# Large Language Models for external disruption scenarios generation for shop floors' resilience assessment

Tanel Aruväli[1] and Emmanuel Francalanza[1]

[1]University of Malta, Msida MSD 2800, Malta

tanel.aruvali@um.edu.mt

**Short Abstract.** The study contributes to the field of operational resilience management and AI-driven decision support in manufacturing systems. The aim is to investigate how local Large Language Models (LLMs), enhanced with Retrieval-Augmented Generation (RAG), can generate forward-looking disruption scenarios for shop floors under volatile external conditions. A general architecture combining prompting, local LLMs, and RAG based on historical event datasets was developed and tested using a hypothetical manufacturing company, comparing standalone LLM and RAG-augmented setups. Results show that the standalone model successfully identified relevant historical analogs and generated plausible narrative scenarios with qualitative likelihood assessments. However, RAG integration revealed challenges in data structuring, particularly in chunking event-based datasets without losing contextual integrity. The study concludes that local LLMs can support resilience scenario generation, offering practical value for proactive risk assessment. The study presents a novel architecture that integrates LLM reasoning with historical event retrieval, addressing the gaps in forward-looking shop floor resilience analysis.

**Keywords:** RAG, Event Reasoning, Manufacturing Resilience

## 1 Introduction and research purpose

Current resilience assessment approaches on shop floors are largely fragmented, generic, and retrospective, providing limited capability to anticipate future resilience behavior. This limitation is further compounded by the increasing volatility and uncertainty in the external environment, which complicates the making of resilience-oriented strategic investment decisions. Due to external environmental volatility, it is important to focus on risks arising from changes that could weaken resilience in the coming periods.

In the financial sector, a successful example of resilience management is banks' stress-testing approach. The Basel III framework currently in use was developed in response to the 2008 financial crisis. The crisis highlighted many weaknesses in the previous stress test version: a lack of flexibility to respond quickly as the crisis evolved, reliance on historical statistical relationships, failure to cover extreme scenarios, and

reliance on single-event scenarios [1]. The reality showed that many adverse events can occur simultaneously, and historical data alone is insufficient to design adverse scenarios for the stress tests. Instead, emerging risks should be included, and plausible forward-looking multi-event scenarios developed. For instance, in Europe, the European Systemic Risk Board applies a number of models to develop a narrative-based macro-financial stress scenario to assess its impact at the bank level [2].

The capabilities of Large Language Models (LLMs) have significantly advanced over the last few years, enabling access to massive amounts of information and the ability to reason to generate text. LLM combines creativity, which is crucial for generating extreme future scenarios, with analysis of the current context to keep the scenarios realistic and context-based. The Lamp project showed good potential for employing LLM-based reasoning to predict events [3], with the model trained on authentic news datasets. They used abductive reasoning, where a contextualized LLM proposed potential causes for events, matching these patterns to real prior events.

The aim of this work is to explore how local LLMs, supported by Retrieval-Augmented Generation (RAG) functionality and historical event databases, can help generate plausible disruption scenarios for shop floors.

## 2 Methodology

To explore the predictive capability of LLMs in generating plausible external disruption scenarios for shop floors, we developed a general architecture that applies local LLMs with RAG capabilities [see Fig. 1]. This architecture combines prompting with RAG and an LLM, while using external data to create a context for the shop floor, historical news feed patterns, and the current realm. Thus, in prompt generation, in addition to a general query, company-specific knowledge must be included, such as the company's location, field of activity, and type of manufacturing system. This enables targeting a specific geographical location and recognizing events that affect this type of manufacturing, including input components and output products. The vectorized prompt is thereafter sent to RAG knowledge library to find the most significant similarities. RAG is carried out on historical event datasets that enable the identification of past patterns that match the prompt. Event databases, such as ICEWS and GDELT, contain processed data on past events, including actors, event types, geographic locations, sentiment values, tone, links to news articles, and more. For instance, GDELT [4] includes news from 65 different languages and is updated every 15 minutes. Specific data fields selection and preparation for RAG are crucial for making the dataset findable and reflecting its context. RAG preparation includes chunking the content into smaller parts, applying embeddings to convert the data into numerical representations, and storing it in a vector database. The prompt is vectorized in the same way to retrieve the best matches from the vector database. Thereafter, the prompt is enriched with relevant information to provide enhanced context that guides the LLM in analyzing the pattern match between historical events and the current context, laying the groundwork for reasoning about potential upcoming external disruptions.

We conducted the preliminary testing using a combination of software LM Studio version 0.4.9 and Anything LLM v1.12.0. LM Studio provided LLM Gemma-4-26b-a4b-it, which has a knowledge cutoff time of the end of January 2025. Anything LLM provided RAG functionality through all-MiniLM-L6-v2 embedder model and LanceDB vector database. For the RAG knowledge library, a fragment from the GDELT 1.0 first datasets (01.01.1979 – 23.06.1979) was selected and embedded. The prompt requested information for a company in France to identify possible energy-shock-based disruption scenarios. Testing was performed across two distinct setups: one utilizing a standalone local model and another utilizing a local model augmented with RAG.

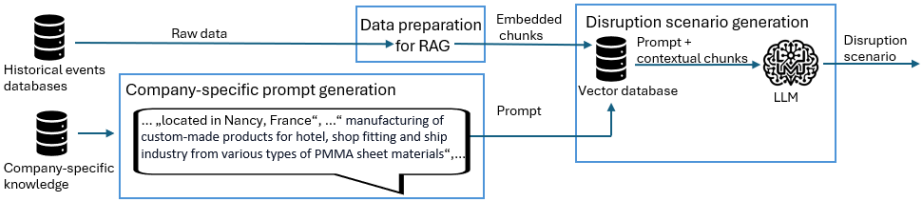



**Fig. 1.** General architecture for the generation of disruption scenarios using LLM and RAG.

## 3 Results and contributions

In the preliminary testing phase, both setups identified the 1979 Iranian Revolution as an historical event which caused the energy shock later that year. The models defined the enablers and the event's subsequent impacts. Both setups provided a comparison of historical findings with the contemporary context, noting where patterns matched or differed between the two periods. By reasoning these comparisons, the models then generated narrative energy-shock-based disruption scenarios and provided estimates on the likelihood of different outcomes for periods after the model's cutoff time.

However, enriching LLM knowledge with RAG-based historical data showed that further study is needed to ensure that the retrieved information and the model's internal logic work together consistently. The difficulty lies in a database-like data structure in which each row represents an individual event. This requires a well-optimized chunking and embedding strategy to avoid breaking data rows while keeping unrelated events in separate chunks.

Overall, preliminary results indicate that local LLMs have the ability to generate reasoned scenarios by pattern-matching past events. In addition, RAG has the potential to enrich reasoning and generate plausible disruption scenarios for the shop floors' forward-looking operational resilience assessment.

## Acknowledgment




This work was funded by the European Union under the Horizon Europe grant number 101207066.

## References

1. Basel Committee on Banking Supervision: Principles for sound stress testing practices and supervision, https://www.bis.org/publ/bcbs155.pdf, (2009).
2. Gross, M., Henry, J., Rancoita, E.: Macrofinancial stress test scenario design—for banks and beyond. In: Farmer, J.D., Kleinnijenhuis, A.M., Schuermann, T., and Wetzer, T. (eds.) Handbook of Financial Stress Testing. Cambridge University Press (2022).
3. Shi, X., Xue, S., Wang, K., Zhou, F., Zhang, J.Y., Zhou, J., Tan, C., Mei, H.: Language models can improve event prediction by few-shot abductive reasoning. In: Proceedings of the 37th International Conference on Neural Information Processing Systems. pp. 29532–29557 (2023).
4. Leetaru, K., Schrodt, P.A.: Gdelt: Global data on events, location, and tone, 1979–2012. In ISA annual convention, 2, pp. 1–49 (2013).