International Civil Aviation Organization
Asia and Pacific Office

Fourteenth Meeting of the Regional Aviation Safety Group – Asia and Pacific Regions (RASG-APAC/14)

(Bangkok, Thailand, 28-29 November 2024)

Agenda Item 4: **ICAO/ Member State / Industry Presentations**

LEVERAGING GENERATIVE PRE-TRAINED LARGE LANGUAGE MODELS TO SUPPORT DATA-INFORMED SAFETY DECISION MAKING IN AVIATION

(*Presented by the United States)*

SUMMARY

This paper presents a novel approach to revolutionize aviation safety management by leveraging the power of Large Language Models (LLMs) to extract safety event information from textual reports. The Federal Aviation Administration's (FAA) is conducting research to address the challenges of manual incident mapping and integrated risk modelling development by employing LLMs to automate and streamline these critical processes. Our prototype solutions demonstrate significant improvements in efficiency, consistency, and transparency, enabling prognostic risk identification and mitigation. Through rigorous verification and validation, we ensure the accuracy and reliability of our LLM-powered tools. Preliminary results highlight the transformative potential of this approach in enhancing aviation safety, paving the way for a more resilient and secure National Airspace System (NAS).

1. INTRODUCTION

1.1 The U.S. Federal Aviation Administration (FAA) is leading an ambitious undertaking involving the integration of innovative technologies and procedures to enhance safety, efficiency, capacity, and environmental sustainability in aviation. However, the introduction of such novel concepts and technologies necessitates a robust framework for proactively identifying and mitigating potential risks.

a) To address this critical need, the FAA aims to augment traditional, retrospective, compliance-based safety decision-making by developing a prognostic, data-informed safety assessment methodology. This shift is crucial as the aviation system, while remarkably safe, still harbours inherent risks that require continuous improvement efforts. The emergence of new entrants also warrants the opportunity for the FAA to evaluate the potential change in risk consequences for the aviation safety system. FAA seeks to identify and address these risks before they can result in accidents or incidents.

b) Central to this prognostic approach is the development of an integrated risk assessment model, a tool designed for data-informed, risk-based decision-making in aviation safety. The traditional model for identifying safety hazards relied upon forensics—learning from past mistakes. The model discussed in this paper improves upon the traditional method by proactively identifying potential hazards. This new model operates on the principle of "Safety II" which shifts the focus from learning solely from past mistakes to also identifying potential hazards through studying successful operations. By mapping historical incidents onto precursor-barrier models, the model gains insights into risk prevalence and the effectiveness of existing safety barriers and controls. These insights are then used to identify precursors that might be detectable and quantifiable in normal flight data, allowing us to understand not only where safety barriers failed, but also where they succeeded in preventing incidents. This shift enables a prognostic approach to risk identification and mitigation, learning from what we do right to further enhance aviation safety.

1.2 The traditional process of manually mapping vast amounts of data to the risk assessment models is a significant bottleneck, hindering the full realization of its potential. Assembling a body of Subject Matter Experts to manually read through thousands of accident and incident reports is an extremely high workload task that takes more resources (funding and personnel) than are normally provided. Natural Language Processing (NLP) provides an economical way to extract safety data from narrative reports.

a) This paper presents a groundbreaking solution to this challenge by harnessing the power of LLMs. We discuss how LLMs can be effectively employed to automate and streamline the incident mapping process, enhancing its efficiency, consistency, and repeatability. Additionally, we showcase the application of LLMs in refining and improving the risk models themselves, leading to increased clarity and comprehensiveness.

b) Combining the risk models with LLMs generates a more efficient and more cost-effective means of calculating the rates of occurrence of precursor events. Another advantage of using NLP is the automatic extraction of the proportion of events resulting in different end states from the entire population of narratives, instead of only using a smaller, manually analyzed sample.

2. DISCUSSION

2.1 The aviation industry’s unwavering commitment to safety necessitates continuous advancement in risk assessment and mitigation strategies. Despite being a data-rich domain, the sheer volume of information has historically overwhelmed conventional analysis methods, limiting the ability of subject matter experts to fully exploit this resource due to the prohibitive manual workload. However, the rapid advancement of Artificial Intelligence (AI) technology now offers a transformative opportunity to overcome these limitations. By integrating AI, we can unlock the potential of this vast data, driving meaningful improvements in safety outcomes.

a) Mapping incidents and occurrences to safety models constructed within risk assessment models plays a pivotal role in leveraging the tool for aviation safety management. It serves as a bridge between the theoretical framework of risk modeling and real-world events, enabling safety professionals to extract actionable insights from historical data within the context of the tool.

b) The primary objective of incident mapping is to identify the specific sequence of events and contributing factors that led to a safety incident or occurrence. By tracing the incident pathway through the precursor-barrier model, analysts can pinpoint the precursors that were not effectively intercepted by the existing safety barriers. This information is crucial for understanding the root causes of incidents and identifying areas where safety enhancements are needed.

c) Moreover, mapping many incidents onto the same model reveals patterns and trends, highlighting the most frequent precursors and the barriers that most often fail to catch and correct deviations. Additionally, through the inclusion of child nodes to identify distinct risk states that lead to a given parent, emergent issues can be easily identified: When multiple mappings with the same characteristics are recorded against that node, the analyst can enhance the safety/risk model to capture the new risk situation. This allows safety professionals to prioritize their efforts and resources, focusing on the areas of greatest risk and vulnerability.

2.2 AI and NLP have advanced remarkably in recent years, culminating in the emergence of LLMs. These sophisticated models, trained on vast datasets of text and code, possess an impressive capacity to understand, generate, and manipulate human language. They have proven adept at a wide array of tasks, from translation and summarization to question answering and creative writing. This versatility has ignited a wave of innovation across numerous fields, and their potential to transform aviation safety management is particularly compelling.

a) LLMs offer a unique set of characteristics that make them well-suited to tackle the complexities of incident mapping and risk assessment. Their ability to process and comprehend large volumes of textual data, combined with their contextual understanding and logical reasoning capabilities, enable them to dissect intricate incident reports, identify key patterns, and uncover causal relationships. Moreover, their proficiency in generating coherent and informative text allow them to produce detailed justifications for their decisions, promoting transparency and facilitating human oversight.

b) The integration of LLMs into safety management processes has the potential to revolutionize the way aviation safety professionals approach risk assessment and mitigation. By automating and streamlining labor-intensive tasks such as incident mapping, LLMs can free up valuable human resources, allowing experts to focus on higher-level analysis and strategic decision-making. Furthermore, LLMs' abilities to process and learn from both historical incidents and normal operations data foster a proactive, data-driven approach to safety management.

2.3 A crucial function of the risk assessment models is the mapping of incident reports onto precursor-barrier models to facilitate risk analysis. We have revolutionized this process by augmenting the models with meticulously crafted True/False statements linked to each precursor and barrier. These statements are designed to be highly specific, concise, unambiguous, and stand-alone, capturing the essence of each node's occurrence or non-occurrence in an incident. Where necessary, we expand acronyms and provide descriptions of technical jargon to ensure clarity and comprehensiveness.

a) The development of these statements is an iterative process, involving extensive testing with numerous incident reports to refine their accuracy and precision. We compare sibling node statements to ensure completeness and avoid overlap, aiming to replicate the decision-making processes of SMEs as closely as possible. This meticulous approach minimizes ambiguity and facilitates the LLM's task in providing accurate and consistent answers.

b) The automated mapping process leverages an LLM-based tool that navigates the risk models by iteratively posing these True/False questions based on the incident report's content. The LLM not only determines the most likely pathway through the model but also generates detailed justifications for each decision, promoting transparency and facilitating human review. To ensure robustness, the mapping process is designed to be idempotent, consistently producing the same results for a given report and set of statements. We rigorously test the statements against various

LLMs, including both large commercial models and open-source options, to ensure compatibility and avoid hallucination.

c) Mapping an individual incident to a risk model provides a retrospective understanding of the causal chain leading to a reportable event. However, when numerous incidents are mapped onto the same model, it transforms into a comprehensive risk landscape. This landscape reveals which undesirable states or precursors occur most frequently and which safety barriers and controls are being most heavily relied upon. This allows analysts to proactively identify areas within the system that may warrant further investigation, confident that sufficient data exists to support a deeper dive into the causal chain.

d) The output of the risk modelling tool offers groundbreaking insights into the distribution of risk within the aviation safety system. For the first time, analysts can efficiently mine historical incident reports to generate quantitative data on the causal factors contributing to safety incidents. This enables a more nuanced understanding of how specific precursors influence safety outcomes and allows for a rigorous assessment of the effectiveness of the safety barriers and controls currently in place.

2.4 Verification and Validation: In safety-critical domains such as aviation, the accuracy and reliability of AI-powered tools are paramount. To ensure the robustness of our LLM-based incident mapping process, we have implemented a two-step verification and validation procedure. First, the FAA developed a synthetic event generation and mapping LLM-powered tool to generate realistic, synthetic incident reports based on predefined mappings through the risk model. By controlling parameters such as report length, severity, location, and aviation jargon complexity, we generate diverse scenarios to thoroughly test the LLM mapping tool. Any discrepancies between the generated mapping and the original mapping used for report generation are flagged for review, ensuring the tool's accuracy and highlighting areas for potential model refinement. Second, we conduct random sampling and review by subject matter experts (SME). To complement the automated validation, we employ a random sampling approach where SMEs review a subset of LLM-generated mappings and their associated justifications. This human-in-the-loop validation ensures that the LLM's decisions align with expert judgment and domain knowledge. The generated justifications facilitate efficient SME review, as they provide clear insights into the LLM's reasoning process without requiring in-depth knowledge of the risk models. SME feedback is then incorporated into the continuous improvement of the LLM mapping tool. This iterative refinement ensures that the system remains aligned with evolving safety standards and expert knowledge.

2.5 Pitfalls and Challenges: States seeking to employ LLM’s using the methods described in this paper should be aware of several pitfalls and challenges that analysts may face when deploying these novel technologies. First, third party models may change without warning. When using public models, idempotency (when operations applied multiple times do not change the result) remains a challenge when we do not control the LLM version. The recommendation to resolve this issue is to self-host a LLM that organizations have full version control of, even if the model is not the best available. Secondly, States should maintain awareness that LLM’s interpret text quite literally and that LLM’s use weighting, not context, to ‘predict’ answers. To give an example, we must be careful to word questions to indicate that aircraft are vehicles and people are in the vehicles. When analyzing safety reports, LLMs will identify aircraft on runways as “vehicles on the runway” and include people inside of vehicles as “people on the runway”. Prompt engineering and specificity when asking questions to LLMs is very important. Iteration of the process requires some knowledge of how the reports are written and common aviation terms. Finally, beware that interpreting the response from the LLM requires the analyst to ensure the LLM is not making “assumptions” based on lack of information in the report. For example, the LLM may not correctly interpret questions about ATC’s detection and correction of unauthorized use of runways if the report does not mention ATC in the narrative. This can lead to incorrect conclusions about incidents.

3. ACTION BY THE MEETING

3.1 The Meeting is invited to:

a) Review this information paper and the FAA welcomes any feedback or interest in further knowledge sharing.
b) Expand reporting sources to provide a richer dataset for all risk modelling tools, allowing for more comprehensive risk modelling and a deeper understanding of everyday operations. Voluntary reporting systems gather a wider range of safety-related data, including near-misses and minor incidents that may not trigger mandatory reporting.
c) Encourage a Positive Reporting Culture: By demonstrating the value of voluntary reports in identifying and mitigating risks, organizations can foster a stronger safety culture where employees feel empowered to report safety concerns without fear of blame. This aligns with an emphasis on collaboration and learning.

— END —