Towards Cognitively-aligned AI with Analogical Reasoning Abilities

Mohammadhossein Khojasteh
Vrije Universiteit Amsterdam
m.khojasteh@vu.nl

Abstract

Analogical reasoning, the ability to map different domains using their structural similarity, is a core mechanism of human cognition that supports tasks such as problem-solving, argumentation, and creativity. Despite recent advances in AI systems, including Large Language Models (LLMs), they still struggle to perform analogical reasoning reliably. In this project, we get inspiration by cognitive theories of analogy and try to develop models that can better capture and perform analogical reasoning. As a key component, we investigate abstraction levels and study how to best extract and represent them. In addition, we systematically analyze the limitations of LLMs by developing a dataset annotated with analogical structure and explore fine-tuning approaches to improve their analogical reasoning abilities.

1 Introduction

Analogical reasoning is the process of identifying a common structure between two situations or domains and using that connection to draw further conclusions or insights [Gentner and Smith, 2012]. The use of analogical reasoning varies from simple everyday learning to complex decision-making, explanation, creativity, and even interpreting metaphors [Penn *et al.*, 2008]. Analogy is considered a core process of human cognition due to its broad applicability [Gentner and Smith, 2012].

In recent years, LLMs have achieved human-level or superior performance in various tasks [Achiam *et al.*, 2023]. However, analogical reasoning remains a challenge for them. Several recent studies in different areas, including evaluations of LLMs on variant analogy problems [Lewis and Mitchell, 2024], psychological research [Stevenson *et al.*, 2024], and analyses of their internal representations [Opiełka *et al.*, 2025], have shown that LLMs lack the ability to perform analogical reasoning in the way humans do. Furthermore, researchers have developed datasets that evaluate analogical reasoning in narratives [Sourati *et al.*, 2024]. Narratives reflect real-life situations, are multi-layered, and can be abstracted, properties that make them especially valuable for studying analogical reasoning [Sourati *et al.*, 2024]. However, LLMs perform poorly on far analogies in these benchmarks, often below human level, and sometimes even below random performance. Meanwhile, cognitively inspired frameworks for analogical mapping and reasoning, such as SME (Structural Mapping Engine) [Forbus *et al.*, 2017] and FAME (Flexible, scalable analogy mappings engine) [Jacob *et al.*, 2023], are designed only for structured inputs and cannot be directly applied to unstructured data or diverse analogical reasoning tasks.

The main goal of this research project is to further investigate analogical reasoning in AI models and develop methods with improved analogical ability. A model that can perform analogical reasoning in a human-like manner could be highly beneficial, either by directly solving complex tasks or by serving as a module within more complex systems of machines and humans. Based on this ambition, we outline our main research questions:

RQ1: Which LLM-derived narrative representations enable cognitively inspired frameworks to perform analogical reasoning in narratives beyond surface similarity?

RQ2: How do explicitly annotated analogical structures in narratives support more systematic evaluation of analogical reasoning in LLMs?

RQ3: How can we design effective abstraction representations in terms of their level of generalization and scope for analogical reasoning?

RQ4: How does fine-tuning on structured analogical reasoning tasks shape LLMs' ability to generalize, transfer relational structures, and avoid surface-level reasoning?

2 Contributions

This research focuses on analogical reasoning in narratives, aiming to improve LLMs while gaining a deeper understanding of how they perform this fundamental ability. We evaluate this ability through target-selection accuracy and by analyzing intermediate steps of analogical reasoning. In contrast to end-to-end baselines such as prompt-based LLMs, we aim to develop modular, cognitively inspired models that make the reasoning process more transparent, integrate insights from cognitive theories, and reveal where and why failures occur. The first objective of this work is to develop a cognitively inspired framework for analogical reasoning over narratives, combining cognitive theories of analogy with the strengths of

LLMs. This work has been completed and is currently under review [Khojastah *et al.*, 2026]. In this paper, we propose a modular framework for analogical reasoning that follows a step-by-step pipeline inspired by cognitive theories. Each narrative is first decomposed into events, which are then used to derive four levels of abstraction. These abstractions are subsequently used for structurally mapping stories, following cognitive theories of analogy. The framework provides a structured alternative to prompt-based methods by explicitly identifying events and capturing their general meaning through abstraction levels, offering clearer insight into the reasoning process while revealing key limitations and challenges.

The second objective is to develop a dataset carefully annotated with analogical structure and use it to evaluate LLMs. In our first paper, we identified several limitations in existing datasets, motivating the need for a new benchmark that addresses these issues and includes explicit structural mappings. To this end, we will combine our previously developed framework with human annotations to capture the structural alignment between narratives. The resulting dataset will provide a well-structured resource for studying and evaluating analogical reasoning. The third objective focuses on abstraction. In our first paper, we showed that abstractions are crucial for analogical reasoning, but our four-level design is only an initial step and leaves room for improvement. First, additional abstraction dimensions, such as moral aspects, should be explored. Second, we will investigate deriving abstractions jointly from pairs of stories rather than from a single story. While individual events may allow multiple valid abstractions, considering both stories together may lead to more consistent representations. We will study how to extract such joint abstractions and how to adapt our structural mapping framework to incorporate them. In the final objective, we aim to improve LLMs themselves through fine-tuning, to strengthen their analogical reasoning ability. Recently, Reasoning Language Models (RLMs) have demonstrated improved performance on reasoning tasks [Besta *et al.*, 2025]. We plan to leverage this technique together with our previously developed dataset to enhance analogical reasoning in LLMs and evaluate how this impacts their performance on complex tasks, such as moral reasoning and creativity.

3 Future Works and Impacts

The goal of this research is twofold: first, to better understand how both humans and AI models perform analogical reasoning, and second, to improve the analogical ability in AI systems. A key impact of our work is developing methods to enhance their analogical reasoning capabilities beyond black-box models. This research also opens several directions for future work. As we finished our first paper, we found challenges for the state-of-the-art models in different steps of analogical reasoning, such as abstraction and causal modeling, which should be explored further. Beyond research, this work has potential applications in areas such as scientific discovery, more reliable decision-making and problem-solving systems, and educational tools that support learning. We do not anticipate any specific ethical risks beyond those generally associated with AI systems.

References

[Achiam *et al.*, 2023] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

[Besta *et al.*, 2025] Maciej Besta, Julia Barth, Eric Schreiber, Ales Kubicek, Afonso Catarino, Robert Gerstenberger, Piotr Nyczyk, Patrick Iff, Yueling Li, Sam Houliston, et al. Reasoning language models: A blueprint. *arXiv preprint arXiv:2501.11223*, 2025.

[Forbus *et al.*, 2017] Kenneth D Forbus, Ronald W Ferguson, Andrew Lovett, and Dedre Gentner. Extending sme to handle large-scale cognitive modeling. *Cognitive Science*, 41(5):1152–1201, 2017.

[Gentner and Smith, 2012] D. Gentner and L. Smith. Analogical reasoning. In V.S. Ramachandran, editor, *Encyclopedia of Human Behavior (Second Edition)*, pages 130–136. Academic Press, San Diego, second edition edition, 2012.

[Jacob *et al.*, 2023] Shahar Jacob, Chen Shani, and Dafna Shahaf. FAME: Flexible, scalable analogy mappings engine. In Houda Bouamor, Juan Pino, and Kalika Bali, editors, *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 16426–16442, Singapore, December 2023. Association for Computational Linguistics.

[Khojastah *et al.*, 2026] M. Khojastah, Y. Jiang, S. De Giorgis, F. van Harmelen, and F. Ilievski. Enhancing structural mapping with llm-derived abstractions for analogical reasoning in narratives, 2026. Under review at the AI Journal.

[Lewis and Mitchell, 2024] Martha Lewis and Melanie Mitchell. Using counterfactual tasks to evaluate the generality of analogical reasoning in large language models. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 46, 2024.

[Opiełka *et al.*, 2025] Gustaw Opiełka, Hannes Rosenbusch, and Claire E Stevenson. Analogical reasoning inside large language models: Concept vectors and the limits of abstraction. *arXiv preprint arXiv:2503.03666*, 2025.

[Penn *et al.*, 2008] Derek C Penn, Keith J Holyoak, and Daniel J Povinelli. Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and brain sciences*, 31(2):109–130, 2008.

[Sourati *et al.*, 2024] Zhivar Sourati, Filip Ilievski, Pia Sommerauer, and Yifan Jiang. ARN: Analogical reasoning on narratives. *Transactions of the Association for Computational Linguistics*, 12:1063–1086, 2024.

[Stevenson *et al.*, 2024] Claire E Stevenson, Alexandra Pafford, Han LJ van der Maas, and Melanie Mitchell. Can large language models generalize analogy solving like people can? *arXiv preprint arXiv:2411.02348*, 2024.