INTERNATIONAL DESIGN CONFERENCE – DESIGN 2024
https://doi.org/10.1017/pds.2024.204

Integrating large language models for improved failure mode and effects analysis (FMEA): a framework and case study

Ibtissam El Hassani [1,2], Tawfik Masrour [1,2], Nouhan Kourouma [1], Damien Motte [3] and Jože Tavčar [3,✉]

[1] Moulay Ismail University, Morocco, [2] University of Quebec at Rimouski, Canada, [3] Lund University, Sweden

✉ joze.tavcar@design.lth.se

Abstract

The manual execution of failure mode and effects analysis (FMEA) is time-consuming and error-prone. This article presents an approach in which large language models (LLMs) are integrated into FMEA. LLMs improve and accelerate FMEA with human in the loop. The discussion looks at software tools for FMEA and emphasizes that the tools must be tailored to the needs of the company. Our framework combines data collection, pre-processing and reliability assessment to automate FMEA. A case study validates this framework and demonstrates its efficiency and accuracy compared to manual FMEA.

Keywords: failure mode and effect analysis (FMEA), large language model (LLM), generative AI, product quality, knowledge management

1. Introduction

Failure mode and effects analysis (FMEA) has long been a cornerstone of the product development process (PDP). It provides a systematic approach to identifying potential failure modes and their effects, and help mitigating risks. By proactively addressing risks during the engineering design phase, FMEA plays a critical role in ensuring product quality, reliability and customer satisfaction. However, the traditional manual execution of FMEA has its own challenges: a labor-intensive process, susceptibility to human error and difficulty in comprehensively analyzing complex designs. State-of-the-art generative artificial intelligence (AI) techniques in the field of FMEA could be an answer to these challenges: Integrating AI into the FMEA process would make it possible to automate the identification of failure modes, and ultimately enable a more efficient and reliable PDP.

Among the available generative AI techniques, large language models (LLMs) have attracted considerable attention, highlighted by the introduction of ChatGPT (Zhao et al., 2023). LLM systems have the potential to extract, process and generate valuable data from both unformatted and formatted documents. LLMs therefore seem to be very relevant for FMEA: FMEA uses very diverse datasets, from previous FMEA reports and product history files to formal complaints and customer reviews, which often need to be processed manually and therefore cannot always be fully utilized. More generally, FMEA tools based on LLMs could save time, reduce errors and help in the development of robust designs.

However, while LLMs can be used for knowledge-intensive tasks with little or no training in prompting engineering (e.g. Dell'Acqua et al., 2023), FMEA tasks require the development of specialized tools and data management. Therefore, a framework (i.e., a process model and an information system model) for integrating LLMs into the FMEA process is proposed in this paper, along with a case study. The

case study presented does not cover the entire framework but serves to illustrate the potential benefits of integrating LLMs into FMEA and to evaluate the accuracy and relevance of the generated results.
This paper is structured as follows: A literature review presents the state of the art, followed by a more detailed compilation of the advantages and limitations of using LLMs in the context of FMEA. On this basis, a framework is proposed, and the case study is reported on. Finally, the lessons learned are presented along with recommendations for future research in this area.

2. Literature review

2.1. Evolution of FMEA

In the 1950s and 1960s, the aerospace and defence industries intensified their efforts to identify potential failure modes in complex systems and to understand their effects. In the early 1960s, the United States Department of Defense issued a military standard called MIL-STD-1629A. In the 1960s and 1970s, the National Aeronautics and Space Administration (NASA) introduced FMEA as part of its mission-critical processes. As FMEA gained acceptance in various industries, efforts to standardize it began. In 1985, the International Electrotechnical Commission (IEC) published the first edition of IEC 60812 and in 1994, the Society of Automotive Engineers (SAE) published the first edition of SAE J1739, which provided guidelines for conducting FMEA during the design phase. The current standard for automotive industry, IATF 16949:2016, requires companies to record methods for the management of product safety-related products and manufacturing processes, including FMEA. FMEA is also an established process for improving production quality and minimizing the severity and occurrence of defects through the use of corrective actions (Huang, et al., 2019). Over the years, FMEA has evolved and branched out into different variants, including design FMEA, process FMEA and system FMEA, each tailored to specific aspects of processes and systems (AIAG and VDA, 2019; Soltanali and Ramezani, 2023). The methodology has become an integral part of quality management systems and risk assessment procedures in industries such as aerospace, automotive, manufacturing, healthcare, and others (IEC 60812:2018). Process and design FMEA are now an integral part of product and process development.
However, FMEA workshops are time-consuming and labor-intensive (Tavčar and Duhovnik, 2014; Thomas, 2023). Therefore, over the years, various methods and automation techniques to improve the efficiency and effectiveness of FMEA, including computerized FMEA generation, have been developed. Among others, the application of AI in engineering and risk analysis has gained considerable attention (see next section) and opens up new opportunities for the improvement of FMEA processes, including the perspective of meeting the needs of the industry and realizing semi- and full automation of the complex work of FMEA (Wu et al., 2021).

2.2. Advances in AI for FMEA

Several studies have investigated AI-driven approaches for FMEA or related risk assessment processes. As early as 1996, Wirth et al. (1996) argued that a knowledge-based approach to FMEA could improve the traditional way of performing an FMEA by using various knowledge bases to support accurate product descriptions with controlled vocabulary and facilitate the subsequent reuse of knowledge gathered during an FMEA.
Recent developments in AI have opened up new opportunities to improve FMEA. Liu et al. (2019) highlighted how multi-criteria decision making methods can support risk assessment in FMEA, while Soltanali et al. (2023) proposed a smart FMEA platform with hybrid FMEA models that combines uncertainty quantification, machine learning techniques and multi-criteria decision making. Na'amnh et al. (2021) present improved risk assessment models using fuzzy inference and neural networks that outperform classical methods, with the fuzzy model proving superior for decision making.
Furthermore, researchers have explored data-driven approaches using machine learning to continuously update and predict risk priority numbers (RPNs) for new failure modes (Peddi et al., 2023). Hassan et al. (2023) successfully used historical data and convolutional neural networks (CNNs) to automate the prioritization of contract requirements. Yucesan et al. (2021) used fuzzy best-worst and fuzzy Bayesian network methods to evaluate risk parameters in FMEA. Data-driven design has also gained attention as

https://doi.org/10.1017/pds.2024.204 Published online by Cambridge University Press

Filz et al. (2021) demonstrated the benefits of combining data from past maintenance events with employee experience to support maintenance planning. In addition, Hodkiewicz et al. (2021) have presented the application of ontological approaches to improve the explicit representation of FMEA-related concepts.

The integration of LLMs, in particular ChatGPT, into the FMEA process has attracted interest. ChatGPT's ability to understand context and learn from new data offers potential benefits in FMEA tasks (Thomas, 2023). Implementing ChatGPT in FMEA is about utilizing its core functionality and enhancing it with company-specific knowledge (Diemert and Weber, 2023). Combining AI tools such as ChatGPT and human expertise would enhance the strengths of both in the FMEA process. Studies that combine FMEA with the use of LLM techniques are still few. One example is the study by Spreafico and Sutrisno (2023), which presents a method in which a chatbot is used for automatic social failure analysis in product sustainability. Three case studies confirm the potential of the method and at the same time show the limitations of the chatbot.

The literature reviewed emphasizes the importance of the integration of AI to further improve FMEA. Combining AI tools with human expertise is seen as a way to achieve better results in FMEA. However, despite the progress made in AI-driven approaches to FMEA, there are still gaps that need to be addressed. Firstly, while some studies have examined specific AI techniques for FMEA, there is a need for a comprehensive framework that provides a systematic approach to utilizing AI capabilities throughout the entire FMEA process. Secondly, existing studies often focus on the technical aspects of AI integration but may overlook the practical challenges of its implementation. Therefore, there is a need for case studies that present implementations of LLM along the different steps of FMEA in different contexts.

To address these gaps, this research paper proposes the development of a framework (process model and information system model) that integrates LLM into the FMEA process and illustrates its practical implementation in a case study.

3. Advantages and limitations of LLMs in FMEA

By applying the general advantages and limitations of LLMs found in the literature (e.g. Bommasani et al., 2021; Hu et al., 2023; Thirunavukarasu et al., 2023) to the context of FMEA and based on the authors' experience in LLMs and FMEA, a set of advantages and limitations of LLMs in FMEA could be compiled.

3.1. Advantages and Contributions of LLMs in FMEA

The main potential advantages of integrating LLMs into FMEA and their contribution to risk analysis in the PDP are listed below:

- Knowledge and expertise: LLMs can be trained on a large amount of textual data, including technical and engineering information. They can provide accurate and up-to-date knowledge related to FMEA methods, best practices and industry standards. By utilizing the knowledge base of LLMs, engineers can gain insight on FMEA concepts, processes and techniques.
- Data analysis support: LLMs can help engineers analyze and interpret data relevant to FMEA. They can assist in pre-processing data and identifying patterns or correlations within the data. The capabilities of LLMs can be particularly valuable when it comes to extracting failure-related information from textual data, which can help identify failure modes.
- Supporting risk assessment: LLMs can support risk assessment calculations by providing guidance on factors such as severity, occurrence, and detection. They can help engineers estimate the likelihood and effect of failure modes and prioritize their risk mitigation efforts.
- Decision support: LLMs can help engineers make informed decisions during the FMEA process. They can provide recommendations for prioritizing failure modes, selecting appropriate corrective measures, and evaluating the effectiveness of mitigation strategies.
- Documentation: LLMs' support in organizing and documenting relevant information ensures that valuable findings and knowledge are properly recorded and shared with team members.

https://doi.org/10.1017/pds.2024.204 Published online by Cambridge University Press

By leveraging AI capabilities, engineers can streamline their FMEA activities, reduce manual effort and improve the overall quality of the analysis.

3.2. Limitations of LLMs in FMEA

Although LLMs provide valuable support in the area of FMEA, it is important to recognize their limitations and potential drawbacks. These limitations should be considered when using LLMs in the FMEA process. The following list outlines some notable limitations:

- Security limitations: LLMs may have vulnerabilities in terms of security, such as susceptibility to hostile attacks or data privacy concerns. Ensuring the security of data and the local LLM model itself is critical when integrating LLMs into FMEA processes, especially considering that LLMs are often not used locally and information can be transferred to external servers or to the owners of LLMs. This brings the additional challenge of protecting sensitive data transmission and assuring that external parties such as LLMs owners have appropriate security practices.
- Lack of contextual understanding: LLMs operate based on patterns and associations learned from training data that may not provide a deep understanding of the specific context and nuances of FMEA in different industries or technical domains. This limitation can lead to incomplete or inaccurate answers being. That require careful interpretation and verification by domain experts.
- Potential biases: Like any machine learning model, LLMs can unintentionally produce biased or subjective responses based on biases in the training data. These biases can influence the guidance or recommendations provided by LLMs in the FMEA process. Critical evaluation and cross-referencing to different sources of information is necessary to mitigate potential biases.
- Limited adaptability: LLMs' responses are based on pre-existing patterns in the training data and may not be adaptable to unique or highly specialized FMEA scenarios. Unconventional or complex cases may require expertise or tailored approaches beyond the capabilities of LLMs.

4. Framework for integration of LLM in FMEA

Based on the advantages of integrating LLMs into FMEA, a process model and an information system model are proposed. The proposed process model consists of the following:

1. **Data collection:** relevant data is collected from various sources, including design data, historical failure data and any additional contextual information. This data serves as the basis for training AI algorithms and provides insights for risk analysis.
2. **Data pre-processing:** The collected data undergoes pre-processing to ensure its quality and compatibility with the LLM. This step may include data cleaning, normalization, feature extraction and the handling of missing values or outliers. The aim is to carry out pre-processing automatically using computer-aided tools.
3. **Model training:** Different data subsets (previous FMEAs, external reviews, etc.) are labelled, i.e. the expected output (failure mode, effect, risk assessment, corrective action, etc.) from each subset is manually entered by engineers. The system is then trained on the set of labelled data until a satisfactory accuracy is obtained.
4. **Extraction of specific elements of FMEA information**: Once trained, the system is applied to the entire dataset and suggests failure modes, effects, risk assessment calculations and corrective actions.
5. **Decision support and regular improvement of FMEA information:** The generated FMEA information is part of a knowledge management system (KMS). The KMS can generate FMEA reports, visualizations, comprehensive summaries, and trends that can be used for current and future analysis and decision-making in FMEAs. Beyond FMEA, the KMS can be part of the quality assurance system of the company and can for example help in detecting need for competencies, recurring failure modes, identify company processes bottlenecks, etc.

The information system model for the integration of LLM tools for data collection, extraction, knowledge management and use in an industrial environment is shown in Figure 1. It assumes that company-specific knowledge can be extracted from key documents in its product lifecycle management (PLM) system such as previous FMEAs, engineering changes (ECs), product history files, etc. (Tavčar et al., 2019). In addition, AI analysis takes into account external sources related to the company's production program and technology-specific information. Especially when a new product is being developed, external sources of knowledge are of greater importance. The selected documents are included in the AI analysis and the extracted information is reused in the FMEA activity and other (PDPs). Systematic data analysis is performed at regular intervals and corrective actions are defined for the PDP, the engineering change management (ECM) and FMEA process (Figure 1). LLM tools are very powerful for analysis as they can identify different patterns in huge amounts of data.

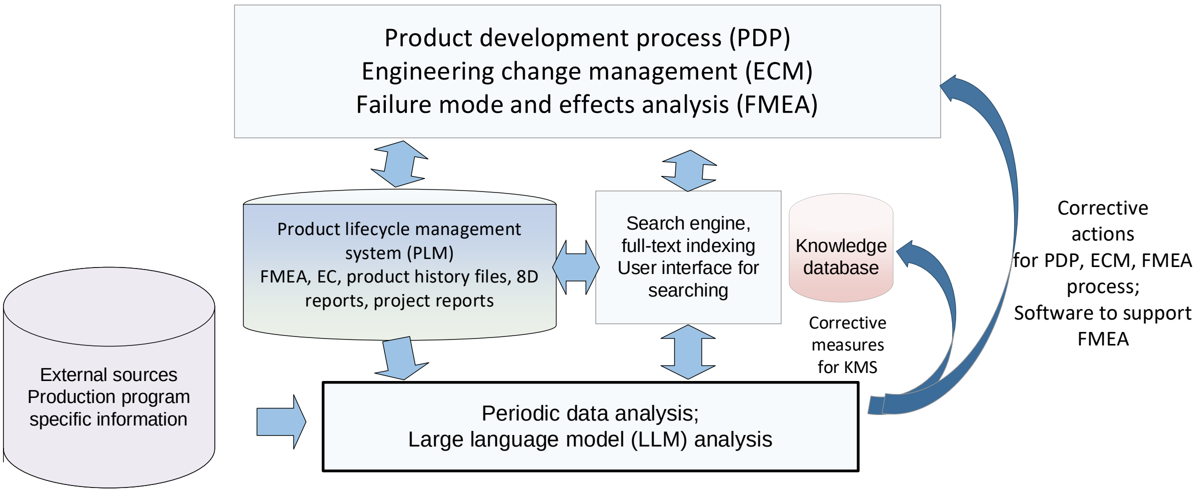


Figure 1. Information system model for application of LLM for data collection, extraction, knowledge management and use in FMEA

The use of the process model and the information system model of the framework must be consistent with the phase of the product life cycle. The maturity level of the product influences the type of optimal support in the FMEA process: external knowledge sources would be prioritized for new products with which a company does not yet have much experience, while internally generated knowledge sources would dominate for more mature products with several years of experience in development, manufacturing and sales.

5. Case study

The applicability of the framework is illustrated in a first step by a case study using publicly available data from the automotive industry, more precisely data from reviews of vehicle by private persons. This type of data is much less structured than company FMEA reports and other product-specific documents and therefore presents a greater challenge in this respect. This also allows the first steps of the framework process model to be implemented and tested without being limited by the security issues that would arise when using proprietary company data. Even though this approach has some limitations (the company context is not fully represented, e.g. no corrective action can be extracted, and the case is limited to design FMEA), it still allows us to evaluate different approaches such as automatic data pre-processing, model training and information extraction.
More specifically, we wanted to test three aspects where LLMs can be useful for FMEA:

- how accurately an LLM can extract information from such data with training,
- how accurately an LLM can extract information from such data without training,
- how relevant an LLM's suggestions can be for FMEA.

We tested these different aspects using the five-step framework process model. The first aspect was tested for data pre-processing, the second aspect with failure mode extraction and the third aspect mainly with effects. The applied process is described in detail below and is illustrated in Figure 2.
We opted for an approach based on the fine-tuning of transformer models using the LLM GPT-3 and later GPT-3.5 Turbo. This approach offers the advantage of high initial performance with limited data, but comes at a higher cost as training with OpenAI's API comes with a token fee. Figure 2 shows the

https://doi.org/10.1017/pds.2024.204 Published online by Cambridge University Press

interface of the platform developed for the case study (right side). The platform was developed in Google's Colab (https://colab.research.google.com/).

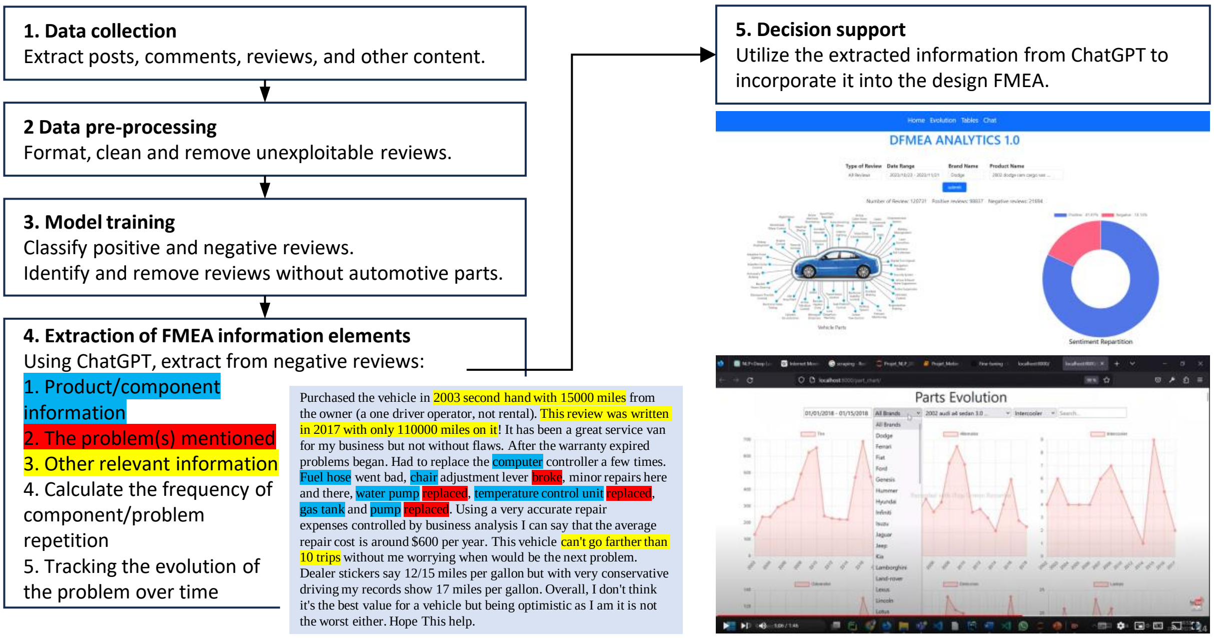


Figure 2. The process of our automatic FMEA platform

1. Data collection

We used a dataset of car reviews from the Kaggle platform (www.kaggle.com), which itself included fifty different datasets of customer reviews for fifty vehicle brands (AnkurJain, 2019) for a total of about 227,000 reviews.

2 Data pre-processing

The data was cleaned, formatted, and merged in a single dataset. The FMEA being performed on vehicle parts, the reviews without parts needed to be removed. For that, we created a list of parts by performing web scraping on websites such as Wikipedia ("List of auto parts", 2023) and List Explained (Dan, 2022). Using string comparison, the reviews without identified part were removed, resulting in a dataset of ca. 100,000 reviews.

3. model training

The model training was carried out in two steps.

First, the reviews needed to be divided into negative and positive ratings. The negative comments often contain valuable information about potential problems and concerns related to automotive components and systems. An existing dataset in which the reviews were already labelled as positive and negative (Maas et al., 2011) was used for training. Several deep learning algorithms were tested. With Tensorflow's CNN (Abadi et al., 2016), which uses bidirectional short-term memory (BiLSTM), the final accuracy was 87%. When fine-tuning with GPT-3 (Curie model), the final accuracy was 97% (Figure 3, top left). The model obtained by GPT-Curie could then be used to sort the negative and positive car reviews.

Secondly, string comparison for identifying reviews without parts present some limitations (spelling errors, non-exhaustivity, use of different languages, etc.). Thus, a model was developed to extract reviews with parts using LLM instead of string comparison. The training of the model was performed with GPT-3.5 Turbo (Figure 3, bottom left). 18,000 reviews were used for training and the accuracy achieved was 98-99%, including French and Spanish reviews.

4. Extraction of FMEA information elements

100 reviews were randomly selected using the sentiment analysis model and the part finder model: 20 reviews with the part "door", "tire", "seats", "wheel" and "window", respectively.

The LLM prompt asked to extract the failure modes as accurately as possible, while the system was allowed to make suggestions for the other FMEA information elements. The prompt used for the

https://doi.org/10.1017/pds.2024.204 Published online by Cambridge University Press

extraction of the FMEA information elements from these reviews was only developed in 2-3 iterations. Similarly, to coding methods used in qualitative research (e.g. Campbell et al., 2013), the failure modes were first manually analysed and evaluated independently by two researchers, and were subsequently compared one-to-one and agreed upon after discussion, to ensure accuracy of the analysis and avoid bias. As the data were reviews and not technical reports, the failure modes were viewed from the user's perspective, which does not always correspond to a 'technical' failure mode. The failure modes were rated on a scale of 0 to 2, where 0 means that the answer was incorrect and 2 means that the answer was deemed completely correct. Accuracy was calculated as the average of all scores. The system showed an accuracy of 84% (0: 11%, 1: 11%, 2: 78%). The two main reasons for a score of 0 were that the system could not extract a failure mode (four cases) or suggested a failure mode that was not actually included in the review (three cases). The main reason for a score of 1 was that the system summarized the failure mode (example: "the two rear windows could not be opened" is reported as "window failure"). In cases where the review did not mention the part in question (e.g., "steering wheel" instead of "wheel") or did not report any associated failure mode, the system did not report a failure mode with 100% accuracy. Six reviews were excluded from the assessment: In five cases, it was not possible to decide whether ChatGPT or the prompt was the cause of the problem. In the sixth case, the evaluators could not understand the review itself.

Effects were assessed on a 0-2 relevance scale. The effects were rated as good in 77% of cases (score of 2), acceptable in 17% of the cases (score of 1) and incorrect in 6% of cases (score of 0), for a total of 85.5%. Five reviews were discarded, for similar reasons as above. The following example illustrates a score of 2: From the failure mode "Water leakage at the doors", the system inferred the effects "Inconvenience, potential damage to interior, and dissatisfaction with the workmanship and quality of parts". "Dissatisfaction with the workmanship and quality of parts" was written in the review, only in a different wording. "Potential damage to interior" was not mentioned at all, although this is a very reasonable effect.

Finally, we have asked the system to generate, for a given failure mode, potential causes, current controls, severity, occurrence and detection (examples can be found in Figure 3, right). Several suggested values were deemed consistent with the reviews, but no quantitative analysis was performed as there was no specific data to justify the proposed numbers.

5. Decision support

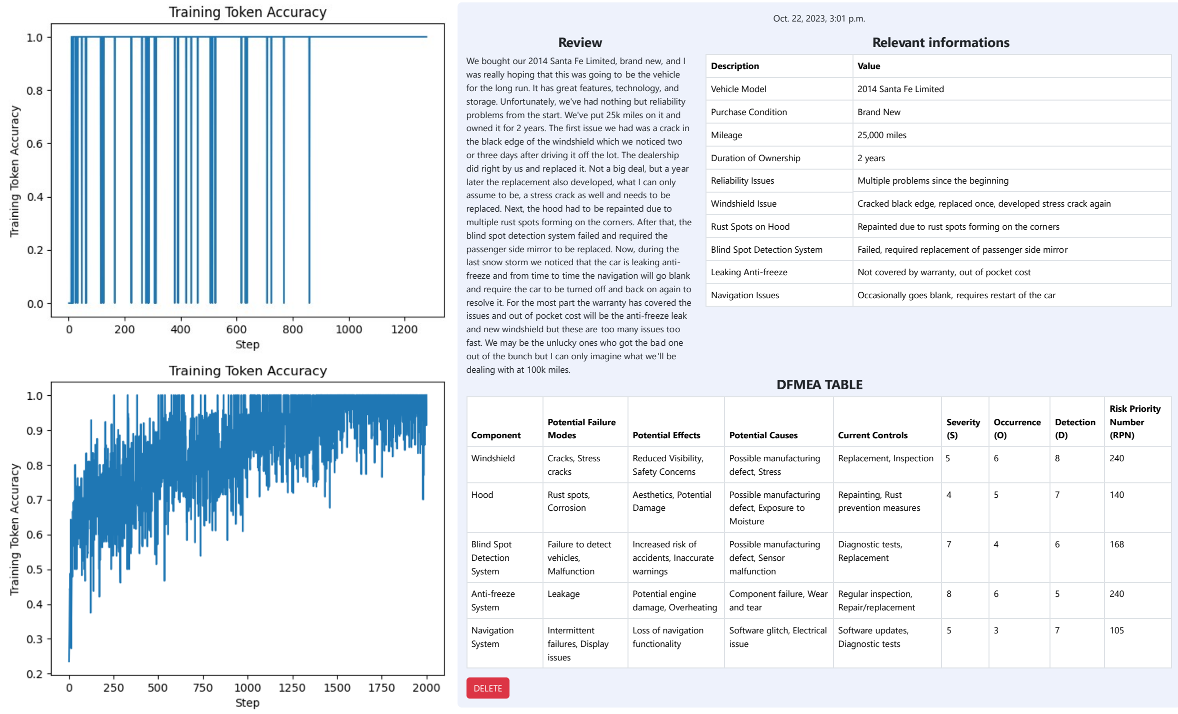
The elements of data resulting from the sentiment analysis model and the part finder model were organized in an accessible way for the subsequent FMEA analysis (see the images of the interface in Figure 2, right side). The extraction of FMEA information elements was done through a prompt that generated a table of relevant information and a DFMEA table for each review. The information extracted through this process is shown in Figure 3 (right). This is an example of what the LLM can do as part of a KMS.

6. Concluding discussion and future work

Two gaps were identified in the literature study, the lack of a comprehensive framework, and the lack of studies of the practical challenges of implementing LLM-based systems in the FMEA process. This paper has presented a framework with a process model and an information system model based on the advantages and contributions that LLMs can provide for FMEA activities and has presented a case study that illustrated different aspects of the framework. The case study itself was used to test some practical challenges taking advantage of the potential of LLMs .

The case study illustrates well the potential benefits of using LLMs along the FMEA process. The two training activities for classifying positive and negative reviews and identifying parts delivered excellent results, better than the other deep learning algorithms tested. The evaluation of the failure modes generated by the LLM without training yielded very satisfactory results and shows the value of LLMs in data extraction and analysis support. The generated effects and the resulting data are promising for the use of LLMs to support risk assessment and decision making. Finally, the generated FMEA tables show the potential of LLMs in organizing, documenting, and visualizing the generated information. In this sense, the case study illustrates the benefits and contributions of LLMs in FMEA presented in Section 3.1.

https://doi.org/10.1017/pds.2024.204 Published online by Cambridge University Press



Oct. 22, 2023, 3:01 p.m.

Review

We bought our 2014 Santa Fe Limited, brand new, and I was really hoping that this was going to be the vehicle for the long run. It has great features, technology, and storage. Unfortunately, we've had nothing but reliability problems from the start. We've put 25k miles on it and owned it for 2 years. The first issue we had was a crack in the black edge of the windshield which we noticed two or three days after driving it off the lot. The dealership did right by us and replaced it. Not a big deal, but a year later the replacement also developed, what I can only assume to be, a stress crack as well and needs to be replaced. Next, the hood had to be repainted due to multiple rust spots forming on the corners. After that, the blind spot detection system failed and required the passenger side mirror to be replaced. Now, during the last snow storm we noticed that the car is leaking anti-freeze and from time to time the navigation will go blank and require the car to be turned off and back on again to resolve it. For the most part the warranty has covered the issues and out of pocket cost will be the anti-freeze leak and new windshield but these are too many issues too fast. We may be the unlucky ones who got the bad one out of the bunch but I can only imagine what we'll be dealing with at 100k miles.

Relevant informations

Description	Value
Vehicle Model	2014 Santa Fe Limited
Purchase Condition	Brand New
Mileage	25,000 miles
Duration of Ownership	2 years
Reliability Issues	Multiple problems since the beginning
Windshield Issue	Cracked black edge, replaced once, developed stress crack again
Rust Spots on Hood	Repainted due to rust spots forming on the corners
Blind Spot Detection System	Failed, required replacement of passenger side mirror
Leaking Anti-freeze	Not covered by warranty, out of pocket cost
Navigation Issues	Occasionally goes blank, requires restart of the car

DFMEA TABLE

Component	Potential Failure Modes	Potential Effects	Potential Causes	Current Controls	Severity (S)	Occurrence (O)	Detection (D)	Risk Priority Number (RPN)
Windshield	Cracks, Stress cracks	Reduced Visibility, Safety Concerns	Possible manufacturing defect, Stress	Replacement, Inspection	5	6	8	240
Hood	Rust spots, Corrosion	Aesthetics, Potential Damage	Possible manufacturing defect, Exposure to Moisture	Repainting, Rust prevention measures	4	5	7	140
Blind Spot Detection System	Failure to detect vehicles, Malfunction	Increased risk of accidents, Inaccurate warnings	Possible manufacturing defect, Sensor malfunction	Diagnostic tests, Replacement	7	4	6	168
Anti-freeze System	Leakage	Potential engine damage, Overheating	Component failure, Wear and tear	Regular inspection, Repair/replacement	8	6	5	240
Navigation System	Intermittent failures, Display issues	Loss of navigation functionality	Software glitch, Electrical issue	Software updates, Diagnostic tests	5	3	7	105

DELETE

Figure 3. Top left: Training accuracy evolution for the sentiment analysis model; Bottom left: Training accuracy evolution for the part finder model; Right: Illustration of automatically generated FMEA table using GPT-3.5 Turbo

The case also provides other insights into the use of LLMs in FMEA. The amount of programming and prompting required was relatively small (2-3 iterations), and only the default settings of the fine-tuning parameters were used. It is also important to reiterate that the end-user reviews, with their non-technical language, added a layer of complexity the data analysis. When working with company documents (FMEA reports, etc.), even more precise information extraction can be expected. If the data extraction of FMEA elements is complemented by training, additional improvement is assured.

The study shows that LLMs can enable faster pre-processing and identification of FMEA elements, significantly reduce manual effort and can potentially save a considerable amount of time in the analysis phase. This study also provides insights into the future ability of LLMs and generative AI in general to support activities in engineering design that could not otherwise be automated.

This study also highlights some limitations and weaknesses of LLM results, such as the tendency to generalize or extrapolate). In order to implement the FMEA framework in an industrial context, some aspects still need to be considered as presented in the next section.

6.1. Future work

Regarding the current results, it has already been noted that the extracted FMEA elements are sometimes presented as a generalized description or extrapolated. Very often this is well done. However, it is an open question to what extent we want such generalization. It is necessary to investigate in more detail how well this can be controlled during training and prompting.

The present study needs to be extended to all FMEA elements, such as selecting corrective actions and evaluating the effectiveness of risk mitigation strategies. The data management of the system can also be made more dynamic (linking the documents from which the data is extracted and dynamically updating the results as new information becomes available) and structured as in Figure 1. New functionalities in LLMs will enable this type of task.

The case study dealt with relatively generic types of car parts. It is reasonable to assume that the LLM has extensive knowledge of automotive parts, which may have helped it analyse the reviews. There is a need to extend this type of analysis to lesser known products and components.

https://doi.org/10.1017/pds.2024.204 Published online by Cambridge University Press

The problem of confidentiality and privacy is crucial for companies. To solve this problem, it will be necessary to set up a local LLM infrastructure (train specific data and develop custom models in isolation) to facilitate control over proprietary knowledge and data security. This problem is also related to the various LLMs available, most of which are proprietary and tend to be less transparent in their handling of data. Some open access LLMs, such as OpenLLaMA (Geng and Liu, 2023), have already emerged and may help ensure privacy.

Finally, it will be important to evaluate the impact of LLMs on the performance of FMEAs in an industrial context. In a recent study, Dell'Acqua and colleagues (2023) showed that consulting tasks within the scope of LLMs capabilities can be performed 25% faster and with an increase in quality of more than 40% while the use of LLMs for activities beyond their capabilities reduces the probability of obtaining correct results by 19%. Similarly, it will be necessary to evaluate the use of LLMs with realistic FMEA tasks to determine which tasks benefit from LLMs and which are better performed without such support.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A. et al. (2016). "TensorFlow: A system for large-scale machine learning", *arXiv*, 1605.08695. https://doi.org/10.48550/arXiv.1605.08695

AIAG and VDA (2019), Failure Mode and Effects Analysis—FMEA Handbook: Design FMEA, Process FMEA, Supplement FMEA for Monitoring and System Response, Automotive Industry Action Group and Verband der Automobilindustrie, Southfield, MI.

AnkurJain (2019), Edmunds-Consumer Car Ratings and Reviews (Version 3), Kaggle. Available at: https://www.kaggle.com/datasets/ankkur13/edmundsconsumer-car-ratings-and-reviews (accessed 4 February 2024).

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S. et al. (2021), "On the opportunities and risks of foundation models", *arXiv*, 2108.07258. https://doi.org/10.48550/arXiv.2108.07258

Campbell, J. L., Quincy, C., Osserman, J. and Pedersen, O. K. (2013), "Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement", *Sociological Methods & Research*, Vol. 42 No. 3, pp. 294-320. https://doi.org/10.1177/0049124113500475

Dell'Acqua, F., McFowland, E. III, Mollick, E., Lifshitz-Assaf, H., Kellogg, K.C. et al. (2023), "Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality", *HBS Working Paper Series*, Working Paper No. 24-013, Harvard Business School. Available at: https://www.hbs.edu/faculty/Pages/item.aspx?num=64700 (accessed 19 November 2023).

Department of Defense (1980), MIL-STD-1629A, Procedures for performing a failure mode, effects and criticality analysis, Department of Defense, Washington, D.C.

Diemert, S., and Weber J.H., (2023), "Can large language models assist in hazard analysis?", *arXiv*, 2303.15473. https://doi.org/10.48550/arXiv.2303.15473

Filz, M.-A., Langner, J.E.B., Herrmann C. and Thiede. S. (2021), "Data-driven failure mode and effect analysis (FMEA) to enhance maintenance planning", *Computers in Industry*, Vol. 129, 103451. https://doi.org/10.1016/j.compind.2021.103451

Geng, X. and Liu, H. (2023). "OpenLLaMA: An open reproduction of LlaMA". Available at: https://github.com/openlm-research/open_llama (accessed 11 November 2023).

Hassan, F., Nguyen, T., Le, T. and Le, C. (2023), "Automated prioritization of construction project requirements using machine learning and fuzzy failure mode and effects analysis (FMEA)", *Automation in Construction*, Vol. 154, 105013. https://doi.org/10.1016/j.autcon.2023.105013

Hodkiewicz, M., Klüwer, J.W., Woods, C., Smoker, T. and Low E. (2021), "An ontology for reasoning over engineering textual data stored in FMEA spreadsheet tables", *Computers in Industry*, Vol. 131, 103496. https://doi.org/10.1016/j.compind.2021.103496

Hu, X., Tian, Y., Nagato, K., Nakao, M. and Liu, A. (2023), "Opportunities and challenges of ChatGPT for design knowledge management", *Procedia CIRP*, Vol. 119, pp. 21-28, https://doi.org/10.1016/j.procir.2023.05.001

Huang, J., Xu, D.-H., Liu, H.-C. and Song, M.-S. (2019), "A new model for failure mode and effect analysis integrating linguistic Z-numbers and projection method", *IEEE Transactions on Fuzzy Systems*, Vol. 29, No. 3, pp. 530–538, http://doi.org/10.1109/TFUZZ.2019.2955916

IATF (2016), IATF 16949:2016, Automotive quality management system standard, International Automotive Task Force.

IEC (1985), IEC 60812, Analysis techniques for system reliability – Procedure for failure mode and effects analysis (FMEA), International Electrotechnical Commission, Geneva.

IEC (2018), IEC 60812, Failure modes and effects analysis (FMEA and FMECA), International Electrotechnical Commission, Geneva.

Kan, D. (2022, November 7), “Car parts that start with A to Z – A thorough exploration”, *List Explained*. Available at: https://listexplained.com/car-parts-that-start-with-a-to-z/ (accessed 3 June 2023).

“List of auto parts” (2023, March 3), *Wikipedia*. Available at: https://en.wikipedia.org/wiki/List_of_auto_parts (accessed March 2023).

Liu, H.-C., Chen X.-Q., Duan, C.-H. and Wang, Y.-M. (2019), “Failure mode and effect analysis using multi-criteria decision-making methods: A systematic literature review”, *Computers & Industrial Engineering*, Vol. 135, pp. 881-897. https://doi.org/10.1016/j.cie.2019.06.055

Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y. and Potts, C. (2011), “Learning word vectors for sentiment analysis”, in: Lin, D., Matsumoto, Y. and Mihalcea, R. (Eds.), *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Association for Computational Linguistics, Stroudsburg, PA, pp. 142-150. Available at: https://aclanthology.org/P11-1015. Dataset derived from original available at: https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews (accessed 4 February 2024).

Na’amnh, S., Salim, M.B., Husti, I. and Daróczi, M. (2021). “Using artificial neural network and fuzzy inference system based prediction to improve failure mode and effects analysis: A case study of the busbars production”, *Processes*, Vol. 9, No. 8, 1444. https://doi.org/10.3390/pr9081444

Peddi, S., Lanka, K. and Gopal, P.R.C. (2023) “Modified FMEA using machine learning for food supply chain”, *Materials Today: Proceedings*, in Press. https://doi.org/10.1016/j.matpr.2023.04.353

SAE (1994), SAE J1739, Potential failure mode and effects analysis in design (design FMEA) and potential failure mode and effects analysis in manufacturing and assembly processes (process FMEA) reference manual, Society of Automotive Engineers, Warrendale, PA.

Soltanali, H. and Ramezani, S. (2023), “Smart failure mode and effects analysis (FMEA) for safety–Critical systems in the context of Industry 4.0”, in: Garg, H. (Ed.) *Advances in Reliability, Failure and Risk Analysis*, Springer, Singapore, pp. 151-176. https://doi.org/10.1007/978-981-19-9909-3_7

Spreafico, C. and Sutrisno, A. (2023). “Artificial intelligence assisted social failure mode and effect analysis (FMEA) for sustainable product design”. *Sustainability*, Vol. 15, No. 11, 8678. https://doi.org/10.3390/su15118678

Tavčar, J., Benedičič, J. and Žavbi, R. (2019), “Knowledge management support in the engineering change process in small and medium-sized companies”, *International Journal of Agile Systems and Management*, Vol. 12, No. 4, pp. 354-381. https://doi.org/10.1504/IJASM.2019.104587

Tavčar, J. and Duhovnik, J. (2014), “Tools and methods stimulate virtual team co-operation at concurrent engineering”, *Proceedings of the 21st ISPE Inc. International Conference on Concurrent Engineering, Beijing, China, September 8–11, 2014*, IOS Press, Amsterdam, pp. 457-466. https://doi.org/10.3233/978-1-61499-440-4-457

Thirunavukarasu, A. J., Ting, D. S. J., Elangovan, K., Gutierrez, L., Tan, T. F. and Ting, D. S. W. (2023), "Large language models in medicine", *Nature Medicine*, Vol. 29 No. 8, pp. 1930-1940. https://doi.org/10.1038/s41591-023-02448-8

Thomas, D. (2023), “Revolutionizing failure modes and effects analysis with ChatGPT: Unleashing the power of AI language models”, *Journal of Failure Analysis and Prevention*, Vol. 23, No. 3, pp. 911–913. https://doi.org/10.1007/s11668-023-01659-y

Wirth, R., Berthold, B., Krämer, A. and Peter, G. (1996), “Knowledge-based support of system analysis for the analysis of failure modes and effects”, *Engineering Applications of Artificial Intelligence*, Vol. 9, No. 3, pp. 219-229. https://doi.org/10.1016/0952-1976(96)00014-0

Wu, Z., Liu, W. and Nie, W. (2021), “Literature review and prospect of the development and application of FMEA in manufacturing industry”, *The International Journal of Advanced Manufacturing Technology*, Vol. 112, No. 5-6, pp. 1409–1436. https://doi.org/10.1007/s00170-020-06425-0

Yucesan, M., Gul, M. and Celik, E. (2021), “A holistic FMEA approach by fuzzy-based Bayesian network and best-worst method”, *Complex & Intelligent Systems*, Vol. 7, No. 3, pp. 1547–1564. https://doi.org/10.1007/s40747-021-00279-z

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X. et al. (2023). “A survey of large language models”, *arXiv*, 2303.18223, https://doi.org/10.48550/arXiv.2303.18223