

Leveraging Cloud-based OpenAI's LLMs to Create Learning-as-a-Service (LaaS) Solutions for Culturally Rich Conversational AI: A Study Using the Legacy of Slavery Dataset (Remote Presentation)

Tuesday, 26 March 2024 14:00 (30 minutes)

Abstract:

In scientific applications, integrating artificial intelligence (AI) and machine learning (ML) has revolutionized research methodologies and workflows. This study delves into an innovative application of cloud-based OpenAI's Large Language Models (LLMs) in developing a conversational AI chatbot, drawing exclusively from the culturally significant Legacy of Slavery (LoS) datasets maintained by the Maryland State Archives. This initiative deviates from conventional chatbots that rely on a vast, generalized corpus for training. Instead, it focuses on harnessing the LoS datasets as the sole source for responses, thereby ensuring the authenticity and contextual relevance of the historical content. At the heart of this research are cloud-hosted digital notebooks designed as Learning-as-a-Service (LaaS) solutions. These notebooks are designed to elucidate the methodology behind employing OpenAI's LLMs to engineer a chatbot that not only engages in meaningful dialogues but is also constrained to using verified data from the LoS collection. The intention is to create a chatbot that supports educational and research-focused interactions, offering users insights rooted directly in the archival material. Additionally, the project integrates LangChain agents, such as CSV agents, to empower the chatbot with capabilities for data aggregation and analytical tasks, thereby extending its functionality beyond standard conversational interfaces. A pivotal aspect of this study is the comparative analysis between the outcomes produced by the LLM-based chatbot and those obtained using traditional data analysis and visualization tools like Tableau. This comparative study is essential to assess the effectiveness and accuracy of AI-driven analysis compared to conventional data analysis methods. It aims to illuminate the potential benefits and drawbacks of employing LLMs in scientific and research settings, particularly in the context of historical and cultural data analysis. The convergence of cloud computing and AI in this project exemplifies an innovative approach to digital humanities and archival research. It stands as an exemplar of the possibilities of using AI in the curation, exploration, and dissemination of cultural and historical data. The cloud-based digital notebooks serve as a model for LaaS solutions, showcasing how AI can transform the access, analysis, and dissemination of cultural and historical data. This research contributes significantly to the ongoing discourse on AI-enabled scientific workflows, offering new perspectives on applying ML and Deep Learning techniques in data-rich domains of humanities research. This project, through its unique use of AI, opens up new pathways for interacting with, analyzing, and learning from historical datasets. It demonstrates the transformative potential of AI in reshaping educational and scholarly approaches to digital humanities. The insights gleaned from this study are poised to influence a range of disciplines, promoting a deeper understanding of how AI can be tailored to respect and amplify the nuances of cultural and historical datasets in the digital era.

Primary author: GNANASEKARAN, Rajesh Kumar (the University of Maryland)

Co-author: Prof. MARCIANO, Richard

Presenter: GNANASEKARAN, Rajesh Kumar (the University of Maryland)

Session Classification: Artificial Intelligence (AI)

Track Classification: Track 10: Artificial Intelligence (AI)