Artificial Intelligence & Machine Learning

November 17-18, 2025 | London, UK

Mattia Beretta [1], Michele Leporati [2], Giuseppe Lo Presti [1], Gabriele Pece [1]

[1] Pirelli & C. S.p.A., Milano, Italy
[2] Università di Milano Bicocca, Department of Informatics, Systems and Communication (DISCo), Milano, Italy

Graph-Based Clustering and Large Language Models for Scalable Summarization of Safety Reports in Manufacturing Environments

Manufacturing environments generate large volumes of safety-related reports covering incidents, near-misses, and unsafe conditions—often written in unstructured natural language. Notably, only a small fraction of these records pertain directly to injuries; the majority highlight contributory factors or unsafe conditions which, if left unresolved, may ultimately result in incidents. Reviewing and summarizing these reports is a time-consuming task. We introduce an automated system that leverages graph-based clustering and Large Language Models (LLMs) to analyze and summarize extensive safety incident documentation efficiently. Developed in collaboration with the Pirelli Digital Solution Center of Bari as part of the "OR6 – Big Data Platform – GenAI platform" initiative, our solution helps Pirelli "Health, Safety and Environment" department identify patterns, emerging risks, and improve workplace safety by promptly and effectively processing safety records. The system constructs similarity graphs using semantic embeddings and applies the Louvain community detection algorithm to group related reports into meaningful clusters. Each cluster is then summarized by an LLM, producing concise overviews and one-line titles that provide rapid insights into core safety issues. The system also delivers basic statistics, such as event counts and time-based distributions, supporting data-driven safety management. Scalability is a key feature: the solution can process years of historical data encompassing thousands of events in just seconds. This allows safety professionals to spot trends, uncover hidden patterns, and reduce individual bias in manual evaluations. By integrating advances in natural language processing and graph theory, our approach offers manufacturing organizations an effective decision-support tool—enhancing situational awareness and supporting continuous improvement for safer workplaces.

Keywords
safety incident reports, workplace safety, text summarization, graph-based clustering, Large Language Models (LLM)