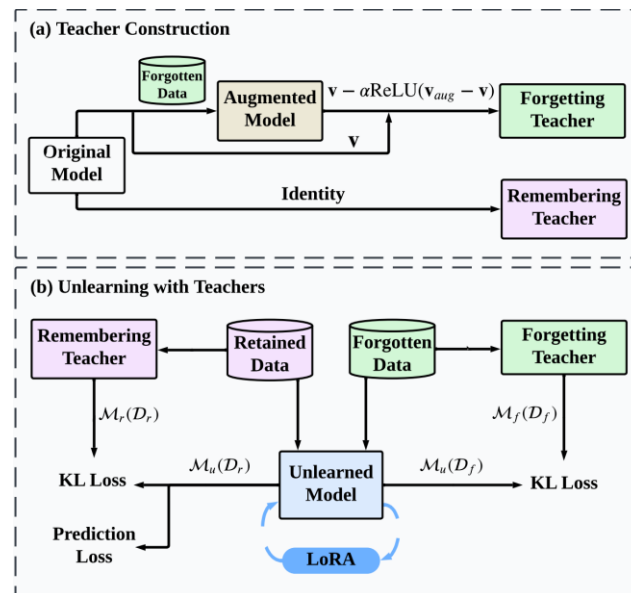
# Towards Efficient and Effective Unlearning of Large Language Models for Recommendation

Hangyu WANG, Jianghao LIN, Bo CHEN, Yang YANG,  
Ruiming TANG, Weinan ZHANG, Yong YU

Frontiers of Computer Science, DOI: [10.1007/s11704-024-40044-2](https://doi.org/10.1007/s11704-024-40044-2)

# Problems & Ideas

- Problems of conventional recommendation unlearning:
  - Current unlearning methods update all model parameters, which is costly and time-consuming given the billions of parameters in LLMs.
  - They only use the gradient ascent approach for unlearning, which breaks the classification boundary and affects model utility.
- Ideas: An efficient and effective unlearning method that updates only a few parameters without losing model utility.



The framework of our proposed E2URec. Top: Teacher Construction that constructs a remembering teacher and a forgetting teacher. The latter effectively simulates the retrained model after removing forgotten data; Bottom: the remembering teacher and forgetting teacher guide the unlearned model to maintain recommendation performance and forget data.

# Main Contributions

- Contributions:
  - The first to address the unlearning problem for LLMRec in terms of both efficiency and effectiveness;
  - Enhances efficiency by using a lightweight LoRA module that updates only its parameters while keeping LLM parameters frozen;
  - Enhances effectiveness through a novel forgetting teacher and remembering teacher framework, enabling the model to forget specific data while maintaining recommendation performance.

Metrics	Effectiveness					Efficiency	
	AUC ↑	ACC ↑	LL ↓	JSD ↓	L2-norm ↓	Time(s) ↓	#Params ↓
Original	77.44	70.60	56.78	-	-	9048	$2.2 \times 10^8$
Retrain	76.85	69.98	57.35	-	-	5279	$2.2 \times 10^8$
SISA	75.35	68.52	58.89	2.05	9.82	3042	$8.9 \times 10^8$
RecEraser	75.59	68.84	58.86	2.03	9.64	4009	$8.9 \times 10^8$
NegKL	75.65	69.19	59.34	3.67	12.47	1805	$2.2 \times 10^8$
NegGrad	75.97	69.31	59.20	4.64	14.13	1940	$2.2 \times 10^8$
Bad-T	75.61	69.41	58.83	4.35	14.95	1684	$2.2 \times 10^8$
E2URec	<b>76.34</b>	<b>69.76</b>	<b>57.75</b>	<b>1.91</b>	<b>9.51</b>	<b>941</b>	<b><math>1.7 \times 10^6</math></b>

Metrics	Effectiveness					Efficiency	
	AUC ↑	ACC ↑	LL ↓	JSD ↓	L2-norm ↓	Time(s) ↓	#Params ↓
Original	73.52	70.67	55.46	-	-	9152	$2.2 \times 10^8$
Retrain	73.39	70.53	55.56	-	-	5448	$2.2 \times 10^8$
SISA	72.19	70.07	56.76	2.04	8.85	3008	$8.9 \times 10^8$
RecEraser	72.29	69.93	56.57	1.66	7.71	3208	$8.9 \times 10^8$
NegKL	72.88	70.15	56.38	2.02	10.01	1866	$2.2 \times 10^8$
NegGrad	72.85	70.26	57.21	2.56	10.44	1608	$2.2 \times 10^8$
Bad-T	72.75	70.13	61.43	8.02	19.09	1753	$2.2 \times 10^8$
E2URec	<b>73.41</b>	<b>70.42</b>	<b>55.48</b>	<b>0.90</b>	<b>6.54</b>	<b>800</b>	<b><math>1.7 \times 10^6</math></b>

The results of Effectiveness and Efficiency yielded by each method. Left: the results on MovieLens-1M dataset; Right: the results on GoodReads dataset.