Proceedings of the Design Society, Volume 5: ICED25
https://doi.org/10.1017/pds.2025.10192

Extracting FMEA information from publicly available datasets using large language models

Rahul Sharan Renu,✉

Austin College, USA

✉ rrenu@austincollege.edu

ABSTRACT: The objective of this research is to explore the use of publicly available recall data from the National Highway Transportation Safety Administration to extract Failure Modes and Effects Analysis data. This large data set was analysed using a Large Language Model chatbot. To assess the usefulness of priming the chatbot with this data, the chatbot was also asked to generate data without priming it with the recall data. This was performed on two specific products. It was found that primed-chatbot results were more specific and used technical terminology appropriate to the product being analysed. The proposed approach can be used by designers in the forward design process during new product development. The proposed approach provides designers with insight into potential failures, the associated consequences, their severity, and root causes as well.

KEYWORDS: artificial intelligence, design engineering, design practice

1. Motivation

Engineering design is an inherently iterative process where designers are constantly battling the need to ensure high quality designs are realized in as short a time as possible. The number and length of iterations can be reduced if designers are able to predict failures in the early stages of the design process. In today's world of big data and artificial intelligence, determining trends from historical accounts has become more feasible. A company can learn from its previous mistakes in many ways. The most popular mechanism for this to occur is through institutional (sometimes implicit) knowledge of long-tenured employees. Another mechanism for companies to keep sight of previous errors is by analyzing their previous product recalls (that is, defects related to performance and/or safety non-compliance resulting in manufacturers replacing/reworking defective components). Specific to automotive design, the National Highway Traffic Safety Administration requires "manufacturers file a defect and noncompliance report as well as quarterly recall status reports, in compliance with Federal Regulation 49 (the National Traffic and Motor Safety Act) Part 573" [https://www.nhtsa.gov/nhtsa-datasets-and-apis]. The NHTSA publishes said recall data for open access to the public.

Given the pressure on automotive companies to hasten new product development cycles while maintaining high product quality, and in lieu of the abundant product recall made publicly available by the NHTSA, there is an opportunity to learn from previous errors committed by a multitude of automotive original equipment manufacturers. A systematic method of enumerating and analyzing failures is known as Failure Modes and Effects Analysis.

2. Failure Modes and Effects Analysis

Failure Modes and Effects Analysis (FMEA) is a tool widely used by engineers (Stamatis, 2003) to:

1. Identify how a system/product/process might fail (failure mode)
2. Analyze the reason that the failure occurs (failure cause)
3. Predict the impacts of the failure (failure effects)

https://doi.org/10.1017/pds.2025.10192 Published online by Cambridge University Press

The recent case of the Takata airbag recall (Segal, 2019) can be viewed from the lens of FMEA which allows for a better understanding of the usefulness of FMEA as a tool. In this case, the failure was that the airbag inflator (a metal device that produces gas causing the airbag to inflate) exploded during a car crash. This makes "explosion" the failure mode. This was found to be caused by use of an inappropriate chemical that resulted in an explosive chemical reaction when exposed to heat for certain periods of time (the failure cause). The failure effect is that metal shards flew towards the passenger during airbag deployment which, in some cases, resulted in loss of life. Unfortunately, this failure was not predicted during design and only observed after the production and sale of millions of cars. As a result of this failure, 67 million airbags have been recalled and 26 deaths have been reported (ConsumerReports.org, 2024).

There are various types of FMEAs. Those performed on products are referred to as Design FMEAs. In this case individual components or specific subsystems are the subject of the analysis. An analysis of the failure modes of movement of a product through its supply chain (including its production process) is referred to as Process FMEA. A third category of FMEAs are those performed with the subject of analysis is the product and its interactions with all internal and external entities – System FMEA. The focus of the research presented in this paper is Design FMEAs.

FMEA provides a systematic method for engineers to determine weaknesses in a design (see Table 1 for an example). This tool has the power to reduce product design costs and time-to-market, thus increasing a company's competitive edge(Stamatis, 2003). More importantly, when FMEA is used during the design of new products, it can alert designers to safety issues and provide the opportunity to prevent critical accidents.

Table 1. Example failure modes and effects analysis

Potential Failure Mode	Effect of Failure	Severity	Potential Cause	Occurrence	Current Controls	Detection	RPN
Burned-out Bulbs	Reduced visibility to drivers	10	Bulb end of life	5	Regular maintenance checks	8	400
Faulty Wiring	Intermittent light function	10	Poor installation	2	Visual inspection	4	80

2.1. Computing RPN

FMEAs allow for objective decision-making by requiring the quantification of the frequency of failure mode occurrence (O), severity of the failure effects (S), and the detectability of the failure mode before it occurs (D). These values (O, S, and D) are typically assigned based on numeric scales (see Tables 2, 3 and 4 for an example) developed by the team performing the FMEA. Once assigned, a Risk Priority Number (RPN) for each failure mode-failure effect instance is computed by multiplying O, S, and D values.

Table 2. Example rating scale for severity (adapted from (Siemens, 2019))

Rating	Description
1	No effect, no danger
2	Very minor – usually noticed only by discriminating or very observant users
3	Minor – only minor part of the system affected; noticed by average users
4-6	Moderate – most users are inconvenienced and/or annoyed
7-8	High – loss of primary function; users are dissatisfied
9-10	Very high – hazardous. Product becomes inoperative, customers angered. Failure constitutes a safety hazard and can cause injury or death.

https://doi.org/10.1017/pds.2025.10192 Published online by Cambridge University Press

Table 3. Example rating scale for occurrence (adapted from (Siemens, 2019))

Rating	Description
1	No documented failures on similar products/processes
2-3	Low – relatively few failures
4-6	Moderate – some occasional failures
7-8	High – repeated failures
9-10	Very high – failure is almost certain

Table 4. Example rating scale for detectability (adapted from (Siemens, 2019))

Rating	Description
1	Fault is certain to be caught by testing
2	Fault almost certain to be caught by testing
3	High probability that tests will catch fault
4-6	Moderate probability that tests will catch fault
7-8	Low probability that tests will catch fault
9-10	Fault will be passed undetected to user/customer

Upon a thorough enumeration of all failure modes and effects, the list can be sorted in multiple ways. This includes a simple descending-order sort based on RPN to focus on those failure modes and effects that have the highest RPNs first. A critical disadvantage of this method is that high criticality failure effects may be pushed down the list if they have lower occurrence and detectability scores compared to failure effects that may not be critical at all but occur frequently or may not be easily detected.

2.2. Challenges with FMEA

While FMEAs result in quantified assessments of failure modes and their effects, the process of performing an FMEA can be subjective. The scales used to categorize Occurrence, Severity, and Detectability can lead to disagreement amongst raters unless the scales' definitions are quantified thoroughly.

Another challenge associated with FMEAs relates to the efforts and person-hours expended in performing them. A team of members that represent all stakeholder departments must first be assembled. Objectives for performing the FMEA must be communicated and well-understood. Scales developed for the FMEA must be discussed to minimize any misinterpretations. And most importantly, team members will be kept away from performing other value-added work during the time that they are involved with the FMEA task. These challenges are inherent to FMEAs and have been the motivation for other research efforts. These will be reviewed and critiqued next. The next section provides an overview of data made available from NHTSA. This is followed by the process employed to extract FMEA information from NHTSA recall data and is followed by sample FMEAs performed with NHTSA information supplementing the process. The last section discusses conclusions and future work.

3. Prior work

Wu and colleagues (Wu et al., 2021) presented a thorough literature of FMEA in the context of the manufacturing industry. A key finding is that computational approaches must be taken to provide decision support for complex tasks and tedious work as they relate to FMEAs. They also recognize that much of the prior work on FMEAs relates to the quantification of RPNs and focusses highly on theoretical concepts as opposed to industrial needs/standards. They recommend that automated recognition of FMEA components (such as root cause, failure mode identification, and so on) should be the focus of future research to improve application efficiency of FMEAs.

Rehman and Kifor (Rehman & Kifor, 2016) evaluated an ontological approach to manage semantic knowledge embedded in Process FMEAs. They recognize that a critical challenge with successfully conducting FMEAs over many years is its reliance on institutional knowledge that is predominantly implicitly known by experienced employees. The method proposed to mitigate this challenge involves the establishment of an ontological representation of all entities of an FMEA which inherently requires the establishment of

https://doi.org/10.1017/pds.2025.10192 Published online by Cambridge University Press

ontological rules to govern the relationship between entities. This requirement poses a challenge since it requires implicit knowledge to be made explicit which requires time commitment and willingness to participate. The proposed method does not address similar challenges in Product FMEAs (focus of the research presented in this paper.)

Chin and colleagues (Chin et al., 2008) have developed a fuzzy system to evaluate design concept alternatives based on FMEA information. Their system, scoped to evaluate only simple products, takes fuzzy linguistic inputs of O, S, and D, and returns an RPN based on fuzzy logic. This analysis is performed on a prepopulated information set related to known components and subsystems. While their research does address aspects of challenges associated with Product FMEAs, it is not agile nor easily scalable.

Price and colleagues developed the FLAME system that automates the FMEA process related to electrical systems. This approach uses ontological descriptions and reasoning to retrieve RPN values for new designs based on patterns recognized in historical data. While this approach does mitigate the issues related with FMEAs, it is an application-specific approach whose scalability requires high effort.

Renu and colleagues (Renu et al., 2016) developed a knowledge-based FMEA approach that uses a knowledge base of previously conducted FMEAs. They use decision trees to develop inference capabilities and use Product Data Management information to predict failure modes and RPNs for subsystems/components. Much like other previous research efforts, their work depends on manual constructed database of historical FMEAs – an effort-intensive task.

Spreafico and Sutrisno (Spreafico & Sutrisno, 2023) investigated the use of ChatBots (LLMs) to aid in the performance of FMEAs from a social sustainability perspective. They recognize that to have social aspects thoroughly considered in the design and development of a new product, a multidisciplinary team of engineers, sociologists, and psychologists will be needed. This is challenging since some companies do not have such a broad employee skillset, and because this will increase product-to-market times. The authors explore the feasibility of using an LLM-based chatbot to serve as a surrogate for the multidisciplinary team members. They evaluated the effectiveness of using a predefined lexicon in questions to solicit FMEA-related feedback from the chatbot. Their research showed limited success and indicated the need to provide the chatbot with more context than a technical lexicon around which prompts are framed.

While these research efforts have begun to address the various challenges associated with FMEAs, some gaps remain. Ontological approaches require time-intensive work to develop ontological representations and rules of the systems under consideration. Fuzzy approaches and other knowledge-based approaches require large amounts of historical data. These approaches have limited scalability and extensibility. The goal of the proposed research is to use LLMs to generate FMEA information while grounding the generative AI by using NHTSA recall data.

4. National Highway Transportation Safety Administration data

The dataset made available by the National Highway Traffic Safety Administration contains approximately 247,000 reports (as of December 2024) with recalls reported from 1949 to the present day. The number presented is approximate only because the dataset is updated periodically.

The dataset covers recalls related to vehicles, tires, child safety seats, and other automotive equipment [https://www.nhtsa.gov/nhtsa-datasets-and-apis#recalls]. The NHTSA offers an interactive web-based dashboard to explore the data and APIs to interact with the data through custom code. The raw data is also made available for bulk download.

5. Extracting FMEA information from NHTSA dataset

The proposed process of extracting information from the NHTSA dataset for use in FMEA is shown in Figure 1 (Left). The process is enumerated below.

1. NHTSA recall data is downloaded from the website.
2. Recall data is filtered to isolate the component/subsystem of interest.
3. The filtered recall data is fed to a Large Language Model (Microsoft Copilot, in the case of this research).
4. The LLM is asked to extract failure mode, failure consequence severity, and failure root cause for each recall for top ten defects. The prompt used is *"Extract the top ten defects (combine semantically similar defects), paraphrase their consequence, rate the severity (low, medium, high), provide a failure mode, and provide a root cause. Format this as table."*.
5. The extracted data is fed back to the designer for a critical review. This is a crucial step to ensure that AI hallucinations and inaccuracies are not present.

https://doi.org/10.1017/pds.2025.10192 Published online by Cambridge University Press

The structure and content of the prompts have an influence on the results generated by the LLM. It was critical to specify that the top ten defects needed to be extracted. In the absence of this detail, the LLM was inconsistent with the number of defects it extracted. The number (ten, in this case) can also be changed as needed. It was also critical to ask the LLM to combine semantically similar defects otherwise it could provide results that are redundant. Paraphrasing the consequence was found to be useful from the perspective of readability of the results generated. Without explicitly asking for failure modes and root causes, the LLM will not generate this critical information.

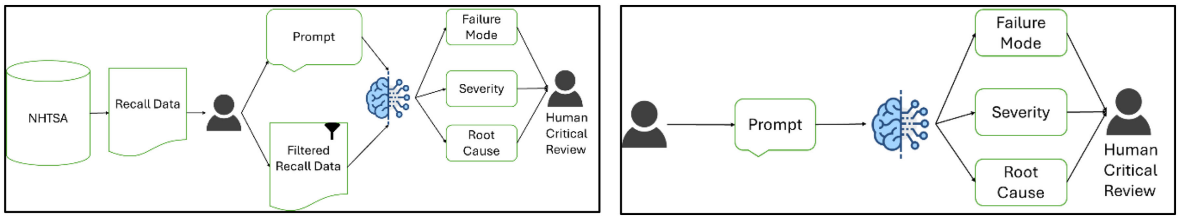


Figure 1. (Left) Process overview for generating FMEA information while priming with NHTSA data. (Right) Process overview for generating FMEA information without priming with NHTSA data

6. FMEA information generated from proposed process

To demonstrate and validate the proposed process, two products have been considered – child car seats, and brake lights. The process outlined in the previous section was followed to analyse the recall data associated with these two products. The perspective assumed was that of a new product designer trying to pre-empt common failure modes, consequences, and root causes. This data was downloaded from NHTSA on December 01, 2024.

6.1. Child car seats

When NTHSA data was filtered on column 'COMPONENT NAME' for any component that had 'child seat' in its name, 1097 rows of data were found. This data had 262 unique pairs of 'DESC_DEFECT' and 'CONSEQUENCE_DEFECT'.

These unique pairs were uploaded to Copilot, and it was asked to extract the top ten defects, and its failure modes, severity ratings, and root causes using the prompt presented in the previous section. The response from Copilot is shown in Table 5.

Table 5. LLM chatbot results for child seats when primed with NHTSA data

Defect	Consequence	Severity	Failure Mode	Root Cause
Tether Webbing Abrasion	Degraded webbing may not protect the child in a crash.	High	Abrasion of webbing	Use of substandard materials
Handle Separation	Handle may detach, causing the seat to flip forward.	High	Handle detachment	Improper assembly or loose screws
Incorrect Labelling	Misleading information can lead to improper use of the seat.	Medium	Incorrect or missing labels	Manufacturing error
Structural Cracks	Cracks in the plastic shell can lead to seat failure in a crash.	High	Shell cracking	Insufficient material strength
Harness Slippage	Harness may loosen, failing to restrain the child properly.	High	Harness slippage	Faulty adjuster mechanism
Excessive Buckle Release Force	Difficulty in releasing the buckle can delay child removal.	Medium	High release force	Design flaw
Flammability Issues	Increased risk of injury in the event of a fire.	Medium	Material flammability	Non-compliance with flammability standards
Buckle Engagement Failure	Buckle may not latch properly, risking ejection in a crash.	High	Buckle failure	Manufacturing defect
Improper Assembly	Various parts may be missing or incorrectly positioned.	Medium	Assembly errors	Quality control issues

(*Continued*)

https://doi.org/10.1017/pds.2025.10192 Published online by Cambridge University Press

Table 5. Continued.

Defect	Consequence	Severity	Failure Mode	Root Cause
Insufficient Padding	Inadequate padding can lead to head injuries in a crash.	High	Lack of sufficient padding	Design oversight

In addition, Copilot was asked to ignore the uploaded file and then generate ten common defects for child car seats and the associated failure modes, severity ratings, and root causes. The prompt used was *"Ignore any data that I provided. Generate top ten defects for brake lights and the associated defect consequences, failure modes, root causes and severity rankings (low, medium, high). Format in a table."* This is shown in Table 6. The process overview for this is shown in Figure 1 (Right).

Table 6. LLM chatbot results for child seats when not primed with NHTSA data

Defect	Consequences	Failure Modes	Root Causes	Severity
Broken Latches	Seat detachment in a crash	Latch failure	Poor manufacturing quality	High
Defective Buckles	Child ejection during impact	Buckle release	Design flaw or material fatigue	High
Weak Frame	Frame breakage under stress	Structural failure	Inadequate material strength	High
Flammable Materials	Increased fire risk	Ignition during a fire	Use of non-compliant materials	High
Torn Straps	Inadequate restraint	Strap tearing	Material degradation	Medium
Defective Chest Clips	Improper harness positioning	Clip breakage	Design flaw or poor material quality	Medium
Inadequate Padding	Increased injury risk	Insufficient impact absorption	Poor design or material choice	Medium
Defective Base Units	Seat not properly secured	Latch failure	Manufacturing defect	High
Poor Installation Instructions	Incorrect seat installation	User error	Inadequate instructions	Medium
Defective Carry Handles	Seat tipping and child injury	Handle release	Design flaw or material fatigue	Medium

A comparison of these two responses shows that Copilot results generated based on NHTSA data has more specifics and details than the results generated without uploading NHTSA data. For instance, results extracted from NHTSA data shows that "buckle engagement failure" and "excessive buckle release force" are two defects as opposed to "defective chest clips". Another instance of this is related to the defect consequence of structural cracks. The results from NHTSA data show that consequence is "Cracks in the plastic shell can lead to seat failure in a crash". Whereas, in the results generated with NHTSA data, "weak frame" is listed as the defect and "Frame breakage under stress" is listed as the consequence.

Another difference in the two result sets is related to terminology used. The results from NHTSA data uses terminology that is less pedestrian than the other result set, as expected. For instance, NHTSA results use the term "tether webbing", while the other results refer to it as a "strap". NHTSA results also refer to non-compliance with standards while the general results do not.

6.2. Brake lights

When NTHSA data was filtered on column 'COMPONENT NAME' for any component that had 'brake light' in its name, 2571 rows of data were found. This data had 234 unique pairs of 'DESC_DEFECT' and 'CONSEQUENCE_DEFECT'.

These unique pairs were uploaded to Copilot, and it was asked to extract the top ten defects, and its failure modes, consequences, and root causes using the prompt presented in the previous section. The response from Copilot is shown in Table 7.

https://doi.org/10.1017/pds.2025.10192 Published online by Cambridge University Press

Table 7. LLM chatbot results for brake lights when primed with NHTSA data

Defect	Consequence	Severity	Failure Mode	Root Cause
Brake lights fail to illuminate	Other drivers may not be aware the vehicle is stopping, increasing the risk of a crash.	High	Brake light switch failure	Faulty switch, electrical issues
Brake lights stay on continuously	Continuous illumination can confuse other drivers, increasing the risk of a crash.	High	Brake light switch failure	Electrical short, software error
Brake lights illuminate intermittently	Inconsistent brake light operation can mislead other drivers, increasing the risk of a crash.	Medium	Wiring issues	Incorrect wiring, loose connections
Brake lights flash instead of staying steady	Flashing brake lights can cause confusion among other drivers, increasing the risk of a crash.	Medium	Software malfunction	Software error
Brake lights do not activate with emergency brake	Lack of brake light activation during emergency braking can increase the risk of a crash.	High	Brake light switch failure	Design flaw, software error
Brake lights too bright	Excessive brightness can impair the vision of other drivers, increasing the risk of a crash.	Medium	LED light malfunction	Manufacturing defect
Brake lights fail due to water intrusion	Water damage can cause brake lights to fail, increasing the risk of a crash.	High	Corrosion of electrical components	Poor sealing, design flaw
Brake lights fail due to contamination	Contaminated switches can prevent brake lights from illuminating, increasing the risk of a crash.	High	Contaminated switch	Manufacturing process issue
Brake lights fail due to mechanical damage	Mechanical damage to the switch can prevent brake lights from working, increasing the risk of a crash.	High	Physical damage to switch	Improper installation, design flaw
Brake lights fail due to software error	Software errors can prevent brake lights from functioning correctly, increasing the risk of a crash.	High	Software malfunction	Software bug

In addition, Copilot was asked to ignore the uploaded file and then generate ten common defects for brake lights and the associated failure modes, consequences and root causes. This is shown in Table 8.

Table 8. LLM chatbot results for brake lights when not primed with NHTSA data

Defect	Consequences	Failure Modes	Root Causes	Severity
Burned-out Bulbs	Reduced visibility to other drivers	Light failure	Bulb wear and tear	High
Faulty Wiring	Intermittent or no light function	Electrical failure	Poor installation or damaged wires	High
Broken Light Housing	Water ingress and light failure	Housing crack	Physical damage or poor material	Medium
Defective Brake Light Switch	Brake lights not activating	Switch failure	Wear and tear or manufacturing defect	High
Corroded Connectors	Poor electrical connection	Connector corrosion	Exposure to moisture	Medium
Blown Fuses	Complete brake light failure	Fuse blowout	Electrical overload	High
Malfunctioning Control Module	Inconsistent light operation	Module failure	Software or hardware defect	High
Poor Ground Connection	Dim or non-functional lights	Grounding issue	Improper installation or corrosion	Medium
Incorrect Bulb Type	Inadequate brightness or fit	Bulb mismatch	Incorrect replacement	Low
Lens Discoloration	Reduced light output	Lens aging	UV exposure or material degradation	Medium

https://doi.org/10.1017/pds.2025.10192 Published online by Cambridge University Press

A comparison of these two result sets reveals a similar pattern as observed with child car seats. The defects from NHTSA data are far more detailed, whereas the result set without NHTSA data has defects that blur the lines between defect description and failure modes. Failure consequences are also more detailed when NHTSA data is provided to the LLM.
The root causes generated from NHTSA data, much like with child car seats, have fewer "or" clauses. This demonstrates the higher confidence in the model's own diagnosis of root causes when NHTSA data is provided.
It must be noted that Copilot was asked to generate top ten defects (based on frequency counts and by combining semantically similar defects) only to ensure comprehensibility of the responses generated. If the designer chooses, it is possible to have Copilot generate failure modes, severity ratings, and root causes for all rows of data filtered from the NHTSA database.

7. Comparing with human FMEA

The NHTSA dataset was analysed by a subject matter expert to provide a basis to compare LLM-generated results against. This human analysis represents a "traditional" method of analysis which is effort intensive. The analysis was performed by the author who has experience with performing FMEAs in industry settings as well as conducting research on FMEAs (Renu et al., 2016). Table 9 and 10 show results for FMEA performed by the human on the top ten (by frequency) defects for both components.

Table 9. Human analysis results for child seats

Defect (count)	Consequence	Severity	Failure Mode	Root Cause
Incorrect installation information.	Car seat not installed securely. It may come loose in the event of a crash and lead to severe injury/death.	High	Incorrect instructions	Lack of thorough documentation review
Seat buckle does not mate properly with the tongue	Child not properly secured. The child may be ejected and could lead to severe injury/death risk.	High	Improper mating of buckle and tongue	Poor geometric design/ dimensioning/ tolerances
Harness lock/release button sticks	Harness can become loose/ tight unintentionally. This can lead to severe injury/ death risk in the event of a crash	High	Harness button malfunction	Lack of sufficient lubrication at site of button
Parts missing that allow for base to mate with seat	Car seat not installed securely. It may come loose in the event of a crash and lead to severe injury/death.	High	Missing parts	Assembly defects
Tether webbing degraded	Child not restrained sufficiently during a crash. This can lead to severe injury/death risk	High	Tether webbing mechanical wear	Poor material selection
Handle not attached correctly to car seat	Handle can come loose during a crash and injure passengers	Medium	Handle detaches	Incorrect fastners specified
Handle lock releases unexpectedly	Seat and infant can flip forward leading to severe injury	High	Handle lock releases	Poor geometric design/ dimensioning/ tolerances
Shoulder strap release button can move during a crash	Excessive force may be needed to release the child from the seat	Low	Strap release button malfunction	Poor geometric design/ dimensioning/ tolerances

(*Continued*)

https://doi.org/10.1017/pds.2025.10192 Published online by Cambridge University Press

Table 9. Continued.

Defect (count)	Consequence	Severity	Failure Mode	Root Cause
Tether webbing degraded due to UV light	Child not restrained sufficiently during a crash. This can lead to severe injury/death risk	High	Tether webbing chemical wear	Poor material selection
Buckles release upon impact	Child not restrained sufficiently during a crash. This can lead to severe injury/death risk	High	Buckles release upon impact	Assembly defects

Table 10. Human analysis results for brake lights

Defect (count)	Consequence	Severity	Failure Mode	Root Cause
Rear Power Distribution Module has internal damage	Brake light does not work leading to increased risk of crash	Medium	Fuse harness malfunction	Incorrect fuse harness specified
Brake light pressure switch not working	Brake light does not illuminate when brake pedal is depressed resulting in increased risk of crash	Medium	Brake light pressure switch broken	Brake light pressure switch specification error
Brake light software malfunctioning	Brake light illuminates even when brake not depressed resulting in increased risk of crash	Medium	Software error	Coding mistake
Brake light switch overheating	Brake light does not illuminate when brake pedal is depressed resulting in increased risk of crash. Or, vice versa.	Medium	Overheating	Insufficient heat insulation provided
Multi-function switch's slider deforms	Brake light does not illuminate when brake pedal is depressed resulting in increased risk of crash. Or, vice versa.	Medium	Deformation of slider	Incorrect geometry/ material specification
Body Control Module resistance increases	Brake light does not illuminate when brake pedal is depressed resulting in increased risk of crash. Or, vice versa.	Medium	Increased electrical resistance	Insufficient space from neighboring components and insufficient electrical insulation
Stop lamp switch malfunctions	Brake light does not illuminate when brake pedal is depressed resulting in increased risk of crash. Or, vice versa.	Medium	Stop lamp switch malfunctions	Supplier defect
Brake light switch stuck	Brake light illuminates even when brake not depressed resulting in increased risk of crash	Medium	Sticking switch	Improper geometric design of switch
Brake light switch not working	Brake light does not illuminate when brake pedal is depressed resulting in increased risk of crash.	Medium	Electrical contact not made	Improper assembly

In conducting the human analysis, the following observations were made. 1) DESC_DEFECT are large texts which include numerous model numbers. Manually reviewing each one to extract the actual defect description is time consuming and error prone. 2) In some cases, DESC_DEFECT listed multiple defects

https://doi.org/10.1017/pds.2025.10192 Published online by Cambridge University Press

within a single description. 3) Root cause for the defects can be extrapolated by analysing the CORRECTIVE_ACTION data, and 4) In some cases, further research into the physics behind the component's operation is needed to extract root cause of the failure.

8. Conclusions and future work

The research presented in this paper explores the use of publicly available recall data from the NHTSA for generating FMEA information. An LLM chatbot was used to analyse the data extracted. Specifically, the LLM was asked to extract top ten defects and generate failure modes, root causes, and severity rankings for these defects.

A qualitative comparison of LLM results generated based on NHTSA data against LLM results generated without using NHTSA data was performed. It was found that using NHTSA data results in more detailed FMEA information being produced while using technical terminology specific to the specific product being analysed. It is recommended that engineering designers prime the LLM chatbot with NHTSA data and extract FMEA information from it.

Another qualitative comparison was performed between NHTSA-primed LLM-generated results and human FMEA. It was found that DESC_DEFECT text length and similar descriptions with minor differences do not merge, leading to inaccuracies in frequency counts. It was also observed that the LLM-generated results covered a wider spread of components with single failure modes for each, while the human analysis was focused on fewer components but with multiple failure modes.

The effect of contextual knowledge provided to the LLM must be investigated. In this research, NHTSA data was provided and the LLM's generic training provided all other information related to FMEAs. Future research must be conducted to assess the effect of providing the LLM with FMEA course notes/ lectures.

Future work also includes automating the extraction of NHTSA data from their website and developing a customized, application-specific chatbot for engineering designers. In addition, verification of the usefulness of the proposed approach must be conducted by performing case studies of designers using the proposed approach in the forward design process of new products. To do this, metrics to assess the validity of the LLM-generated results must be established.

References

Chin, K. S., Chan, A., & Yang, J. B. (2008). Development of a fuzzy FMEA based product design system. *International Journal of Advanced Manufacturing Technology*, 36(7–8), 633–649. https://doi.org/10.1007/s00170-006-0898-3

ConsumerReports.org. (2024, August). Takata Airbag Recall: Everything You Need to Know, 2023 . https://www.Consumerreports.Org/Cars/Car-Recalls-Defects/Takata-Airbag-Recall-Everything-You-Need-to-Know-A1060713669/.

Rehman, Z., & Kifor, C. V. (2016). An Ontology to Support Semantic Management of FMEA Knowledge. *International Journal of Computers Communication & Control*, 11(4), 507–521.

Renu, R., Visotsky, D., Knackstedt, S., Mocko, G., Summers, J. D., & Schulte, J. (2016). A Knowledge Based FMEA to Support Identification and Management of Vehicle Flexible Component Issues. *Procedia CIRP*, 44, 157–162. https://doi.org/10.1016/j.procir.2016.02.112

Siemens, P. L. M. (2019). Software. *How to conduct a failure modes and effects analysis (FMEA)*.

Spreafico, C., & Sutrisno, A. (2023). Artificial Intelligence Assisted Social Failure Mode and Effect Analysis (FMEA) for Sustainable Product Design. *Sustainability*, 15(11). https://doi.org/10.3390/su15118678

Stamatis, D. H. (2003). Failure mode and effect analysis. *Quality Press*.

Wu, Z., Liu, W., & Nie, W. (2021). Literature Review and Prospect of the Development and Application of FMEA in Manufacturing Industry. *The International Journal of Advanced Manufacturing Technology*. https://doi.org/10.1007/s00170-020-06425-0/Published

https://doi.org/10.1017/pds.2025.10192 Published online by Cambridge University Press