*Article*

# A Generative Artificial-Intelligence-Based Workbench to Test New Methodologies in Organisational Health and Safety

**Andrea Falegnami [1,\*], Andrea Tomassi [1], Giuseppe Corbelli [2], Francesco Saverio Nucci [3] and Elpidio Romano [1]**

[1] Management Engineering Faculty, Uninettuno University, 00186 Rome, Italy; andrea.tomassi@uninettunouniversity.net (A.T.); elpidio.romano@uninettunouniversity.net (E.R.)
[2] Psychology Faculty, Uninettuno University, 00186 Rome, Italy; giuseppe.corbelli@uninettunouniversity.net
[3] Research and Innovation Area, Uninettuno University, 00186 Rome, Italy; francesco.nucci@uninettunouniversity.net
**\*** Correspondence: andrea.falegnami@uninettunouniversity.net

**Featured Application**: LLM integration in generating new methodologies in OHS.

**Abstract:** This paper introduces a novel generative artificial intelligence workbench specifically tailored to the field of safety sciences. Utilizing large language models (LLMs), this innovative approach significantly diverges from traditional methods by enabling the rapid development, refinement, and preliminary testing of new safety methodologies. Traditional techniques in this field typically depend on slow, iterative cycles of empirical data collection and analysis, which can be both time-intensive and costly. In contrast, our LLM-based workbench leverages synthetic data generation and advanced prompt engineering to simulate complex safety scenarios and generate diverse, realistic data sets on demand. This capability allows for more flexible and accelerated experimentation, enhancing the efficiency and scalability of safety science research. By detailing an application case, we demonstrate the practical implementation and advantages of our framework, such as its ability to adapt quickly to evolving safety requirements and its potential to significantly cut down development time and resources. The introduction of this workbench represents a paradigm shift in safety methodology development, offering a potent tool that combines the theoretical rigor of traditional methods with the agility of modern AI technologies.

**Keywords:** OHS; ChatGPT; AI; chain of thought; complexity; design science research; methodology design

## 1. Introduction

The development and validation of novel methodologies in the field of safety science, particularly in the areas of organisational resilience and occupational health and safety (OHS), requires rigorous testing and a substantial amount of high-quality data [1,2]. However, obtaining this data presents a number of practical and ethical challenges, from obtaining approval for human-centred studies to navigating complex data protection regulations [3]. These barriers often limit researchers' abilities to efficiently prototype and evaluate new methods, hindering progress in areas where timely innovation is critical. To address these limitations, this study presents an integrated approach that uses generative artificial intelligence (AI) tools to create a flexible and accessible workflow workbench for testing and refining early-stage methods, specifically in the areas of occupational safety and organisational resilience.

The impetus for this research stems from the development of an integrated Resilience Analysis Grid-RASCH method, a novel approach to improving organisational resilience by quantifying and analysing adaptive capacity [4–7]. To assess the viability of this method without incurring the full cost and complexity of data collection and extensive

real-world testing, we employ a structured workflow based on generative AI tools, such as large language models (LLMs). These tools have shown great promise in simulating complex environments and generating synthetic data, allowing researchers to model and test hypotheses in a controlled and reproducible manner before committing to extensive field trials [8,9].

This article demonstrates how generative AI can serve as a powerful tool for prototyping and iterating new methods in an OHS context. While LLMs and other advanced neural networks typically require extensive computational resources and expertise to develop, recent advances have made accessible and relatively user-friendly tools available to non-specialist researchers and practitioners [10,11]. The core contribution of this paper is thus twofold: (1) to illustrate a generalisable model for the use of generative AI in method development in safety science; and (2) to present a practical walkthrough of a conceptual workflow workbench that can support researchers in the preliminary testing and refinement of OHS methods.

In the literature on the application of AI in OHS, one fundamental study is by Jason Wei et al. (2022) [12], which outlines the use of traditional machine learning algorithms for predicting workplace accidents based on historical incident data. While this approach has proven moderately effective, it fundamentally lacks the ability to adapt to novel or evolving safety conditions, a limitation that our research addresses through the use of LLMs [12]. Moreover, at the present time (December 2024), the Scopus database does not report any document regarding OHS and Chain-of-thought concurrently. Therefore, the main outcome of this paper is to propose a structured methodology to generate specific methodologies for OHS based on Chain-of-thoughts prompting. Our workbench leverages the generative capabilities of LLMs not only to predict but also to simulate potential safety scenarios, thereby filling a critical gap in dynamic risk assessment. This method significantly extends the utility of AI in safety science, offering a more robust, adaptable framework that is capable of keeping pace with changes in workplace environments and safety protocols.

Given the high stakes of safety-related research and the potential ethical constraints on data collection, this framework serves as a valuable resource for those seeking to validate the applicability and robustness of new approaches. In describing the proposed model, we detail how AI tools can be tailored to specifically support OHS applications while remaining accessible to researchers without specific training in machine learning or artificial intelligence. This workbench not only addresses current challenges of data scarcity and ethical considerations but also provides a scalable model for the broader field of safety science, bridging the gap between concept and implementation under the lens of a design science approach.

The article is structured as follows: The Section 2 contextualises the research against the background of recent international literature; Section 3 details the methodological framework proposed by the research team; Section 4 guides the reader through an example application on how to implement the framework in their own OHS projects; finally, Section 5 draws conclusions and outlines some potential applications of the framework proposed here.

## 2. Literature Background

Recent research highlights the significant impact of artificial intelligence (AI) on occupational health and safety (OHS) and organisational resilience. AI-driven technologies are revolutionising workplace safety by providing predictive insights, real-time monitoring, and risk mitigation strategies [13]. In high-risk sectors such as construction, mining, and oil and gas, AI applications include computer vision, sensor networks, and machine learning for hazard identification and risk management [14]. Generative AI and LLMs are transforming several industries, with applications ranging from healthcare to finance [15]. In the construction industry, generative AI offers innovative solutions to productivity challenges, with retrieval-augmented generation improving basic LLM performance

when querying contract documents [16]. These advances underscore the potential of AI to improve workplace safety, productivity, and decision-making across multiple sectors while also highlighting the need for responsible innovation and ethical considerations.

AI and machine learning are increasingly being applied to OHS in high-risk industries. These technologies offer real-time monitoring, risk assessment, and personalised training solutions. However, the development and validation of OHS methodologies face challenges, including the need for accurate and ethically collected data and barriers due to data protection laws [13]. Synthetic data generation using AI has emerged as a potential solution to these data-related issues, enabling data sharing and use where real-world data are insufficient or inaccessible [17]. Various ML approaches are being explored for OHS applications, with ongoing research aimed at identifying the most appropriate techniques for specific problems in the field [18]. Despite its potential, barriers to the adoption of AI in OHS include high costs, the lack of skilled personnel, and ethical concerns [14].

Recent advances in generative AI and LLMs have shown significant potential for synthetic data generation, especially in scenarios with limited data availability [19]. These technologies are revolutionising OHS by providing predictive insights, real-time monitoring, and risk mitigation strategies. Generative AI, exemplified by models such as ChatGPT and DALL-E, has rapidly developed and found applications in various sectors, including healthcare and finance [15]. Synthetic data generation is emerging as a promising solution to address challenges such as data quality, insufficient data points, and privacy concerns in machine learning applications [17,20]. This approach enables data sharing and exploitation in ways that are not possible with real-world data and spans domains such as computer vision, speech, natural language processing, and healthcare. However, privacy and fairness concerns related to synthetic data generation remain critical issues that require further research and consideration [21,22].

The principles of design science provide a structured approach to developing and evaluating methodological frameworks in applied research fields. Johannesson and Perjons provided a comprehensive introduction to design science, outlining its core activities and ethical considerations [23]. This approach can be applied to improve literature review methods, as demonstrated by Galli et al., who developed a tool using latent semantic indexing to improve the efficiency of literature searches [24]. Sturm and Sunyaev further contributed to the field by deriving design principles for systematic search systems that aim to increase comprehensiveness, accuracy, and reproducibility [25]. The integration of artificial intelligence (AI) within such frameworks can address methodological gaps in several fields, including occupational safety and health (OSH). Pishgar et al. present the REDECA framework for reviewing AI applications in OSH, highlighting the potential for AI to improve risk identification and control in various industrial sectors [26]. In light of these advancements, the integration of AI-driven methodologies, particularly through generative AI and design science principles, presents a compelling opportunity to address existing challenges and enhance innovation in occupational health and safety (OHS) and organisational resilience, paving the way for responsible and effective method development in safety science.

### 3. Proposed Workflow Workbench

The research team aimed to develop a new method guided by the principles of Design Science Research (DSR). DSR is an approach used to create and evaluate artefacts (such as models or methods) intended to solve specific problems. This process starts with the conceptual design of an abstract model or method, followed by its practical implementation and testing through a prototype. The artefact created by the team is both idiographic and prescriptive. 'Idiographic' means it is tailored to a specific case or context, focusing on particularities rather than broad generalizations. 'Prescriptive' refers to providing structured models and actionable methods that offer direct guidance for problem-solving processes, with clear guidelines tailored for the unique setting under study. This methodology effectively bridges the gap from general theoretical insights to specific,

practical applications, making a substantial contribution to both research and practice [27]. This work reflects recent trends in constructivist methods, which focus on the emergent construction of artefacts. These methods integrate a mix of qualitative and quantitative data to develop adaptive, data-driven frameworks capable of evolving with new insights. Such approaches are crucial for synthesising complex information systems through ongoing development and fine-tuning specific to the context [28]. By employing AI for tasks such as clustering and knowledge graph creation, the team managed to weave complex data into a unified artefact that organically exposes patterns and insights [29]. Grounded in complexity theory, this method mirrors the dynamics of natural systems, enabling the structure and flexibility necessary for a nuanced, emergent understanding of interconnected elements within the system [30,31].

### 3.1. General Structure

The general structure of the proposed workflow consists of two main building blocks: the use of an LLM and a synthetic data generation package. The former is used to bend the knowledge implicit in unstructured data sources to the specific context in which the method designer needs to operate; the latter is used to generate a set of tests for the operational implementation of the method and to prove its practical feasibility.

### 3.1.1. LLM-Based Block

The design of the LLM-based block is based on several fundamental considerations about LLMs. These models operate as probabilistic tools based on the concept of lexical co-occurrence: the frequency with which two terms occur together in a corpus determines the strength of their semantic association. Therefore, certain widely recognised terms are more likely to have been included in the training of LLMs, as they reflect general or well-known knowledge that is abundant in available data sources; for example, general terms such as the "Likert scale" are common, as evidenced by approximately 896,000 results in Google Scholar. In contrast, less common, specialised terms are less likely to be represented in the LLM's training data; specific concepts such as "Resilience Analysis Grid" yield significantly fewer references—around 896 results, or about $10^{-3}$ of those associated with the "Likert scale". Consequently, the quality and relevance of the LLM's output is intrinsically dependent on the breadth and depth of information contained within its training corpus, essentially reflecting the model's pre-existing knowledge competence. However, this competence can be enhanced by supplementing the model with additional context-specific resources through techniques such as retrieval-augmented generation or by providing documents tailored to the specific context in question [32,33].

In order to optimise the performance of the LLM in this area, we developed a progressive prompt engineering workflow, which acts as a structured form of Socratic dialogue designed to encourage iterative learning and adaptation [34]. This approach alternates between two different types of prompts, called *framework* and *context* prompts.

The framework prompt involves the user providing the artificial agent with broad, general information about the nature of the tasks to be performed, thereby establishing a basic structure for the interaction.

The context prompt enables the user to provide precise, context-specific information, allowing the LLM to refine and apply the framework to the task in an operationally relevant way. This alternation between framework and context prompts helps the model adapt general principles to specific situations, fostering a dynamic, responsive workflow. Figure 1 illustrates the dynamic interaction process between a human user and a system based on a large language model (LLM) through structured prompt engineering. Initially, the user provides a general framework or outline, which furnishes the LLM with high-level information about the task or activity (a). This initial stage sets a broad context and establishes fundamental guidelines for the LLM's operation, effectively shaping its subsequent outputs. Following this, the LLM produces initial results based on the user's framework (b). These preliminary outputs allow the user to assess the LLM's work and provide

feedback, creating an iterative loop of refinement. This feedback is crucial as it guides the LLM towards generating more precise and relevant outcomes.

As the interaction progresses, the user specifies additional, detailed information relevant to the task at hand (c). This step is critical as it enables the user to tailor the initial framework to meet the specific requirements of the task, ensuring that the LLM's responses are appropriately aligned with the task's unique needs.

In the final step, the LLM utilises the enhanced framework and specific context provided by the user to deliver detailed answers or solutions (d). These responses are crafted to integrate general principles with the specific demands of the task, thereby enhancing the LLM's ability to provide accurate, relevant, and contextually tailored answers. This sequential and adaptive process underpins a responsive and context-sensitive workflow that significantly boosts the LLM's effectiveness in practical applications.
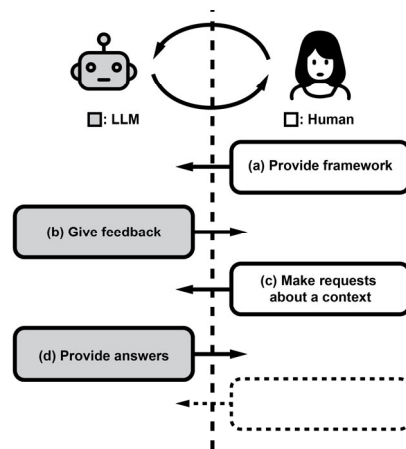


**Figure 1.** Diagram showing the step-by-step interaction between a human user and an LLM-based system through structured prompt engineering. Initially, in phase (**a**), "Provide framework", the user defines a general structure for the LLM, setting broad guidelines. In phase (**b**), "Give feedback", the LLM generates initial outputs, which the user refines through feedback, enhancing precision. Phase (**c**), "Make requests about a context", involves the user specifying details to tailor the framework to the task's unique needs. Finally, in phase (**d**), "Provide answers", the LLM delivers detailed solutions aligned with both general principles and specific task requirements, illustrating a responsive and adaptive workflow that enhances the LLM's effectiveness. The entire process can be subject to continuous feedback loops at each intermediate stage and can virtually extend indefinitely, as summarised by the curved arrows between the two agents and the dashed box, respectively.

### 3.1.2. Synthetic Data Generation Block

The synthetic data generation block is based on several critical considerations that are essential for overcoming the inherent challenges of data collection, particularly in areas such as OHS, where data collection can be both costly and resource intensive. Traditional methods often require extensive data collection sessions, such as semi-structured interviews, naturalistic observations, or questionnaires, all of which require a significant and sometimes impractical level of participant engagement [35]. In addition, real-world data collected using these methods are often scarce, incomplete, and subject to various types of data entry errors, thus compromising data quality. Given these limitations, it is neither practical nor advisable to rely on untested measurement methods in research campaigns, and yet a paradox remains: how can such methods be tested without reliable data? To address this issue, freely available software packages are now available that allow the generation of test data tailored to the specific needs of the research, taking into account different data types (structured or unstructured), specific probabilistic distributions, levels of respondent agreement (e.g., Cronbach's Alpha), and even variable rates of missing or incomplete responses within the data set matrix.

Figure 2 outlines the general process for generating synthetic data, detailing each step from identifying the operating conditions of the test to implementing synthetic data in method trials. The synthetic data generation workflow is detailed through a structured, multi-stage process. The initial stage, "Identify the test's operating conditions", establishes the foundational conditions and requirements of the test environment (a). This involves defining the scope and constraints under which synthetic data will be utilized. Subsequently, "Identify the typology of needed data" determines the specific type of data necessary for the research, distinguishing between structured and unstructured formats, along with any relevant characteristics essential for accurate simulation (b). Following this, the stage titled "Identify the experimental data parameters" involves defining key parameters that the synthetic data must adhere to, such as distribution properties, levels of consistency among variables, and patterns of missing data (c). These parameters are critical as they guide the data generation process to ensure the synthetic data meet the experiment's specific requirements. The next step, "Select the proper synthetic generator", involves choosing the most suitable tool for generating the data based on the previously identified data type and parameters (d). For instance, text-based data may be generated using a language model, while statistical data may employ parametric tools or Monte Carlo simulations. After selecting the appropriate generator, the process moves to "Generate synthetic data according to params." In this phase, synthetic data are produced according to the defined parameters, resulting in a data set that reflects the characteristics specified in earlier steps (e). The final step, "Implement synthetic data in the methods", sees the synthetic data being integrated into the research methods, enabling practical testing and validation of the measurement tools and methodologies under evaluation (f). This systematic approach ensures robust data generation that is tailored to meet the experimental needs, facilitating comprehensive testing under controlled conditions.
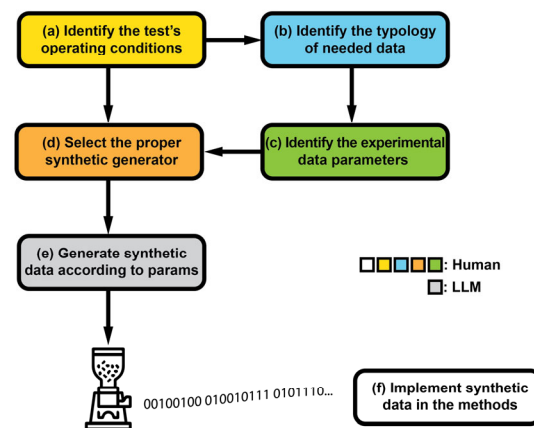


**Figure 2.** Synthetic data generation workflow depicting each stage of the process. In (**a**), yellow box, "Identify the test's operating conditions", initial conditions and requirements are set, establishing the data generation foundation, including scope and constraints. In (**b**), cyan box, "Identify the typology of needed data," the necessary data type for the research is specified, distinguishing between structured and unstructured formats. In (**c**), green box, "Identify the experimental data parameters", key parameters such as data distribution and consistency levels are defined. In (**d**), orange box, "Select the proper synthetic generator", the appropriate tool for data generation is chosen based on the data type and parameters, such as using a language model for text or parametric tools for statistical data. In (**e**), grey box, "Generate synthetic data according to params", data are produced according to the specified characteristics. Finally, in (**f**), white box, "Implement synthetic data in the methods", the generated data are integrated into research methods for testing and validation. This streamlined approach ensures robust, tailored data generation for effective testing under controlled conditions. All the boxes, except the grey one, correspond to actions performed by human agents.

The selection of the appropriate synthetic data generation tool is influenced by the nature of the method—and therefore the type of data—being tested. For example, methods involving unstructured textual data might use LLMs as synthetic data generators, while methods requiring data that follow simple statistical distributions might use standard parametric generators based on Monte Carlo methods. The workflow involves first identifying the operating conditions for the test, followed by determining the required data type and experimental data parameters, selecting the appropriate synthetic generator, and finally generating synthetic data according to these defined parameters before integrating it into the methodological test.

## 4. Walkthrough Application

In developing a comprehensive approach to integrating the Resilience Analysis Grid (RAG) with the Rasch model, an operator must establish a methodological framework that supports both data definition and synthetic data generation tailored to the OHS context.

The methodological walkthrough of this framework begins with the user establishing a basic 'framework' by referring to a broad, generic questionnaire, such as Hollnagel's RAG, which provides a structure around OHS contexts (see Figure 3).

This framework acts as a reference point that the LLM can interpret, initially assessing the outline and feeding back its understanding to the user in a feedback loop. This initial response allows the user to assess whether the LLM has accurately captured the general framework, ensuring that any future refinements are based on a correct interpretation of the high-level goals. Next, the user makes specific contextual requests, starting with the adaptation of the general RAG questionnaire to the specific OHS context of manufacturing [36,37]. This requires the LLM to translate the general questionnaire items into those relevant to the four core competencies in manufacturing and to align them with the unique requirements and characteristics of this sector.

The user then adds further specificity by introducing a Likert scale format for capturing responses, thus modifying the general framework with a further contextual layer specifically tailored to OHS in manufacturing. The LLM adapts the general structure of the Likert scale by interpreting it in the context of each question, and the user refines these interpretations to ensure consistency with manufacturing-specific needs. With the basic questionnaire and response format now customised, the user defines the technical requirements for synthetic data generation and determines that the Likert scale data type is required to simulate realistic response patterns.

The user proceeds by identifying the operating conditions required to integrate a Rasch model with the RAG data, a step that involves ensuring that the synthetic data fit a single latent dimension, an essential criterion for the application of the Rasch model. Once these criteria are established, the user specifies other data parameters, such as the number of fictitious respondents and the dimensions required for each respondent's responses, as well as distributional properties and internal consistency measures, such as Cronbach's alpha, tailored to the experimental needs of the OHS study. On the basis of these specifications, the user selects the appropriate synthetic data generation tool (in this case the "LikertMakeR" package, an R-based tool particularly suited to generating data that meet the predefined distribution and scaling conditions [38]).

The selected generator then produces the synthetic data according to the established parameters, creating a data set that reflects real-world response patterns to OHS-related questions in the manufacturing sector.
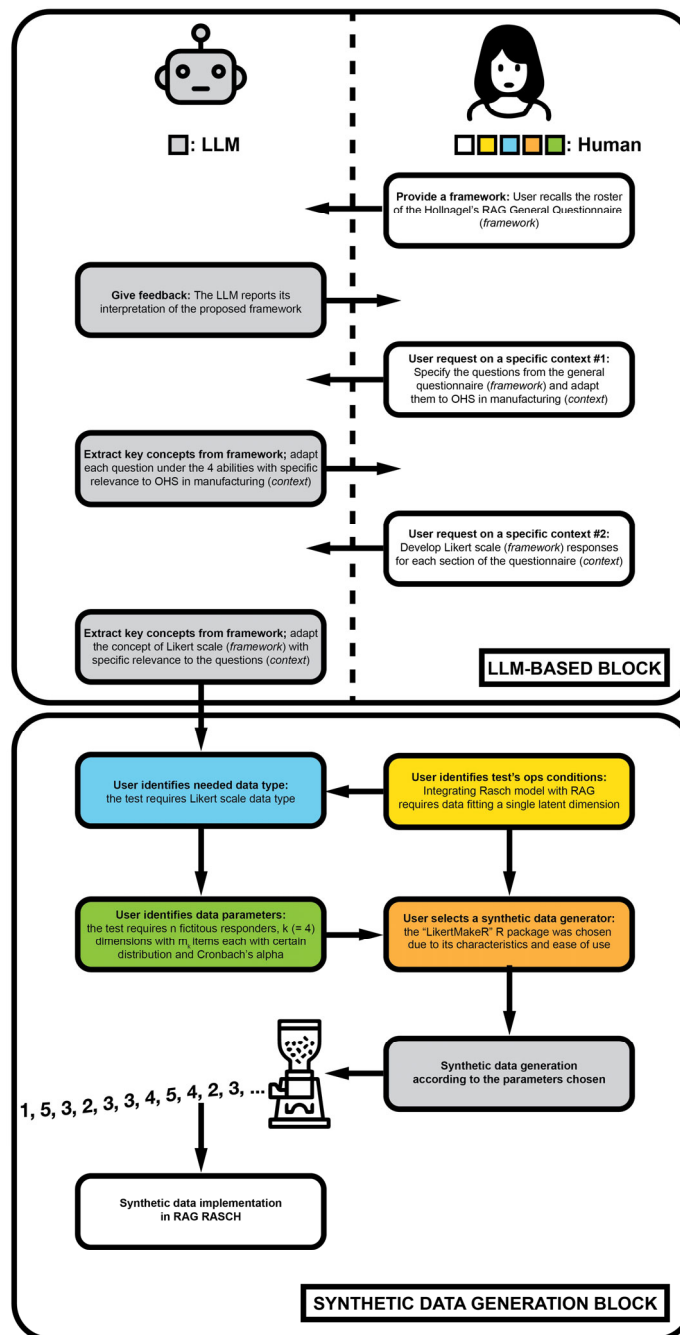
**Figure 3.** Step-by-step workflow illustrates the interaction between a human user and an LLM-based system within a structured prompt engineering and synthetic data generation framework. The two blocks constituting the proposed methodology are clearly highlighted. Initially, the user sets a basic framework based on Hollnagel's RAG general questionnaire for OHS relevance. The LLM then provides feedback to refine its model interpretation. Next, the user requests contextual adaptations to tailor the questionnaire for OHS in manufacturing, leading to the LLM adjusting questions to reflect four core OHS competencies. Further, the user requests Likert scale responses, specifies the type of data required, and sets test conditions to fit a Rasch model, including specifying data parameters like the number of responses and Cronbach's alpha for consistency. The "LikertMakeR" R package is chosen to generate data fitting these criteria, producing a data set that mimics real-world OHS scenarios in manufacturing. These data are then implemented within the RAG-Rasch framework for testing and validation, completing a robust, context-specific methodological workflow for early-stage OHS methodology testing. The colour code in use corresponds to that in Figure 2.

These data are then ready for implementation in the RAG-Rasch methodology testing phase, where they are analysed within the structured framework that has been iteratively refined through the combination of LLM-guided dialogue and synthetic data parameterization. This process allows for a comprehensive simulation of the RAG in an OHS manufacturing context, providing a robust testbed for refining and validating methodologies without the logistical and ethical complexities of large-scale, real-world data collection. Through this iterative workflow, based on alternating framework and context prompts, the LLM is able to dynamically adapt general principles to specific applications, while the synthetic data generation tools support rigorous and replicable testing, creating a powerful platform for early-stage methodological development in OHS research. The resulting questionnaire, as well as the corresponding allowed answers, are reported in Tables A1–A4, collected in Appendix A. Appendix B reports the original Socratic dialogue between a member of the research team and the LLM about the creation of the RAG questionnaire tailored to the OHS domain.

The walkthrough application presented in this paper is based on a real case used to propose the novel combination of RAG with the Rasch Rating Scale model. The resulting RAG-Rasch is presented more in detail in a companion article, entitled "Introducing the RAG-Rasch Rating Scale Model for Measuring Organizational Resilience Potentials", belonging to the same special issue of this journal. To demonstrate the scalability potential of this Chain-of-thought-based approach, a set of other possible use cases are listed:

- Starting from the general structure of the Analytic Hierarchy Process, the human asks the LLM to provide a specific AHP-based cost-benefit analysis on the implementation of Lock Out Tag Out procedures in a dairy factory, also defining the parameters for a synthetic discrete time simulation of the packaging machines;
- Starting from a typical measuring chain for an experimental setting in hydrogen natural storage, the goal is to develop a Functional Resonance Analysis Method model representing the sociotechnical system under the focus of the analysis and also providing the phenotypes for the associated Monte Carlo simulation;
- Starting from the general structure of Bayesian networks, the human asks the LLM to provide a specific model to evaluate emergency response strategies in high-rise buildings. This involves defining parameters for a synthetic simulation, such as building height, population density, and response times, which will demonstrate the effectiveness of various safety measures across different building configurations;
- Starting from Decision Tree analyses, the human asks the LLM to develop a model to assess optimal personal protective equipment (PPE) usage in chemical manufacturing. The model includes defining decision nodes based on chemical hazards and PPE types and parameters for synthetic scenarios that adapt to different plant sizes and exposure levels;
- Starting from dynamic systems modelling, the human agent asks the LLM to create a model for analysing supply chain resilience in the retail industry. This involves defining parameters that simulate disruptions like delivery delays and supply shortages due to the pandemic and tailoring the model to different scales of retail operations and geographic variables;
- Starting from the agent-based modelling notions, the human agent asks the LLM to simulate the spread of infectious diseases within corporate offices and evaluate health interventions. The model will require defining parameters such as office size, interaction patterns, and intervention efficacy, aiming to provide tailored health and safety strategies for different office environments;
- Starting from system dynamics theory, the human asks the LLM to analyse energy consumption patterns and sustainability strategies across various industrial complexes. The setup involves defining dynamic parameters such as energy source types, machinery efficiency, and operational practices, which adapt the model for diverse industrial sectors and scales.

The examples outlined illustrate the scalability of our proposed methodology across diverse industries and conditions, demonstrating its broad applicability. Each implementation, while theoretically sound, must be rigorously validated by researchers to ensure accuracy and reliability. This continuous validation is crucial to effectively adapt and refine the methodology for specific real-world applications.

## 5. Conclusions

This research represents a transformative advance in the field of OHS, providing a novel AI-driven workbench that not only enables the design and testing of innovative methodologies but also addresses long-standing challenges in early-stage method validation. By integrating LLMs and synthetic data generation tools, this framework creates a dynamic platform that can simulate and evaluate new OHS methodologies with unprecedented efficiency and rigour, effectively eliminating the practical and ethical constraints typically associated with real-world data collection. The applications of this approach are far-reaching: it provides OHS researchers and practitioners with a powerful tool for prototyping and refining methodologies, ultimately accelerating the adoption of data-driven safety interventions in industries where rapid, evidence-based responses are critical. This workbench has the potential to fundamentally reshape safety science, enabling the iterative and responsive development of methodologies that can be tailored to different sectors and adapted to complex and evolving workplace environments. The implications of this research extend far beyond OHS, positioning it as a valuable model for other high-risk areas where accurate, context-specific data are difficult to obtain, such as healthcare, environmental monitoring, and disaster management. The ability to test new safety protocols in a simulated environment paves the way for more agile, reliable, and proactive risk management strategies, fostering a safer and more resilient organisational culture. Despite its strengths, there are limitations due to the synthetic nature of the data, which, while highly customisable, may lack the complete unpredictability and variability of real-world scenarios. In addition, the reliance on LLMs requires careful oversight to ensure that the outputs generated are accurately aligned with sector-specific requirements. It is also crucial to recognise the inherent biases present within all LLMs, which can be tracked back to the training stage in which the word embeddings are built. Since LLMs have to be trained on a large amount of textual data, it is very difficult to avoid the presence of biases inherited from the sources. This is a well-documented issue of such models that is currently a trending research topic both for industry and academia. This is another reason to keep the human in the loop and not to make the process fully automated. However, the authors want to warn the audience to take into account any potential discrepancy in AI-generated results that can lead to other biases where common knowledge is overemphasised while niche, specific knowledge pertinent to particular fields, such as safety sciences, is marginalized. To counteract these biases and enhance the quality and relevance of synthetic data, one might employ advanced techniques such as retrieval-augmented generation. This method, not addressed in this paper, could supplement the LLM with targeted, context-specific documents that enrich its understanding and output, ensuring a more balanced and accurate generation of synthetic data that are reflective of the specialised terminology and concepts critical to the OHS research domain.

Future developments could focus on improving the realism of synthetic data models and extending the applicability of the framework to different high-risk domains. This research thus lays the foundation for a new era in occupational health and safety science, providing a scalable, adaptable, and scientifically robust tool that enables organisations to anticipate, assess, and mitigate risk with unprecedented precision and foresight.

**Author Contributions:** A.F.: Conceptualization, Methodology, Software, Data curation, Writing—original draft preparation, Results interpretation, Figures realization, Validation, Reviewing and Editing; A.T.: Conceptualization, Methodology, Software, Data curation, Writing—original draft preparation, Results interpretation, Figures realization, Validation, Reviewing and Editing; G.C.:

Methodology, Formal Analysis, Software, Data Curation, Writing—original draft preparation, Results interpretation, Validation; F.S.N.: Writing—original draft preparation, Results interpretation, Validation; E.R.: Writing—original draft preparation, Results interpretation, Validation. All authors have read and agreed to the published version of the manuscript.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data included in article is referenced in the article. The data sets generated and/or analysed during the current study are available from the corresponding author on reasonable request.

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Appendix A**

*OHS RAG Questionnaire*

**Table A1.** Ability to Respond (OHS in Manufacturing Context).

| ID | Item | Question | Possible Answers | | | | |
|---|---|---|---|---|---|---|---|
| | | | **1** | **2** | **3** | **4** | **5** |
| R1 | Event list | What are the key OHS incidents for which the system has predefined responses (e.g., machinery accidents, chemical spills, fire, electrical hazards, slips/trips)? Are there prepared responses for different types of injuries (e.g., fractures, burns)? | No prepared responses for any events | Very few events have prepared responses | Some events have prepared responses, but important gaps exist | Most events have prepared responses | All critical events have detailed, prepared responses |
| R2 | Background | How were these OHS events identified as critical? Was it based on risk assessments, regulatory requirements (e.g., OSHA, EU regulations), historical incidents within the manufacturing sector, or industry-wide trends? | No clear rationale for event selection | Some events were selected with rationale, but most were not | Selection process is based on partial evidence or incomplete analysis | Selection process is mostly thorough and follows industry standards | Event selection is fully based on comprehensive risk assessments and regulatory requirements |
| R3 | Relevance | When was this list of OHS-related events created? How frequently is it reviewed or updated? Is the list modified after a near-miss, safety audit, or based on new regulatory requirements? Who is responsible for updating it—HR, Safety Officers, or the Compliance Team? | The list has never been revised | The list is rarely revised (e.g., only in response to incidents) | The list is updated occasionally (e.g., after audits or inspections) | The list is regularly updated based on industry trends or regulatory changes | The list is continuously and proactively revised with a formal review process |
| R4 | Threshold | What is the threshold for activating a safety response (e.g., injury severity, exposure to hazardous substances)? Does it vary depending on internal factors like production demands or external factors like weather conditions? Is there a trade-off between safety and maintaining production schedules? | There are no defined criteria for triggering a response | Criteria exist but are vague or rarely followed | Some criteria are clear, but others remain inconsistent or dependent on factors like production pressure | Most criteria are well-defined and generally followed | Criteria are fully defined, specific, and consistently applied across all OHS scenarios |
| R5 | Response list | How was the list of specific OHS responses developed (e.g., use of safety standards like ISO 45001)? Is it based on empirical evidence from accidents in the manufacturing sector, or on model simulations? | The response list is inadequate for almost all events | The response list is inadequate for many events | The response list is somewhat adequate, but significant gaps remain | The response list is adequate for most events, but a few gaps exist | The response list is comprehensive and fully adequate for all expected events |

| | | | | | | |
|---|---|---|---|---|---|---|
| R6 | Speed | How quickly can the safety team respond to different types of incidents (e.g., how fast can machinery be shut down in an emergency)? What measures ensure quick response (e.g., location of first-aid kits, accessibility of fire extinguishers, on-site medical staff)? | Responses are slow and usually delayed | Responses are often slow but occasionally timely | Responses are reasonably quick but may face occasional delays | Responses are quick and timely in most cases | Responses are implemented rapidly and effectively in all cases |
| R7 | Duration | For how long can an emergency response (e.g., handling a chemical spill) be maintained before additional resources are required? What is the minimum acceptable response level (e.g., having a trained first responder on-site) and how long can that level be sustained? | Full response cannot be sustained for any meaningful time | Full response can be sustained only for a short period | Full response can be sustained for a moderate amount of time | Full response can be sustained for an extended period | Full response can be sustained indefinitely, or as long as needed |
| R8 | Stop rule | What criteria are used to determine when it is safe to resume normal operations after an incident (e.g., after machinery has been inspected post-accident or air quality has been verified post-gas leak)? | There are no defined criteria for stopping a response | Stop criteria are vague or inconsistently applied | Stop criteria exist but are not always followed or well understood | Stop criteria are clearly defined and generally followed | Stop criteria are well-defined, consistently applied, and understood by all employees |
| R9 | Response capability | What resources (e.g., personal protective equipment (PPE), fire suppression systems, spill control materials) are allocated to ensure readiness? How many workers are trained for OHS emergency responses, and who is responsible for maintaining this readiness (e.g., Safety Officers, OHS department)? | Resources are wholly inadequate | Resources are often insufficient to maintain readiness | Resources are adequate but stretched thin at times | Resources are mostly sufficient, with minor occasional limitations | Resources are fully adequate and consistently available for all OHS needs |
| R10 | Verification | How is OHS response readiness maintained (e.g., regular drills, inspections)? How often is readiness verified and by whom (e.g., through third-party safety audits, internal audits)? | Readiness is never or rarely verified | Readiness is verified occasionally but without consistency | Readiness is verified, but the process is incomplete or irregular | Readiness is regularly verified and mostly thorough | Readiness is systematically and consistently verified with comprehensive testing |

**Table A2.** Ability to Monitor (OHS in Manufacturing Context).

| ID | Item | Question | Possible answers | | | | |
|---|---|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 | 5 |
| M1 | Indicator list | How are safety performance indicators defined in the manufacturing context? Are they based on historical incident data, regulatory requirements (e.g., incident rates, near-miss reports), or industry best practices? | No indicators are defined | Indicators are poorly defined and lack relevance to OHS risks | Some indicators are well-defined, but important areas are missing | Most indicators are clearly defined and relevant to OHS risks | Indicators are comprehensively defined and fully relevant to all OHS risks |
| M2 | Relevance | When was the OHS indicator list created? Is it regularly updated based on audits, risk assessments, or changing legal requirements? Who is responsible for maintaining the list (e.g., the Safety Officer or Compliance Manager)? | The list is outdated and irrelevant | The list is mostly outdated with minimal relevance to current risks | The list is somewhat relevant but needs regular updates | The list is relevant and updated fairly regularly | The list is fully up-to-date and relevant, with a clear process for regular revisions |
| M3 | Indicator type | Are the OHS indicators leading (e.g., number of near-misses, safety training completion rates) or lagging (e.g., injury frequency, lost workdays)? Are they based on individual incidents or aggregated data from across multiple workstations or facilities? | Only lagging indicators are used | Mostly lagging indicators, with very few leading ones | A balance of leading and lagging indicators is present, but needs improvement | A strong balance between leading and lagging indicators is maintained | Leading and lagging indicators are fully integrated and balanced |
| M4 | Validity | How is the validity of OHS indicators ensured? Are they tied to an articulated risk management model (e.g., hazard identification and control) or just general safety guidelines? | Indicators lack validity and are not tied to any systematic model | Indicators have limited validity, often relying on assumptions | Some indicators are valid, but others are based on informal processes | Most indicators are valid and based on articulated OHS models | Indicators are fully valid, systematically derived from risk models or empirical data |
| M5 | Delay | What is the typical lag between a safety incident (e.g., injury, equipment malfunction) and the reporting of the corresponding indicator? Is the delay acceptable for timely corrective actions? | The delay is unacceptable and causes significant problems | The delay is too long and hampers corrective actions | The delay is manageable but should be improved | The delay is acceptable in most cases | The delay is minimal and fully acceptable for all indicators |
| M6 | Measurement type | Are OHS measurements qualitative (e.g., employee feedback on safety culture) or quantitative (e.g., incident rates, exposure levels)? If quantitative, what scales are used (e.g., accident severity scales)? | Measurements are exclusively qualitative with no quantifiable data | Measurements are mostly qualitative with minimal quantitative data | A mix of qualitative and quantitative measurements is used, but it's inconsistent | Measurements are mostly quantitative, with some qualitative inputs | Measurements are systematically quantitative, with qualitative inputs where necessary |

| M7 | Measure-ment fre-quency | How often are safety indicators measured? Is it con-tinuous (e.g., real-time monitoring of air quality or noise levels) or periodic (e.g., monthly safety audits)? | Measurements are rarely taken and are incon-sistent | Measurements are taken occa-sionally but not regularly | Measurements are taken periodically, but not with enough frequency | Measurements are taken regularly and with sufficient fre-quency | Measurements are taken continuously or at very regular intervals, ensuring constant monitoring |
| --- | --- | --- | --- | --- | --- | --- | --- |
| M8 | Analy-sis/inter-pretation | What is the time gap between collecting safety data (e.g., incident reports) and analyzing them? How many indicators require deeper analysis (e.g., trend analysis of near-misses), and how are the results com-municated to workers and management? | There is a signifi-cant delay be-tween measure-ment and analy-sis | Analysis is slow and often incom-plete | Analysis is timely but requires im-provement for deeper insights | Analysis is timely and mostly effective | Analysis is prompt, thor-ough, and leads to action-able insights quickly |
| M9 | Stability | Are the measured safety effects temporary (e.g., re-duced accidents during a safety campaign) or long-lasting (e.g., consistently low incident rates over years)? | The measured ef-fects are highly transient and do not last | The effects are mostly transient with occasional lasting impacts | The effects are somewhat stable, but not consistently long-lasting | The effects are mostly stable and sustained over time | The effects are highly sta-ble and consistently per-manent |
| M10 | Organiza-tional sup-port | Is there a structured schedule for safety inspections (e.g., weekly equipment checks, quarterly safety au-dits)? Is there adequate resource allocation for ongo-ing monitoring (e.g., dedicated Safety Officers, budget for safety improvements)? | There is little to no support for OHS monitoring | The organization provides mini-mal support for OHS monitoring | There is some sup-port, but it is incon-sistent or insuffi-cient | The organization pro-vides adequate sup-port, with room for improvement | The organization provides full and consistent sup-port, ensuring the moni-toring process is well-re-sourced and effective |

**Table A3.** Ability to Learn (OHS in Manufacturing Context).

| | | | Possible answers | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ID | Item | Question | 1 | 2 | 3 | 4 | 5 |
| L1 | Selection criteria | What types of OHS incidents or near-misses are investigated (e.g., only severe accidents or also near-misses)? What criteria (e.g., injury severity, frequency) are used for selecting cases for detailed investigation? | No clear criteria for selecting events to investigate | Events are rarely selected based on clear criteria | Events are sometimes selected based on frequency or severity, but inconsistently | Most events are selected for investigation based on relevant criteria | All OHS events are consistently selected based on well-defined and appropriate criteria |
| L2 | Learning basis | Does the organization learn from positive events (e.g., instances where employees avoided accidents through proper safety behavior) as well as negative ones (e.g., actual accidents)? | The system learns only from failures, never from successes | The system rarely learns from successes, primarily focusing on failures | The system sometimes learns from successes, but mostly from failures | The system often learns from both successes and failures | The system fully integrates learning from both successes and failures in a balanced manner |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| L3 | Classification | How are OHS incidents described, classified, and categorized (e.g., based on type of injury, location, equipment involved)? | There is no system for classifying or categorizing incidents | Incidents are rarely classified or categorized in a meaningful way | Incidents are classified inconsistently or with limited structure | Most incidents are systematically classified and categorized | All incidents are comprehensively described, classified, and categorized in a standardized manner |
| L4 | Formalization | Are there formal procedures for collecting, analyzing, and learning from OHS incidents (e.g., root cause analysis, safety debriefs)? | There are no formal procedures in place | Procedures are informal and inconsistently applied | Some formal procedures exist, but they are not always followed | Procedures are mostly formalized and followed | Fully formalized procedures are in place and consistently applied for all aspects of data collection and learning |
| L5 | Training | Is there formal training on how to collect and analyze safety data? Are employees and supervisors trained on how to apply lessons learned from past incidents? | There is no training provided | Training is minimal and insufficient | Some training is provided, but it needs improvement | Adequate training is provided, but with occasional gaps | Comprehensive and regular training is provided to all relevant staff, ensuring effective learning from incidents |
| L6 | Learning style | Is learning from safety incidents continuous (e.g., integrated into regular safety meetings) or only triggered by specific events (e.g., after a major accident)? | Learning only occurs after major incidents or accidents | Learning is mostly event-driven, with little focus on continuous improvement | Learning is a mix of event-driven and continuous, but mostly reactive | Learning is often continuous, with some event-driven elements | Learning is fully continuous and integrated into everyday practices |
| L7 | Resources | Are there adequate resources (e.g., dedicated investigation teams, software tools for incident tracking) allocated for investigating safety incidents and facilitating learning? | There are no dedicated resources for investigation and learning | Resources are very limited and inadequate | Resources are available but are often stretched too thin | Adequate resources are available, but occasional limitations occur | Resources are fully sufficient and consistently available for thorough investigation and learning |
| L8 | Delay | What is the delay between an incident and learning from it (e.g., how quickly are root causes identified, and corrective actions implemented)? How are the outcomes communicated internally (e.g., through safety bulletins) and externally (e.g., regulatory bodies)? | There is a significant delay that prevents timely learning | Delays are often too long to ensure effective learning | There are occasional delays, but learning is generally timely | The delay is minimal, and learning occurs in a timely manner | Learning happens promptly after every incident, with immediate analysis and corrective actions |

| ID | Item | Question | | | | | |
|----|------|----------|---|---|---|---|---|
| L9 | Learning target | At what level does learning take place? Does it focus on individual workers, specific teams, or at the organizational level? | Learning rarely takes effect at any level | Learning mostly affects individuals, with little collective or organizational impact | Learning has some impact on teams or individuals, but limited organizational changes | Learning impacts both individuals and teams, with some organizational changes | Learning is systematically applied at the individual, team, and organizational levels |
| L10 | Implementation | How are lessons learned from OHS incidents implemented? Are they translated into revised safety procedures, updated training programs, or new workplace designs? | Lessons learned are rarely, if ever, implemented | Implementation of lessons learned is sporadic and ineffective | Lessons are sometimes implemented, but often inconsistently | Lessons learned are implemented effectively most of the time | Lessons learned are always implemented thoroughly and consistently across the organization, with clear impact on procedures and practices |

**Table A4.** Ability to Anticipate (OHS in Manufacturing Context).

| | | | Possible answers | | | | |
|----|------|----------|---|---|---|---|---|
| ID | Item | Question | 1 | 2 | 3 | 4 | 5 |
| A1 | Expertise | What kind of expertise is used to anticipate future OHS risks (e.g., in-house safety experts, external consultants)? Is expertise drawn from cross-industry practices, ergonomics, or specific manufacturing knowledge? | No expertise is used to anticipate future risks | Minimal expertise is available, and it is often insufficient | Expertise is available but is inconsistently applied or limited in scope | Expertise is mostly sufficient and appropriately applied | Expertise is fully sufficient, comprehensive, and consistently applied |
| A2 | Frequency | How often are future OHS threats (e.g., potential risks from new machinery, changes in production lines) assessed? | Future threats and opportunities are never assessed | Assessments are very rare and typically ad-hoc | Assessments are conducted occasionally, but not regularly | Assessments are conducted regularly, with some gaps | Future threats and opportunities are assessed frequently and systematically |
| A3 | Communication | How are forecasts of future OHS risks or opportunities (e.g., new safety technologies) communicated within the organization? | Expectations are never communicated | Communication of future risks is very poor and rarely reaches relevant parties | Communication occurs but is often unclear or inconsistent | Communication is mostly clear and reaches relevant stakeholders | Communication is always clear, consistent, and reaches all relevant stakeholders in a timely manner |
| A4 | Strategy | Does the organization have a clear safety vision or strategy for addressing future OHS | There is no strategy for future OHS risks | There is a very vague or informal strategy | There is a strategy, but it is incomplete or not well formulated | The strategy is well-formulated and mostly clear, with minor gaps | The organization has a clearly formulated and comprehensive |

|  |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|
|  |  | risks (e.g., zero-accident policies, advanced automation for hazardous tasks)? |  | with minimal planning |  |  | strategy for addressing future OHS risks |
| A5 | Model | Is the safety strategy or model of future OHS risks explicitly defined (e.g., through formal risk assessments), and is it qualitative (e.g., expert judgment) or quantitative (e.g., risk probability models)? | There is no explicit model or assumptions about future risks | Models are mostly implicit and qualitative, with little structure | Models are somewhat explicit but lack thoroughness, primarily qualitative | Models are explicit and reasonably detailed, often blending qualitative and quantitative elements | Models are fully explicit, highly detailed, and include both qualitative and quantitative aspects |
| A6 | Time horizon | How far ahead does the system look in terms of OHS risk anticipation? Does this vary between business goals (e.g., production targets) and safety priorities? | No forward-looking assessments are made | The time horizon is very short (e.g., only a few months) | The time horizon is moderate (e.g., 1–2 years), but may not account for long-term risks | The time horizon is substantial (e.g., 3–5 years), with good planning for future risks | The organization looks far ahead (5+ years), with comprehensive risk assessments for the long-term future |
| A7 | Acceptability of risks | What criteria are used to determine which OHS risks are acceptable (e.g., minor injuries) and which are not (e.g., fatalities, permanent disabilities)? | No distinction is made between acceptable and unacceptable risks | The definition of acceptable risks is vague or inconsistently applied | Some criteria exist for defining acceptable risks, but they are incomplete | Acceptable and unacceptable risks are mostly well defined and understood | There are clear, well-documented criteria for distinguishing acceptable and unacceptable risks, consistently applied across the organization |
| A8 | Aetiology | How does the organization define the nature of future OHS threats? Are they seen as inherent to the manufacturing process (e.g., machinery risks) or as preventable through better safety controls? | The nature of future threats is not anticipated at all | Future threats and opportunities are anticipated in a very limited and reactive way | Some effort is made to anticipate future threats, but the approach is incomplete | The organization mostly anticipates the nature of future threats and opportunities in a proactive way | The organization systematically anticipates and plans for future OHS threats and opportunities with a proactive approach |
| A9 | Culture | Is there a strong culture of risk awareness in the organization? Do employees actively contribute to risk assessments and proactive safety measures? | Risk awareness is not part of the culture at all | Risk awareness is minimal and rarely encouraged within the organization | Risk awareness is present but not consistently reinforced across the organization | Risk awareness is a part of the culture and generally reinforced in most areas | Risk awareness is fully embedded in the organizational culture and actively promoted at all levels |

## Appendix B

*OHS RAG Questionnaire Original Socratic Dialogue*

Human: I have a questionnaire from Hollnagel about resilience abilities. Could you help adapt it for Occupational Health and Safety (OHS) in the manufacturing industry?

LLM: Sure! We can contextualize the questions across four resilience abilities—Respond, Monitor, Learn, and Anticipate—specific to OHS in manufacturing. This will make it more applicable to that environment.

Human: Great! Let's start with the ability to respond. Could you adapt the questions for this ability and suggest Likert-scale responses?

LLM: Here's the adapted set of questions for "Ability to Respond" in the manufacturing context, with a 5-point Likert scale. Each question assesses an organization's readiness in OHS event response, from event lists to stop criteria.

Human: Perfect! Let's proceed to the ability to monitor and create the Likert scales for that as well.

LLM: Here's the contextualized set of questions and Likert responses for "Ability to Monitor," assessing how effectively an organization monitors OHS indicators, measurement frequency, and analysis.

Human: Excellent. Now, let's move on to the ability to learn with corresponding Likert scales.

LLM: Here's the tailored list for "Ability to Learn," with questions on data collection, incident classification, formalization, and implementation of lessons learned.

Human: This looks good. Now let's complete the questionnaire with the ability to anticipate questions and the Likert scale options.

LLM: Here's the adapted set of questions and responses for "Ability to Anticipate," focusing on expertise, risk modeling, communication, and risk culture.

Human: This is exactly what I needed! Now, I'd like to generate some synthetic data based on a Resilience Analysis Grid (RAG) assessment using the Likert scale. Could you provide an example?

LLM: Certainly. Here's synthetic data based on the RAG scoring system. Each question has a corresponding Likert score to simulate an organization's OHS assessment.

Human: Perfect. Now let's restructure the steps we've followed as a general LLM-based process.

LLM: Here's the structured process:

1. Understanding the Task—Analyze and contextualize the initial framework.
2. Contextualization of Framework—Adapt questions for OHS in manufacturing.
3. Creation of Measurement Tools—Develop Likert scales for assessment.
4. Synthetic Data Generation—Generate data using RAG scoring.
5. Feedback Iteration and Refinement—Incorporate user feedback to improve the process.

## References

1. Koc, K.; Gurgun, A.P. Scenario-Based Automated Data Preprocessing to Predict Severity of Construction Accidents. *Autom. Constr.* **2022**, *140*, 104351. https://doi.org/10.1016/j.autcon.2022.104351.
2. Tsalidis, G.A. Introducing the Occupational Health and Safety Potential Midpoint Impact Indicator in Social Life Cycle Assessment. *Sustainability* **2024**, *16*, 3844. https://doi.org/10.3390/su16093844.
3. Stepanovic, S.; Naous, D.; Mettler, T. A Privacy Impact Assessment Method for Organizations Implementing IoT for Occupational Health and Safety. 2023. Available online: https://aisel.aisnet.org/icis2023/ishealthcare/ishealthcare/14/ (accessed on 11 November 2024).
4. Bond, T.G.; Fox, C.M. *Applying the Rasch Model*; Psychology Press: London, UK, 2013; ISBN 978-1-135-60265-9.
5. Adriaensen, A.; Costantino, F.; Di Gravio, G.; Patriarca, R. Teaming with Industrial Cobots: A Socio-Technical Perspective on Safety Analysis. *Hum. Factors Ergon. Manuf. Serv. Ind.* **2022**, *32*, 173–198. https://doi.org/10.1002/hfm.20939.
6. Rodríguez, M.; Lawson, E.; Butler, D. A Study of the Resilience Analysis Grid Method and Its Applicability to the Water Sector in England and Wales. *Water Environ. J.* **2020**, *34*, 623–633. https://doi.org/10.1111/wej.12539.

7.   Duros, S.E. Supporting Resilience in Distributed Work Systems through Modeling Adaptive Capacity. 2022. Available online: https://rave.ohiolink.edu (accessed on 11 November 2024).

8.   Brown, T.B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language Models Are Few-Shot Learners. *arXiv* **2020**, arXiv: 2005.14165

9.   Hellwig, N.C.; Fehle, J.; Wolff, C. Exploring Large Language Models for the Generation of Synthetic Training Samples for Aspect-Based Sentiment Analysis in Low Resource Settings. *Expert Syst. Appl.* **2025**, *261*, 125514. https://doi.org/10.1016/j.eswa.2024.125514.

10.  Zhao, F.; Yu, F.; Trull, T.; Shang, Y. A New Method Using LLMs for Keypoints Generation in Qualitative Data Analysis. In Proceedings of the 2023 IEEE Conference on Artificial Intelligence (CAI), Santa Clara, CA, USA, 5–6 June 2023; pp. 333–334.

11.  Kalyniukova, A.; Várfalvyová, A.; Andruch, V. Applicability of ChatGPT 3.5 in the Development of New Analytical Procedures. *Microchem. J.* **2024**, *203*, 110787. https://doi.org/10.1016/j.microc.2024.110787.

12.  Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Ichter, B.; Xia, F.; Chi, E.; Le, Q.V.; Zhou, D. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 24824–24837.

13.  Shah, I.A.; Mishra, S. Artificial Intelligence in Advancing Occupational Health and Safety: An Encapsulation of Developments. *J. Occup. Health* **2024**, *66*, uiad017. https://doi.org/10.1093/joccuh/uiad017.

14.  Tang, K.H.D. Artificial Intelligence in Occupational Health and Safety Risk Management of Construction, Mining, and Oil and Gas Sectors: Advances and Prospects. *J. Eng. Res. Rep.* **2024**, *26*, 241–253. https://doi.org/10.9734/jerr/2024/v26i61177.

15.  Linkon, A.A.; Shaima, M.; Sarker, M.S.U.; Nabi, N.; Rana, M.N.U.; Ghosh, S.K.; Rahman, M.A.; Esa, H.; Chowdhury, F.R. Advancements and Applications of Generative Artificial Intelligence and Large Language Models on Business Management: A Comprehensive Review. *J. Comput. Sci. Technol. Stud.* **2024**, *6*, 225–232. https://doi.org/10.32996/jcsts.2024.6.1.26.

16.  Taiwo, R.; Bello, I.T.; Abdulai, S.F.; Yussif, A.-M.; Salami, B.A.; Saka, A.; Zayed, T. Generative AI in the Construction Industry: A State-of-the-Art Analysis. *arXiv* **2024**, arXiv: 2402.09939.

17.  Lu, Y.; Shen, M.; Wang, H.; Wang, X.; van Rechem, C.; Fu, T.; Wei, W. Machine Learning for Synthetic Data Generation: A Review. *arXiv* **2023**, arXiv: 2302.04062.

18.  Maheronnaghsh, S.; Zolfagharnasab, H.; Gorgich, M.; Duarte, J. Machine Learning in Occupational Safety and Health: Protocol for a Systematic Review: (Protocol). *Int. J. Occup. Environ. Saf.* **2021**, *5*, 32–38. https://doi.org/10.24840/2184-0954_005.001_0004.

19.  Guo, X.; Chen, Y. Generative AI for Synthetic Data Generation: Methods, Challenges and the Future. *arXiv* **2024**, arXiv: 2403.04190.

20.  Quadir, A.; Lewis, C.; Rau, R.-J. Generation of Pseudo-Synthetic Seismograms from Gamma-Ray Well Logs of Highly Radioactive Formations. *Pure Appl. Geophys.* **2019**, *176*, 1579–1599.

21.  Vaccari, C.; Chadwick, A. Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Soc. Media+ Soc.* **2020**, *6*, 2056305120903408.

22.  Loiseau, T.; Vu, T.-H.; Chen, M.; Pérez, P.; Cord, M. Reliability in Semantic Segmentation: Can We Use Synthetic Data? In *Computer Vision—ECCV 2024*; Leonardis, A., Ricci, E., Roth, S., Russakovsky, O., Sattler, T., Varol, G., Eds.; Lecture Notes in Computer Science; Springer Nature: Cham, Switzerland, 2025; Volume 15081, pp. 442–459, ISBN 978-3-031-73336-9.

23.  Johannesson, P.; Perjons, E. *An Introduction to Design Science*; Springer International Publishing: Cham, Switzerland, 2021; ISBN 978-3-030-78131-6.

24.  Galli, C.; Cusano, C.; Meleti, M.; Donos, N.; Calciolari, E. Topic Modeling for Faster Literature Screening Using Transformer-Based Embeddings. *Metrics* **2024**, *1*, 2. https://doi.org/10.3390/metrics1010002.

25.  Sturm, B.; Sunyaev, A. Design Principles for Systematic Search Systems: A Holistic Synthesis of a Rigorous Multi-Cycle Design Science Research Journey. *Bus. Inf. Syst. Eng.* **2019**, *61*, 91–111. https://doi.org/10.1007/s12599-018-0569-6.

26.  Pishgar, M.; Issa, S.F.; Sietsema, M.; Pratap, P.; Darabi, H. REDECA: A Novel Framework to Review Artificial Intelligence and Its Applications in Occupational Safety and Health. *Int. J. Environ. Res. Public Health* **2021**, *18*, 6705. https://doi.org/10.3390/ijerph18136705.

27.  Akoka, J.; Comyn-Wattiau, I.; Prat, N.; Storey, V.C. Knowledge Contributions in Design Science Research: Paths of Knowledge Types. *Decis. Support Syst.* **2023**, *166*, 113898. https://doi.org/10.1016/j.dss.2022.113898.

28.  Falegnami, A.; Romano, E.; Tomassi, A. The Emergence of the GreenSCENT Competence Framework: A Constructivist Approach: The GreenSCENT Theory. In *The European Green Deal in Education*; Routledg: London, UK, 2024; ISBN 978-1-00-349259-7.

29.  Tomassi, A.; Falegnami, A.; Romano, E. Mapping Automatic Social Media Information Disorder. The Role of Bots and AI in Spreading Misleading Information in Society. *PLoS ONE* **2024**, *19*, e0303183. https://doi.org/10.1371/journal.pone.0303183.

30.  Chen, L.; Cai, Z.; Jiang, Z.; Sun, L.; Childs, P.; Zuo, H. A Knowledge Graph-Based Bio-Inspired Design Approach for Knowledge Retrieval and Reasoning. *J. Eng. Des.* **2024**, Volume 35, 1–31. https://doi.org/10.1080/09544828.2024.2311065.

31.  Falegnami, A.; Tomassi, A.; Corbelli, G.; Romano, E. Managing Complexity in Socio-Technical Systems by Mimicking Emergent Simplicities in Nature: A Brief Communication. *Biomimetics* **2024**, *9*, 322. https://doi.org/10.3390/biomimetics9060322.

32.  Lewis, P.; Perez, E.; Piktus, A.; Petroni, F.; Karpukhin, V.; Goyal, N.; Küttler, H.; Lewis, M.; Yih, W.; Rocktäschel, T. Retrieval-Augmented Generation for Knowledge-Intensive Nlp Tasks. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 9459–9474.

33.  Ampazis, N. Improving RAG Quality for Large Language Models with Topic-Enhanced Reranking. In *Artificial Intelligence Applications and Innovations*; Maglogiannis, I., Iliadis, L., Macintyre, J., Avlonitis, M., Papaleonidas, A., Eds.; IFIP Advances in

Information and Communication Technology; Springer Nature: Cham, Switzerland, 2024; Volume 712, pp. 74–87, ISBN 978-3-031-63214-3.

34. Exploring TITAN's Approach to Integrating Socratic Thinking and AI in Chatbot Dialogue. Available online: https://www.titanthinking.eu/post/exploring-titan-s-approach-to-integrating-socratic-thinking-and-ai-in-chatbot-dialogue (accessed on 13 November 2024).

35. Flynn, B.B.; Sakakibara, S.; Schroeder, R.G.; Bates, K.A.; Flynn, E.J. Empirical Research Methods in Operations Management. *J Ops. Manag.* **1990**, *9*, 250–284. https://doi.org/10.1016/0272-6963(90)90098-X.

36. Hollnagel, E. *Safety-II in Practice: Developing the Resilience Potentials*; Routledge: London, UK, 2017; ISBN 978-1-351-78076-6.

37. Hollnagel, E. Epilogue: RAG—The Resilience Analysis Grid. In *Resilience Engineering in Practice*; CRC Press: Boca Raton, FL, USA, 2019.

38. Winzar, H. LikertMakeR: Synthesise and Correlate Rating-Scale Data with predefined first & second moments; 2022, 0.4.0., Available online: https://github.com/WinzarH/LikertMakeR (accessed on 11 November 2024).