# From BERTopic to SysML: Informing Model-Based Failure Analysis with Natural Language Processing for Complex Aerospace Systems

Seydou Mbaye * and Hannah S. Walsh †
*Intelligent Systems Division, NASA Ames Research Center, Moffett Field, CA, 94035, USA*

Samantha I. Infeld ‡
*NASA Langley Research Center, Hampton, VA, USA*

Misty D. Davies §
*Intelligent Systems Division, NASA Ames Research Center, Moffett Field, CA, 94035, USA*

Garfield Jones ¶
*Department of Industrial and Systems Engineering, Morgan State University, Baltimore, MD, 21251, USA*

**The development of emerging complex aerospace systems will require new approaches for capturing safety incident scenarios as early as possible in the design phase. However, for novel systems, relevant data available is limited. In this work, we propose a framework informing model-based mission assurance activities with historical incident reports, lessons learned, or other relevant engineering documents using natural language processing. In doing so, we investigate whether there is useful information in data sets that are relevant, if not identical, to the system under design and whether, through rigorous systems engineering practice, this information can be effectively leveraged through model-based failure analysis. In a worked case study, we apply state-of-the-art topic modeling techniques to two data sets, a mission relevant data set and a system relevant data set. The sets of topics are merged and interpreted to form a preliminary list of failure topics that can be used to inform the identification of off-nominal modes in the model-based failure modes and effects analysis development. Once data from the system in operation is available, it can be used to update the topics identified. By extracting information about likely failures from relevant historical data sets and utilizing model-based mission assurance to ensure relevance and rigor, unanticipated failures can be reduced, and projects can more effectively learn from past missions.**

## I. Introduction

The increasing diversity and complexity of the National Airspace System (NAS) calls for a type of Safety Management System (SMS) that is scalable. Today's model of aviation SMS in the United States is very efficient. This SMS is derived by the International Civil Aviation Organization (ICAO)'s SMS mandate, which is based upon the four pillars of safety: Safety Policy, Safety Risk Management, Safety Assurance, and Safety Promotion (Fig. 1). However, as safe as it is, this type of SMS is very labor intensive, sometimes reactive, and is not scalable [1]. As increasingly autonomous systems continue to enter the NAS ecosystem, calls for a more scalable model of SMS are growing. To that effect, the National Academies of Science, Engineering and Medicine tasked the National Aeronautics and Space Administration (NASA) with developing a prototype of an In-time Aviation Safety Management Systems (IASMS) [2]. An effective IASMS will facilitate the safe integration of complex emerging operations concepts into the NAS through sets of services, functions, and capabilities (SFCs) that will continuously monitor the NAS, assess all safety risks, and mitigate them in time [3]. To accomplish that, most, if not all, safety incident scenarios will need to be taken into account during the design phase, but given the novelty of the operations in question, the availability of the data needed is very limited.

*Researcher, NASA Ames Research Center, AIAA Student Member, seydou.mbaye@nasa.gov.
†Computer Engineer, NASA Ames Research Center, hannah.s.walsh@nasa.gov.
‡Systems Engineer, NASA Langley Research Center, samantha.i.infeld.nasa.gov.
§Project Manager, NASA System-Wide Safety, AIAA Associate Fellow, misty.d.davies@nasa.gov.
¶Adjunct Professor, Industrial and Systems Engineering, Morgan State University, garfield.jones@morgan.edu.

Model-based systems engineering (MBSE) is a potential solution to this challenge, with rigorous modeling processes capturing behaviors of complex systems and systematically documenting assumptions. Additionally, it is often useful when developing systems to reference historical documents such as lessons learned and incident reports to avoid repeating mistakes from past designs. Recently, state-of-the-art natural language processing (NLP) has emerged as an option for analyzing large sets of historical documents, which can help engineers efficiently and effectively find information from past incident reports, lessons learned, and other types of relevant documents [4, 5]. In particular, applying NLP to historical documents enables a broader coverage of safety incident scenarios considered during design phase. Essentially, this process enables the vast amount of knowledge from past missions to be leveraged in new projects. For novel systems, however, failures extracted may be relevant (if the data set is chosen well), but either will not all be relevant or may be similar, but not identical to the failures present in the system under design. When integrated into a rigorous MBSE process, however, it is possible to gain the benefits of learning about failures from past missions while ensuring the identified failures are relevant and rigorously modeled.

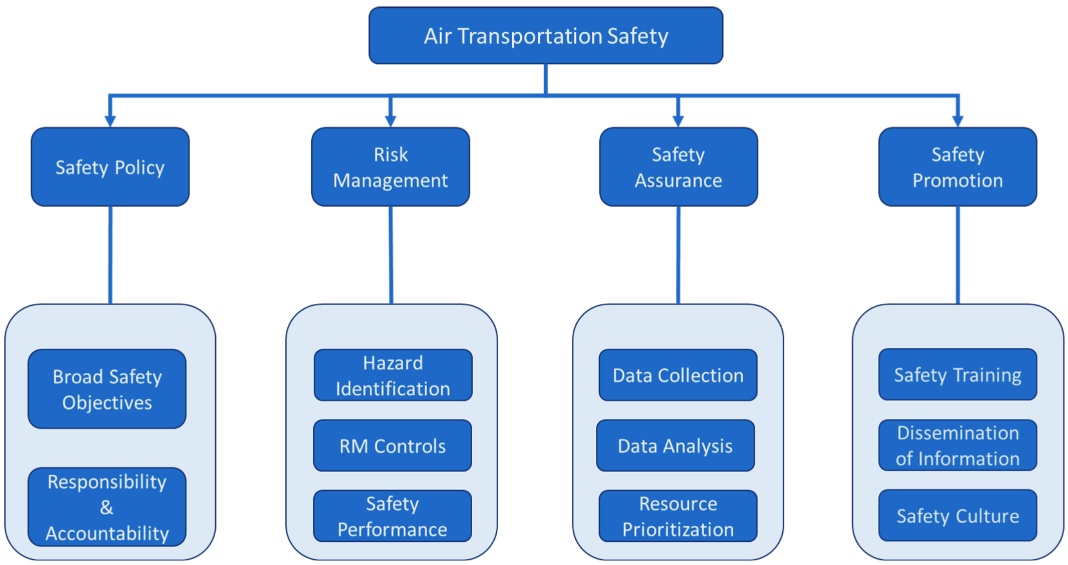


**Fig. 1 The Traditional 'Four Pillars' of a Safety Management System (International Civil Aviation Organization, "Safety Management, Standards and Recommended Practices - Annex 19," in Convention on International Civil Aviation, 2nd Edition, 2016)**

In this paper, we propose a framework for integrating a pipeline for leveraging natural language-based engineering documents to inform model-based failure analysis activities. In particular, this work investigates whether there is useful information in data sets that are relevant, but not identical, to the system under design, and whether, through rigorous systems engineering practice, those results can be integrated into a useful failure modes and effects analysis (FMEA) for the system under design. Additionally, once these reports are available for the system in operation, they can be used to update models and assumptions. During design, we use topic modeling, which is a natural language processing approach to identifying themes in large sets of documents, to learn from relevant sets of engineering documents, which may include incident reports, accident reports, and lessons learned. Then, we process those results and apply them to inform modeling activities, specifically in the identification of failure modes in developing a FMEA.

## II. Background

There is a large body of existing research and practice in Model-based Systems Engineering (MBSE) and model-based failure analysis, including the use of Failure Modes and Effects Analysis (FMEA) in MBSE. We consider gaps in current capabilities, in particular as they relate to sources of knowledge used. Additionally, we explore prior and related work applying natural language processing to extract information from historic documents that is useful to MBSE and model-based failure analysis more specifically.

### A. Model-Based Systems Engineering (MBSE)

MBSE is an application of systems engineering that uses models in lieu of documents as a means to collaborate within a project. MBSE provides a more systematic and holistic view of complex engineered systems throughout their life cycle. The International Council on Systems Engineering defines MBSE as the "formalized application of modeling to support system requirements, design, analysis, verification and validation activities beginning in the conceptual design phase and continuing throughout development and later life cycle phases" [6]. The effective practice of MBSE relies on foundations of the theory of modeling and models, including that models are by definition abstractions of reality and may be produced at different levels of abstraction depending on the purpose of their use [7]. As MBSE has been gaining traction within the NASA community, questions surrounding its application to mission assurance have emerged [8]. To that effect, the Office of Safety and Mission Assurance developed Model-Based Mission Assurance (MBMA) [9]. The overarching goal of MBMA is to leverage the system architecture captured through MBSE and facilitate the automatic generation of reliability diagrams such as failure mode and effects analysis (FMEA) and fault tree analysis (FTA). Early during the design phase, possible off-nominal conditions are modeled and studied, in an effort to figure out their possible outcomes and inform barrier measures, as shown in Fig. 2. Recent advances to the area have included efforts to introduce a computational framework for system-level behavior [10], integrating MBSE with Multidisciplinary Design Analysis and Optimization (MDAO) [11], and easing the process of building out a safety case using models [12].

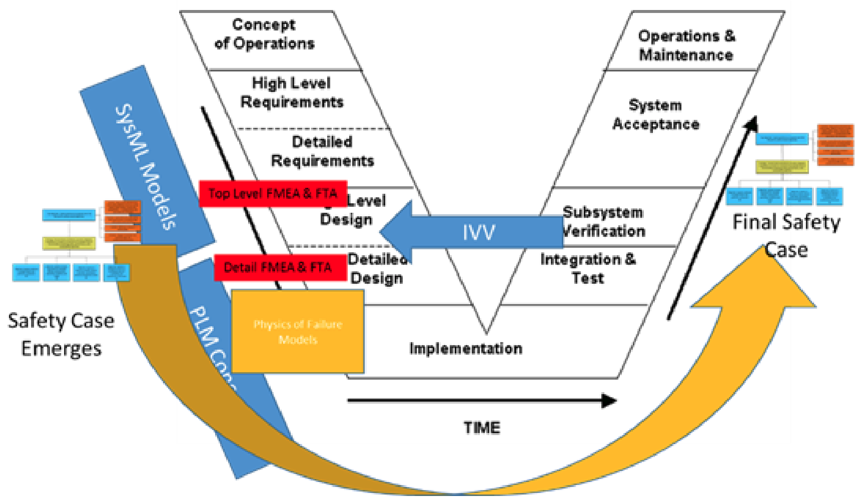


**Fig. 2 The System Engineering 'V Model' With Safety Mission Assurance inputs (NASA material, not subject to copyright in the US).**

### B. Failure Mode and Effects Analysis (FMEA)

The FMEA was first developed by the aerospace industry in the mid-1960s. The standard reference is U.S. MIL-STD-1629 [13]. As a part of the reliability-centered maintenance concept, FMEA is a method to identify where and how an asset might fail and to assess the relative impact of different failures. Traditionally, it is used at all stages of system development and failure analysis, from concept to implementation. The FMEA analysis describes inherent causes of events that lead to a system failure, determines their consequences, and devises methods to minimize their occurrence or recurrence [14]. As part of a criticality rating or the risk priority number (RPN) rating may also be determined for each failure mode and its resulting effect [14]. The rating is normally based on the probability of the failure occurrence, the severity of its effect(s), and its detectability. Failures that score high in this rating represent areas of greatest risk, and their causes should be mitigated.

More recently, FMEA may be applied within an MBSE process. Functional failures may be identified for a functional decomposition in an MBSE tool such as Magic Draw [15], which can then guide or generate an FMEA and other reliability analysis artifacts [16–19]. Specialized plugins have been developed for this purpose [20]. Tool pipelines for generating FMEAs have also been proposed for other MBSE tools such as GENESYS [21]. In general, existing tools automatically generate an FMEA primarily either by reasoning using model elements or by using a database (or using some combination of the two approaches). These approaches, while improving efficiency and consistency of system development processes, still require the practitioner to utilize sources of knowledge to complete certain steps. There have been some efforts to update existing FMEAs using operational data [22], but these have primarily been limited to updating quantitative aspects of the FMEA with limited ability to capture qualitative aspects such as failure modes and effects. With recent advancements in natural language processing, it is becoming possible to evaluate large sets of unstructured, narrative-based documents to build and update failure modes and effects in an FMEA.

### C. Natural Language Processing (NLP) for Failure Analysis

A grand challenge to novel complex engineered systems is anticipating probable failures and testing out a given system's resiliency against those failure modes. The "learn-as-you-go" method can be costly both financially and from a fatality standpoint as well [23]. With regards to that, it is of utmost importance that alternative means of identifying risks inherent to the system of interest and recommend mitigative plans. One proposed solution is to leverage historical data from closely similar systems to identify sources of failures and recommend how to mitigate them. However, said data has a tendency to be unstructured, and voluminous, thus requiring non-trivial manpower to sort through. NLP application has emerged as a credible remedy. The usage of NLP techniques to leverage lessons learned in an effort to increase robustness of systems has been in applied in different areas such as, aviation [24, 25], manufacturing [26], maintenance [27, 28], and space flights [18], and its efficiency proven.

Aviation incident reports play a crucial role in facilitating failure analysis within the aviation industry. These reports provide valuable data and insights into the root causes of accidents or incidents, enabling early reliability and resiliency analyses during the design process. However, conducting thorough failure analysis from these reports can be a labor and resource-intensive process, given the vast amount of data and technical jargon involved. Recent studies have shown that applying NLP techniques, such as text mining and sentiment analysis, to these reports, the process becomes more efficient and effective, allowing for quicker identification of patterns and trends that could prevent future incidents and enhance aviation safety. Tanguy et al. [24] proposed applying NLP to sets of aviation reports, written both in English and French, to demonstrate its effectiveness and extract lessons learned. Yun, Carlone and Liu [28] applied a Dirichlet allocation-based topic modeling method on a database of 20+ years of maintenance logs of a medical device to identify failure trends and their frequency. Their experimental application of NLP techniques to a large, unstandardized database proved successful. Topic modeling was used to automatically mine failure modes from a database of interest, and temporal analysis was then applied to identify trends. The study showed a considerable decrease in time spent compared to the traditional manual methods, thus, increasing efficiency. Prior work by the authors has extended such work to a process for assembling NLP findings into useful systems engineering artifacts, in particular a fishbone diagram [29].

### D. Natural Language Processing for Model-Based Systems Engineering

Efforts to leverage NLP capabilities to generate MBSE models have gained traction in recent years [30]. Oftentimes, the migration from traditional, document-based systems engineering towards a more digitized, model-based systems engineering can be labor-intensive, and system modelers rely on human performance to parse relevant documents to inform model generation. Such an already non-trivial task can even require more attention when the documents in question are not properly structured. NLP techniques offer the ability to automatically extract information from said documents, regardless of their level of structure. Previous studies have shown that NLP can be used not only to extract information, but also to automatically generate SysML models [31]. Researchers in this work demonstrated the efficiency of the "text-to-model" theory by comparing the models generated from texts to manually developed models. The results showed a precision rate of roughly 86% and a recall rate of about 66%, showing the accuracy of their methodology [31]. Similarly, Zhong et al. proposed a methodology in which NLP is used to extract information from different textual datasets and convert them into SysML diagrams, particularly structure and requirement diagrams [32].

## III. Methodology

The proposed framework is summarized in Fig. 3. The methodology can be divided into three sections: (1) Natural Language Processing Pipeline, (2) Hazard and Failure Mode Identification, and (3) Model-Based Failure Modes and Effect Analysis Development. Within the Natural Language Processing Pipeline, topic modeling is performed in parallel for both sets of documents. The purpose of topic modeling is to extract meaningful information from the sets of documents. Topics are labeled and merged into a single set of topics – this is primarily a concatenation but if, for example, similar topics are found in each data set, they are manually merged into the same topic. Expert interpretation of the topics is then required in Hazard and Failure Mode Identification before integrating the results into the Model-Based Failure Modes and Effects Analysis Development. In particular, the topic modeling results are used to inform the Failure Modes Identification step in the FMEA development process. After the system is deployed, it is possible to update the set of topics with new data. We withhold a portion of the original data sets and add the withheld portion back in to demonstrate this process for updating the results once the system is in operation. These steps will be explained in more detail in the remainder of this section.

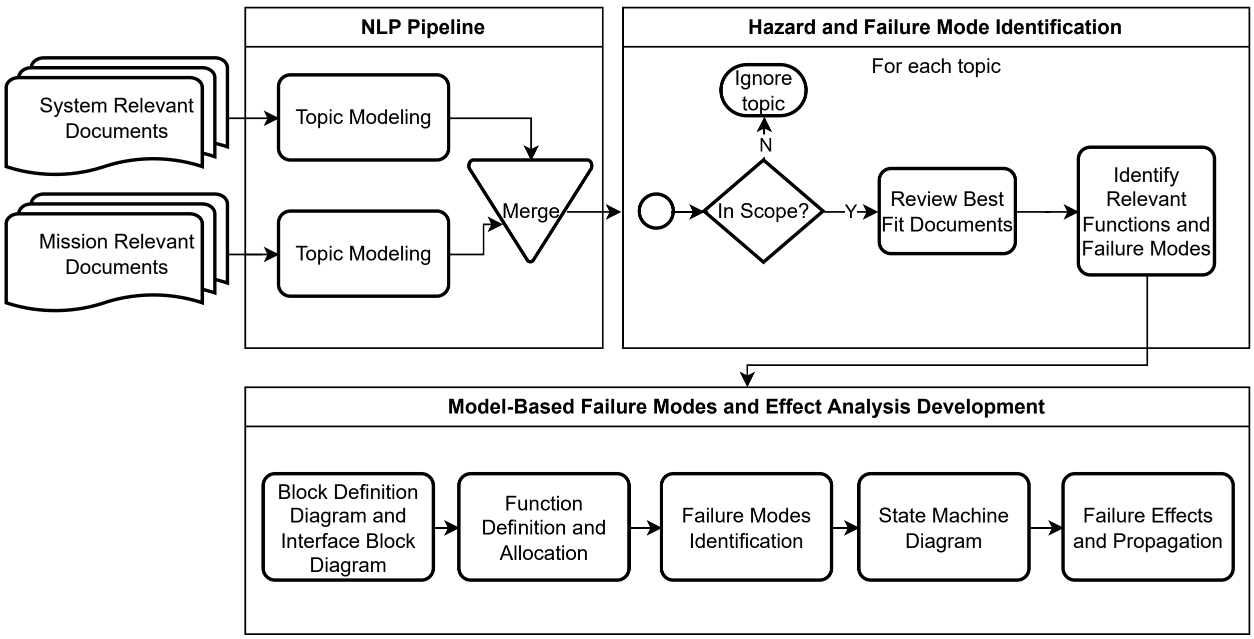


**Fig. 3 Proposed framework for informing and updating model-based FMEA development with topic modeling of relevant sets of documents. A natural language processing pipeline is set up to inform the failure modes identification step in model-based FMEA development. Topic models can be updated when current mission documents are available.**

We apply the proposed framework to a wildfire response mission in which an unmanned aircraft system (UAS) is used for surveillance and fire line monitoring, which has been described in detail in prior work by the authors [33]. Specifically, we consider the design of software for a UAS that may be used in such a mission. For the identified use case, two general categories of documents will be useful: documents from (1) mission relevant datasets and those from (2) system relevant data sets. In other words, since the selected use case is about applying a given system (UAS) to a new mission (wildfire response), two different data sets are considered. In this case, designing software for an unmanned aircraft system (UAS) that may be used in a wildfire response mission, data sets about the mission (wildfire response incidents) as well as about the system (UAS incidents, particularly with similar software) are relevant. For the former, we use documents from the ICS-209-PLUS data set from 2013 to 2014 (13,936 documents) [34]. For the latter, we use documents selected from NASA's Aviation Safety Reporting System (ASRS) * (filtered for UAS-relevant reports, 139 documents). If applying the proposed framework to a database without filtering capabilities, it is possible to build a machine learning model to extract relevant documents [4].

### A. Natural Language Processing Pipeline

Once appropriate data sets are identified, a natural language processing pipeline is used to extract information. Topic modeling has been applied in previous research to extract hazards from large sets of engineering documents [4]. In particular, in this study, we apply BERTopic topic modeling [35]. Topic modeling produces a list of themes, each represented by a list of words. There are several approaches to topic labeling, with the simplest method being

*https://asrs.arc.nasa.gov/

to select the top $n$ words to represent the topic. Alternatively, topic labels may be generated manually, or through more sophisticated means such as a pointwise mutual information (PMI) extractor [36]. Literature has suggested a preference for short phrases to represent topics, as single words tend to suggest too-broad themes and sentences can overly constrain a theme [36]. In this study, we select the top $n$ words with $n = 5$ to represent a topic. This provides a balance of appropriate specificity while avoiding biasing or over-constraining a human analyst who will interpret the topic for inclusion in the subsequent model-based steps.

The topic models are applied separately since the two sets of documents are sufficiently different such that high-quality topics containing documents from both sets would be unexpected or rare. It is possible to merge topics later in the process if such a situation is found. Once the two preliminary lists of failures are generated, they are merged and interpreted by an expert. Similar topics are merged according to expert judgment. Low quality or irrelevant topics representing failures are filtered out at this point and similar failures are combined.

### B. Hazard and Failure Mode Identification

The NLP pipeline results in a list of topics extracted from the two data sets. Next, an expert reviews the list of topics to determine whether each topic is in scope. If the topic is out of scope, it may be ignored. Remaining topics of interest can be further explored by reviewing the best fit documents associated with each topic. Reviewing the original documents that best describe the topic provides details that may help an expert define failure modes associated with a topic and may provide context to better understand the topic. This context and detail is then used, alongside expert judgment, to identify relevant functions and failure modes.

### C. Model-Based Failure Modes and Effects Analysis Development

Once the failure topic list is finalized, it is used to assist in identification of off-nominal modes within the model-based FMEA development. This is done systematically across the components and relationship of the modeled vehicle or traffic management system architectures to help elicit failure modes for each function as well as to consider whether there are external factors or subsystem interactions that may cause failure. The authors of this paper have demonstrated the development of FMEA and fault tree analysis (FTA) in a descriptive system model of an aircraft using SysML, onto which FMEA and FTA were applied in previous work [37]. Said analyses were done using the Tietronix [**?** ] FMEA and FTA plugins. The methodology is as follows:

- Identification of system architecture and interfaces using Block Definition Diagrams (BDD) and Interface Block Diagrams (IBD)
- Identification of the nominal functions of the system and allocation functions to the components performing those functions
- Identification of the possible failures of the system's components and creation of failure mode and effects signals
- Capture of components' behavior and creation of States and State Machine Diagrams
- Identification of the activity occurring on entering each state as an entry activity to the state
- Identification of triggers causing transition between nominal or from nominal to fail states
- Identification of the effects of component failure and propagation of this failure to other components

While the aforementioned paper's main contribution was to present a methodology for developing the FMEA and FTA using system model elements, this work focuses more on using NLP to inform key model elements. The Tietronix MagicDraw plugins to use with SysML require that the failure modes and their causes be modeled as signals stereotyped accordingly as "Failures" or "Effects," which we were able to identify by using NLP in this work.

### D. Updating with New Data

After the system is in operation, the natural language processing pipeline can be re-run with new documents and the results updated. In this study, to simulate this process, we withhold documents from one of the sets used in the initial topic modeling step (ICS-209-PLUS) and add them back in to simulate new reports being generated. Then, we show how this can be used to update the topic repository which, where appropriate, can be used to update the system model and the FMEA.

## IV. Results

Thirty-seven topics are initially extracted from the ICS-209-PLUS data set and eleven from the ASRS data set for a total of forty-eight topics. Ten topics are merged and one removed for being low-quality, leaving thirty-five final topics,

**Table 1 Topic counts from each step of the post-processing stage. Final counts as a result of the processing are italicized.**

| | ASRS | ICS-209-PLUS | Total |
|---|---|---|---|
| Original | 11 | 37 | 48 |
| Merged | 3 | 7 | 10 |
| Removed | 0 | 1 | 1 |
| *Final* | *7* | *28* | *35* |

**Table 2 Example merged topic, resulting from three original topics sufficiently similar to warrant treating as a single topic.**

| No. | Topic Words |
|---|---|
| 3 | demobilization, resourced, demobilized, released, crew<br>resources, released, resources released, resources demobilized, demobilization<br>declared, controlled, called, declared controlled, contained |

with twenty-eight being from the ICS-209-PLUS and seven from the ASRS. Topic counts from original set of results, merged topics, removed topics, and from the final set of topics after merging and removing low quality topics are given in Table 1. An example of a merged topic (i.e., at least two topics sufficiently similar to warrant treating as one) is given in Table 2. The merged topic describes a theme of demobilizing resources due to a fire being declared controlled.

Of the thirty-five topics resulting from this post-processing, twenty-nine are considered in-scope based on the chosen use case. This list of topics is provided in Table 3. In the case of merged topics, the top words of the one original topic are shown. This table includes light edits (e.g., defining acronyms and removing redundant words). As a whole, the topics are fairly human-readable for this method and do not take extensive time or effort to understand. More extensive effort, specifically reviewing individual documents associated with a topic, is required in order model one of the topics.

To demonstrate the method for updating results with new data, we re-run the topic modeling algorithm to include data from the latter half of 2012 in addition to the original set of 2013-2014 documents from the ICS-209-PLUS. Using the same parameter settings, we gain an additional two topics for a total for thirty-nine. Topics generally are not identical when running the topic modeling algorithm multiple times due to the stochastic nature of the algorithm, and therefore it is not possible to clearly delineate new topics from old. However, analysis of the new set of results indicates a new topic about landowners inadvertently starting fires when burning debris (*destroyed*, *debris*, *landowner*, *pile*, *burning*) and another about mobilizing involved forestry department resources (*forestry*, *mobilized*, *department forestry*, *forestry incident management*, *forestry incident*) with other topics being of the same general themes as the previous set of results.

In Fig. 4 and Fig. 5, we demonstrate how the NLP results can be used to inform modeling artifacts. Here, we used an example where a geofence failure led to a flight in controlled airspace hazard. Topic number twenty-eight in Table 3, “airspace, controlled, controlled airspace, discussed, management”, points to this hazard. Documents associated with this topic pointed to reports of said hazard, and after taking a detailed look at the associated narratives, we could then assess the causes leading up to that. This process is outlined in Table 4, in which the narratives from the relevant documents are shown alongside the expert identified failure modes, potential causes, and effects that are translated into model elements. We uncovered that "Pilot Error" and/or "Faulty Equipment" can lead to a failure of the geofence instrument, which would eventually lead to the undesired outcome (flight in controlled airspace).

## V. Discussion

Performance of the NLP algorithms on this data set is as expected and aligned with previous work. Because of the difference in size in the data sets, the results are more heavily focused on the ICS-209-PLUS (thirty-seven versus eleven topics); however, this is expected to be a realistic situation if the proposed methodology were to be used in a real engineering design project, and with separate topic models that allow for the ICS-209-PLUS-generated topics to be comparatively broader than the ASRS-generated topics, is manageable. Both data sets resulted in high-quality topics, with only one topic being removed due to low-quality (specifically, repetitive topic words with little meaningful

**Table 3 Topics identified from the data sets used in the study, after merging and removing out-of-scope results.**

| No. | Topic Words |
|---|---|
| 1 | type, team, transition, type team, organization |
| 2 | acres, acreage, mapping, acre, accurate |
| 3 | demobilization, resourced, demobilized, released, crew |
| 4 | numerous spot overs, numerous spot, spot overs, overs, numerous |
| 5 | destroyed, structure, lost, shed, debris |
| 6 | rain, received, precipitation, light, area |
| 7 | evacuation, road, highway, closure, effect |
| 8 | injury, firefighter, injuries, hospital, reported |
| 9 | rain, winds, thunderstorms, chance, weather |
| 10 | line, crews, critical, operations, division |
| 11 | surveillance, zone surveillance, flew, zone, reported |
| 12 | helicopter, type, type helicopter, helicopters, crews |
| 13 | wilderness, creek, area, located, burning |
| 14 | smoke, interior, smoke impacts, smoke lifted |
| 15 | responded, fire department responded, responded fire department, fire department |
| 16 | wind, winds, spread, relative humidity, low |
| 17 | district, grass, brush, miles |
| 18 | national, national forest, timber, timber, miles |
| 19 | flag, red flag, red flag warning, warning, effect |
| 20 | monitor, continue monitor, continue, monitoring, ground |
| 21 | using indirect tactics, negative impacts, using indirect, negative |
| 22 | box, service, timber |
| 23 | drone, flying, flight, regulations, registration |
| 24 | drone, set, drones, mission, GPS |
| 25 | aircraft, helicopter, drone, feet, altitude |
| 26 | authorization, time, Low Altitude Authorization and Notification Capability (LAANC), request, flight |
| 27 | bridge, link, drone, lost, lost link |
| 28 | airspace, controlled, controlled airspace, discussed, management |
| 29 | operator, tree, adverse, state, left |

**Table 4 Narratives from documents associated with topic twenty-eight identified using the NLP pipeline mapped to failure modes, potential causes, and effects.**

| Narrative | Failure Mode | Potential Cause | Effect |
| --- | --- | --- | --- |
| I originally checked the VFR sectional and misinterpreted the dashed magenta line and did not realize I was flying in controlled airspace. I was not aware I was flying in controlled airspace until management reviewed my flight and made me aware of the situation. Management discussed with me how to properly research controlled airspace. I also flew this same location without prior authorization on [previous dates]. | Geofence Failure | Pilot Error | Flight in controlled airspace |
| Got to location failed to research the airspace. Was relying on the manufacture's Geo fencing to guide me with airspace. Violation was discovered by company UAS coordinator. Violation was discussed and corrective action for further use is understood. We discussed and now understand the difference between Geo fencing and the NAS. I going further; will look up airspace before flights. I will not fly in controlled airspace without proper authorization. | Geofence Failure | Faulty Equipment | Flight in controlled airspace |
| I made a residential structural sUAS using DJI GO4. Relied on DJI's GEO Fencing to alert me when I was in controlled airspace instead of checking with LAANC provider or FFA's UASFM. The mistake was caught during monthly flight review by manager. We have discussed the error in flight planning and shall verify all airspace before every flight using either FFA LAANC provider or UASFM. | Geofence Failure | Pilot Error | Flight in approved airspace |

information).

The NLP tools are used in an assistive manner, in such a way as to help an expert brainstorm possible model elements to then formalize. A possible barrier to this process is the human readability of the NLP results, especially for a user who is not well-practiced in understanding this kind of machine-generated result. This problem is particularly salient when using topic modeling. The authors have found results from BERTopic to be comparatively easier to understand than other approaches, which can help mitigate this problem. Additionally, usage of an information retrieval approach alongside topic modeling can help the user fine-tune results and quickly and accurately retrieve documents or passages for more context and detail to a topic in question. Moreover, the assistive approach provides the benefits of the large amount of information available in large data sets alongside rigorous modeling practice and careful engineering judgment. However, it is also important that the information is presented to the user in such a way as to not imply that the NLP results are complete or formalized in any way. Details of human interaction with an assistive NLP-powered tool suitable for such an application will be explored in future work.

In addition to providing brainstorming input, the NLP-generated topics and accompanying original documents can provide detail needed to fully define the model. For instance, specific causes or failure conditions may be readily available in historical incident reports for a particular failure mode which may have, for example, been identified by an expert independently of the NLP pipeline. Moreover, the original documents may provide justification for representation of failure modes in the model, in addition to expert judgment.

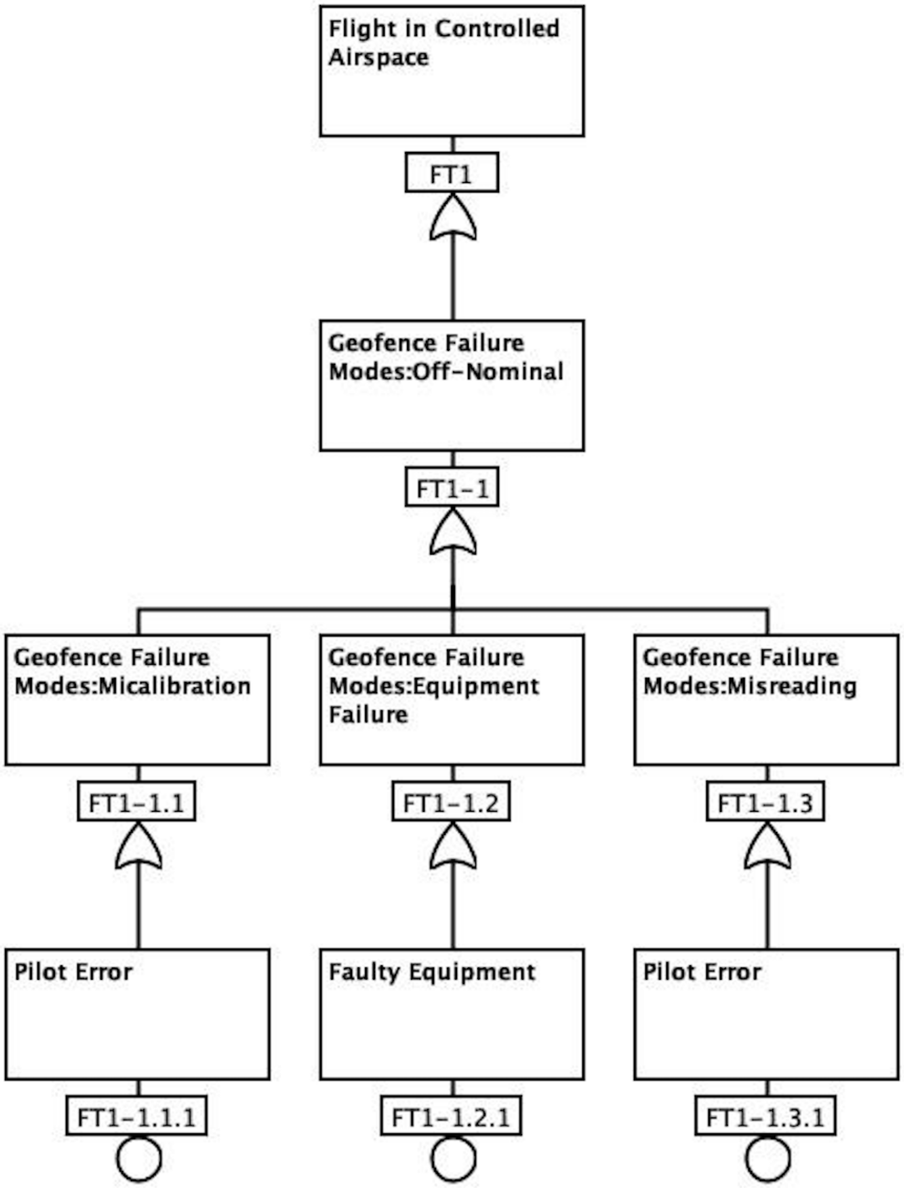


**Fig. 4 Flight in Controlled Airspace Fault Tree**

Hierarchy: ALL Row Count: 3

| Subsystem | Item | Potential Failure Mode | End Effect | Potential Cause(s) |
|---|---|---|---|---|
| Navigation Control | Geofencing Sensor | Miscalibration | EFFECT: Geofencing Sensor Flight in Controlled Airspace | EVENT: Geofencing Sensor Pilot Error |
| Navigation Control | Geofencing Sensor | Equipment Failure | EFFECT: Geofencing Sensor Flight in Controlled Airspace | EVENT: Geofencing Sensor Faulty Equipment |
| Navigation Control | Geofencing Sensor | Misreading | EFFECT: Geofencing Sensor Flight in Controlled Airspace | EVENT: Geofencing Sensor Pilot Error |

**Fig. 5 Flight in Controlled Airspace FMECA Table**

# VI. Conclusions and Future Work

This paper has presented a framework for informing model-based failure modes and effects analysis development with a natural language processing pipeline that uses BERTopic to extract preliminary failure modes from historical documents. The generated topics can be updated once the system is in operation and new documents are available. We show, through a worked case study, that it is possible to extract failures relevant to a new mission from documents pertaining to related, but not identical, missions. By extracting failures from relevant data sets and filtering them through the rigorous model-based development process, it is expected that unanticipated failures can be reduced. This will support the development of (and ultimately improve the safety of) emerging complex aerospace systems which are prone to difficult-to-predict subsystem interactions and emergent behavior.

Future work will extend the proposed methodology to building a digital assistant that interacts with an analyst to carry out this process of gathering and translating NLP results into meaningful model elements and, ultimately, entries in a finished FMEA. To support this effort, more extensive studies will also be performed to test this process on a larger scale, particularly with more NLP results translated into FMEA elements and with applications in other domains. Studies will also be performed with experienced analysts to measure and improve relevance of NLP results and presentation of those results.


# Acknowledgments

This research is supported by the System-Wide Safety (SWS) project in the Airspace Operations & Safety program (AOSP) in the NASA Aeronautics Research & Mission Directorate (ARMD). Any opinions or findings of this work are the responsibility of the authors and do not necessarily reflect the views of the sponsors or collaborators.


# References


[1] Ellis, K., Prinzel, L., Davies, M., Krois, P., Mah, R., Oza, N., Stephens, C., Vincent, M., Ackerson, J., and Infeld, S., "The In-time Aviation Safety Management System Concept for Part 135 Operators," 2022. https://doi.org/10.1109/DASC55683.2022.9925800.

[2] National Academies of Sciences, Engineering, and Medicine, *In-Time Aviation Safety Management: Challenges and Research for an Evolving Aviation System*, The National Academies Press, Washington, DC, 2018. https://doi.org/10.17226/24962, URL https://nap.nationalacademies.org/catalog/24962/in-time-aviation-safety-management-challenges-and-research-for-an.

[3] Ellis, K., Krois, P., Koelling, J., Prinzel, L., Davies, M., and Mah, R., "A Concept of Operations (ConOps) of an In-time Aviation Safety Management System (IASMS) for Advanced Air Mobility (AAM)," 2021. https://doi.org/10.2514/6.2021-1978.

[4] Andrade, S. R., and Walsh, H. S., "Discovering a Failure Taxonomy for Early Design of Complex Engineered Systems Using Natural Language Processing," *Journal of Computing and Information Science in Engineering*, Vol. 23, No. 3, 2022. https://doi.org/10.1115/1.4054688.

[5] Walsh, H. S., and Andrade, S. R., "Semantic Search With Sentence-BERT for Design Information Retrieval," 2022. https://doi.org/10.1115/DETC2022-89557, URL https://doi.org/10.1115/DETC2022-89557, v002T02A066.

[6] Friedenthal, S., Moore, A., and Steiner, R., *A Practical Guide to SysML, Third Edition: The Systems Modeling Language*, 3rd ed., Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2014.

[7] Rauzy, A. B., and Haskins, C., "Foundations for model-based systems engineering and model-based safety assessment," *Systems Engineering*, Vol. 22, No. 2, 2019, pp. 146–155. https://doi.org/https://doi.org/10.1002/sys.21469, URL https://incose.onlinelibrary.wiley.com/doi/abs/10.1002/sys.21469.

[8] Evans, J., Cornford, S., and Feather, M. S., "Model based mission assurance: NASA's assurance future," *2016 Annual Reliability and Maintainability Symposium (RAMS)*, 2016, pp. 1–7. https://doi.org/10.1109/RAMS.2016.7448047.

[9] "OSMA Creates MBMA Program to Improve Integration of Assurance Considerations in MBSE," https://sma.nasa.gov/news/articles/newsitem/2018/06/27/osma-creates-mbma-program-to-improve-integration-of-assurance-considerations-in-mbse, 2018.

[10] Gharbi, A., Fischer, O., and Mavris, D. N., "Towards a Robust Computational Solution for the Verification and Validation of Complex Systems in MBSE using Wymore's Tricotyledon Theory of System Design," *AIAA SCITECH 2022 Forum*, ???? https://doi.org/10.2514/6.2022-0094, URL https://arc.aiaa.org/doi/abs/10.2514/6.2022-0094.

[11] Ciampa, P. D., and Nagel, B., "Accelerating the Development of Complex Systems in Aeronautics via MBSE and MDAO: a Roadmap to Agility," *AIAA AVIATION 2021 FORUM*, 2021. https://doi.org/10.2514/6.2021-3056, URL https://arc.aiaa.org/doi/abs/10.2514/6.2021-3056.

[12] Wei, R., Kelly, T. P., Dai, X., Zhao, S., and Hawkins, R., "Model based system assurance using the structured assurance case metamodel," *Journal of Systems and Software*, Vol. 154, 2019, pp. 211–233. https://doi.org/https://doi.org/10.1016/j.jss.2019.05.013, URL https://www.sciencedirect.com/science/article/pii/S0164121219301062.

[13] Doganaksoy, N., "Practical Reliability Engineering, 4th edition, Patrick D. T. O'Connor, Wiley, 2002, 540 pages," *Quality and Reliability Engineering International*, Vol. 21, 2005, pp. 841–841. https://doi.org/10.1002/qre.703.

[14] Rao, S., *Reliability Engineering*, 1$^{st}$ ed., Pearson, 2014.

[15] Planas, E., and Cabot, J., "How are UML class diagrams built in practice? A usability study of two UML tools: Magicdraw and Papyrus," *Comput. Stand. Interfaces*, Vol. 67, 2020. URL https://api.semanticscholar.org/CorpusID:198333579.

[16] Huang, Z., Swalgen, S., Davidz, H., and Murray, J., "MBSE-assisted FMEA approach — Challenges and opportunities," *2017 Annual Reliability and Maintainability Symposium (RAMS)*, 2017, pp. 1–8. https://doi.org/10.1109/RAM.2017.7889722.

[17] Chong, J., Zhou, H., Wang, M., and Chen, Y., "A Design Framework for Complex Spacecraft Systems with Integrated Reliability Using MBSE Methodology," *Signal and Information Processing, Networking and Computers*, edited by J. Sun, Y. Wang, M. Huo, and L. Xu, Springer Nature Singapore, Singapore, 2023, pp. 165–173.

[18] Izygon, M., Wagner, H., Okon, S., Wang, L., Sargusingh, M., and Evans, J., "Facilitating R&M in spaceflight systems with MBSE," *2016 Annual Reliability and Maintainability Symposium (RAMS)*, 2016, pp. 1–6. https://doi.org/10.1109/RAMS.2016.7448031.

[19] Evans, J. W., Groen, F. J., Wang, L., Austin, R., Witulski, A., Cornford, S. L., Feather, M., and Lindsey, N., "Towards a Framework for Reliability and Safety Analysis of Complex Space Missions," *19th AIAA Non-Deterministic Approaches Conference*, 2017. https://doi.org/10.2514/6.2017-1099, URL https://arc.aiaa.org/doi/abs/10.2514/6.2017-1099.

[20] Castet, J.-F., Bareh, M., Nunes, J., Okon, S., Garner, L., Chacko, E., and Izygon, M., "Failure analysis and products in a model-based environment," *2018 IEEE Aerospace Conference*, 2018, pp. 1–13. https://doi.org/10.1109/AERO.2018.8396736.

[21] Winton, D. B., and Carl Huang, Z., "MBSE and FMEA Integration Using GENESYS," *2021 Annual Reliability and Maintainability Symposium (RAMS)*, 2021, pp. 1–7. https://doi.org/10.1109/RAMS48097.2021.9605748.

[22] Yang, C., Letourneau, S., Zaluski, M., and Scarlett, E., "APU FMEA Validation and Its Application to Fault Identification," 2010, pp. 959–967. https://doi.org/10.1115/DETC2010-28438, URL https://doi.org/10.1115/DETC2010-28438.

[23] Haskins, B., Stecklein, J., Dick, B., Moroney, G., Lovell, R., and Dabney, J., "8.4.2 Error Cost Escalation Through the Project Life Cycle," *INCOSE International Symposium*, Vol. 14, No. 1, 2004, pp. 1723–1737. https://doi.org/https://doi.org/10.1002/j.2334-5837.2004.tb00608.x, URL https://incose.onlinelibrary.wiley.com/doi/abs/10.1002/j.2334-5837.2004.tb00608.x.

[24] Tanguy, L., Tulechki, N., Urieli, A., Hermann, E., and Raynal, C., "Natural language processing for aviation safety reports: From classification to interactive analysis," *Computers in Industry*, Vol. 78, 2016, pp. 80–95. https://doi.org/https://doi.org/10.1016/j.compind.2015.09.005, URL https://www.sciencedirect.com/science/article/pii/S0166361515300464.

[25] Kuhn, K. D., "Using structural topic modeling to identify latent topics and trends in aviation incident reports," *Transportation Research Part C: Emerging Technologies*, Vol. 87, 2018, pp. 105–122. https://doi.org/https://doi.org/10.1016/j.trc.2017.12.018, URL https://www.sciencedirect.com/science/article/pii/S0968090X17303881.

[26] Kulkarni, A., Terpenny, J., and Prabhu, V., "Leveraging Active Learning for Failure Mode Acquisition," *Sensors*, Vol. 23, 2023. https://doi.org/10.3390/s23052818.

[27] Bouabdallaoui, Y., Lafhaj, Z., Yim, P., Ducoulombier, L., and Bennadji, B., "Natural Language Processing Model for Managing Maintenance Requests in Buildings," *Buildings*, Vol. 10, 2020, p. 160. https://doi.org/10.3390/buildings10090160.

[28] Yun, H., Carlone, M., and Liu, Z., "Topic modeling of maintenance logs for linac failure modes and trends identification," *Journal of Applied Clinical Medical Physics*, Vol. 23, 2021. https://doi.org/10.1002/acm2.13477.

[29] Mbaye, S., Walsh, H. S., Jones, G., and Davies, M., "BERT-Based Topic Modeling and Information Retrieval to Support Fishbone Diagramming for Safe Integration of Unmanned Aircraft Systems in Wildfire Response," *2023 IEEE/AIAA 42st Digital Avionics Systems Conference (DASC)*, 2023. In Press.

[30] Chami, M., Zoghbi, C., and Bruel, J.-M., *A First Step towards AI for MBSE: Generating a Part of SysML Models from Text Using AI*, 2019, pp. 123–136.

[31] Chen, M., and Bhada, S. V., "Converting natural language policy article into MBSE model," *INCOSE International Symposium*, Vol. 32, No. S2, 2022, pp. 73–81. https://doi.org/https://doi.org/10.1002/iis2.12897, URL https://incose.onlinelibrary.wiley.com/doi/abs/10.1002/iis2.12897.

[32] Zhong, S., Scarinci, A., and Cicirello, A., "Natural Language Processing for systems engineering: Automatic generation of Systems Modelling Language diagrams," *Knowledge-Based Systems*, Vol. 259, 2023, p. 110071. https://doi.org/https://doi.org/10.1016/j.knosys.2022.110071, URL https://www.sciencedirect.com/science/article/pii/S0950705122011649.

[33] Walsh, H. S., Spirakis, E., Andrade, S. R., Hulse, D., and Davies, M., "SMARt-STEReO: Preliminary Concept of Operations," Tech. Rep. NASA/TM-20205007665, NASA, 2020.

[34] St. Denis, Lise A., Mietkiewicz, Nathan P., Short, Karen C., Buckland, Mollie, and Balch, Jennifer K., "All-hazards dataset mined from the US National Incident Management System 1999–2014," *Scientific Data*, Vol. 7, 2020, p. 64. https://doi.org/10.1038/s41597-020-0403-0.

[35] Grootendorst, M., "BERTopic: Neural topic modeling with a class-based TF-IDF procedure," *arXiv preprint arXiv:2203.05794*, 2022.

[36] Mei, Q., Shen, X., and Zhai, C., "Automatic Labeling of Multinomial Topic Models," *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, New York, NY, USA, 2007, p. 490–499. https://doi.org/10.1145/1281192.1281246, URL https://doi.org/10.1145/1281192.1281246.

[37] Mbaye, S., Jones, G., Infeld, S. I., Okon, S., and Davies, M. D., "A Model-Based Systems Engineering Evaluation of the Evolution to an In-Time Aviation Safety Management System," *AIAA AVIATION 2022 Forum*, 2022. https://doi.org/10.2514/6.2022-3423, URL https://arc.aiaa.org/doi/abs/10.2514/6.2022-3423.