BERT-based Topic Modeling and Information Retrieval to Support Fishbone Diagramming for Safe Integration of Unmanned Aircraft Systems in Wildfire Response

Seydou Mbaye, Hannah S. Walsh
Intelligent Systems Division
NASA Ames Research Center
Moffett Field, CA, USA
seydou.mbaye@nasa.gov;
hannah.s.walsh@nasa.gov

Garfield Jones
Department of Industrial and Systems Engineering
Morgan State University
Baltimore, MD, USA
garfield.jones@morgan.edu

Misty Davies
Intelligent Systems Division
NASA Ames Research Center
Moffett Field, CA, USA
misty.d.davies@nasa.gov

***Abstract*—Recent concepts for emerging wildfire response operations have included unmanned aircraft systems (UAS) due to their increasing accessibility and capabilities. To integrate UAS into wildfire response safely, researchers have studied the use of large repositories of historic incident reports to improve the scope of root cause analysis. Recent work has emphasized applying state-of-the-art natural language processing techniques to extract useful information from these repositories. However, it has not yet been studied how these results can be interpreted and integrated into the systems engineering process. In this work, we propose a process in which Bidirectional Encoder Representations from Transformers (BERT)-based topic modeling and information retrieval are applied to a relevant set of documents in order to support the development of a fishbone diagram in a semi-automated process. High-level themes in the document set are identified using topic modeling, which are then refined and interpreted by a human analyst. Then, the themes are used to guide a finer search using information retrieval, which returns specific incident reports of relevance. This provides traceability to specific incidents as well as broader categorizations that comprise the fishbone branches. We apply the proposed process to relevant documents from NASA's Aviation Safety Reporting System (ASRS). The proposed process is widely applicable when relevant documents are available, and the results from this study will be useful to identifying potential causes of wildfire response UAS incidents.**

***Index Terms*—hazard analysis, wildfire response, systems engineering, fishbone diagram, system safety**

I. Introduction

The increase in wildfire incidents and their complexity across the United States continues to put at risk the lives of wildland firefighters [1]. As unmanned aircraft systems (UAS) become more accessible, there are calls for their wider integration in wildfire management to offload inherent risk to humans. With their broad capabilities, UAS can be used for reconnaissance, heat mapping, search and rescue, and other tasks [2]. However, it is critical that their integration be done safely by leveraging available data that pertain to UAS-related incidents.

A possible direction in improving the coverage of root cause analysis is to leverage knowledge available from past incidents. While there are rich repositories of historic incidents available, they are often in natural language format, unstructured, and written in an inconsistent manner due to having many contributors. Metadata may be available but may be inconsistently filled out or be insufficiently descriptive for failure analyses. Therefore, it is difficult to efficiently capture knowledge from past incidents during the design process. A potential solution is the application of state-of-the-art natural language processing techniques to extract information efficiently [3], [4]. While several techniques have been developed to automatically or semi-automatically extract relevant information from such data sets, questions remain regarding how these results can be integrated into the systems engineering process. Specifically, to improve the adoption of such technologies, they should be presented in such a form that will be familiar, valid, and explainable within a systems engineering context.

In this paper, we present a process in which Bidirectional Encoder Representations from Transformers (BERT)-based topic modeling and information retrieval applied to a relevant data set can be used to support the development of a fishbone diagram depicting hazards that can lead to incidents. Then, we apply this process to the development of a fishbone diagram for a UAS in wildfire response mission context. Building upon prior work applying natural language processing techniques to extracting hazards and failures in large data sets, this work provides a process in which such techniques can be applied and integrated in such a way as to generate a widely used systems engineering diagram. Moreover, the process provides explainability by explicitly tracing results to specific, relevant historic examples.

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.

II. Background

Emerging concepts for extended use of UAS in wildfire response missions [5] will require identification of hazards and mitigation strategies. Prior work has developed capabilities to support and extend these efforts, in particular applying natural language processing to identify and analyze hazards from relevant historical documents [6]. In this work, we propose a process by which these capabilities can be integrated into the systems engineering process.

A. Artificial Intelligence and Machine Learning in Root Cause Analysis

Artificial Intelligence and Machine Learning (AI/ML) are being widely used in a variety of engineering applications, including root cause analysis (RCA). More recently, the AI/ML application has gained significant traction in RCA-processes due to its time-saving potential. Conducting traditional RCA is a time consuming and labor-intensive task where teams of engineers and analysts gather to manually go through databases of sometimes poorly structured documents, making it nearly impossible to learn meaningful information. Relevant work has determined that about 80% of information relevant to businesses are contained in semi-structured and/or unstructured data [7]. Researchers from the National Institute of Standards and Technology (NIST) emphasized how maintenance tickets used for RCA are typically in fragmented sentence structures and written in domain-specific language, making it hard for those unfamiliar with a given domain to make sense out of said ticket [8]. They proposed using natural language processing (NLP) to sort out a database and categorize entries that can otherwise seem dissimilar in an effort to identify emerging themes [8].

B. Fishbone/Ishikawa Diagrams

The Fishbone Diagram, also commonly referred to as an Ishikawa Diagram, traces its roots back to the 1960s [9]. Its names derives from its fish skeleton-like look and its initial pioneer, Japanese statistician Kaoru Ishikawa. The Fishbone diagram is a root cause analysis tool that allows stakeholders to identify and visualize existing or potential problems in a given system, in an effort to devise mitigation techniques. Fishbone diagramming is a common technique used in Systems Engineering, particularly in the design, development and deployment of complex systems. Developing a Fishbone diagram generally involves six generally accepted major steps [10]:

1) Main Event identification
2) Main Event formalization
3) Identification of causes leading to Main Event
4) Prioritization of causes
5) Diagram development
6) Diagram analysis and implementation of barrier measures

C. Natural Language Processing for Systems Engineering Applications

Natural language processing is a branch of artificial intelligence specialized for understanding human language in text or spoken form. Broadly, there is a strong precedent for applying natural language processing and text mining techniques for solving systems engineering and design problems. In particular, it has been used to automate learning about function knowledge [11], design ideas [12], and patent topics [13] from natural language text. Recent advances in the area, including the success of state-of-the-art methods such as Bidirectional Encoder Representations from Transformers (BERT) [14] have increased interest in applying these techniques to systems engineering and design. A Masked Language Model (MLM) is used to train a BERT model, where the model is trained to predict masked words. This allows the model to learn context from both directions, left and right ("bi-directional"). These models are trained over very large data sets, on the order of hundreds of thousands or millions of documents. The Manager for Intelligent Knowledge Access (MIKA) toolkit [6] packages state-of-the-art knowledge discovery [15]–[17] and information retrieval [18] techniques that are specifically tuned for technical engineering documents for use in systems engineering and design applications.

III. Method

The proposed method is divided into three parts: Theme Identification, Theme Refinement, and Fishbone Diagram Generation. Prior to implementing the method, a use case is defined and a data set identified. In the first part of the method, themes are identified using a BERT-based topic modeling approach. In the second part of the method, themes are refined and assigned specific historic evidence using information retrieval to search the data set for relevant documents. In the third part of the method, the results from the first two parts are interpreted and synthesized with the context of the use case in order to build a fishbone diagram in a systematic way that links elements to specific historic evidence. The proposed method is summarized in Fig. 1 and will be detailed in the remainder of this section. The natural language processing techniques used in this study are applied using the Manager for Intelligent Knowledge Access (MIKA) toolkit [6].

A. Use Case and Data

The proposed method is applied to a UAS mission in a wildfire response context. Possible uses for a UAS in wildfire response include surveillance, logistics delivery, and fire line monitoring [5]. We use data from NASA's Aviation Safety Reporting System (ASRS) [1], which we specifically filter for reports that describe incidents related to UAS. If applying the proposed method to a database without these filtering capabilities, it is possible to build a machine learning model to filter for relevant documents [15].

[1]https://asrs.arc.nasa.gov/

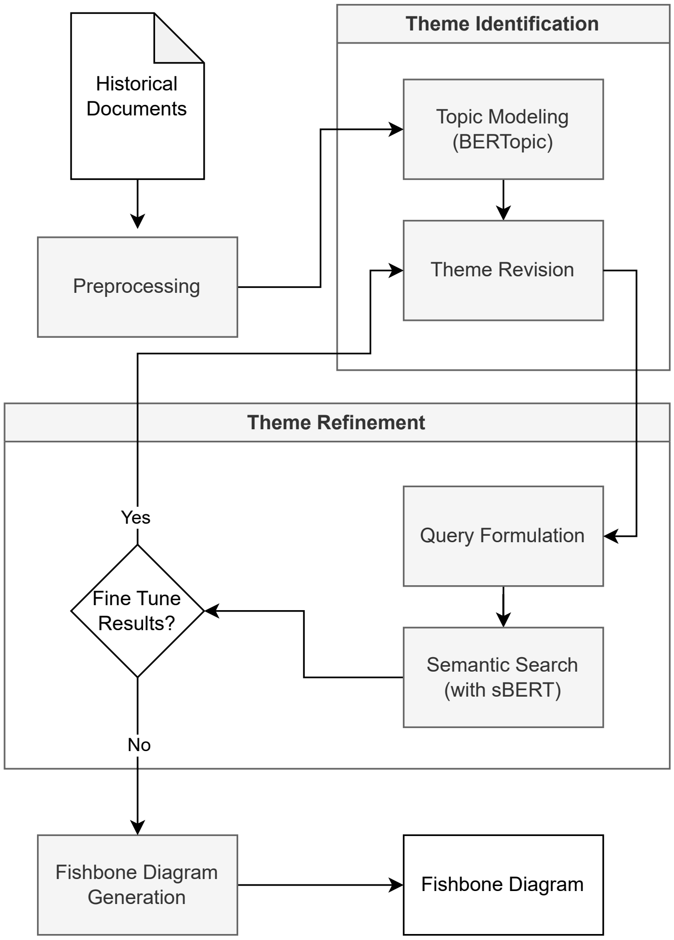

Fig. 1. Flowchart describing the proposed method.

B. Theme Identification

Themes are identified using a state-of-the-art topic modeling approach and then refined using the context of the use case and systems engineering judgment. Identifying the themes serves two purposes. First, it guides the development of the primary branches of the fishbone diagrams. Second, it guides the user's search queries for the theme refinement step later in the proposed method.

1) Topic Modeling: After identifying a main event of interest and a relevant data set, we first perform topic modeling. Specifically, we apply BERTopic via MIKA's interface. Topic modeling returns general themes present in the data set represented as a list of words shared by groups of documents representing those themes. The use of topic modeling for knowledge discovery in technical data sets has been covered extensively in past research [15]. There are multiple possible algorithms for topic modeling, including Latent Dirichlet Allocation (LDA) [19], Hierarchical LDA (hLDA) [20], and Hierarchical Dirichlet Process (HDP) [21]. More recently, state-of-the-art Bidirectional Encoder Representations from Transformers (BERT) models have been used for topic modeling in BERTopic [22]. In this work, we apply BERTopic to extract themes from the selected set of documents, though in general it would be equally possible to apply a different topic modeling approach such as LDA instead. We select BERTopic in this case because it requires minimal to no preprocessing and produces high-quality, easy-to-interpret results. BERTopic is implemented using the BERTopic package in Python [22].

2) Theme Revision: Once obtained from topic modeling, themes are revised based on relevance to the use case. They are also labeled manually by reviewing the list of words representing the topic as well as representative documents in the topic. These labels will become primary branches of the fishbone diagram.

C. Theme Refinement

Once the high-level themes in the data set are identified, they are refined using targeted information retrieval techniques. Specifically, while high-level themes guide the development of primary branches of the fishbone, the theme refinement step identifies secondary and tertiary branches of the primary branches. In contrast to the theme identification step, which extracts high-level topics from the data set, information retrieval allows users to search the data set using specific queries to retrieve individual documents that can be used to inform a particular element in the fishbone diagram. Connecting the themes to specific, historical evidence improves explainability of the results.

1) Query Formulation: The themes are used to elicit queries used to search the data set for specific historic incidents containing causes that may lead to the main event of interest. This allows a more informed and guided search than would otherwise be possible without deep knowledge of the data set. To fully capture the information need associated with each theme, we allow multiple queries per theme, if needed. The queries should also be targeted specifically towards the selected use case as much as possible. For example, if there is a theme about bird strikes, a possible query for a UAS-focused use case could be: "Do bird strikes affect UAS?". Well-formulated queries should accurately represent the information need, so we apply basic trial-and-error processes to improve the relevance of each query to the associated information need.

2) Semantic Search: Semantic search using sentence-BERT (sBERT) on a custom fine-tuned model available via MIKA is used for information retrieval. Sentence-BERT is a modification of BERT in which siamese and triplet structures compute sentence embeddings, making it more efficient for certain NLP tasks, including semantic search. The method has been covered extensively in prior work applying the technique to NASA's Lessons Learned Information System (LLIS), which also developed a model that has been fine-tuned on NASA projects [18]. Essentially, given a query entered by a user, the search system returns a ranked list of relevant documents from the data set. This is most useful in large data sets and when specific documents are required to be returned. Recently, state-of-the-art sentence-BERT methods have been applied to semantic search applications [23]. In contrast to searches that use lexical matching to retrieve results, semantic search additionally considers context of words.

The semantic search model used in this study has been selected and fine-tuned for technical, NASA-relevant applications in prior work [18]. The model is specialized for asymmetric search applications, which means the query and result pair are of significantly differing lengths. This option is suitable for search applications in which the user wants to input a short phrase or question in order to obtain a longer document. It is fine-tuned using data from NASA's Lessons Learned Information System (LLIS), which describes lessons learned from NASA projects. The fine-tuning leads to higher performance for the engineering application studied in prior work [18]. The model is used to generate embeddings for the corpus (set of documents to search), which can be stored and referenced whenever a search is run. Embeddings for the query are also computed and compared to the corpus embeddings using cosine similarity. Cosine similarity is defined in Eq. 1, where cosine similarity $cos\theta$ is the cosine of the angle between two vectorized documents D_i and D_j. The search returns a ranked list of k hits. The search is implemented using sentence-transformers in Python [23].

$$cos\theta = \frac{D_i \cdot D_j}{||D_i||||D_j||} \tag{1}$$

At this point, the user may assess the search results and refine the themes and queries. In particular, the user may wish to add a theme if they believe there may be one missing or may add a query to better represent a theme. This is a step in the process in which systems engineering judgment can be used to improve the results, while maximizing the utility of the information retrieval process.

D. Fishbone Diagram Generation

The themes and query results are then used to build a fishbone diagram. The use case definition is used to identify the main event in the fishbone diagram, which is the rightmost element into which all the branches flow. Since themes represent higher-level causes and individual documents retrieved via a query describe specific causes, themes represent primary branches of the fishbone diagram while causes described in individual documents fill the sub-branches. The revised set of themes identified in the Theme Identification step are first used to fill primary branches of the fishbone diagram. Primary branches of the fishbone are often organized with the most frequently occurring or impactful causes being nearest to the main event. We apply a simple counting scheme to the information retrieval results by summing the occurrences of each cause within the retrieved documents, which is determined by a human reader. Then, we arrange the primary branches in a ranked order, with the cause that is most frequently occurring in the reviewed documents being placed closest to the main event.

Secondary and tertiary branches are determined based on the search results from the Theme Refinement step in the proposed method. Most frequently, multiple sub-categories of the main themes identified in the Theme Identification step are identified through the search results. Sometimes, when multiple queries are associated with a single theme, each query becomes its own secondary branch. Tertiary branches are used in the case that a third level category emerges from this process. Like the primary branches, the secondary and tertiary branches are labeled by a human analyst based on the content of the search result(s) represented by each branch. After reviewing all results, we review whether the set of queries is complete and may iterate the process as needed.

IV. Results

A sample of topics identified using BERTopic is given in Table I. Thirty-eight total themes are identified. Themes are provided with the top five words used to describe the associated topic. In the study, the top ten words are returned and analyzed; however, Table I shows only the top five. In Table I, note that the algorithm sometimes returns acronyms which, in the results in this paper, are presented with the acronym definition. Topics are interpreted as general themes in the dataset. For example, the topic "object, small, appeared, quickly, passed" can be interpreted as a small object (UAS or suspected UAS) being spotted flying near a manned aircraft. Other topics, upon review, are excluded from further analysis due to being out of scope. For example, the topic "aircraft, runway, pattern, fuel, downwind" includes a document in which a pilot assumed a communications station was unmanned (rather than including description of an unmanned aircraft system). It is possible that a finer pre-analysis review could remove these documents, but the proposed process includes manual review (thereby still removing irrelevant information from the final results) and topic modeling can make this review more efficient than reviewing each document individually.

Queries are developed from reviewing the themes. Continuing with the example of the topic "object, small, appeared, quickly, passed", we generate the following relevant queries: "Do drones ever collide with aircraft?", "Is evasive action required when a drone flies near an aircraft?", "What near misses have happened with UAS?", and "Are drones a risk to helicopters?". The wording of these queries need not be explicitly stated in the theme, but rather flow logically while considering the theme and its implications in the chosen use

TABLE I
A sample of topics identified using BERTopic with the top five words used to describe the associated topic.

Topic Words
airspace, flight, authorization, low altitude authorization and notification capability (LAANC), application
aircraft, flight, runway, engine, airport
aircraft, airspace, sector, controller, supervisor
aircraft, runway, pattern, fuel, downwind
drone, runway, feet, approximately, air traffic control (ATC)
altitude, aircraft, descent, FL-190, climb
drone, officer, final, tower, runway
object, small, appeared, quickly, passed
drone, landing, left, public aircraft operations (PAO), drones
uav, aircraft, approach, flying, route orientation scheme (ROS)

case. Table II shows the list of queries developed from the identified themes. Multiple similar queries are included to elicit a more complete set of documents. Twenty-two queries in total are used. Initially, only thirteen queries are considered, and the remainder are added after reviewing the results from the original thirteen. Additional queries can be added if the results from the first set elicit further safety concerns that need to be investigated or if certain queries do not elicit expected results.

TABLE II
LIST OF QUERIES DEVELOPED FROM THE IDENTIFIED THEMES.

Query
Do bird strikes affect UAS?
Are there issues with drones landing on certain surfaces?
Do drones ever collide with aircraft?
Do drones ever interfere with landing?
Do drones interfere with weather balloons?
Is evasive action required when a drone flies near an aircraft?
What near misses have happened with UAS?
When there are waivers for drones, are there additional risks to the mission?
Are drones a risk to helicopters?
What happens when there is a lost link with a drone?
Is drone battery reliable for the mission?
What if there is a lost link with a drone?
What if a pilot loses contact with a drone?
Do drones interfere with takeoff?
Do drones fly above 400 ft. above ground level?
Do drones fly in restricted areas?
Do drones collide with terrain?
How often do drones fly without waivers when a waiver is required?
Do drones have difficulties landing on certain surfaces?
Have drones collided with buildings?
Have drones collided with trees?
Have drones collided with objects?

In Table III, the frequencies of occurrence of causes described in each primary branch in the reports studied are presented. These frequencies are counted by manually assessing the primary cause present in the identified document out of all search results. Irrelevant search results are ignored. The frequencies are used to organize the ordering of the primary branches of the fishbone diagram, although they are only representative of the frequencies within the documents studied and are not necessarily representative of the true value of the frequencies of these causes.

TABLE III
FREQUENCY OF OCCURRENCE OF CAUSES DESCRIBED IN EACH PRIMARY BRANCH OF THE FISHBONE.

Rank	Branch	Count
1	Flying in Restricted Area	58
2	Proximity to Threat	18
3	Communications Loss	13
4	Traffic Control	7
5	External Actors	6
6	Equipment Failure	4
7	Power Loss	3

Table IV presents the top three search results from two queries, the first one a continuation of the example above and the second a query from another theme. In the study, the top five search results are returned and analyzed. Text is reduced for brevity. The first query shows three fairly straightforward cases of drones flying close to aircraft. For the second query, in the first result, the battery rapidly discharged and the UAS did not have sufficient battery to complete the mission or return to base. The second result describes a scenario in which the UAS pilot “lost track of time”, which in this case led to an airspace violation but in other cases could lead to a depletion of battery life. In the third instance, the UAS battery disconnected from the UAS in a crash and had to be manually located by the pilot, posing a fire risk if not located. Each represents a different series of events that could (or did) lead to an incident related to the drone battery. The search portion of the process is intended to find unanticipated risks that are present in large sets of historical documents.

TABLE IV
TOP THREE SEARCH RESULTS FROM A TWO QUERIES USED IN THE STUDY.

No.	Result
Query: Do drones ever collide with aircraft?	
1	I saw something below us that made me do a double-take... realized it was remote-controlled quadcopter-style helicopter drone!
2	...an Air Marshal informed me he saw a drone just off the left wingtip on short final... It was close enough for him to see the propeller blades on the drone.
3	...something caught my eye ahead... As we got closer; I realized I was seeing a ... drone.
Query: Is drone battery reliable for a mission?	
1	During a routine mission; the drone battery rapidly discharged and uncontrollably landed... To prevent this in the future drones will not remain in flight with less than 25% battery life.
2	I was flying drone within 3 miles of the sporting event... I had lost track of time and didn’t realize I had violated the time restriction...
3	...The drone was lost in grass seed bushes that were at least 4-5ft tall... if the battery is disconnected while it is in the ’on’ mode; the battery will continue to discharge...

The fishbone diagram generated using the proposed method is provided in Fig. 2. The final fishbone has seven primary branches: Flying in Restricted Area, Proximity to Threat, Communications Loss, Traffic Control, External Actors, Equipment Failure, and Power Loss. Each has between three and four secondary branches, with two secondary branches having one or two tertiary branches.

The “Flying in Restricted Area” branch represents the potential causes related to flying a UAS in an area where it is not permitted. The majority of restricted area violations, as shown by the sub branches, were “flying above 400 feet above ground level (AGL)” and/or “flying near an airport.” As it stands, the Federal Aviation Administration, under 14 Code of Federal Regulation (CFR) Part 107, stipulates that no UAS flying in the National Airspace System can fly over 400 feet AGL or near an airport without a Certificate of Waiver or Authorization [24]. “Proximity to Threat” depicts the the potential hazards inherent to flying a UAS to in close proximity to a threat such as birds, tress, or other objects like weather balloons. The next branch, “Communication Loss,” represents causes related to lost link, control unresponsiveness, application or video feed failure, which make up its sub branches. Morevover, the Lost Link sub branch has tertiary

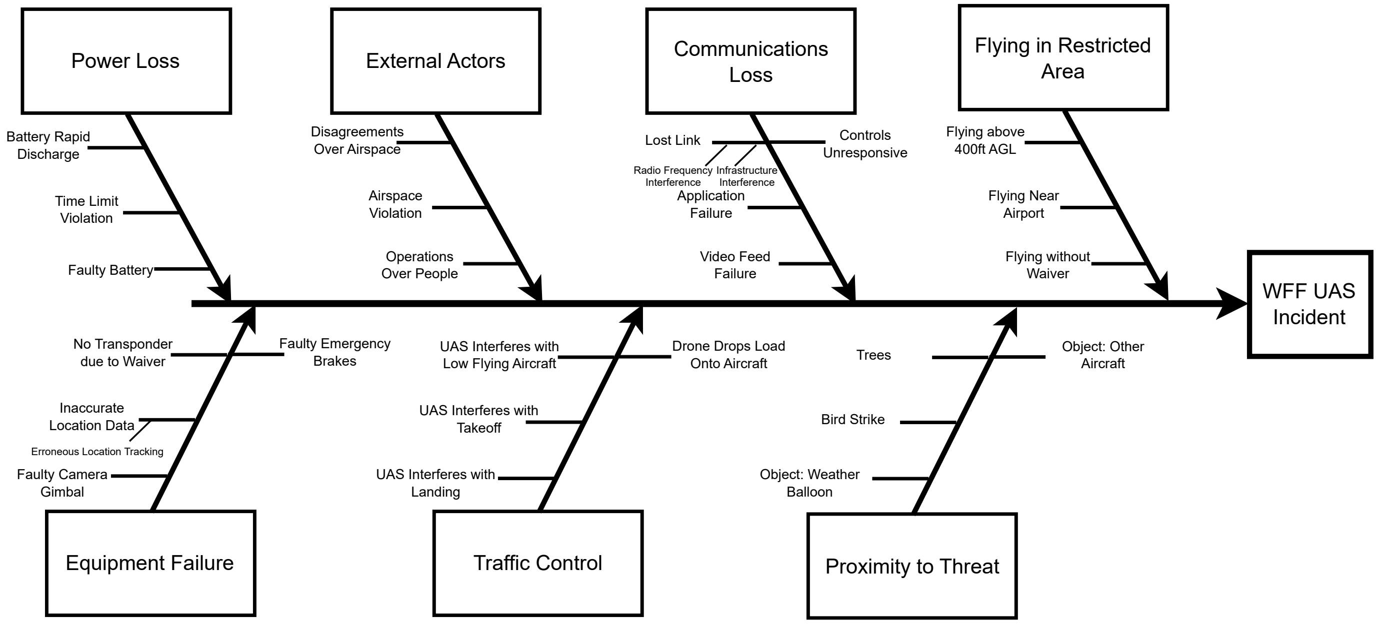

Fig. 2. Fishbone diagram describing potential causes of a wildfire fighting UAS incident generated using the proposed method.

branches (RF interference, infrastructure interference) that can cause it. The branch "Traffic Control" shows how a UAS violating restricted airspaces can interfere with takeoff and landing, as well as low flying aircraft such helicopters. External factors, which extend beyond the control of a UAS or ground control crew, such as weather, violation of airspace specific to UAS can also lead to an incident during a wildfire fighting mission. "Equipment Failure", another branch on the proposed fishbone diagram, can be the product of erroneous data, faulty camera gimbal, for example. Finally, the last branch of the diagram, "Power Loss" is a failure state caused in part by battery rapid discharge, time limit violation, or a faulty battery. Here, it is worth noting there may be some other external factors, such as high wind gusts, that can contribute to a healthy battery rapidly discharging and leading up to a total power loss.

V. Discussion

Wildfire incidents pose a significant challenge due to their complexity and dynamic nature, which can have fatal consequences for humans and lasting negative impacts on the environment. While the majority of wildfire management efforts are carried out by humans on the ground or airborne, roughly 26% of wildfire management fatalities are linked to aviation incidents [25]. This has prompted fire management entities and NASA to explore novel unmanned ways to fight wildfires, such as unmanned aircraft systems (UAS), which can offload inherent risks to humans.

Integrating UAS into wildfire management presents benefits such as better fuel mapping, situational awareness, and real-time data, but it also introduces new risks that must be carefully managed and mitigated [26]). One way to mitigate these risks is to proactively leverage historical data and conduct thorough root cause analysis to 1) identify areas of failure in wildfire fighting, 2) identify failure causes, and 3) draw barrier measures. The fishbone diagram produced through the proposed method identifies potential risks that could lead to a wildfire fighting UAS incident by using data available in the Aviation Safety Reporting System (ASRS). While the ASRS offers a large variety of safety reports, those reports can be extensive, often duplicates, and making sense out of and using them can be labor intensive.

Natural language processing (NLP) can be used to analyze that large and rich database and extract relevant information to develop fishbone diagrams. By using NLP, relevant data can be extracted from various sources, such as news articles and incident reports, and analyzed to identify factors that can potentially contribute to UAS wildfire fighting incidents. By analyzing and having a holistic view of the causes and effects relative to the vulnerabilities that exist or have the potential to be present in the system of interest, effective mitigation strategies can be put in place early in the design process to increase reliability and resiliency. These factors can then be used to develop a fishbone diagram that maps out the chronological order of events leading to a UAS incident. This approach, when integrated with systems engineering principles, can enable the design and deployment of UAS operations with a risk management approach, leading to effective and efficient wildfire response.

VI. Conclusions and Future Work

In this study, we propose a process for generating a fishbone diagram using topic modeling and semantic search. We apply the process to a use case of a UAS used in a wildfire response scenario and apply the technique to relevant ASRS documents about UAS incidents, using human systems engineering judgment at multiple points in the partially automated process. The results represent a first step towards a more formal integration of natural language processing-enabled analysis of historical technical documents in the systems engineering process that is more easily understood by systems engineers than raw natural language processing output and explainable in the sense that the diagram elements can be traced to specific historical documents. Moreover, the results can support root cause analysis in emerging wildfire response missions.

In future work, we plan to study the use of the proposed process within a real systems engineering context in order to validate the results and measure the improvements to root cause analysis. Additionally, we plan to further investigate improvements to the process - i.e., how to most efficiently perform the proposed steps by the user in a real project and whether more steps can and should be automated. We also plan to extend this process so it can be used with multiple data sets, which would be particularly useful for the novel UAS mission for wildfire response use case that we describe in this paper (e.g., a second data set describing wildfire response incidents could be considered).

REFERENCES

[1] T. Schoennagel, J. K. Balch, H. Brenkert-Smith, P. E. Dennison, B. J. Harvey, M. A. Krawchuk, N. Mietkiewicz, P. Morgan, M. A. Moritz, R. Rasker, M. G. Turner, and C. Whitlock, "Adapt to more wildfire in western North American forests as climate changes," *Proceedings of the National Academy of Sciences*, vol. 114, no. 18, pp. 4582–4590, May 2017. [Online]. Available: https://pnas.org/doi/full/10.1073/pnas.1617464114

[2] M. A. Akhloufi, N. A. Castro, and A. Couturier, "Unmanned Aerial Systems for Wildland and Forest Fires," *Drones*, vol. 5, no. 1, p. 15, Feb. 2021, arXiv:2004.13883 [cs, eess]. [Online]. Available: http://arxiv.org/abs/2004.13883

[3] K. Bot and J. G. Borges, "A systematic review of applications of machine learning techniques for wildfire management decision support," *Inventions*, vol. 7, no. 1, 2022. [Online]. Available: https://www.mdpi.com/2411-5134/7/1/15

[4] B. Baharudin, L. H. Lee, K. Khan, and A. Khan, "A review of machine learning algorithms for text-documents classification," *Journal of Advances in Information Technology*, vol. 1, 02 2010.

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

[6] S. Andrade, D. Hulse, L. Irshad, and H. S. Walsh, "Supporting Hazard Analysis for Wildfire Response Using fmdtools and MIKA," NASA, Technical Memorandum NASA/TM—20220014099, 2022.

[7] A. Soti and Z. Dobos, "NLP Analysis of Incident and Problem Descriptions," *25th Annual EurOMA Conference*, 6 2018.

[8] M. Sharp, T. Sexton, and M. Brundage, "Toward semi-autonomous information extraction for unstructured maintenance data in root cause analysis," in *Advances in Production Management Systems*, 07 2017.

[9] E. M. Loredana, "The Analysis of Causes and Effects of a phenomenon by Means of the "Fishbone" Diagram," *Annals of Constantin Brancusi University of Targu Jiu, Economics Series*, 6 2017.

[10] C. Ciocoiu and G. Ilie, "Application of fishbone diagram to determine the risk of an event with multiple causes," *Management Research and Practice*, vol. 2, pp. 1–20, 01 2010.

[11] H. Cheong, W. Li, A. Cheung, A. Nogueira, and F. Iorio, "Automated Extraction of Function Knowledge From Text," *Journal of Mechanical Design*, vol. 139, no. 11, 10 2017. [Online]. Available: https://doi.org/10.1115/1.4037817

[12] F. Ahmed, M. Fuge, and L. D. Gorbunov, "Discovering Diverse, High Quality Design Ideas From a Large Corpus," in *IDETC-CIE2016*, Volume 7: 28th International Conference on Design Theory and Methodology, 8 2016, v007T06A008. [Online]. Available: https://doi.org/10.1115/DETC2016-59926

[13] U. H. Govindarajan, A. J. Trappey, and G. Kumar, "Latent Dirichlet Allocation Modeling for CPS Patent Topic Discovery," in *Proceedings of the 2018 International Conference on Industrial Enterprise and System Engineering (IcoIESE 2018)*. Yogyakarta, Indonesia: Atlantis Press, 2019. [Online]. Available: https://www.atlantis-press.com/article/55914786

[14] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," 2018, publisher: arXiv Version Number: 2. [Online]. Available: https://arxiv.org/abs/1810.04805

[15] S. R. Andrade and H. S. Walsh, "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, 8 2022.

[16] ——, "Wildfire Emergency Response Hazard Extraction and Analysis of Trends (HEAT) through Natural Language Processing and Time Series," in *2021 IEEE/AIAA 40th Digital Avionics Systems Conference (DASC)*, 10 2021, pp. 1–10, iSSN: 2155-7209.

[17] ——, "What Went Wrong: A Survey of Wildfire UAS Mishaps through Named Entity Recognition," in *2022 IEEE/AIAA 41st Digital Avionics Systems Conference (DASC)*, 9 2022, pp. 1–10, iSSN: 2155-7209.

[18] H. S. Walsh and S. R. Andrade, "Semantic Search With Sentence-BERT for Design Information Retrieval," in *IDETC-CIE2022*, Volume 2: 42nd Computers and Information in Engineering Conference (CIE), 8 2022, v002T02A066.

[19] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 3 2003.

[20] D. M. Blei, M. I. Jordan, T. L. Griffiths, and J. B. Tenenbaum, "Hierarchical Topic Models and the Nested Chinese Restaurant Process," in *Proceedings of the 16th International Conference on Neural Information Processing Systems*, ser. NIPS'03. Cambridge, MA, USA: MIT Press, 2003, pp. 17–24.

[21] Y. Teh, M. Jordan, M. Beal, and D. Blei, "Sharing Clusters among Related Groups: Hierarchical Dirichlet Processes," in *Advances in Neural Information Processing Systems*, L. Saul, Y. Weiss, and L. Bottou, Eds., vol. 17. MIT Press, 2004.

[22] M. Grootendorst, "BERTopic: Neural topic modeling with a class-based tf-idf procedure," 2022.

[23] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," 2019, publisher: arXiv Version Number: 1. [Online]. Available: https://arxiv.org/abs/1908.10084

[24] "14 CFR Part 107 – Small Unmanned Aircraft Systems." [Online]. Available: https://www.ecfr.gov/current/title-14/chapter-I/subchapter-F/part-107

[25] C. R. Butler, M. B. O'Connor, and J. M. Lincoln, "Aviation-Related Wildland Firefighter Fatalities — United States, 2000–2013." [Online]. Available: https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6429a4.htm

[26] S. Mbaye, G. Jones, and M. D. Davies, "Analysis of input from wildfire incident experts to identify key risks and hazards in wildfire emergency response," *AIAA SCITECH 2023 Forum*, Jan. 2023. [Online]. Available: https://arc.aiaa.org/doi/abs/10.2514/6.2023-2710