Original Article

Artificial intelligence–based monitoring of risk scenarios

Felippe Ribas Barboza [1]*
Guilherme Sausen Welter [1]
Lorena Vieira Rezende [2]
Fabricio Fracaroli Cola [3]
Ramon Bonela [1]

Abstract

This paper presents an initiative developed between Samarco and the startup Simple Safety, focused on building an intelligent system for monitoring safety risk scenarios. The main objective was to validate the technical and operational feasibility of applying large language models (LLMs) and artificial intelligence agents as a semantic enrichment layer over existing safety data. The methodology included a six-month proof of concept (PoC), involving technical site visits, stakeholder alignment meetings, and operational context mapping, combined with data and semantic engineering techniques. The resulting system enabled the identification of risks and vulnerabilities in critical controls, with automated categorization and the prescription of actions based on detected patterns. The model achieved an 81% coherence rate in risk and control classification, alongside perceived improvements in decision-making processes and in strengthening the organization's safety culture. The study supports the feasibility of generative AI-based solutions to expand safety capacity, support real-time decision-making, and promote a higher level of digital maturity in risk management.

Keywords: Safety; Artificial intelligence; Language models; Risk management.

1 Introduction

Digital transformation is a significant milestone in the industrial context, enabling companies to gain competitive advantage, foster innovation, and strengthen organizational resilience to absorb and adapt within an environment of uncertainty. Understanding the context, enhanced by the application and sharing of information and knowledge extracted from multiple available sources, can drive learning and increase the capacity to anticipate, plan, and respond to the changes inherent to industrial operations [1]. When the goal is to promote a safe working environment, this transformation becomes an essential precursor for identifying inherent hazards, error-prone conditions in operations, managing critical controls, and supporting the development of a strong safety culture.

However, this digital leap introduces a subtle and often overlooked side effect: cognitive overload. As cited by Strutzel [2], Hebert Simon observes: "In an information-rich world, the wealth of information means a dearth of something else: a scarcity of whatever it is that information consumes. What it consumes is rather obvious: the attention of its recipients." In other words, an abundance of data paradoxically creates a poverty of attention. In the context of safety, the impact of this effect manifests in the compromised allocation of resources, reduced decision-making effectiveness, and difficulty in prioritizing critical risks.

Within this scenario, Samarco Mineração, one of Brazil's leading mining companies, which holds safety as an organizational value, has made consistent progress in applying digital technologies to enhance risk and safety management. However, despite significant advances in systematization and monitoring aimed at preventing work-related injuries and health issues, the vast amount of data is not fully leveraged for its strategic potential, revealing a real example of the challenges inherent in operational routines.

Given this context, the present work aims to present an initiative developed between Samarco Mineração and Simple Safety, a technology startup focused on helping industrial organizations reduce human exposure to risk. The purpose was to validate the technical and operational feasibility of a solution oriented toward monitoring safety risk scenarios using large language models (LLMs).

[1]*Simple Safety, São Paulo, SP, Brasil.*
[2]*Samarco Mineração, Belo Horizonte, MG, Brasil.*
[3]*Samarco Mineração, Anchieta, ES, Brasil.*

***Corresponding author:** *felippe.ribas@simplesafety.ai*
E-mails: guilherme.welter@simplesafety.ai; lorena.vieira@samarco.com; fabricio.cola@samarco.com; ramon.bonela@simplesafety.ai

2 Development

2.1 The challenge

Risk and safety management at Samarco, although structured and guided by essential enablers that drive the creation of a solid foundation of sustainable practices promoting positive results in organizational culture, faces challenges related to the continuous generation of large volumes of information.

Among the main obstacles are:

- Excessive volumes of safety data, hindering prioritization and timely response to critical situations.
- Decentralized data across systems and spreadsheets, limiting integrated risk visibility and increasing misinterpretation; and
- Limited data correlation, restricting the identification of patterns, trends, and operational vulnerabilities.

These elements, when combined, directly affect the effectiveness of decision-making and reduce what is known as safety capacity, the degree to which the organization, its operations, and its professionals can absorb both the constructive and destructive consequences of failures [3].

2.2 The solution

A project was proposed with the objective of leveraging opportunities arising from the identified challenges and enhancing safety capacity, more specifically, operational capacity, which concerns how effectively the organization identifies inherent and emerging hazards, as well as conditions prone to errors in operations, creating barriers and controls against them [3].

The project was conducted over six months as a Proof of Concept (PoC) at Samarco Mineração's Ubu unit, including two on-site visits to assess safety culture and digital maturity in Risk and Health and Safety Management. More than thirty working sessions were held with Safety and IT teams to understand how data was used for decision-making, planning, and learning.

During this period, an agile SCRUM methodology was applied, enabling immersion in Samarco's operational reality while establishing predefined milestones to ensure PoC success and alignment with organizational needs (Figure 1).

2.2.1 Data overview

Samarco has more than 23,000 risks registered in its management system and generates, monthly and from various data sources, approximately 2,000 records of unsafe conditions, 80 near-miss reports, 1,000 inspections of critical and general controls, and 4,000 behavioral observations, totaling around 240 safety records per day.

From these data sources, along with information related to the risk inventory and organizational hierarchy (Table 1), a sample of anonymized data was made available to begin the analysis process, which involved stages of extraction, cleaning, and transformation.

2.2.2 Understanding the information

From a data engineering and data science perspective, the project faced typical challenges such as dealing with multiple databases originating from systems developed for different purposes and by different vendors, with no direct key compatibility among them.

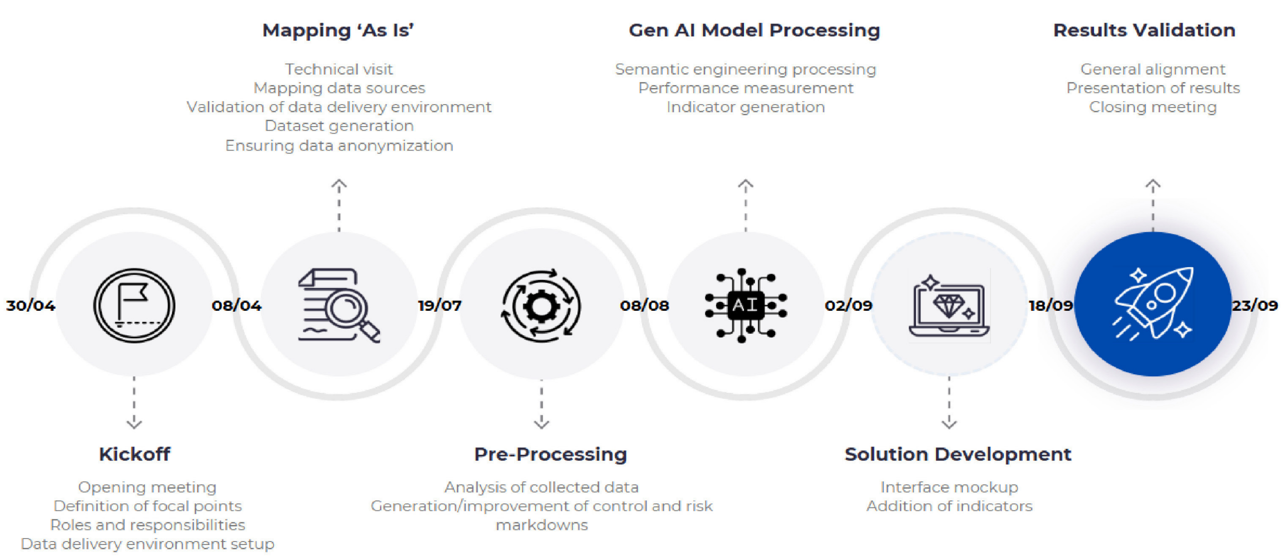

Figure 1. POC's Roadmap.

Table 1. Data Sources Used in the PoC

Category	Data Sources
Occurrences	Accidents, Near Misses, Unsafe Conditions, Incidents
Inspections	Critical Controls, General Controls, Behavioral Observations
Risk Inventory	Risk Situations and Control Measures
Organizational Hierarchy	Departments / Management Areas

The datasets contained between 20 and 30 fields, mostly categorical (with highly unbalanced distributions due to the large number of possible categories), numerical, and free-text types, with some missing values.

To standardize the data, cleaning and preprocessing techniques based on regular expressions and natural language processing (NLP) were applied, accounting for abbreviations, acronyms, technical terminology, material and equipment names, and spelling errors that could affect solution development.

2.2.3 Semantic engineering

From the initial stage of understanding the information contained in the databases, it became evident that there was a need to establish a strategy for the development of the solution, since, even with the use of sophisticated artificial intelligence resources, the sparse nature of the information and the lack of contextualization would limit the achievable outcomes.

As the foundation of this strategy, the principles based on the review article by Rowley [4] were followed, which describes the Data–Information–Knowledge–Wisdom hierarchy. From this perspective, it can be observed that there are certain types of information that can be objectively deduced from data through processes that are often automatable, such as sorting, classification, selection, and mathematical and computational operations. However, there are other types of information, closer to what is understood as Knowledge, that simply cannot be deduced from data; this type of information is more subjective and necessarily dependent on a process of contextualization and on specific knowledge to acquire meaning and utility.

The PoC challenge corresponds to the second case described above. To address this problem, a structured process referred to as Semantic Engineering was developed to organize, represent, and retrieve knowledge by constructing, contextualizing, and aggregating information that complements database-extracted data, enabling computational systems to understand meaning, context, and relationships between entities [5].

To orchestrate this engine, a set of tools was employed which, together, enhance the ability to categorize, relate, and compare situations, behaviors, and other relevant aspects that contribute to the understanding of operational risk scenarios, in the form of a conceptual or semantic network.

The foundations of this approach trace back to 1956, with Richard Richens' work at the Cambridge Language Research Unit, which introduced conceptual graphs for machine translation based on the premise that knowledge should be structured as networks of associations, a cornerstone of Natural Language Processing that mirrors human semantic memory through logical and hierarchical connections [6].

Following this milestone, tools emerged to represent meaning through symbolically structured knowledge networks. In this work, LLMs, embedding models, and RAG (Retrieval Augmented Generation) were employed to extract elementary attributes from data and construct an informational base. The choice of these three elements was not random, and each plays a fundamental role in the Semantic Engineering process:

LLMs act as the cognitive layer of the system, interpreting natural language data beyond the variability of free-text fields. They enable the association of meanings, inference of latent relationships, and generation of explainable syntheses across multiple information sources, providing the generalization and contextualization required to handle the ambiguity inherent in operational records and human narratives in safety-related data. [7].

Embedding models provide the system's semantic infrastructure by converting event texts and operational descriptions into dense vector representations. This enables the measurement of semantic similarity, conceptual proximity, and recurring patterns across situations [8], supporting large-scale comparison of historical events, behaviors, and emerging risk scenarios, and mathematically linking language, meaning, and structure beyond the context limits of LLMs.

RAG integrates LLMs and embedding models with knowledge retrieval mechanisms, grounding analyses and insights in domain-specific knowledge bases rather than relying solely on implicit model knowledge. This approach reduces bias, improves inference traceability, and enhances the explainability of recommendations, which is critical in high-risk industrial environments [9].

In an integrated manner, LLMs, embeddings, and RAG form the technological pillars that operationalize the proposed Semantic Engineering. Embeddings structure the semantic space, LLMs enable cognitive interpretation and synthesis, and RAG maintains alignment between reasoning and contextualized knowledge, supporting informed, operationally grounded decisions and the evolution toward an informed safety culture.

2.2.4 Risk assessment and vulnerability insights

Using the informational base and the mapped contexts and risk situations obtained through Semantic Engineering, a system was developed to act as an intelligence layer over existing data, designed to assess risks and identify vulnerabilities in safety controls.

Its purpose is to enhance monitoring capability and problem identification, considering information about incidents, hazards, and risks [10], while providing decision support through prescribed actions that stimulate critical thinking and accelerate decision-making for restoring barriers.

In general terms, the system operates on four fundamental pillars:

- Collect: Processing existing organizational data.
- Interpret: Enrichment of data through semantic layers.
- Prioritize: Identification of imminent risk scenarios and vulnerable controls.
- Recommendation: Prescription of actions for the most critical risk scenarios.

2.3 Result

To evaluate the results both technically and in terms of their impact on operational routines, a validation was conducted focusing on two aspects: the semantic coherence of risk and control categorizations and a qualitative assessment of the solution's potential performance within the organizational context.

2.3.1 Validation of results

To assess the outcomes, a methodology was adopted in which a random and homogeneous sample of records was selected. Due to the uneven representation of classes (such as Incident Type / Critical Potential and Inspection Type / Inspection Group), a restriction was applied to ensure that all risk and control categories were equally represented in the dataset. Consequently, 138 records from a 12-month data collection were used for evaluation.

The results of the risk and control categorizations produced by the AI model were compared with two independent classifications performed by safety specialists. One reference categorization was conducted by a specialist from Simple Safety, and the other by the Samarco Safety team.

It is important to note that these categorizations are, to some extent, interpretative, meaning discrepancies may naturally occur, which also explains the absence of a standardized benchmark in literature. However, the two independent expert classifications were found to be highly consistent with each other.

Taking the specialists' classifications as the reference (the "correct" results), the model achieved a weighted accuracy rate of 81%. This weighting corresponds to the results of the risk and control categorizations across both incident and inspection datasets.

The validation demonstrates the coherence and consistency of the automated risk and control classification achieved through an AI-based model.

2.3.2 Discussion

To provide traceability to the weighted accuracy rate of 81% obtained in the semantic validation, three representative records were selected (one correct classification, one partial correct classification, and one noncorrect classification) from the sampling-based evaluation stage, conducted with human reference (subject-matter experts) as shown in Table 2.

For confidentiality purposes, site, management unit, and company identifiers were anonymized, preserving only the technical elements necessary for discussion.

In Case A, the record describes a near-miss event associated with work at height without the use of fall protection. The model generated a consistent simulated response by characterizing the problem, listing probable causes (for example, failure to use PPE and gaps in training or supervision), and identifying risks compatible with falls from height.

Table 2. Evidence of correct, partial correct, and incorrect classifications

Case	Type	Actual situation (record excerpt)	Simulated response (summary)	Model classification (Risk / Control)	Result
Case A (Correct)	Near miss	Work at height without fall protection device (work on elevated structure)	Model describes the problem, probable causes, and fall-related risks, suggesting controls typical of the scenario	Risk: fall from a different level. Controls: safety harness; training	100%
Case B (Partial correct)	Accident	Event in a pressurized system with component rupture and fluid projection	Characterizes the problem, causes, and projections risks and secondary effects; suggests coherent controls, with partial divergence in the control set	Risks: struck by projected materials. Controls: maintenance plan; training; collective protection	83% (risk valid; 2 of 3 controls validated)
Case C (Noncorrect)	Unsafe condition	Excess accumulation of material on conveyor section (material handling operation)	Model describes risk of spillage/obstruction and possible operational causes, but drifts to catalogue classes weakly related to the asset/process	Risks: silo overflow; impact/ collision/overturning. Controls: work instruction; training	0%

At the classification stage, there was direct adherence to the catalog: the risk classified as fall from a different level and the controls safety harness and training are coherent with the damage mechanism of the event and with typical barrier practices for this scenario.

In Case B, the record describes an accident involving pressurized equipment, where a component rupture caused fluid projection that struck the worker. The model correctly identified the main risk mechanism and associated secondary effects, achieving clear adherence in risk classification as struck by projected materials. For controls, three classes were suggested, of which two were validated, resulting in an 83% score. This case illustrates that errors may arise not from misinterpretation of the scenario, but from judgment-dependent selection among multiple plausible control barriers.

In Case C, the record describes an unsafe condition involving material accumulation on a conveyor, for which the model produced a generally plausible textual interpretation. However, when mapping this interpretation to formal catalog classes, the model drifted to misaligned categories, such as silo overflow and collision-related risks, along with generic controls. This reflects a context disambiguation gap likely driven by semantic proximity during retrieval and the absence of structured asset attributes, resulting in zero percent accuracy for both risk and control classifications.

2.3.3 Validation of solution

The validation of the solution's performance within Samarco, although not formally measured through quantitative indicators, was evidenced through stakeholder interactions, which revealed clear value and transformational potential in day-to-day operations. These findings can be represented across four key dimensions:

- Greater Efficiency: Automation of analysis and continuous improvement initiatives optimizes time, reduces rework, and makes processes faster and more agile.
- Cost Savings: Identifying recurring failures enables preventive actions, avoiding operational losses and reducing incidents as well as costs associated with inefficiency and recurrence.
- Actionable Insights: Artificial intelligence identifies patterns across multiple safety monitoring data sources and recommends practical, targeted solutions, enabling faster and more informed decisions.
- Focus on Critical Risks: The system automatically prioritizes the points with the greatest potential impact, ensuring efforts are concentrated where operational and life-safety risks are highest.

2.4 Future steps

Although the solution exhibited a 19% noncorrect classification rate, this result represents both a technical feasibility milestone and a baseline for continuous improvement. The findings highlight that sparse and weakly contextualized records inherently constrain system performance, as missing essential attributes limit the semantic layer's ability to disambiguate scenarios, infer risks and controls, and anchor or validate generated outputs. In some cases, the limitation is even more structural: the absence of records or traceable information reduces the system's capacity to anchor and validate the generated output.

Alongside this fragility, certain risks may materialize, as shown in Table 3, if the use and application of this type of technology are not properly governed.

Therefore, it is evident that, to improve system performance and mitigate these potential risks, a bidirectional strategy is required (Organization to System and System to Organization), structured around three fundamental pillars:

2.4.1 1st Pillar – Safety Culture and Data Quality

The relationship between safety culture and the quality of monitoring data is direct and structural and is aligned with what James Reason defines as an informed culture. The quality of data feeding a monitoring system depends not only on technology, but primarily on human behavior and the level of trust within the organization [11].

In practical terms, data quality and integrity depend on two complementary subcultures:

- Reporting Culture: the organization's ability to record and analyze failures, incidents, and learning opportunities. When reporting is not encouraged, data becomes incomplete or nonexistent, and the organization loses real visibility of its risks [11].
- Just Culture: the foundation that sustains reporting. It ensures that individuals report errors and threats without fear of punishment. In punitive environments, records tend to be distorted or omitted, generating biases that compromise monitoring and limit preventive capacity [11].

Table 3. Potential risk and possible operational impact

Potential Risk	Possible operational impact
Incorrect risk prioritization	Misallocation of focus and resources to secondary causes
Generic controls dominating the dialogue	Superficial actions with low effectiveness on critical risks
Improper use as a final decision tool	Decisions made based on incomplete information

Table 4. Comparison Between Conventional Practice and Proposed Approach (PoC)

Aspect	Conventional Practice	Proposed Approach (PoC)
Data Integration	Data fragmented across multiple systems and spreadsheets	Data integrated and organized within a semantic structure
Type of Analysis	Predominantly descriptive	Descriptive and prescriptive analysis supported by artificial intelligence
Insight Generation	Dependent on analyst judgment and prior knowledge	Automated by large language models (LLMs) and AI, stimulating critical thinking
Action Prescription	Nonexistent or based on tacit expertise	Recommendations suggested based on recognition of contextual patterns
Focus on Priorities	Generalist, with risk of dispersed efforts	Directed toward critical risks and vulnerabilities in controls

2.4.2 2nd Pillar – Integration of New Semantic Components

At the system level, the natural path forward is to reinforce the engine with additional semantic components capable of supporting the understanding of operational contexts [5], namely:

- Taxonomy (controlled vocabulary and hierarchies) to reduce naming variations and noise.
- Ontology (rules, axioms, and consistency constraints) to limit incoherent combinations and increase semantic precision.
- Knowledge Graph to connect assets, conditions, energies, and barriers, increasing contextual disambiguation through relational structure.

2.4.3 3rd Pillar – Semantic Contingency

This pillar involves building a containment barrier oriented toward lexical anchoring, which in practice means defining a stable reference point for the meaning of what is being described. In linguistics, this concept relates to the process by which the meaning of a word or expression is fixed and stabilized within a specific context [12]. Semantic contingency would therefore ensure coherence between what was observed and what was recorded, reducing divergent interpretations and increasing consistency in the mapping to risk and control classes.

2.5 Lessons for industries in digital transformation

Technological solutions are increasingly prevalent in industrial environments, generating large volumes of data that require specialized processing. In many cases, this results in Business Intelligence outputs that still depend on human interpretation to construct narratives and integrate insights for decision-making.

For some industrial segments, this approach is stable and effective; however, in areas such as Risk Management and Health and Safety, where critical insights reside in contextual interpretation of narratives, generating more data and dashboards does not necessarily lead to greater effectiveness.

The results demonstrate that a knowledge-based approach offers clear advantages in building smarter decision-support systems, reducing cognitive overload and improving management effectiveness.

The paradigm shifts and the contrast between traditional and proposed approaches are illustrated in Table 4.

3 Conclusion

This paper presented a joint initiative between Samarco and Simple Safety to develop an intelligent system for monitoring safety risk scenarios, using Semantic Engineering as a knowledge-based layer over existing data to enhance decision-making and strengthen safety culture.

Although part of the identified benefits is associated with the technical perception of those involved, the results achieved so far reinforce the potential of this approach to transform the way industrial organizations manage risks.

Among the key aspects of the project, the following stand out:

- The developed solution functions as an intelligent layer capable of recognizing risk patterns and recommending preventive actions automatically.
- It demonstrated the potential to accelerate decision-making and focus attention on critical risks, based on integrated and contextualized data.
- The application of semantic engineering contributed to greater consistency and traceability in risk analyses.
- The initiative represents a significant advancement in the digital maturity of risk and safety management and a step toward consolidating a more proactive, informed, and resilient safety culture.

Acknowledgements

Innovation is always a demanding endeavor, and when it involves bridging different realities, the challenge becomes even greater. It requires the willingness to face uncertainty and the courage to take risks [13]. We extend our gratitude to Samarco Mineração and Simple Safety for their partnership and collaboration in the development of this initiative and in the preparation of this paper.

References

1 International Organization for Standardization. ISO 22316:2017 – Security and resilience – Organizational resilience – Principles and attributes. Geneva: ISO; 2017.

2 Strutzel T. Presença digital. Rio de Janeiro: Alta Books; 2015. p. 39-52.

3 Acosta ML, Conklin T. Safety capacity: leadership practices for failing safely. Boca Raton: CRC Press; 2024.

4 Rowley J. The wisdom hierarchy: representations of the DIKW hierarchy. Journal of Information Science. 2007;33(2):163-180.

5 Hart C. Taxonomy, ontology, knowledge graph, and semantics. Ontology Explained: Philosophy and AI [Internet]. YouTube; 2025 [cited 2025 Jan 10]. Available at: https://www.youtube.com/watch?v=sr257blfdY8

6 Richens RH. Preprogramming for mechanical translation. Machine Translation. 1956;3(1):20-25.

7 Devlin J, Chang MW, Lee K, Toutanova K. 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 (NAACL-HLT); 2019 June 3-5; Minneapolis, MN. Kerrville: Association for Computational Linguistics; 2019. p. 4171-86.

8 Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. arXiv. 2013.

9 Lewis P, Perez E, Piktus A, Petroni F, Karpukhin V, Goyal N, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. In: Advances in Neural Information Processing Systems (NeurIPS); 2020 December 6-12. San Diego: Neural Information Processing Systems Foundation; 2020.

10 Dekker S, Tooma M. A capacity index to replace flawed incident-based metrics for worker safety. Safety Science. 2022;150:105718.

11 Cooper MD. The safety culture construct: theory and practice. In: Gilbert C, Journé B, Laroche H, Bieder C, editors. Safety cultures, safety models: taking stock and moving forward. Cham: Springer Nature; 2018. p. 47-62. doi:10.1007/978-3-319-95129-4_5.

12 Koch IV, Marcuschi LA. Estratégias de referenciação. In: Koch IV, organizer. Referenciação e orientação argumentativa. 2. ed. São Paulo: Contexto; 2002.

13 Carlomagno M, Scherfer F. Simbiose corporativa. Rio de Janeiro: Alta Books; 2025.

Received: 24 Oct. 2025
Accepted: 11 Feb. 2026

Editor-in-charge: André Luiz Vasconcellos da Costa e Silva