MILITARY UNIVERSITY OF TECHNOLOGY
FACULTY OF CYBERNETICS
INSTITUTE OF INFORMATION SYSTEMS

Wyznaczanie ukrytych relacji w tekstach w oparciu o wzorce.

Doctoral thesis
submitted by

mgr inż. Michał Gałusza

supervised by:
prof. Andrzej Walczak
Warsaw, 2024

Institute of Information Systems
Military University of Technology

Declaration

I, the undersigned .... hereby declare that I am the sole author of this thesis. To the best of my knowledge, this thesis contains no material previously published by anyone except where due acknowledgment has been made. This thesis contains no material that has been accepted as part of the requirements of any other academic degree or non-degree program in English or any other language. This is a true copy of the thesis, including final revisions.

....

Warsaw, 31st Jan 2024 Your signature

Acknowledgements

Podziękowania

Abstract

The thesis proposes a solution to the problem of recognizing and modeling risk transmission contained in documents describing a system's operation or failure. Relationships are determined in the context of entire documents and go beyond the currently dominant approach of identifying relationships within a single sentence and using classifiers trained on collected dedicated training examples.

The problem being addressed is significant because information about the flow of threats can create complex interactions between system elements described in such a way that they may be scattered throughout the document and even between sources. There is generally a lack of dedicated training sets for classifiers, and existing solutions are limited to selected areas, such as railways, or description formats, such as HAZOP or FMEA.

The proposed solution involves a gradual decomposition of descriptions. The examined text is decomposed into a Semantic Frames Graph (SFG) in the first step. In the second step, the pattern of threat propagation relationships is used to recognize propagation. Recognized propagations are stored in an Intermediate Relationship Graph (IRG). In the final step, propagations are aggregated into the form of an Asset-Vulnerability-Hazard (A-V-H) graph, which allows for a network analysis of the risk contained in the description of the operation of a given system.

The proposed approach allows for modeling risk propagation without needing a dedicated relationship detection mechanism, as this method is based on verbalizing the relationship pattern. Another reason for eliminating dedicated classification is the extension of pattern analysis to analyze the dialog coherence in the path between nodes in the SFG graph. The detection results obtained by combining both methods are verified using current language models (Large Language Models such as chatGPT) and prompt engineering. The threshold above which relationships are accepted is a solution to the multi-criteria optimization task.

Overall, this work presents a new method for detecting relationships and its application in risk analysis. It also explores the potential of semantic pattern methods, dialogic coherence, and prompt engineering in constructing a network risk model, which facilitates modeling complex threat propagation dependencies.

Streszczenie

Rozprawa proponuje rozwiązanie dla problemu rozpoznawania i modelowania transmisji ryzyka zawartego w dokumentach opisujących działanie systemu lub opisujących jego awarię. Relacje wyznaczane są w konteście całych dokumentów i wykraczają poza aktualnie dominujące podejście wyznaczania relacji w ramach jednego zdania oraz przy użyciu klasyfikatorów wytrenowanych na zebranych dedykowanych przykładach trenujących.

Rozwiązywany problem jest istotny jako, że informacje o przeływie zagrożenia mogą tworzyć skomplikowane interakcje pomiędzy elementami systemu opisanymi w taki sposób, że mogą być rozproszone po całym dokumencie a nawet pomiędzy źródłami i na ogół brakuje dedykowanych zbiorów trenujących klasyfikatory a istniejące rozwiązania są ograniczone do wybranych obszarów np.: kolei, lub formatów opisow np.: HAZOP lub FMEA.

Proponowane rozwiązanie zakłada stopniową dekopmozycję opisów. W pierwszym kroku badany tekst dekomponowany jest do postaci Grafu Ramek Semantycznych (ang. Semarntic Frames Graph, SFG). W drugim kroku, wzorzec relacji propagacji zagrożenia używany jest do rozpoznania propagacji. Rozpoznane propagacje zapisywane są w grafie Pośrednich Relacji Semantycznych (ang. Intermediate Relathionship Graph, IRG). W ostatnim kroku, propagacje są agregowane do postaci grafu Zasób-Podatność-Zagrożenie (ZPZ) (ang. Asset-Vulnerability-Hazard, A-V-H), który pozwala na sieciową analizę ryzyka zawartego w opisie dzialania danego systemu.

Zaproponowane podejście pozwala na modelowanie propagacji ryzyka bez konieczności stosowanie dedykowanego mechanizmu detekcji relacji jako, że metoda ta opiera sie na werbalizacji wzorca relacji (ang. pattern verbalization). Drugim powodem, który pozwana na eliminację dedykowanej klasyfikacji jest rozszerzenie analizy wzorca o analizę spójnościa dialogowej w scieżce pomiedzy węzłami w grafie SFG. Wyniki detekcji uzyskanych poprzez zestawienie obu metod weryfikowane są poprzez wykorzystanie aktuanych językowch modeli generatywnych (Large Language Models np.: chatGPT) oraz inżynierii podpowiedzi (ang. prompt engineering). Próg powyżej którego relacje są akceptowane jest rozwiązaniem zadania optymalizacji wielokryterialnej.

Ogólnie nieniejsza praca przedstawia nową metodą detekcji relacji i jej zastosowanie w obszarze analizy ryzyka. Przedstawia potencjał metody wzorców semantycznych, spójności dialogowej oraz inżynierii podpowiedzi w konstruowaniu sieciowego modelu ryzyka, który ułatwia modelowanie złożonych zależności propagacji zagrożenia.

Acronyms

RA Risk Analysis

LM Language Model

LLM Large Language Model

RM Risk Management

SFG Semantic Frames Graph

SRL Semantic Role Labelling

IRG Intermediate Relationships Graph

A-V-H Asset-Vunerability-Hazard

OIE Open Information Extraction

KG Knowledge Graph

KA Knowledge Acquisition

KAP Knowledge Acquisition Pipeline

SN Semantic Network

PHA Pre-Hazard Analysis

HAZOP Hazard and Operability Study

OIE Open Information Extraction

TL Tranfer Learning

NER Named Entity Recognition

NLI Natural Language Inference

RE Relationship Extraction

Contents

Chapter 1

Introduction

Cambridge English Dictionary defines the adjective "hidden" as «not easy to find». Although "not easy" is connected with "difficult", in the domain of risk analysis, it should instead refer to the fact of not being directly detectable. The «hidden» nature of risk relations results from human language's flexibility in expressing the risk descriptions. In addition, the semantics' of risk is characterized by extreme contextuality.

Consider a sample sentence: *A hunter shot a raging bear*. Both *a hunter* and *a bear* are *threats* depending on the context. For an object *bear*, *hunter* is a *Threat* or *Hazard* as he shot it eventually; however, for an object *hunter a bear* is a *Threat* therefore it was shot.

Contextuality is even more visible if we expand the scope. Let's consider another sentence: *"A hunter shot a raging bear attacking a woman"*. Within this single sentence, object *hunter* should be assigned two roles simultaneously: *Threat* from the bear's perspective and *Savior* from the woman's. Therefore, which concept should represent a word *hunter*?

In the risk domain, such contextual situations are not uncommon and require a contextual approach to detect them correctly. For example, the contextual role of the package (Fig. 1.1) depends on whether we consider a human underneath - in this case, it will belong to class *Hazard*. However, given that the package is valuable, it will be considered as *Asset*, which *Vaulnerability* would be a line carrying it and *Hazard* an event of the line snapping.

The contextuality challenges a classical knowledge graph construction as the KG construction relies on the notion of *"concept"*. The hierarchy of concepts and the relations between them is the foundation of the representation, of the model of the domain of interest. That organized structure forms an *Ontology* a foundation for *Inference*. In the classical approach, a concept is a component of human thought and is the thinking unit that refers to objective things and their peculiar properties. A concept's formation is a procedure with the direction "from special to general". Considering various objects that are "special" cases, one determines a "general" set of properties that form the concept. This implies that we can define the concepts only through their properties and how linguistic expressions of concepts exist within the narrative. As with the word "apple," we can associate the information related to its shape, color, taste, and the context in which it usually appears in any narrative.

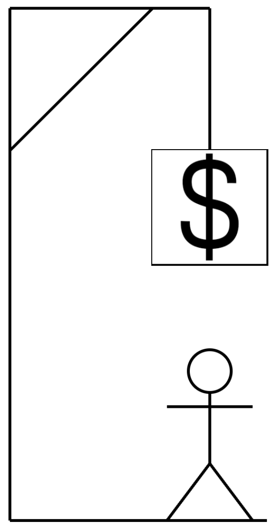


Figure 1.1: Risk Contextuality

We can observe that other words, such as "peach" and "banana," share the same linguistic properties; therefore, they are similar. All of these allow classifying "apple", "peach", and "banana" to the concept of "fruit".

Using a similar approach to identifying *Hazard* concept would require enumerating its features or effects to detect them in the parsed descriptions. Due to the unmanageable number of combinations, such an approach is impossible as the concept of hazard manifests itself in various domains differently. For example, in the medical domain, the hazard manifests through adverse drug effects, impact on the organs, or general deterioration of patients' medical conditions. In the financial realm, risk will be manifested through capital loss. In software engineering, through a data breach or unexpected system malfunction.

The thesis aims to extract risk-related interactions from the text. It uses a specific risk structure, a triple Asset-Vulnerability-Threst [1], that constrains how the interactions are identified. The approach proposed relaxes the problem of direct detection of risk-related concepts and formulates the methodology, which allows constructing the comprehensive representation of risk interactions in the form of a specific Knowledge Graph called Asset-Vulnerability-Hazard graph (pol. graf Zasób-Podatość-Zagrożenie) [1].

1.1 Thesis structure

The thesis is structured as follows:

- Chapter 1 – **Introduction**

  This chapter explains the motivations behind the research. It also outlines challenges in Risk Analysis and requirements for a solution to construct a comprehensive risk representation.

- Chapter 2 – **Related Work**
  The chapter summarizes related work in relationship extraction, entity recognition, and network representation of risk interactions.
- Chapter 3 – **Proposed Solution**
  This chapter discusses a proposed solution for a risk modeling system. It explains the main challenge in the naive approach to solving triple identification based on entity recognition. It provides a solution via analysis of risk propagation and describes the main solution concepts: Semantic Frame Graphs, Dialog Coherence, and Intermediate Relationship Graphs. It formulates a multi-objective optimization to establish the threshold on dialog coherence scores on the verbalization of the risk-relevant relationship templates.
- Chapter 4 - **Results**
  This chapter presents the results, provides insights into how current LLM is used in validating relations, and explains how the solution can be used in domains other than risk.
- Chapter 5 - **Conclusions**
  This chapter concludes the dissertation and outlines future work.
- Appendix - **Academic achievements**
  This chapter lists publications and conferences at which the results were presented.

1.2 Motivations

1.2.1 Risk Analysis

Critical factors support the continuous improvement of Risk Analysis and Risk Management methods. First, it is necessary due to the growing complexity of new systems, objects, and processes in which risk-related interactions are increasingly complex to find and model. Second, risk management must be an integral part of the overall management process, describing risk interactions in a meaningful and standardized way across system elements to allow for informed decisions on risk-preventing strategies. Third, legal regulations already impose risk management strategies on corporations, i.e., Seveso Directives.

However, few structured data sources collect risk interaction comprehensively, even though collecting, analyzing, and storing data relating to accidents and incidents, given the regulations, is mandatory in some industries. There are just a few examples of official government-managed semi-structured repositories:

- *Nuclear Power Industry.* In this industry, the data collection is rooted in the International Convention on Nuclear Safety. According to this convention, each contracting party commits to taking the appropriate steps to ensure that:

  > «incidents significant to safety are reported in a timely manner by the holder of the relevant license to the regulatory body; [and that] programs to collect and analyze operating experience are established, the results obtained and the conclusions drawn are acted upon and that existing mechanisms are used to share important experience with international bodies and with other operating organizations and regulatory bodies» [2]

- *Aviation.* According to EU directive 2003/42/EC on "Occurrence reporting in civil aviation" data related to all civil aviation incidents and accidents must be collected, reported, and analyzed. The organization European Co-ordination Centre for Accident and Incident Reporting Systems (ECCAIRS) has been established to «assist national and European transport entities in collecting, sharing and analyzing their safety information to improve public transport safety.»
- *Process industries covered by Seveso II Directive.* Companies in Europe that comply with the Seveso II directive must collect and report data in a specified format to the national authorities and the eMARS database.

A key element in managing and preventing a disaster is modeling and understanding the connections between components, hazards, and consequences in the system's domain. There are methodologies to construct such representation focusing on specific risk modeling approaches, namely qualitative and quantitative risk assessment methods.

Current risk analysis methods are based primarily on the decomposition of the structure of system elements (most often down to a graph form) that presents the impact of hazards and the type of applied safeguards. Depending on the degree of mathematical formalism and available information, the model may represent a general qualitative risk assessment based on, for example, the identification of causal links between failure and its effect. The first method developed on this basis in the 1940s was the Failure Mode and Effect Analysis (FMEA) method) [3]. Unfortunately, the quality of the FMEA model is limited by the experience of experts and represents a de facto subjective assessment of the security system. The FMEA method is also very labor-intensive. It requires the identification of all potential events and does not allow for a comprehensive analysis of possible combinations of these hazards [4]. Therefore, the comprehensiveness and objectivity of FMEA are inevitably limited.

A systematic approach and, consequently, one that reduces subjectivity in modeling is the Hazard and Operability (HAZOP) analysis proposed in the 1960s for the chemical industry [5]. The formalism of this analysis revolves around guidewords, e.g., LESS / MORE PRESSURE / TEMPERATURE, based on which a team of experts analyses the consequences of deviation from the structurally established process flow. The identified drawbacks of this method are labor-intensiveness, a descriptive presentation of relationships between the object and hazards, and – despite the proposed formalism of the analysis – excessive reliance on the knowledge of experts. Similarly, as in the case of FMEA, labor intensiveness, and the related costs harm the comprehensiveness of the model [4], [6].

In general, qualitative methods are the beginning of risk analysis, whereas the examples provided only illustrate the basic problems associated with them. Similar problems also exist in quantitative risk assessment models, which aim to estimate and concentrate attention on the most relevant hazards from the perspective of risk [4]. In quantitative models, the problem of identifying the structure of the system hazards and their interactions is compounded by the problem of estimating the probability of an event, e.g., in the Fault Tree Analysis (FTA) method or methods based on Bayesian networks [4]. In other methods, there is a problem with estimating potential losses resulting from safeguard malfunctions, e.g., Event Tree Analysis (ETA) [6].

In reality, a limited model is usually constructed through an iterative, time-consuming process involving subject matter experts with a focus on a selected area of operation of the system [6], [4]. This leads to a situation where most of the risk-related data is written down and stored in various descriptions, either of the failure events, i.e., railway accident report [7], or "near-misses" reports in industrial cases [8] or as descriptions of safety operations of the given system. Such representation restricts the possibility of analyzing the risk interactions comprehensively, as documents need to be read and interpreted by experts.

This defines the first set of requirements for building comprehensive risk representation. Such a system shall be able to consume information written in natural language, normalize it, and store it in a format allowing standardized analysis.

Another critical factor in risk prevention is the subjective nature of risk itself. It is influenced by a broad set of phenomena beyond the mere technical conception of risk as a combination of accident scenarios, probabilities, and adverse outcomes. In fact, excessive reliance on subject matter experts is a risk of the risk analysis methods presented.

The cognition bias of experts performing the risk analysis, regardless of methodology, given the system's complexity and time or budget constraints, is responsible for underestimating or even omitting scenarios that lead to a catastrophe. An excellent example of such accidents would be "Herald of Free Enterprise" or "Jan Heweliusz" - roll-on roll-off car and passenger ferries capsizing. In the first case, missing bow door indicators allowed the ship to depart with the bow doors unlocked. In the second, an inefficient weight-balancing mechanism was inadequate for weather conditions.

Therefore, the second set of requirements is to mitigate subjective risk perception. The system shall be constructed so that a clear template related to risk propagation should be used against the set of documents while searching for risk relations. The role of experts shall be switched from performing complete risk analysis manually to collecting relevant documentation and validating the results of risk detection. The system shall automatically ingest and transform the documents to normalized representation incrementally so the risk representation is augmented when new facts arrive.

The final requirement relates to representing risk interaction in the modeled system. The natural candidate would be the network model of risk interaction. Its advantage is the simplicity of interpretation, which means that the graphical form is understandable also for those not involved in construction. In the case of risk assessment, the undoubted advantage of the network model is the ability to visualize the links between the effects of threats, which is an initial step towards more complex quantitative risk estimation as Bayes Nets [1]. Defining dependencies in the Bayesian network is troublesome as it requires initial decomposition of risk interaction and then estimating conditional risk probabilities [1]. Therefore, developing methods for building network security models may prove to be a foundation for cost-effective ways of security analysis.

1.2.2 Risk - Asset - Vulnerability Dilemma

The contextuality of risk means that the same element can belong to all classes. For example, *engine* is an asset impacted by the *droplet* risk in *fuel*. However, the *airplane* asset is impacted by the *engine* as the airplane's flying capability relies on it. In this case, *engine* is the airplane's Vulnerability. Therefore, *engine* must be assigned two concepts simultaneously.

The current class detection, Named Entity Recognition (NER), in NLP pipelines relies on a text span classification approach, in which both span and the class are detected [9]. The approach is limited in two ways. First, it is limited by the selection of span sizes to capture the interaction, and second, by lack of referring between entities. In the second case, the A-V-T triple would require the classifier to assign multiple classes to the same span, denoting either entity depending on which element is considered an *Asset*. This means that variant entity classification is required in a single context, which is impossible in current NER solutions.

1.2.3 Large Language Models

Late rapid progress on Large Language Models (LLMs) was initiated with the publication of the Transformer architecture [10]. Two elements are behind LLMs' success. First is the attention mechanism, which allows weighted access to the fragments of the context. The other is the model's architecture, which enables easy expansion of the model's parameter space.

With increased computing and training resources, LLMs have demonstrated increased semantic capabilities, making perfect use of both the attention mechanism and its scalable architecture. Since Alan Turing's seminal paper on "Computing Machinery and Intelligence" and his famous Turing Test, we have progressed to the state of the art, which sparks ongoing discussions on threats posed by the uncontrolled growth of capabilities of such models [11]. Such discussions are academic no more, and on 31st October 2023, the UK Prime Minister, Mr. Rishi Sunak, hosted the first global summit on the risks associated with artificial intelligence.

The capabilities of the current general language models, called *foundation models* [12], are, in fact, staggering. However, they are limited in several aspects.

First, the performance is a function of their parameters [13]. A perfect example explaining the improvement with the increase of model parameters is a question answering where the zero-shot setting comes close to the current state-of-the-art performance of the fine-tuned models [14] (Fig. 1.2).

Increased parameter space comes with computing requirements. The scaling is visible if we compare models' training compute power utilization [14] (Fig. 1.3).

Lastly, significant parameter space requires a significant amount of data. OpenAI's GPT-3 model was trained on a filtered Common Crawl dataset, WebText, two internet-based books corpora, and English Wikipedia. All three elements, significant parameter space, compute resources, and abundant training data, enable identifying representations of the knowledge encoded in corpora at scale. Still, the model learns from the data it used; Therefore, OpenAI had to undertake a deliberate training strategy to counter data

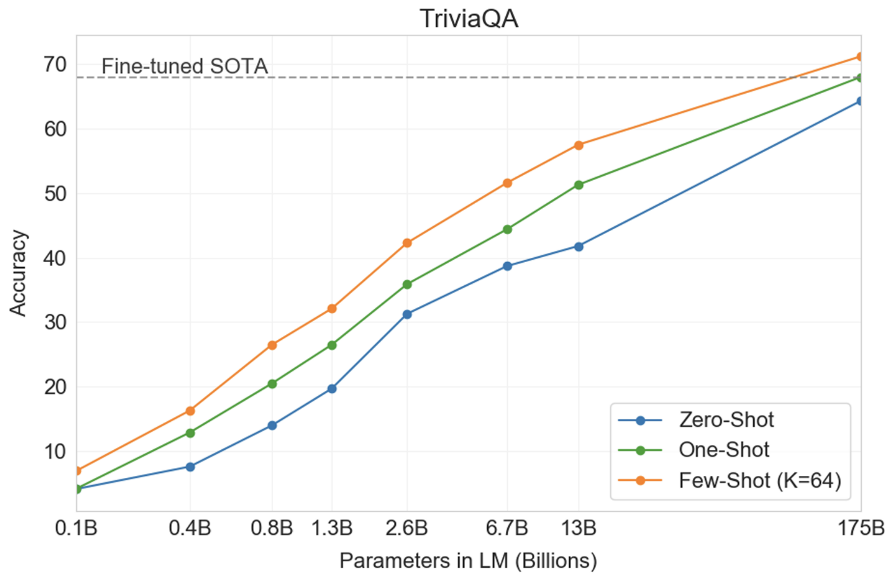


Figure 1.2: Question Answering Performance [14]

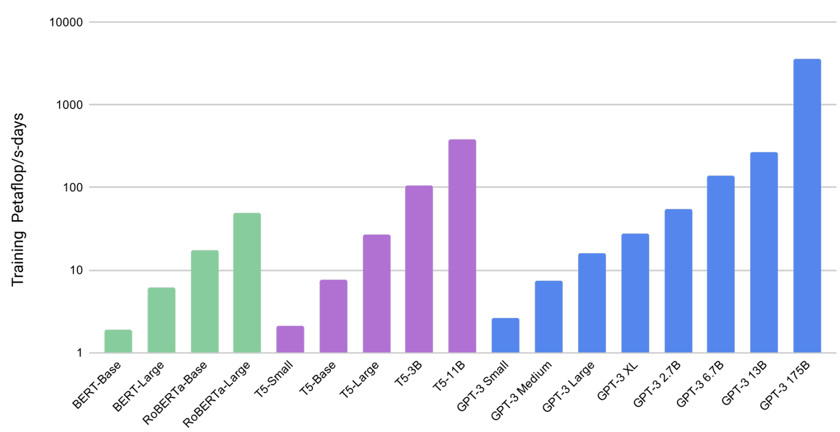


Figure 1.3: Model Training Compute Cost Comparison [14]

contamination [14].

Due to data requirements, only a few areas have enough textual resources to train the dedicated LLMs. In the judicial domain, LLMs encode unstructured textual resources that comprise the legal system. The reason why specific training is required is that the nature of the language in the domain deviates from the 'common' language used daily. A very good example of such a case is the word *consideration*, which in general English means *the act of thinking about something carefully* and in legal terms, it is *to describe the benefit each party to a contract receives.* This is often payment in exchange for goods or services. In the judicial domain, LLMs can perform, among others, the following tasks [15]:

- they can quickly extract key points from legal documents, combine them with the judgment outcomes, and generate concise and accurate case summaries,
- they can generate a draft legal document that complies with legal standards,
- through interactive Q&A sessions with users, they can provide convenient and efficient legal consultation services, while also reducing the workload of professional lawyers,
- they can summarize and extract the key features of a given case, which can contain significant legal documentation.

In the financial area, LLMs are trained in specialized financial textual resources to perform [16]:

- Regulatory Compliance to assists financial institutions in analyzing and interpreting complex regulatory documents, compliance requirements, and legal agreements.
- Investment Research and Due Diligence to help analyze vast amounts of financial reports, company filings, analyst research, and news articles to identify investment opportunities and conduct due diligence on potential investments.
-

. By connecting historical textual data and financial data, LLMs can perform rudiment financial analysis such as:

- Market Analysis and Forecasting in which LLMs analyze historical financial data, market trends, and economic indicators to generate insights and forecasts,
- Fraud Detection in which LLMs can assist in detecting fraudulent activities, suspicious transactions, and potential money laundering activities by analyzing textual data such as transaction records, customer communications, and public records.

Last but not least is the medical domain. It has a portfolio of dedicated language models, including MedPalm and MedPlam2 [17], which encode medical knowledge and have been developed specifically for medical question-answering tasks. However, potential use cases for such models are more advanced and apply to the areas such as:

- Clinical Trial Matching and Recruitment where LLMs can help match patients, based on their medical history, to relevant clinical trials,
- Healthcare Chatbots and Virtual Assistants where the chatbots and virtual assistants can be used to interact with patients to schedule appointments, answer medical questions, provide medication reminders, and offer health coaching,
- Integration of Electronic Health Records (EHR) where LLMs can improve the efficiency and accuracy of analysis of EHR by interpreting and extracting relevant information from unstructured clinical notes, physician narratives, and patient histories.

These models share a common trait: they are significantly founded and have enough good-quality data for training. The risk domain has incomparably smaller data sets for several reasons:

- textual resources are scarce as once an adverse event occurs, corrective measures are taken to prevent it from reoccurring,
- subjects analyzed are not as massive compared to, for example, human health data in medicine, as risk analysis focuses on dedicated areas,
- security descriptions are usually explicitly prepared by subject matter experts (SMEs) during the dedicated analysis tasks performed regularly but rarely, specifically for selected critical infrastructure elements, i.e., power plants.

On the other hand, risk prevention is a more challenging task as, given the current knowledge, we would like to predict possible future risk events to prevent them *before* they actually happen. Hence, taking the medical domain example, given current diseases and how we treat them, we would like to predict future diseases and design their treatment *before* before anybody actually gets sick.

1.2.4 Validation and Explainablility

Explainable artificial intelligence (XAI) is a set of processes and methods that allows human users to comprehend and trust the results and output created by machine learning algorithms. From a risk analysis perspective, explainability is essential as a network representation of risk, which in many cases contains aggregations, must be constructed from descriptions in a verifiable way. For example, sentences: *"Astra Zeneca was first to develop a Covid19 vaccine. COVID-19 was a serious threat to global health in 2020 and 2021"* may generate a relationship between *Astra Zeneca* and *threat to global health* through

Covid19. Therefore, a situation must be reconstructed from the original text to validate the relationship [18]. Explainability is even more critical in the case of distributed evidence, as collecting and linking them from various sources is even more challenging. From the specific NLP case perspective, for example, NLI, explainability means that not only decision: *entailment, contradiction, or neutral* is provided, but also sentences or text spans that justify it [19].

Validation is a method for evaluating the model's performance. In a basic approach, validation is performed through a dedicated data set to validate the classifier. Validation is meant to increase the model's trustworthiness and, in the risk scenario, the trustworthiness of risk-related detection. Validation is usually performed based on evaluating the classifier's performance metrics, such as Accuracy, Precision, Recall, and Sensitivity. Unfortunately, many areas, including risk analysis, lack extensive dedicated training or validation sets, and relying on 'some validation' sets designed for different cases may falsely increase the trustworthiness. The reason for this is that the meanings of words across domains differ. An example of such a case is the usage of the word "boot". It means footwear and a process of restarting electronic equipment. Validating a system on a dataset with the first "boot" meaning can harm understanding the system operating in the electronic equipment domain.

1.2.5 Goal of Reasearch

Natural language processing is a rapidly changing domain nowadays. Although the progress is exceptional, we are still far from the position to have a portfolio of components to *"plug them in"* to build a dedicated risk analysis system. Therefore, several elements are driving the research described in the presented thesis:

- First, as it is financially prohibitive, the research shall answer if it is possible to construct a risk detection system without creating a dedicated LLM for Risk Analysis.
- Second, it shall evaluate available trained and language-specialized classifiers and construct the pipeline to identify risk relations without a dedicated training set.
- Third, a validation method shall be provided.
- Finally, the pipeline shall produce the A-V-H graph, a chosen network model of risk interaction.

Overall, the pipeline shall rely on the Large Language Model's capabilities as much as possible, relaxing the problem of missing training examples and providing means to verify the facts supporting the risk relations identified.

Chapter 2

Related Work

Drawing lessons from historical accidents and proactively identifying and even predicting the potential risks of various hazards improves the safety assurance level of any system. The thesis focuses on two aspects of this endeavor: first, consuming the risk information from natural language resources, and second, representing the risk interactions as a labeled property graph.

The network representation of risk interaction is not a novel approach. The driving reason is to enable comprehensive analysis of unstructured or semi-structured descriptions of various risk-related events, either the actual malfunction or "near-miss" events that did not end with a malfunction but posed a threat in the future. Expert systems are already targeting this task. Although they represent risk interaction in a network manner, integrating various sources together for combined analysis, they differ in framing the problem.

Formal solutions rely on representing the system, its structure, and the connectivity of its components through an ontology. This approach augments existing hazard identification processes, i.e., Failure Mode and Effect Analysis (FMEA), by defining the problem as a reasoning task. The reasoning combines facts collected through the system analysis with concepts and axioms implemented in the ontology.

Ontology is a powerful approach to detecting the impact of events on the system. This approach allows one to evaluate which components are affected by the occurrence of a specific hazard. However, it does not provide a quantitative indication of which components are the most vulnerable, meaning the largest number of hazards impacts them. The Labelled Property Graph (LPG) allows such analysis through network representation of risk interactions. The LPG allows modeling the flow of hazard in the model of the system, allowing graph algorithms, i.e., centrality measures, to indicate which nodes form, for example, hubs. Such hubs are system components that are the most vulnerable as they aggregate the impact of many hazards.

This chapter provides examples of ontology and LPG approaches to risk-interaction detection and discusses how narratives were provided for analysis.

2.1 Ontology in Risk Analysis

The definition of ontology has evolved over time. It started with one proposed by Gruber: "explicit definition of a conceptualization" [20], which emphasizes the notion of conceptualization - a structure of concepts and relations between them that is abstracted away from the real-world objects. In this aspect, conceptualization shall represent the simplified, abstract model of the area of interest. In 1997, Borst defined an ontology as a “formal specification of a shared conceptualization” in which a "shared" feature underlines that the ontology shall be interoperable, forming a backbone of a common interpretation of the area of interest. The assumed formalism of ontology should facilitate the machine interpretability of the ontology itself, allowing automated reasoning. In 1998, Studer provided a combined definition of ontology as a "formal, explicit specification of a shared conceptualization" used today [21]. Ontology in the risk domain serves two purposes:

- to normalize (through shared conceptualization) and integrate risk information represented in various narratives and,
- to automate (as conceptualization is formal) and therefore support the impact analysis itself, as manual validation of the impact is inefficient

2.1.1 Direct Application of Reasoning on Ontologies in Risk Analysis

Application of the ontology in a formal query-based analysis

Processes and procedures can be very complex, and the description of events and their context is difficult to interpret. The interpretation depends on the experience of the human experts involved in safety assessments. To simplify the "general risk ontology" creation, the ontology is constructed specifically around the critical analysis goals called *comptence questions*. The role of the ontology is to put a narrative into a semantic framework around the goals, for example [22]:

- "What are the hazardous events that involve a specific substance and equipment?"
- "What are potential causes of a specific hazardous event, based on the involved substance and equipment (location)?"
- "What are the potential consequences of a specific hazardous event, based on the involved substance and equipment (location)?"

Constructing the ontology directly from the descriptions is difficult [23]. Therefore, it contains the initial, created manually, structure of concepts and relations (a terminology box -

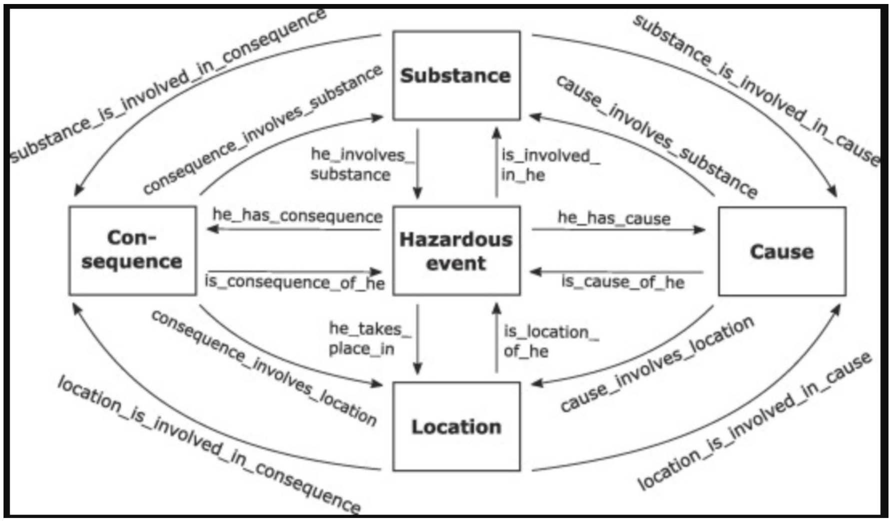


Figure 2.1: Core Concept Structure [22]

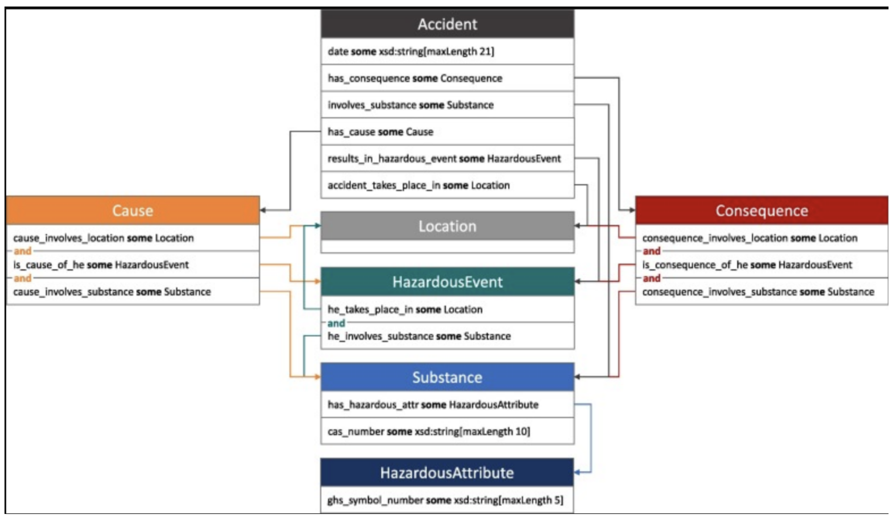


Figure 2.2: Expanded Concept Structure [22]

TBOX) (Fig. 2.1):

- HazardousEvent: potentially harmful event,
- Location: involved equipment, unit or plant component,
- Substance: any involved chemical substance,
- Cause: potential causes that led to the hazardous event,
- Consequence: events resulting from hazardous events.

Core concepts are used to define derived concepts, such as *Accident* (Fig. 2.2). The narrative is processed in a semi-automatic manner (as human validation and correction are needed) to link the extracted terms with specific concepts (Fig. 2.3). The goal is to put all extracted terms in relation to each other to identify cause-effect relationships: cause → hazardous event → consequence.

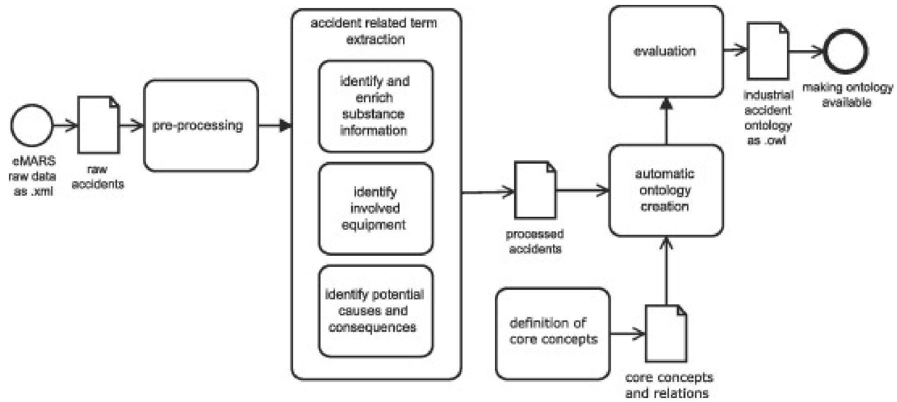


Figure 2.3: Creating the ontology from narratives by expanding core concepts and relations [22]

The cornerstone of the solution is correctly identifying individuals for core and expanded concepts in the text (the assertion of terms). In the first step, a custom pattern-based tag recognition is performed to simplify their expression for seeding concept terms. It is assumed that concepts can be described with one, two, or three words that occur in a certain sequence. For example, for the hazardous event *FIRE*, various expressions of *fire* such as *jet fire*, *pool fire*, or *flash fire* will contain *FIRE* tag. For the cooling equipment, various expressions for cooling like *cooling system* or *cooling jacket* will contain *COOL* tag. The semantic relationship between terms, for example, that specific *hazardous event* took place at the *location* with specific *consequence* is assumed if they co-occur in the pre-defined context. The relationship is then manually validated.

In this way, a set of ontology individuals is created. The risk-related reasoning is performed directly on the constructed ontology with a formal query designed to answer the specific competence question and executed through the HermiT reasoner [22].

Application of Ontology to Integrate Data from Various Documents

One of the challenges in performing the risk assessment is to perform it holistically when the overall task is split into subtasks distributed across teams of experts. Failure Mode and Effect Analysis (FMEA) [24] is a good example of such a situation. From the data acquisition perspective, an established methodology constrains data to be provided in the specific FMEA format (Fig. 2.4). The format specifies that each component is provided with a together with its specific function within the system, its failure mode that defines how it breaks, and an associated failure effect. Failure analysis is performed in separate documents, and human reasoning across separate spreadsheets is laborious. Therefore, there is a need for an ontology to support it.

A hierarchical decomposition of the system into its functional sub-systems precedes the application of FMEA. Therefore, the relationships between components, which are both physical and functional, can be explicitly coded as axioms in an ontology and hence become interpretable by logical reasoner software such as OWL, Pellet, or HermiT. As each component, having multiple failure modes, has relationships to other components, then the effect of a failure at a low level can become a failure cause of a higher level. These effects are propagated through the system hierarchy until the final failure effect is identified.

The constructed ontology is the extension of ISO 15926-14 ontology [25]. It contains a target functional system decomposition (FSO Ontology), specific FMEA ontology, which represents failure effects, failure mode observations ontology (FMO Ontology), and systems and components of a particular asset (ASO Ontology) (Fig. 2.5). The FMEA ontology defines concepts denoting the system's inferred state, for example: *ObjectInFaultState*. The reasoning is performed for a specific individual in a specific state. For example, for the object *"heater"* and the state *"heaters malfunction in the overcurrent"*. The inferred state of the *"heater"* is *ObjectInFaultState* (Fig. 2.6). Assuming the hierarchy of components within the system, the failure propagates across the hierarchy where *ObjectInFaultState* is deducted for *"heater system"* and *"heating, ventilation and cooling system"* (Fig. 2.7).

The FMEA approach is formal. It relies on formally specified ontology and the well-structured input format. Still, the linguistic aspect of the content of the FMEA files itself blocks the general applicability [24]:

- it isn't easy to ensure that terms and relationships are used consistently, particularly when the tables are large,
- the language used is often specific to those involved in a particular FMEA development exercise,
- a spreadsheet typically has no explicit semantics, making it difficult to find, share or reuse the knowledge acquired during the analysis.

2.1.2 Application of Ontology in the Knowledge Graph Construction

The ontology provides a shared understanding of a domain that can be used to structure information and enable interoperability between different systems and applications. The conceptualization is defined explicitly and normalizes risk interaction, which is then analyzed quantitatively through KG representation of risk instead of direct reasoning on the ontology. The ontology is then a model of the data that allows it to be stored coherently in the Knowledge Graph implemented as Labeled Property Graph (LPG).

Ref.	Component	Function	Failure mode	Failure effect
20.1.1	Heaters	To heat up unit	(a) overcurrent	Loss of all heating.
			(b) short circuit	Loss of all heating.
			(c) earth fault	Loss of all heating
20.1.2	Terminal box	Connect supply to heaters	(a) overcurrent	Loss or reduction of heating
			(b) short circuit	Loss of all heating
			(c) cable failure	Loss or reduction of heating

Figure 2.4: FMEA Spreadsheet format [24]

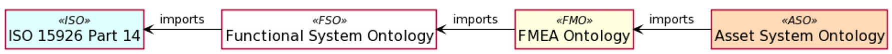


Figure 2.5: FMEA Ontology Hierarchy [24]

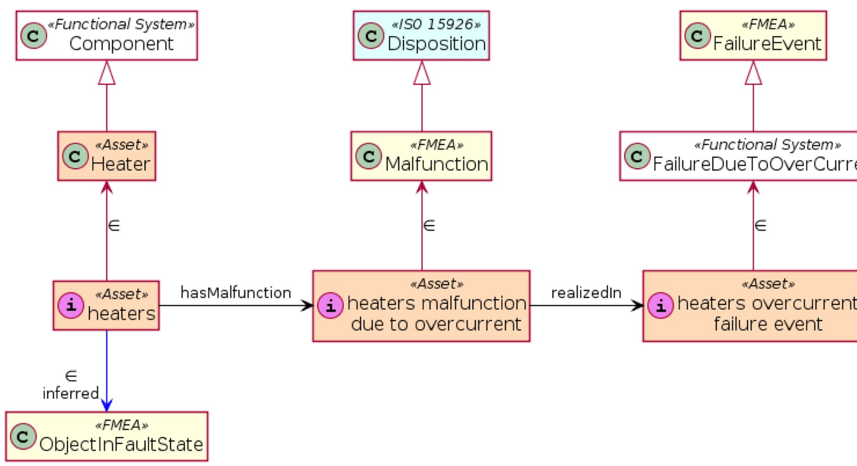


Figure 2.6: Sample Reasoning in the target ontology [24]

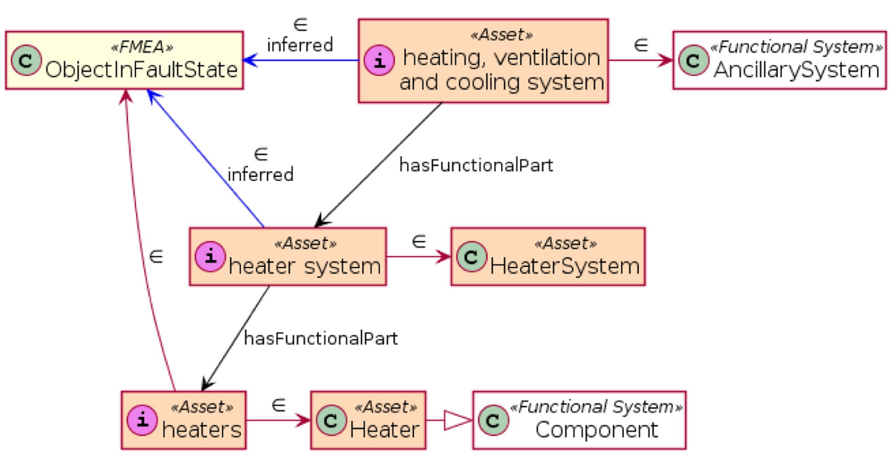


Figure 2.7: FMEA Failure propagation across the hierarchy [24]

Ontology for Railway Accident Analysis

The analysis of railway accident reports in Switzerland faces the problem of multilingual expressions of the same events or concepts (Fig. 2.9). In this scenario, the ontology normalizes the information across German, French, and Italian. Then, at the conceptual level, the risk interactions are represented in the form of a knowledge graph (Fig. 2.9) [26].

The ontology is constructed to normalize risk representation for the following *competence questions* [22]: identify incident reports in any of the source languages that relate to an injury occurring as a result of passengers:

- alighting vehicles,
- falling down stairs,
- boarding vehicles,
- being trapped by closing doors,
- being struck by falling bags.

It defines reports' content and structure (Fig. 2.8). The structure is modeled as the explicit relations between the document, the *Record* node, the sentences *Sentence* node, and the words *Word* node. The *ontology* nodes represent the concept hierarchy connected by a default relation *a_type_of*, for example, *"train" is a_type_of "vehicle"* (Fig. 2.8). The relations between *terms* and *words* are confirmed manually. The analyst connects variant grammatical forms of *words* representing the same *term* through relation *forms*. The relationship *next* indicates the next word in the sentence used to identify bigram concepts. Unigram and bigram terms are identified through a measure of importance (Tf_Idf) of terms in the corpus (Eq. 2.1). Additionally, the analyst defines the hierarchy concept-term (Tab. 2.1).

The proposed approach allowed the analysis of incident reports to collectively identify the safety-related statistics around *competence questions* (Fig. 2.10). Although the proposed text processing is a rudiment dictionary of railway safety-specific terms in the three languages used in reporting, the benefits of the proposed approach are:

- significantly simplified query design executed on the KB instead of ontology,
- quantitative analysis of cases of injuries.

$$TF\_IDF = \frac{f_T}{N} ln(\frac{R}{R_T}) \tag{2.1}$$

where:

- f_T is the frequency of occurrence of term T within a record,
- N is the total number of terms in the corpus,

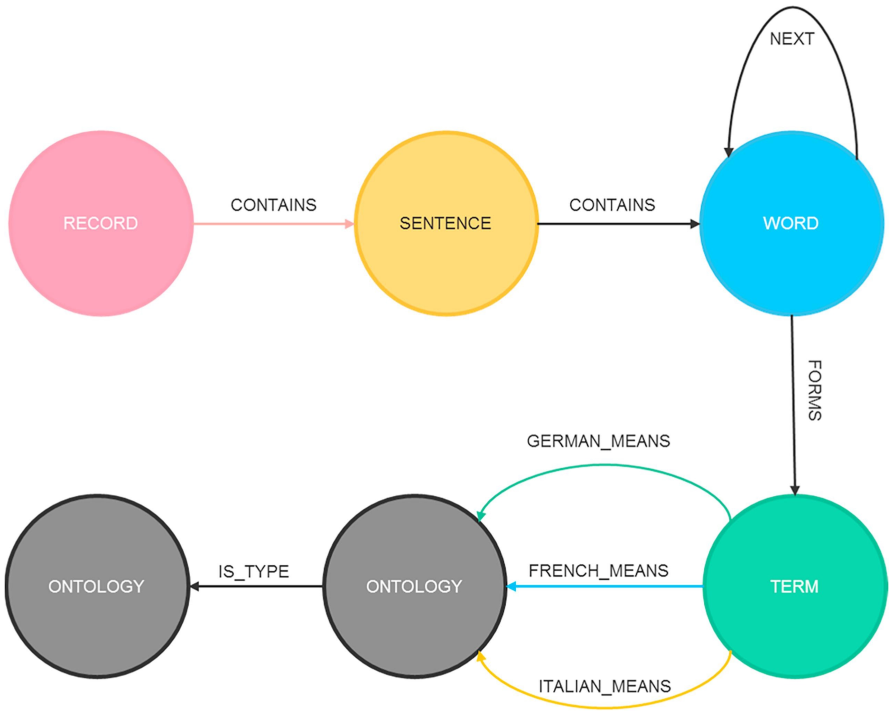


Figure 2.8: Multilingual Ontology [26]

- R is the number of records in the corpus,
- R_T is the number of records that contain the term T.

Ontology and Knowledge Graph Construction for Near-Misses

Descriptions of "near-misses" provide detailed information on complex hazardous situations, including their initialization, evolution in the system, and barrier effectiveness. This data is critical from a safety analysis perspective, as barriers may eventually fail in similar situations.

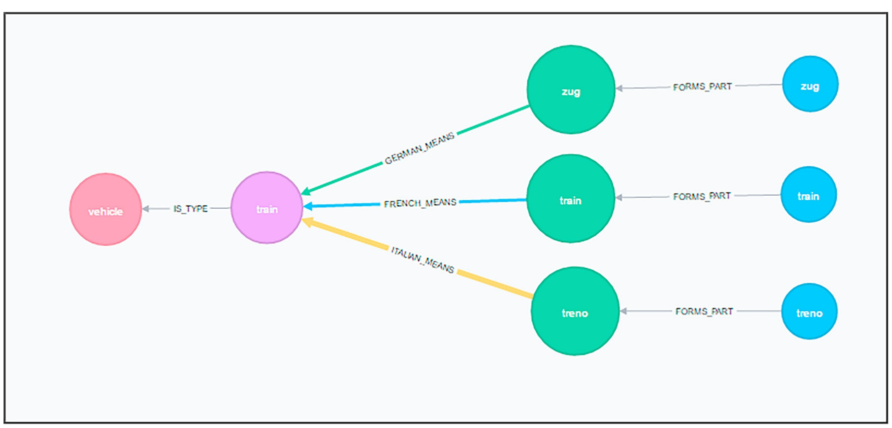


Figure 2.9: Multilinqual Concept Expression [26]

Concept	**Term**
vehicle	carriage, vehicle, ambulance, tram, train, bus
person	doctor, self, customer, person, driver, passenger, months old, years old, baby, young, old, female, male
object	Bag, alcohol, drugs, stairs, footboard, customer information system, ticket, door

Table 2.1: Concept - Term mapping [26]

Query	**Language**		
	German	**French**	**Italian**
1. Alighting	693	296	34
2. Falling down stairs	73	9	1
3. Boarding	1100	349	16
4. Closing doors	1220	464	409
5. Falling bags	3	1	0

Figure 2.10: Multi-linqual railway incident incident results [26]

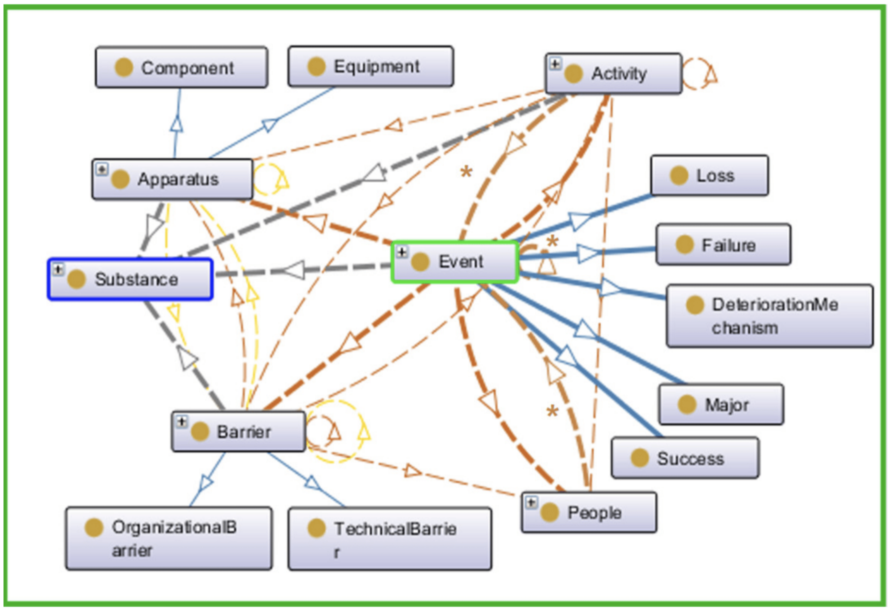


Figure 2.11: EsOpAI Ontology Structure [27]

To properly extract risk interactions from narratives, the use case for ontology is to explicitly model the event's evolution in the system and the effectiveness of the barriers [8]. The goal of the KG representation is to answer the following competence questions:

- "Which are the more vulnerable apparatus in the system (e.g., refinery)?
- "How do they usually fail?"
- "Which are the consequences of such failures?".
- "Which is the most critical barrier preventing the catastrophe?".

The ontology (Fig. 2.11) is constructed using concepts defined in EsOpAI (Operational Experiences via Artificial Intelligence) [27] to represent a near-miss. It contains the following entities: EVENT, APPARATUS, SUBSTANCE, ACTIVITY, BARRIER, and PEOPLE. The concepts are further specified to represent more granular concepts. For example, EVENT is split into:

- LOSS related to losses of containment (e.g., leakage, overfilling), FAILURE, which contains both failure and damages,
- DETERIORATION that includes all the mechanisms of deterioration (e.g., corrosion, pitting, creep) that caused integrity problems,
- MAJOR depicting all those events which have the potential to generate other incidents (e.g., fire)
- SUCCESS to indicate the positive barrier action that contributes to interrupting the incident escalation

The APPARATUS entity is divided into EQUIPMENT, which indicates a whole, and COM-

Relation	Head Concept	Tail Concept
causes	EVENT, ACTIVITY, PEOPLE	EVENT
involves	EVENT, ACTIVITY, APPARATUS, BARRIER	SUBSTANCE
part_of	APPARATUS, BARRIER	APPARATUS, BARRIER
related_to	EVENT, BARRIER, ACTIVITY	ACTIVITY, PEOPLE, APPARATUS, BARRIER

Table 2.2: EsOpAI Relations [27]

PONENT, which indicates a part of the EQUIPMENT connected with a PART_OF relation. EsOpAI has been designed with four relationships:

- RELATED_TO, which is a generic relation between two entities,
- PART_OF describing a physical connection between two entities,
- INVOLVES, that relates an entity to a substance,
- CAUSES, which states a causal connection between two entities

The relations are predefined and established between the entities (Tab 2.2). Supervised learning is used to detect concepts and relations in the narrative. The annotation strategy strictly follows the structure of the ontology, and the entities and relations are annotated exactly to the intended concepts and relations (Fig. 2.12). Finally, detected triples i.e.: *head_concept - relation - tail_concept* are extracted directly into the KG. The structure of the KG is aligned with the ontology as the triplets loaded are detected according to the ontology definition of entities and relations between them. The KG allows for "ontological explorative analysis" [8], a combined network analysis of numerous near-misses across installations and plants. It allows quantitative answers to the competence questions, for example, "most frequent causes of failures upon the most frequent apparatus that fail in refineries" (Fig. 2.13).

2.2 Knowledge Graphs and Network Representation of Risk

The network representation of risk enables the topological analysis of its interactions: degrees, shortest paths, and centralities [1]. For example, the betweenness centrality measure will indicate the degree to which the node transmits the impact of the hazard to other

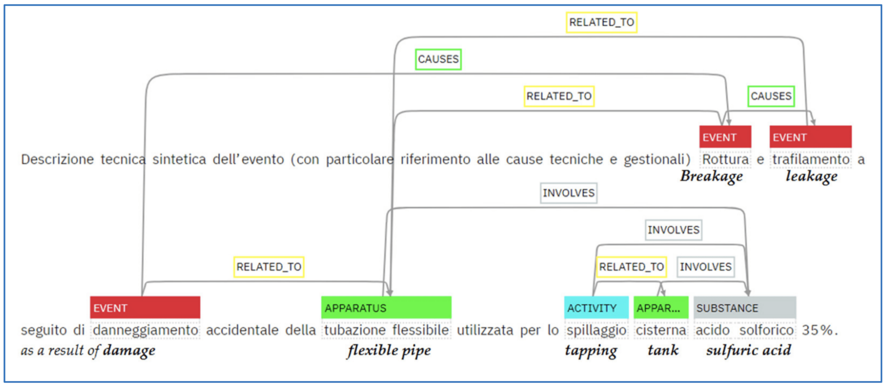


Figure 2.12: EsOpAI Annotation Example [27]

AE	f_{AE}	$f^{\%}_{AE}$
FLANGE LOSS	7	3.30%
PIPELINE LOSS	7	3.30%
GASKET BREAKAGE	6	2.90%
VALVE LOSS	4	1.90%
PIPE DAMAGE	3	1.40%

Figure 2.13: Top five apparatus failures (AE) causes in refineries in Italy [8]

nodes. Identifying such nodes in the model would help focus the protection measures around that specific system element. In general, network representation of risks will allow [1]:

- to characterize the risk-related impact of each system component, represented as a node, in the system's overall structure,
- to simulate the impact of the system's structural changes from the risk control perspective. For example, the analysis of the impact of change in strictly law-regulated IT infrastructure due to the statutory or standards change of the selected component.
- to build a simulator of real situations, which are not repetitive (crisis situations). The simulation can occur under varying conditions of risk.

There are several examples of application of such risk representation strategy. They differ in the method for normalization of risk relations between the intended concepts. These methods do not rely on prior ontology definitions. The assumed KG structure defines the Knowledge Acquisition Pipeline, which consumes textual information directly and transforms it into a required representation. Ingestion and transformation of textual data are complex tasks. Entities and relations between them are key elements to detect.

2.2.1 Representing HAZOP Safety Reports as Knowledge Graphs

Significant safety knowledge is encoded in textual sources, for example, formalized safety reports, such as the HAZOP, near-misses or FMEA. However, this knowledge remains undiscovered because it is not transformed into a representation suitable for comprehensive analysis. The Knowledge Graph (KG) representation allows such analysis. The structure of the KG depends on the modeling goals, namely, which aspect of the narrative is useful and shall be represented. For example, the risk interaction network (Fig. 2.14) consists of four node types: suggestion, result, cause, and equipment/assets. The goal of the representation is *evaluate and combine* hazard causes, effects, and prevention measures and, therefore, validate the gaps in the safety design. Apart from that, in this specific case, the graph allows employees to picture and understand the safety and the operational requirements for the process [28].

The acquisition pipeline assumes a specific input data format, such as HAZOP. HAZOP, which a semi-structured analysis methodology, uses a defined set of guidewords to help evaluate the consequences of deviation from the regular process flow [4] simplifying entity detection [28]. In this case, processing is performed as follows. There are three application layers. The conceptual layer encodes knowledge of the processes used in the extraction performed in the extraction layer. It aims to normalize HAZOP reports to combine information from different processes and represent industrial safety knowledge. The goal of the layer is to define the node types (concepts) used in the storage layer, like cause, result, equipment, or suggestion. For example, a cause entity could be: "coil blockage", result: "crude oil in the vaporizer will boil" and the recommended entity: "clean the coils". The extraction layer extracts all entities and relationships among them. In short, it extracts the knowledge triple: subject, relationship, and object and saves them into a storage layer for reporting. The storage layer contains the network representation of risk (Fig. 2.14).

2.2.2 Representing Free Text Risk-Related Narratives as Knowledge Graphs

Reusing and releasing the value of industrial safety knowledge is a step towards a comprehensive risk repository. The idea of representing risk interaction as a heterogeneous graph, where the node type represents a concept of risk, cause, or asset, increases the clarity of representation. However, the approach proposed for the HAZOP analysis is limited because it assumes a specific input format; hence, the conceptual layer is aligned.

It is possible to relax the constraint of a specific input format. A solution representing

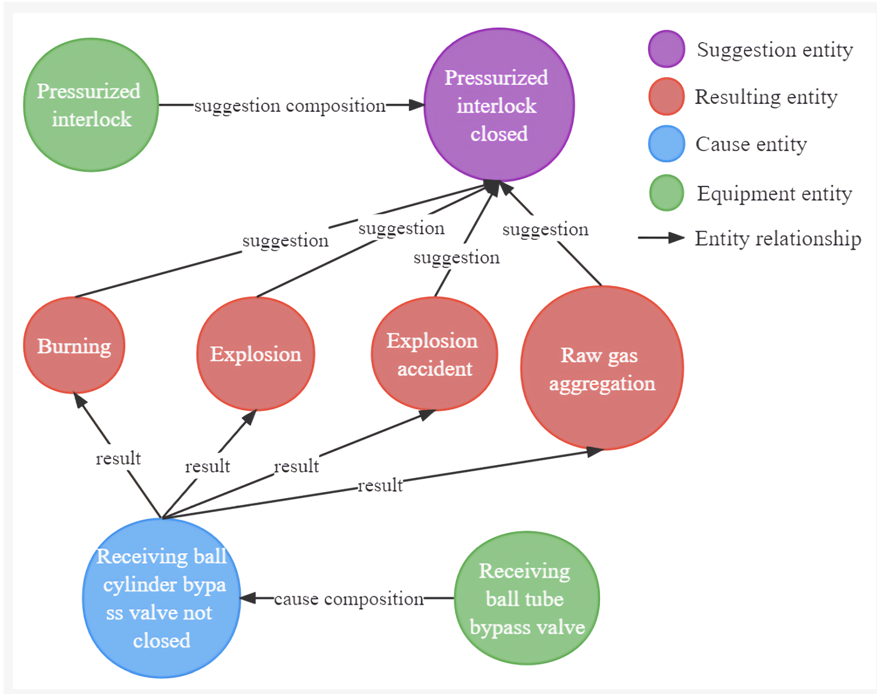


Figure 2.14: Exemplary Storage Layer Content for Hazop Analysis [28]

risk relations for railway safety parses natural text description of British rail accidents [7]. The processing is step-wise (called the knowledge extraction steps), where each step is responsible for detecting, linking, normalizing representation across documents and mentions (knowledge fusion step), and eventually representing risk interactions in the form of the knowledge graph, namely RKGRS (Fig. 2.15).

The resulting structure represents risk interactions as a heterogeneous graph with defined node types representing different concepts. For example, C nodes indicate Cause, D - danger, and K consequences. For example, C01 (imperfect management of maintenance practices) will result in danger nodes: D01: struck by object or collision, D07: derailment. Consequences are K03: damage to structure, component, or device, K01: injuries (Fig. 2.16).

The ability to parse free text reports is a step in the right direction, as the input format shall not limit the comprehensiveness of the risk repository. However, relaxing the format results in a complicated knowledge acquisition pipeline. In this case, the ensemble of classifiers is trained in a supervised manner [7], and the training set was designed specifically for the railway safety scenario. The training uses **seventeen** text augmentation algorithms to enrich the representation (through embeddings) or disturb the text insignificantly by adding or replacing characters or synonyms to achieve acceptable training results. To apply the proposed approach to other domains, the training set and text augmentations shall be

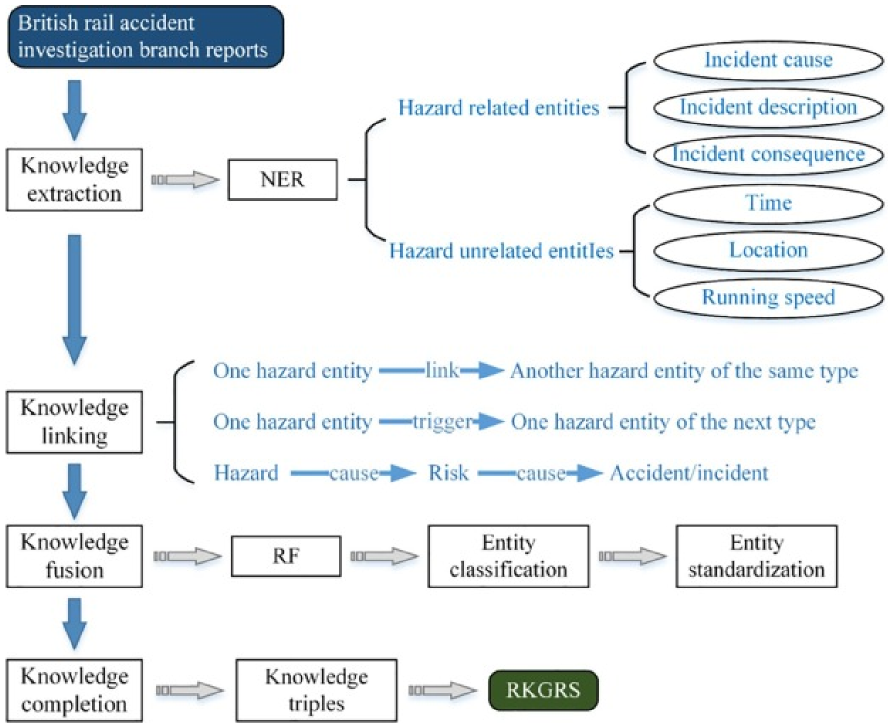


Figure 2.15: Railway Accident Reports Processing Pipeline [7]

adjusted [7].

2.3 Summary

The methodology for integrating and representing risk interactions is not established yet. In fact, more solutions are specific and target defined areas, exactly as the examples provided. The limited applicability of current solutions does not come from the fact that creating a comprehensive risk ontology is potentially impossible, and we have to focus on a well-defined *competence question* only. The ontology is a *shared* conceptualization, and it is interactive and updated in an iterative manner.

It is the flexibility of human language in risk description that is a limiting factor. This flexibility translates into a general lack of comprehensive training sets, and domain-specific solutions are required to prepare their training and validation sets. There are approaches to address the training set issue by providing additional constraints on data, i.e. hierarchy [29]. In this approach, the hierarchy structures the training as the training example for an element lower in the hierarchy will be used to train higher-level concepts. Still, the training set is required.

Therefore, it is required to focus on the linguistic aspect of risk modeling, which will

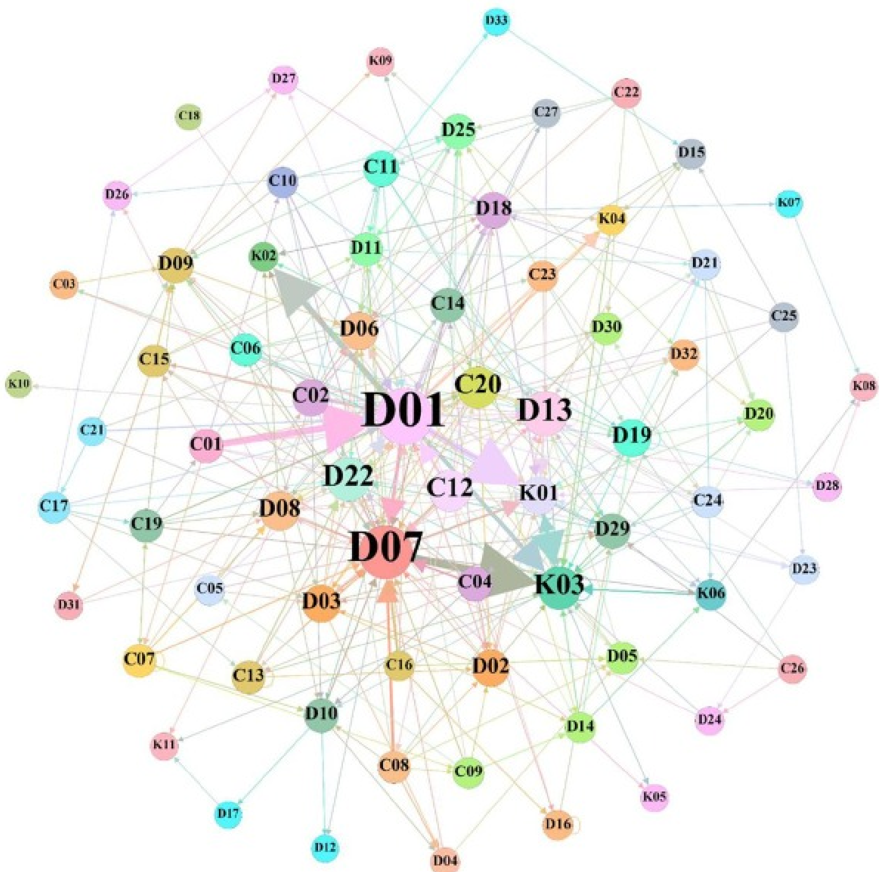


Figure 2.16: Railway Safety Knowledge Graph [7]

then become a foundation for an application that will comprehensively represent risk interaction.

Chapter 3

Relevant Natural Language Processing Techniques

The examples described in the previous chapter relied on two key natural language processing elements. First, Entity Recognition was responsible for detecting a concept in the narrative. Second, Relationship Extraction aimed to confirm risk-related semantic relations between them in the narrative. This section describes NLP technics relevant to the proposed solution and discusses their limitations.

3.1 Entity Recognition

Entity Recognition, or Named Entity Recognition (NER), is an NLP task that aims to assign a text span to one of the pre-defined semantic categories like 'PERSON', 'COUNTRY', 'ORGANIZATION' [30]. NER is useful in applications requiring such classification. In KG construction, NER assigns the concept and classifies nodes. For example, to represent railway risk interaction, the knowledge acquisition pipeline is responsible for identifying Danger, Consequence, and Cause categories. However, detecting risk-related categories is difficult to obtain using current NER approaches.

Sequence labeling and span recognition are the standard approaches to solving the NER task. The sequencing and span detection relies on BIO (and BILOU being the variant) tagging [31]. This method allows the treatment of NER like a word-by-word sequence labeling task via tags that capture both the boundary and the entity type. In the BIO tagging, B indicates the beginning of the sequence I - inside, O-outside. For example, a sentence *Joseph Robinette Biden is the 48th President of the United States of America.*, yields the sequence of BIO tags related to entities *PERSON* and *LOCATION* as *B-PER, I-PER, I-PER, O, O, O, O, O, O, O B-LOC, I-LOC, I-LOC, I-LOC, I-LOC* .

Formally, the NER classification is the sequential prediction problem to estimate the probabilities of predicting the ith BIO tag given the *context* being the history of k, the future of l words, and the history of m past BIO tags:

$$P(y_i|x_{(i-k)},...,x_{(i+l)},y_{(i-m)},...,y_{(i-1)})$$

where k,l, and m are small numbers. Algorithms used to estimate the probabilities of the tags are:

- Conditional Random Fields (CRFs): CRFs model the conditional probability of label sequences given input features, capturing dependencies between neighboring labels,
- Bidirectional LSTM (BiLSTM): BiLSTM networks are recurrent neural networks that can capture sequential dependencies in the input data,
- Transformer-Based Models: Transformer architectures such as BERT [32], GPT [14], and their variants which have shown state-of-the-art performance in NER by leveraging self-attention mechanisms to capture contextual information

However, to design an efficient NER detection system, the following design questions need to be answered [33]:

- How to model non-local dependencies,
- How to use external knowledge resources or construct a training set.

From the RA perspective, both pose significant difficulties.

The non-local dependencies mean that the classification of the text span depends on the selection of the context and changes with the context size. The local dependency means that syntactic and semantic features of the entity must be present in the current context of k past l future words and m past tags. Such a definition of context is fixed across the narrative to perform the classification. For example, in the sentence *"Joseph Robinette Biden is the 48th President of the United States of America."* term *America* is a part of the *LOCATION*, and all features required to classify *America* this way are present in the context. Likewise, in the sentence *"Brian Moynihan is the President of the Bank of America"*, the term *America* is a part *CORPORATION* local features of the same context definition allows classifying the terms correctly though the different class is assigned.

The fixed, local context assumption harms the classification in the risk-analysis domain. In the railway case [7], in order to classify a text span to a "Danger" category, the context shall contain features expressing some harmful effect of it, as the token "Bank" differentiates the classification of *America* in previous examples. In general, such features may not follow the locality assumption as effect, e.g., injuries and damage to the infrastructure, can follow the "derailment" danger in an arbitrary long context.

Another difficulty in a direct application of the Entity Recognition approach is the *contextuality* of risk-related concepts. The *contextuality* means that the classification of the text span depends on the classification of other text spans in the context. Formally, current tag classification depends on the fixed history of m tags. Depending on the intended focus, the same text span can have multiple classifications in the same context. For example, in the sentence *"A spark ignited a container supplying fuel to the engine"* depending on the focus, the *container* is a "Danger" as it can be ignited by the spark potentially destroying the engine. Simoultenousely, it is a "Component" as, without it, proper functioning of the engine is impossible as it is responsible for fuel supply.

Last but not least, supervised learning is the dominant approach to solving the classification task. In the risk analysis domain, standardized training sets are not available. Preparing a training set means solving the entity recognition task manually in many, especially rare, extraordinary malfunction cases.

3.2 Relationship Extraction

The efficient, algorithmic Relationship Extraction started with introducing the Hearts Patterns. Hearst proposed to detect hyponymy, namely "is-a", relations, e.g., "rose is a flower". The approach used a set of regular expressions on the syntactical decomposition of the sentence. It was motivated by two goals:

- to avoid the need to pre-encode the extensive knowledge and
- to apply the same approach across a wide range of text.

[34]. The key assumption was that the syntactical patterns are enough to define relations within a sentence effectively. It quickly turned out that the flexibility of human language was underestimated, and the pattern approach proved ineffective. The Hearst Patterns, however, initiated intensive research in the RE area. This section will evaluate the applicability of existing RE methods in the context of scarce resources in the Risk analysis domain.

3.2.1 Intra-Sentence Relationship Extraction

Sentences are composed of smaller linguistic units, such as words and phrases, and the meaning of a sentence is determined by the meanings of its constituent parts and the way they are combined. Sentences convey semantic relationships between their parts, words, and phrases that can be detected and identified. Their syntactical decomposition provides additional features: parts of speech, grammatical relations, or named entities within the sentence (Fig. 3.2) [35]. These features were used extensively in the initial approaches to the RE.

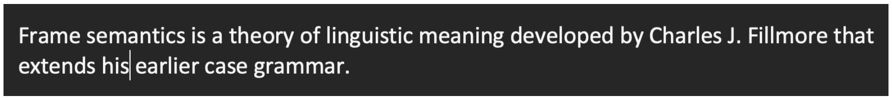


Figure 3.1: Exemplary Sentence

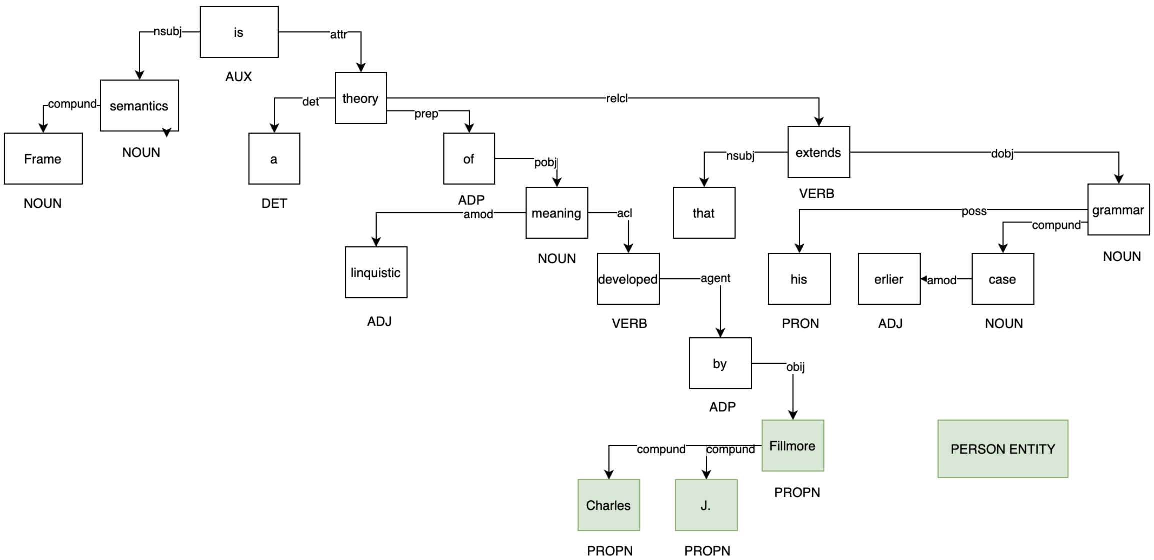


Figure 3.2: A sample sentence decomposition to a feature-rich dependency tree

Rapid progress in the RE started after launching a dedicated relationship extraction task between pairs of nominal within a single sentence [36]. Initial solutions relied on heavily syntactic features and formulated the problem as kernel-based classification of the shortest path in the sentence dependency tree [37], maximum entropy models over syntactic features in the sentence [38] or graphical models [39]. Currently, two approaches, a transformer-based [40], and dependency decomposition, graph-based [41] trained in a supervised manner, are currently the best models.

The main drawback of the state-of-the-art approaches that limit their direct applicability in the Risk Analysis domain comes directly from the training strategy. A supervised approach, although very effective [40], [41], does not guarantee the same performance on the vocabulary that is out of domain [19]. It is hard to imagine a scenario in which each new failure event is provided with a dedicated training set that would contain annotated elements. This would mean we solve RE classification tasks manually each time we analyze risk interactions for a specific case.

Relaxing the supervised training strategy and relying on an unsupervised approach is tempting. However, the unsupervised approach falls short of applicability to risk analysis due to large corpus requirements or reliance on auxiliary classifications, such as entity

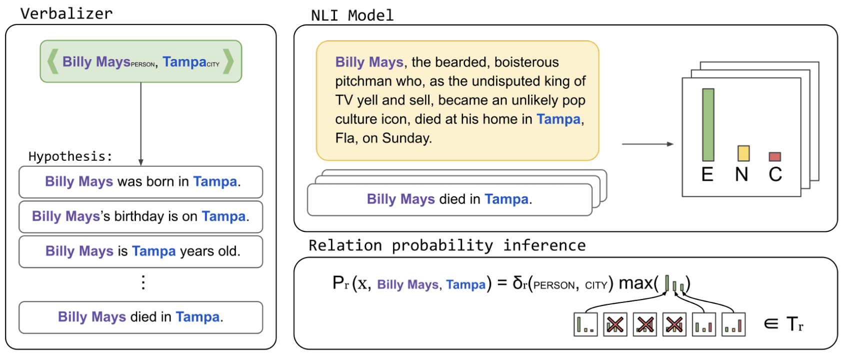


Figure 3.3: Verbalizer Approach to RE [45]

recognition, to improve clustering results [42].

An approach that relaxes the training set and the corpus requirements uses the direct application of textual entailment [43], [44] in relation detection. In this approach, a relation classification task is cast as an inference of the hypothesis that the sentence, being a premise, entails the relation pattern of interest [45]. The approach assumes that the subject and the object of the relation are in the sentence, and relationship templates are augmented to validate them against the premise. A template with the highest entailment score is selected (Fig. 3.3) [45].

Although relaxing the most critical constraints on risk analysis, namely training set and corpus, the verbalizer is limited in the assumption that the sentence defines a relationship and the patterns are evaluated for sentences only. In a general corpus, almost 40% of relations are defined across sentences [46]. Unfortunately, the distributed nature of the relations in the risk domain is prevalent as information on hazards is stored across various documents. Moreover, the documents need to be integrated for a comprehensive analysis [28], [7].

3.2.2 Inter-Sentence Relationship Extraction

Inter-sentence relationship extraction is an approach that focuses on identifying relations across the whole document. Identifying relations in an inter-sentence scenario requires modeling semantic interactions between mentions of entities in different sentences. The direction of the relation may not follow a "reading direction." For example, it is possible, given the exemplary text (Fig. 3.4.), to verify that *"droplet"* is in a relationship with the *"engine"*. In this case, the analysis may follow the deduction path: *"droplet"* → *"supercooled water droplets collide with a surface"* → *"if supercooled water droplets collide*

Airplane uses engines for flying.
ATF is a type of aviation fuel designed for use in aircraft powered gas-turbine engines.
If these supercooled droplets collide with a surface they can freeze and may result in blocked fuel inlet pipes.

Figure 3.4: Inter-sentence Relationship Example

with a surface they may result in blocked fuel inlet pipes" → "blocked fuel inlet pipes" → "aviation fuel designed for use in aircraft powered by gas turbine engines" → "engine"

Document-level relationship extraction will be reviewed for two types of models: sequence-based and graph-based.

Sequence-based models

In the sequence-based approaches, the assumption is that most cross-sentence relations are fully defined within the fixed context anyway. The statistics around the general distribution of head and tail relation entities is that they are mostly separated by three sentences in the narrative. Therefore, the RE detection is cast as selecting a combination of at most three sentences from the document. These sentences would form the relation context. The context must include references to both the head and tail entities, and it is used to classify the relationship. The strategies for selecting the sentences are as follows. Sentences may form a "Consecutive Path" (Fig. (3.5) following each other in the "reading order." Another type of context supports multi-hop reasoning, meaning each sentence is connected by a common entity. In this context, sentences may not follow the "reading order"; however, still, their number is limited to three; (Fig. (3.5). The third strategy defines the context as pairs of sentences containing the first, the head entity, and the second, the tail. The set of contexts is the cartesian product of sentences containing the entities (Fig. (3.5). The contexts constructed are used to train a discriminative classifier based on the prepared training set [47].

Another crucial assumption in sequence-based models is to separate local and global relations contexts. The separation of contexts models interactions document-wise and supports cross-sentence reasoning, Mentioned-based Reasoning Network (MRN) [48]. Local contexts capturing close subject-object relations are modeled through the convolution of entity mentions. Each convolution combines a subject and object mention spans; therefore, several convolution layers are used to model several spans in the document. A document-wise representation of a relation, a global context, is achieved through a co-attention of local contexts. Co-attention performs a weighted combination of several local contexts

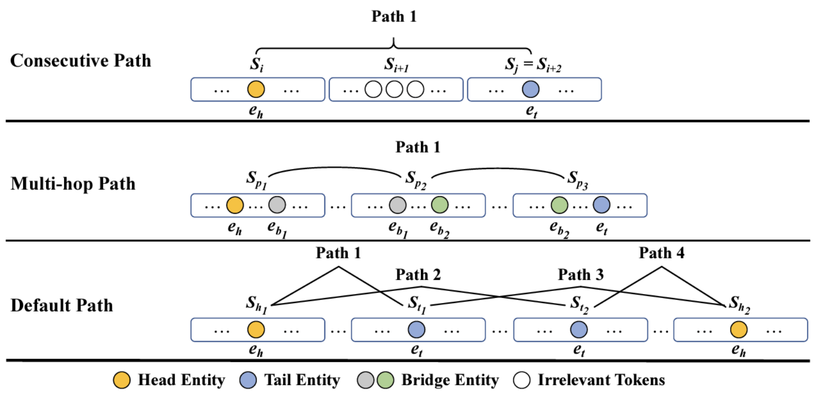


Figure 3.5: Types of contexts in sequence-based document level RE [47]

(local relations) to achieve its global representation. The MRN is a discriminative classifier trained on a dedicated document-level training set.

Graph-based models

A graph-based approach has gained significant attention lately and is considered the most effective approach to document-level RE. It is based on a network representation of the document [49]. The representation captures semantic, syntactic, and positional information on the entities, providing more features for relationship classification. The general processing pipeline involves splitting documents into sentences and **detecting the entities** that will be evaluated for relationships in the document. The document-level relationship classification tasks are cast as a link prediction problem between the nodes representing the entities of interest.

There is a significant number of models solving document-level RE tasks varying in implementation details. For example, the Edge-oriented Graph (EoG), which is the extension of an earlier model for solving sentence level RE [50], represents the document as a heterogenous, undirected graph containing the following node types (Fig. 3.6):

- Entity node represents concepts.
- Mention node is a span describing the entity.
- Sentence node is a sentence that holds the mentions.

Each node representation is computed as the average of the embedding of its constituent words. The embeddings of words are computed in the Sentence Encoding Layer using the BI-LSTM network (Fig. 3.7). The edges represent different roles of the node in the document:

- Entity-Mention (EM) edge indicates that the mention describes an entity,
- Mention-Sentence (SS) edge indicates that specific mention is a part of the sentence,

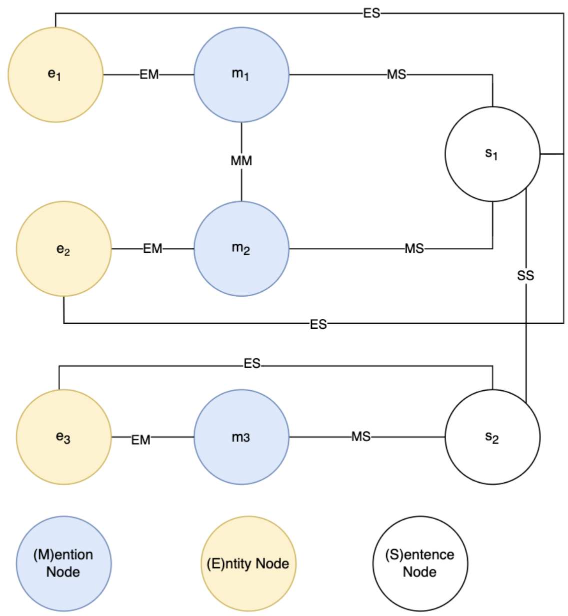


Figure 3.6: Edge-Oriented Graph Structure [50]

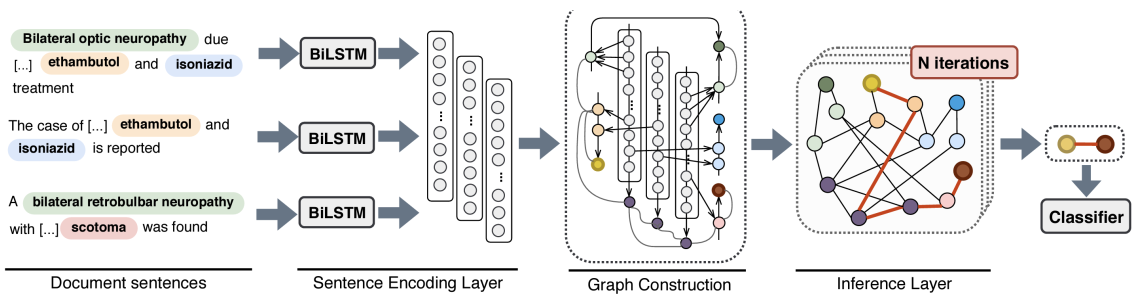


Figure 3.7: EoG Architecture [50]

- Entity-Sentence (ES) edge indicates that the entity is contained in the sentence
- Sentence-Sentence (SS) edge connects the sentence to model non-local dependencies. The edge specifies how many other sentences (the distance) separate the sentences in the document.
- Mention-Mention (MM) edge connects mentions that are a part of the same sentence.

. The edge representation is calculated as the concatenation of the representation of connected nodes. The link prediction task is cast as the classification of the path that links entity nodes in the network. The representation of the path is calculated as the non-linear transformation of embeddings of edges in the path [50].

In the Global-To-Local Neural Network for Relationship Extraction (GLRE) [51], the document structure is also explicitly represented as a network (Fig. 3.8). The structure of the graph remains similar to EoG's with only slight modification in that all sentences are

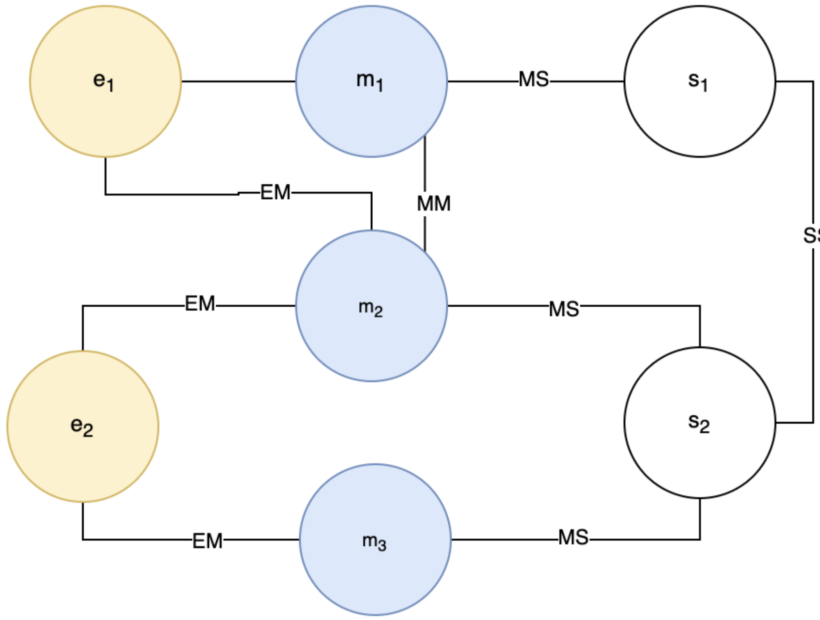


Figure 3.8: GLRE Document Representation and Classification Schema

connected without addition *distance* information:

- Entity-Mention (EM) edge indicates that the mention describes an entity,
- Mention-Sentence (SS) edge indicates that specific mention is a part of the sentence,
- Entity-Sentence (ES) edge indicates that the entity is contained in the sentence,
- Sentence-Sentence (SS) edge‘ connects all sentences regardless of their distance,
- Mention-Mention (MM) edge connects mentions that are a part of the same sentence.

The representation of nodes is an average of the embeddings of its constituent words. Words embeddings are calculated during text transformation, and it is the embedding calculated through the BERT transformer of the sentence or short text fragment for better contextuality [32] [51]. There is no representation associated with the edges. The relation detection problem is cast as a link prediction problem between entity nodes.

There are two representations of entities: global and local. Each entity can have only one global representation that is convoluted with the network structure of the document [52]. The global representation of head and tail entities results from the graph convolution of their network neighborhood [52]. Local head and tail embeddings are weighted combinations of local and global embeddings of their mentions (Fig. 3.8). The dedicated multi-head attention modules (Multi-head Attention 0 for the head and multi-head Attention 1 for the tail) select appropriate mentions. The final link prediction task between head and tail uses concatenated embeddings of both in the logistic regression task. Contrary to the previous

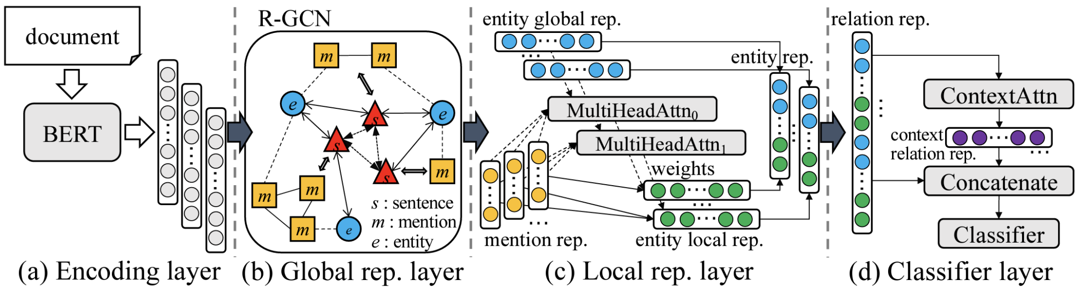


Figure 3.9: Global-to-Local Neural Network for RE Architecture [51]

examples, The GLRE approach explicitly separates the document representation, which drives the global representation of entities, from the cross-influence of local mentions, which is mediated by regular multi-head attention [51].

The efficacy of current document-level extraction is limited specifically in two ways:

- first, it is supervised and may not be generalized onto documents semantically far from the training sets.
- second, the approach relies on the pre-processing which detects *entities* and *"entities span*. Quality of the detection is essential for the overall performance [53].

3.2.3 Transformer-based Relationship Extraction

In LM domain, the RE classification task can be cast as another NLP task. For example, the verbalizer casts RE as an entailment task relying on the generalization capabilities of LM in entailment recognition [45]. The other one, but chronologically the first, frames RE classification as a question-answering (QA) task [54]. This is the first approach to a zero-shot classification scenario in which a classifier is used for cases not presented in the training phase. The leading assumption of the approach is that question formulation generalizes the structure of the relation. Therefore, training on question-answer pairs instead of relation examples is more efficient. In the direct RE training, the classifier is responsible for identifying linguistic elements defining the relation abstracting away from combinations of subject and object [40]. In the QA scenario, the classifier is trained on an extensive QA dataset instead of an RE dataset [54].

The RE classification problem is cast as a parameterized, relation-specific question, and the classifier identifies span as an answer. An answer is an object of the relation (Fig. 3.10). The relation-specific question's parameterization must be specified (Tab. 3.1). For example, given the sentence *Brian Moynihan is the President of the Bank of America*, the

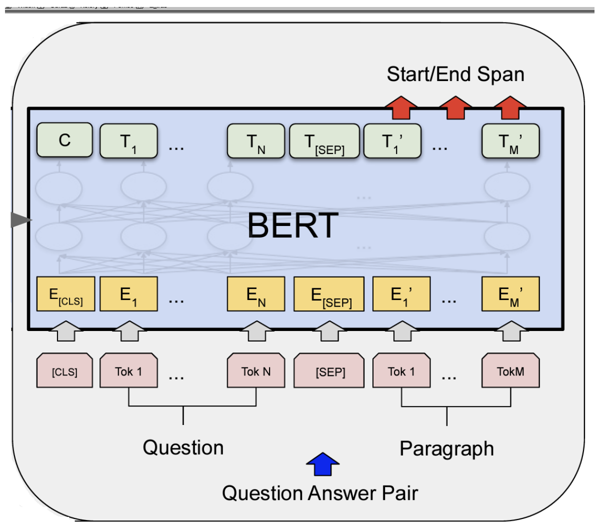


Figure 3.10: Question Answering Architecture for BERT [32]

Relation	**Question Template**
educated_at(x,y)	Where did x graduate from? In which university did x study?
occupation_at(x,y)	What does x do? What is x's job?

Table 3.1: Question-Relation Templates [54]

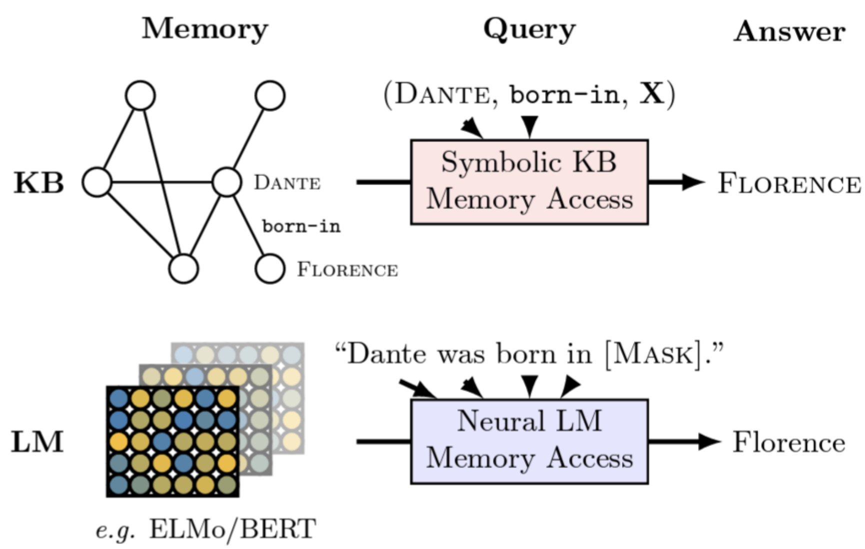


Figure 3.11: Language Models as Knowledge Encoders [55]

goal would be to evaluate the workplace for *Brian Moynihan*. The parametrized question would be *Where does Brian Moynihan work*. In case of a relation not supported by the text, the classifier returns an empty span.

Language Models are trained extensively over a large corpus. While learning, some relational knowledge is encoded and can be retrieved directly through a masking strategy. In this scenario, the LM is considered as a linguistic memory [55]. To access this memory, a simple relation pattern is evaluated. For example, let's assume that the fact that *Brian Moynihan* works for *the Bank of America* and this information has been encoded in the LM during its training. If a templated query using a masking strategy is formulated to query the workplace: *Brian Moynihan works for [MASK]*, then the LM shall substitute *bank* as the highest probable token associated with *[MASK]* (Fig. 3.11).

Large Langauge Models, i.e., GPT3, also encode knowledge of the language. However, the conditional, autoregressive text generation capabilities [14] impact the approach to RE. The conditional, autoregressive generation means that text is generated given the buffer. The buffer may contain some instructions and examples, which is **the prompt** [14] [56]. Conditioning on the buffer, information on the relation will be in the generated output. Specifically, the prompt opens a new avenue to provide either relationship examples or instructions on how relations shall be detected in the provided text.

A naive prompt engineering approach (Fig. 3.12) for RE initially did not yield reasonable results, mostly because they were not addressing the language phenomena, which led the classifier to focus on shallow language features, which could be, for example, a simple

Given the possible relations: [*member of*, *field of work*, *work location*, ..., *father*, *sibling*].
What are the relations between the subject entity and the object entity expressed by the sentence?
Sentence: Savi was born in Pisa, son of Gaetano Savi, professor of Botany at the University of Pisa.
Subject: Gaetano Savi
Object: Botany
Relation: **field of work**

Figure 3.12: Naive Prompting [58]

overlapping of words. The reason for performance significantly below fine-tuned LM (BERT) were identified to be:

- low relevance regarding entity and relation in the existing sentence-level demonstration,
- the lack of explaining the exemplary mappings of demonstrations via precise instructions in natural language.

[57].

A significant improvement in LLM RE has been achieved by designing prompt combining instructions and examples with respect to the target relation [58]. Therefore RE task has been formulated as a prompt-generation task containing:

- head and tail entities of the relation,
- examples of the relations from the repository of relation mentions,
- instruction to retrieve each relation in the provided text according to the examples or "no relation" in case nothing can be matched.

3.3 Textual Entailment

Inference or entailment is a critical ability to draw conclusions. Ido Dagan defines textual entailment to be a directional relationship between pairs of text expressions, denoted by T (the entailing “Text”) and H (the entailed “Hypothesis”). It is considered that T entails H if *a human reading* of T typically infers that H is most likely true [43] [59]. In the past, there used to be several domain-specific applications that were running language inference like textual entailment, and Textual Entailment (TE) is an attempt to provide a generic framework that would define the mechanisms of such semantic inference across many domains and establish a coherent evaluation of the proposed inference mechanisms [43].

In essence, textual entailment is a relaxation of the formal logical entailment and

Text: The purchase of Houston-based LexCorp by BMI for \$2Bn prompted widespread sell-offs by traders as they sought to minimize exposure. LexCorp had been an employee-owned concern since 2008.

Hyp 1: BMI acquired an American company.
Hyp 2: BMI bought employee-owned LexCorp for \$3.4Bn.
Hyp 3: BMI is an employee-owned concern.

Figure 3.13: Entailment Example [59]

comprises several elements that come directly from natural language, or *human* perception of it, namely, "what a person would typically infer from the premise.". The textual entailment traits can be put into the following categories:

- generalization, which represents the hypothesis as a more general statement of the premise, e.g., premise: *"antibiotics inhibit the synthesis of bacterial cell walls."*, hypothesis: *"antibiotics slow down the development of bacterias"*
- inference, which involves deriving *new* facts and grounding them with the provided premise using, e.g., logical reasoning (premise: *"dealer sold 103 cars"*, hypothesis: *"dealer sold over 100 cars"*).
- paraphrasing is the situation in which premise and hypothesis are equivalent and the textual entailment is a bi-directional relationship. For example, premise: *"Frosty situations lead to conflicts"*, hypothesis: *"Unfriendly situations lead to tensions or animosities"*

The textual entailment is a three-way classification task in which the system shall detect if a given premise/hypothesis pair is either *entailed*, *contradicted*, or *neutral*. The exact, formal algorithm that would assign the pair to either of the classes yet does not exist. From the algorithmic point of view, entailment identification is a complex "NLP complete" task [59] that involves multiple techniques to run the inference. For example, to identify an entailment for hypothesis 1 (Fig. 3.13), it is required to run the inference to confirm that BMI acquired an American concern. It requires reasoning that Huston is the capital of Texas and Texas is a part of the USA, a synonym of America. However, this is not enough. The location of headquarters in Huston does not make a company American. To complete the inference, several additional connections must be established: between the LexCorp owners and the fact that the Americans live in Huston, and that concern is a coreference to LexCorp itself.

A series of workshops drove the early attempts to solve the TE problem, "The PASCAL Recognising Textual Entailment Challenge." Initially, solutions relied on the syntactic decomposition of premise and hypothesis to match terms, called *anchors*. Generally, the

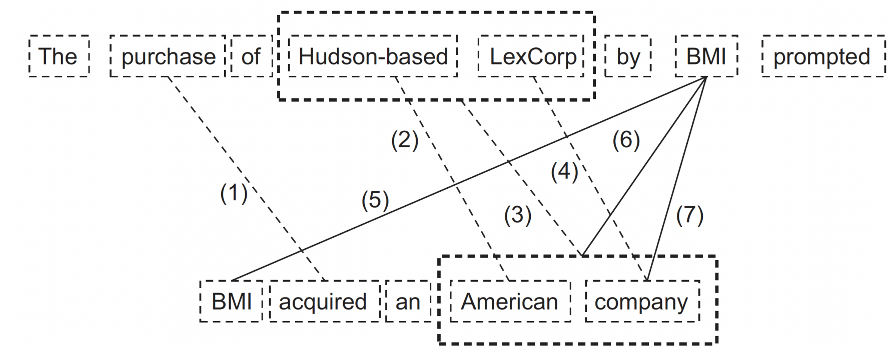


Figure 3.14: Alignment of similar terms for T (top) and H (bottom) [59]

process was structured as follows [59]:

- Candidate Alignment Generation. In this step, premise and hypothesis texts are chunked to select pairs of terms from both (Fig. 3.14). For each pair, the similarity score was calculated. The similarity score could be as simple as 1 for identical terms and 0 otherwise. However, for example, the candidate (1), "purchase","acquired" are synonyms (Fig. 3.14). There is no candidate "purchase","BMI" as the terms are of different parts of speech. Therefore, in reality, more sophisticated similarity functions were used. [59].
- Alignment. This step selects the best alignment between terms in premise and hypothesis through, for example, greedy maximization of similarity score.
- Classification. Given the aligned anchors, a feature vector is constructed. The feature vector represents "the state of similarity" between the premise and hypothesis and is used in supervised training. Given the training set, the classifier learns the decision on entailment / non-entailment given the representation of the premise and hypothesis.

Early attempts to solve the TE problem relied heavily on feature engineering on the lexical and syntactical decomposition of T and H, therefore their generalization capabilities were limited. Transformer [10] approach casts the TE task as a fine-tuning task on top of the pre-trained Language Model (BERT or RoBERT) [32], [60]. Throughout the pre-training phase, the transformer model acquires significant linguistic abilities that relax the extensive manual feature engineering requirement. These abilities are:

1. effective and thorough representations for the meanings of sentences (i.e., their lexical and compositional semantics) which can be observed through contextualized embeddings of words [32]. It means that embedding of the word "bank" would be different in the context of "river" and "money".

2. ability to handle lexical entailment. For example, the relation between "cat" and

"animal".

3. ability to handle quantification, as the transformer is capable of discerning the similarity of sentences based on words such as "none," "some," and "few." These differences will be visible in the sentence embedding.

4. and much more, for example: the ability to handle coreference (through attention mechanism), tense as it is able to distinguish past, present, and future expressions, modality (which expresses the possibility the statement is true i.e. would , could or possibly), and lexical and syntactic ambiguity which means that two sentences expressing same thing but differently will have embeddings close to each other.

. These abilities are encoded into embedding the 'CLS' token, which is the output of the Transformer component of the classifier (Fig. 3.15).

In addition to extensive pre-training that captures linguistic phenomena, LM is fine-tuned to the RTE task. The fine-tuning is performed in a supervised manner (Fig. 3.15), with two large training sets: Stanford Natural Language Inference (SNLI) dataset [44], and The Multi-Genre NLI Corpus (MNLI) [61]. These datasets address the limitations of earlier RTE training sets such as:

- they were limited in number of training examples. Early RTE corpuses had a couple of thousand hand-labeled examples. This is not enough for deep neural and transformer models.
- from the linguistic perspective, the examples were not diverse enough and, in many cases, were simply incorrect. For example, early examples referred to coreference resolution, which was problematic to resolve for even for annotators [44]. An example of such a case which assumes a specific interpretation of "New York" to be a city instead of the state, would be the premise: "A tourist visited New York." and hypothesis: "A tourist visited a city."
- MNLI specifically, provides entailment examples from ten different sources across genres and domains to model the usage of modern American English [61]. It significantly improved the generalization capabilities of the classifier as the training set accounted for more diverse examples of entailment expressions.

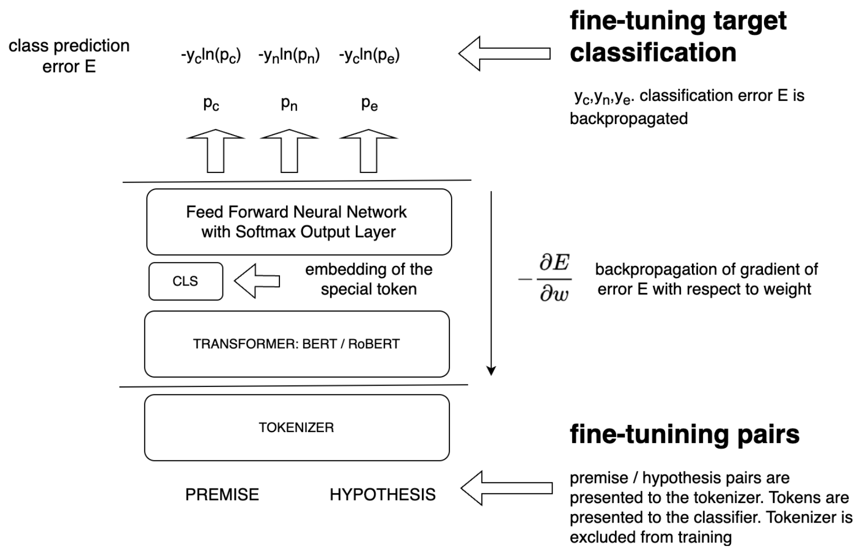


Figure 3.15: Transformer Architecture for RTE with training

3.4 Semantic Frames And Semantic Role Labeling

Charles J. Fillmore originally introduced Frame Semantics with the basic idea that one cannot understand the meaning of a single word without access to all the essential knowledge related to that word, namely, its semantic frame [62]. The semantic frames are strictly associated with the cases expressed in the sentence, and at times, the innovation was that it called for the case organization of sentences, known as the *compositionality assumption*. In other words, sentences consist of cases that define specific roles of nouns, e.g., Patient, Agent, or Instrument, that support cases. Cases were considered events or scenes to study the semantics of the words involved. The approach has led to the creation of FrameNet. This repository is the evidence of the semantic and syntactic structure of the cases considered semantic building blocks for each sentence.

In the FrameNet, each Semantic Frame was defined as the specific case and the list of its arguments. The Frame consists of Frame Elements (FE) and the Lexical Units (LU). The Frame Elements are participants (case roles) defining the frame, and LUs are text fragments that evoke a given frame. For example, the frame *commercial transaction* would be evoked by the LU *John is buying a new car from 20.000 USD*, and its FEs would be:

- *Buyer or Agent*: The person or entity making the purchase, *John* in our example,
- *Seller or Patient*: The person or entity selling the product or service, in our example undefined,
- *Product/Service or Instrument*: What is being bought, *car*

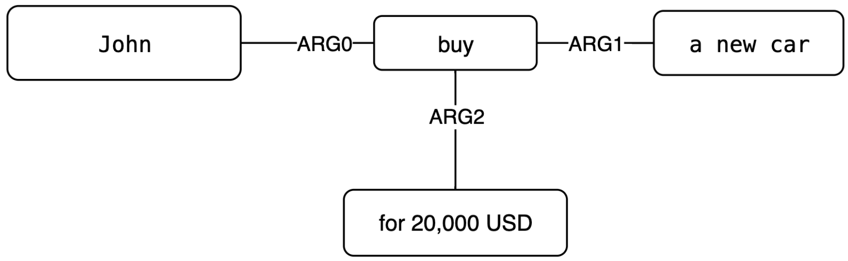


Figure 3.16: General Semantic Roles

- *Price*: The amount of money exchanged for the product or service, *20000USD*.

The same frame *commercial transaction* would be evoked for synonyms of *buy*, i.e., *purchase, acquire* etc. The FEs define "Who, What, Where, When, With What, Why, How" for each frame. The precise semantic detection of frame elements (Buyer, Seller, Price etc), in general, is not possible. For example, in the sentences *I ate dinner with Anna* and *I ate dinner with sticks*, the objects defined with "with" are difficult to distinguish for their specific roles. Therefore, generalized roles, e.g., ARG0, ARG1, ARG2, and others, are currently detected [63] (Fig. 3.16); however, the frame structure is preserved, and the assumption of case compositionality of the sentences holds.

The compositionality assumption means that each sentence is the combination of its frames. Therefore, a semantic relation between the entities can be performed within the frame's arguments after frames, and their arguments are extracted from the sentence.

Summarizing, a semantic frame serves several purposes:

- it is a means to abstract cognitive schemata, and it is this schema computational counterpart [64], which means that each frame encodes the relationship and its arguments,
- The structure of the semantic frame naturally identifies its subject and objects (as they are the frame's arguments) and the relationship between them [18] [65], which simplifies the detection of semantic relations,
- The structure of semantic frames within a sentence represents the semantics of the sentence. Therefore, the relationship detection can be performed directly on elements of the frames, not the cartesian product of elements of the sentence itself.

Semantic Role Labelling (SRL) [66] [67] identifies the frame's structure. A state-of-the-art deep pre-trained SRL classifier [68] detects the simplified structure of a frame where instead of an agent, a patient, or an instrument, it detects generic simplified arguments of a verb: ARG0, ARG1, and others [63].

3.5 Summary

In the Natural Language Processing domain, it is a known fact that solutions explicitly designed to model specific scenarios are not guaranteed to be generalized in another case as they rely on the detection schema (classifier and the training set) targeted for that case only. For example, in the railway scenario [7], the detection needs to be retrained for another case. Although it might sound reasonable to apply more sophisticated language representation as LM's pre-pretraining improves the adaptation to a specific task; the evidence suggests that the generalization achieved under this paradigm can be poor because the model is overly specific to the training distribution and does not generalize well outside it [69], [70]. Additionally, training sets often do not cover linguistic phenomena comprehensively. Therefore, deep neural networks suffer from inductive bias overfitting to shallow syntactical features [71] or words that, as for NLI example, indicate contradiction [72]. Thus, the performance of fine-tuned models on specific benchmarks may exaggerate actual performance, even when it is nominally at the human level. Therefore, an effective solution shall rely on language models' generalization capabilities rather than a dedicated detection scheme. To design a risk-comprehensive system, the example-based approach shall be reduced, and instead, general principles shall be identified and explored.

Chapter 4

Proposed Solution

This chapter describes a solution to a research problem: identifying risk interaction in the narrative. In its essence, it is a document-level relationship extraction. Specifically, the goal is to identify the risk propagation triplets Threat-Vulnerability-Assets and represent the triplets as an Asset-Vulnerability-Hazard graph (the A-V-H graph). The Knowledge Acquisition Pipeline to construct the required network representation is presented. Each processing step is provided with algorithmic complexity to prove the solution's efficacy, which is multinomial as opposed to the combinatorial complexity of the naive approach. Relationship acceptance is formulated as a multicriterial optimization task. The optimization results from an ensemble of deep textual entailement classifier and Large Language Model. At the end, the chapter discusses an approach to model various relationship templates to construct a richer semantic representation of risk propagation.

4.1 Asset-Vunerability-Threat Triplet and A-V-H Graph

The network risk model can be represented as a triplet (Fig. 4.1) modeling the risk propagation [73]. In this representation, the *Asset* is a component of the system, or its part, relevant from the perspective of the analysis context. *Assets* may be at different levels of abstraction. For instance, an asset may be a car at risk of an accident because of a slippery road surface and excessive speed. *Assets* will also be elements of the car, e.g., a tire prone to being punctured or engine components subject to specific failure, e.g., low oil level

The *Vulnerability* is a system's component, its part, or anything that interacts with the system that impacts the performance of the system. It causes the *Asset* to lose its ability to function correctly under the influence of a *Hazard* or *Threat*. In the case of a car, a *Vulnerability* may be a "slippery road surface" that generates the risk of an accident under the influence of a "speed" *Hazard*. In the case of an engine, a *Vulnerability* may be a "low oil level" generated by the "oil leak" *Hazard*. The "Low oil level" is contextual as it will be a *Hazard* for other *Vulnerabilities*, e.g., "high temperature."

The *Threat* or *Hazard* entity is the *Asset's* component, an event, or other *Asset* that

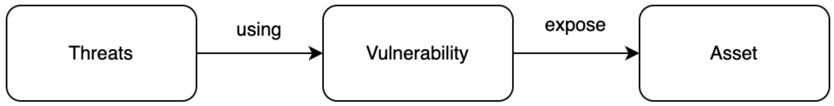


Figure 4.1: Threat - Vulnerability - Asset Triple [73]

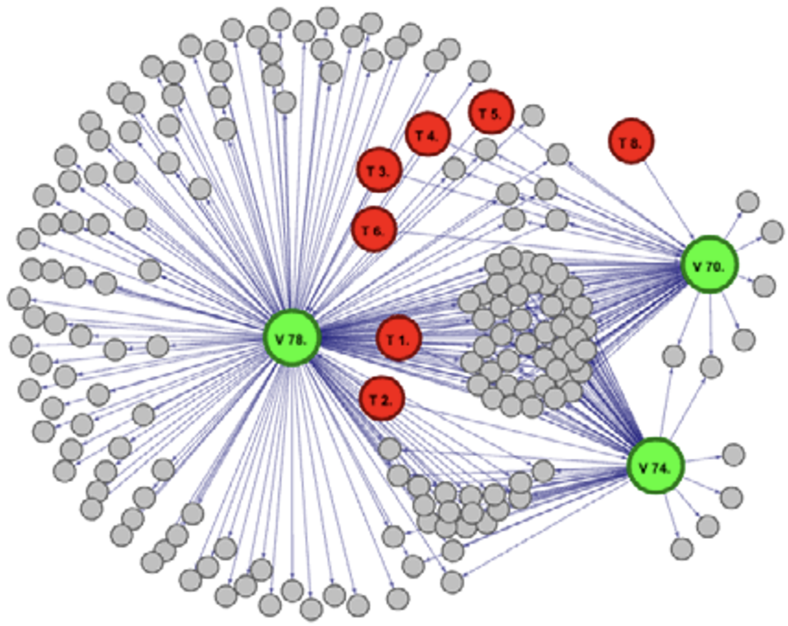


Figure 4.2: The A-V-H Graph, network approach to risk analysis. green : *Assets*, red : *Hazards*, grey: *Vulnerabilities* [1]

exposes the *Asset* to risk due to its interaction with the asset's *Vulnerabilities*

The Asset - Vulnerability - Hazard, the A-V-H graph is a Knowledge Graph that is a comprehensive representation of the AVH triplets in the system domain (Fig. 4.2), [1].

4.2 Problem Statement

Given the system's description contained in the document of D pages, consisting of a total of S sentences and containing a total of N distinct nouns, create a three-element ordered set of nouns denoting *Assets*, *Vulnerability*, and *Threat* (Fig. 4.3) forming Asset-Vulnerability-Threat triple (Fig. 4.1) and aggregate them into an A-V-H graph (Fig. 4.2) for a network analysis of risk interaction in the system.

In other words, *Assets* are system elements identified by *nouns* in the syntactic decomposition of the description. The *Vulnerabilities*, and *Threats* are contextual risk-related

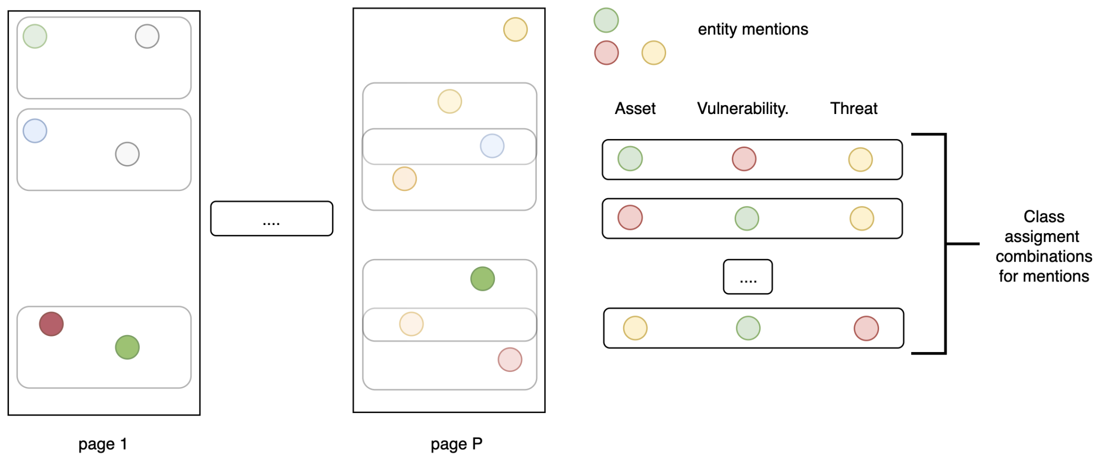


Figure 4.3: Asset-Vulnerability-Threat Extraction Problem

items detected for each *Asset*. They are identified by *nouns* in the syntactic decomposition as well [1], [73]. Therefore, the task is to identify the ordered triple of nouns to which we can assign and confirm the semantics of *Asset*, *Vulnerability*, and *Threat*. The identified triplets must be supported by the narrative they are extracted from.

Computational Complexity of Triplet Extraction

In the document holding the description, there are N distinct nouns to be assigned to either of the categories. Estimating the upper bound of the time complexity of the task in the function of N nouns is straightforward. The function f of the number of triples in the set of N nouns is the number of combinations without repetitions of a subset of 3 elements from the set of N elements:

$$f(N) = 3! * \binom{N}{3}$$

Therefore, the problem of selecting the semantic triple is of polynomial $O(n^3)$ complexity.

The narrative must support the classification of the elements into either of the categories. The NLP's Entity Recognition will not allow for the direct classification of the abovementioned classes as *Vulnerability* and *Threat* are contextualized. To confirm that the narrative supports the triple, we have to verify that there is a risk-related semantic relation among the elements in the triple. Therefore, we will select elements of the triple (Fig. 4.1) that are connected by a risk-specific relation across the document (Fig. 4.4).

Risk-Related Semantic Relation

In general, the semantic relation is defined as a triple containing:

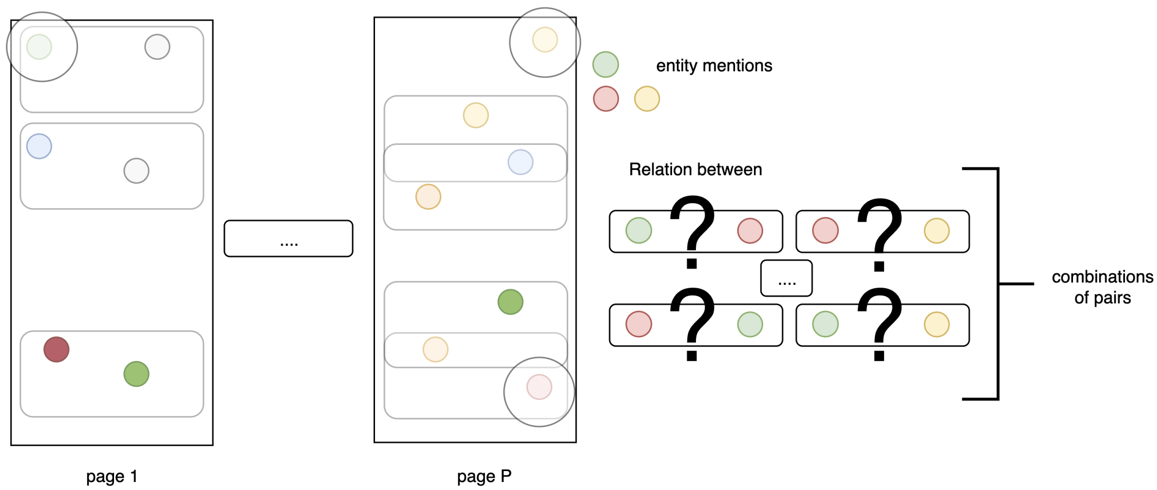


Figure 4.4: Document-Level Validation Problem

- *head entity* e_h which is the subject of the relation,
- *tail entity* e_t which is the object of the relation,
- *predicate* which defines the kind of relation between the head and tail entities, e.g., *"has an impact on"*

The relation is detected within the textual context, which contains head and tail entities and supports the predicate between them. The risk-related relation must have properties that:

- *transitive*, meaning that if a noun a is a head entity e_h in relation r with noun b being a tail entity e_t and noun b is a head relation e_h in relation with noun c is a tail entity e_t, then noun a is a head entity e_h in relation with noun c. From a risk analysis perspective, it means that the hazard propagates across the system, and we can identify the impact of *Threat* through a *Vulnerability* on an *Asset*
- *irreflexive*, meaning that fact that any *Asset* cannot be in relation with itself. From a risk analysis perspective, any system element cannot impact itself. Therefore, external factors must exist to initiate risk propagation, and *Asset* cannot be a simultaneously *Vulnerability* and *Threat* to itself.
- *antisymmetric* meaning that given the context c, a is in the relation with b and b is in the relation with a only if a and b are the same entity. In other words, given the current elements in the Asset-Vulnerability-Threat triple, current *Asset* can be assigned *Vulnerability* role and current *Vulnerability* an *Asset* role if and only if there is another context that supports the switch.

The above assumptions on risk-related relations constrain the network representation of risk propagation to directed graphs without self-loops.

Triplet Validation Problem

The entities of the triple can be distributed across the entire narrative (Fig. 4.4). The AVH triple defines the risk relation as a directed relation following from *Hazard* through *Vulnerability* to *Asset* (Fig. 4.1). In the brute-force approach, the relations are checked pair-wise within the fixed contexts, e.g., sentence. It considers a verbalized risk-related relation for each noun pair in the sentence. Therefore, we must find the combinations of contexts where entities form a risk-related relation such that a chain of relations links the triple elements.

We assume that there are N distinct nouns distributed across C contexts in the document, and there are k entities on average in each of the contexts. The function indicating the number of comparisons to perform to confirm the existence of risk-related relation between any pair of nouns is a function of the number of contexts:

$$f(C) = (k-1)^{C-1} k^C C!$$

Therefore, the time complexity of a brute force validation is $O(k^C C!)$, and it is intractable.

4.3 Proposed Solution Architecture

There are three main issues impacting the modeling of risk propagation using the AVH triple:

- computational complexity to validate the naive approach,
- lack of training sets to use other NLP approaches, for example apply document-level relationship extraction directly,
- no possibility to detect the AVH triple directly in the narrative using Entity Recognition approach

The solution will address these problems by:

- decomposition of the processing into specialized steps, namely, constructing Semantic Frames Graph and Intermediate Relationship Graph (IRG), which will reduce the computational complexity
- ensemble learning that will establish the acceptance threshold for the accepted relations to counter the lack of training examples and standard ROC analysis

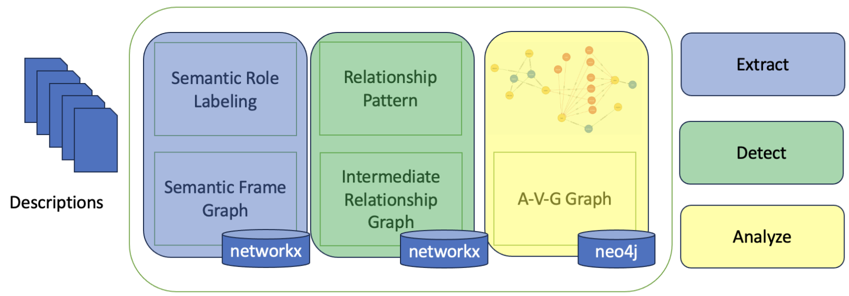


Figure 4.5: Proposed Solution Architecture

- risk-related relation analysis will replace the entity recognition to identify AVH triple elements in the narrative
- path analysis on IRG will provide required validation method

Overall, the solution will answer the research questions:

- First, as it is financially prohibitive, the research shall answer if it is possible to construct a risk detection system without creating a dedicated LLM for Risk Analysis.
- Second, it shall evaluate available trained and language-specialized classifiers and construct the pipeline to identify risk relations without a dedicated training set.
- Third, a validation method shall be provided.
- Finally, the pipeline shall produce the A-V-H graph, a chosen network model of risk interaction.

4.4 Sentence Decomposition and Semantic Role Labelling

Semantic Frames are semantic build blocks of sentences. They are at the heart of the *compositional semantics* that postulates that the overall semantics of the sentence is a composition of its frames [74]. Semantic Role Labeling (SRL) detects frames and their structure in the sentence (Fig. 4.6). The central element identified in each frame's structure is verbs [65], for example, *offering* on (Fig. 4.6). The other elements are generalized semantic roles (ARG0, ARG1, ARGM-LOC, ARGM-TMP, etc.) [63] identified by the SRL and associated with the frame's central verb (Fig. 4.7) We seldom encounter a situation in which an event

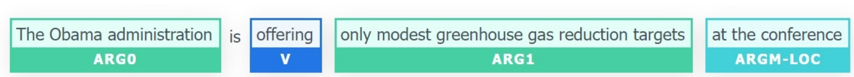


Figure 4.6: Semantic Role Decomposition of a sample sentence

is described in a singular sentence with a subject and object explicitly stated without

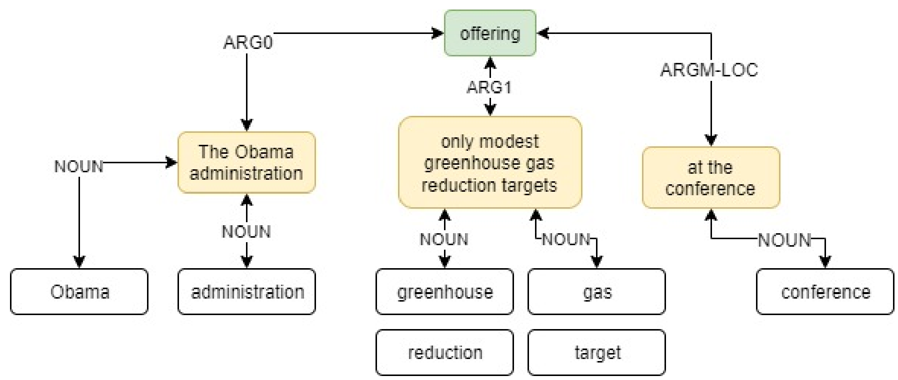


Figure 4.7: Semantic Frames Graph representation of a sample sentence

additional predicates. A singular sentence would transform into a single frame. However, this is not the case for complex subordinate sentences. Therefore, when complex sentences are decomposed, then a hierarchy of semantic frames is created to capture the interaction between frames and each frame's argument separately from other frames' arguments. (Fig. 5.17).

Semantic frame decomposition is essential for relationship detection as it separates sentence components according to the frame to which they belong. This simplifies the analysis as, for example, the relation of interest may be chained in the hierarchy of frames. In a sentence *"A water landing of a jetliner that lost both engines due to hitting birds became known as the Miracle on the Hudson River"* (Figure 5.17), a subject *jetliner* is linked with *engines* and *birds* through frames *lost* and *hitting*. The main frame *known* is skipped as irrelevant from a risk analysis perspective. This simplifies the analysis of the relation between *jetliner* and other nouns - potential hazards of a jetliner:

1. *[jetliner]* - *"a jetliner lost both engines due to hitting birds"* - *[engine]*
2. *[jetliner]* - *"a jetliner lost both engines due to hitting birds"* - *[bird]*

The relation between head and tail entities is often distributed across the description. In this case, the relationship is expressed by a sequence of frames that connects them and forms a reasoning scheme that justifies the existence of the relation (Fig. 4.9). For example, it is possible, given the exemplary text (Fig. 3.4.), to verify that *"droplet"* is in a relationship with the *"engine"*. In this case, the analysis may follow the deduction path: *"droplet"* → *"supercooled water droplets collide with a surface"* → *"if supercooled water droplets collide with a surface they may result in blocked fuel inlet pipes"* → *"blocked fuel inlet pipes"* → *"aviation fuel designed for use in aircraft powered by gas turbine engines"* → *"engine"*

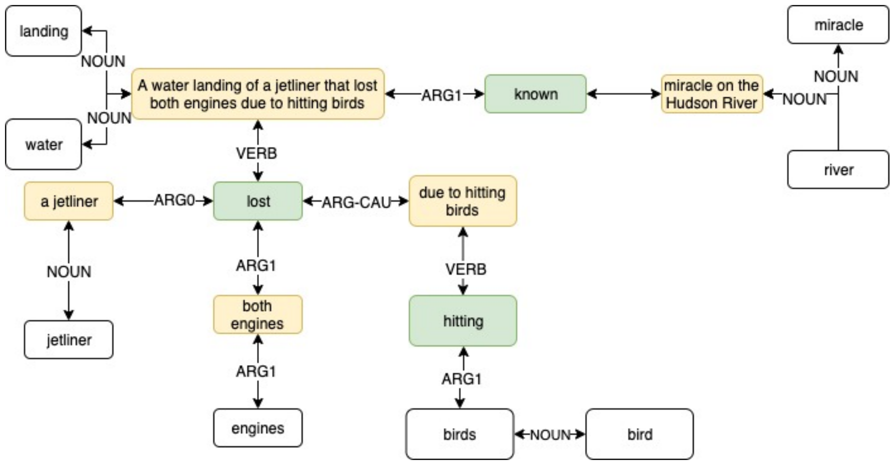


Figure 4.8: SFG Hierarchy of Frames

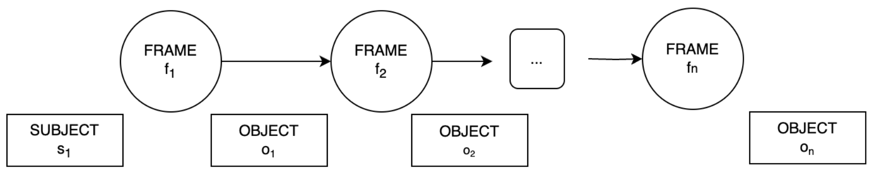


Figure 4.9: Subject - Frame(s) - Object(s)

4.5 Semantic Frames Graph

Formally, a Semantic Frame Graph is an undirected, attributed, heterogeneous graph

$$G = (V, E)$$

where:

- V is a set of nodes of the following types:
  - Noun: nouns detected in the frame's argument. White rectangle
  - Argument: a span of text describing the semantic role of a frame. Yellow rectangle.
  - Verb: verb identifying an event associated with the frame. Green rectangle.
- E is a set of edges representing:
  - *role_type* : a specific semantic role type for example: ARG0, ARG1, etc.
  - *verb*: a connection to the central verb of the role's constituent frame. This happens if the description contains subordinate clauses.
  - *noun*: a connection to a noun that is a part of the frame only.

The graph is constructed by applying recursive semantic decomposition (Algorithm 1) on every sentence in the corpus. At the sentence level, the decomposition is the recursive application of deep SRL identification [68] until none of the identified frame's elements can be further split into frames (Algorithm 1). Nouns are assigned to the lowest-level frame's role and are not repeated at the higher levels.

The results are recorded in a graph structure as a hierarchy of frames (Fig. 4.10). Connections between frames are established through nouns that frames share.

Algorithm 1 Recursive SRL decomposition

procedure PARSE_FRAME($Frame\ f, Graph\ g$)
  $args \leftarrow getSRL(f)$ ▷ get arguments of the frame f
  $verb \leftarrow getVerb(f)$ ▷ get the verb for the frame f
  $g \leftarrow add\_node(verb)$
  for `argument in args` **do**
    if argument is a frame **then**
      parse_frame(argument,g) ▷ further decompose current argument in case it is a frame
    else
      $g \leftarrow add\_link(verb, argument)$ ▷ connect verb's frame with its argument
      for `n in getNouns(argument)` **do**
        $g \leftarrow add\_link(argument, n)$ ▷ connect frame's argument with its nouns
      end for
    end if
  end for
end procedure

Computational Complexity

Given that, on average, a sentence has f frames, the decomposition of S sentences into its frame structure is a function of a number of sentences $f(S) = f * S$, which is linear with the number of sentences.

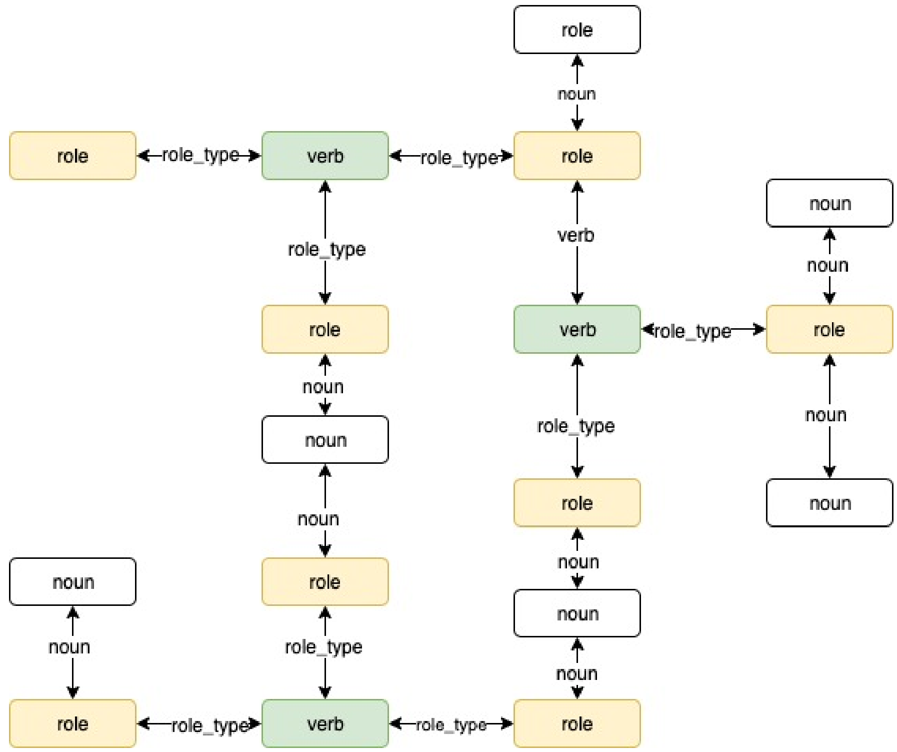


Figure 4.10: Semantic Frame Graph Structure

4.6 Semantic Pattern and Relationship Extraction

Relationship extraction is performed by applying a semantic pattern on frames connecting nouns in the SFG. The semantic pattern, a template t for the relationship context c is a textual core of the relation that requires providing *subject* i.e., *head entity* (e_h) and *object* i.e. *tail entity* (e_t). The substitution is called *verbalization* (Fig. 4.11) [45].

The relation context should support the verbalization. Entailment evaluates how well it is supported. The premise is the relation context, and the hypothesis is the template verbalization.

For example, the frame: "The failure to plan and account for extreme floodwaters resulted in the immediate death of 26000 as a result of the water itself" contains one verb, "resulted" suggesting some form of impact (Fig. 4.11). It is a premise. Assuming that the t: 'has impact on', head entity: e_h: "floodwater" and tail entity: e_t: "death", then the verbalization of the hypothesis is "floodwater has an impact on death". Therefore, the relation detection is cast as NLI problem [44].

Relationship extraction over the SFG graph verbalizes a relationship template t on frames connecting every *Noun* nodes in the graph. The structure of the SFG graph allows traversing it using an explicitly structured walk on the graph, namely, metapaths [75] with a

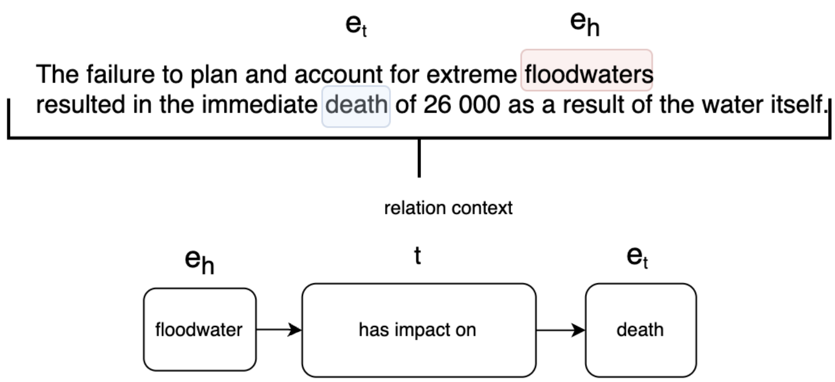


Figure 4.11: Semantic Pattern Verbalization: template t, relation context c

limited number of metapaths:

1. *noun-role-verb-role-noun* where the relationship is contained within a single frame
2. *noun-role-noun* where the relationship is contained within the frame's argument, which is a special case of a single frame,
3. a: *noun-role-verb-role-noun-role-noun*, b: *noun-role-verb-role-noun-role-verb-role-noun*, where the relationship is held in adjacent frames
4. *noun-role-verb-role-verb-role-noun* where the relationship is contained in two frames of a subordinate clause

The structured walks allow the relationship detection specific to each metapath. There are two cases (Fig. 4.12):

- single frame: metapaths 1 and 2 ,
- two frames: metapaths 3a, 3b and 4.

4.6.1 Single Frame Relation Extraction

A single frame case is a basic case in which the relationship between the head and tail entities is detected within the frame. This is template verbalization case only [45], e.g., *water vapour create droplets* is validated against a relationship template T : "$haseffecton$, e.g., *"water has impact on vapour"* (Fig. 4.13).

Computational Complexity

As sentences have, on average, k nouns, in the worst-case scenario, all k nouns would be part of each frame. We have S sentences in the document, which contain, on average,

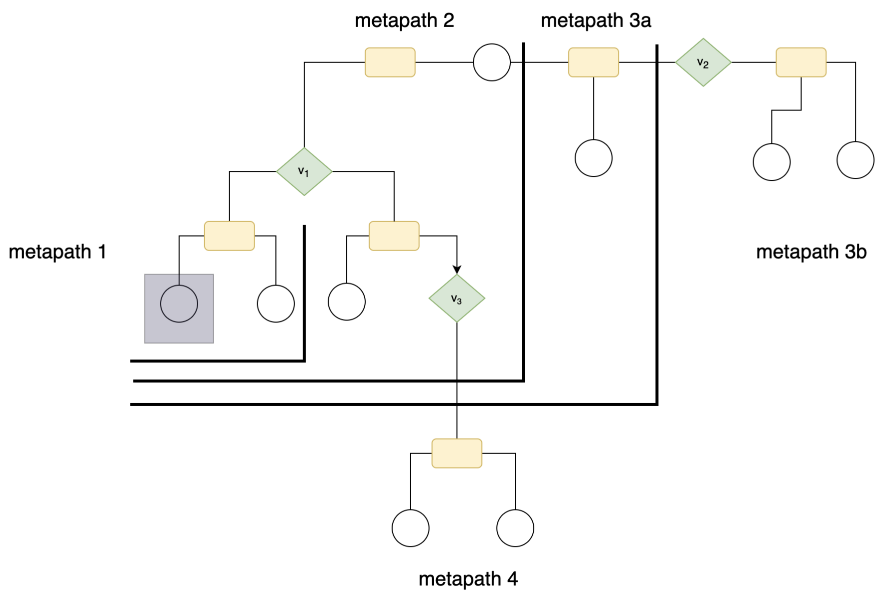


Figure 4.12: Relation Detection per Metapaths

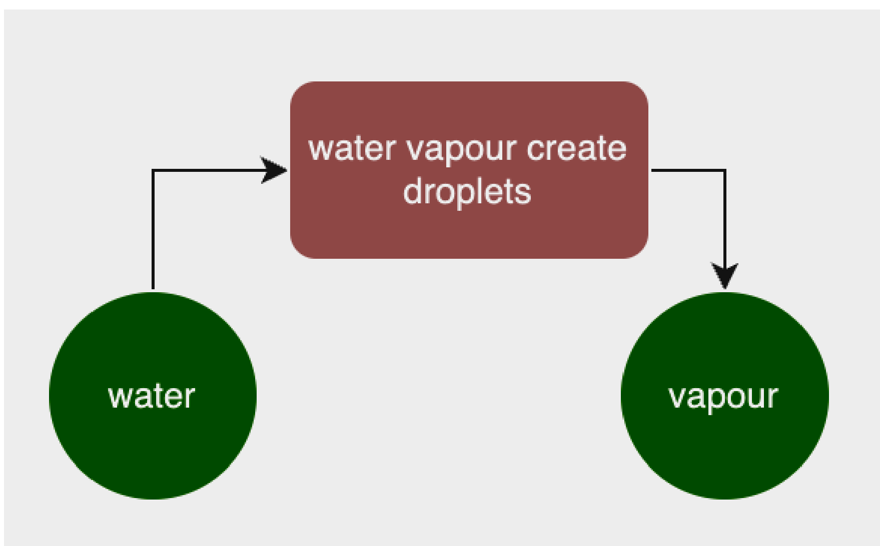


Figure 4.13: Single Frame Case

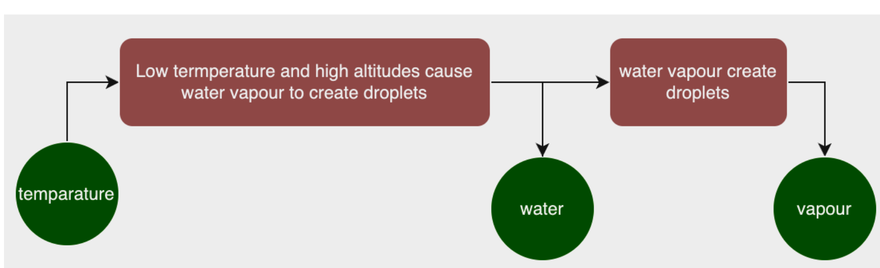


Figure 4.14: Two Frames Case

f frames; therefore, the number of comparisons to evaluate template entailment is given by the linear function of sentences in the document: $f(S) = k*(k-1)*f*S$. The time complexity is then linear with respect to the number of sentences in the description: $O(S)$

4.6.2 Two-frame Relation Extraction

The two-frame case extends the single-frame case for situations where one of the frames entails the relation and the other is semantically coherent. The verbalizer approach cannot be used directly on both frames as it does not comply with NLI training [44]. The relationship between entities (nouns) exists if a path in SFG joins them [37]. A path of semantic frames joining them creates a passage defining a relationship. For example, a path between *water* and an *engine*: *"water"* → *"supercooled water droplets collide with a surface"* → *"if supercooled water droplets collide with a surface they may result in blocked fuel inlet pipes"* → *"blocked fuel inlet pipes"* → *"aviation fuel designed for use in aircraft powered by gas turbine engines"* → *"engine"*

The generated path may indicate that *"water"* impacts *"engine"*. However, we cannot 'entail' the whole path as it would not comply with the training scheme for entailment, which considers inference on a single sentence only [44] [61]. Therefore, the relationship detection task is split into coherency and single-frame template entailment and performed pair-wise for each consecutive frame in the path. The metapath walks 3a, 3b, and 4 perform the required pair-wise validation for each node's adjacent frames in the SFG graph.

Another reason for such specific entailment is that many descriptions in the risk analysis domain assume a correlation of entailment with other elements based on discourse coherence. For example, in the following sentences *"Electricity was cut off in the control room. All electronic equipment in the control room was disabled."* it is natural for a human to infer that *"Electricity has an effect on electronic equipment in the control room"*. However, using the standard NLI approach, we cannot confirm the entailment of such a hypothesis as evidence is distributed across two sentences.

The approach to measuring dialog consistency between two frames is based on the Modified Dialog Coherence Function. The Modified Dialog Coherence Function measures a "semantic drift" between two frames [76]. An entailment of the relation template is calculated using the Combined Entailment Function, which combines dialog consistency with template entailment at the single-frame level, hence complying with the NLI task [44] it applies the NLI task in the two-frame scenario.

4.6.3 Modified Dialog Coherence Function

Let F_i and F_{i+1} be the preceding and succeeding frames in the path. The backward-looking centers of the pair of frames F_i and F_{i+1} are generalized a subject or an object of the frame F_{i+1} Generalized subject and object nouns are linked to ARG0, ARG1, ARG2, ARG4 roles in the SFG decomposition of the frame F_{i+1}

Let $f^c_{i,i+1}$ denote a coherence function between both frames F_i and F_{i+1}.

The coherence functions shall be bounded 0 $f^c_{i,i+1}$ 1 such that the interpretation if, from the linguistic standpoint, is that if $f^c_{i,i+1} \approx 0$ means that both frames are not coherent and if $f^c_{i,i+1} \approx 1$ are coherent.

The function shall evaluate the following pair of frames (Example 1): F_1: *water vapour create droplet* and F_2: *spark causes ignition of fuel* as incoherent.

The following pair (Example 2): and F_1 : *water vapour create droplet* and F_2: *droplet can block the fuel inlet pipe* as more coherent than Example 1.

To calculate the coherence, we will use zero-shot text classification *TC* [77]of the backward-looking centers *blc* in the frames and normalizing the output, hence defining Modified Dialog Coherence Function to be:

$$f^c_{i,i+1} = \frac{1}{|blc|} \sum_{m=1}^{|blc|} \left\{ \begin{matrix} 1 : TC(F_i, m) > TC(F_{i+1}, m) \\ TC(F_i, m) / TC(F_{i+1}, m) : TC(F_i, m) \;\; TC(F_{i+1}, m) \end{matrix} \right\} \tag{4.1}$$

The function (4.1) defined as such has the properties:

- $f^c_{i,i+1} \approx 0$ if there is no reference to any of the backward centers in F_i; hence frames are incoherent. The score for Example 1 is 0.0022
- $f^c_{i,i+1} \approx 1$ if there is a perfect overlap of the centers in both frames, meaning both frames are almost identical semantically.
- $0 < f^c_{i,i+1} < 1$ if there is an overlap of the centers and frames are semantically related. The score of Example 2 is 0.6

Computational Complexity

The metapaths constrain the way the SFG graph is traversed. The worst-case scenario requires the most computation when all relationships are represented as metapath 3b, which connects nouns separated by two frames sharing one noun (Fig. 4.15). Additionally, in the scenario, each sentence in the text is singular and decomposed to a single frame; therefore, each frame has an average number of k nouns. To evaluate all relationships, we must visit all metapaths 3b subgraphs and traverse them for $2^{(2k-2)^2} - 2^{(2k-2)(2k-3)}$ pairs. For

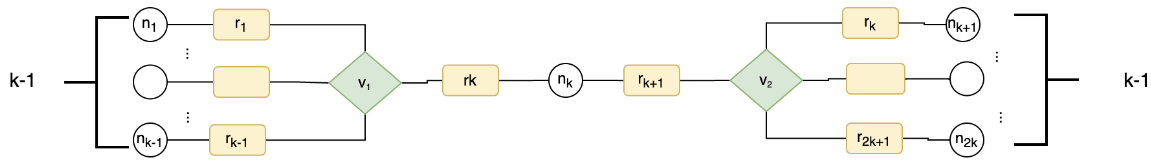


Figure 4.15: Metapath 3b: the worst-case scenario

each pair, we have to find the shortest path connecting them using the Dijsktra algorithm - complexity $O((V+E)log(V))$, where V is the number of vertices in the subgraph. The subgraph is sparse $V^2 >> E$, therefore the complexity has the form: $O(Vlog(V))$. The number of vertices is easy to estimate as there are 2k nouns, 2k roles, and 2 verbs. Therefore, the time complexity to evaluate all relations in a metapath 3b can be estimated as

$$O(k) = (k-1)^2(4k+2)log(2k+2)$$

depends on the average number of k nouns per sentence only In the worst-case scenario, all metapaths 3b will be connected. Therefore, the upper bound on complexity is the number of pairs of sentences

$$O(S) = S*(S-1)$$

.

4.6.4 Combined Entailment Function

Combined entailment $f^e_{i,i+1}(T)$ measures entailment of relationship template T within two adjacent frames in the path F_i and F_{i+1}. It evaluates the entailment of the template of the relationship in both frames and their dialog consistency. The function is bounded: 0 $f^e_{i,i+1}(T)$ 1.

Let n_i be a noun in the path leading to frame F_i like *"temperature"* in example (Fig. 4.14), n a noun linking F_i and F_{i+1} like *"water"* in example (Fig. 4.14), n_{i+1} a target noun like *"vapour"* in example (Fig. 4.14). Let T denote the relationship template *"has an effect on"*, which would mean that $f^e_{i,i+1}(T)$ would measure if *"temparatere has an effect on vapour"* in two frames setup (Fig. 4.14). Let $T(n_1, n_2)$ denote verbalization of the template given pair of nouns (n_1, n_2), for example, *T("temperature","water)* would resolve to *"Temperature has an effect on water"*. We define a Combined Entailment Function given the template T to be:

$$f^e_{i,i+1}(T) = min[f^c_{i,i+1}, max[RTE(F_i, T(n_i, n)), RTE(F_{i+1}, T(n, n_{i+1}))]] \quad (4.2)$$

where RTE is a classifier detecting *entailment* classification probability given frame F and verbalization of the template T. We used transformer RTE implementation [45].

Following the example, we calculate the value of the RTE of the template verbalization *"temperature has effect on water"* of the frame *"Low temperature and high altitudes cause water vapour to create droplets"*; the template verbalization *"water has an effect on vapour* for the frame *"water vapour create droplets"*; dialog consistency between both frames and apply the rule (4.2).

4.6.5 Multiple Templates

The goal of the solution is to model the propagation of hazards. This defines the requirements for semantic relations to be:

- irreflexive
- transitive
- antisymmetric.

Only such relations will model the potential risk propagation within the system. For example, a template *has effect on* meets the criteria as no object can affect itself without external cause. The effect propagates, meaning that if component a has an effect on another component b and the component b has an effect on another component c, then component a has an effect on component c. The effect relation is antisymmetric, meaning that given narrative justifying the effect between components a and b, such as *"airplane uses engines for flying"* does not automatically mean the opposite relation unless detected in the narrative specifically. The example justifies the hazard propagation from *engine* to *airplane* only, not vice versa. If the narrative is expanded with the text: *The engine's performance relies on fuel stored in the airplane's wings.*, only then the *has effect on* relation can be established between *fuel*, *airplane* and *wing* and *engine*.

There are many more ways to express the hazard propagation. For example, *a type of* relation meets requirements for risk propagation relation as it is irreflexive (hardly any object can be a type of itself), transitive (if an object a is a type of the object b and then object b is a type of the object c then an object a is a type of the object c). A *a type of* relation is antisymmetric by definition. Continuing with the fuel example, if the narrative is expanded with *The ATF is a type of aviation fuel* then, *ATF* connects *fuel* and propagates hazard to elements *fuel* would propagate: *engine* and *airplane*.

From the risk propagation perspective, *precise* detection and comprehensive representation of semantic relations between objects is not required. For example, another risk-related template is introduced *is a part of* with the context *Engine is a part of the airplane*. From a risk propagation perspective, it is irrelevant if the risk propagation connection between *engine* and *airplane* is established through *is a part of* or *has effect on*. Therefore, selecting the best risk-related relationship between the entities the narrative supports, namely, with the highest coherence score, is enough.

Formalizing the approach, let $E^k_{i,j}$ denote the set of edges in the IRG graph connecting nouns n_i, n_j with respect to the template k. Let T denote the set of templates. Let $w_k(i, j)$ denote a weight, dialog score, between nodes with respect to the template k Then the relationship $R_{i,j}$ connecting these nodes is defined such a template t from maximizes the total score between the the nouns n_i, n_j across all the templates and metapaths:

$$R_{i,j} = \underset{t \in T}{\arg\max}\, w_t(i, j)$$

4.7 Intermediate Relationship Graph

The Intermediate Relationship Graph (IRG) is a Semantic Network that stores all detected relations between nouns by applying a metapath walks on the SFG. Formally, the IRG a weighted multigraph (Fig 4.16), where nodes are nouns and edges represent the verbalizations of the template of, for example, *"has effect on"* relation; edges' weights are either entailment score or value of the Combined Entailment Function depending on the walk; the metapath that connects the nouns in the original SFG graph is recorded as an edge attribute.

Both verbalized relation and dialog consistency are transitive. Therefore, it is possible to traverse the IRG freely, meaning the NLI problem has been transformed into a graph traversal one. For example, node n_1 is directly connected with node n_3, which indicates a direct impact. It is also connected through node n_2, which could indicate a mediated impact.

While traversing the graph, we can use only edges with scores above the assumed threshold and analyze the evidence. For example, there is a metapath joining nodes n_1 and n_2 (Fig. 4.16) in the SFG graph. Weight w_3 is the score of the template verbalization of the connection, and m_3 means, depending on the metapath, either a frame or two frames connecting them.

From the risk analysis perspective, the graph provides a complete propagation of impact between system elements, where edges provide scores and evidence.

The IRG graph is a central point of representation of risk within a single document and across documents. It allows incremental knowledge acquisition as new relations can be

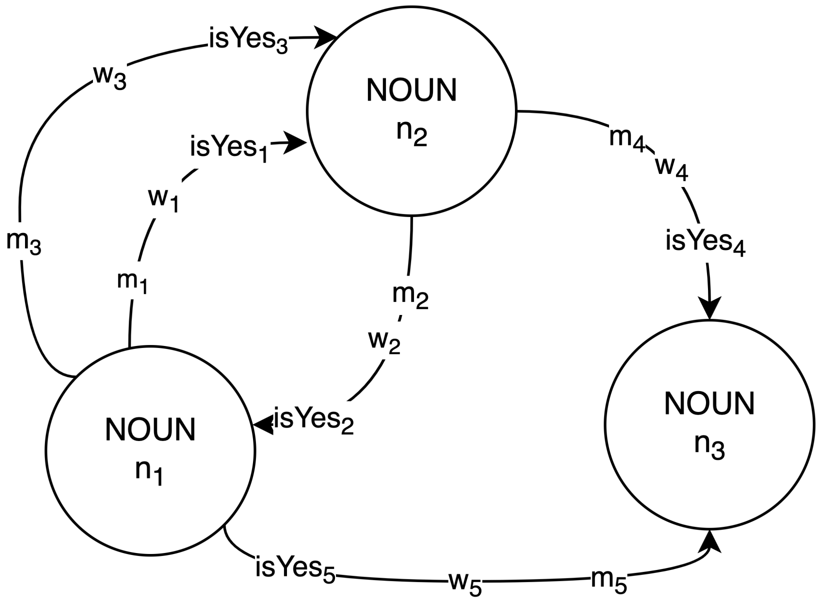


Figure 4.16: Intermediate Relationship Graph Structure

added to the graph.

Computational Complexity

The time complexity of constructing the IRG graph from SFG is bounded by the complexity of the most complex metapath and it is bounded by the number of sentences in the narrative, not the number of nouns:

$$O(S) = S^2$$

4.8 Asset-Vulnerability-Hazard Graph

The Asset - Vulnerability – Hazard (A-V-H) graph aggregates the relationships expressed in the IRG graph that meet the minimum weight criterion for the weights[1]. It isolates three types of nodes essential from the system analysis perspective, namely:

- Asset node: an element of the system that is relevant from the perspective of its proper operation in the analysis context. Assets may be at different levels of abstraction. For instance, an asset may be a car at risk of an accident because of a slippery road surface and excessive speed. Assets will also be elements of the car, e.g., a tire prone to being

[1] the threshold is is solution to a multicriterial optimization task

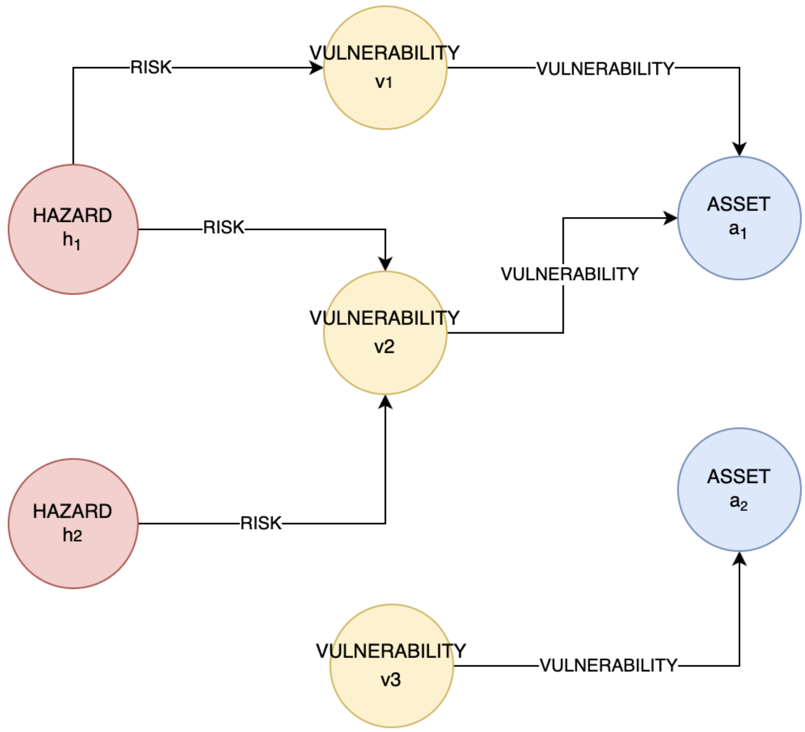


Figure 4.17: A-V-H Graph Structure

punctured. Assets will be engine components subject to specific failure, e.g., low oil level

- Vulnerability node: an element of an object, an event, or any other object that causes the Asset to lose its ability to function correctly under the influence of a Hazard. In the case of a car, a Vulnerability may be a "slippery road surface" that generates the risk of an accident under the influence of a "speed" Hazard. In the case of an engine, a Vulnerability may be a "low oil level" generated by the "oil leak" Hazard. "Low oil level" will be a Hazard for a Vulnerability, e.g., "high temperature."
- Hazard node: a system element, an event, or other asset that exposes the Asset to a risk due to the Vulnerability

Relying on the transitivity of relations in the IRG graph, the construction of the specific A-V-R nodes is performed as follows:

- Asset nodes will be all IRG nodes by default.
- Vulnerability modes will IRG nodes that directly affect Assets. Hence, these are the nodes that directly neighbor with Assets.
- Hazard nodes will be IRG nodes that connect with the Asset through Vulnerability. These are nodes from which the Asset node is reachable through its Vulnerability.

Therefore, the aggregation transforms a weighted, directed multigraph IRG into an unweighted graph A-V-H (Fig. 4.17). The connection between nodes in A-V-H is established if the IRG graph connects them with an edge whose weight is above the assumed threshold or if there is a path between nodes connected with all edges' weight above the assumed threshold.

The A-V-R graphs enable risk analysis using a graph analysis approach, i.e., nodes' centrality measure, to identify hazard-related hubs that aggregate and transmit influence to other system elements [1].

Computational Complexity

The time complexity to establish *Asset* - *Vulnerability* pair is given by the time complexity of evaluating neighbors of *Vulnerability*. In the worst-case scenario, when IRG is a complete graph, each node will have N-1 neighbors. Therefore, the upper bound on the time complexity is

$$O(N) = N * (N-1)$$

.

The relation between *Vulnerability* and *Risk* exists if a path links them in the IRG. Therefore, the time complexity is estimated by the complexity of finding the shortest path. The shortest part, however, shall skip the *Asset* node as it is already connected with the *Vulnerability*. The IRG graph is a sparse graph therefore, the complexity of a single path between selected *Asset* and *Vulnerability* is $O((N-2)log(N-2))$. Again, in the worst-case scenario, we must scan all IRG nodes for *Asset* and *Vulnerability* pairs to find the shortest paths to all remaining nodes. Therefore the overall time complexity is

$$O(N) = N * (N-1) * (N-2)log(N-2)$$

which is

$$O(N^3)$$

4.9 Validation and Explainability

4.9.1 Validation and Threshold Calculation

Validation in the context of machine learning and data science means evaluating the model on unseen data. In the proposed solution, as no specific dataset holds unseen data,

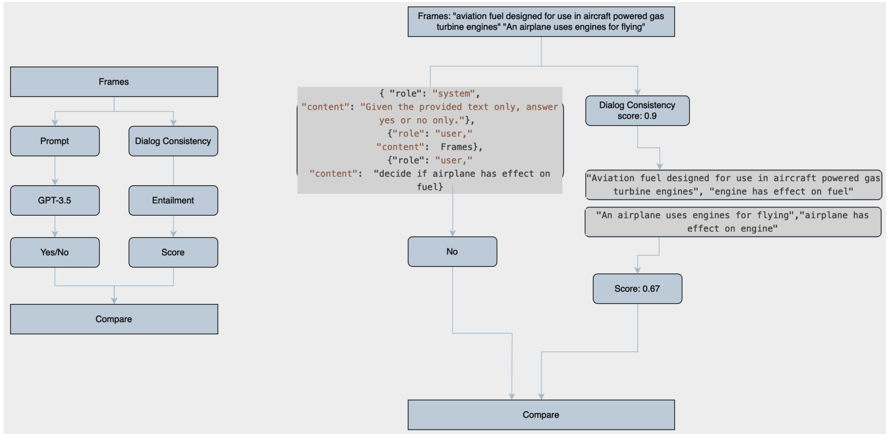


Figure 4.18: Validation Approach

validation is performed by another model, LLM, performing the same relation extraction task. It aims to establish the cutoff threshold for the weights on the IRG graph. The *threshold* is the value of the dialog function that maximizes the number of LLM's confirmed relations and minimizes the number of LLM's rejected relations. The approach is based on the ensemble of, in fact, three independent classifiers performing the same Language Inference task ((Fig. 4.18):

1. the entailment and text classification that are combined to provide the estimation of the *coherence function*, which is a *weight* on the IRG edges,
2. the prompted LLMs, which provides another independent decision on whether the edge encodes the relation template and the edge is justified. LLM works on prompt engineering principle in a zero-shot setup [78], [14], [79] which is encoded as a *isYes* parameter on the edge. The value of isYes: 1 indicates that LLM confirms the verbalization, and 0 indicates otherwise.

The LLM prompt is constructed to provide explicit instructions on the task: "Given the premise text only, decide if the premise explicitly and directly entails hypothesis. Answer Yes or No only". Answer 'Yes' is converted to 1, 'No' to 0.

Formulation of the Optimization Task

Let $S = \{s_1, s_2, \ldots, s_n\}$ be the set of elements denoting distinct *weights* in the IRG graph with additional attributes associated with them and each element s_i in the set S contains:

- w_i: the weight value,
- $y_i^{(0)}$: number of edges having weight w_i in the IRG classification unconfirmed by LLM,
- $y_i^{(1)}$: number of edges having weight w_i in the IRG classification confirmed.

The elements of the set S have the property that they are indexed with respect to the *weight*:

$$\forall s_i, s_j \in S, \text{ if } i > j \text{ then } w_i > w_j$$

Let

- $t^{(0)}$ denotes the total number of unconfirmed relations in the IRG and
- $t^{(1)}$ denotes the total number of confirmed relations in the IRG.

Then, we can define functions:

- $f^{(0)}(w)$ returning a fraction of cumulative unconfirmed relations in IRG below or at w
- $f^{(1)}(w)$ returning a fraction of cumulative confirmed relations in IRG below or at w

The values of the functions are calculated as:

$$f^{(0)}(w) = \frac{\sum_{i=1}^{n} y_i^{(0)} | w_i <= w}{t^{(0)}}$$

$$f^{(1)}(w) = \frac{\sum_{i=1}^{n} y_i^{(1)} | w_i <= w}{t^{(1)}}$$

Let

$$\mathbf{F}(w) = (-f^{(0)}(w), f^{(1)}(w))$$

Then, we can formulate the multi-objective optimization task to evaluate the IRG edge cutoff threshold of w_{th}, below which we will maximize the number of unconfirmed and minimize the number of confirmed relations to be removed from the IRG.

$$\min_{w} F(w) \tag{4.3}$$

The optimization task Eq. 4.3, can be solved by selecting the element of the Pareto Front: $(f^{(0)}(w_{th}), f^{(1)}(w_{th})$ such that

$$\min_{w} \|\mathbf{F}(w) - \mathbf{z}^{ideal}\|$$

where

$$\mathbf{z}^{ideal} = (0, 0)$$

4.9.2 Explainablity

Explainability in AI refers to the possibility that a human can understand, interpret, and accept decisions made by artificial intelligence agents. It goes past performance scores, e.g., ROC curve [80] or measures such as dialog coherence or textual entailment. In the

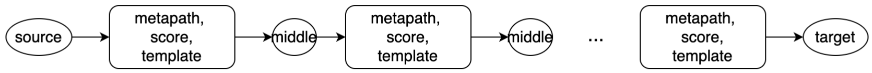
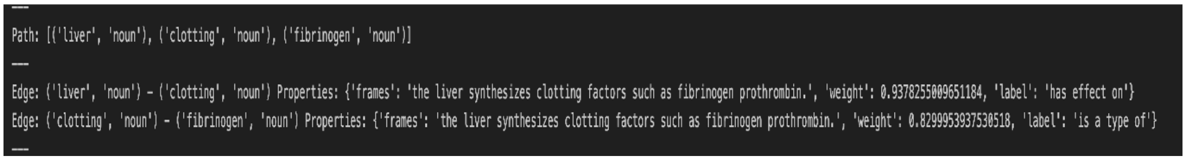


Figure 4.19: Explaiablity Chain



```
---
Path: [('liver', 'noun'), ('clotting', 'noun'), ('fibrinogen', 'noun')]
---
Edge: ('liver', 'noun') - ('clotting', 'noun') Properties: {'frames': 'the liver synthesizes clotting factors such as fibrinogen prothrombin.', 'weight': 0.9378255009651184, 'label': 'has effect on'}
Edge: ('clotting', 'noun') - ('fibrinogen', 'noun') Properties: {'frames': 'the liver synthesizes clotting factors such as fibrinogen prothrombin.', 'weight': 0.8299953937530518, 'label': 'is a type of'}
---
```

Figure 4.20: Explainability Example

proposed solution, the IRG edges store the metapath, dialog coherence score, and relation template used to detect the relation between the designated SFG nodes. The explainability means that the solution is able to provide the exact chain of frames connecting the required nodes (Fig 4.19), providing all required information supporting the existence of the specific relation between them. For example, the sample narrative describing the function of the liver is analyzed to detect how the liver impacts other human body mechanisms, such as hemostasis. Such an approach would allow us to evaluate the potential impact of the medicine, which mechanism of action targets the liver. This functionality is important as medicine can indirectly affect other essential processes. The example (Fig. 4.20) provides an explanation of how the liver impacts fibrinogen, given the narrative on the liver.

4.10 Summary

The proposed solution identifies the risk interactions contained in the narrative that are obviously *hidden*; hence, there has been limited success in designing a Knowledge Acquisition Pipeline for a comprehensive network representation of risk interactions. Unfortunately, most current solutions rely on the dedicated detection schema targeting either a specific input format or analysis domain such as railway [7]. The proposed solution relaxes the training set constraints. It applies existing and available deep NLI classifier BART [77] to infer the relation through the entailment between relationship context expressed combination of frames.

In several publications on relations detection, especially in document-level relationship detection, the researchers specifically raise the idea to "explicitly explore reasoning" [46]. Current LLMs have gained significant capabilities, however, their reasoning ones are still area of intensive research. The solution proposes the ensemble of deep NLI classifiers and prompts LLM. The risk relation acceptance threshold is a solution to the multicriteria optimization task that combines the decision of both.

Compared with the direct approach to identify risk-interaction triples (Asset-Vulnerability-Risk) in the narrative, the proposed detection and validation are performed in multinomial time complexity which is feasible for large descriptions.

Chapter 5

Results

The solution aims to process selected text(s), detect risk propagation relations defined by the templates, and construct the Asset-Vulnerability-Hazard graph to represent the risk interactions in the described system. Initially, a synthetic and easy case will demonstrate how the solution works step by step. The synthetic case is a risk modeling scenario. Also, A-V-H graphs will be constructed to show results on other types of corpora (financial and medical) to show that the proposed approach provides meaningful insides into the propagation of risk in other domains. The multi-template example will show that it is possible to construct more comprehensive IRG and A-V-H graphs. Lastly, the impact and selection of various language models will be discussed.

5.1 Knowledge Acquisition Pipeline

The synthetic example of the risk-related report contains a description of the behavior of the fuel and its impact on the engine and aircraft (Fig. 5.1). Although they are not explicitly expressed, there is a logical connection indicating that potential risk interaction between *"water"* and *"fuel"* in the context of the *"airplane"* exists. The SFG decomposition of the description (Fig. 5.2) confirms that such a path exists:

"water" → "supercooled water droplets collide with a surface" → "if supercooled water droplets collide with a surface they may result in blocked fuel inlet pipes" → "blocked fuel inlet pipes" → "aviation fuel designed for use in aircraft powered by gas turbine engines" → "engine".

It also shows that the path is relevant from a risk interaction perspective, i.e., “water” - “collide” - “with a surface,” which means that both water and surface affect each other. There is a similar relationship between “droplets” and “surface”. In the second element, the relationship between “droplets” - “results” - “pipes” is also a “has effect on” relationship between “droplets” and “pipe”. The “blocked fuel inlet pipes” are connected to the last frame, where it is possible to designate a relationship between “engine” - “use” - “fuel”. By reading the entire path, it is also possible to confirm that, in terms of consistency and

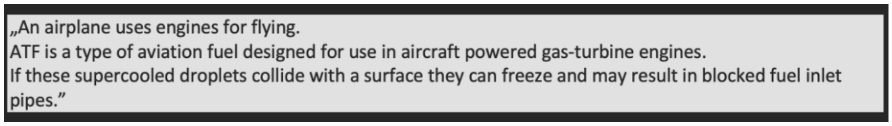

„An airplane uses engines for flying.
ATF is a type of aviation fuel designed for use in aircraft powered gas-turbine engines.
If these supercooled droplets collide with a surface they can freeze and may result in blocked fuel inlet pipes."

Figure 5.1: ATF fuel description

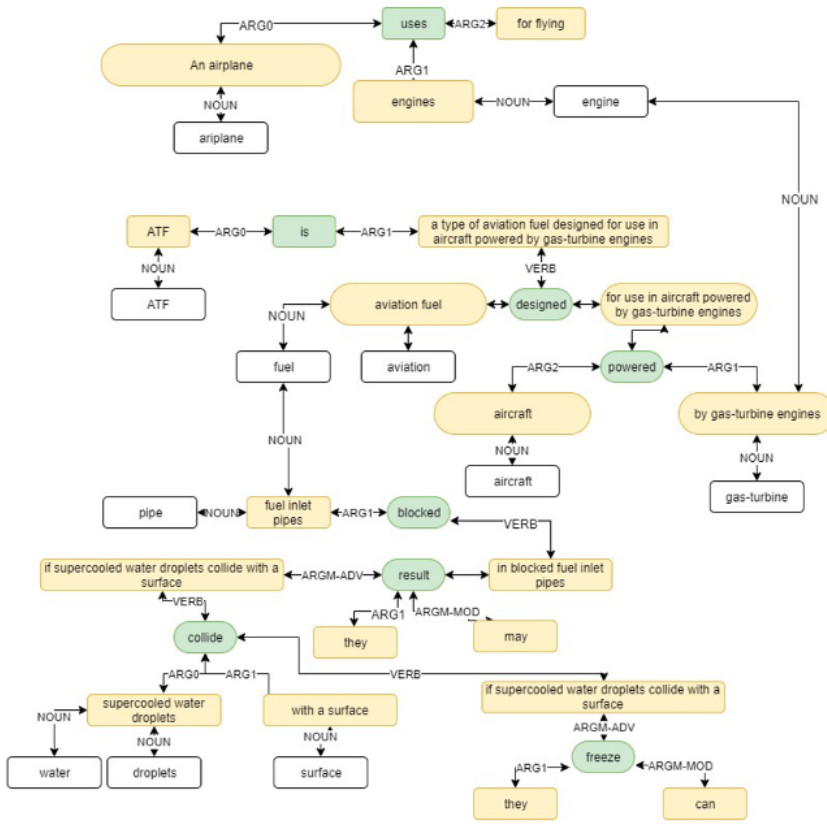


Figure 5.2: ATF fuel SFG decomposition

general operational safety, the "droplet" (Hazard) affects the "engine" (Asset) through "fuel" (Vulnerability).

The synthetic template used to detect the risk flow in the system will be "has an effect on". Such relation meets the requirement for a non-reflexive relation (as the system can hardly have an effect on itself), a transitive one as "effect flows" through the system. The effect flow is antisymmetric, as the direction of the flow must be supported explicitly by the relation context.

All specified metapaths perform the SFG graph (fig. 5.3):

- isRTE: indicate that metapath 1 and metapath 2 and these are intra-frame cases,
- isDialogRTE: indicate metapth 3a and 3b, which are inter-frame cases,
- isDialogRTE2: indicate metapaht4, a subordinate clause decomposition case.

The synthetic example shows that most of the relations evaluated were *isRTE* (over

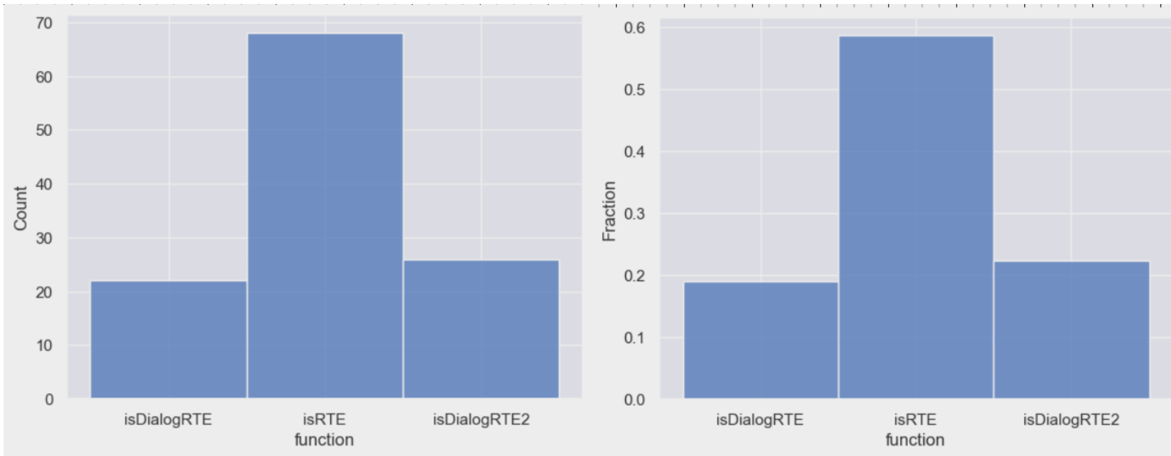


Figure 5.3: Distribution of relations per detection function assigned to a metapath

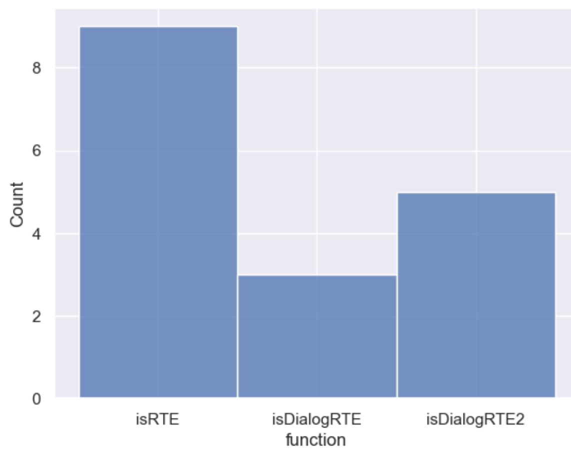


Figure 5.4: Distribution of **valid** relations per detection function assigned to a metapath

50%), meaning they were identified within a single frame together. However, dialog-based strategies denoted as *isDialogRTE* for two adjacent frames and *isDialogRTE2* for subordinate clause decomposition, all together add a significant part of them (Fig. 5.3). In many cases, they are irrelevant as their dialog coherence score is low (Fig. 5.5), but not considering them at all would deteriorate the risk model significantly as they are a significant part of confirmed relations (Fig. 5.4). A multicriterial optimization performed for chatGPT 3.5 LLM (Eq. 4.3) establishes the treashold $w_{th} = .18$ and relations accepted will have dialog scores above it (Fig. 5.7). It is worth noting that the prompted chatGPT 3.5 itself does not provide perfect decisions, and solely relying on the LLM's Yes/No answer will not produce a trustworthy risk model.

source	target	function	GPT Decision	weight
('airplane', 'noun')	('aviation', 'noun')	isDialogRTE	Yes	0.5047957187956318
('airplane', 'noun')	('fuel', 'noun')	isDialogRTE	Yes	0.5047957187956318
('airplane', 'noun')	('type', 'noun')	isDialogRTE	No	0.5688557397630128
('airplane', 'noun')	('use', 'noun')	isDialogRTE	No	0.0
('airplane', 'noun')	('gas', 'noun')	isDialogRTE	No	0.0
('airplane', 'noun')	('aircraft', 'noun')	isDialogRTE	Yes	0.0
('airplane', 'noun')	('turbine', 'noun')	isDialogRTE	No	0.0
('airplane', 'noun')	('engine', 'noun')	isRTE	No	0.7587770819664001

Figure 5.5: Sample of relations detected for the 'airplane'

source	target	path	verbalizer	score	prompt results
droplet	pipe	these supercooled droplets collide with a surface. If these supercooled droplets collide with a surface they may result in blocked fuel inlet pipes.	droplet has an effect on pipe	0.7	No
airplane	fuel	An airplane uses engines for flying.aviation fuel designed for use in aircraft powered gas turbine engines.	airplane has effect on fuel	0.5	No
fuel	airplane	An airplane uses engines for flying.aviation fuel designed for use in aircraft powered gas turbine engines.	fuel has effect on airplane	0.9	Yes
fuel	pipe	If these supercooled droplets collide with a surface they may result in blocked fuel inlet pipes	fuel has effect on pipe	0.6	Yes
pipe	fuel	If these supercooled droplets collide with a surface they may result in blocked fuel inlet pipes	pipe has effect on fuel	0.6	Yes

Figure 5.6: Examples of classification decisions

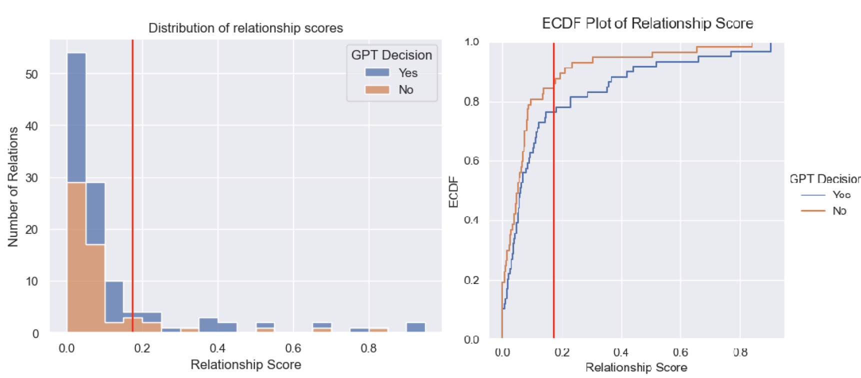


Figure 5.7: The distributions of chatGPT prompted decision together with the threshold (red vertical line) (left). Empirical cumulative distribution of relations wrt dialog function score (right)

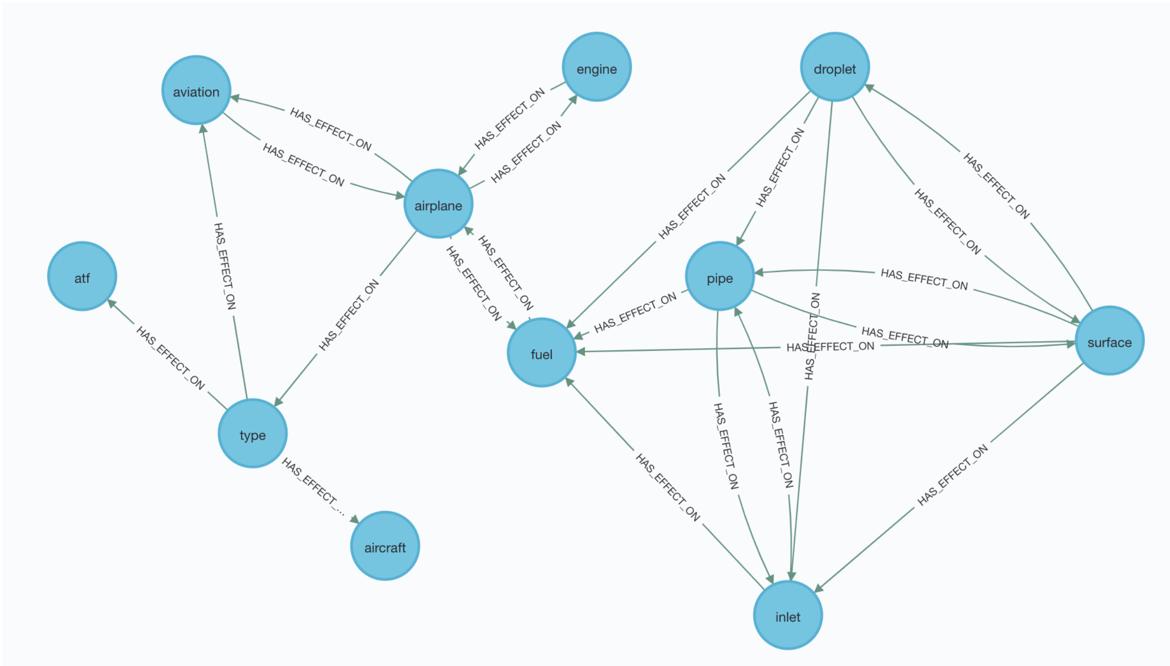


Figure 5.8: The IRG for the synthetic example

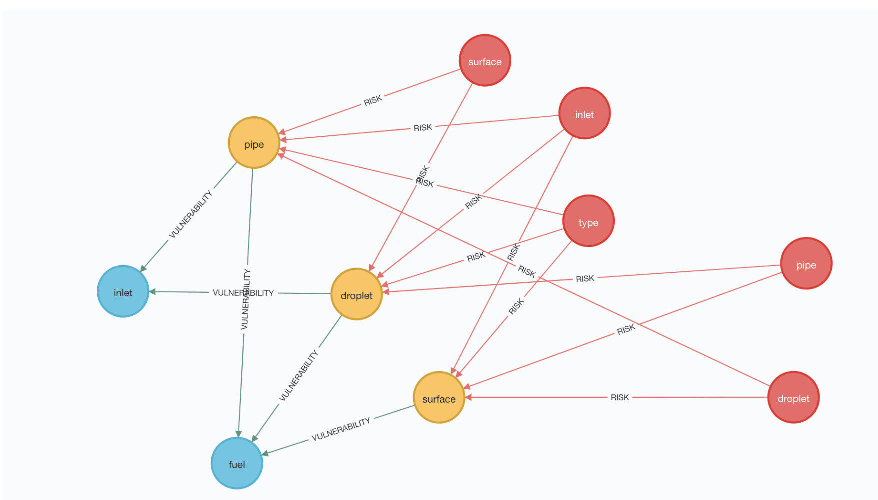


Figure 5.9: The A-V-H graph for the synthetic example.Blue nodes: Assets, Orange: Vulnerabilities, Red: Risks

A spark can cause fuel to ignite.
Low temperature and high altitudes cause water vapour to create droplets.

Figure 5.10: Additional contextual information for the synthetic examples

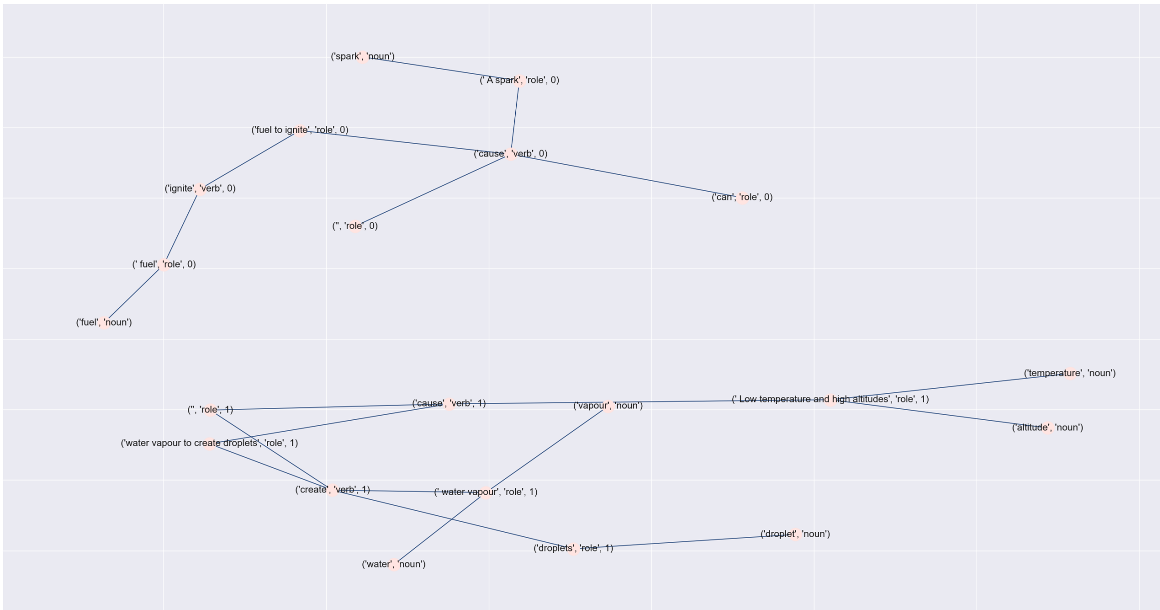


Figure 5.11: SFG decomposition of additional synthetic context

Incremental Knowledge Acquisition Pipeline

The synthetic example can be expanded with additional contextual information. This simulates the scenario when the risk model is constructed gradually once new descriptions are available. A new SFG graph is constructed for a new description (Fig. 5.11). The IRG graph is *augmented* with incoming new relations. Eventually the A-V-H graph is updated with new interactions.

The augmented A-V-H graph shows that *fuel* Vulnerability becomes a hub that connects with eight risks in the context of the *airplane* Asset.

Adding additional contextual information enriches the A-H-V graph further. For example, adding elements directly impacted by fuel, like the APU, air conditioning, electricity, navigation instruments, and others, is possible. The centrality analysis of Vulnerability nodes can help identify areas requiring specific precautions given connected risks [1].

5.2 Impact of Large Langauge Models

As mentioned in the introduction, a few LLMs are explicitly trained in a dedicated domain, i.e. medical or legal. Unfortunately, the risk analysis domain does not have enough

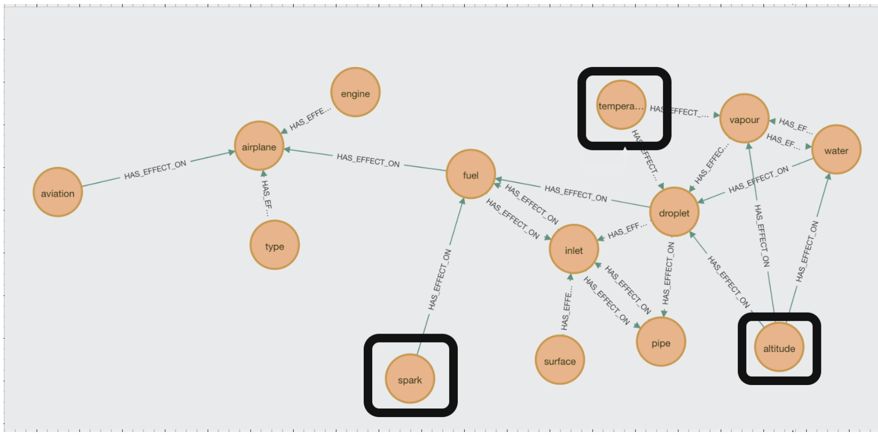


Figure 5.12: Augmented IRG Graph for the synthetic example. New connections are highlighted.

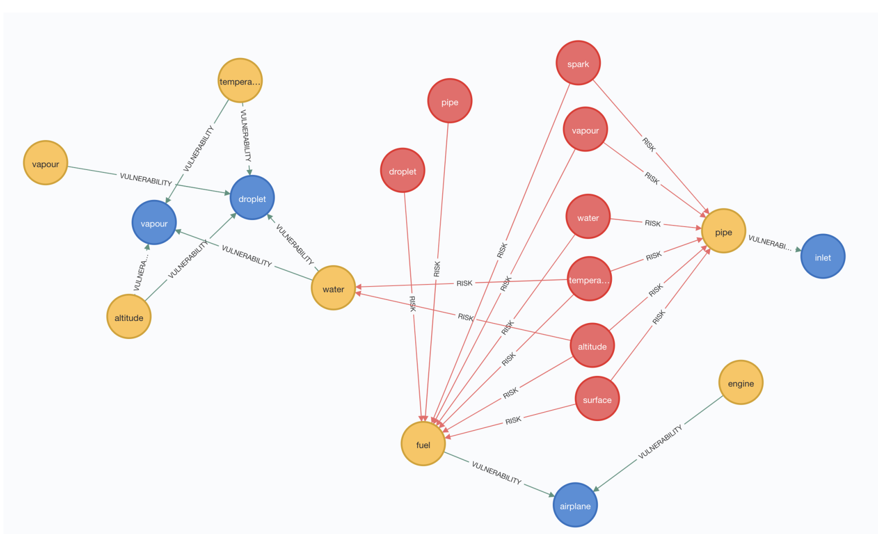


Figure 5.13: Augmented A-V-H Graph indicating new Risks: "temperature" and "altitude"

textual resources to perform such comprehensive training. On the other hand, LLMs e.q.: FLAN [56], [81], GPT3 [14] or Mistral [82] have achieved remarkable success lately establishing new reference results in a variety of general NLP tasks i.e. MMLU [1]. LLMs have achieved such results following the latest in-context learning approaches such as instruction-tuning [56], chain-of-thoughts [83], or self-consistency [84]. The performance of the LLM is also thanks to good quality textual resources such as Wikipedia and thousands of books used in their training.

Therefore, a selection of LLMs has been evaluated for entailment score threshold cutoff. The threshold selection approach selects the cutoff for the entailment score so that the distance between the ratios of rejected and accepted relations is the largest. Therefore, it will select the point on the weight axes for which the distance between empirical distributions (ECDF) of rejected (blue curves) and accepted relations (orange curves) are as far as possible (Figs. 5.19, 5.21, 5.23). The statistics suggests that the vanilla models' performance (promoted in a general way, not specifically for the task) varies significantly. Therefore, a valid question arises: which one to choose, and how does LLM performance relate to the RTE classification performed by the other classifier?

The experiment setup was as follows. The narrative described the Teton Dam collapse in 1975 (Fig. 5.18). The text was decomposed to the SFG representation. A single relation template "has a direct impact on" was used, and metapaths 1 and 2 were executed only to remove the effect of dialog consistency. The statistics (aggregated counts and empirical distribution) as a function of template entailment score - denoted as weight- were collected for LLMs (FLAN, LLAMA, Mistral) available on huggingface.com. Each model was prompted according to its interface with precisely the same prompt as others.

Together with the ECDFs, the distributions of LLM's *Yes* / *No* decisions are plotted against the entailment score for each model (Figs. 5.20, 5.22, 5.23, 5.25. The cutoff threshold for each model is provided in Fig. (5.26). It is impossible to assume that the threshold with a higher value is better. Therefore, we have to analyze the performance of each model and select the one based on the analysis, not the threshold value itself.

The semantic frame sentence decomposition and the SFG representation suggest an approach. We can rely on the following observations to decide which LLM better handles the relation verbalization validation task:

- the overall distribution of the number of relations per weight is exponential (Fig. 5.14)
- Most LLMS rejections shall be in low-scoring areas as most frames connecting nouns

[1] https://paperswithcode.com/sota/multi-task-language-understanding-on-mmlu

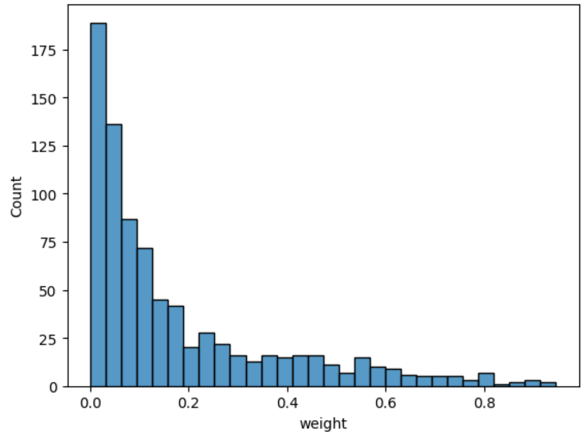


Figure 5.14: Histogram of relations given entitlement score (weight) for the exemplary narrative (Fig. 5.18)

will not verbalize the relation template. In the example for the jetliner, only two pairs (*bird* - *jetliner*) and (*bird* - *engine*) would be connected by a template *has a direct impact of* (Fig. 5.17),

- the LLMS should agree with the RTE classifier therefore, the acceptance probability should be correlated with the weight and increase with it.(Fig. 5.15). In the perfect scenario, all rejected relations shall be at 0 and accepted at 1.

Therefore, the rudimentary criterion for selecting the performing model would be the difference d between the mean value of the score for acceptance and the mean value for rejection. Let w^0 denote the entitlement score (weight) associated with the LLM's rejection of the transaction and w^1 the acceptance, then the distance

$$d = E[w^1] - E[w^0]$$

will denote the quality of LLMs decisions:

- if d>0, the LLM correlates with the RTE and properly discerns the rejections and acceptances. Solving the optimization task (Eq. 4.3) will establish the threshold,
- if d=0, the LLM is independent of RTE and cannot be used to estimate the threshold (Fig. 5.16),
- if d<0, the LLM is negatively correlated to the RTE and cannot be used to estimate the threshold (Fig. 5.19).

The comparison of differences between the means for the models analyzed (Fig. 5.27) suggests using the mistral-v0.1 model. The acceptance threshold for this model has been calculated and is 0.273 (Fig. 5.26)

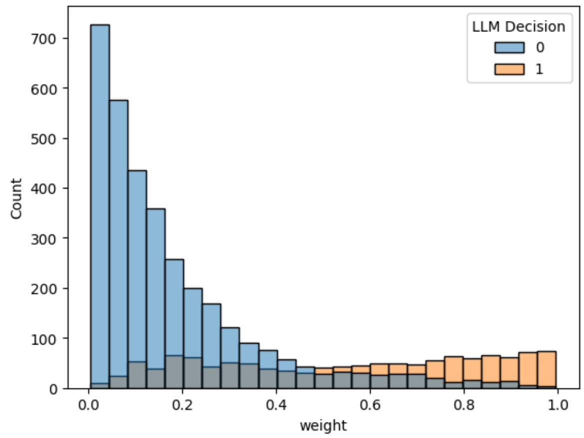


Figure 5.15: Ideal histogram of exponential distribution of the number of relations given entailment score (weight) where class assignment correlates with entitlement score (weight)

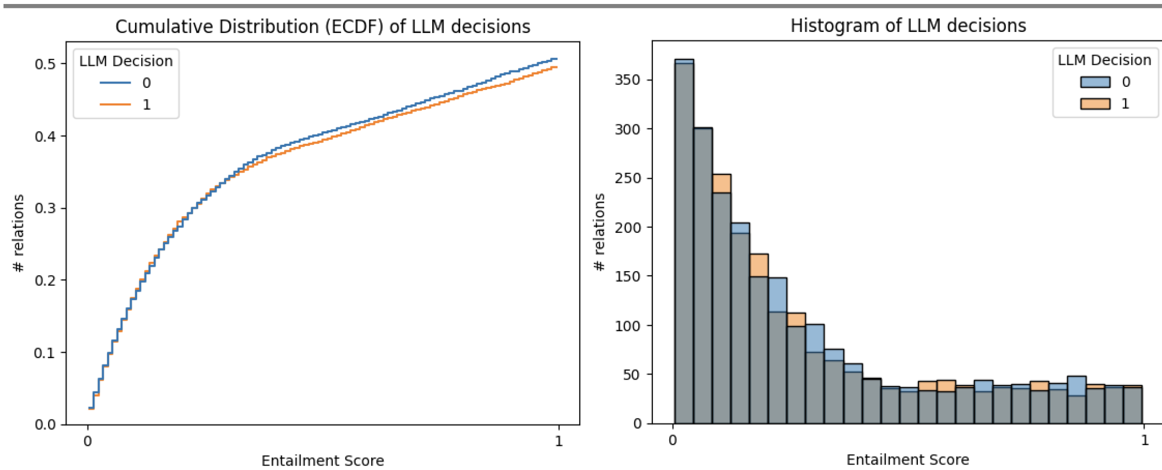


Figure 5.16: Synthetic empirical distribution (left) and histogram (right) of the case where LLM classification is uncorrelated with the entailment score

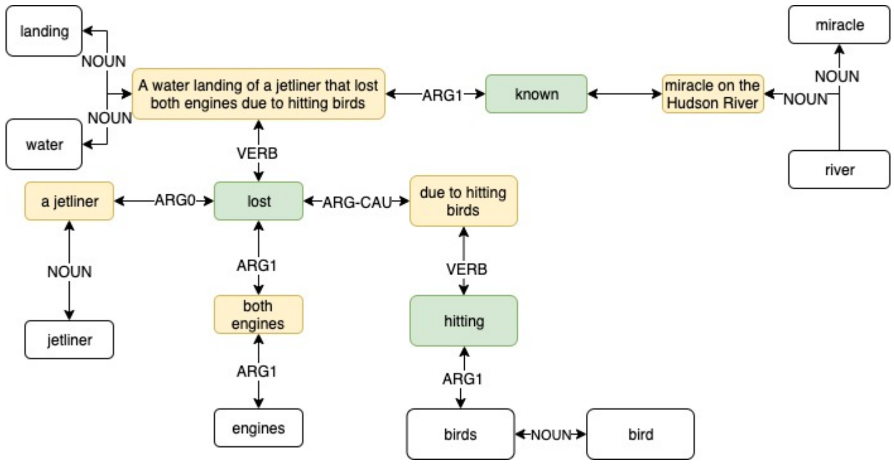


Figure 5.17: SFG Hierarchy of Frames

5.3 Examples

The proposed solution can construct that A-V-H in various scenarios where it is possible to detect the risk-related relation and propagation of hazard.

5.3.1 Financial Scenario

In this scenario, the description of price increase, i.e., inflation, is analyzed for potential impact on other components of the economy. The flow of inflation impact is like hazard propagation in the description of the system. The generated narrative (Fig. 5.28) describes the context surrounding the effect of inflation on the economy. In the IRG graph, we can represent the closest neighborhood of *inflation* (Fig. 5.29).

5.3.2 Medical Scenario

In the medical scenario, we can model the propagation of a drug's effect. Identifying such indirect impact can help reduce the side effects of therapies. The sample text describes the role of the liver in the human body (Fig. 5.31). Assume therapy impacts the liver directly, and the goal is to identify potentially impacted elements of the human body. Provided

5.4 Summary

The main research goal was to propose a method for constructing a Knowledge Acquisition Pipeline that builds an A-V-H representation of the flow of hazards in the system

Teton Canyon ends about six miles below the dam site, where the river flows onto the Snake River Plain. When the dam failed, the flood struck several communities immediately downstream, particularly Wilford at the terminus of the canyon, Sugar City, Salem, Hibbard, and Rexburg. Thousands of homes and businesses were destroyed.

The small agricultural communities of Wilford and Sugar City were wiped from the river bank. Five of the eleven deaths attributed to the flood occurred in Wilford.

The similar community of Teton, on the south bank of the river, is on a modest bench and was largely spared. One Teton resident was fishing on the river at the time of the dam failure and was drowned. An elderly woman living in the city of Teton died as a result of the evacuation. One estimate placed damage to Hibbard and Rexburg area, with a population of about 10,000, at 80% of existing structures. The Teton River flows through the industrial, commercial, and residential districts of north Rexburg. A significant reason for the massive damage in the community was the location of a lumber yard directly upstream. When the flood waters hit, thousands of logs were washed into town. Dozens of logs hit a bulk gasoline-storage tank a few hundred yards away. The gasoline ignited and sent flaming slicks adrift on the racing water.The force of the logs and cut lumber and the subsequent fires practically destroyed the city.The flood waters traveled west along the route of the Henrys Fork of the Snake River, around both sides of the Menan Buttes, significantly damaging the community of Roberts. The city of Idaho Falls, even further down on the flood plain, had time to prepare. At the older American Falls Dam downstream, engineers increased discharge by less than 5% before the flood arrived. That dam held and the flood was effectively over, but tens of thousands of acres of land near the river were stripped of fertile topsoil. The force of the failure destroyed the lower part of the Teton River, washing away riparian zones and reducing the canyon walls.This seriously damaged the stream's ecology and impacted the native Yellowstone cutthroat trout population. The force of the water and excessive sediment also damaged stream habitat in the Snake River and some tributaries, as far downstream as the Fort Hall bottoms. In August 1975, the region experienced an extreme flood, resulting in quantities of water falling that had not been considered during the construction of the dam. More than a year's worth of rain fell in only 24 hours, and the dam failed on August 8.

Early on August 8, the dam was breached and 700 million cubic meters of floodwater was released, flooding communities and homes downstream. After this burst, a chain reaction began and the other 61 reservoirs located in the area collapsed—releasing another six billion cubic meters of floodwater. The water covered an area equal to 10 000 square kilometers.

The failure to plan and account for extreme floodwaters resulted in the immediate death of 26 000 as a result of the water itself. 145 000 more people died as a result of epidemics and famine following the flood.The cause of the incident was reported to be unsafe vibrations coming from one of the turbines, which caused the turbine to break apart violently.

Water that had been entering the turbine, flooded the turbine hall—flooding the room and levels below. The ceiling of the hall also broke apart from the impact from the turbine.

At this point, power failed in the power station resulting in a blackout.

Gasoline can cause fires.Steel gates to the water intake pipes of the turbines were manually closed and spillways were opened to prevent more damage.

Figure 5.18: Teton Dame Collapse Description

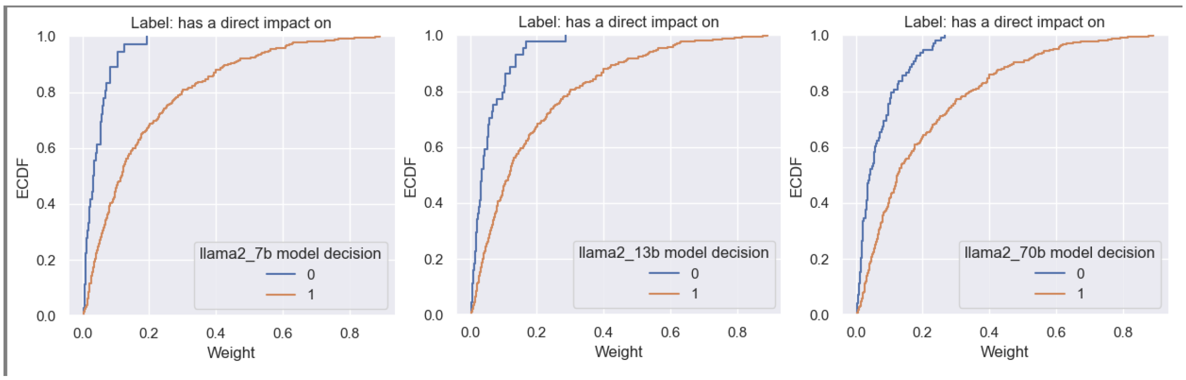


Figure 5.19: Emprical Distribution (ECDF) of LLama models per entailment score

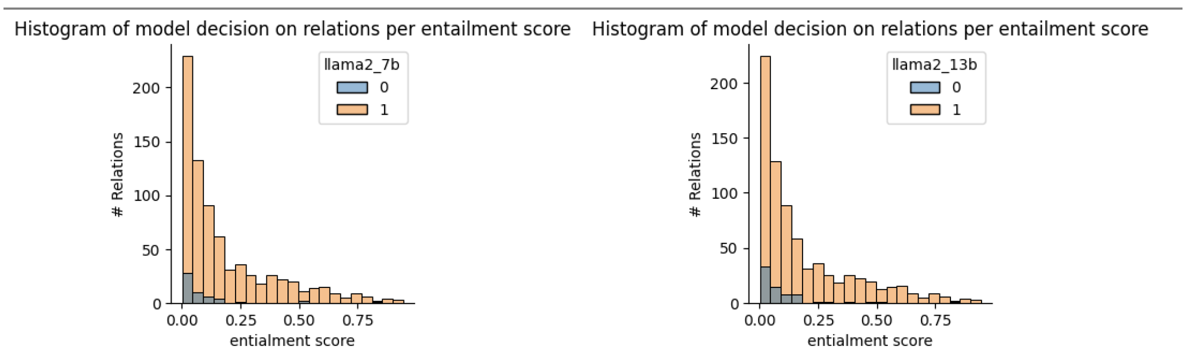


Figure 5.20: Histograms of model's decisions on the relation per entailment score

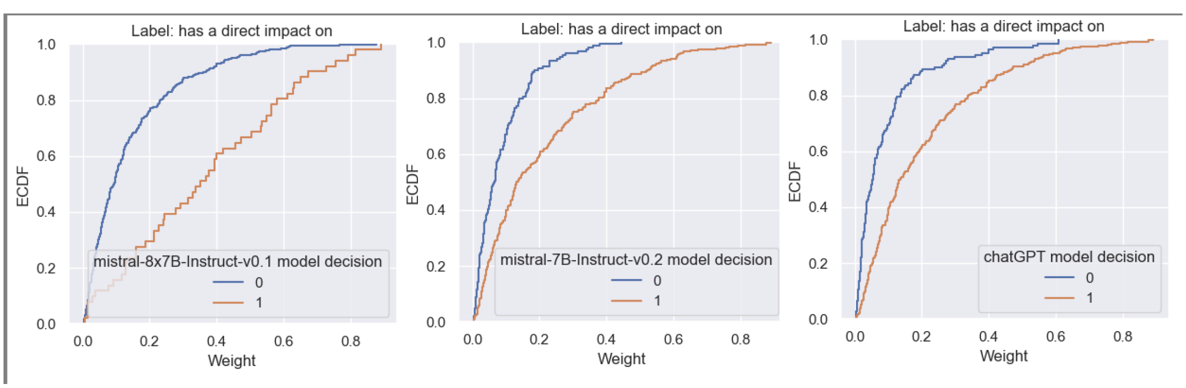


Figure 5.21: Emprical Distribution (ECDF) of Mistral and chatGPT models per entailment score

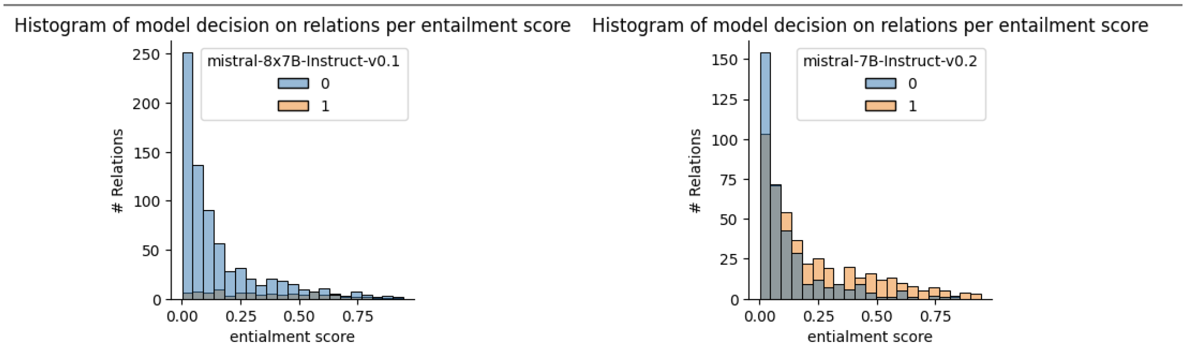


Figure 5.22: Histograms of model's decisions on the relation per entailment score

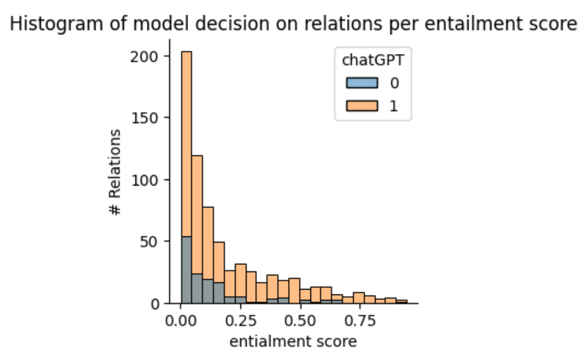


Figure 5.23: Histograms of model's decisions on the relation per entailment score

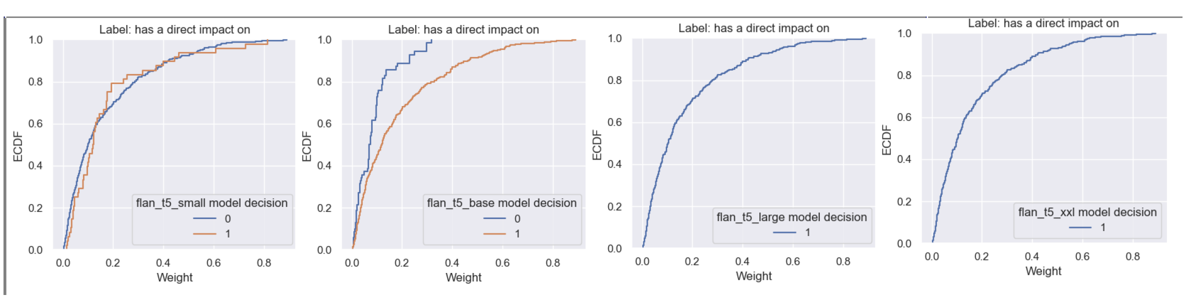


Figure 5.24: Emprical Distribution (ECDF) of FLAN models per entailment score

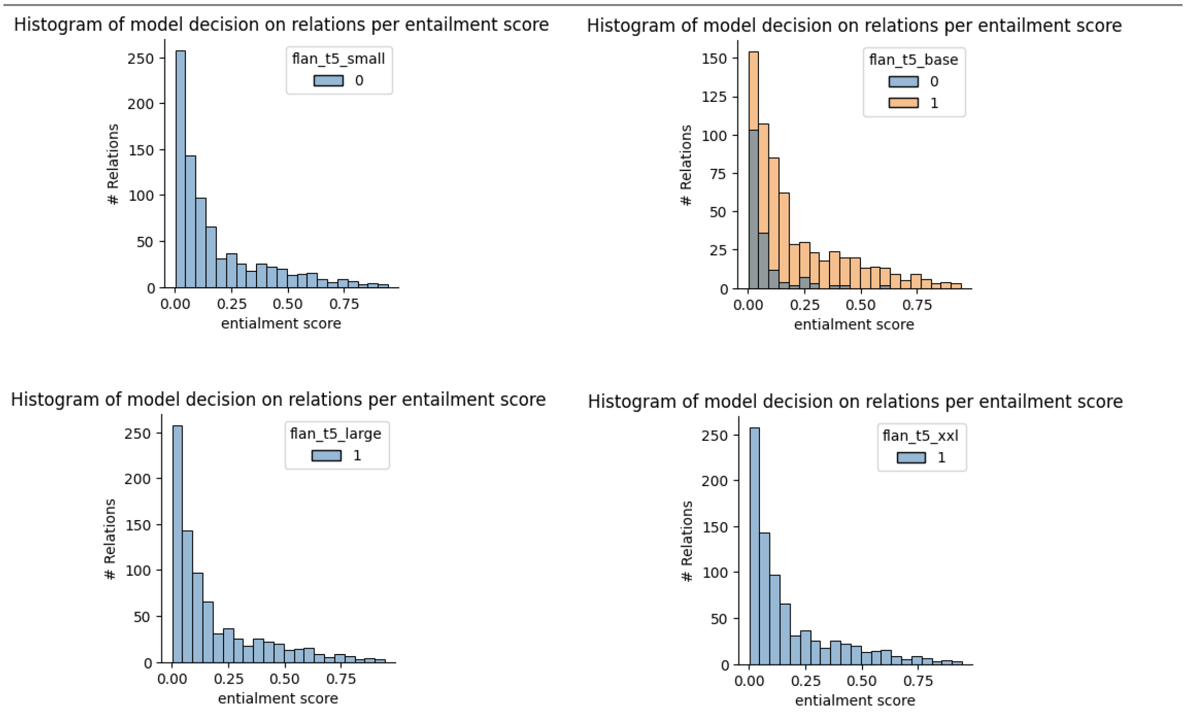


Figure 5.25: Histograms of model's decisions on the relation per entailment score

model	relation	treshold
mistral-8x7B-Instruct-v0.1	has a direct impact on	0,273590982
chatGPT	has a direct impact on	0,213499486
mistral-7B-Instruct-v0.2	has a direct impact on	0,183974847
llama2_13b	has a direct impact on	0,158783749
llama2_7b	has a direct impact on	0,158696875
flan_t5_base	has a direct impact on	0,106075577

Figure 5.26: The cutoff thresholds for each LLMs

model	distance
mistral-8x7B-Instruct-v0,1	0,207326455
flan_t5_base	0,137173552
mistral-7B-Instruct-v0,2	0,106743931
llama2_13b	0,094051495
llama2_7b	0,092510897
chatGPT	0,059997983

Figure 5.27: Calculated distances between expected values for accepted and rejected relations per LLM

Consumer price increase, often referred to as inflation, is the general increase in the prices of goods and services over time. It means that, on average, the cost of purchasing goods and services rises, leading to a decrease in the purchasing power of money. Demand-Pull Inflation occurs when demand for goods and services exceeds supply. When consumers have more money to spend or when there is increased demand due to factors like economic growth or government stimulus, businesses may raise prices to capitalize on the higher demand.Cost-Push Inflation happens when the cost of production for goods and services increases. Factors such as rising wages, increased raw material costs, or higher taxes can push up production costs, prompting businesses to pass these costs onto consumers through higher prices. Central banks control inflation by adjusting interest rates and the money supply. If a central bank increases the money supply excessively or keeps interest rates too low for too long, it can lead to higher inflation as more money chases the same amount of goods and services. Disruptions in the supply chain, such as natural disasters, geopolitical tensions, or pandemics, can lead to shortages of certain goods and services. When supply is limited and demand remains constant or increases, prices tend to rise. As prices rise, the same amount of money buys fewer goods and services, reducing the purchasing power of consumers' incomes.

Figure 5.28: Financial Example: Description of inflation

Figure 5.29: Financial Scenario: the "inflation" neighborhood in the IRG.

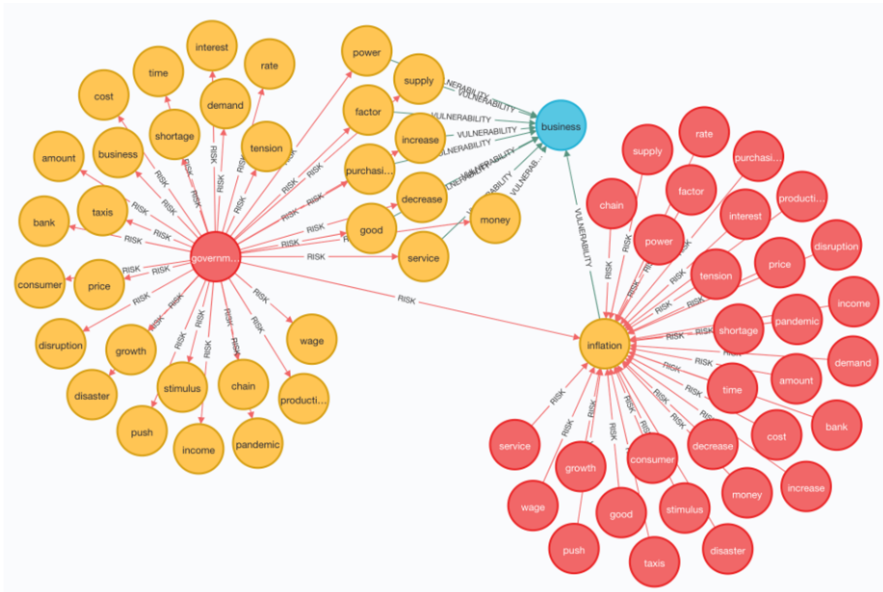


Figure 5.30: Financial Scenario: the A-V-H graph for "inflation" as Vulnerability of "business" Asset

The liver is responsible for metabolizing nutrients from the food we eat, including carbohydrates, fats, and proteins. The liver converts glucose into glycogen for storage and releases it when blood sugar levels drop. Additionally, the liver metabolizes fats and produces bile, which aids in fat digestion and absorption in the intestines. Moreover, the liver detoxifies harmful substances, such as drugs, alcohol, and metabolic waste products, by breaking them down and facilitating their excretion from the body. The liver synthesizes albumin, which helps maintain blood volume and pressure. The liver synthesizes clotting factors, such as fibrinogen, prothrombin. The liver serves as a storage site for various nutrients and vitamins. It stores glycogen, which can be converted back into glucose when the body needs energy. The liver stores vitamins A, D, and B12, as well as iron and copper, which are essential for various metabolic processes. The liver stores excess glucose as glycogen during periods of high blood sugar and releases glucose into the bloodstream when blood sugar levels drop, ensuring that cells have a constant supply of energy. The liver produces bile, a greenish-yellow fluid that helps emulsify fats and facilitate their digestion and absorption in the small intestine. Bile is stored in the gallbladder and released into the small intestine when needed to aid in the digestion and absorption of dietary fats and fat-soluble vitamins. The liver plays a crucial role in the body's immune system by filtering and removing bacteria, viruses, and other pathogens from the bloodstream. The liver synthesizes various hormones and cholesterol necessary for maintaining hormonal balance and cell membrane integrity such as insulin-like growth factor 1 (IGF-1). Fibrinogen is a precursor to fibrin, which is the main protein component of blood clots.

Figure 5.31: Medical Example: Description of the role of a liver

description.

Chapter 6

Conclusions

The method described has demonstrated the following characteristics. First, it detects intra-sentence relations without training sets through a dialog consistency and verbalizing relationship pattern. Second, it shows that defining transitive relations applies to the risk analysis domain, as it is possible to construct the A-V-H, which is a risk-specific interaction graph. Third, we show that in the absence of training sets, a prompt-based classification using a language model can be used to provide a validation method. The proposed solution addresses the contextual entity classification problem and can be used to construct a comprehensive risk representation incrementally once new narratives are available.

Although there is a breadth of work on how current language models encode "knowledge" and how they can be used to either extract it directly or validate the engineered hypothesis, limitations still hinder their direct applicability in various areas, including risk analysis. Hallucinations pose the greatest difficulty in their practical application. It seems that the relationship detection and classification task, quoted in this paper as the Verbilizer, although being the most straightforward use case as it requires only properly defined prompt [45], [85] does not provide definite decisions as there are accepted relations that are significantly below the acceptance threshold. It seems reasonable to fuse such classification with methods working on different principles.

Another limitation is the fixed size of the context window. In in-context few-shot training, the context window would be used to provide system descriptions. Therefore, descriptions not fitting in would have to be split, resulting in possible lost relations. In addition, it is unclear how expanding the context window impacts the detection performance. It is also unclear how prompts would have to be constructed to solve the contextual representation classification (Risk-Asset-Vulnerability Dilemma). It seems that some sort of intermediate graph representation of text, similar to the SFG graph, which provides a method to limit the context (to a path of defined length) together with a specific graph traversing strategy, like metapaths, can help to fuse distant text fragments which would not fit into context if a description is provided directly. Linking prompts with a graph traversing strategy can also provide a method to improve the relation classification.

We want to continue research in the following directions. First, we would like to focus on the capabilities of the SFG graph to check if it is possible to include more risk-oriented text operations. We want to augment coreference resolution to go beyond simple preposition-noun or similar noun substitution to perform contextual substitution for attributes of frames expressing the same event. For example, the sentences *"Electricity cutoff disables all electronic devices. In such situations, emergency UPSs provide backup power supply."*, *situation* refers to *electricity cutoff*. In the SFG graph, these sentences are unconnected as no nouns are in common. Therefore, either a specific connection or a complete replacement of *situation* with *electricity cutoff* should be performed to reduce the graph distance between *electronic devices* and *emergency UPS*.

Second, the IRG represents a single transitive relation, simplifying real-world scenarios. We want to establish a methodology that would include additional relation types. In our example, we mention *ATF*, a type of *fuel*, which is not expressed as *type-of* relationship in the IRG. In fact, ATF is represented as a completely separate node. We hope that adding more relation types and specifying how the IRG graph is traversed will bring even more clarity to risk representation and help construct a comprehensive risk Knowledge Graph.

Third, we would like to explore if it is possible to construct graphical representations of risk interaction other than A-V-H. We used *has-effect* template only. We would like to verify if *path algebra* [86] can be applied to risk analysis. We believe we could model Event-Tree-Analysis by adding different templates. By combining them to detect if events shall co-occur to propagate hazard, we could create Fault-Tree-Analysis models [4].

Bibliography

[1] A. Walczak, J. Napiórkowski, P. Adamczyk, and G. Kiryk, "Network model of risk analysis in the technical structures," in *21st International Conference on Circuits, Systems, Communications and Computers (CSCC 2017)*, 2017.

[2] I. A. E. Agency, *Convention on nuclear safety*. International Atomic Energy Agency, 1994.

[3] I. E. Comission, "Analysis techniques for system reliability – procedure for failure mode and effects analysis, iec 60812," 2018.

[4] M. Rausand and S. Haugen, *Risk Assessment: Theory, Methods, and Applications*. WileySons, 2020.

[5] T. A. Kletz, *Hazop Hazan: Identifying and Assessing Process Industry Hazards*. CRC Press, 1999.

[6] F. Crawley, *A Guide to Hazard Identification Methods*. Elsevier, 2020.

[7] C. Liu and S. Yang, "Using text mining to establish knowledge graph from accident/incident reports in risk assessment," *Expert Systems with Applications*, vol. 207, p. 117991, 2022.

[8] F. Simone, S. M. Ansaldi, P. Agnello, and R. Patriarca, "Industrial safety management in the digital era: Constructing a knowledge graph from near misses," *Computers in Industry*, vol. 146, p. 103849, 2023.

[9] V. Yadav and S. Bethard, "A survey on recent advances in named entity recognition from deep learning models," in *Proceedings of the 27th International Conference on Computational Linguistics* (E. M. Bender, L. Derczynski, and P. Isabelle, eds.), (Santa Fe, New Mexico, USA), pp. 2145–2158, Association for Computational Linguistics, Aug. 2018.

[10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds.), vol. 30, Curran Associates, Inc., 2017.

[11] C. D. Manning, "Human Language Understanding amp; Reasoning," *Daedalus*, vol. 151, pp. 127–138, 05 2022.

[12] R. Bommasani, D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, M. S. Bernstein, J. Bohg, A. Bosselut, E. Brunskill, E. Brynjolfsson, S. Buch, D. Card, R. Castellon, N. Chatterji, A. Chen, K. Creel, J. Q. Davis, D. Demszky, C. Donahue, M. Doumbouya, E. Durmus, S. Ermon, J. Etchemendy, K. Ethayarajh, L. Fei-Fei, C. Finn, T. Gale, L. Gillespie, K. Goel, N. Goodman, S. Grossman, N. Guha, T. Hashimoto, P. Henderson, J. Hewitt, D. E. Ho, J. Hong, K. Hsu, J. Huang, T. Icard, S. Jain, D. Jurafsky, P. Kalluri, S. Karamcheti, G. Keeling, F. Khani, O. Khattab, P. W. Koh, M. Krass, R. Krishna, R. Kuditipudi, A. Kumar, F. Ladhak, M. Lee, T. Lee, J. Leskovec, I. Levent, X. L. Li, X. Li, T. Ma, A. Malik, C. D. Manning, S. Mirchandani, E. Mitchell, Z. Munyikwa, S. Nair, A. Narayan, D. Narayanan, B. Newman, A. Nie, J. C. Niebles, H. Nilforoshan, J. Nyarko, G. Ogut, L. Orr, I. Papadimitriou, J. S. Park, C. Piech, E. Portelance, C. Potts, A. Raghunathan, R. Reich, H. Ren, F. Rong, Y. Roohani, C. Ruiz, J. Ryan, C. Ré, D. Sadigh, S. Sagawa, K. Santhanam, A. Shih, K. Srinivasan, A. Tamkin, R. Taori, A. W. Thomas, F. Tramèr, R. E. Wang, W. Wang, B. Wu, J. Wu, Y. Wu, S. M. Xie, M. Yasunaga, J. You, M. Zaharia, M. Zhang, T. Zhang, X. Zhang, Y. Zhang, L. Zheng, K. Zhou, and P. Liang, "On the opportunities and risks of foundation models," 2022.

[13] J. Kaplan, S. McCandlish, T. Henighan, T. B. Brown, B. Chess, R. Child, S. Gray, A. Radford, J. Wu, and D. Amodei, "Scaling laws for neural language models," 2020.

[14] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, "Language models are few-shot learners," in *Advances in Neural Information Processing Systems* (H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, eds.), vol. 33, pp. 1877–1901, Curran Associates, Inc., 2020.

[15] J. Lai, W. Gan, J. Wu, Z. Qi, and P. S. Yu, "Large language models in law: A survey," 2023.

[16] Y. Li, S. Wang, H. Ding, and H. Chen, "Large language models in finance: A survey," 2023.

[17] K. Singhal, S. Azizi, T. Tu, S. S. Mahdavi, J. Wei, H. W. Chung, N. Scales, A. Tanwani, H. Cole-Lewis, S. Pfohl, P. Payne, M. Seneviratne, P. Gamble, C. Kelly, N. Scharli,

A. Chowdhery, P. Mansfield, B. A. y Arcas, D. Webster, G. S. Corrado, Y. Matias, K. Chou, J. Gottweis, N. Tomasev, Y. Liu, A. Rajkomar, J. Barral, C. Semturs, A. Karthikesalingam, and V. Natarajan, "Large language models encode clinical knowledge," 2022.

[18] A. G. M. Alam, "Semantic role labeling for knowledge graph extraction from text.," in *Progress in Artificial Intelligence*, 2021.

[19] Y. Koreeda and C. Manning, "ContractNLI: A dataset for document-level natural language inference for contracts," in *Findings of the Association for Computational Linguistics: EMNLP 2021*, (Punta Cana, Dominican Republic), pp. 1907–1919, Association for Computational Linguistics, Nov. 2021.

[20] T. R. Gruber, "A translation approach to portable ontology specifications," *Knowledge Acquisition*, vol. 5, no. 2, pp. 199–220, 1993.

[21] R. Studer, V. Benjamins, and D. Fensel, "Knowledge engineering: Principles and methods," *Data Knowledge Engineering*, vol. 25, no. 1, pp. 161–197, 1998.

[22] J. I. Single, J. Schmidt, and J. Denecke, "Knowledge acquisition from chemical accident databases using an ontology-based method and natural language processing," *Safety Science*, vol. 129, p. 104747, 2020.

[23] P. Cimiano, "Ontology learning and population from text - algorithms, evaluation and applications," 2006.

[24] M. Hodkiewicz, J. W. Klüwer, C. Woods, T. Smoker, and E. Low, "An ontology for reasoning over engineering textual data stored in fmea spreadsheet tables," *Computers in Industry*, vol. 131, p. 103496, 2021.

[25] I. O. for Standardization, *Industrial Automation Systems and Integration-Integration of Life-cycle Data for Process Plants Including Oil and Gas Production Facilities.* ISO, 2003.

[26] P. Hughes, R. Robinson, M. Figueres-Esteban, and C. van Gulijk, "Extracting safety information from multi-lingual accident reports using an ontology-based approach," *Safety Science*, vol. 118, pp. 288–297, 2019.

[27] S. M. Ansaldi, P. Agnello, A. Pirone, and M. R. Vallerotonda, "Near miss archive: A challenge to share knowledge among inspectors and improve seveso inspections," *Sustainability*, vol. 13, no. 15, 2021.

[28] Z. Yin, L. Shi, Y. Yuan, X. Tan, and S. Xu, "A study on a knowledge graph construction method of safety reports for process industries," *Processes*, vol. 11, no. 1, 2023.

[29] X. Zhao, H. Yan, and Y. Liu, "Hierarchical multi-label classification for fine-level event extraction from aviation accident reports," 2024.

[30] R. Grishman and B. Sundheim, "Message Understanding Conference- 6: A brief history," in *COLING 1996 Volume 1: The 16th International Conference on Computational Linguistics*, 1996.

[31] L. Ramshaw and M. Marcus, "Text chunking using transformation-based learning," in *Third Workshop on Very Large Corpora*, 1995.

[32] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (J. Burstein, C. Doran, and T. Solorio, eds.), (Minneapolis, Minnesota), pp. 4171–4186, Association for Computational Linguistics, June 2019.

[33] L. Ratinov and D. Roth, "Design challenges and misconceptions in named entity recognition," in *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL-2009)* (S. Stevenson and X. Carreras, eds.), (Boulder, Colorado), pp. 147–155, Association for Computational Linguistics, June 2009.

[34] M. A. Hearst, "Automatic acquisition of hyponyms from large text corpora," in *COLING 1992 Volume 2: The 14th International Conference on Computational Linguistics*, 1992.

[35] M. Honnibal and I. Montani, "spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing." To appear, 2017.

[36] I. Hendrickx, S. N. Kim, Z. Kozareva, P. Nakov, D. Ó Séaghdha, S. Padó, M. Pennacchiotti, L. Romano, and S. Szpakowicz, "SemEval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals," in *Proceedings of the 5th International Workshop on Semantic Evaluation* (K. Erk and C. Strapparava, eds.), (Uppsala, Sweden), pp. 33–38, Association for Computational Linguistics, July 2010.

[37] R. Bunescu and R. Mooney, "A shortest path dependency kernel for relation extraction," in *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, (Vancouver, British Columbia, Canada), pp. 724–731, Association for Computational Linguistics, Oct. 2005.

[38] N. Kambhatla, "Combining lexical, syntactic, and semantic features with maximum entropy models for information extraction," in *Proceedings of the ACL Interactive Poster and Demonstration Sessions*, (Barcelona, Spain), pp. 178–181, Association for Computational Linguistics, July 2004.

[39] B. Rosario and M. Hearst, "Classifying semantic relations in bioscience texts," in *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04)*, (Barcelona, Spain), pp. 430–437, July 2004.

[40] L. B. Soares, N. FitzGerald, J. Ling, and T. Kwiatkowski, "Matching the blanks: Distributional similarity for relation learning," 2019.

[41] Y. Tian, G. Chen, Y. Song, and X. Wan, "Dependency-driven relation extraction with attentive graph convolutional networks," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, (Online), pp. 4458–4471, Association for Computational Linguistics, Aug. 2021.

[42] T. T. Tran, P. Le, and S. Ananiadou, "Revisiting unsupervised relation extraction," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (D. Jurafsky, J. Chai, N. Schluter, and J. Tetreault, eds.), (Online), pp. 7498–7505, Association for Computational Linguistics, July 2020.

[43] I. Dagan, O. Glickman, and B. Magnini, "The pascal recognising textual entailment challenge," in *Machine Learning Challenges. Evaluating Predictive Uncertainty, Visual Object Classification, and Recognising Tectual Entailment* (J. Quiñonero-Candela, I. Dagan, B. Magnini, and F. d'Alché Buc, eds.), (Berlin, Heidelberg), pp. 177–190, Springer Berlin Heidelberg, 2006.

[44] S. R. Bowman, G. Angeli, C. Potts, and C. D. Manning, "A large annotated corpus for learning natural language inference," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, (Lisbon, Portugal), pp. 632–642, Association for Computational Linguistics, Sept. 2015.

[45] O. Sainz, O. Lopez de Lacalle, G. Labaka, A. Barrena, and E. Agirre, "Label verbalization and entailment for effective zero and few-shot relation extraction," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, (Online and Punta Cana, Dominican Republic), pp. 1199–1212, Association for Computational Linguistics, Nov. 2021.

[46] Y. Yao, D. Ye, P. Li, X. Han, Y. Lin, Z. Liu, Z. Liu, L. Huang, J. Zhou, and M. Sun, "DocRED: A large-scale document-level relation extraction dataset," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, (Florence, Italy), pp. 764–777, Association for Computational Linguistics, July 2019.

[47] Q. Huang, S. Zhu, Y. Feng, Y. Ye, Y. Lai, and D. Zhao, "Three sentences are all you need: Local path enhanced document relation extraction," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)* (C. Zong, F. Xia, W. Li, and R. Navigli, eds.), (Online), pp. 998–1004, Association for Computational Linguistics, Aug. 2021.

[48] J. Li, K. Xu, F. Li, H. Fei, Y. Ren, and D. Ji, "MRN: A locally and globally mention-based reasoning network for document-level relation extraction," in *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021* (C. Zong, F. Xia, W. Li, and R. Navigli, eds.), (Online), pp. 1359–1370, Association for Computational Linguistics, Aug. 2021.

[49] C. Quirk and H. Poon, "Distant supervision for relation extraction beyond the sentence boundary," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers* (M. Lapata, P. Blunsom, and A. Koller, eds.), (Valencia, Spain), pp. 1171–1182, Association for Computational Linguistics, Apr. 2017.

[50] F. Christopoulou, M. Miwa, and S. Ananiadou, "Connecting the dots: Document-level neural relation extraction with edge-oriented graphs," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (K. Inui, J. Jiang, V. Ng, and X. Wan, eds.), (Hong Kong, China), pp. 4925–4936, Association for Computational Linguistics, Nov. 2019.

[51] D. Wang, W. Hu, E. Cao, and W. Sun, "Global-to-local neural networks for document-level relation extraction," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (B. Webber, T. Cohn, Y. He, and Y. Liu, eds.), (Online), pp. 3711–3721, Association for Computational Linguistics, Nov. 2020.

[52] M. Schlichtkrull, T. N. Kipf, P. Bloem, R. van den Berg, I. Titov, and M. Welling, "Modeling relational data with graph convolutional networks," 2017.

[53] J. Li, Y. Wang, S. Zhang, and M. Zhang, "Rethinking document-level relation extraction: A reality check," 2023.

[54] O. Levy, M. Seo, E. Choi, and L. Zettlemoyer, "Zero-shot relation extraction via reading comprehension," in *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)* (R. Levy and L. Specia, eds.), (Vancouver, Canada), pp. 333–342, Association for Computational Linguistics, Aug. 2017.

[55] F. Petroni, T. Rocktäschel, S. Riedel, P. Lewis, A. Bakhtin, Y. Wu, and A. Miller, "Language models as knowledge bases?," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (K. Inui, J. Jiang, V. Ng, and X. Wan, eds.), (Hong Kong, China), pp. 2463–2473, Association for Computational Linguistics, Nov. 2019.

[56] J. Wei, M. Bosma, V. Y. Zhao, K. Guu, A. W. Yu, B. Lester, N. Du, A. M. Dai, and Q. V. Le, "Finetuned language models are zero-shot learners," 2022.

[57] Z. Wan, F. Cheng, Z. Mao, Q. Liu, H. Song, J. Li, and S. Kurohashi, "GPT-RE: In-context learning for relation extraction using large language models," in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (H. Bouamor, J. Pino, and K. Bali, eds.), (Singapore), pp. 3534–3547, Association for Computational Linguistics, Dec. 2023.

[58] G. Li, P. Wang, and W. Ke, "Revisiting large language models as zero-shot relation extractors," 2023.

[59] I. Dagan, D. Roth, M. Sammons, and F. M. Zanzotto, *Recognizing Textual Entailment: Models and Applications*. Synthesis Lectures on Human Language Technologies, Morgan & Claypool Publishers, 2013.

[60] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," 2019.

[61] A. Williams, N. Nangia, and S. R. Bowman, "A broad-coverage challenge corpus for sentence understanding through inference," 2018.

[62] C. Fillmore, "Scenes-and-frames semantics," *Linguistic Structures Processing*, 1977.

[63] M. Palmer, D. Gildea, and P. Kingsbury, "The Proposition Bank: An annotated corpus of semantic roles," *Computational Linguistics*, vol. 31, no. 1, pp. 71–106, 2005.

[64] V. P. A. Gangemi, "Towards a pattern science for the semantic web.," in *Semantic Web Journal*, IOS Press, 2010.

[65] D. Davidson, "The logical form of action sentences," in *The Logic of Decision and Action* (N. Rescher, ed.), pp. 81–95, University of Pittsburgh Press, 1967.

[66] D. D.Gildea, "Automatic labeling of semantic roles," in *Association for Computational Linguistics*, ACL, 2000.

[67] J. D.Jurafsky, *Speech and Language Processing.* Stanford: DRAFT, 2nd ed., 2022.

[68] P. Shi and J. Lin, "Simple bert models for relation extraction and semantic role labeling," 2019.

[69] D. Hendrycks, X. Liu, E. Wallace, A. Dziedzic, R. Krishnan, and D. Song, "Pretrained transformers improve out-of-distribution robustness," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (D. Jurafsky, J. Chai, N. Schluter, and J. Tetreault, eds.), (Online), pp. 2744–2751, Association for Computational Linguistics, July 2020.

[70] D. Yogatama, C. de Masson d'Autume, J. Connor, T. Kocisky, M. Chrzanowski, L. Kong, A. Lazaridou, W. Ling, L. Yu, C. Dyer, and P. Blunsom, "Learning and evaluating general linguistic intelligence," 2019.

[71] T. McCoy, E. Pavlick, and T. Linzen, "Right for the wrong reasons: Diagnosing syntactic heuristics in natural language inference," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (A. Korhonen, D. Traum, and L. Màrquez, eds.), (Florence, Italy), pp. 3428–3448, Association for Computational Linguistics, July 2019.

[72] S. Gururangan, S. Swayamdipta, O. Levy, R. Schwartz, S. Bowman, and N. A. Smith, "Annotation artifacts in natural language inference data," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)* (M. Walker, H. Ji, and A. Stent, eds.), (New Orleans, Louisiana), pp. 107–112, Association for Computational Linguistics, June 2018.

[73] Adamczyk, Piotr, Kiryk, Grzegorz, Napiórkowski, Jarosław, and Walczak, Andrzej, "Network model of security system," *MATEC Web Conf.*, vol. 76, p. 02002, 2016.

[74] C. J. Fillmore, "The case for case," in *Universals in Linguistic Theory* (E. Bach and R. T. Harms, eds.), pp. 0–88, New York: Holt, Rinehart and Winston, 1968.

[75] Y. Dong, N. V. Chawla, and A. Swami, "Metapath2vec: Scalable representation learning for heterogeneous networks," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '17, (New York, NY, USA), p. 135–144, Association for Computing Machinery, 2017.

[76] B. J. Grosz, A. K. Joshi, and S. Weinstein, "Centering: A framework for modeling the local coherence of discourse," *Computational Linguistics*, vol. 21, no. 2, pp. 203–225, 1995.

[77] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, and L. Zettlemoyer, "BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, (Online), pp. 7871–7880, Association for Computational Linguistics, July 2020.

[78] "Chatgpt prompt engineering." `http://https://www.promptingguide.ai/models/chatgpt`. Accessed: 2023-09-10.

[79] P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, and G. Neubig, "Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing," *ACM Comput. Surv.*, vol. 55, jan 2023.

[80] A. P. Bradley, "The use of the area under the roc curve in the evaluation of machine learning algorithms," *Pattern Recognition*, vol. 30, no. 7, pp. 1145–1159, 1997.

[81] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, A. Rodriguez, A. Joulin, E. Grave, and G. Lample, “Llama: Open and efficient foundation language models,” 2023.

[82] A. Q. Jiang, A. Sablayrolles, A. Mensch, C. Bamford, D. S. Chaplot, D. de las Casas, F. Bressand, G. Lengyel, G. Lample, L. Saulnier, L. R. Lavaud, M.-A. Lachaux, P. Stock, T. L. Scao, T. Lavril, T. Wang, T. Lacroix, and W. E. Sayed, “Mistral 7b,” 2023.

[83] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. Chi, Q. Le, and D. Zhou, “Chain-of-thought prompting elicits reasoning in large language models,” 2023.

[84] X. Wang, J. Wei, D. Schuurmans, Q. Le, E. Chi, S. Narang, A. Chowdhery, and D. Zhou, “Self-consistency improves chain of thought reasoning in language models,” 2023.

[85] T. Schick and H. Schütze, “Exploiting cloze-questions for few-shot text classification and natural language inference,” in *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume* (P. Merlo, J. Tiedemann, and R. Tsarfaty, eds.), (Online), pp. 255–269, Association for Computational Linguistics, Apr. 2021.

[86] R. Bakker, *Knowledge Graphs : Representation and Structuring of Scientific Knowledge*. PhD thesis, University of Twente, Enschede, The Netherlands, 1987.

Appendix

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

- Item number one,
- item number two,
- item number three,

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Table 1: List of conferences at which I presented the results of investigations in the form of an oral presentation.

Date	Place	Conference	Title of contribution
01.2026	virtual	A fancy Seminar	Really long name of the presentation
06.2027	virtual	Conference number two	something equally interesting
06.2029	New York	11th International Conference	The same speech
06.2029	Copenhagen	Nanotechnology conference	A revolutionary presentation