Improving Loss Prevention In High Hazard Industries Through The Evaluation Of Safety Culture and Error Traps From Structured And Unstructured Data Using Machine Learning

Nyala Noë, Data Scientist, Empirisys, 2-4 Packhorse Road, Gerrards Cross, Buckinghamshire SL9 7QE
Emina Tuzovic, Data Scientist, Empirisys, 2-4 Packhorse Road, Gerrards Cross, Buckinghamshire SL9 7QE
Frederik Simon McArthur, Data Scientist, Empirisys, 2-4 Packhorse Road, Gerrards Cross, Buckinghamshire SL9 7QE
Angus Thompson Carroll, CEO, Empirisys, 2-4 Packhorse Road, Gerrards Cross, Buckinghamshire SL9 7QE

Unplanned outages are the bane of operational life. They need immediate attention and often require production to shut down primarily for safety reasons, leading to significant loss of economic activity. Operations data and academic research show that about 56% of events are due to latent system weaknesses that "hide in plain view" and thus are notoriously difficult to spot in advance of an incident. Latent system weaknesses create the preconditions for errors and reduce the ability of an operator to mitigate the consequences. They lie dormant until uncovered, typically during incident investigations, and are often referred to as performance influencing factors or "error traps". An error trap is a situation where the circumstances in combination with human cognitive limitations make errors more likely. To understand error traps, we need to identify their presence, strength, and potential impact. To achieve this insight, we look for distinct markers for each error trap in a company's structured and unstructured datasets. The benefit of this approach is that it enables a high proportion of the overall datasets to be analysed. It allows us to describe the prevailing safety culture and extract, classify, and index the data into suitable data visualisation outputs, providing users with actionable insight. This analysis advances the current approach of identifying, assessing, and reporting error traps by using a wider pool of data and improving the visualisation of any findings. We believe this approach will help establish a more holistic view of safety culture applicable across a wide array of high hazard sectors. Suppose we can reduce unplanned outages resulting from error traps. In that case, there will be a beneficial impact on loss prevention by helping to reduce major accidents and unplanned outages. This, in turn, will help improve production and directly benefit bottom-line performance, safety, and reputational and societal risks. This paper reports on research in collaboration with Centrica Storage Ltd.

Keywords: Error trap, User Testing, Human & Organisational factors, Natural Language Processing

Introduction

Managing and preventing major accidents in processing industries is undeniably a significant moral and economic concern that has been the focus of many academic, industrial, and regulatory studies over the years. However, major industrial accidents still occur with depressing regularity, often with very similar causes to historical events. Therefore, it is crucial to study the root causes of major accidents and analyse their impacts. Most root causes are human errors (80%). But, ultimately, these errors result from flaws in the context in which the human operated. Only ~30% result from a genuine lapse in judgment or willful (albeit usually well-intentioned) violations, with the remaining ~70% due to system errors (Wachter & Yorio, 2013).

These system errors, also known as error traps, performance-influencing factors, or latent defects, are factors that make accidents more likely to occur. Note that in this paper, we use the term error traps. The challenge in effectively detecting error traps is that they often hide in plain sight but are usually discovered only during the accident investigation after the event. Because most human-induced errors are directly related to error traps, addressing them is essential to avoid incidents. In this context, incidents include accidents, injuries, casualties, and less severe occurrences, such as near-misses or system outages.

Error traps encompass a wide variety of factors at the individual, task, and organizational level. Examples of individual error traps include fatigue and stress. Task-related error traps include the use of inappropriate tools and the absence of proper equipment. Finally, some examples of organizational error traps are peer pressure or lack of leadership.

A key aspect of error traps is that they rarely occur in isolation. Fatigue, stress, and time-pressure is an often-found combination that induces humans to commit errors. For example, think of the Deepwater Horizon disaster in 2010, where an explosion led to the death of 11 people, with 17 other injured. The Transocean platform exploded as gas was freed during a flawed cement job. A blowout preventer that should have been activated to prevent the explosion failed.

In the subsequent investigation, some of the error traps that came to light are the following:

- ***Inadequate preparation for task:*** the cement used was against best practices. Only six centralizers were used instead of 21
- ***System/equipment interface failures:*** the gas alarm system failed, which led to severe cognitive overload as the incident gathered pace
- ***Unusual occurrence:*** the pressure test was misinterpreted because the results seemed contradictory (confirmation bias)
- ***Time and money pressure:*** the well was 43 days behind schedule, resulting in skipping an important cement bond log test which would have taken another 9-12 hours and cost more money

- ***Absence of psychological safety:*** fear of speaking out about safety concerns among rig workers if these concerns led to more delays
- ***Communication, leadership level:*** there was a recurring pattern from leadership to focus on occupational safety concerns from workers and contractors

Literature review

Error traps

Error traps have many different names in the academic literature. In a sub-sample of 15 representative papers on safety performance in organisations, only two terms appeared in more than one paper: 'factors' appeared in three, and 'human factors' in two. A total of eleven different terms refer to the same concept as error traps, including critical success factor, human error contributing factor, performance influencing factors, risk influencing factor, hazard reducing measures, root causes, and variables/clusters. Most error trap research focuses on high-hazard industries, but error traps also apply to other sectors.

Beyond terminology, there is also a lack of consensus on how to classify and define error traps. Researchers have attempted to consolidate the different frameworks used in human reliability analysis into one unifying taxonomy (Kim and Jung, 2003). Kim and Jung (2003) combined 18 disparate taxonomies and condensed them into 11 main items with 39 sub-items across four categories. Their final taxonomy consisted of '*Human*' (training and experience), '*Task*' (availability and quality of procedures, simultaneous goals and tasks, task type and attributes), '*System*' (availability and quality of information, status and trend of critical parameters, status of safety system and components, time pressure), and '*Environment*' (working environmental features, team cooperation and communication, plant policy and safety culture). The Health and Safety Executive (HSE), on the other hand, have defined a non-exhaustive list of 26 error traps (called performance-influencing factors in their taxonomy) under three categories: person, task, and organisation ('Performance Influencing Factors (PIFs)', 2019).

One neglected aspect in the taxonomy debate is how to measure error traps. Assessment of error traps mainly consists of qualitative measures, such as surveys and interviews by assessors using human reliability analysis frameworks, either before or during accident reports. This approach has some problems. It relies heavily on the subjective perception of employees (which may not be an accurate reflection of the actual state of the error trap in question), it is cumbersome and time-consuming, and it misses any error traps that employees have not identified. The last point is crucial to consider when we realise that a defining feature of error traps is that they are hiding in plain sight.

The gap between academia and industry

The gap between academia and industry on error traps has not escaped the wider safety community (Glendon and Stanton, 2000). Academia has tried to consolidate thinking around different safety frameworks, the differences between safety climate and culture, and taxonomy completeness (Kim and Jung, 2003). Industry uses KPIs to measure safety performance. Often different companies use different KPIs, and sometimes business units within the same company use varying KPIs. The absence of standardisation makes it hard to establish industry benchmarks and gain insight into the most common error traps. This could stem from a lack of consensus and definition of what constitutes an error trap, a lack of understanding and tools to effectively measure error traps, or under-estimation of the role of error traps in incidents. In any case, we believe that narrowing the gap between academia and industry benefits the safety community, organisations, and -most of all- the people that work in them.

Data collection as a hurdle to effective safety management

Both academia and industry rely on manually collected data when studying the impact of human factors on safety. Frequently used tools are self-reporting methodologies like questionnaires and interviews. Human reliability assessments also commonly consider error traps, although their use and inclusion vary with industry and framework (Kim and Jung, 2003). The systematic use of system data to assess the presence and severity of error traps is novel. Consider, for example, the person-related error trap of fatigue. One can ask people whether they experience fatigue as part of a one-off questionnaire. But system data about shift patterns and overtime hours could provide additional information. These readily available data points offer a proxy measure for fatigue. System data can highlight areas of concern when surveying employees is complex or cumbersome. It can also complement the subjective perspectives collected in interviews and surveys. Unstructured data, such as text in reports, is currently underused altogether because it is not easy to interpret. And interviews are conducted on samples because it is too hard to analyse all the information. Natural Language Processing (NLP) technology now makes this vast and rich data source accessible for insights.

The importance of feedback loops

After identifying error traps, the question of what to do next remains. A system that tracks error traps and alerts the user when there is a problem is a practical first step. Next, each error trap needs its own sets of recommendations specific to the industry, company, and asset that it affects. We consult external bodies such as the HSE to define recommendations, use historical records where the asset has suffered the same problem before, or gain insights from similar experiences in other sites and companies.

Leadership decisions that negatively impact safety measures tend to be rooted in the pursuit of profits over safety. Employees perform best in an environment where production and safety goals are compatible and aligned (McLain and Jarrell, 2007). Therefore, error traps should be monitored and dealt with as soon as they arise: they signal a vulnerability in a barrier or system that, if acted upon, can reduce the likelihood of errors, leading to fewer outages and incidents. Being safe is being profitable.

Frameworks used in academia and in industry

Academia and industry have proposed several strategies and frameworks to mitigate error traps. While some of them are industry-specific, many frameworks are developed as a general solution for detecting, dealing with, and preventing these issues from occurring. Note that the severity of an error trap varies by organisation, circumstance, and general work environment. While some error traps can have a broad impact across the organization (e.g., communication issues), others are more specific (such as using a particular material). But it is the combination of these two types that creates an unsafe working environment.

Expedition Engineering Ltd (2016) developed a strategy to manage errors related to skill deficiencies in the construction industry. The firm used the grounded theory method to collect and analyse the causes of error and the methods used to avoid errors. They then used the Delphi method to rank results and assess financial impact of errors, latent defects and unrecorded process waste. Note that we have learned from post-Grenfell inquiries that this approach is flawed. A steering group made up of individuals from the construction industry supervised the process. They provided regular feedback and participated in the analyses.

Darabont, Badea, and Trifu (2020) develop an approach to understand the causes of accidents without blaming any one individual. The Human Factors Analysis and Classification System (HFACS) relies on four levels of deficiencies that lead to accidents: unsafe acts, pre-conditions for unsafe acts, unsafe supervision and organizational failures.

Gisquet, Beauquier, and Poulain (2020) propose to improve safety in nuclear facilities via organisational culture. Their cultural analysis framework has three levels. Firstly the Macro level, which includes analysis of the economic, political, regulatory, institutional, and cultural context, Secondly, the Meso level to review analysis of the organizational culture and the social hierarchy within the company. Thirdly, the Micro-level, where analysis of the professional culture, collective life and social relations takes place.

Shehata and Faroun (2018) aim to address the root causes of the Oil and Gas industry incidents. They propose a tool to identify corrective actions after an incident. The framework has multiple steps including selecting equipment type, identifying key accident contributors, identifying timing error per project life cycle, identifying direct causes and root causes of the event, and finally selecting the cost-effective corrective actions.

Lastly, the HSE proposes a method for classifying performance influencing factors, equivalent to error traps, into three categories: job factors, person factors, and organisational factors. They suggest four approaches for different types of analyses, including human factors in risk assessments, analysing incidents or accidents and near misses, analysing human factors in design and procurement, or human factors in other aspects of health and safety management.

Research Questions

We mentioned that a practical first step to reducing human error is a system that tracks error traps and alerts the user when there is a problem. We design the DETECT (Detection of Error Trap Effect and Consequence Tool) for this purpose. DETECT MVP1 is a non-functional prototype built to test user desirability and user experience. Our initial questions are the following:

- How would DETECT be used, and for which purposes?
- What information does the user want to see in DETECT?
- How does DETECT contribute to safety performance?

In summary, we aim to understand what insights we can gain from a company's structured and unstructured data and how the results can improve safety performance.

DETECT (Detection of Error Trap Effect and Consequence Tool)

We propose a tool that uses organisational data that is readily available to augment and enhance the existing analysis performed manually during human reliability assessments. This addresses all three issues outlined above as it (a) enhances the subjective perception of employees with objective system data (b) uncovers error traps missed by employees (c) is fast and can process large amounts of data.

The tool, called DETECT (Detection of Error Trap Effect and Consequence Tool), plugs into an organisation's existing system and uses statistical and heuristics-based algorithms to detect the presence and severity of error traps in an organisation. Each error trap would therefore get a score, which can be used as a KPI in the organisation. The interaction and interpretation with employees at all levels of the organisation is an important part of the tool. It relies on human input in addition to system data to augment an organisation's understanding of which areas to address.

For a tool like DETECT to be useful, error traps need to be made measurable. An important distinction we therefore wish to make is between what constitutes an error trap and what constitutes a marker, as this distinction has not been made in the error trap literature so far, to the best of our knowledge. We define the two concepts as follows:

1. Error Trap: a composite (KPI) metric that is derived from multiple marker scores, and belongs to one of three factors: organisation, job, or person
2. Marker: a simple (KPI) metric that is assigned to one or more error traps

Let us consider an example of an error trap we wish to measure: 'Clarity of signs, signals, instructions and other information' under the 'Job factors' category, as defined by the HSE ('Performance Influencing Factors (PIFs)', 2019). Within this error trap, let us assume we only consider instructions - written instructions specifically.
Written instructions can be measured by two markers:

1. Clarity of written instructions (as measured by the Flesch test)
2. The language ability of workers (as measured by a language test score)

The composite measure for the error trap as derived from its markers works as follows 1: two marker scores roll up to a single derived score for the error trap, which in turn contributes to a derived score for the error trap category they belong to, for example, 'Job factors'. One marker can also be assigned to more than one error trap within the same set of factors. For example, the language ability of workers also contributes to the error trap 'Communication, with colleagues, supervision, contractor, other' under 'Job factors'. Finally, one marker can be assigned to more than one error trap across sets of factors. For example, the language ability of workers also contributes to the error trap 'Communication' under 'Organization factors' ('Performance Influencing Factors (PIFs)', 2019).

By defining each error trap with an associated marker pattern, it allows DETECT to make inferences about their presence and severity without relying on whether the error trap is already known to employees or not.
We need to be aware that some error traps are much more difficult to measure than others, and that not all markers will be measurable in every organisation, because of differences in data availability and organisational structure. This means that DETECT, while having a generalisable framework, will need to be made bespoke to every organisation through a baselining exercise. Once launched, however, it offers continuous feedback on the state of the organisation with regards to its potential error traps.

DETECT Prototype

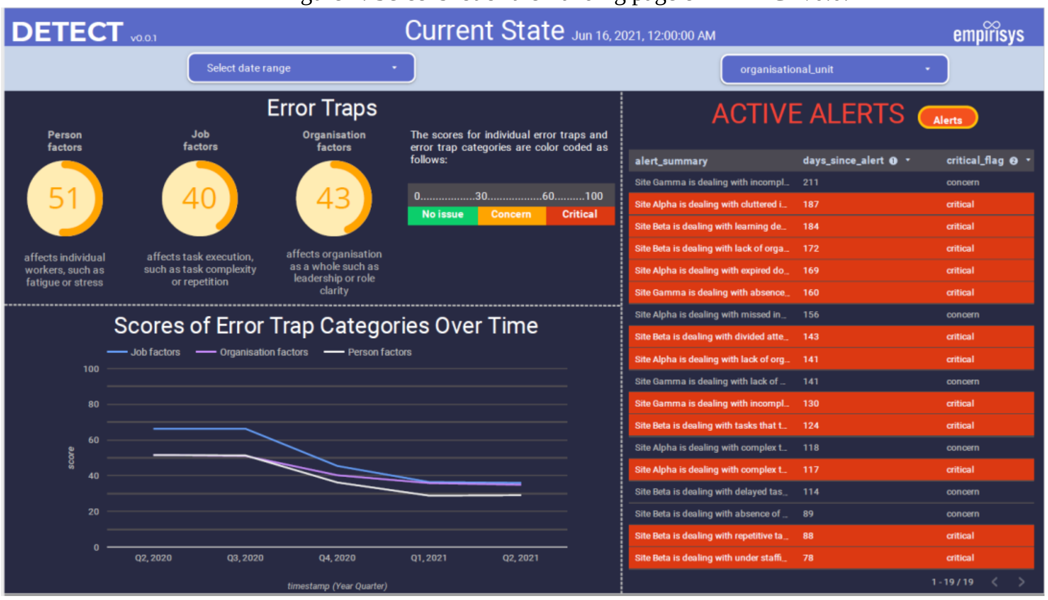
The prototype used artificially-generated data and was built in Google Data Studio as an interactive dashboard. The dashboard was connected to a MySQL server containing the data displayed in DETECT.

The DETECT prototype consists of six pages:

- **The Landing page** gives an overview score of the three error trap categories (individual, job, and organisation) in the form of gauges. It also provides the user with these same three scores over time and an alert table that shows them the specific areas of concern in the organisation, ordered by date and severity.
- **The Alerts page** gives an overview of the active alerts that need to be addressed. An example of an alert is "appropriate tools unavailable for task" or "task instructions too complex". An alert gets triggered when DETECT first becomes aware of the presence of a problem and assigns a severity score to it. For example, if "instructions too complex" is only a problem for a few documents, the severity score will be lower than if this problem exists for almost all documents. Whether instructions are too complex or not is measured by DETECT using various statistical, machine-learning, and heuristics-based algorithms.
- **The Feedback page** allows users to take actions based on the alerts seen on the **Alerts page**. The limited functionality of Google Data Studio did not allow for an interactive experience, but the desired functionality was displayed to the user with the help of a gif. It showed how the user can select a specific alert and submit an intervention based on it. The page gives some general recommendations and statistics on the alert so the user can submit an appropriate intervention.
- **The Interventions page** gives an overview of how the interventions perform over time. It displays a graph of the alerts with the interventions overlayed so users can monitor whether their interventions are having any effect. If interventions are effective, they should see the alerts disappear or reduce.
- **The Error Trap Deep Dive page** has a single time-based graph that displays the scores for the different error traps. It has many drill-down opportunities, with the ability to slice the data by error trap category, error trap, organisational unit, and date range. It also gives an overview of the best and worst-performing error traps in the organisation.
- **The Comparison page** shows how we can compare organisational performance against a baseline. One graph shows a bar chart version with the performance of different sites alongside a baseline. The other graph displays this

same performance over time, with the baseline indicated through a dotted line. This page allows users to monitor whether they are improving on their baseline performance.

Figure 1: Screenshot of the Landing page of DETECTv0.0.1



User Testing Sessions - Methodology

5 highly experienced asset, operations and safety experts from the oil&gas and chemicals sectors were each presented with the DETECT prototype during a 1-hour structured interview. The goal of the user testing sessions was to understand the need for a tool like DETECT and how information about error traps should be presented to be useful and actionable. All participants received minimal information on the product before the testing sessions, as we were interested in their immediate impressions. In attendance at the session were the interviewer, an observer (who was off-screen and silent the whole session), and the participant. The interviewer was leading the session, while the observer solely focused on taking notes, without intervening in the session or interacting with the interviewee. The interview sessions were carried out between the 24th of June and the 29th of June over Teams.

The interview had the following structure:

- Introduction and begin recording session after getting participant consent
- Confirm consent to participate in the study, reminding participants they can stop the session anytime they want and for any reason
- Brief introduction from the interviewer about error traps and human error in the context of industrial incidents in high-hazard industries. Terminology around error traps is clarified, as the interviewees might be more familiar with the terms latent defects, or performance-influencing factors
- The interviewee is asked about their background in high-hazard industry and their experiences in safety
- The interviewee is taken through the prototype. After sharing the link to the prototype in the Teams chat, the interviewee was asked to open the link and to share their screen. All the interviewees were seeing the prototype for the first time. The interviewer takes the interviewee through each page, asking the following questions:
    - What are your first impressions of the page?
    - What do you notice on this page?
    - If you could only keep one graph/section of the page, which one would it be?

- The interview was concluded with the following wrap-up questions:
    - What is the one visualisation you remember from the prototype?
    - What did you like about this experience? Why?
    - What did you not like about this experience? Why?
    - How would using this dashboard fit into your day?
    - What information do you need to address a safety risk/error trap?
    - Do you feel there is a need for an error trap detection tool?

- Finally, the recording was ended, the interviewee was thanked for their time, and the Teams call was ended.

User Testing Sessions – Results

We focus first on interviewees' answers to the last question of the interview: Do you feel there is a need for an error trap detection tool like DETECT?

All participants responded positively. They gave the following reasons.

- It would be possible to use at all levels of the organisation due to the drill-down capabilities of the tool.
- The tool provides context for observations made by employees. The tool would serve as a foundation for discussing safety-related issues during daily, weekly, or monthly meetings.
- The tool allows for interventions and actions to be taken based on the information provided.
- The tool gives all the information needed to make a decision "at your fingertips" while still providing a concise image of the essential information necessary to react to safety concerns."
- By relying on system data, the factual nature of the tool allows for a solid argument on why interventions are being carried out: "It is not somebody's opinion, it is data."

The participants also suggested the following features which were not present in the prototype:

- Provide recommendations based on historic interventions made under similar circumstances or on industry benchmarking
- Provide the ability to sign-off interventions in the tool directly to avoid duplication with other systems
- Give the tool the ability to integrate with existing systems

Most participants would use the tool weekly or monthly. One participant wanted to use it at different meetings on a daily, weekly, and monthly basis but at varying levels of granularity. The overview would be presented and discussed at monthly meetings. Daily meetings would use more granular information, focusing on specific sites and error traps.

Users highlighted that the tool requires a certain level of data maturity to be helpful. For example, companies with mostly paper-based systems would get less value from the tool than companies that store volumes of data in a data lake. DETECT would allow an organisation to see what data is used by the tool at any time and identify any blind spots. This can help guide the organisation's data strategy to feed DETECT and improve safety performance by extension.

The tool's ability to drill down from a high-level overview into the detail of the error trap scores and alerts is considered crucial by all interviewees. This feature allows it to become an interactive tool for the whole organisation. It presents an objective, unbiased view of the state of the entire organisation, with the data to back it up. Engaging workers with the tool allows them to take ownership and responsibility for their own safety and that of others is an essential aspect of safe company culture.

Another critical aspect that several interviewees noted was that the dashboard presents information in a simple and easy-to-understand way. They felt that this is especially important for managers and c-level employees who need to understand a situation quickly. Transparency and simplicity are some of the core design principles of DETECT.

Finally, the direct feedback the tool offers is a key aspect. In an effective intervention, the tool should reflect improved scores for error traps in the targeted areas. If the improvement is not visible, employees know that the intervention was ineffective or targeted the wrong aspect of an error trap. The intervention team can understand which interventions were successful and not, allowing for learning and a more tailored response next time.

The above has a few implications. DETECT must be deployed under the expert guidance of human and organisational factors specialists. Data derived from maintenance or asset integrity management systems must be deployed with operational specialists. However, whilst not underestimating the complexity of achieving this goal, DETECT provides a viable means for systematically collecting and reporting a set of "weak signals" to process safety vulnerabilities, which are often described within the context of high-reliability organisations.

An Example of Unstructured Data Analysis

Analysis often excludes qualitative data simply because it is difficult to make sense of unstructured information like words, documents and topics. Written reports, incident investigations and maintenance logs are usually carefully stored but rarely analysed. Centrica Storage Limited shared with us a sample of log datasets, including over 21,600 observations, 14,200 notifications, and 450 events spanning three years at three separate sites.

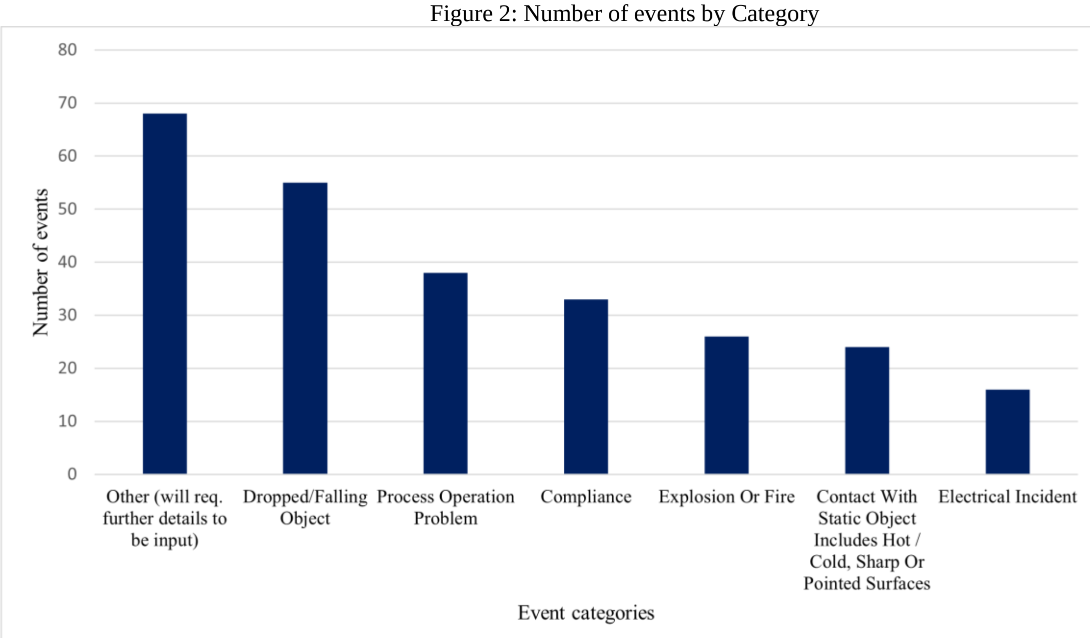
Unstructured Data Analysis - Descriptive statistics

Looking first at the observations dataset, the most common Observation falls in the category of "Safety Systems / Plant Equipment", with just under 3000 observations. This is followed by "Communication", "Housekeeping" and "Practices / Behaviours" in that order, with 2380, 2305, and 2261 observations respectively under these categories. There are a total of 38 Observation categories.

Within the observation data, there was a lot of information that could be found from the word associations. Some of this data is to be expected, for example, the word "deck" commonly occurring with the word "skid", or the word "team" commonly occurring with "leader". There were some interesting insights found in these relations - the word "gas" occurring commonly with "smell", or "access" commonly occurring with "restricted", with both of these having the possibility to cause an incident or make one much worse.

Events are a bit more complicated than observations, as they are divided into categories (Figure 2) and sub-categories. This means that the category of Compliance could contain many sub-categories like Potential Injury, Potential Material Loss, and Loss of Product. This makes the task of subdividing them more difficult. The most common event in our data is classed as "Other (will require further descriptions)". To gather more insight into this data, we can look at how it breaks down by sub-type. The most common sub-type is "Potential material Loss", corresponding to just under half the events in this "Other" category. This is followed by "Potential Injury" and "First Aid". After the "Other" category, the most common categories are "Dropped/Falling Object" and "Process Operation Problem".

Figure 2: Number of events by Category



Unstructured Data Analysis - Methodology

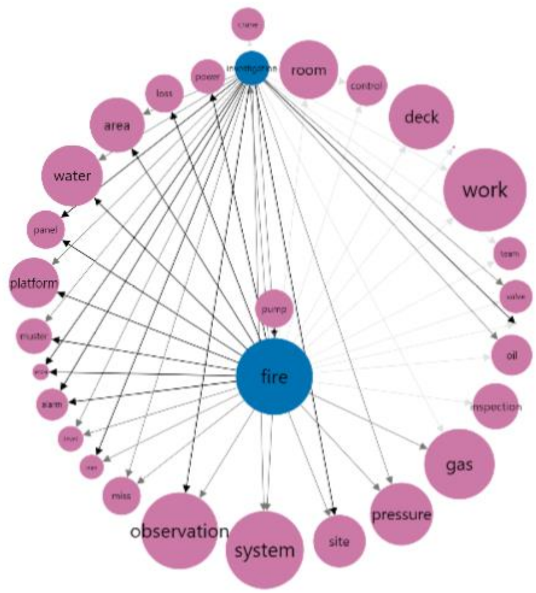
The analysis of unstructured data can provide a powerful tool for highlighting key areas. Using artificial intelligence and especially natural language processing makes it now possible to investigate great amounts of textual data in minimal time. For this purpose, it is important not only to understand each particular word in a given context, but also independently. In order to quantify the meaningfulness and relationship between words in a given context, the Pointwise Mutual Information (PMI) criterion is a highly reliable tool (Church and Hanks, 1990). It allows for the quantification of the likelihood of having two words co-occur together by using a probabilistic approach.

For this analysis, we focused on the 'Events' dataset from CSL, where each row describes a particular event. We used the description of the events as our main data input, because we were carrying out our analysis on unstructured data. First, we needed to build a co-occurrence matrix where we focused on the relationship between center and context words. For this analysis, we considered that a center word can only be a noun, but a context word can either be a verb or an adjective. Finally, we converted the co-occurrence values to PMI scores.

Unstructured Data Analysis - Results

We then built a word network based on the results of the unstructured data analysis, where the nodes are words and the edges are their PMI score, which quantifies how strongly associated two words are. The Events word network comprised of 3,478 unique word pairs. Since we have a large amount of textual data and an abundance of words, it is important to filter the network down when focusing on a specific topic of interest. For example, if we were to focus on events related to fires or investigations, we can filter the initial network to only include specific terms, resulting in a much more focused, and readable, network as shown in Figure 3.

Figure 3: Compact 'Events' word network for 'Fire' and 'Investigation'



In this small network, the nodes colored in blue are center words. The nodes colored in pink are context words. The links are in different shades of black according to PMI scores. In particular, the darker the link, the stronger the relationship between the two considered words. The diameter of each node is relative to the frequency of the word in the context of the whole textual dataset.

The network requires human interpretation to be useful. For example, if the word 'Fire' occurs a lot with 'Gas' and 'Pressure', this might be an indication that there is a re-occurring issue surrounding these terms. This is especially relevant if the co-occurrence of these words does not diminish over time, despite numerous investigations and reports. These networks often require subject-matter experts to be properly interpreted, highlighting the need for continuously having humans in the loop, even when using machine-learning powered tools such as DETECT.

We can apply the same kind of analysis to the observations as well as the maintenance datasets and visualise the corresponding word networks with the possibility to filter by location, relationship strength, timeframe, source words or even target words to name a few. Furthermore, the same analysis of unstructured data can also be a powerful tool for investigating answers to open-ended questions within surveys. Therefore, it is now possible to have a better grasp of how people talk and relate to certain topics by extracting word pairs with strong relationships in survey answers.

Discussion

To conclude, users see a strong need for a product like DETECT, which could be used across the whole organisation to address error traps. Furthermore, it appears that users are particularly interested in DETECT as they consider that it would be an effective way to encourage employees to take action and form the appropriate intervention plans to mitigate error traps.

As a result, users would ideally want DETECT to offer the possibility to drill-down through different layers of data leading to the raw data sources. Therefore, users require complete transparency of the information displayed by the tool. In addition, a separate section with the ability to input interventions and monitor the effectiveness of each plan is a highly desirable feature sought by most users.

DETECT can strongly contribute in improving safety performance within the industry. In the majority of cases, unstructured data is available in high volumes but hardly ever analysed thoroughly. Using data science and particularly natural language processing to extract insights from usually overlooked sources of data can save a large amount of time while focusing on key areas relevant to the organisation. Word co-occurrence outputs can become valuable inputs for tools like DETECT. Word frequencies and PMI scores are measures that can be used as variables in error trap detection algorithms. This is especially valuable when words can be associated with specific error traps and where changes in frequency over time can indicate changes in prevalence of a specific error trap.

It is important to point out the limitations of our work. The degree of reliability regarding the results strongly depends on the quality and the quantity of data. This in turn relies on the accuracy and rigour with which reports are submitted, itself an indicator of leadership and culture in an organisation. The tool requires human validation and supervision to be useful, as outputs need to be interpreted before becoming actionable.

Industrial disasters do not only affect the victims. They also impact the friends, families, and colleagues of those involved in these incidents. The psychological trauma and material loss to the company are important burdens to address. People across academia and industry are working ceaselessly on developing new tools and strategies to ensure safer work environments. We hope that by building DETECT, we too can help in the fight against industrial disasters.

Acknowledgments

We would like to thank Mike Orley and Centrica Storage Limited for providing us with valuable datasets for the analysis reported on in this paper.

References

Church, K. W. and Hanks, P., 1990, Word Association Norms, Mutual Information, and Lexicography, *Computational Linguistics*, 16(1): 22–29.

Darabont, D. C., Badea, D. O. and Trifu, A., 2020, Comparison of four major industrial disasters from the perspective of human error factor, *MATEC Web of Conferences*, 305: 00017.

Expedition Engineering Ltd, 2016, Get It Right Initiative - Improving Value by Eliminating Error', 24: 1–74. Available at: https://getitright.uk.com/app/uploads/2017/03/GIRI-Research-Report-Revision-3.pdf.

Gisquet, E., Beauquier, S. and Poulain, E., 2020, Revisiting Safety Culture: The Benefits of a New Cultural Analysis Framework for Safety Management, *Nuclear Technology*.

Glendon, A. I. and Stanton, N. A., 2000, Perspectives on safety culture, *Safety Science*, 34: 193-214..

Health and Safety Executive (no date) 'Performance Influencing Factors (PIFs)'. Available at: https://www.hse.gov.uk/humanfactors/topics/pifs.pdf.

Kariuki, S. G. and Löwe, K., 2007, Integrating human factors into process hazard analysis, *Reliability Engineering and System Safety*, 92(12): 1764–1773.

Kim, J. W. and Jung, W., 2003, A taxonomy of performance influencing factors for human reliability analysis of emergency tasks, *Journal of Loss Prevention in the Process Industries*, 16: 479–495.

Kyriakidis, M., Majumdar, A. and Ochieng, W. Y., 2018, The human performance railway operational index—a novel approach to assess human performance for railway operations, *Reliability Engineering and System Safety*, 170:226–243.

McLain, D. L. and Jarrell, K. A., 2007, The perceived compatibility of safety and production expectations in hazardous occupations, *Journal of Safety Research*, 38(3): 299–309.

Shehata, W. M. and Faroun, H. A., 2018, A Tool to Identify the Proactive Corrective Actions after the Accidents in Oil and Gas Industry', *International Journal of Petroleum and Petrochemical Engineering*, 4(1): 1–12.

Sherry, L. *et al.*, 2000, Formal method for identifying two types of automation surprises. Honeywell, Phoenix, AZ, Tech. Rep, 1–6.

Wachter, J.K. and Yorio, P.L., 2013, Engaging Workers as the Best Defense Against Errors & Error Precursors, *Safety Management*, 54-64.