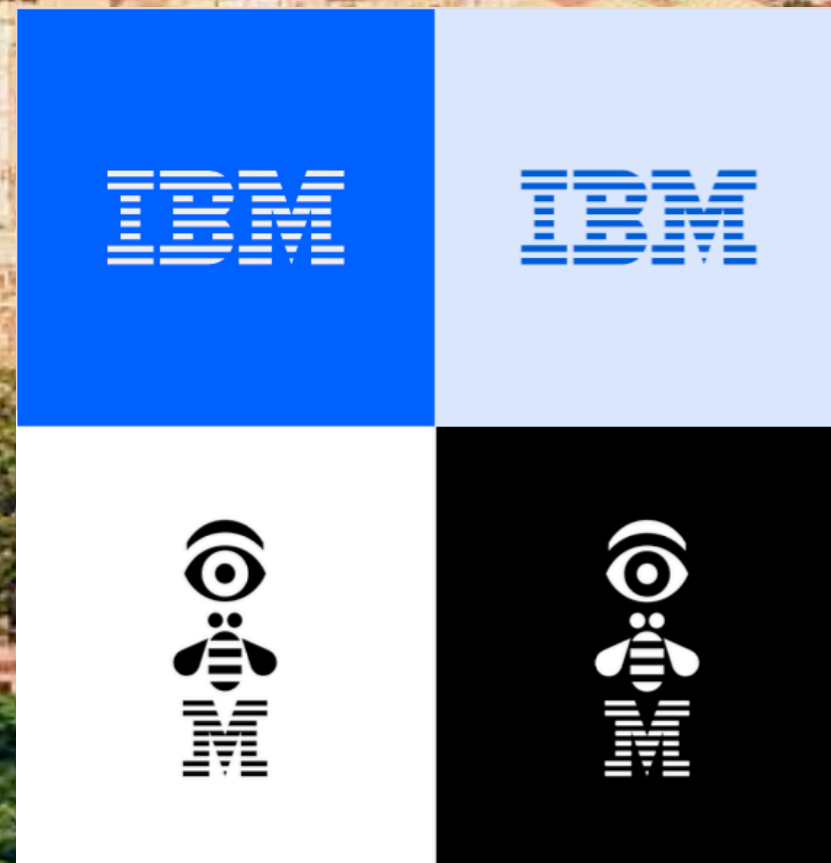




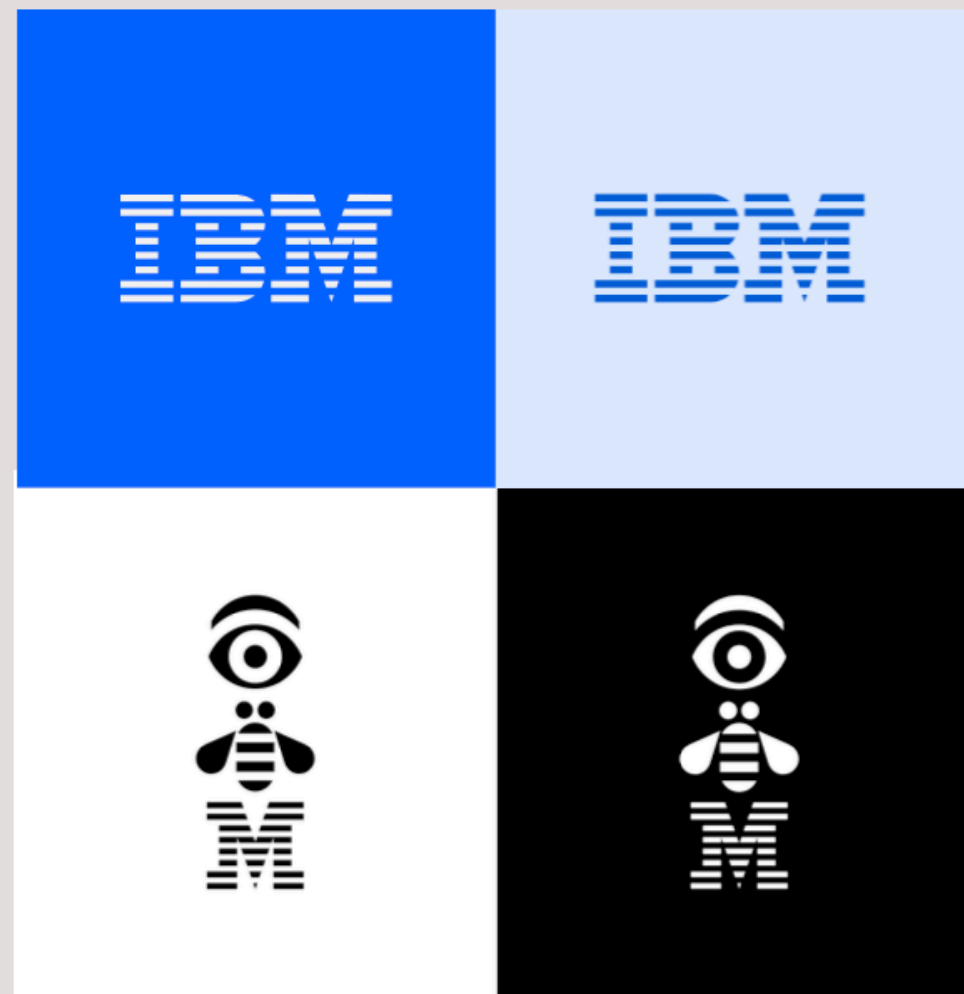
# Open Framework for Synthetic Fraud Datasets via Generative AI



**Micol Olocco**  
Cagliari 15-19 June 2025

European Coalition for AI in Fundamental Physics





# Open Framework for Synthetic Fraud Datasets via Generative AI

Insights from **Industrial Secondment at IBM**

Micol Olocco (TU Dortmund), Pierre Feillet (IBM France Lab Saclay)

micol.olooco@cern.ch, FEILLET@fr.ibm.com



- The [SMARTHEP Network](https://www.smarthep.org/) is a European research project funded by the European Commission, focused on applying **real-time analysis to particle physics and broader industry challenges**.
- SMARTHEP received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement n. 956086

LinkedIn



Pierre Feillet • 1°

Artificial Intelligence Architect for Digital Automation at IBM. Chief A...  
3m • Modificato •

IBM France Lab had the pleasure of welcoming **Micol Olocco** for her PhD secondment as part of the European SMARTHEP project. <https://www.smarthep.org/> project initiative focuses on applying machine learning to High Energy Physics and business challenges. Micol's secondment at IBM France Lab aimed to bridge research between anomaly detection in particle physics collisions and industrial use cases. IBM's involvement is driven by the ambition to explore innovative techniques for fraud detection in business applications.





# Open Framework for Synthetic Fraud Datasets via Generative AI

Insights from Industrial Secondment at IBM

Micol Olocco (TU Dortmund), Pierre Feillet (IBM France Lab Saclay)

[micol.olocco@cern.ch](mailto:micol.olocco@cern.ch), [FEILLET@fr.ibm.com](mailto:FEILLET@fr.ibm.com)



Data scarcity

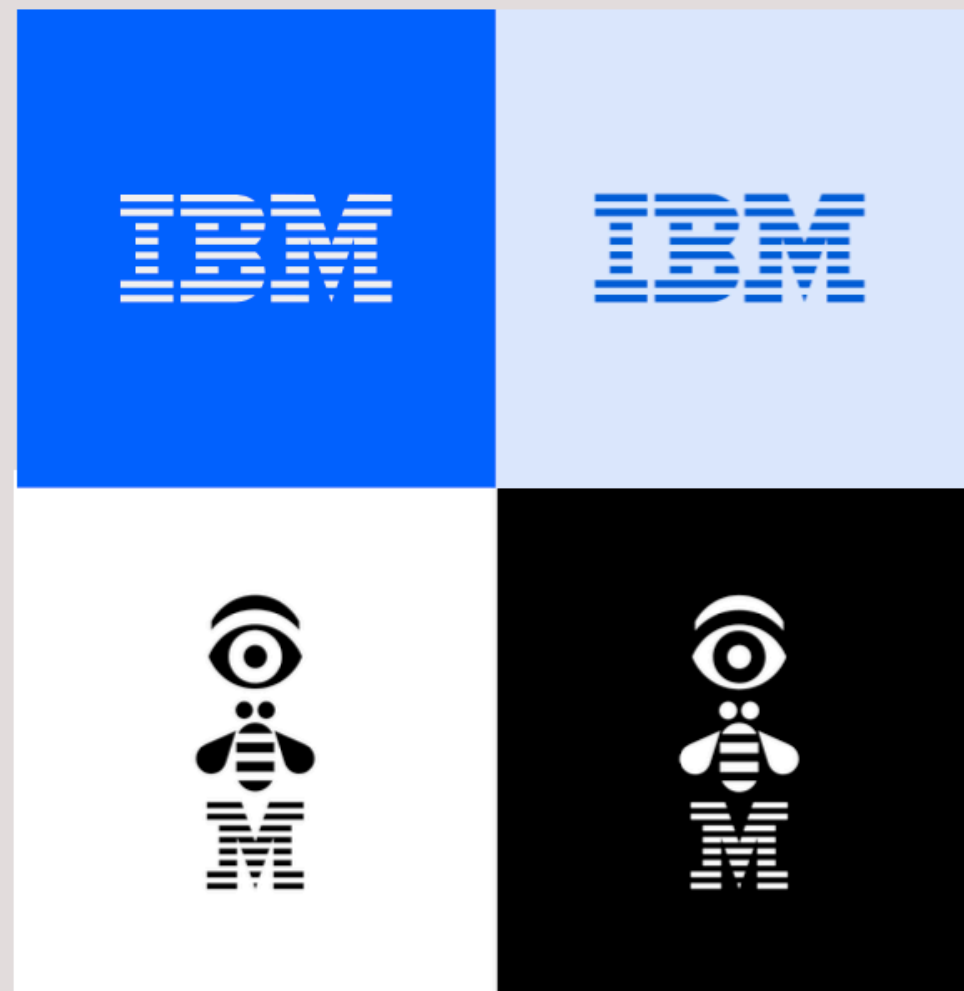
Dataset control

Why a Synthetic Dataset?

Sensitive data

Simulation control





# Open Framework for Synthetic Fraud Datasets via Generative AI

Insights from Industrial Secondment at IBM

Micol Olocco (TU Dortmund), Pierre Feillet (IBM France Lab Saclay)

[micol.olocco@cern.ch](mailto:micol.olocco@cern.ch), [FEILLET@fr.ibm.com](mailto:FEILLET@fr.ibm.com)



Data scarcity

Dataset control

Why a Synthetic Dataset?

Sensitive data

Simulation control

Creativity

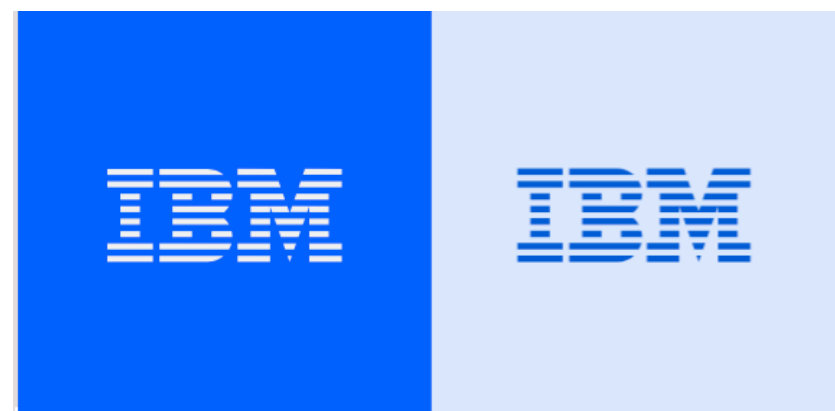
Why Generative AI (specifically LLMs)?

No domain-expertise

Chain-of-thoughts

**It's cool!**

Scalability



# The Simulation Framework

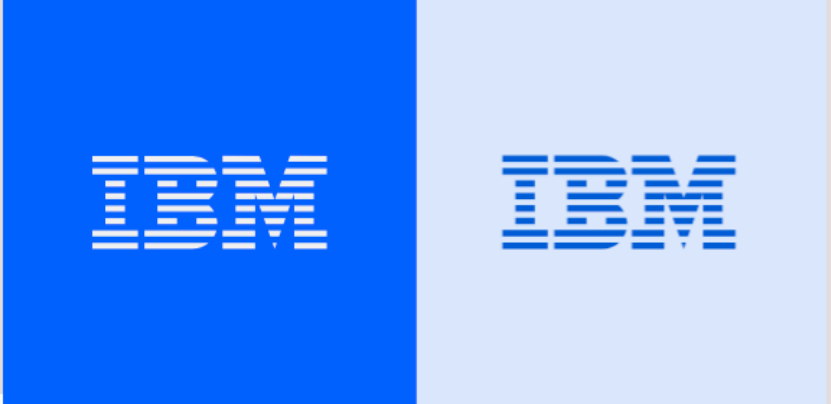
- Build a bank transaction (activity) log
- > Predict the next most probable activity

Bank activity sequence = Markov Chain { Model a huge and sparse transition matrix

Query the LLM to get the next most probable activity based on a user past history







# For each task, its LLM



## Open Framework for Synthetic Fraud Datasets via Generative AI

Insights from Industrial Secondment at IBM

Micol Olocco (TU Dortmund), Pierre Feillet (IBM France Lab Saclay)

micol.olocco@cern.ch, FEILLET@fr.ibm.com



### Motivations

Why an **open-source synthetic** dataset?

- Data scarcity
- Sensitive data
- Flexibility

Why **Large Language Models (LLMs)**?

- Traditional approaches treat a sequence of activities as a Markov Chain, implying a huge and sparse activity transition matrix.
- Idea: replace the probability matrix with an LLM
  - Doesn't require the developer to have domain expertise.
  - Exploit LLM creativity to generate new fraud patterns, possibly.
  - Provide fraud interpretability by analysing the LLM chain-of-thought.

### The Simulation Framework

#### The Strategist LLM

- **Task:** Produce the behavioural pattern of common legitimate and fraudulent profiles (e.g. everyday spender, traveller, identity theft, money laundering).
- **Result:** Catalogue with specifications defining each "strategy", including: number of accounts involved, time between transactions, geographic distribution, transaction amounts, requirement for hijacking, common devices, network types, recipients, and more.
- **LLM requirements:** Creative, discursive, capable of chain-of-thought reasoning.
- **Chosen Model:** llama-3-405b-instructor, served via IBM Watsonx.ai

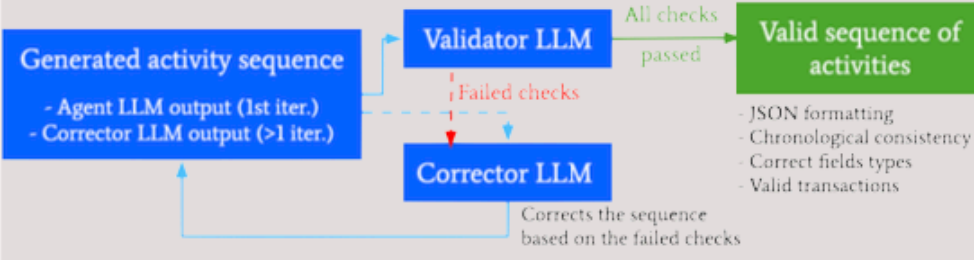
	Deepseek-r1:7b	Mistral:7b
Avg Response Time	5.7 minutes	1 minutes
Min Response Time	48 seconds	18 secondes
Max Response Time	28.3 minutes	2.2 minutes
Failure rate	~40%	~10%

Preliminary performance study for strategy generation. Deepseek-r1:7b VS Mistral:7b run locally via Ollama on MacBook Pro (Apple M1 Pro chip, 16GB memory).

### Fallback systems

To handle incorrect output generation:

- **Manual Corrections:** Implement ad-hoc corrections based on the most observed errors. Eventually, reject the generated sequence and create a new one.
- **Adaptive Learning:** Reinforce the prompt with past errors.
- **Self-correcting system via LLMs:** Generated sequences are validated and eventually corrected by lightweight LLMs. This approach recovers invalid outputs, reducing overall operational costs.



### Applications at CERN

- Simulate realistic insider threats and social engineering (e.g. phishing, impersonation) to test and improve LHC cybersecurity defences.
- Simulate user behaviour across computing and storage systems to predict peak loads and optimise resource allocation and scheduling.
- LLMs can help translate complex or ML-based trigger logic into human-readable explanations, supporting debugging and interpretability.

### The Agent LLM

- **Task:** Given a behavioural strategy (generated by the Strategist LLM) and a user's prior activity history, predict the next most probable activity/chain of activities.
- **Result:** Set of structured activities, each with specifications such as type, timestamp, geographical location, amount, and other contextual attributes. The dataset is constructed by aggregating the predicted activities.
- **LLM requirements:** Concise output, strong JSON structuring, balanced creativity with adherence to prompt.
- **Chosen Model:** mistral-large, served via IBM Watsonx.ai

We acknowledge funding from the European Union Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

## The Strategist LLM

## The Agent LLM

<b>TASK:</b> produce a “strategy” for different legitimate and fraudulent behaviors (ex: everyday spender, traveller / identity theft, money laundering..)	<b>TASK:</b> Given the user profile described by the strategy and the relative account history, predict the next most probable activity/chain of activities.
<b>RESULT:</b> behavior catalogue with specifications about number of accounts, time between transactions, geography distribution, amounts involved, if it involves hijacking, common devices, network types, recipients etc	<b>RESULT:</b> predicted activity with associated type, time, geographical location, amount etc
<b>LLM REQUIREMENTS:</b> creative and discursive, chain-of-thoughts	<b>LLM REQUIREMENTS:</b> concise, good at JSON structuring, good balance between creativity and adherence to prompt
<b>MODEL:</b> llama-3-405b-instructor	<b>MODEL:</b> mistral-large

- Follow up in tomorrow’s poster session for more details and discussions about how we can use LLMs in HEP!