

COMPUTER AND CONTROL ENGINEERING

AMMIN/DAUIN - Agentic AI for Advanced Temporal Reasoning

Funded By	Dipartimento DAUIN Politecnico di TORINO [P.iva/CF:00518460019]
Supervisor	CAGLIERO LUCA - luca.cagliero@polito.it
Contact	CAGLIERO LUCA - luca.cagliero@polito.it CHIUSANO SILVIA ANNA - silvia.chiusano@polito.it
Context of the research activity	Addressing Temporal Reasoning using Large Language Models (LLMs) requires understanding not only the general concepts of time and time relations, such as ordering or duration, but also more intricate aspects, such as task planning or causal relation discovery. The scholarship aims to explore the use of Agentic frameworks based on Multimodal LLMs to manage complex temporal aspects and effectively and efficiently address challenges related to content misalignment and long-context reasoning. The research activities will be carried out under the FONDO ITALIANO PER LA SCIENZA 2022 – 2023 (BANDO FIS 2) National Project “TA-LLM - Large Language Models: a matter of time?” funded by MUR. CUP: E53C25001820001
	Context Temporal Reasoning (TR) requires a combination of various skills including mathematical and logical reasoning as well as commonsense knowledge [1,2]. Addressing TR on multimodal databases including, for instance, time series, audio tracks, documents, and images, involves training and fine-tuning multimodal learning architectures, such as transformers and Multimodal Large Language Models (MLLMs), to incorporate the notion of time and effectively handle time-evolving scenarios. Challenges Multimodal content misalignment: in real-world applications paired inputs in different modalities (e.g., an audio track and a video content, a time series and its textual annotations) are sometimes partially misaligned due to, for instance, the presence of subtle errors in the audio speech transcription, issues in the audio-video synchronization, or inaccuracies in the human annotations. Since state-of-the-art Multimodal LLMs are usually trained on aligned image-text pairs, the quality of the MLLM outputs could degrade in

the presence of content misalignment. Long context: incorporating historical data in time-evolving scenarios entails prompting MLLMs with long-context inputs, which are either costly using large proprietary models or practically unmanageable using small ones. Tool calling for mathematical reasoning: to efficiently address mathematical reasoning for TR, a common practice is to use LLM calls to external tools. However, as a matter of fact, most MLLMs are specialized neither for TR nor for coding.

Research objectives

Benchmark state-of-the-art Agentic AI frameworks on TR tasks;
Propose innovative approaches to detect and mitigate misalignment issues in paired multimodal inputs;
Explore the adoption of recent approaches to long-context modeling and reasoning (e.g., [4,5]) for TR;
Develop Agentic AI and Reinforcement Learning solutions incorporating innovative TR approaches;
Apply the proposed solutions in real-world application domains (finance, IoT, predictive maintenance, telecommunications, energy).

Objectives

Tentative work plan

During the first year, the PhD student will study existing AI agent frameworks suited for TR tasks. Focusing on the issue of content misalignment between modalities in multimodal sources, the goal is to also propose ad hoc testing benchmarks, drift detection mechanisms, and mitigation strategies.

In the second year, the PhD student will extend the study of Agentic AI methods, specifically addressing the issues of long-context modeling and reasoning for TR, exploring the application of novel methods to real-world scenarios.

In the last year, the PhD student will propose new AI agents and RL approaches to tackle TR tasks and further investigate their application in real industrial scenarios.

Bibliography

- [1] Large Language Models Can Learn Temporal Reasoning. Siheng Xiong, Ali Payani, Ramana Kompella, Faramarz Fekri. ACL 2024. <https://aclanthology.org/2024.acl-long.563.pdf>
- [2] Timedial: Temporal commonsense reasoning in dialog. Lianhui Qin, Aditya Gupta, Shyam Upadhyay, Luheng He, Yejin Choi, and Manaal Faruqui. ACL 2021.
- [3] H Liu, C Li, Q Wu, and YJ Lee. Visual instruction tuning. CoRR, abs/2304.08485, 2023. <https://doi.org/10.48550/arXiv.2304.08485>
- [4] Recursive Language Models. Alex L. Zhang, Tim Kraska, Omar Khattab. <https://arxiv.org/abs/2512.24601>
- [5] Xixi Wu, Kuan Li, Yida Zhao, Liwen Zhang, Litu Ou, Huifeng Yin, Zhongwang Zhang, Xinmiao Yu, Dingchu Zhang, Yong Jiang, Pengjun Xie, Fei Huang, Minhao Cheng, Shuai Wang, Hong Cheng, Jingren Zhou. ReSum: Unlocking Long-Horizon Search Intelligence via Context Summarization. <https://arxiv.org/abs/2509.13313>

List of possible publication venues

- Conferences: ACL, EMNLP, ACM Multimedia, NeurIPS, AAAI, KDD, IEEE ICDM, ECML PKDD, ACM CIKM
- Journals: IEEE TKDE, ACM TKDD, IEEE TAI, ACM TIST, IEEE/ACM TASLP, TACL

**Skills and
competencies
for the
development of
the activity**

The PhD candidate is expected to

- Have the ability to critically analyze complex systems, model them and identify weaknesses;
- be proficient in Python programming;
- know data science fundamentals;
- have a solid background on machine learning and deep learning;
- have natural inclination for teamwork;
- be proficient in English speaking, reading, and writing.