

FALL 2024

Lehigh AI is a digital hub for collaborative efforts to increase AI literacy and explore how AI technologies can enhance learning, teaching, research, and community engagement. Supported by the [Center for Innovation in Teaching and Learning](#) and funded by a Lehigh University Future Makers Grant from the [Office of Strategic Planning & Initiatives](#), Lehigh AI creates and sustains an online resource with instructions for classroom activities, teaching modules integrable to multidisciplinary classes, technical how-tos, demonstrations, and advice for leveraging tools based on artificial intelligence.

Project title:

Domain-Specific Dialogue: Fine-Tuning Large Language Models for Enhanced Research Outcomes and Insights

Project Development: in Progress/Prototyped

Contributor 1**Name:**

Yao, Zheng

Role: Staff

Affiliation(s) and department(s):

Energy Research Center, Lehigh University

Contributor 2**Name:**

Landskron, Kai

Role: Faculty

Affiliation(s) and department(s):

Department of Chemistry, College of Arts & Sciences, Lehigh University



Description: *A brief description of the project*

Fine-tuning Large Language Models for enhanced research outcomes and insights with domain-specific dialogue.

Abstract: *A more complete description of the project*

Challenge (or opportunity): Large Language Models (LLMs) have become pivotal technology, enabling machines to understand and generate human-like replies to questions. Popular pre-trained LLMs have a general understanding of language and capture a wide range of linguistic patterns, but may not perform in specific tasks or domains. Specialized training or fine-tuning is needed to help improve their performance and accuracy. Fine-tuning LLMs allows us to customize the model to a specific domain, enable LLMs to better understand domain-specific terminology, jargon, and context, and could provide enhanced and expedited research outcomes and insights. This use-case will be designed to serve as a foundation for building a versatile documents-to-LLM pipeline, capable of being adapted and reused across various academic disciplines. In addition, education opportunities will be provided to participating students with in-depth knowledge of AI and LLMs, enabling them to develop expertise in designing, and applying these powerful tools to drive innovation and discovery across various academic disciplines.

Proposed solution: First, search, collect, extract and organize pertinent references library from multiple domains. Then, utilizing LLMs to generate diverse, relevant, and coherent questions from reference library documents. Fine-tuning will be performed by training the model with a domain-specific dataset through RAG. Finally, the fine-tuned model will be utilized to make predictions and results will be verified.

Delivery: (1) Provide training for participating students; (2) web harvesting tool for domain-specific dialogue; (3) fine-tuned domain specific model; and (4) use case as a foundation across various academic disciplines.

Outcomes: Enhanced and expedited research outcomes and insights.



Toolkit: *A description of what software, hardware, and programming languages were used for the project. Includes a description of relevant repositories, datasets, and code.*

Not Available

Keywords: *Terms/tags/keywords/phrases related to the project.*

large language model, fine tuning, domain-specific, enhanced research

