ACCSCIENCE PUBLISHING

International Journal of AI for Materials and Design

PERSPECTIVE ARTICLE

Utilizing artificial intelligence for National Transportation Safety Board unmanned aerial vehicle accident analysis and categorization

Eugene Pik[1]* and Joao S. D. Garcia[2]

[1]Mevocopter Aerospace, Vaughan, Ontario, Canada

[2]DB-School of Graduate Studies, Embry-Riddle Aeronautical University, Daytona Beach, Florida, United States of America

Abstract

The rapid increase in unmanned aerial vehicle (UAV) usage has introduced significant safety challenges, including issues such as system failure, loss of control, transmission failures, and collisions. Analyzing these incidents has been challenging due to the absence of a dedicated category field in the National Transportation Safety Board (NTSB) data. This research tackles this problem by utilizing artificial intelligence (AI) to automate the classification of UAV accident reports collected between 2006 and 2023. Using natural language processing techniques, we categorize NTSB reports to improve the analysis and interpretation of incident data. We also employ advanced data visualization tools to reveal geographic and temporal patterns, offering a detailed view of UAV accident trends. The results indicate that system and component failures unrelated to propulsion systems (system/component failure or malfunction [non-powerplant]) and abnormal contact upon landing (abnormal runway contact) are predicted as the primary categories (37%) of UAV accidents for the period. These insights suggest the potential value of AI-driven categorization and visualization techniques in enhancing UAV safety standards and supporting policy development. Initial results provide promising insight into the use of language models for text classification in aviation safety problems.

Keywords: UAV accident analysis; AI categorization; GPT-4 analysis; Data visualization in safety; NTSB accident data; Accident trend analysis

***Corresponding author:**
Eugene Pik
(eugene.pik@mevocopter.com)

Citation: Pik E, Garcia JSD. Utilizing artificial intelligence for National Transportation Safety Board unmanned aerial vehicle accident analysis and categorization. *Int J AI Mater Design*. 2025;2(1):1-7. doi: 10.36922/ijamd.8544

Received: January 15, 2025

Revised: February 11, 2025

Accepted: February 18, 2025

Published online: February 28, 2025

1. Introduction

The use of unmanned aerial vehicles (UAVs) has seen a dramatic increase in recent years. The commercial UAV fleet in the United States expanded from 42,000 in 2016 to 349,000 in 2023, representing a staggering 731% increase.[1] This surge in UAV usage brings with it some safety concerns, including loss of control, transmission failures, navigation system malfunctions, and collisions with aircraft, buildings, and power lines.[2] In addition, severe weather events, take-off and landing incidents, and rotor failures have also been mentioned as relevant to safety in UAV operations.

With the increase in operations, UAV-related accidents have also escalated, creating the need for improved categorization and understanding of these incidents to support decision-making and risk management.[3] The textual nature of traditional accident reports can often hamper more automated structured categorization efforts, making it challenging to analyze and interpret the data efficiently. Advanced methods are gaining prominence in aviation research and practical industry application. They can contribute to the need to process large volumes of unstructured data and extract meaningful insights.[4] The National Transportation Safety Board (NTSB) investigates UAV-related accidents and makes detailed reports openly available to support safety analysis. Still, at times, some of these reports may have missing fields, inconsistent taxonomies and data formats, and can require natural language processing (NLP) to extract critical information and support statistical analysis.[5]

Artificial intelligence (AI), specifically NLP and machine learning approaches, can enhance the analysis of UAV accident reports and support the identification of safety improvements. By leveraging OpenAI's GPT-4, this study aims to evaluate the feasibility of automating the categorization of UAV accident reports and the identification of probable causes and patterns in the data. AI-driven categorization coupled with data visualization techniques can provide a deeper understanding of UAV accidents, enabling proactive measures to address safety issues and inform policy decisions. Given proper consideration of accuracy challenges that could be explored in further studies, the approach can also complement the resource-intensive manual categorization tasks or serve as a capability augmentation tool for such activities. This research highlights the potential of AI to revolutionize the way UAV accident data are processed and analyzed, ultimately contributing to improved safety standards and practices in the UAV industry.

2. Literature review

The exponential increase in UAV usage has led to a corresponding rise in safety challenges, necessitating comprehensive studies to mitigate associated risks. Key safety concerns include loss of control, navigation failures, and collisions with other aircraft or infrastructure. These issues often stem from operator error, mechanical failure, and adverse environmental conditions.[2,3] Regulatory and technical challenges further complicate the safe integration of UAVs into national airspace systems. Nguyen *et al.*[6] highlight difficulties in establishing consistent regulatory frameworks and the technical limitations of current UAV systems, which contribute to navigation and control issues. Synthesizing these studies, it becomes evident that addressing UAV safety requires a multifaceted approach involving enhanced risk assessments, robust regulatory measures, and technological advancements.

AI, particularly machine learning and deep learning, has shown significant promise in UAV data analysis. Nguyen *et al.*[6] demonstrated the effectiveness of multitask deep learning in analyzing UAV multisensory data, enhancing crop productivity and safety by accurately predicting equipment malfunctions and environmental conditions. Similarly, the application of large language models (LLM), particularly GPT-4, for categorizing UAV accident reports, as proposed in this study, leverages the technology to systematically classify incident causes. This structured approach offers a more nuanced understanding of UAV-related risks. Other researchers have also employed AI models to parse and interpret large datasets of accident reports, achieving higher accuracy in identifying patterns and probable causes than manual methods.[4,5,7-9] For instance, AI categorization techniques have been applied in manned aviation and automotive industries, demonstrating the versatility and robustness of AI-driven analysis.[10]

Effective data visualization can also be crucial for interpreting complex UAV data and communicating findings. Techniques such as cluster maps and interactive charts are employed to identify geographic and temporal patterns in UAV accidents. These visualization methods facilitate a better understanding of spatial distributions and trends, aiding in the development of targeted safety measures.[11,12] Research underscores the importance of visualization in making complex data accessible and actionable, particularly for policymakers and regulatory bodies aiming to implement data-driven safety interventions.[13,14]

The integration of AI and advanced data visualization techniques offers a powerful framework for enhancing UAV safety research. AI-driven categorization streamlines data analysis while visualization tools make the data more accessible and interpretable. Studies have shown that this integrated approach not only improves the accuracy of UAV data analysis but also supports the development of proactive safety policies. For instance, combining AI categorization with visual analytics has been used to identify critical safety issues and inform regulatory measures in aviation and other high-risk industries.[4] This synergy is essential for addressing the evolving challenges in UAV operations and ensuring robust safety standards. By leveraging the strengths of both AI and data visualization, researchers can provide deeper insights into accident data, enabling more effective and timely interventions to enhance UAV safety.

3. Methods

The UAV accident reports (n = 34) were sourced from the NTSB database, which provides detailed records of aviation incidents. These reports, covering accidents from April 2006 to August 2023, were obtained through the NTSB Aviation Investigation Search platform.[15] The dataset includes various fields such as event dates, probable causes, and geographic coordinates, which are crucial for analyzing and categorizing the accidents.

The collected data underwent several preprocessing steps to ensure its quality and usability. Dynamic encoding detection was employed to accurately read files with different encodings, preventing data corruption.[16] Text sanitization was performed using the Python and Unidecode library, which converts text to ASCII, making it uniform and easier to process.[17,18] Python's NumPy and Pandas together with the error handling mechanisms were implemented to manage missing values, inconsistent formats, and other anomalies in the dataset, ensuring the integrity and reliability of the processed data.[19,20]

3.1. AI categorization

OpenAI's GPT-4 application programming interface (API) was utilized to categorize the probable causes of UAV accidents.[21] The categorization process involved feeding the cleaned and sanitized report text data into the GPT-4 model, which then assigned an aviation occurrence category from the CAST-ICAO common taxonomy team (CICTT) to each accident report.[22] This approach leveraged the advanced NLP capabilities of GPT-4 to accurately interpret and classify the narrative descriptions of the accidents, thereby automating the categorization process and reducing manual effort.[23]

3.2. Visualization

Various data visualization tools and techniques were used to illustrate the findings and trends in the UAV accident data. Libraries such as Matplotlib, Seaborn, and Folium were employed to create charts, graphs, and maps.[12-14] These visualizations helped in identifying geographic distributions, temporal patterns, and key insights from the data. For instance, Matplotlib and Seaborn were used to generate detailed plots showing seasonal variations and accident trends over the years, while Folium was utilized to create interactive maps highlighting accident hotspots across the United States.

3.3. Python scripts for NTSB accident report analysis

The categorization script, *NTSB_analysis_with_gpt4_V3.0.py*, employs OpenAI's GPT-4 model to categorize UAV accident reports from the NTSB.[23] The process begins with data preprocessing, where the script reads raw accident reports and cleans and normalizes the text to ensure uniformity. This preprocessing step is crucial for removing inconsistencies and preparing the text for analysis. After preprocessing, the sanitized text is input into the GPT-4 model, which categorizes each report into predefined aviation occurrence categories based on the narrative descriptions. The categorized data is then saved for subsequent analysis.

The visualization script, *NTSB_reports_visualisation_and_map_V1.1.py*, focuses on visualizing the categorized UAV accident data.[23] The script begins by loading the categorized accident data. It then uses various data visualization tools, such as Matplotlib, Seaborn, and Folium, to generate visual representations of the data. The visualizations include line graphs and bar charts to illustrate temporal trends and accident frequencies over time. In addition, interactive maps created with Folium highlight geographic distributions and accident hotspots. These visualizations enable the identification of key insights, trends, and patterns in the accident data, facilitating a deeper understanding of UAV accident occurrences.

Together, these scripts streamline the processes of categorizing and visualizing NTSB accident reports, enhancing the efficiency and accuracy of data analysis.

4. Results

Following, OpenAI API was used to assign accident categories from the CICTT taxonomy to the NTSB UAV accident reports.[22] Table 1 lists names and codes of the categories assigned by the AI using the ICAO list.

The analysis revealed that the primary classification of UAV accidents is system and component failure, specifically categorized as "System and Component Failure or Malfunction (SCF-NP)." This category encompasses issues such as loss of control, transmission failures,

Table 1. UAV accident codes and category names

Code	Category name
FUEL	Fuel related
ICE	Icing
RAMP	Ground handling
WSTRW	Wind shear or thunderstorm
LOC-G	Loss of control – ground
SEC	Security related
SCF-PP	System/component failure or malfunction (powerplant)
MAC	Airprox/TCAS alert/loss of separation/near midair collisions/midair collisions
NAV	Navigation errors
LOC-I	Loss of control – inflight
ARC	Abnormal runway contact
SCF-NP	System/component failure or malfunction (non-powerplant)

and rotor malfunctions. Figure 1 shows the number of accidents per category.

Geographic distribution analysis showed that UAV accidents are widespread across the United States, with certain areas exhibiting higher concentrations of incidents. Hotspots were identified in regions with dense urban development and higher UAV activity. Figure 2 shows the map of the incidents.

This spatial analysis is crucial for targeted safety interventions and regulatory measures in specific areas. Visualizations, including heat maps and cluster maps created using Folium, effectively illustrate these geographic patterns, providing clear insights into regional accident trends.

Figure 3 shows notable yearly variations, with a significant spike in accidents observed in 2019. However, a downward trend in accidents per UAV after 2019 suggests that recent safety measures and technological advancements are beginning to have a positive impact.

The visualizations and tables generated from the data analysis provided a comprehensive view of the accident data. For instance, a table summarizing the frequency of different accident causes offered a quick reference to the most common issues, while a series of bar charts illustrated the monthly and yearly accident trends. Interactive maps highlighted accident hotspots, enabling a more intuitive understanding of the geographic distribution. These visual tools not only supported the findings but also enhanced the overall presentation of the data, making it accessible and interpretable for a broader audience.

Some limitations associated with the work are worth mentioning. A significant one is the limitation of the number of reports that supported the analysis. Expanding this initial exploratory approach to a larger set of reports would allow the testing of model predictive performance with added confidence. Developing an SME validated dataset with assigned CICTT categories could create opportunities for estimating predictive precision of the proposed models.

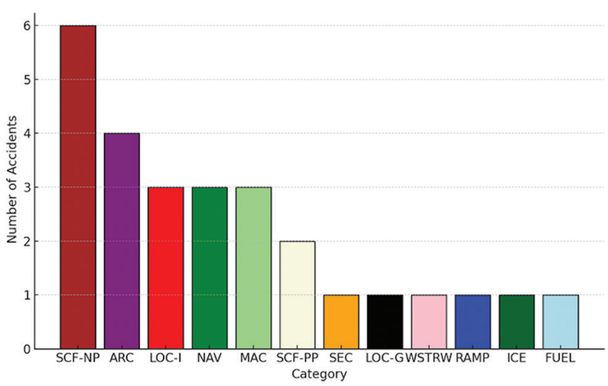

Figure 1. Number of accidents in each category

5. Discussion

The results of this study underscore the critical importance of addressing system and component failures to enhance UAV safety. The prevalence of issues such as loss of control and navigation system failures suggests that technological improvements and stringent maintenance protocols are essential. The geographic distribution of accidents further highlights the need for localized safety interventions, particularly in urban areas where UAV operations are more frequent and complex. Monitoring the annual evolution can provide valuable insights for anticipating specific risk trends, allowing for more effective preemptive safety measures.

When compared with previous studies, our findings align with the broader consensus that UAV safety is predominantly compromised by technological failures. However, our use of AI-driven categorization offers a more nuanced understanding of these issues. Ferrigan[2] identified similar hazards but his findings lacked the granularity provided by AI categorization. In addition, our identification of specific geographic and temporal patterns offers new dimensions for understanding UAV safety, which were less explored in prior research. This highlights the added value of using advanced AI and data visualization techniques.

The integration of AI and data visualization has significant implications for improving UAV safety policies. AI, particularly NLP through GPT-4, enables efficient and accurate categorization of accident reports, which is essential for large-scale data analysis. This automation reduces the manual workload and increases the consistency of data interpretation. Data visualization tools like Matplotlib, Seaborn, and Folium transform raw data into insightful visual representations, making it easier for policymakers to identify trends and patterns. Together, these technologies provide a powerful framework for developing data-driven safety strategies, enhancing regulatory measures, and ultimately reducing the incidence of UAV accidents.

6. Future work

Future research should explore the potential of AI techniques for more sophisticated analysis of UAV accident data. These techniques can uncover complex patterns and correlations simpler models might miss. Further development of NLP methods is crucial for extracting deeper insights from accident reports. Advanced NLP models can analyze

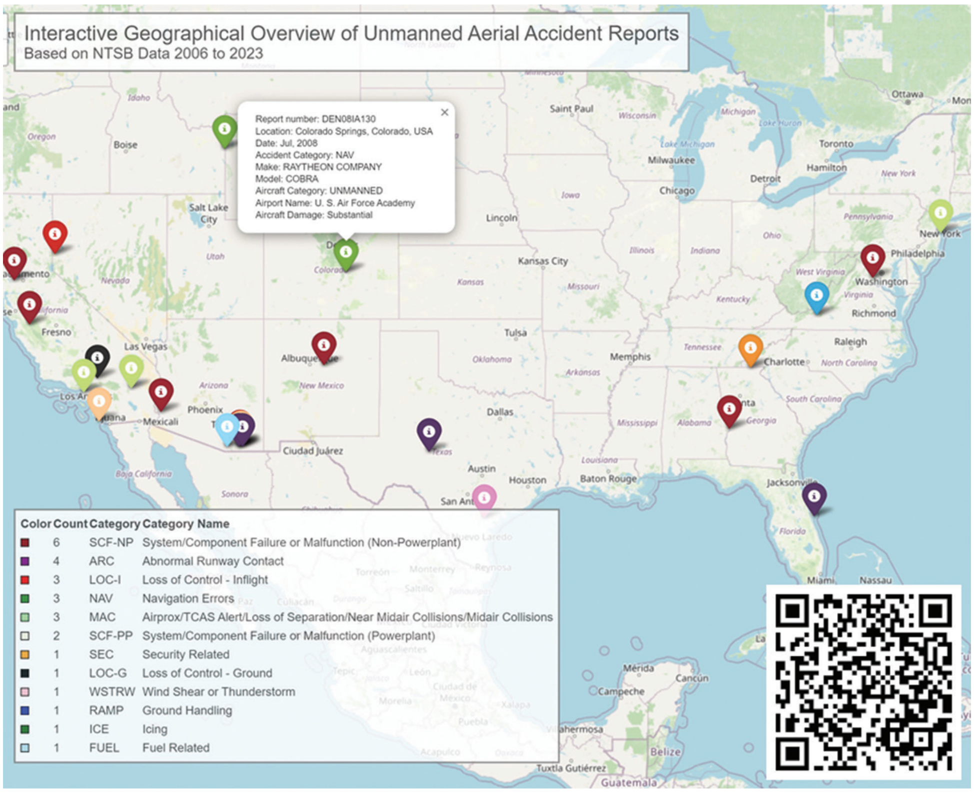

Figure 2. Interactive map of the UAV accidents[23]

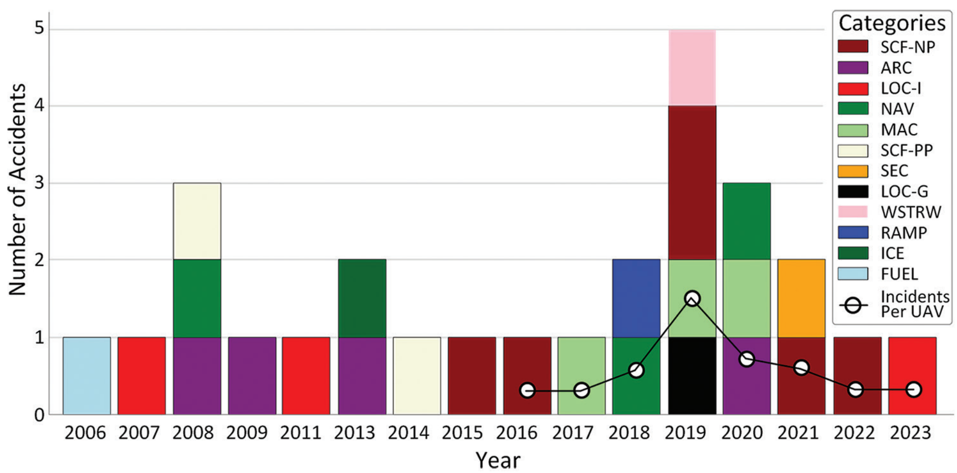

Figure 3. Accidents per year by category

narrative sections of reports to identify underlying causes and contributing factors more accurately. Machine learning algorithms should be developed and refined to classify accident scenarios and predict potential risks. These algorithms can enhance proactive safety measures by providing early warnings based on historical data and emerging trends. Interdisciplinary research combining expertise from aviation, AI, and data science can lead to comprehensive safety frameworks. Collaboration with regulatory bodies and industry stakeholders will be vital in implementing these advanced technologies effectively. Future work will involve a larger number of NTSB reports to enable robust statistical analysis. Future research can enhance reliability by ensuring consistent results across different initial conditions and demonstrating model convergence as more data is added, confirming stability and robustness. Additional research could employ a similar process to explore socioeconomic factors associated with such occurrences.

7. Conclusion

This research demonstrates the significant potential of AI in categorizing and analyzing UAV accident reports. By leveraging OpenAI's GPT-4 and various data visualization tools, we have identified patterns and trends in UAV accidents, highlighting the primary causes and their geographic and temporal distributions. Our findings underscore the importance of addressing system and component failures to improve UAV safety. The use of AI and data visualization not only streamlines the analysis process but also provides valuable insights that can inform safety policies and regulatory measures. The integration of these technologies is crucial for advancing UAV safety research and enhancing the overall safety standards in the UAV industry. Moving forward, continuous improvement of AI models and data collection methods will be essential to keep pace with the rapidly evolving UAV landscape, ensuring that safety measures remain robust and effective.

Acknowledgments

None.

Funding

None.

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization: Eugene Pik
Formal analysis: Joao S. D. Garcia
Investigation: Eugene Pik
Methodology: Eugene Pik
Writing – original draft: Eugene Pik
Writing – review & editing: Joao S. D. Garcia

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

1. The original data presented in this study are openly accessible at https://www.ntsb.gov/Pages/AviationQueryV2.aspx.
2. The data analysis scripts for this paper are available at https://doi.org/10.5281/zenodo.10576209.

Further disclosure

Part of findings has been presented in a conference "ERAU Discovery Days" (https://commons.erau.edu/discovery-day/db-discovery-day-2024/poster-session-2/54/).

References

1. Pik E. *Commercial and Recreational Active UAV Fleet in the U.S. 2016-2022*; 2024.

   doi: 10.5281/zenodo.10573914

2. Ferrigan J. *Safety Risk Assessment for UAV Operation*; 2022. Available from: https://www.regulations.gov/document/faa-2022-0426-0004 [Last accessed on 2024 Jan 15].

3. Zhang X, Srinivasan P, Mahadevan S. Sequential deep learning from NTSB reports for aviation safety prognosis. *Saf Sci*. 2021;142:105390.

   doi: 10.1016/j.ssci.2021.105390

4. Yang C, Huang C. Natural language processing (NLP) in aviation safety: Systematic review of research and outlook into the future. *Aerospace*. 2023;10(7):600.

   doi: 10.3390/aerospace10070600

5. Kasprzyk PJ, Konert A. Reporting and investigation of unmanned aircraft systems (UAS) accidents and serious incidents. Regulatory perspective. *J Intell Robot Syst*. 2021;103(1):3.

   doi: 10.1007/s10846-021-01447-6

6. Nguyen C, Sagan V, Bhadra S, Moose S. UAV multisensory data fusion and multi-task deep learning for high-throughput maize phenotyping. *Sensors* (*Basel*). 2023;23(4):1827.

   doi: 10.3390/s23041827

7. Nanyonga A, Wild G. Impact of Dataset Size and Data Source on Aviation Safety Incident Prediction Models with Natural Language Processing. In: *2023 Global Conference on Information Technologies and Communications (GCITC)*; 2023. p. 1-7.

   doi: 10.1109/GCITC60406.2023.10426284

8. Lázaro FL, Madeira T, Melicio R, Valério D, Santos LFFM. Identifying human factors in aviation accidents with natural language processing and machine learning models. *Aerospace*. 2025;12(2):106.

   doi: 10.3390/aerospace12020106

9. New MD, Wallace RJ. Classifying aviation safety reports: Using supervised natural language processing (NLP) in an Applied Context. *Safety*. 2025;11(1):7.

   doi: 10.3390/safety11010007

10. Sarkar NI, Gul S. Artificial Intelligence-based autonomous

UAV networks: A survey. *Drones*. 2023;7(5):322.

doi: 10.3390/drones7050322

11. Kovanič Ľ, Topitzer B, Peťovský P, Blišťan P, Gergeľová MB, Blišťanová M. Review of photogrammetric and lidar applications of UAV. *Appl Sci*. 2023;13(11):6732.

doi: 10.3390/app13116732

12. Filipe, Anema F, Story R, *et al*. *Python-Visualization/Folium: V0.15.1*; 2023.

doi: 10.5281/zenodo.10255171

13. Caswell TA, Droettboom M, Lee A, *et al*. *Matplotlib REL: V3.5.1*; 2021.

doi: 10.5281/zenodo.5773480

14. Waskom M. *Seaborn: Statistical Data Visualization*; 2021.

doi: 10.5281/zenodo.4645478

15. National Transportation Safety Board. *NTSB Aviation Investigation Search*; 2023. Available from: https://www.ntsb.gov/pages/aviationqueryv2.aspx [Last accessed on 2024 Jan 11].

16. Pilgrim M. *Chardet: Universal Encoding Detector for Python 3*; 2023. Available from: https://pypi.org/project/chardet [Last accessed on 2024 Jan 15].

17. Solc T. Unidecode: ASCII Transliterations of Unicode Text; 2024. Available from: https://pypi.org/project/unidecode [Last accessed on 2024 Jan 15].

18. Van Rossum G, Drake FL. *The Python Language Reference*; 2012. Available from: https://docs.python.org/3/reference/index.html [Last accessed on 2024 Jan 15].

19. Harris CR, Millman KJ, Van der Walt SJ, *et al*. Array programming with NumPy. *Nature*. 2020;585(7825):357-362.

doi: 10.1038/s41586-020-2649-2

20. The Pandas Development Team. *Pandas-dev/Pandas: Pandas*; 2023.

doi: 10.5281/zenodo.3509134

21. OpenAI. *OpenAI Platform-Documentation*; 2024. Available from: https://platform.openai.com/docs/api-reference/introduction [Last accessed on 2024 Jan 11].

22. ICAO. *Aviation Occurrence Categories-Definitions and Usage Notes*; 2013. Available from: https://www.ntsb.gov/safety/data/documents/datafiles/occurrencecategorydefinitions.pdf [Last accessed on 2024 Jan 11].

23. Pik E. *GPT-4 Assisted Categorization and Visualization of NTSB UAV Accident Reports-Python Scripts*; 2024.

doi: 10.5281/zenodo.10576209