The 64th Israel Annual Conference on Aerospace Sciences 2025

Leveraging Foundation Model approach in Fluids Mechanics Systems engineering

*Shaul Eliahou-Niv, +Yotam Gardush, +Assaf Shiloach

This paper aims to explore how current Artificial Intelligence (AI) technology contributes to the enhancement and acceleration of systems engineering tasks. While systems engineering has broader roots, aeronautics has been a pivotal area for its application and evolution. The integration of Artificial Intelligence (AI) in systems engineering marks a significant evolutionary step in managing complex projects. AI can optimize system designs and processes, leading to cost savings and efficiency improvements. Foundation Models can perform a wide range of disparate tasks with a high degree of accuracy based on input prompts. Unfortunately, despite all the capabilities listed above, Large language models (LLM's) have significant limitations in field of fluid mechanics. The limitation of LLMs in dealing with tasks in fluid dynamics stems from their lack of specialized knowledge and inability to perform complex numerical simulations required for accurate predictions in this field. Our conjecture is that these limitations arise because these generative models have not been trained or refined with required information and data on fluid dynamics phenomena. In our research we employed several strategies to enhance the accuracy of LLMs, including Prompt Engineering, Retrieval-Augmented Generation (RAG) and Error Detection and Correction. RAG is an AI framework that combines the strengths of traditional information retrieval systems with the capabilities of LLM's. Actually, the RAG system is an information retrieval approach designed to overcome the limitations of using a LLM directly. By combining unique and designated data of the user and the world knowledge with LLM language skills grounded generation is more accurate, up-to-date, and relevant to the user specific needs. Due to the compute cost, data preparation time and the required resources, using RAG without training or fine-tuning is an attractive proposition. However, challenges arise when using large language models for information extraction such as performance with long text. In Prompt tuning one adjusts the content of the prompt that is passed to the model to guide the model to generate output that matches a pattern that was specify. The basic foundation model and its parameters are not edited. Only the prompt input is altered. Prompt-tune model, the underlying foundation model can be used to address different engineering needs without being retrained each time. This reduces the computational needs and inference costs. By using both methods and our unique corpus in the RAG Framework we were able to reduce the error rate of different LLM's by 9.07% - 33.33% depending on the model. In our research we outline a structured approach to building a domain-specific RAG database, applying techniques such as text splitting, cleaning and relevance checking to enhance retrieval effectiveness. Key sections cover the construction and optimization of the RAG system, followed by an evaluation using the LM-Harness framework to assess model accuracy and confidence on system engineering benchmarks. Initial results provide insights into the model's capabilities and limitations, while future work will expand the database, improve relevance filtering, and test larger models to further understand LLM applicability in systems engineering contexts. We hope that this technical report will serve as a starting point for further collaboration and efficient knowledge sharing between reserchers from the fluid mechanics community to bridge the existing gap between the pre-trained models and the expert required information for the accurate rag infrastructure. This can accelerate the onboarding process for quickly collect and distill relevant information, documentation and best practices for accurate querying of the LLMs. This document represents a mid-stage report, setting the foundation for ongoing and future developments in the field.

*Director, Applied Research IAI. Senior Research Fellow, The faculty of Aeronautics Technion
+Undergraduate student, Faculty of Computer Science, HUJI.

This research was made possible thanks to the generous support of the Bernard M. Gordon Center for Systems Engineering at the Technion.