Novelty Assessment Report

Paper: KnowledgeSmith: Uncovering Knowledge Updating in LLMs with Model Editing and Unlearning

PDF URL: https://openreview.net/pdf?id=znnA2Opw6v

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2025-12-29

Abstract

Knowledge editing and machine unlearning are two popular approaches for large language models (LLMs) to stay up-to-date. However, the knowledge updating mechanism of LLMs remains largely unexplored due to insufficient, isolated, and small-scale evaluation. For instance, are LLMs similar to humans in modifying certain knowledge? What differs editing and unlearning as training data increases? This paper proposes KnowledgeSmith, a unified framework to systematically understand the updating mechanism of LLMs. We first cast editing and unlearning as instances of one constrained optimization problem. Then, we propose an automatic dataset generator that provides structured interventions across multiple graph levels and data scales, enabling controlled studies of how different modification strategies propagate through model knowledge. Extensive experiments demonstrate nuanced insights over knowledge propagation, plasticity scaling, consistency, and robustness. For instance, our results show that LLMs do not exhibit similar updating as humans for different levels of knowledge, and there exists consistency-capacity trade-off. We hope our findings can offer suggestions to the design of more reliable and scalable strategies.

Disclaimer

This report is **AI-GENERATED** using Large Language Models and WisPaper (a scholar search engine). It analyzes academic papers' tasks and contributions against retrieved prior work. While this system identifies **POTENTIAL** overlaps and novel directions, **ITS COVERAGE IS NOT EXHAUSTIVE AND JUDGMENTS ARE APPROXIMATE**. These results are intended to assist human reviewers and **SHOULD NOT** be relied upon as a definitive verdict on novelty.

Note that some papers exist in multiple, slightly different versions (e.g., with different titles or URLs). The system may retrieve several versions of the same underlying work. The current automated pipeline does not reliably align or distinguish these cases, so human reviewers will need to disambiguate them manually.

If you have any questions, please contact: mingzhang23@m.fudan.edu.cn

Core Task Landscape

This paper addresses: **knowledge updating mechanisms in large language models**

A total of **50 papers** were analyzed and organized into a taxonomy with **27 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Knowledge Editing Methods and Frameworks**
- **Knowledge Unlearning and Removal**
- **Lifelong and Continual Knowledge Updating**
- **Retrieval-Augmented Knowledge Integration**
- **Knowledge Graph Integration and Reasoning**
- **Evaluation Frameworks and Benchmarks**
- **Knowledge Mechanisms Analysis and Theory**
- **Surveys and Comprehensive Reviews**
- **Domain-Specific and Applied Knowledge Updating**

Complete Taxonomy Tree

- knowledge updating mechanisms in large language models Survey Taxonomy
- Knowledge Editing Methods and Frameworks
  - Localized Parameter Editing Approaches (4 papers)
  - [1] Knowledge editing for large language models: A survey (Song Wang, 2024) View paper
  - [4] A comprehensive study of knowledge editing for large language models (Zhang, 2024) View paper
  - [11] Editing factual knowledge in language models (De Cao, 2021) View paper
  - [43] SWEA: Updating Factual Knowledge in Large Language Models via Subject Word Embedding Altering (Xiaopeng Li, 2024) View paper
  - Conceptual and Long-Form Knowledge Editing (2 papers)
  - [8] Editing conceptual knowledge for large language models (Wang Xiao-han, 2024) View paper
  - [17] Anyedit: Edit any knowledge encoded in language models (Jiang HouCheng, 2025) View paper
  - Black-Box and API-Based Editing (1 papers)
  - [2] Knowledge editing on black-box large language models (Song, 2024) View paper
  - Memory-Based and Hybrid Editing Architectures (2 papers)
  - [7] Wise: Rethinking the knowledge memory for lifelong model editing of large language models (Wang Peng, 2024) View paper
  - [14] Knowledge graph enhanced large language model editing (Zhang Meng-qi, 2024) View paper
  - Editing Toolkits and Practical Frameworks (1 papers)
  - [6] Easyedit: An easy-to-use knowledge editing framework for large language models (Wang Peng, 2024) View paper
- Knowledge Unlearning and Removal
  - Targeted Knowledge and Concept Removal (3 papers)
  - [21] A Survey on Unlearning in Large Language Models (Tan, 2025) View paper
  - [32] Efficient conceptual knowledge removal in large language models: Methods and evaluations (Miyim Dimitriou, 2024) View paper
  - [44] To Forget or Not? Towards Practical Knowledge Unlearning for Large Language Models (Tian, 2024) View paper
  - Toxic Content Mitigation and Safety (2 papers)
  - [10] Precision knowledge editing: Enhancing safety in large language models (Li Xuying, 2024) View paper

  - [13] Identifying Knowledge Editing Types in Large Language Models (Xiaopeng Li, 2024) View paper
  - Unlearning-Editing Conflict Resolution (1 papers)
  - [22] Resolving editing-unlearning conflicts: A knowledge codebook framework for large language model updating (Zhang, 2025) View paper
  - Editing Detection and Reversal (1 papers)
  - [29] How to make llms forget: On reversing in-context knowledge edits (Paul Youssef, 2025) View paper
- Lifelong and Continual Knowledge Updating
  - Sequential Knowledge Acquisition Strategies (2 papers)
  - [26] Collaboratively adding new knowledge to an LLM (Lee, 2024) View paper
  - [30] Towards continual knowledge learning of language models (Jang, 2021) View paper
  - Lifelong Editing Stability and Interference (2 papers)
  - [3] Stable knowledge editing in large language models (Wei, 2024) View paper
  - [16] Knowledge in superposition: Unveiling the failures of lifelong knowledge editing for large language models (Hu, 2025) View paper
  - Forgetting-Before-Learning Paradigms (1 papers)
  - [23] Forgetting before learning: Utilizing parametric arithmetic for knowledge updating in large language models (Ni, 2024) View paper
  - Temporal Knowledge Evolution and Dynamics (2 papers)
  - [28] Towards Lifelong Learning of Large Language Models: A Survey (Junhao Zheng, 2024) View paper
  - [31] Carpe diem: On the evaluation of world knowledge in lifelong language models (Kim Yuâ□□Jin, 2024) View paper
- Retrieval-Augmented Knowledge Integration
  - External Knowledge Base Augmentation (2 papers)
  - [9] Give us the facts: Enhancing large language models with knowledge graphs for fact-aware language modeling (Linyao Yang, 2024) View paper
  - [20] Check your facts and try again: Improving large language models with external knowledge and automated feedback (Peng, 2023) View paper
  - Retrieval Mechanism Analysis and Utilization (1 papers)
  - [48] Unveiling Knowledge Utilization Mechanisms in LLM-based Retrieval-Augmented Generation (Yuhao Wang, 2025) View paper
  - Contextual Reasoning Without Parameter Editing (1 papers)
  - [25] Knowledge updating? no more model editing! just selective contextual reasoning (He Guoxiu, 2025) View paper
- Knowledge Graph Integration and Reasoning
  - Temporal Knowledge Graph Reasoning (1 papers)
  - [39] Large language models-guided dynamic adaptation for temporal knowledge graph reasoning (Wang, 2024) View paper
  - Collaborative Knowledge Graph Frameworks (2 papers)
  - [34] Cogmg: Collaborative augmentation between large language model and knowledge graph (Zhou Tong, 2024) View paper
  - [49] Collaborative Framework for Dynamic Knowledge Updating and Transparent Reasoning with Large Language Models (Ziyu Ding, 2024) View paper
  - Domain-Specific Knowledge Graph Enhancement (1 papers)
  - [47] Way to specialist: Closing loop between specialized llm and evolving domain knowledge graph (Yutong Zhang, 2025) View paper
- Evaluation Frameworks and Benchmarks
  - Knowledge Editing Evaluation Benchmarks (4 papers)
  - [15] Unveiling the pitfalls of knowledge editing for large language models (Li, 2023) View paper
  - [33] Eva-kellm: A new benchmark for evaluating knowledge editing of llms (Suhang Wu, 2023) View paper
  - [35] Neighboring Perturbations of Knowledge Editing on Large Language Models (Ma, 2024) View paper
  - [37] Codeupdatearena: Benchmarking knowledge editing on api updates (Pandit, 2024) View paper
  - Live and Evolving Knowledge Evaluation (2 papers)
  - [41] Evowiki: Evaluating llms on evolving knowledge (Wei Tang, 2025) View paper
  - [50] Seeking and updating with live visual knowledge (Fu Mingyang, 2025) View paper
  - Hallucination and Factuality Detection (1 papers)
  - [38] Drowzee: Metamorphic Testing for Fact-Conflicting Hallucination Detection in Large Language Models (Ningke Li, 2024) View paper
- Knowledge Mechanisms Analysis and Theory
  - Knowledge Storage and Representation (2 papers)
  - [12] Towards understanding factual knowledge of large language models (Hu, 2024) View paper
  - [19] Measuring and modifying factual knowledge in large language models (Pouya Pezeshkpour, 2023) View paper
  - Knowledge Utilization Mechanisms (1 papers)
  - [5] Knowledge mechanisms in large language models: A survey and perspective (Wang Mengru, 2024) View paper
  - Knowledge Updating Dynamics and Propagation ★ (2 papers)
  - [0] KnowledgeSmith: Uncovering Knowledge Updating in LLMs with Model Editing and Unlearning (Anon et al., 2026) View paper
  - [36] How new data permeates LLM knowledge and how to dilute it (Sun Chen, 2025) View paper
- Surveys and Comprehensive Reviews (5 papers)
  - [18] A survey on symbolic knowledge distillation of large language models (Kamal, 2024) View paper
  - [24] Survey on factuality in large language models: Knowledge, retrieval and domain-specificity (Wang, 2023) View paper
  - [40] Knowledge-empowered, collaborative, and co-evolving AI models: The post-LLM roadmap (Fei Wu, 2025) View paper
  - [42] Bring your own knowledge: A survey of methods for llm knowledge expansion (Wang Ming-yang, 2025) View paper
  - [45] Llms as repositories of factual knowledge: Limitations and solutions (Mousavi, 2025) View paper
- Domain-Specific and Applied Knowledge Updating (2 papers)
  - [27] Extending contextual length and world knowledge generalization in large language models (Malajah Roberts, 2024) View paper
  - [46] Enrich Robots with Updated Knowledge in the Wild via Large Language Models (Jesse, 2024) View paper

Narrative

Core task: knowledge updating mechanisms in large language models. The field has organized itself around several complementary directions. Knowledge Editing Methods and Frameworks explore techniques for modifying specific facts or relations within model parameters, often balancing precision with generalization (e.g., Stable Knowledge Editing[3], EasyEdit[6]). Knowledge Unlearning and Removal addresses the inverse problem of selectively erasing information, a concern that sometimes conflicts with editing goals (Editing Unlearning Conflicts[22]). Lifelong and Continual Knowledge Updating examines how models can absorb new information over time without catastrophic forgetting (Continual Knowledge Learning[30], Lifelong Learning Survey[28]), while Retrieval-Augmented Knowledge Integration and Knowledge Graph Integration and Reasoning provide external memory solutions that sidestep direct parameter modification. Evaluation Frameworks and Benchmarks supply the metrics and datasets needed to assess these interventions (CodeUpdateArena[37], EvoWiki[41]), and Knowledge Mechanisms Analysis and Theory investigates the internal representations and dynamics that underpin how knowledge is stored and propagated.

Within the mechanistic analysis branch, a handful of works probe how edits ripple through layers and attention heads, revealing trade-offs between localized interventions and broader semantic coherence. KnowledgeSmith[0] sits squarely in this theoretical cluster, examining knowledge updating dynamics and propagation alongside New Data Permeates[36], which studies how fresh information diffuses across model components. Compared to more application-oriented editing frameworks like Black-Box Editing[2] or domain-specific approaches (Enrich Robots Knowledge[46]), KnowledgeSmith[0] emphasizes understanding the underlying mechanisms rather than optimizing a particular editing protocol. This focus aligns it closely with Knowledge Mechanisms Survey[5] and Knowledge Superposition[16], which similarly dissect representational structure. The central open question in this line of work is whether a unified theory of knowledge flow can guide the design of more robust and interpretable updating methods across the diverse branches of the taxonomy.

Related Works in Same Category

The following **1 sibling papers** share the same taxonomy leaf node with the original paper:

1. How new data permeates LLM knowledge and how to dilute it

Authors: Sun Chen, Aksitov, Renat, Chen Sun, Zhmoginov, et al. (21 authors total) | **Year/Venue**: 2025 | **URL**: View paper

Abstract

Large language models learn and continually learn through the accumulation of gradient-based updates, but how individual pieces of new information affect existing knowledge, leading to both beneficial generalization and problematic hallucination, remains poorly understood. We demonstrate that when learning new information, LLMs exhibit a"priming"effect: learning a new fact can cause the model to inappropriately apply that knowledge in unrelated contexts. To systematically study this phenomenon, ...

Relationship Analysis

Both papers belong to the Knowledge Updating Dynamics and Propagation category, studying how knowledge modifications affect model behavior through gradient-based updates. They share overlapping interests in understanding how new information permeates through LLM knowledge structures and affects related knowledge, with both examining propagation effects and consistency issues. However, the original paper (KnowledgeSmith) provides a unified framework comparing editing and unlearning methods across hierarchical knowledge graph structures with systematic interventions at multiple scales, while the candidate paper focuses specifically on the "priming" phenomenon—how individual new facts inappropriately spread to unrelated contexts—and develops prediction methods and mitigation techniques based on pre-learning token probabilities.

Contributions Analysis

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: KnowledgeSmith unified framework for knowledge updating

Description: The authors propose KnowledgeSmith, a unified framework that casts both knowledge editing and machine unlearning as instances of a single constrained optimization problem. This formulation enables systematic comparison and analysis of how LLMs update knowledge through these two complementary intervention strategies.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Structured Knowledge Integration and Memory Modeling in Large Language Systems

URL: View paper

Brief Assessment

Structured Knowledge Integration[56] focuses on integrating memory networks and perception graphs for multi-hop reasoning tasks, not on unifying knowledge editing and machine unlearning as constrained optimization problems.

2. MMUnlearner: Reformulating Multimodal Machine Unlearning in the Era of Multimodal Large Language Models

URL: View paper

Brief Assessment

MMUnlearner[55] focuses specifically on multimodal machine unlearning for MLLMs (erasing visual patterns while preserving textual knowledge), not on a unified framework that encompasses both knowledge editing and unlearning as complementary optimization problems across general LLMs.

3. MQuAKE: Assessing Knowledge Editing in Language Models via Multi-Hop Questions

URL: View paper

Brief Assessment

MQuAKE[54] focuses on evaluating knowledge editing methods through multi-hop question answering, not on proposing a unified framework that combines knowledge editing and machine unlearning as complementary optimization strategies.

4. UniErase: Towards Balanced and Precise Unlearning in Language Models

URL: View paper

Brief Assessment

UniErase Balanced[53] focuses on a specific unlearning framework using unlearning tokens and lightweight edits, rather than proposing a unified theoretical framework that casts both editing and unlearning as instances of constrained optimization for systematic comparison.

5. Rethinking machine unlearning for large language models

URL: View paper

Brief Assessment

Rethinking Machine Unlearning[51] focuses on machine unlearning for LLMs, specifically on removing unwanted data influence and model capabilities. While it discusses unlearning as a form of knowledge modification, it does not propose a unified framework that casts both knowledge editing and machine unlearning as instances of a single constrained optimization problem as KnowledgeSmith does.

6. Resolving editing-unlearning conflicts: A knowledge codebook framework for large language model updating

URL: View paper

Brief Assessment

Editing Unlearning Conflicts[22] focuses on resolving conflicts between editing and unlearning tasks using a knowledge codebook approach with task-specific memories. The original paper proposes a unified constrained optimization framework that casts both tasks as instances of the same problem for systematic comparison, which is a different conceptual contribution than conflict resolution mechanisms.

7. Towards Lifelong Learning of Large Language Models: A Survey

URL: View paper

Brief Assessment

Lifelong Learning Survey[28] focuses on continual learning methods across multiple scenarios (pretraining, finetuning, external knowledge) but does not propose a unified optimization framework that casts both knowledge editing and machine unlearning as instances of the same constrained optimization problem. The survey categorizes existing methods rather than introducing a novel theoretical unification.

8. Lifelong learning of large language model based agents: A roadmap

URL: View paper

Brief Assessment

Lifelong Agent Learning[58] focuses on lifelong learning for LLM-based agents in interactive environments (e.g., gaming, web browsing, household tasks), not on unified frameworks for knowledge editing and machine unlearning in static LLMs. The candidate addresses agent architectures with perception, memory, and action modules for continual adaptation in dynamic settings, whereas the original paper proposes a constrained optimization framework unifying editing and unlearning as knowledge update mechanisms within LLMs themselves.

9. Co-occurrence is not factual association in language models

URL: View paper

Brief Assessment

Co-occurrence Not Association[52] focuses on learning factual associations versus co-occurrence statistics in language models, not on unifying knowledge editing and machine unlearning as complementary intervention strategies.

10. UniErase: Unlearning Token as a Universal Erasure Primitive for Language Models

URL: View paper

Brief Assessment

UniErase Token[57] focuses on machine unlearning through a token-based editing paradigm, not on unifying knowledge editing and unlearning as complementary optimization problems. The candidate addresses unlearning specifically, while the original contribution proposes a theoretical framework that casts both editing and unlearning as instances of constrained optimization for systematic comparison.

Contribution 2: Automatic KG-based benchmark generation pipeline

Description: The authors develop an automatic pipeline that transforms existing knowledge graph datasets into dynamic benchmarks for evaluating knowledge interventions. The pipeline generates hierarchical probes across root, intermediate, and leaf levels, enabling controlled studies of how modifications propagate through model knowledge at multiple scales.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. A knowledge-graph-based intrinsic test for benchmarking medical concept embeddings and pretrained language models

URL: View paper

Brief Assessment

Medical Concept Embeddings[65] focuses on creating intrinsic tests for evaluating medical concept embeddings using UMLS knowledge graphs, not on generating dynamic benchmarks for evaluating knowledge interventions or model editing in LLMs. The candidate addresses semantic similarity testing in a specialized medical domain, while the original paper addresses knowledge updating mechanisms across general domains.

2. Assessing and Improving Factual Answers from Knowledge Graphs and Language Models

URL: View paper

Brief Assessment

Factual Answers Assessment[63] focuses on assessing factual reliability in knowledge graphs and language models through error detection and evaluation frameworks, not on automatic benchmark generation for knowledge interventions or model editing.

3. Towards Dynamically Generated KGQA Benchmark Datasets for Memorization-Resistant Evaluations

URL: View paper

Brief Assessment

Dynamically Generated KGQA[67] focuses on KGQA benchmark generation for memorization-resistant evaluations, while the original paper develops a pipeline specifically for evaluating knowledge interventions (editing/unlearning) in LLMs with hierarchical probes across root, intermediate, and leaf levels.

4. Benchmarking large language models in complex question answering attribution using knowledge graphs

URL: View paper

Brief Assessment

Complex QA Attribution[64] focuses on evaluating attribution in question answering using knowledge graphs, not on generating benchmarks for evaluating knowledge interventions or model editing as in the original paper.

5. GAPS: A Clinically Grounded, Automated Benchmark for Evaluating AI Clinicians

URL: View paper

Brief Assessment

GAPS[66] focuses on clinical AI evaluation using medical guidelines and evidence neighborhoods, not on knowledge graph interventions or model editing. The pipeline generates clinical questions and rubrics, not probes for testing knowledge modifications in LLMs.

6. Biokgbench: A knowledge graph checking benchmark of ai agent for biomedical science

URL: View paper

Brief Assessment

BioKGBench[61] focuses on constructing a benchmark for biomedical agents using an existing knowledge graph (CKG) with manually crafted question templates and expert annotations, rather than developing an automatic pipeline that transforms KG datasets into dynamic hierarchical benchmarks for evaluating knowledge interventions at multiple scales.

7. Towards verifiable generation: A benchmark for knowledge-aware language model attribution

URL: View paper

Brief Assessment

Verifiable Generation[60] focuses on generating benchmarks for knowledge attribution from knowledge graphs to evaluate citation quality in LLM outputs. The original paper's pipeline generates hierarchical probes for evaluating knowledge interventions (editing/unlearning), not attribution tasks. These are fundamentally different evaluation objectives.

8. A comprehensive study of knowledge editing for large language models

URL: View paper

Brief Assessment

Comprehensive Editing Study[4] focuses on evaluating existing knowledge editing methods rather than generating benchmarks. While it constructs a benchmark (KnowEdit), the paper does not describe an automatic pipeline that transforms knowledge graphs into hierarchical probes across root, intermediate, and leaf levels as claimed in the original contribution.

9. LENS: Layers of Evaluation of Hallucination in GenAI Systems

URL: View paper

Brief Assessment

LENS[68] focuses on hallucination evaluation through hierarchical query decomposition for GenAI systems, not on knowledge graph-based benchmark generation for evaluating knowledge interventions in LLMs. The technical domains and objectives are fundamentally different.

10. RARE: Retrieval-Aware Robustness Evaluation for Retrieval-Augmented Generation Systems

URL: View paper

Brief Assessment

RARE[62] focuses on retrieval-augmented generation (RAG) robustness evaluation using knowledge graphs to generate questions from domain-specific documents. The original paper addresses knowledge editing and unlearning in LLMs using knowledge graphs to create hierarchical probes for evaluating knowledge interventions. These are distinct application domains with different evaluation objectives.

Contribution 3: Empirical insights on LLM knowledge updating mechanisms

Description: Through extensive experiments across multiple model families and domains, the authors uncover fundamental properties of knowledge updating in LLMs, including propagation asymmetry, plasticity scaling laws, consistency-capacity tradeoffs, subject-dependent update behavior, and unified failure modes. These findings reveal how editing and unlearning differ in their effects on model knowledge.

This contribution was assessed against **1 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Theories of error back-propagation in the brain

URL: View paper

Brief Assessment

Error Back-Propagation Brain[59] focuses on biological neural networks and how error back-propagation might be implemented in the brain through local plasticity rules. This is fundamentally different from the original paper's empirical study of knowledge editing and unlearning in large language models, which examines propagation asymmetry, plasticity scaling laws, and consistency-capacity tradeoffs in artificial neural networks.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] KnowledgeSmith: Uncovering Knowledge Updating in LLMs with Model Editing and Unlearning View paper
- [1] Knowledge editing for large language models: A survey View paper
- [2] Knowledge editing on black-box large language models View paper
- [3] Stable knowledge editing in large language models View paper
- [4] A comprehensive study of knowledge editing for large language models View paper
- [5] Knowledge mechanisms in large language models: A survey and perspective View paper
- [6] Easyedit: An easy-to-use knowledge editing framework for large language models View paper
- [7] Wise: Rethinking the knowledge memory for lifelong model editing of large language models View paper
- [8] Editing conceptual knowledge for large language models View paper
- [9] Give us the facts: Enhancing large language models with knowledge graphs for fact-aware language modeling View paper
- [10] Precision knowledge editing: Enhancing safety in large language models View paper
- [11] Editing factual knowledge in language models View paper

- [12] Towards understanding factual knowledge of large language models View paper
- [13] Identifying Knowledge Editing Types in Large Language Models View paper
- [14] Knowledge graph enhanced large language model editing View paper
- [15] Unveiling the pitfalls of knowledge editing for large language models View paper
- [16] Knowledge in superposition: Unveiling the failures of lifelong knowledge editing for large language models View paper
- [17] Anyedit: Edit any knowledge encoded in language models View paper
- [18] A survey on symbolic knowledge distillation of large language models View paper
- [19] Measuring and modifying factual knowledge in large language models View paper
- [20] Check your facts and try again: Improving large language models with external knowledge and automated feedback View paper
- [21] A Survey on Unlearning in Large Language Models View paper
- [22] Resolving editing-unlearning conflicts: A knowledge codebook framework for large language model updating View paper
- [23] Forgetting before learning: Utilizing parametric arithmetic for knowledge updating in large language models View paper
- [24] Survey on factuality in large language models: Knowledge, retrieval and domain-specificity View paper
- [25] Knowledge updating? no more model editing! just selective contextual reasoning View paper
- [26] Collaboratively adding new knowledge to an LLM View paper
- [27] Extending contextual length and world knowledge generalization in large language models View paper
- [28] Towards Lifelong Learning of Large Language Models: A Survey View paper
- [29] How to make llms forget: On reversing in-context knowledge edits View paper
- [30] Towards continual knowledge learning of language models View paper
- [31] Carpe diem: On the evaluation of world knowledge in lifelong language models View paper
- [32] Efficient conceptual knowledge removal in large language models: Methods and evaluations View paper
- [33] Eva-kellm: A new benchmark for evaluating knowledge editing of llms View paper
- [34] Cogmg: Collaborative augmentation between large language model and knowledge graph View paper
- [35] Neighboring Perturbations of Knowledge Editing on Large Language Models View paper
- [36] How new data permeates LLM knowledge and how to dilute it View paper
- [37] Codeupdatearena: Benchmarking knowledge editing on api updates View paper
- [38] Drowzee: Metamorphic Testing for Fact-Conflicting Hallucination Detection in Large Language Models View paper
- [39] Large language models-guided dynamic adaptation for temporal knowledge graph reasoning View paper
- [40] Knowledge-empowered, collaborative, and co-evolving AI models: The post-LLM roadmap View paper
- [41] Evowiki: Evaluating llms on evolving knowledge View paper
- [42] Bring your own knowledge: A survey of methods for llm knowledge expansion View paper
- [43] SWEA: Updating Factual Knowledge in Large Language Models via Subject Word Embedding Altering View paper
- [44] To Forget or Not? Towards Practical Knowledge Unlearning for Large Language Models View paper
- [45] Llms as repositories of factual knowledge: Limitations and solutions View paper
- [46] Enrich Robots with Updated Knowledge in the Wild via Large Language Models View paper
- [47] Way to specialist: Closing loop between specialized llm and evolving domain knowledge graph View paper
- [48] Unveiling Knowledge Utilization Mechanisms in LLM-based Retrieval-Augmented Generation View paper
- [49] Collaborative Framework for Dynamic Knowledge Updating and Transparent Reasoning with Large Language Models View paper
- [50] Seeking and updating with live visual knowledge View paper
- [51] Rethinking machine unlearning for large language models View paper
- [52] Co-occurrence is not factual association in language models View paper
- [53] UniErase: Towards Balanced and Precise Unlearning in Language Models View paper
- [54] MQuAKE: Assessing Knowledge Editing in Language Models via Multi-Hop Questions View paper
- [55] MMUnlearner: Reformulating Multimodal Machine Unlearning in the Era of Multimodal Large Language Models View paper
- [56] Structured Knowledge Integration and Memory Modeling in Large Language Systems View paper
- [57] UniErase: Unlearning Token as a Universal Erasure Primitive for Language Models View paper
- [58] Lifelong learning of large language model based agents: A roadmap View paper
- [59] Theories of error back-propagation in the brain View paper
- [60] Towards verifiable generation: A benchmark for knowledge-aware language model attribution View paper
- [61] Biokgbench: A knowledge graph checking benchmark of ai agent for biomedical science View paper
- [62] RARE: Retrieval-Aware Robustness Evaluation for Retrieval-Augmented Generation Systems View paper
- [63] Assessing and Improving Factual Answers from Knowledge Graphs and Language Models View paper
- [64] Benchmarking large language models in complex question answering attribution using knowledge graphs View paper
- [65] A knowledge-graph-based intrinsic test for benchmarking medical concept embeddings and pretrained language models View paper
- [66] GAPS: A Clinically Grounded, Automated Benchmark for Evaluating AI Clinicians View paper
- [67] Towards Dynamically Generated KGQA Benchmark Datasets for Memorization-Resistant Evaluations View paper
- [68] LENS: Layers of Evaluation of Hallucination in GenAI Systems View paper