GPT-4 Assisted Categorization and Visualization of NTSB UAV Accident Reports

Eugene Pik

pike@my.erau.edu

Embry-Riddle Aeronautical University, College of Aviation

Background

Increased UAV Usage

Commercial fleet 42,000 (2016) → 349,000 (2023), a 731% increase.[1]

UAVs Introduce Safety Hazards [2]

- Loss of altitude, control, transmission
- Failure or loss of navigation systems
- Collision with aircraft, buildings, power lines
- Severe weather or climatic events
- Take-off and landing incidents
- Rotor failures

Rising Safety Concerns

Escalating safety issues due to increased UAV usage.

Source Data Utilization Challenges

NTSB investigates UAV-related accidents and prepares reports for public dissemination.[3]

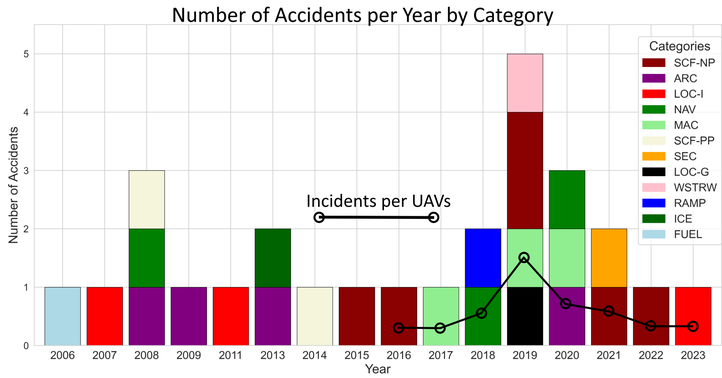
- NTSB reports missing accident category field
- Data columns have missing values
- Numeric fields represented as objects
- Narrative text requires natural language processing
- Date fields require date parsing

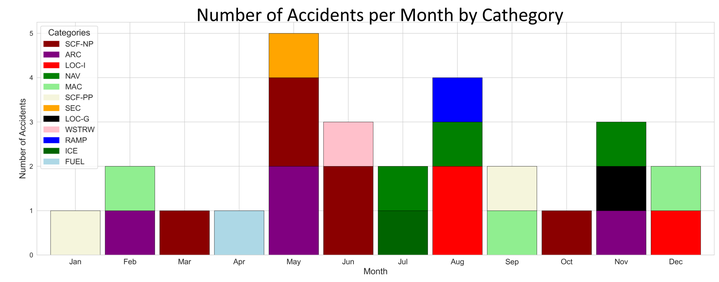
Role of AI and Data Visualization

AI can be used to dissect and understand UAV accidents by AI-facilitated categorization coupled with data visualization techniques. Together, these can be used to proactively address UAV safety policies.

References & Acknowledgements

Results





Color	Count	Category	Category Name
	6	SCF-NP	System/Component Failure or Malfunction (Non-Powerplant)
	4	ARC	Abnormal Runway Contact
	3	LOC-I	Loss of Control - Inflight
	3	NAV	Navigation Errors
	3	MAC	Airprox/TCAS Alert/Loss of Separation/Near Midair Collisions/Midair Collisions
	2	SCF-PP	System/Component Failure or Malfunction (Powerplant)
	1	SEC	Security Related
	1	LOC-G	Loss of Control - Ground
	1	WSTRW	Wind Shear or Thunderstorm
	1	RAMP	Ground Handling
	1	ICE	Icing
	1	FUEL	Fuel Related

Methods

Data Sourcing: NSTB UAV Accident Reports (n=34; between 04/2006 and 08/2023).[3]

AI-Driven Categorization: Identified matching category from NTSB document that corresponds to "possible cause" field of each report.[3-6]

Software and Tools Used: Python 3.x, pandas, numpy, joblib, chardet, unidecode, matplotlib, seaborn, and folium.[7-15]

Data Handling: Dynamic encoding detection used for reading and processing CSV input data files. Unidecode sanitized text data in parallel for enhanced performance.[12] Python scripts used error handling and logging to ensure data integrity.[4]

Summary

Key Findings

- The primary cause of UAV accidents is System and Component Failure or Malfunction (SCF-NP).
- UAV accidents are geographically widespread in the U.S. with some localized clusters.
- Seasonal variation with spikes in May and August.
- Accidents show some variations across years, with notable spike in 2019.
- A normalized per UAV downward trend in UAV accidents suggests improvements in safety after 2019.[1]

Future Recommendations

- Deep Learning Techniques to analyze complex patterns in accident data.
- Natural Language Processing for extracting insights from accident reports.
- Machine Learning Algorithms for classifying and predicting accident scenarios.