Uncovering Resilient Behavior in the Aviation Safety Reporting System Using Large Language Models

1st Bryan Matthews
KBR, Inc.
NASA Ames Research Center
Mountain View, CA, USA
bryan.l.matthews@nasa.gov

Abstract*—Resiliency is present in everyday life, both in system design and exhibited by the operators that function within these systems. This includes the National Airspace System (NAS) where pilots and controllers make positive decisions and take preventative or corrective actions every day even in unsafe situations. Pilot safety reports filed after an event are rich text narratives that detail the conditions around an event and can provide additional context leading up to, during, and describe how the situation was resolved. This yields useful insights into the resilient positive actions and corrective steps that may have transpired to prevented a safety incident from degrading further. Analyzing large archives of these reports can be impractical for subject matter experts to properly extract evidence of resilient behavior. However, Large Language Models have demonstrated the potential to extract useful insights from extensive bodies of text. This work proposes to utilize the Llama3.1 Instruct model to identify examples of resilient behavior within four categories on over 250,000 narratives from NASA's Aviation Safety Reporting System (ASRS). The analysis will reveal how similar and different resilient behaviors are present within various ASRS anomaly categories such as airborne conflict, near mid air collision, altitude deviation overshoot, runway excursion/incursions, and responses to external factors such as weather turbulence. Additionally, the analysis will compare resilient behavior between general aviation and commercial operation events as well as temporal trends within the archive of reports. The analysis aims to uncover how operators are practicing positive resilient behavior in situations described within the corpus of the report archive. This method provides a new lens into these valuable safety reports that can be used to inform and improve safety monitoring systems from this human resiliency perspective. The benefit can lead to highlighting operator proficiencies within the community and identify any knowledge gaps to ultimately improve safety within the NAS.

***Index Terms*—Aviation Safety,Large Language Models**

I. Introduction

Since the 2020 COVID-19 pandemic the National Airspace System (NAS) has quickly returned to and exceeded the historically high traffic volume levels of the pre-pandemic era. The U.S. Bureau of Transportation Statistics reported that U.S. Airline traffic for July of 2024 hit an all-time high of 91.8M passengers and August of 2024 was up 4.5% for the same month of the previous year [1]. This increasing demand along with other emerging operations that include both piloted and unmanned aircraft requires a need to identify resilient factors that, when present in the system, add to the critical layers of safety that millions of air travelers rely upon. Additionally, quantifying resilient actions performed by humans can further highlight requirements that will be needed for autonomous systems to function safely within a future hybrid human/autonomous airspace. While pilots and air traffic controllers are managing aircraft within the NAS they are constantly facing a variety of pressures that require them to adapt and pivot from their original flight plan to safely navigate aircraft to their destination. These pressures range from environmental factors, to managing traffic separation, to handling issues with equipment and automation. To successfully address the impact of these pressures to the operations, the pilots and controllers depend upon the resiliency of the system to ensure the safety of the NAS. Components of the system's resiliency rely upon standard operating procedures, safety nets such as the traffic collision avoidance system (TCAS) and other warning systems, and human decision making and responses that are both strategic and tactical. Furthermore, human factors such as: effective communication on the flight deck, monitoring for readback errors during air traffic controller communication, or adherence to standard operating procedures are well known skills and qualities that are covered in training and are considered good airmanship. However, situations arise during the operations that require pilots and controllers to lean upon their years of experience and training to address a situation that, if not handled properly, could lead to something more serious such as damage to the aircraft or even injuries or loss of life. Often pilots and controllers will have to adapt to a situation and implement a work around or utilize a countermeasure [2] when automation or procedural failures occur within daily line operations. This points to a need to learn from historical events to better understand which resilient behaviors are present in the existing NAS and how they are associated with different safety critical events. NASA's Aviation Safety Reporting System (ASRS) is an established archive where pilots and controllers voluntarily and anonymously report safety incidents that they have observed on recent flights. The reports are comprehensive narratives that describe the events and conditions leading up to the safety event. Since the reports are narratives and self-reflective in nature, there are insights into the pilot's thinking at the time of the event and the author will often recount lessons learned from their experience. Further examination of an event leading up to and in response to a threat, can yield an understanding of the positive actions that averted a more serious adverse outcome. This approach of viewing a situation is commonly referred to as Safety II [3] as opposed to Safety I where the focus is more on the factors that went wrong during an event. These Safety II aspects of the operations can be analyzed to highlight successful elements with the objective to gain a better understanding and ultimately reinforce these systems and behaviors in future operations. Historically Safety I has focused on the threats, errors, and adverse conditions that lead to a safety incident or accident and aims to better understand how to prevent such an event in the future. However, since adverse safety events are relatively rare in the operations a Safety II perspective has the potential to draw from significantly more examples. Positive actions taken by the operators exemplify resilient qualities that when objectively identified can help augment current safety monitoring systems and ultimately reinforce operator proficiencies. Although ASRS reports are primarily focused on the adverse event that the author is reporting on, these rich text narratives also offer the opportunity to extract evidence of resilient behaviors and characterize them within four constructs: *anticipate, monitor, respond, and learn*. Given the large corpus of narratives that exist within ASRS, this poses a daunting task for a subject matter experts to manually sift through, extract these examples, and organize into actionable recommendations to improve safety. However, recent advances in Large Language Models (LLMs) have demonstrated the ability to sift through large bodies of text and extract out key insights that the prompt engineer has requested. Once the resilient behavior evidence is extracted the models can also be used to summarized and help organize the factors. This knowledge, can in turn, be shared within the aviation safety community and increase the effectiveness of existing safety monitoring systems. This paper will describe the concepts of human resilience and demonstrate how ASRS can be leveraged with the assistance of LLMs to uncover aspects of both similar and different resilient qualities across various operations and incident types.

A. Resiliency

Bertoni [4] describes a resilient framework concept where qualities of human resilience fall under four categories: *anticipate, monitor, respond,* and *learn*. Everyday actions taken by the operators demonstrate these resilient qualities to maintain safety. This is also apparent when work imagined does not always translate to work done such as in the case where decision making is required to manage new developments in the operations. Mapping these actions to their respective resilient categories can better capture the characteristics of work done which can, in turn, uncover positive behavior that is required to continue to maintain a safe NAS. The NAS has continued to grow year over year since the COVID-19 pandemic and with the expected integration of emerging Unmanned Aircraft System (UAS) operations, there is a need to identify what aspects of the current system demonstrates resilience. These behaviors are important to ascertain so that they can be considered when defining new requirements in future concepts of operations.

B. Related Work

One program proposed and implemented by American Airlines is the Learning Improvement Team (LIT) [5]. This framework has the ability to capture pilot proficiencies that fall within the Safety II paradigm. In this program, pilot peer observers take notes in the jump seat during a flight to identify the pressures the pilots face during line operations. The observer is trained to recognize pilot responses and code their proficiencies after the flight is completed. Although the data is a rich source for capturing these resilient qualities, some shortcomings, are that there is a limit to how many observable flights that can be collected every year and not all airlines have the resources to implement such a program. However, the idea proposed by LIT of first capturing a narrative that describes activities within the context of operations and then assessing for coded proficiencies can be extended to other types of narratives to yield other forms of resiliency. In [6] a novel way to examine Aviation Safety Reporting System (ASRS) reports was proposed using sentiment analysis and clustering of positive pilot actions to uncover themes of resilient behavior. However, the approach is unsupervised in the sense that it does not specifically categorize the behavior into a predetermined resilient framework. Advancements in Large Language Models (LLM) have expanded the capabilities to review and analyze considerably large amounts of text in a highly scalable way where it would have previously been infeasible for subject matter experts. Additionally the LLMs have the ability to be instructed with prompts and use context to help guide the output towards the desired objective, making this task potentially more feasible to scale.

C. Aviation Safety Reporting System

NASA's ASRS [7] database contains a repository of nearly 2 million safety reports. The reports are voluntarily submitted by members of the aviation community including pilots, air traffic controllers, and flight attendants from both commercial airline operations [8] and general aviation (GA) [9]. Pilots that file Safety Action Program (ASAP) reports and controllers that file Air traffic Safety Action Program (ATSAP) with their Safety Management System (SMS) can choose to send a de-identified version of the report to ASRS to preserve the anonymity of the people and organizations involved. The reports consist of a text narrative describing a safety event of concern including the conditions leading up to the event, what happened during the event, actions taken by those involved, and the outcome. In addition to the report narrative, demographics such as whether a report involved GA aircraft or a commercial airline. ASRS analysts also provide labels for the types of events as coded anomaly categories that can be used to group the reports for analyzing and understanding common characteristics around specific types of events. The actions taken by the operators in the reports as well as the thought process and self reflective nature of the narratives hold potential examples of resilient behavior.

```
{"Anticipate":
   ["We expected some wake turbulence
    ↪  following a heavy.",
   "Thunderstorms were reported in the
    ↪  area so we planned to carry
    ↪  extra fuel."],
"Monitor":
   ["We were listening to ATC
    ↪  communications to understand
    ↪  which runways to expect.",
   "We were watching the temperature
    ↪  and inspecting the wings for
    ↪  potential icing."],
"Respond":
   ["We followed the TCAS RA and
    ↪  leveled off to avoid a loss of
    ↪  separation.",
   "ATC set us up high on a final
    ↪  approach so we had to bleed off
    ↪  some altitude and speed."],
"Learn":
   ["Next time I will make sure to ask
    ↪  for clarification when in
    ↪  doubt.",
   "I will be sure to look over the
    ↪  airport taxiways in case there
    ↪  is an unexpected change to
    ↪  them."]
}
```

Fig. 1: In-context JSON schema

D. Large Language Models

With the recent advancements in LLMs such as: OpenAI's ChatGPT, Meta's Llama, and Google's Gemini, the potential for extracting and summarizing insights from large bodies of text have now become more feasible, opening the door to novel ways of investigating report narratives on safety events at scale. An advantage that the open sourced Llama model has is that it can be run locally with on premises hardware allowing for tighter data control. Although ASRS reports are publicly available, the proposed approach described in this paper can be applied to sensitive ASAP reports as well. In the case of ASAP reports, airlines require that the data not be shared externally to preserve the confidentiality of the pilots who have filed the reports. For the purposes of this work the Llama3.1 8B Instruct Model with post training 6 bit quantization was used [10]. Although it can be useful to compare the performance between different LLMs, the objective of this paper is to demonstrate an approach to extract resilient behavior and illustrate how the tool can support Safety II objectives.

Apart from being able to provide answers to queries by drawing from an extremely large corpus of training data, these models have the ability to answer specific questions within the scope of a provided set of text. In general, this methodology is called Retrieval-Augmented Language Models (RALM) [11] and in specific implementations, referred to as Retrieval Augmented Generation (RAG). This technique can improve the relevant retrieval performance since the information extracted by the LLM is limited to the scope of the provided text and less likely susceptible to hallucinations where the model infers an inaccurate response due to the fact that it is focused on next word completion, drawing from the enormous variety of text and domains that it was trained upon that may or may not be relevant to the prompted task. Using the RALM technique to focus the text completion task, the user can prompt the LLM to perform a specific task such as identify resilient behavior constrained to only information from the provided report narrative. Additionally, the prompt can be augmented with in-context [12] examples of what the user would like to retrieve from the provided report. In the domain of resilient behavior, the user may provide a different categories of resiliency in the form of an example sentence and category. This serves two functions: 1) the in-context examples help guide the LLM's response towards the desired information that the user is requesting and 2) it defines a structured output such as JSON schema formatting to help with parsing and organizing the output. Fig. 1 shows the in-context augmented text provided with the LLM prompt that was used to guide the model to retrieve examples for the resilient categories in each of the reports.

Prompting is an important component in guiding the LLM to extract the information that the user is attempting to retrieve and may take some iterations to refine. Prompt priming is one way to give the LLM a subject matter expertise background so that the responses align with the domain the user is asking the LLM to respond. The prompt should provide some description of the RAG text the LLM will be examining. Additionally explicit instructions are needed to ensure the LLM does not stray from the task or provide extraneous information. The prompt for the Llama3.1 Instruct model to analyze each ASRS report was:

> *You are an aviation expert with a human factors background. Review a report written by a member of the aviation community. The report describes a safety event that took place. The report may or may not contain resilient behavior exhibited by the pilot or controller in the report. If the report does not contain resilient behavior leave the list blank. If there is resilient behavior, quote each example. The text following the EXAMPLE SCHEMA: are some examples of resilient behavior but the findings should not be limited to only these specific exam-*

ples. Provide output in a valid JSON format using the EXAMPLE SCHEMA. Only quote instances of resilient behavior. Do not summarize or explain why the cited text is characterized as resilient behavior. Do not perform other tasks.

This LLM task was run independently for each ASRS narrative spanning January 1988 through December 2024, comprising a total of 254,126 narratives. For each narrative the LLM produced a set of extracted evidence sentences from the narrative that were categorized into the 4 different resilient categories. Narratives would yield multiple categories of evidence sentences or none at all. The total number for each category were as follows: Respond (146,517), Learn (101,816), Monitor (48,359), and Anticipate (244).

II. Results

A. Resilient Category Summaries

Once the LLM identified the evidence sentences from each report that support the four resilient behavior categories the LLM was prompted to summarize the top 10 themes for each of these categories. The sentences were supplied to the LLM as a list across all the reports for each of the categories. For some categories the total number of sentences exceeded the LLM's token limit and the summarization task had to be performed in smaller batches. After multiple top 10 lists were compiled across the smaller batches, the LLM was used again to combine these summary list. The top 10 overall themes for each of the four resilient categories are as follows:

Anticipate:

- Weather forecasting and anticipation
- Fuel management and planning
- Air traffic control and communication
- System malfunctions and failures
- Emergency procedures and protocols
- Altitude and terrain considerations
- Wake turbulence and airspeed deviations
- Communication with dispatch and maintenance
- Alternate routes and airports planning
- Pre-flight inspections and preparations

Monitor:

- Monitoring ATC communications
- Watching other aircraft's position and movement
- Inspecting aircraft systems and instruments
- Monitoring weather conditions
- Watching for potential hazards and obstacles
- Listening to radio communications for traffic information
- Monitoring fuel levels and consumption
- Scanning for traffic in the vicinity
- Watching taxiways and runways for potential hazards
- Listening to radio communications with other aircraft or ground control

Respond:

- Responding to ATC instructions
- Correcting mistakes or errors
- Declaring an emergency or taking evasive action

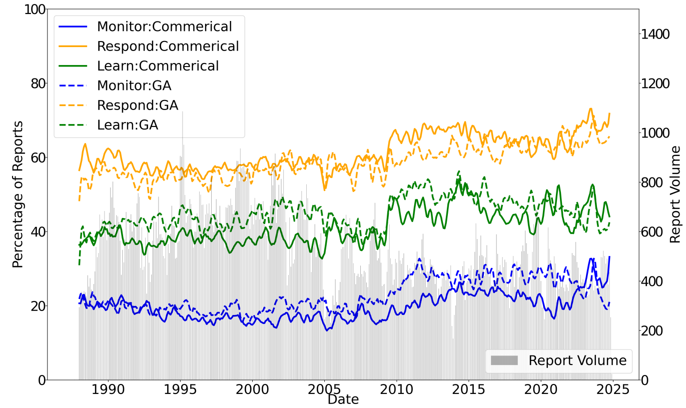


Fig. 2: Rate of categories of resilient behavior over time and volume of ASRS reporting

- Communicating with ATC and other aircraft
- Taking evasive action to avoid collisions
- Following procedures for abnormal situations
- Responding to system malfunctions or failures
- Correcting navigation errors or misunderstandings
- Coordinating with crew members and air traffic control
- Reporting incidents or issues to authorities

Learn:

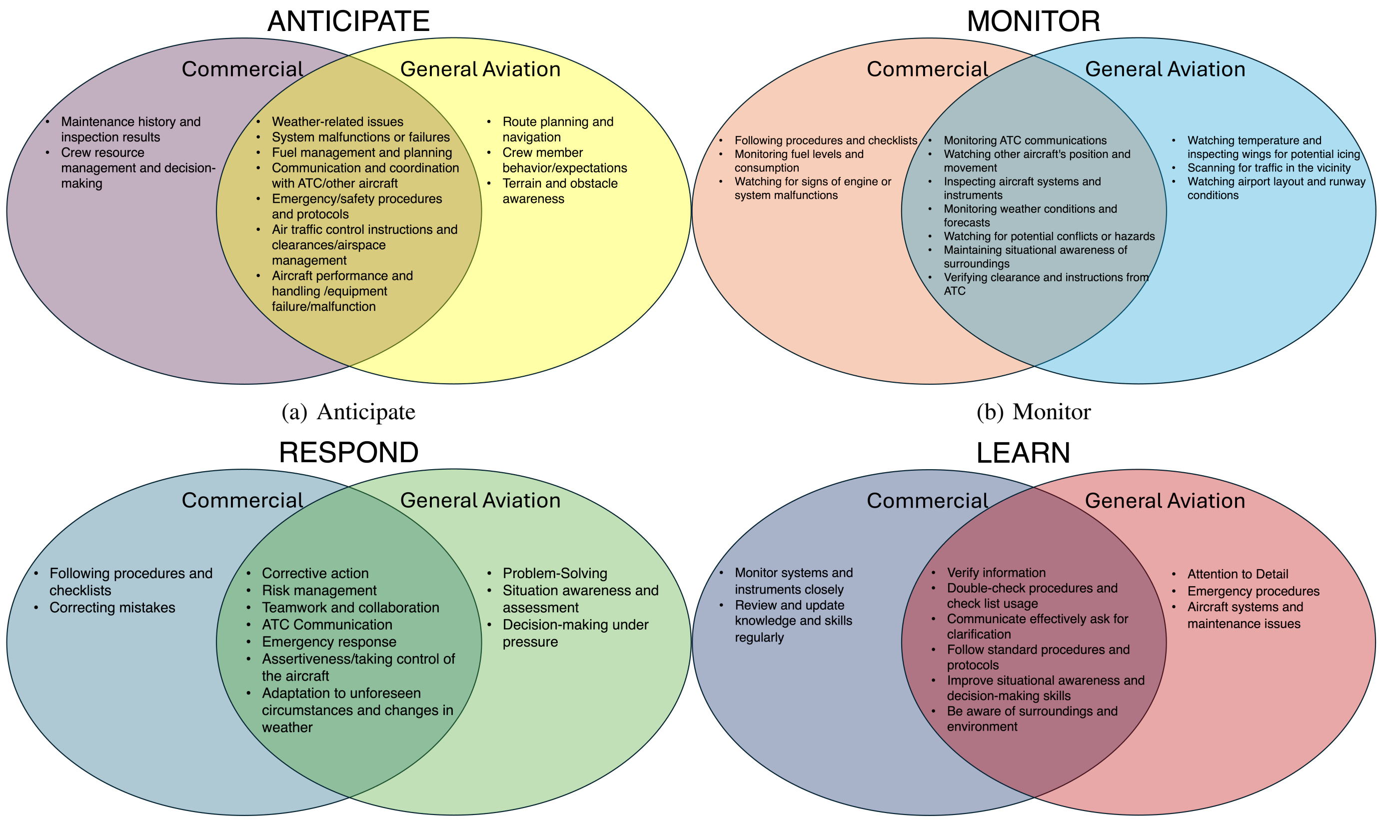
- Verify Information
- Double-check Procedures
- Communicate Clearly
- Stay Vigilant and Focused
- Follow Standard Procedures
- Be Aware of Surroundings and Environment
- Monitor Systems and Instruments
- Take Responsibility for Actions and Decisions
- Improve Situational Awareness
- Review and Analyze Data

B. Temporal Trends

Even though the exact date of the ASRS reports are de-identified, the month is still preserved. This allows analysts to look at seasonal trends or the ability to track emerging topics such as the rise of the term COVID-19 in 2020. In addition to the month there are demographics on what type of operation was involved in the observed safety situation. This includes GA aircraft as well as commercial airline operations. Fig 2 show the trend over time of the rate of the three resilient categories *monitor, respond,* and *learn* for both GA and commercial operations. Note: *anticipate* is not displayed due to the low overall count. A three month moving average window was used to smooth the trends. The gray bars in the background and corresponding right hand y-axis illustrate the volume of reports over time.

C. Commercial vs General Aviation

To compare the summary themes between the commercial airline reports vs the GA reports the evidence sentences were summarized for each type of operation by the LLM



(a) Anticipate

(b) Monitor

(c) Respond

(d) Learn

Fig. 3: Venn diagrams showing resilient behavior between commercial vs general aviation reports

separately. When comparing the two sets of themes there was good overlap, with a majority of the themes being shared by both commercial and GA operations, however, there were some unique to each across the four resilient categories. Fig. 3 illustrates the shared and unique themes across the four resilient categories as Venn diagrams.

D. Anomaly Categories

The ASRS archive has coded anomaly categories for each report. The top five most frequent anomaly categories were airborne conflict (21,778), weather turbulence (16,529), runway ground excursion/incursion (12,399), Near Mid Air Collision (NMAC) (11,677), and altitude deviation overshoot (10,664). The LLM was prompted to summarize the top 10 themes within the four different resilient categories for each anomaly category. Using the top 10 themes, the LLM was prompted to determine which themes were shared between each of the anomaly categories. TABLE I lists the anomaly categories and overlapping resilient themes across *anticipate, monitor, respond* and *learn*. In addition to identifying where themes overlapped between anomaly categories, the LLM also identified themes that were unique to each anomaly category. The following lists the anomaly categories and the unique themes identified by the LLM.

Airborne Conflict:

- *Anticipate:* Traffic conflicts and altitude management.
- *Anticipate:* Expectation bias and loss of radio contact/communication breakdowns.
- *Monitor:* Monitoring TCASII system to understand traffic situation
- *Monitor:* Inspecting surroundings for potential traffic or hazards
- *Monitor:* Maintaining visual contact with other aircraft and monitoring altitude and position of other aircraft
- *Respond:* Maintaining visual separation with other aircraft
- *Respond:* Issuing traffic alerts or warnings
- *Respond:* Reporting near misses or conflicts
- *Learn:* Equipment issues and automation misuse
- *Learn:* Airspace design/management and controller-pilot coordination

NMAC:

- *Anticipate:* Monitoring communications, anticipating pilot behavior, and assessing situational awareness.
- *Anticipate:* Includes items related to recognizing potential conflicts and adjusting flight plans.
- *Monitor:* Making position reports and announcements on CTAF/Unicom frequency
- *Monitor:* Using TCAS or ADS-B to track traffic

	ASRS Anomaly Categories				
	Air-borne Conflict	NMAC	Over-shoot	Run-way	Weather Turbu-lence
Anticipate					
Weather Conditions	x				x
Air Traffic Control Instructions	x			x	
Safety Precautions	x				
Situational Awareness		x	x		
Monitor					
Monitoring ATC communications	x	x	x	x	x
Scanning for traffic visually or using radar/ TCAS/Traffic Awareness System (TAS)	x	x	x	x	x
Watching other aircraft's position and altitude	x	x	x	x	x
Listening to CTAF communications or Unicom frequency	x	x	x	x	x
Monitoring weather conditions and wind direction/speed	x	x	x	x	x
Respond					
Communicating with ATC	x	x	x	x	x
Taking evasive action or corrective action	x	x	x		
Changing course or altitude		x			x
Reporting incidents or near misses to ATC	x	x			
Responding to TCAS RA or traffic alerts	x		x		
Learn					
Communication	x	x	x	x	x
Human Error or Mistake	x	x	x		
Situational Awareness	x	x			
Fatigue and Workload Management	x	x	x		x
Procedural Errors and Adherence to Procedures	x		x		

TABLE I: Top 5 anomaly categories and overlapping resilient behavior.

- *Respond:* Go Around/Missed Approach
- *Respond:* Increase Rate of Descent/Decelerate
- *Learn:* Using technology, such as TCAS, effectively
- *Learn:* Reporting incidents and near-misses to authorities

Overshoot:

- *Anticipate:* Factors that can contribute to an aircraft overshooting its intended target or destination.
- *Anticipate:* Includes items like autopilot reliance, power management, and climb rate management.
- *Monitor:* Inspecting Autopilot Performance and Settings
- *Monitor:* Monitoring Flight Instruments and Systems for Potential Issues, Monitoring Autopilot and Flight Director Systems
- *Respond:* Recovering from Autopilot Failure or Malfunction
- *Respond:* Correcting for Altimeter Setting Error
- *Learn:* Altitude Management Errors
- *Learn:* Pilot Experience and Familiarity with Aircraft/Systems

Runway:

- *Anticipate:* Situational awareness during takeoff and landing.
- *Anticipate:* Includes items related to runway and taxiway situations, fuel management, and emergency situations.
- *Monitor:* Watching taxiway and runway layout
- *Monitor:* Using airport diagrams, charts, or maps to understand taxi routes and runways
- *Respond:* Misunderstanding or Miscommunication with ATC / Failure to Follow ATC Instructions
- *Respond:* Loss of Control or Directional Issues / Failure to Follow Standard Procedures
- *Learn:* Overconfidence and Arrogance

Weather Turbulence:

- *Anticipate:* Reviewing weather forecasts, assessing potential hazards, and planning for contingencies.
- *Monitor:* Monitoring weather radar
- *Monitor:* Inspecting aircraft systems for potential issues
- *Monitor:* Coordinating with ATC
- *Respond:* Declaring an emergency
- *Respond:* Requesting assistance or clarification
- *Respond:* Declaring minimum fuel or emergency fuel
- *Learn:* Weather conditions
- *Learn:* Airspace and navigation issues

III. Analysis And Discussion

One consistent reoccurring theme across the different conditions used to split up the reports was ”communication”. All four resilient categories in the overall summaries (in section II-A) had one form or another of communicating with ATC. In the comparison between commercial and GA operations (in Fig. 3) all of the categories had some type of communication as a shared theme. And in the anomaly categories (in TABLE I) communication was also highlighted across the five anomaly categories in *monitor, respond*, and *learn*. This aligns well with resiliency since communication allows pilots and controllers to share knowledge and intention with other operators, making decisions and actions clear and more predictable for other actors in the airspace. Additionally the temporal trends report rates shown in Fig. 2 across both commercial and GA in the resilient categories of *monitor, respond* and *learn* have similar patterns.

Another observation of note is that in the anomaly category analysis in TABLE I, all the *monitor* category themes were shared across all five of the anomaly categories. These specific themes point to a more general theme of enhancing situational awareness around what is happening in the airspace.

However, in contrast to having shared themes, there were some unique behaviors found primarily in commercial operations that were not highlighted in the GA reports. For example in both resilient categories: *monitor* and *respond* the theme of "following procedures and checklist" was unique in the commercial reports. This is a understandable, since airlines have developed procedures and checklists as a cornerstone to their pilot training and culture.

IV. Conclusion

This proposed approach demonstrated a ways to leverage LLMs to extract evidence of resilient behavior within pilot and controller safety reports and that this same tool can successfully organize the evidence into themes of resilience. Previously this type of analysis would be impractical for subject matter experts due to the amount of time it would take and volume of reports that would need to be examined to extract these themes. Using this process can yield insights into operator proficiencies in routine situations, as well as in safety critical events. In these situation, both pilots' and controllers' training and experience is evident in how they: 1) anticipate threats and pressures to the original flight plan and/or procedures, 2) constantly survey the environment and other actors to gather information about how the situation will unfold, 3) take proper action and relay information to the appropriate people involved, and 4) learn from the event to better understand how to prepare for a similar situation in the future. These qualities exemplify how the human component in a complex system demonstrates the flexibility to ensure the primary mission of keeping the aircraft and all the souls aboard safe. SMS programs can benefit from identifying what proficiencies exist in their own operations so that they can ensure that these qualities are maintained either through pilot bulletins, training, or other means to disseminate these characteristics throughout the community.

V. Future Work

Potential areas to extend this work may include comprehensively examination of all anomaly categories and other conditional demographics to identify unique and consistent resilient behavior across the various event report types. Other steps would be to work with stakeholders to discuss how this could be incorporated into existing SMS programs for ongoing assessment of these proficiencies. The tool could be used to examine other report types such as the National Transportation Safety Board accident investigation reports, LIT observation narratives, or Line Operations Safety Audit observation narratives. Other domains include the Confidential Close Call Reporting System [13] which is an ASRS style safety reporting system for rail road operations or the Patient Safety Reporting System [14]. All of these diverse types of narratives may capture a different perspective of resilient behavior, and highlighting these qualities can serve to improve the overall resiliency of their prospective domains.

VI. Acknowledgments

To be added out upon acceptance.

References

[1] U. B. of Transportation, "August 2024 u.s. airline traffic data up 4.5% from the same month last year," Bureau of Transportation Statistics website, Nov 2024, accessed 2025-04-24. [Online]. Available: https://www.bts.gov/newsroom/august-2024-us-airline-traffic-data-45-same-month-last-year

[2] M. Stewart and B. Matthews, "Resilient strategies in commercial aviation," in *22nd International Symposium on Aviation Psychology*, ser. Proceedings, Rochester, NY, 2023.

[3] E. Hollnagel, R. L. Wears, and J. Braithwaite, "From safety-i to safety-ii: a white paper," *The resilient health care net: published simultaneously by the University of Southern Denmark, University of Florida, USA, and Macquarie University, Australia*, 2015.

[4] V. B. Bertoni, T. A. Saurin, F. S. Fogliatto, A. Falegnami, and R. Patriarca, "Monitor, anticipate, respond, and learn: Developing and interpreting a multilayer social network of resilience abilities," *Safety Science*, vol. 136, p. 105148, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925753520305440

[5] American Airlines' Department of Flight Safety (2020). Trailblazers into Safety-II: American Airlines' Learning and Improvement Team, A White Paper Outlining AA's Beginnings of a Safety-II Journey.

[6] B. Matthews, I. Barshi, and J. Feldman, "An approach to identifying aspects of positive pilot behavior within the aviation safety reporting system," in *42nd AIAA/IEEE Digital Avionics Systems Conference Proceedings*, 2023.

[7] NASA Aviation Safety Reporting System, "Asrs database online," https://asrs.arc.nasa.gov/search/database.html, 2023.

[8] F. A. Administration, "14 cfr part 121 operating requirements," Federal Aviation Administration website, 04 2025, accessed 2025-04-24. [Online]. Available: https://www.ecfr.gov/current/title-14/chapter-I/subchapter-G/part-121

[9] ——, "14 cfr part 135 operating requirements," Federal Aviation Administration website, 04 2025, accessed 2025-04-24. [Online]. Available: https://www.ecfr.gov/current/title-14/chapter-I/subchapter-G/part-135

[10] "bartowski/meta-llama-3.1-8b-instruct-gguf," accessed: 2024-08-07. [Online]. Available: https://huggingface.co/bartowski/Meta-Llama-3.1-8B-Instruct-GGUF/blob/main/Meta-Llama-3.1-8B-Instruct-Q6_K_L.gguf

[11] K. Guu, K. Lee, Z. Tung, P. Pasupat, and M. Chang, "Retrieval augmented language model pre-training," in *Proceedings of the 37th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, H. D. III and A. Singh, Eds., vol. 119. PMLR, 13–18 Jul 2020, pp. 3929–3938. [Online]. Available: https://proceedings.mlr.press/v119/guu20a.html

[12] O. Ram, Y. Levine, I. Dalmedigos, D. Muhlgay, A. Shashua, K. Leyton-Brown, and Y. Shoham, "In-context retrieval-augmented language models," *Transactions of the Association for Computational Linguistics*, vol. 11, pp. 1316–1331, 11 2023. [Online]. Available: https://doi.org/10.1162/tacl_a_00605

[13] Confidential Close Call Reporting System, "C^3RS database online," https://c3rs.arc.nasa.gov/, 2025.

[14] Patient Safety Reporting System, "Psrs database online," https://psrs.arc.nasa.gov/programoverview/index.html, 2025.