

FINE-TUNING LARGE LANGUAGE MODELS FOR CASE RESOLUTION IN SUPPORT CLOUD

Srikanth Balla

Christian Brothers University, Memphis, TN, USA

ABSTRACT

Large Language Model (LLM) utilization in cloud customer support platforms has opened up new possibilities for case resolution process automation and speedup. Existing research is focused on general language understanding or restricted domain adaptation, disregarding specific needs of real-time high-accuracy support situations. There remains a vast research gap for efficient LLM fine-tuning into domain-level case resolution, especially in Support Cloud platforms, where diverse dynamic user requests require contextual understanding, continuous learning, and fusion with historical case history. This study explores fine-tuning strategies optimized for Support Cloud application use cases with special focus on the use of domain-specific data sets, prompt engineering, and reinforcement learning with human feedback (RLHF) to increase resolution correctness and response coherence. Our approach employs historical ticket logs, resolution pattern learning, and feedback channels to enhance the model's understanding and response to difficult user queries at low latency. The study also addresses hallucination, stale responses, and task drift through the use of real-time data streams and retrieval-augmented generation (RAG) mechanisms. Experiments on key metrics such as resolution rate, customer satisfaction, and average handling time achieve significant improvements over baselines. This paper tries to bridge the gap between generic LLM capability and practical deployment needs in enterprise-class support systems, thus enabling the creation of more intelligent, context-aware, and autonomous case resolution models in the Support Cloud space.

KEYWORDS: *Support Cloud, Large Language Models, Case Resolution, Reinforcement Learning, Fine-Tuning, Domain Adaptation, Customer Support Automation, Prompt Engineering, Enterprise AI, and Retrieval-Augmented Generation.*

Article History

Received: 04 Nov 2019 | Revised: 07 Nov 2019 | Accepted: 09 Nov 2019
