



Contribution ID: 163

Type: Poster + Flashtalk

## Open Framework for Synthetic Fraud Datasets via Generative AI: Insights from Industrial Secondment at IBM

This work presents an open framework for generating synthetic transactional datasets, addressing the twin challenges of data scarcity and privacy in fraud research. Conducted as an industry secondment at IBM France Lab Saclay within the SMARThEP Network—a European project fostering collaboration between High Energy Physics and Industry—our approach leverages Generative AI Agents to simulate both legitimate and fraudulent behaviors within banking time series data. Initially, our methodology employed Markov chain-based simulations to generate baseline transactional patterns. However, to capture the nuanced dynamics of real-world activities, we transitioned to an LLM-based Chain of Thought approach, enabling the creation of adaptive fraud scenarios that more realistically mimic complex banking behaviors. The resulting synthetic dataset provides a resource for researchers to develop and benchmark fraud detection methods while mitigating issues related to proprietary data constraints. Future plans include releasing the simulator code and datasets to foster targeted collaboration on developing anomaly detection techniques for fraud detection.

In addition, exploratory thinking on applying these automation principles to High Energy Physics will be presented, particularly for automation during data taking. Drawing a conceptual parallel with previous work on the infrastructure for deployment and evaluation of LHCb trigger configurations, this idea suggests that similar strategies might be adapted to manage complex operational processes in experimental settings.

### AI keywords

Generative AI; Simulation; Automation

**Primary authors:** OLOCCO, Micol (TU Dortmund); Mr FEILLET, Pierre (IBM)

**Presenter:** OLOCCO, Micol (TU Dortmund)

**Track Classification:** Datasets & Ethics