FOPAM 2026
Proceedings of the 3rd Foundations of Process/Product Analytics and Machine Learning
Decatur, USA, July 26-30 (2026)

Sketch2Simulation: Automating Flowsheet Generation Via Multi-Agent Large Language Models

Emma Pajak

Imperial College London, The Sargent Centre for Process Systems Engineering, Department of Chemical Engineering

ABSTRACT

Converting process flow diagrams into complete simulation models remains a persistent bottleneck in process systems engineering (PSE), requiring significant manual effort and simulator-specific expertise. Although advances in diagram interpretation and automated model generation have been made, these tasks are typically addressed in isolation, limiting the automation of end-to-end workflows. This work introduces Sketch2Simulation, a unified computational framework that automates flowsheet generation directly from raw engineering diagrams using a multi-agent large language model (LLM) architecture. The proposed framework integrates three coordinated layers: (i) Diagram Parsing and Interpretation, (ii) Simulation Model Synthesis, and (iii) Multi-level Validation.

In the first layer, multimodal LLM agents extract process semantics, identify unit operations and stream connectivity, and resolve implicit structural features. This information is encoded into a directed graph-based intermediate representation that captures process topology while enforcing simulator-compatible constraints. This intermediate representation serves as a formal interface between diagram interpretation and simulator execution, enabling consistent translation of unstructured visual inputs into simulator-compatible models.

The second layer translates this representation into a simulation model through sequential agents responsible for thermodynamic specification, object instantiation, and operating condition assignment, culminating in simulation execution within Aspen HYSYS. The use of a multi-agent architecture enables decomposition of the workflow into specialised reasoning tasks spanning multimodal interpretation, structured model synthesis, and simulator interaction, improving scalability, interpretability, and robustness compared to monolithic LLM approaches.

The final layer introduces validation at multiple stages, including schema enforcement and an execution-and-correction loop that iteratively resolves runtime errors to ensure model validity. The framework is evaluated across four case studies of increasing complexity, including industrial-scale flowsheets with recycle loops.

Results demonstrate consistent generation of simulation models with high structural fidelity, achieving near-complete recovery of process topology (e.g., connection consistency ≥ 0.93, stream consistency ≥ 0.96). Performance degradation is primarily associated with increased diagram complexity and dense interconnections. This work demonstrates that diagram-to-simulation transformation can be formulated as a unified computational problem, reducing reliance on manual model construction and advancing the digitalisation of PSE workflows. Crucially, this enables faster iteration between conceptual design and simulation, lowering the barrier to deploying high-fidelity models in both research and industrial settings.