4th International Academic Conference on Research in Engineering and Technology

18 – 20 September 2025
Prague, Czech Republic

What are the Internal and External Applications of Conversational Data in Large Language Models?

Haesung Bae

Northern Valley Regional High School, The United States

Abstract

As Large Language Models (LLMs) have become increasingly integrated into the Internet, the collection and use of conversational data from LLM interactions has introduced new potentials and dangers. The paper reviews both internal and external applications of LLM conversational data in light of current privacy policies and practices. Internally, this paper describes uses of conversational data for fine-tuning and model improvement. Conversational data first undergoes a preprocessing step of filtering and labeling, to enhance its quality as a fine-tuning training set. This paper will cover the main types of fine tuning, including supervised fine-tuning, reinforcement learning from human feedback, and multi-turn dialogue supervised fine-tuning. Fine-tuning with preprocessed conversational data has demonstrated great improvements in LLM's performance level, but is heavily dependent on certain elements of conversational data's quality and quantity. Externally, conversational data may be used for user analytic purposes, to refine behavior predictions for both large and small populations. Nevertheless, constraints to accessing high-quality data have slowed the development of conversation-based user analytics. Across both internal and external applications, conversational data introduces new potentials for future applications, but their impact is limited by accessibility constraints. Specifically, this paper highlights a disparity between large and small AI companies in their ability to benefit from conversational data that will only grow if the trends discussed in this paper are not addressed. The areas of improvement that emerge in this paper point towards a future where thoughtful use of conversational data can leverage LLMs' predictive capabilities for user benefits.

Keywords: Artificial intelligence; conversational data; large language models; prediction; privacy