Quantum Computing for Predicting Environmental Impacts: A Focus on Hydrogen Safety in Industry

July Bias Macedo[a*], Plinio Ramos[a], Caio Souto Maior[a], Isis Didier Lins[a], Márcio das Chagas Moura[a]
[a]Universidade Federal de Pernambuco – CEERMA, Recife, Brazil

Abstract: The growing interest in hydrogen as a clean energy solution underscores the need for rigorous safety assessments to mitigate potential environmental impacts. Safety concerns persist as a significant barrier due to hydrogen's characteristics, notably its high flammability and propensity for inducing material degradation. Consequently, incidents involving hydrogen carry the potential for catastrophic outcomes, as evidenced by the Fukushima nuclear disaster, which points out the severe consequences of hydrogen combustion and explosion (SUN et al., 2024). Regulatory mandates necessitate the reporting of accidents, and Natural Language Processing (NLP) serves as a valuable tool for extracting actionable insights from these reports. Despite the increasing use of quantum approaches in various fields, Quantum Natural Language Processing (QNLP) applications in Risk Analysis (RA) remain understudied, particularly in the hydrogen fuel industry, where risk analysis is critical. To fill this gap, we build a QNLP model capable of effectively classifying hydrogen-related accidents based on their potential environmental consequences, as inferred from event descriptions. By using Lambeq's pipeline, we convert textual accident descriptions into quantum circuits, enabling precise classification of accidents for environmental impact. Our model's application to the Hydrogen Incidents and Accidents Database (HIAD 2.1) highlights its efficacy in discerning environmental repercussions from incident narratives. Identifying whether a hydrogen-related accident leads to environmental impact is crucial for regulatory compliance and industry sustainability. This research not only advances the capabilities of QNLP but also underscores the imperative of leveraging quantum computing to address environmental concerns in emerging energy sectors.

Keywords: Accident Reports, Hydrogen, Natural Language Processing, Quantum Computing.

1. INTRODUCTION

The growing interest in hydrogen as a clean energy solution underscores the need for rigorous safety assessments to mitigate potential environmental impacts. Safety concerns persist as a significant barrier due to hydrogen's characteristics, notably its high flammability and propensity for inducing material degradation (Shen *et al.*, 2023). For instance, hydrogen is frequently stored as a compressed gas, which is more volatile than in other storage options; leaks due to impact or compartment failure may result in a rapid discharge of a highly explosive gas. Moreover, there are other issues including material weakening, which causes premature degradation of mechanical properties and results in cracking and loss of the storage cylinders' rigidity (Vidas and Castro, 2021). Consequently, incidents involving hydrogen carry the potential for catastrophic outcomes, as evidenced by the Fukushima nuclear disaster, which points out the severe consequences of hydrogen combustion and explosion (Sun *et al.*, 2024).

Thus, risk analysis (RA) emerges as a fundamental tool to guide investments and efforts to avoid undesirable events. Moreover, identifying whether a hydrogen-related accident leads to environmental impact is crucial for regulatory compliance and industry sustainability. Indeed, to address these safety issues, safety protocols and regulations have been put in place for hydrogen production. These regulations require the use of safety equipment and training for workers, as well as the monitoring and reporting of safety incidents (Silva and Jacinto, 2012; Hämäläinen, Takala and Kiat, 2017). By documenting and analyzing incidents, companies can identify patterns or trends indicating systemic issues that need attention. This knowledge informs training programs and safety protocols to minimize the likelihood of similar accidents in the future. However, the large volume of accident reports makes human review impractical (Macêdo *et al.*, 2022).

Given the challenge of reviewing a large volume of accident reports, advanced data analysis techniques, such as Natural Language Processing (NLP), have proven valuable for extracting relevant insights. These advancements are particularly important because risk models have evolved, one reason is the advance in computing performance and the ability to record, store and process massive amounts of data. Another reason is the breakthroughs in the field of artificial intelligence that have enabled the efficient extraction of information from complex, high-dimensional and unstructured datasets (Nateghi and Aven, 2021).

The advancement of NLP, particularly Transformers like BERT (Devlin *et al.*, 2018) and GPT (Radford *et al.*, 2018), has significantly enhanced performance across various NLP tasks, surpassing previous models like LSTM recurrent neural networks. However, this progress comes with a challenge: these models are becoming increasingly complex (e.g., OpenAI's GPT model has hundreds of billions of parameters), demanding vast amounts of data for efficient training (e.g., entire Wikipedia corpora across multiple languages) (Guarasci, De Pietro and Esposito, 2022). This complexity and resource requirement represent significant drawbacks of current Transformer-based approaches.

As NLP advances, new technologies like quantum computing offer additional promises. Specifically, the emerging field of Quantum NLP (QNLP) could overcome current limitations by leveraging quantum mechanics, particularly categorical quantum mechanics, to relate sentence structure to quantum states and grammar to entanglement. Quantum computing harnesses the quantum mechanics properties, such as superposition and entanglement to manipulate and represent data. Recent developments have seen the emergence of quantum algorithms for machine learning applications. Some approaches aim to enhance machine learning performance using quantum properties, while others reformulate machine learning problems using quantum theory. This intersection of quantum computing and machine learning has led to Quantum NLP (QNLP), which implements natural language on quantum hardware (Pandey *et al.*, 2023).

Motivated by the scarcity of research in this area, with only 259 QNLP-related studies identified in the Web of Science and none exploring its application in industry, this research endeavors to fill this gap. The core idea behind QNLP lies in harnessing categorical quantum mechanics, a branch of quantum physics that connects complex mathematical structures with linguistic meanings, allowing for a richer representation of grammatical relationships to effectively integrate language meaning and grammatical structure, promising a deeper comprehension of language nuances compared to classical dependency-based representations. In particular, QNLP offers the potential to efficiently represent complex linguistic structures through quantum states and entanglement. This approach could enable a more nuanced understanding of language, especially in capturing intricate relationships between words and phrases. Additionally, QNLP may provide a more seamless integration of semantic meaning and grammatical structure, leading to models that better capture the richness and complexity of human language. These benefits could be especially valuable in tasks requiring deep contextual understanding, further underscoring the promise of QNLP in advancing NLP capabilities (Guarasci, De Pietro and Esposito, 2022).

Thus, to fill this research gap, we propose the development of a QNLP model to classify hydrogen-related accidents based on their potential environmental implications, which are crucial for regulatory compliance and risk mitigation in emerging industries. Adopting Lambeq's pipeline, we will translate textual accident narratives into quantum circuits, facilitating precise accident classification in terms of environmental impact. Our model's deployment on the Hydrogen Incidents and Accidents Database (HIAD 2.1), established by the European Commission-funded Network of Excellence on Hydrogen Safety (Daniele, Jennifer Xiaoling and Moretto, 2019), will showcase its effectiveness in discerning environmental consequences from incident descriptions. Given the significance of identifying environmental impacts for regulatory compliance and industry sustainability in the burgeoning energy sectors, this research not only pushes the boundaries of QNLP capabilities but also underscores the imperative of harnessing quantum computing to address environmental concerns.

2. THEORETICAL BACKGROUND

The integration of Natural Language Processing (NLP) with quantum theory relies on a fundamental assumption: the direct correlation between linguistic attributes, such as syntactic structures and semantic meanings, and quantum states. This connection is facilitated by the DisCoCat framework, which uses string diagrams as a network-like language to model these relationships (Yeung and Kartsaklis, 2021). Traditionally, two main approaches have been used to represent language structures and meanings: the distributional approach, which relies on statistical analyses of word contexts, and the symbolic approach, which focuses on compositional meanings of sentences. Distributional models, though widely adopted and effective, face significant limitations, including a reliance on extensive datasets and parameters, which grows as models evolve, posing resource and efficiency challenges. Additionally, their interpretability is limited, making it difficult to gain meaningful insights into linguistic phenomena (Guarasci, De Pietro and Esposito, 2022).

To address these limitations, efforts have been made to integrate language structure, particularly syntax, into distributional models. This has led to the development of models like DisCoCat, which draws on concepts from theoretical linguistics such as Universal Grammar (Montague, 1970), categorical grammar (Lambek, 1958), and pre-group grammar (Lambek, 1999). The compositional distributional model, as proposed by (Coecke, Sadrzadeh and Clark, 2010), introduces a graphical framework using string diagrams. This model aims to deepen the understanding of word interactions and sentence meanings by integrating structural components of language with statistical methodologies.

DisCoCat employs string diagrams to represent sentence meanings. In a sentence like “Hydrogen produces electricity”, the string diagram connects the noun “Hydrogen”, the verb “produces” and the noun “electricity” to construct the overall meaning. The sentence meanings are computed using the tensor product of vector spaces representing word meanings and their grammatical roles.

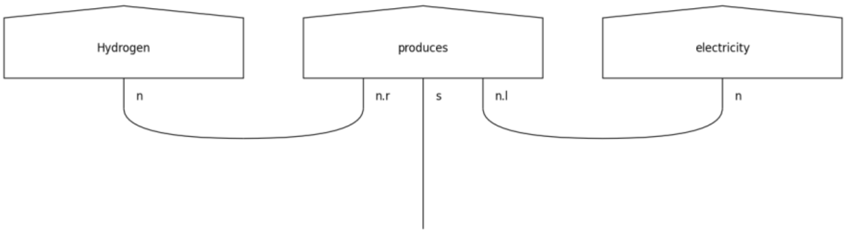

Figure 1. Example of a simple sentence represented using a string diagram.

DisCoCat has been further reformulated in quantum terms, where pentagons represent quantum states and wires represent Bell-effect connections. This reformulation has demonstrated the potential compatibility of DisCoCat with quantum computing characteristics (Coecke *et al.*, 2020). DisCoCat’s reformulation in quantum terms showcases its potential compatibility with quantum-native characteristics. Consequently, the compatibility between natural language and quantum hardware suggests that QNLP could find a more conducive environment in quantum computing systems rather than classical ones. This suggests that translating linguistic structures into quantum circuits could harness quantum advantages, particularly speedup when implemented on proper quantum hardware.

To build a QNLP model to perform a specific task, in our case a text classification, text data needs to be converted to quantum circuits. Thus, a QNLP model to classify hydrogen-related accidents based on their potential environmental implications, inferred from event descriptions, was developed according to a pipeline as shown in Figure 2, based on Lambeq’s pipeline (Yeung and Kartsaklis, 2021).

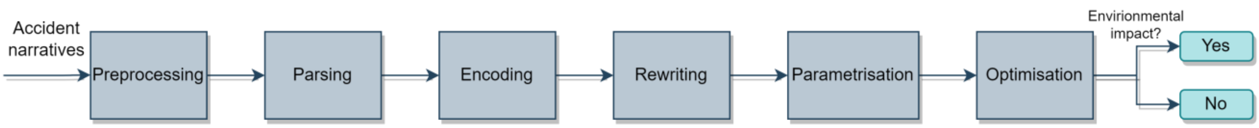

Figure 2. Pipeline to implement our QNLP model.

The first step consists of parsing the input sentences, which is the process of assigning a syntactic structure to a given grammar. In other words, in this step, a statistical tool converts a sentence into a hierarchical representation that reflects the syntactic relationships between the words (a syntax tree) based on a specific grammar formalism (Clark, 2021). In the example presented in Figure 3, “Hydrogen” is a type composed only of the group “n” (noun), “produces” is a type composed of two adjuncts, “n.r.”, which indicates that the word is adjoined to the right of “Hydrogen”, and “n.l.”, which indicates that the word is adjoined to the left of “electricity”, and an “s” group (sentence). Finally, “electricity” also is a type composed only of the “n” group.

Next, the resulting syntactic trees are encoded into String Diagrams and, as they respect the order of the words and indicate the compositional relationships between them, they are close to Quantum Circuits, while still being free of low-level decisions, such as the types of gates to be used and so on (Bonchi *et al.*, 2021). Due to the current state of quantum computing, characterized by devices featuring 50–100 qubits, it becomes crucial to consider the size and intricacy of String Diagrams. As their dimensions increase, so does the demand for qubits in modeling, consequently amplifying the associated computational costs (Sá, Oliveira and Roditi, 2023). For this reason, various diagram rewriting techniques can be used, from techniques that simply “distort” the diagram to techniques that exclude connectors and auxiliary verbs (Kartsaklis *et al.*, 2021).

The String Diagrams are then transformed into Parameterized Quantum Circuits, also called Ansatz, where the desired computation is carried out (Kharsa, Bouridane and Amira, 2023). This is achieved by mapping the pre-groups defined in the previous stage and the corresponding quantum gates. The circuit Ansatz is characterized by its inclusion of adjustable parameters, which can be systematically tuned throughout the execution of a quantum algorithm to optimize the configuration for solving a designated computational (Biamonte *et al.*, 2017). The type of Ansatz chosen defines the number of qubits used, and the configuration of the gates in the circuit.

Finally, once the quantum circuits have been defined, measurements can be made, and results verified. However, in most cases, a “newly created” circuit will not give an optimal result in the first measurement. This is because it is necessary to change the circuit parameters, e.g., angles of the rotation gates, according to the task being performed. There are several algorithms, both classical and quantum, that can perform optimization. In the quantum context, Lambeq supports algorithms such as SPSA and Rotosolve, implemented in the SPSAOptimizer and RotosolveOptimizer classes (Acampora, Chiatto and Vitiello, 2023).

Thus, to perform classification with the quantum models we must optimize the parameters of the quantum circuit to minimize a certain cost or error function. The quantum model is trained on a labeled dataset, and during the training process, it learns to map input data to the correct output class. Once trained, the quantum model can be used for classification by applying the learned quantum circuit to unseen data. The quantum operations in the circuit transform the input data, and the final measurement outcomes are used to make predictions about the class of the input (Ostaszewski, Grant and Benedetti, 2021).

3. METHODOLOGY

3.1. HIAD 2.1

The analysis in this study is based on data from the Hydrogen Incidents and Accidents Database (HIAD 2.1), a comprehensive repository established to collect reports on industrial accidents involving hydrogen and its derivatives. HIAD was originally created by the Joint Research Centre (JRC) of the European Commission as part of the Hydrogen Safety Excellence Network (HySafe) initiative, which ran from 2004 to 2009 (JRC, 2004). The primary goal of HIAD was to facilitate the sharing of insights from hazardous incidents involving hydrogen to improve safety and prevent similar events in the future (Kirchsteiger, Vetere Arellano and Funnemark, 2007).

In 2017, a significant update was undertaken by the JRC in collaboration with the Fuel Cell and Hydrogen Joint Undertaking (FCH 2 JU), resulting in the evolution of HIAD 2.0 into HIAD 2.1. This updated version was subsequently integrated into the framework of the European Hydrogen Safety Panel (EHSP) activities spanning the period from 2009 to 2022 (Daniele, Jennifer Xiaoling and Moretto, 2019). HIAD 2.1 includes a detailed set of parameters for understanding and analyzing reported incidents. The database is organized into an Excel workbook, last updated on January 1, 2022, with different sheets, including:

- Events: Contains classified events, narrative summaries, involved systems, dates, locations, and cause classifications.
- Installation: Provides descriptions of applications, storage conditions, location types, and pre-event conditions.
- Consequences: Details the impact in terms of human and property losses.
- Lessons Learned: Includes insights into corrective measures taken.
- Event Nature: Offers quantitative data on emergency responses, leakage characteristics, leak types, and fire-related consequences.
- Reference: Lists primary sources of information.

Additionally, the dataset includes information on the number of injuries, fatalities, environmental damage, property loss (onsite and offsite), post-event summaries, legal actions, and investigation comments. This study focuses on classifying environmental damage based on the descriptions within this dataset.

3.2. Modelling

Our model implementation utilizes Python's Lambeq library, an open-source tool developed by Quantinuum. Lambeq facilitates the integration of various tools from different libraries, such as Pytorch and Pennylane (Kartsaklis *et al.*, 2021). Pennylane is a key component in our approach, providing a versatile platform for executing and training quantum programs on different backends. It also integrates with classical machine learning libraries such as PyTorch, TensorFlow, Keras, and NumPy, supporting hybrid quantum-classical algorithms. Pennylane's features include quantum circuit optimization, error correction, and a range of quantum machine learning algorithms.

The accident descriptions from HIAD 2.1 were preprocessed to remove stop words, punctuation, and null entries. Descriptions were then labeled based on their environmental impact. To manage the limitations of qubits and keep the model applicable to a smaller corpus, each description was randomly reduced to 10 words. The dataset was divided into training, testing, and validation sets.

The preprocessed descriptions were then converted into diagrammatic representations using the BobcatParser. These diagrams were then rewritten to streamline their structures and improve computational efficiency. Techniques applied included specific rewrite categories such as 'auxiliary', 'postadverb', and 'preadverb', and redundant cup elements were removed using the RemoveCupsRewriter.

Quantum circuits were generated from the rewritten diagrams using an IQPAnsatz. This ansatz was configured with parameters for NOUN and SENTENCE atomic types, each assigned to one qubit, and included two layers with three single-qubit parameters. A model was instantiated from these circuits, utilizing the AerBackend for quantum simulation with a default compilation pass and 8192 shots. In addition, a sequential neural network was integrated with the quantum model. The architecture of this network included:

- An input layer corresponding to the output of the PennyLaneModel.
- A dense layer with 128 units and ReLU activation.
- A dropout layer with a rate of 0.5 to prevent overfitting.
- A second dense layer with 64 units and ReLU activation.
- Another dropout layer with a rate of 0.5.
- A final dense layer with 1 unit and a sigmoid activation function for binary classification.

The hybrid model was trained using the Adam optimizer with a binary cross-entropy loss function for 100 epochs, with a learning rate of 0.005. Training involved periodic evaluation on a validation set to monitor performance and applied early stopping criteria based on accuracy to prevent overfitting. Datasets were created with batch sizes of four..

To analyze the QNLP-based model's performance, three distinct datasets were built to train our QNLP model to classify the environmental impact based on the accident description. The first dataset comprised accident narratives as input sentences, preprocessed as described earlier. The second dataset included additional information from HIAD 2.1, such as 'Release type', 'Release substance', 'Release amount', 'Ignition source', 'High-pressure explosion', and 'Flame type', to build more informative input sentences. The third dataset was built using data augmentation to address the imbalance in the dataset given the limited number of instances labeled as having an environmental impact (12 instances) compared to those labeled as having no environmental impact (54 instances). Data augmentation was performed by generating new sentences through random word swapping to synonyms.

4. RESULTS AND DISCUSSION

The QNLP-based model's performance was evaluated through a 5-fold cross-validation across all datasets. Figure 1 displays the confusion matrices showing the best and worst fold results of each model on the test data. It is noteworthy that the test data remains consistent across all models.

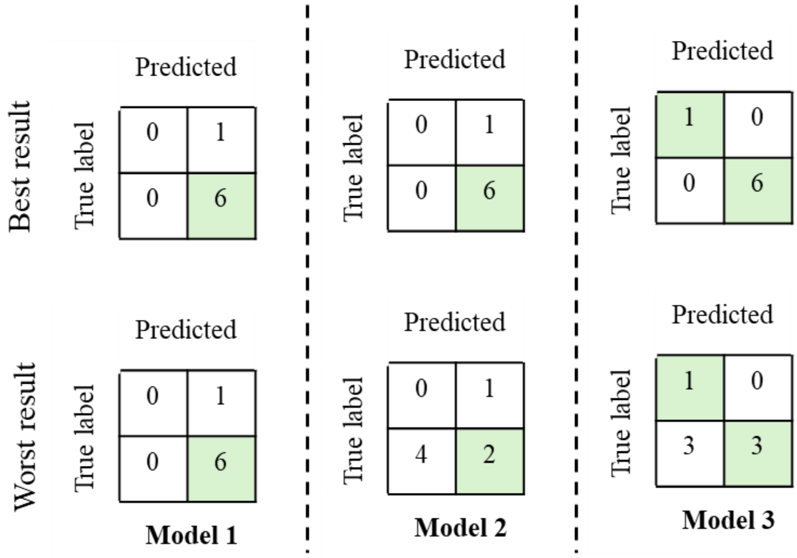

Figure 1. The best and worst fold results of each model on the test data.

The model trained on the first dataset, comprising only preprocessed accident narratives as input sentences, achieved an average accuracy of 85.71%. Notably, the model correctly classified all accidents that did not lead to environmental impact. However, it consistently misclassified instances that led to environmental impact in all folds. This discrepancy indicates potential limitations in the model's ability to discern nuanced linguistic features related to environmental implications.

Similarly, the model trained on the second dataset, enriched with additional information from HIAD 2.1, was not able to correctly identify the only instance that led to environmental impact. The second model also achieved an average accuracy of 85.71%. This consistent misclassification underscores the need for further refinement in capturing contextual cues related to environmental factors.

The third dataset, augmented to address class imbalance, yielded an average accuracy of 77.13%. Despite the lower overall accuracy compared to the previous datasets, this dataset demonstrated more balanced results. Notably, the model misclassified the instance leading to environmental impact only once across the five folds. This improvement suggests that data augmentation techniques contribute to enhancing the model's robustness in identifying environmental implications. Importantly, despite the worst result (shown in Figure 1), Model 3 successfully classified the instance leading to environmental impact.

Comparing these results to those presented in (Macêdo *et al.*, 2024), he current QNLP-based model shows significant improvements. Earlier experiments with a purely quantum approach had lower accuracy, with models taking more than 8 hours to train and achieving less than 63% accuracy on test data. This comparison underscores the computational intensity and relatively lower performance of quantum-only models.

Classical approaches, while faster, also showed improved results with significantly reduced training times. A classical model trained for about 1 hour achieved an average accuracy of 73% on test data. Although this classical model performed better than the initial quantum models, it still fell short of the accuracy achieved by the current QNLP-based model.

The hybrid QNLP-based model, combining quantum circuits with a sequential neural network, demonstrates enhanced classification capabilities, particularly in recognizing environmental impacts. This approach leverages the strengths of both quantum and classical methodologies, resulting in improved accuracy and computational efficiency. The model's consistent performance across various datasets and its success in identifying critical instances of environmental impact highlight the potential of QNLP in advancing accident analysis and environmental impact assessment.

5. CONCLUSION

The inclusion of additional information from HIAD 2.1 in the second dataset aimed to provide a more comprehensive context for accident classification. However, the marginal improvement in accuracy suggests the need for more nuanced feature engineering or alternative data representation techniques to leverage additional information. The third dataset, augmented to address the class imbalance, demonstrated improved balance in classification results. This highlights the efficacy of data augmentation techniques in mitigating the imbalanced dataset challenges. Future research could explore advanced augmentation strategies to further improve model performance.

By pioneering the application of QNLP in the domain of RA, this research establishes a foundation for future investigations. The development and evaluation of a QNLP-based model for accident classification represent a novel contribution to the field, offering insights into the feasibility and efficacy of quantum-inspired approaches in addressing complex real-world challenges.

Quantum computing introduces unique challenges, including qubit limitations and circuit depth constraints, which can impact the scalability and complexity of QNLP models. While our study demonstrates promising results within a controlled environment, the practical implementation of QNLP models on larger datasets may encounter computational bottlenecks. Addressing these limitations requires continued advancements in quantum hardware and algorithmic development to optimize efficiency and scalability. It is worth emphasizing the exploratory nature of this research. As such, the observed performance of the QNLP model reflects initial insights rather than definitive conclusions.

Acknowledgements

The authors thank the following Brazilian research funding agencies for the financial support: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), 310892/2022-8, 402761/2023-5; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Finance Code 001. We also thank the Fundação de Amparo a Ciência e Tecnologia de Pernambuco (FACEPE), APQ-0394-3.08/22 and Human Resources Program 38.1 (PRH 38.1) entitled Risk Analysis and Environmental Modeling in the Exploration, Development and Production of Oil and Gas, financed by Agência Nacional de Petróleo, Gás Natural e Biocombustíveis (ANP) and managed by Financiadora de Estudos e Projetos (FINEP), 044819.

References

Acampora, G., Chiatto, A. and Vitiello, A. (2023) 'Genetic algorithms as classical optimizer for the Quantum Approximate Optimization Algorithm', *Applied Soft Computing*, 142, p. 110296. Available at: https://doi.org/10.1016/j.asoc.2023.110296.

Biamonte, J. *et al.* (2017) 'Quantum machine learning', *Nature*, 549(7671), pp. 195–202. Available at: https://doi.org/10.1038/nature23474.

Bonchi, F. *et al.* (2021) 'Bialgebraic foundations for the operational semantics of string diagrams', *Information and Computation*, 281, p. 104767. Available at: https://doi.org/10.1016/j.ic.2021.104767.

Clark, S. (2021) 'Something Old, Something New: Grammar-based CCG Parsing with Transformer Models'. Available at: http://arxiv.org/abs/2109.10044.

Coecke, B. *et al.* (2020) 'Foundations for Near-Term Quantum Natural Language Processing', (1), pp. 1–43. Available at: http://arxiv.org/abs/2012.03755.

Coecke, B., Sadrzadeh, M. and Clark, S. (2010) 'Mathematical Foundations for a Compositional Distributional Model of Meaning'. Available at: http://arxiv.org/abs/1003.4394.

Daniele, M., Jennifer Xiaoling, W. and Moretto, P. (2019) 'HIAD 2.0 - hydrogen incidents and accidents database', in *Proceedings of the International Conference on Hydrogen Safety ICHS*.

Devlin, J. *et al.* (2018) 'BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding', *arXiv preprint arXiv:1810.04805* [Preprint].

Guarasci, R., De Pietro, G. and Esposito, M. (2022) 'Quantum Natural Language Processing: Challenges and Opportunities', *Applied Sciences (Switzerland)*, 12(11). Available at: https://doi.org/10.3390/app12115651.

Hämäläinen, P., Takala, J. and Kiat, T.B. (2017) *Global Estimates of Occupational Accidents and Work-Related Illnesses 2017*, *Workplace Safety and Health Institute*.

Kartsaklis, D. *et al.* (2021) 'lambeq: An Efficient High-Level Python Library for Quantum NLP', *arXiv preprint*, arXiv:2110, pp. 1–19. Available at: http://arxiv.org/abs/2110.04236.

Kharsa, R., Bouridane, A. and Amira, A. (2023) 'Advances in Quantum Machine Learning and Deep Learning for Image Classification: A Survey', *Neurocomputing*, 560, p. 126843. Available at: https://doi.org/10.1016/j.neucom.2023.126843.

Kirchsteiger, C., Vetere Arellano, A.L. and Funnemark, E. (2007) 'Towards establishing an International Hydrogen Incidents and Accidents Database (HIAD)', *Journal of Loss Prevention in the Process Industries*, 20(1), pp. 98–107. Available at: https://doi.org/10.1016/j.jlp.2006.10.004.

Lambek, J. (1958) 'The Mathematics of Sentence Structure', *The American Mathematical Monthly*, 65(3), p. 154. Available at: https://doi.org/10.2307/2310058.

Lambek, J. (1999) 'Type grammar revisited', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 1582, pp. 1–27. Available at: https://doi.org/10.1007/3-540-48975-4_1.

Macêdo, J.B. *et al.* (2022) 'Identification of risk features using text mining and BERT-based models:

Application to an oil refinery', *Process Safety and Environmental Protection*, 158, pp. 382–399. Available at: https://doi.org/10.1016/j.psep.2021.12.025.

Macêdo, J.B. *et al.* (2024) 'Quantum Natural Language Processing To Human Factor Identification In Accident Reports', in K. Kołowrocki and E. Dąbrowska (eds) *Advances in Reliability, Safety and Security ESREL 2024 Collection of Extended Abstracts Part 2*. Krakow.

Montague, R. (1970) 'Universal grammar', *Theoria*, 36(3), pp. 373–398. Available at: https://doi.org/10.1111/j.1755-2567.1970.tb00434.x.

Nateghi, R. and Aven, T. (2021) 'Risk Analysis in the Age of Big Data: The Promises and Pitfalls', *Risk Analysis*, 41(10), pp. 1751–1758. Available at: https://doi.org/10.1111/risa.13682.

Ostaszewski, M., Grant, E. and Benedetti, M. (2021) 'Structure optimization for parameterized quantum circuits', *Quantum*, 5, pp. 1–13. Available at: https://doi.org/10.22331/Q-2021-01-28-391.

Pandey, S. *et al.* (2023) 'Quantum machine learning for natural language processing application', *Physica A: Statistical Mechanics and its Applications*, 627, p. 129123. Available at: https://doi.org/10.1016/j.physa.2023.129123.

Radford, A. *et al.* (2018) 'Improving Language Understanding by Generative Pre-Training'. Available at: https://doi.org/10.4310/HHA.2007.v9.n1.a16.

Sá, N., Oliveira, I.S. and Roditi, I. (2023) 'Towards solving the BCS Hamiltonian gap in near-term quantum computers', *Results in Physics*, 44. Available at: https://doi.org/10.1016/j.rinp.2022.106131.

Shen, Y. *et al.* (2023) 'Preliminary hazard identification for qualitative risk assessment on onboard hydrogen storage and supply systems of hydrogen fuel cell vehicles', *Renewable Energy*, 212, pp. 834–854. Available at: https://doi.org/10.1016/j.renene.2023.05.037.

Silva, J. and Jacinto, C. (2012) 'Finding occupational accident patterns in the extractive industry using a systematic data mining approach', *Reliability Engineering and System Safety*, 108, pp. 108–122. Available at: https://doi.org/10.1016/j.ress.2012.07.001.

Sun, J. *et al.* (2024) 'Research on hydrogen risk prediction in probability safety analysis for severe accidents of nuclear power plants', *Nuclear Engineering and Design*, 417(November 2023), p. 112798. Available at: https://doi.org/10.1016/j.nucengdes.2023.112798.

Vidas, L. and Castro, R. (2021) 'Recent Developments on Hydrogen Production Technologies: State-of-the-Art Review with a Focus on Green-Electrolysis', *Applied Sciences*, 11(23), p. 11363. Available at: https://doi.org/10.3390/app112311363.

Yeung, R. and Kartsaklis, D. (2021) 'A CCG-Based Version of the DisCoCat Framework'. Available at: http://arxiv.org/abs/2105.07720.