

# Enhancing Enterprise Knowledge Base Construction with Fine-Tuned Generative Language Models

Liana Mikaelyan<sup>1</sup>

<sup>1</sup>Research Software Development Engineer, Alexandria team, Microsoft Research Cambridge UK

## Abstract

In this talk, we will present our latest work on leveraging the power of generative language models for knowledge base construction. We have fine-tuned a generative LLM to extract entities and their relevant properties from text passages and represent them in a structured JSON format. This task was accomplished by creating a dataset of short passages and corresponding JSON outputs using GPT4, which was then used to fine-tune the OpenLlama 3B model on a single A100 GPU. Our approach has demonstrated superior performance compared to the existing template matching algorithm in Alexandria, both in terms of precision and coverage, as well as extracting a richer set of properties from the text. Furthermore, the addition of new properties to the knowledge base has been significantly simplified. Future work involves exploring ways to improve the generation time as well as investigating other models to further enhance our system's performance

## Keywords

Large Language Model, Generative LLM,

## Biography

Ms. Liana Mikaelyan is a Research Software Development Engineer in the Alexandria team at Microsoft Research Cambridge UK . Before joining Microsoft Research Cambridge she worked on various machine learning projects mainly in speech synthesis and recognition. She completed her MSc in Machine Learning at UCL with a background in mathematics.

---

Workshop on Enterprise Knowledge Graphs using Large Language Models, Oct 22, 2023, Birmingham, UK



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)