Temporal Minds: A Time-Bound Conversational System with Data Unlearning

Sheetal Sharma and **Divita Phadakale** and **Onkar Apte** and **Nirmit Karkera**
Department of Computer Science
University of Colorado Boulder
Boulder, US
{Sheetal.Sharma, Divita.Phadakale, Onkar.Apte, NirmitShekar.Karkera}@colorado.edu

Abstract

We propose Temporal Minds, a time-aware conversational AI system featuring dual historical personas, each constrained to knowledge available only up to their respective eras (e.g., Aristotle and Alan Turing). Our system integrates era-specific Knowledge Graphs and Retrieval-Augmented Generation (RAG) pipelines to support historically grounded, context-rich dialogue. To ensure persona fidelity, we implement a data unlearning module that filters anachronistic content and models era-specific forgetfulness. We introduce novel evaluation metrics for temporal consistency and realism, and demonstrate that our framework significantly improves the historical accuracy and contextual coherence of persona-driven conversations.

1 Introduction

The task addressed in this work is the development of a time-bound conversational AI system that embodies two distinct historical personas—such as Aristotle and Alan Turing—each constrained to knowledge available only up to their respective eras. We propose a novel framework, Temporal Minds, that combines Knowledge Graphs (KGs) and Retrieval-Augmented Generation (RAG) pipelines to deliver historically accurate, context-rich responses. Additionally, the system incorporates a data unlearning module that actively removes or avoids anachronistic knowledge which a persona should not possess, ensuring historically faithful simulation.

The motivation for this task arises from the increasing use of conversational agents . Despite their popularity, most existing AI systems are trained on large, undifferentiated corpora, making them prone to temporal hallucinations and generating responses with knowledge inconsistent with the supposed era of the persona. While prior work has investigated knowledge filtering and data unlearning, few systems have tackled the challenge of temporally grounded dialogue. Our approach uniquely focuses on both era-specific knowledge grounding and intentional forgetfulness, aiming to build more authentic historical conversational agents.

This task presents multiple challenges. Designing an efficient and flexible knowledge base schema that supports both structured reasoning and fluid conversation is inherently complex, especially when temporal constraints must be enforced. Furthermore, recognizing the user's intent becomes more difficult when historical context and era-specific semantics influence interpretation. Another challenge lies in dynamically filtering out-of-era information, especially when modern pretrained models may infer or hallucinate content based on associations beyond the persona's time. Finally, defining robust evaluation metrics—such as anachronism detection rate, knowledge consistency, and user realism scores—poses its own difficulty. To mitigate these issues, we plan to use temporally tagged datasets, train custom intent classifiers, and implement a filtering module based on entity-time mappings. If any component under performs, we will adopt fallback strategies such as rule-based filters, and prompt tuning to iteratively improve system performance.

2 Related Work

Evaluating Machine Unlearning: Applications, Approaches, and Accuracy(2024) by Zulfiqar Ali, Asif Muhammad, Rubina Adnan, Tamim Alkhalifah, and Sheraz Aslam. This paper investigates various machine unlearning (MU) approaches, focusing on their accuracy and potential applications. The authors present experiments showing that a data-driven approach is the most efficient in terms of both time and accuracy, achieving high precision with minimal training epochs. The problem addressed in this paper is closely related to ensuring data unlearning in the Temporal Minds project, where knowledge is actively forgotten based on time-bound personas. However, our project involves incorporating RAG pipelines and knowledge graphs, which adds complexity compared to the straightforward unlearning of models in this paper.

Machine Unlearning for Traditional Models and Large Language Models: A Short Survey (2018) by Yi Xu. This survey paper explores machine unlearning in both traditional models and large language models (LLMs), offering a detailed classification of unlearning methods and their evaluation criteria. It also discusses the challenges of implementing machine unlearning in different environments and proposes solutions to overcome these challenges. The Temporal Minds project shares a similar focus on LLMs and their ability to forget knowledge, particularly when it concerns data associated with specific time periods. The difference lies in the integration of dual-persona systems and historical accuracy, which is a novel aspect not covered in this survey.

An Overview of Machine Unlearning (2020) by Chunxiao Li, Haipeng Jiang, Jiankang Chen, Yu Zhao, Shuxuan Fu, Fangming Jing, and Yu Guo. This paper provides a comprehensive overview of machine unlearning, discussing its formulation, process, and various challenges involved in implementing unlearning algorithms. It also presents potential applications of machine unlearning across different domains. While the paper offers a solid foundation for understanding unlearning, the Temporal Minds project introduces an additional challenge by ensuring the separation of knowledge based on time-bound personas, which is a more complex scenario than general unlearning tasks.

Learn to Unlearn: Meta-Learning-Based Knowledge Graph Embedding Unlearning" (2024) by Naixing Xu, Qian Li, Xu Wang, Bingchen Liu, and Xin Li. This paper proposes a Meta-Learning-Based Knowledge Graph Embedding Unlearning framework (MetaEU) designed specifically for knowledge graph (KG) embedding unlearning. By leveraging meta-learning, the method adapts the embeddings to unlearn specific knowledge while preserving the performance on remaining data. The Temporal Minds project shares a similar emphasis on using knowledge graphs but goes beyond the unlearning of graph embeddings by applying a dual-persona framework to simulate historical personas. This adds an additional layer of complexity to the unlearning task, integrating both time-bound knowledge and persona-specific data retention.