Text to Process Simulation

Francesco de Fusco [a], Mattia Vallerio [a], Flavio Manenti [a]

[a] CMIC Department "Giulio Natta", Politecnico di Milano, Piazza Leonardo da Vinci 32, Milano, 20133, Italy

E-mail: mattia.vallerio@polimi.it

Large Language Models (LLMs) have shown remarkable capabilities in generating code from natural language prompts across general-purpose programming languages. However, their performance is still limited when applied to domain-specific languages used in chemical process simulation environments, such as Aspen Plus or Aspen HYSYS. These tools require structured, syntax-sensitive input and/or specialized code that reflects detailed process knowledge. This research investigates the potential of LLMs to automatically generate simulation models in Aspen from natural language prompts. The approach involves fine-tuning an LLM on a dataset comprising simulations such as chemical reactors, separation systems, and heat exchangers, each paired with a descriptive textual prompt. This supervised training allows the model to learn the syntax, structure, and logic of Aspen simulation files, enabling it to translate process descriptions into executable models. The fine-tuned model is evaluated not only on its ability to replicate training scenarios but also on its generalization to unseen cases. The goal is to establish whether LLMs can assist engineers in rapidly prototyping process simulation models or helping them in troubleshooting them, without requiring deep familiarity with simulation software syntax. An additional layer of innovation is introduced through integration with *Tiresias*, a tool capable of generating surrogate models from simulation outputs. The LLM is prompted to produce Python code snippets that define input variables and conditions required by *Tiresias* for initialization. This creates a seamless, automated pipeline: from natural language to Aspen simulation, and from simulation to surrogate model. This work represents a step toward intelligent modeling environments that combine the power of natural language interfaces with advanced process simulation and modeling tools, aiming to reduce development time and democratize access to complex engineering software.

Keywords: *LLMs, process simulation, surrogate models*