Digital Avionics Systems Conference

Montreal, Canada - September 14-18, 2025

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

2025-09-18

Bryan Matthews KBR, Inc.

Data Science Group NASA Ames Research Center

Outline

- Background and Concept of Resiliency
- Aviation Safety Reporting System
- Approach using LLMs
- Findings
- Discussion
- Future Work

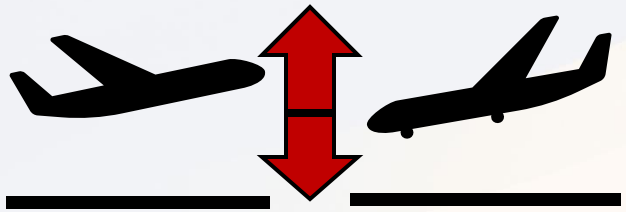
What is Resiliency?

- Types of resilient behavior defined by Erik Hollnagel et al*.

Anticipate

Monitor

React



Learn

*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.

How Do We Currently Capture Resilient Behavior?

- Human factors studies:
  - Surveys
  - Interviews
  - Eye tracking
  - Simulation scenarios
- Flight deck observations:
  - Line Operational Safety Audit (LOSA)
  - Learning Improvement Team (LIT)

Can We Leverage Existing Data Sources In New Ways?

ASRS has over 2M voluntarily submitted reports.

- Capture safety events that operators were concerned about.
- Context includes before during and after the event.
- Potentially contains corrective actions and preventative measures that pilots and controllers exhibit to make the system more resilient.

ASRS Report 1. *"SOCAL Approach Control cleared our flight for the ILS 24R via the CRCUS transition. We were following a B787-9. To help increase the space between our airplanes the Los Angeles Center* ***Controller instructed us to slow to 250 KIAS*** *while on the ANJLL4 arrival which we complied with. Looking at our TCAS display, I estimated the 787 was approximately 5 miles ahead of us.* ***SOCAL approach appropriately cautioned us for wake turbulence*** *since we were following the heavy 787. Our flight was normal until we reached CRCUS waypoint where we encountered the 787's wake".*

How Can We Leverage Large Language Models?

- Large Language Models (LLM)
  - LLMs are designed for token/word predictions given prior context and instructions
  - Retrieval Augmented Generation (RAG)
    - Limits the task's scope to use specific provided text when crafting response
- Open sourced Llama 3.1 8B Instruct Q6 8 Billion trained parameters
  - Multilingual Text
  - 128k context length (number of tokens)
  - Model uses 6 bit post quantization (~6.6GB fits within laptop memory)

What Are We Looking For?

- Extracting evidence of resilient characteristics in the text
- Use specific prompt components
  - Persona
  - Task
  - Context
  - Format

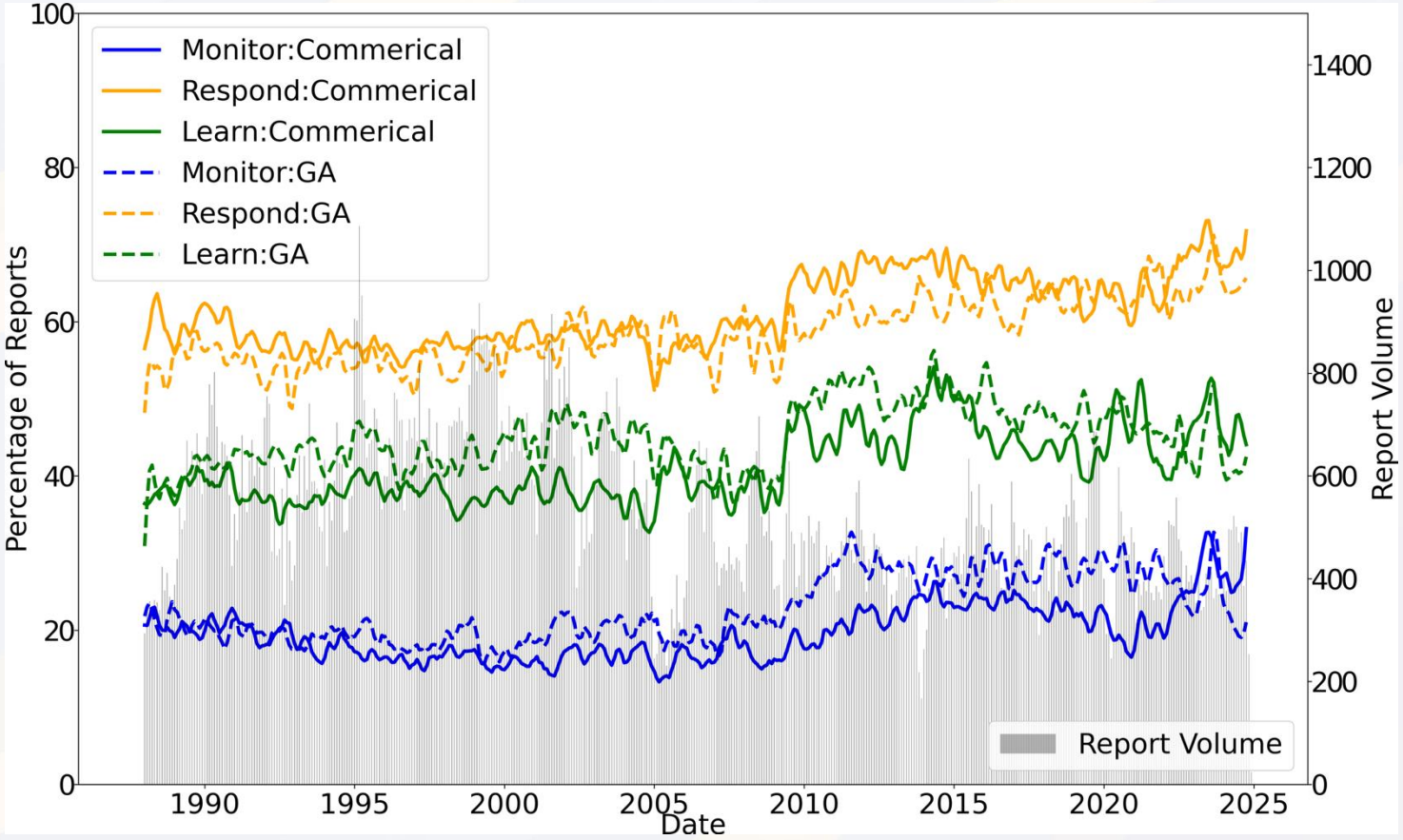
You are an aviation expert with a human factors background. Review a report written by a member of the aviation community. The report describes a safety event that took place. The report may or may not contain resilient behavior exhibited by the pilot or controller in the report. If the report does not 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 examples. 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.

Example schema and context

```
{"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."]
}
```

What Are the Trends?

- GA and Commercial follow similar reporting trends.
  - GA has slightly higher monitoring rates.



What Does Resiliency Look Like?

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

What Does Resiliency Look Like?

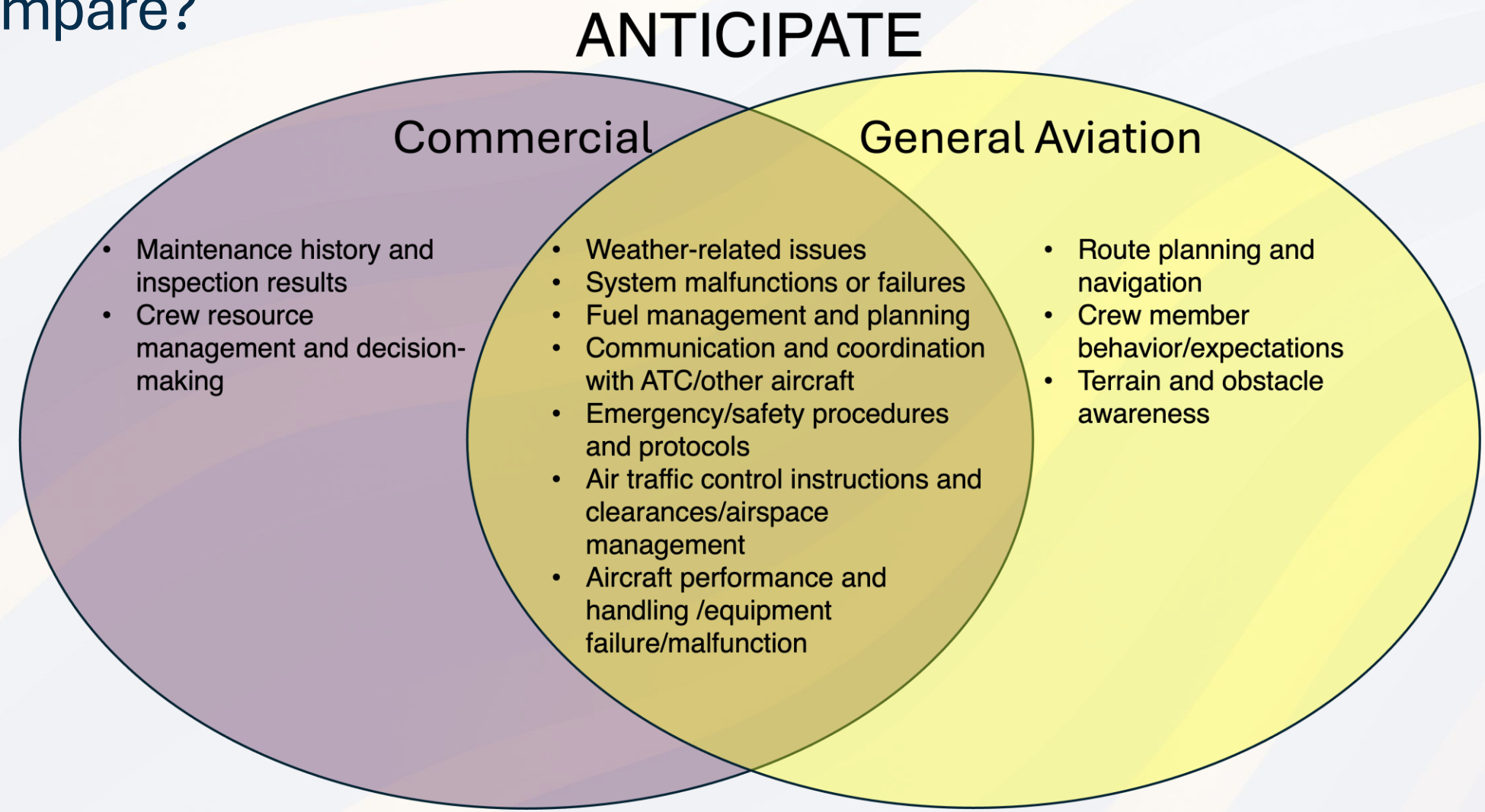
Respond:

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

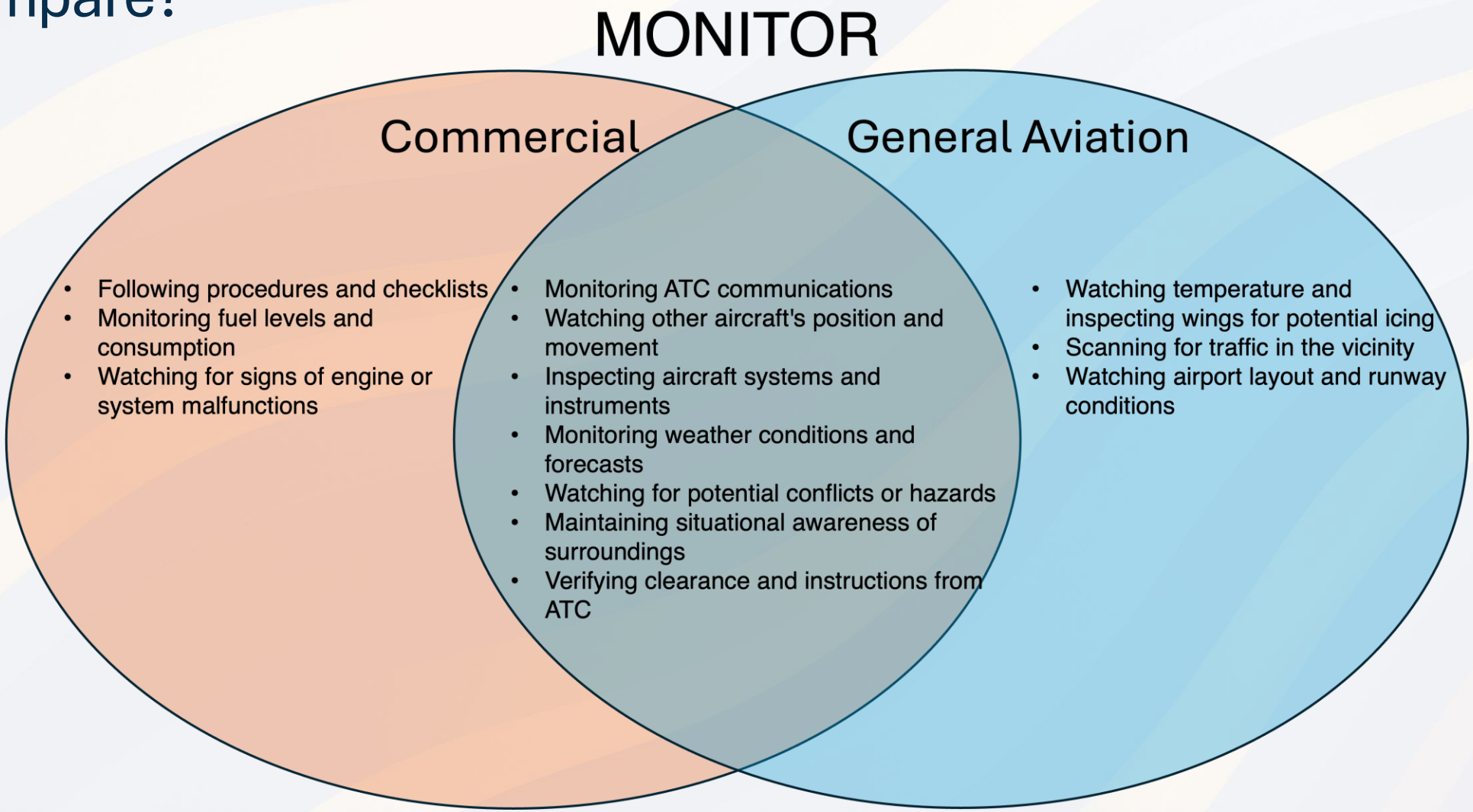
Learn:

- 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

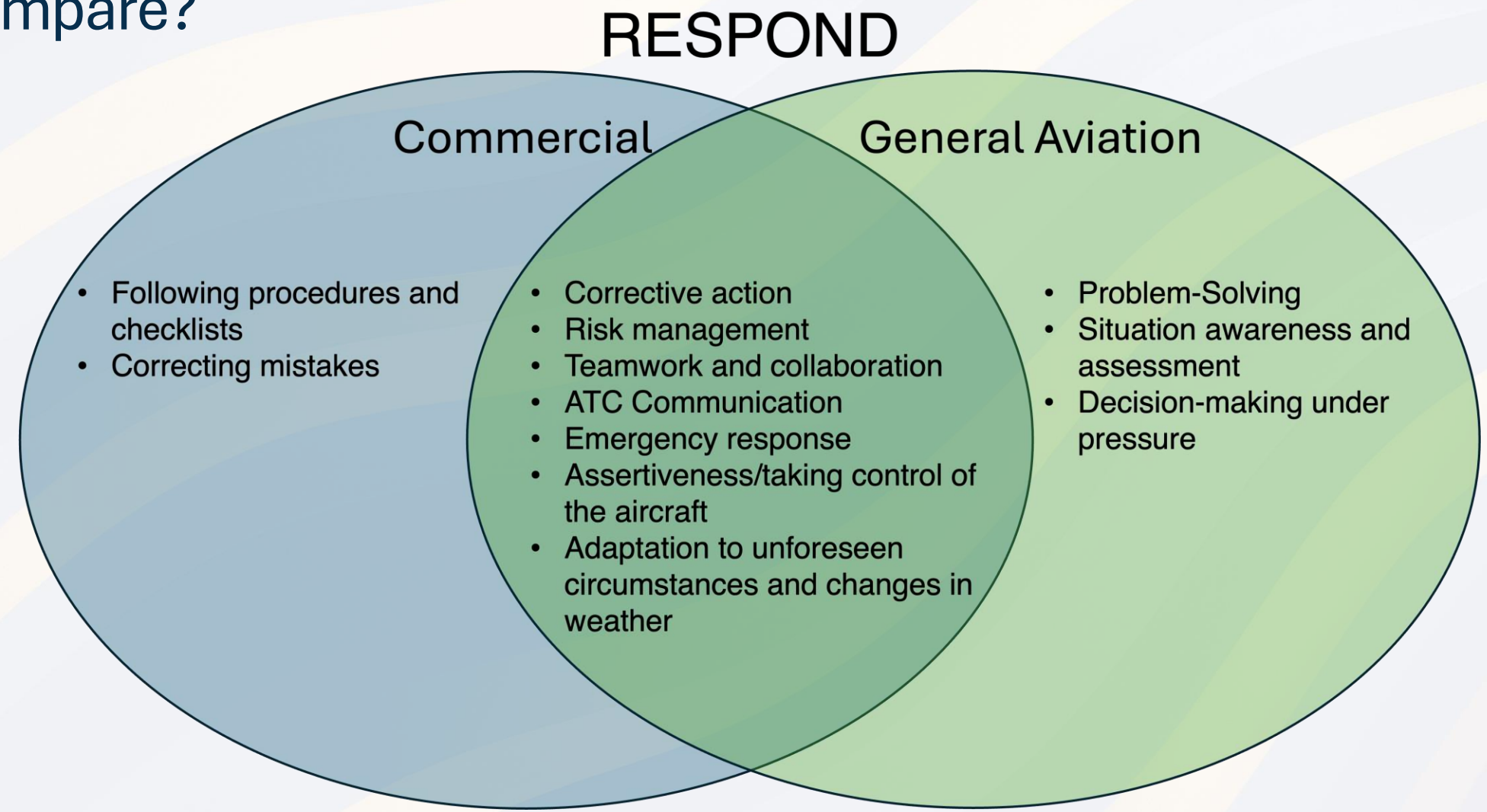
How Do Commercial and General Aviation Compare?



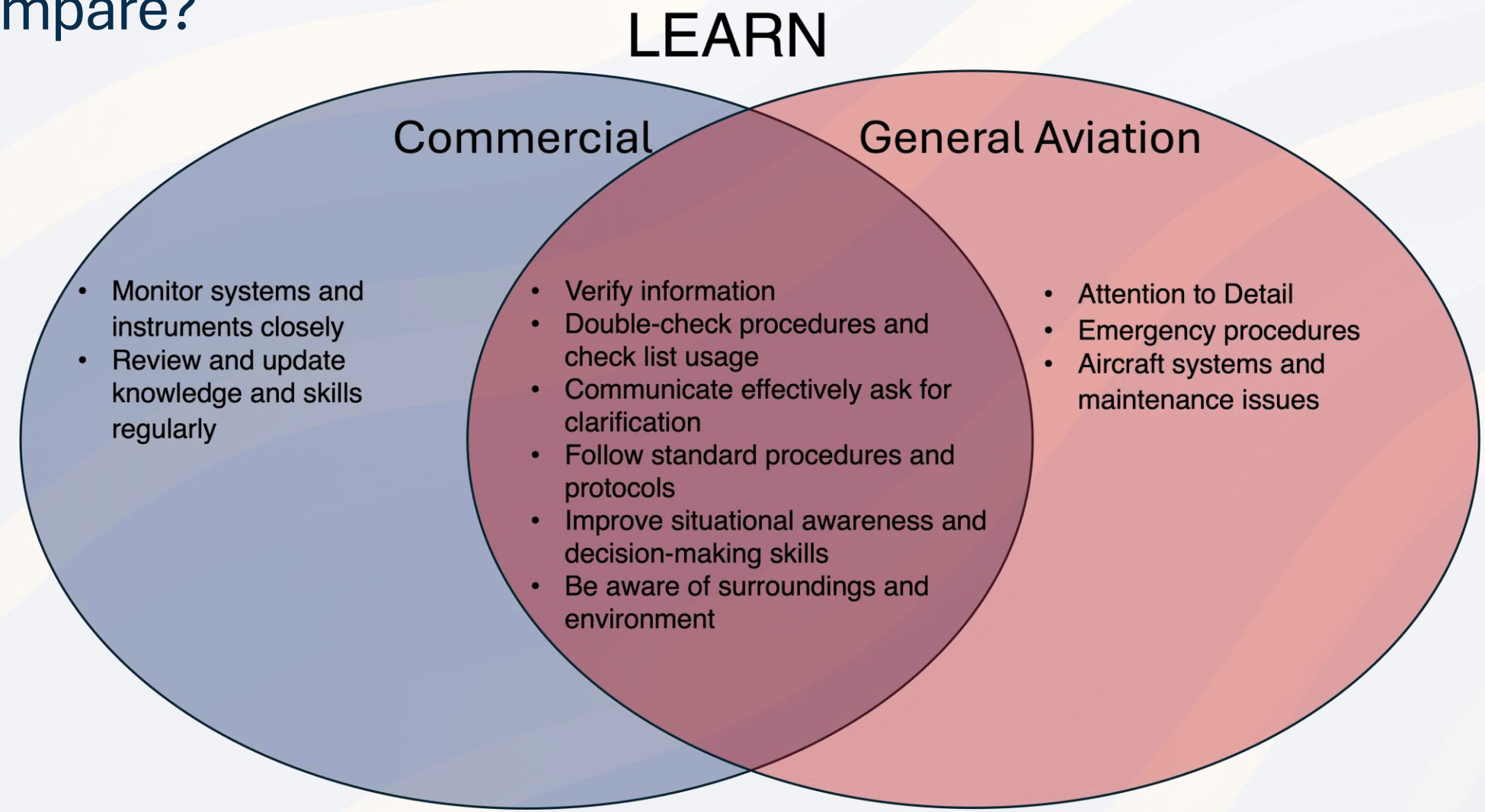
How Do Commercial and General Aviation Compare?



How Do Commercial and General Aviation Compare?



How Do Commercial and General Aviation Compare?



What Does Resilient Behavior Look Like Across Different Safety Events?

Airborne Conflict [**21,778**]

Near Mid Air Collision (NMAC) [**11,677**]

Altitude Deviation Overshoot [**10,664**]

Runway Ground Excursion/Incursion [**12,399**]

Weather Turbulence [**16,529**]

	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

	ASRS Anomaly Categories				
	Air-borne Conflict	NMAC	Over-shoot	Run-way	Weather Turbu-lence
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		

What Can We Learn From This?

- Communication:
  - Across all 4 categories of resilience.
  - Both in GA and Commercial operations.
  - Across different safety events.
  - Sharing knowledge, intensions, asking for assistance.
- GA vs Commercial
  - Commercial operations had more focus on “following procedures and checklist” in both **monitor** and **respond**.
  - GA operations relied upon “problem solving” in the unique **respond** category.

- **Anticipate:** threats and pressures to the original flight plan and/or procedures.
- **Monitor:** constantly survey the environment and other actors to gather information about how the situation is unfolding.
- **Respond:** take proper action and relay information to the appropriate people involved.
- **Learn:** from the event to better understand how to prepare for a similar situation in the future.

What Are The Next Steps?

- Comprehensive examination of all anomaly categories in ASRS.
- Complement existing SMS programs for ongoing assessment of resilient proficiencies.
- Analyze other safety reports:
  - National Transportation Safety Board accident investigation reports.
  - LIT or LOSA observation narratives.
- Apply this technique to other safety reporting domains.

Acknowledgments

- Human Contributions to Safety Team
- NASA's Aviation Safety Reporting System
- Work was funded by NASA's System-Wide Safety Project under NASA's Aeronautics Mission Directorate's Airspace Operations and Safety Program.

Questions?