

Exploring the use of Large Language Models in Risk Analysis: Opportunities and Challenges

Zdenko Šimić

Paul Scherrer Institute

Forschungsstrasse 111, 5232 Villigen PSI, Schweiz

zdenko.simic@psi.ch

ABSTRACT

The rapid advancement of Large Language Models (LLMs, new generation of generative artificial intelligence models) has opened novel possibilities for their application in various fields, including risk analysis (RA). This paper explores the potential of LLMs to enhance the consistency, efficiency, and comprehensiveness of RA processes. The paper begins by outlining the key elements of a RA framework and discussing the limitations of traditional AI methods in this field. It then examines the unique capabilities of LLMs, such as their ability to quickly process vast amounts of textual data, identify patterns, and generate insights, potentially suitable for various RA tasks.

The paper presents findings from a targeted literature review, presenting diverse applications of LLMs in RA across different domains and areas, including information extraction, hazard identification, scenario generation and semi-automatic application of different RA methods. Additionally, a simple assessment of LLM capabilities through illustrative queries related to reliability analysis is presented, demonstrating their foundational knowledge of RA concepts and questions.

The paper discusses the potential benefits of LLMs in RA, such as improved efficiency, scalability, and the ability to handle complex, context-dependent scenarios. It also identifies the challenges and limitations, including data quality and bias, reliable contextual interpretation, explainability, and the need for validation and verification.

The paper concludes by highlighting the potential benefits of using LLMs in RA, emphasizing the importance of further research and development to better realize their potential while addressing ethical considerations and ensuring responsible integration into RA practices.

Keywords: Risk Analysis, Large Language Model (LLM), Artificial Intelligence, Augmented analysis

1 INTRODUCTION

Risk analysis (RA) plays a crucial role in ensuring the safety and reliability of critical systems across diverse domains, from nuclear power plants to aviation and healthcare. A comprehensive RA framework typically encompasses several key elements: hazard identification, consequence characterization, uncertainty quantification, modelling, knowledge management, and the development of mitigation strategies. However, traditional RA approaches often face challenges due to the subjective nature of expert judgment, modelling challenges, and the complexity of synthesizing vast amounts of information.

Artificial Intelligence (AI) has been increasingly applied to enhance various aspects of risk analysis and mitigation. Traditional AI methods have focused on ([1], [2], [3]):

1. Pattern Recognition and Anomaly Detection: Identifying potential risks or hazards in large datasets.
2. Predictive Modelling: Forecasting future events based on historical data and current conditions.
3. Decision Support Systems: Assisting analysts in evaluating complex scenarios and identifying optimal courses of action.

However, these methods often require extensive data pre-processing and lack the flexibility to handle realistic, context-dependent information [4].

Recent advancements in AI, particularly in Large Language Models (LLMs), present new potential for enhancing risk analysis. LLMs, such as GPT-4, Claude, Gemini, and Llama are transformer-based neural networks trained on vast amounts of textual data and have hundreds of billions of parameters [5]. These models have demonstrated significant capabilities in natural language processing, reasoning, and knowledge synthesis ([6], [7]).

Key characteristics of LLMs that make them potentially suitable for risk analysis include:

1. Ability to process, integrate, and interpret large volumes of diverse textual¹ data.
2. Capacity to identify patterns and generate new insights.
3. Flexibility in handling complex, context-dependent information.

While LLMs show promise, it's important to note their known limitations, such as potential and unknown biases in training data, limited domain-specific knowledge, and the 'black box' nature of their working [8]. Addressing these challenges is crucial for their effective application in risk analysis ([9], [10]).

This paper explores the potential applications of LLMs in risk analysis, investigating their capacity to enhance consistency, efficiency, and comprehensiveness in RA processes. The aim is to examine how LLMs can be leveraged to support RA methodology application, validation, and advancement, ultimately contributing to enhanced safety and reliability in high-stakes environments such as nuclear power and aviation.

The rest of the paper is organized as follows: Section 2 describes the exploration approach; Section 3 presents results; Section 4 discusses findings; and Section 5 offers conclusions. The Appendix presents results from illustrative queries demonstrating LLMs fundamental knowledge in selected risk analysis questions.

2 EXPLORATION APPROACH

This preliminary exploration of the potential of Large Language Models (LLMs) use in Risk Analysis (RA) is based on a targeted literature review, and illustrative trial usage of LLMs. Preliminary findings are then formulated based on review and trial usage.

2.1 Scoping Literature Review

A limited scoping literature search was conducted using Google Scholar with the following boundaries:

- Keywords: "large language models," "risk analysis," "human reliability analysis," "artificial intelligence," "machine learning".
- Time frame: Last five years.

¹ New generative AI models are most developed as LLMs dealing with textual data. However, increasingly these models are developing as multimodal with capacity to process speech and visual data as well. This paper is focused on LLMs because they are most developed and used. Multimodality will only improve potential benefit in the future.

- Selection criteria: Relevance to RA applications, citation count, and journal impact factor.

This review aimed to capture a representative sample of most recent works rather than an exhaustive analysis, focusing on LLMs risk analysis applications.

2.2 Illustrative Trial Usage

Knowledge and capabilities of LLMs significantly depend on data used for training. More complete LLMs training data in certain domain are important for easy and useful application. Preliminary testing of the base knowledge and capabilities of LLMs in RA related topics was done with two zero-shot prompt queries including two state-of-the-art models:

- Models tested: GPT-4 (OpenAI) and Gemini Advanced (Google).
- Approach: Zero-shot (single query without details, examples, or additional documentation) prompting on human reliability analysis topics.
- Evaluation: Qualitative assessment based on accuracy, relevance, coherence, and alignment with expert expectations.

This simple testing serves to illustrate LLMs' inherent knowledge and inference on RA concepts, providing a foundation for potential future, more comprehensive explorations.

2.3 Preliminary Findings

Based on insights from the literature review and trial usage, initial ideas were formulated for potential LLM applications in RA. These proposals consider known challenges in current RA methods and suggest areas where LLMs might offer support, such as:

- Semi-automated analysis of operational experience reports.
- Augmented application of RA methods.
- Enhanced analysis coverage.
- Advanced applications like LLM-generated scenarios.

These preliminary proposals are intended to stimulate discussion and guide future research directions.

This exploratory approach aims to provide initial insights into the potential of LLMs in enhancing RA practices, laying the groundwork for more comprehensive studies and practical applications in the future.

3 RESULTS

This section presents the findings from our investigation into the potential use of Large Language Models (LLMs) in risk analysis (RA). It includes (1) results from a limited scoping literature review of relevant applications, and (2) an exploratory illustration of LLMs capabilities through two risk analysis related queries.

3.1 Findings from Limited Scoping Literature Review

A Google Scholar search produced a substantial number of references, even with a five-year publication restriction. However, the majority are related to traditional AI applications in risk analysis. By refining the search to focus on "large language models" and manually filtering the results, the number of relevant references was significantly reduced. The search also identified numerous studies on the risks associated with the use of large language models, which were not considered in this review. Ultimately, nine references were selected for detailed analysis, with seven focused on practical applications. The application domains, specific problems addressed, and results, including identified limitations, are described below. The remaining two references are discussed later in the discussion section.

Summarization of diverse information sources [11]: Study applied LLMs for summarizing diverse sources relevant to decisions, including risk related, in the medical domain, comparing the quality and completeness of these summaries with those produced by experts. The study encompassed various clinical documents, including radiology reports, patient queries, progress notes, and doctor-patient dialogues. Both general LLMs pretrained on diverse text corpora and adapted LLMs fine-tuned with medical data were evaluated. The adapted LLMs achieved or exceeded expert-level quality, with larger LLMs generally performing better. The use of in-context examples (ICE) within prompts enhanced the results. Lightweight fine-tuning using quantized low-rank adaptation (QLoRA) provided marginal improvements, primarily for smaller LLMs. Notable limitations included the potential for hallucinations, the necessity of human evaluation for clinical readiness, and challenges in ensuring the accuracy and relevance of generated summaries.

Risk analysis for Autonomous Underwater Vehicle (AUV) [12]: study investigated the use of LLMs for risk estimation regarding the survival of AUV during specified missions. Various text-embedding techniques, including BERT (Bidirectional Encoder Representations from Transformers), were compared. The study highlighted the benefits of transfer learning, where pre-training on large corpora enhances performance on a limited fault description dataset. Local Interpretable Model-agnostic Explanations (LIME) were employed to analyse the influence of specific words on risk estimations, effectively emulating expert judgments. The models demonstrated accuracy comparable to expert assessments. However, training the LLMs required data elicited from experts using formal methods. BERT's bidirectional architecture contrasts with today more commonly used autoregressive architectures in LLMs.

Support in Failure Mode and Effects Analysis (FMEA) [13]: An LLM was utilized to support the FMEA process by automating the identification of failure modes. The developed framework employed LLMs for data collection, extraction, and knowledge management, specifically fine-tuning a GPT-3 model with a large dataset of negative customer reviews. Integrating LLMs into FMEA processes showed potential in enhancing efficiency and accuracy. LLMs provided valuable support in data extraction and analysis, indicating their capability to augment traditional FMEA approach. However, the application requires specialized tools and robust data management practices, alongside challenges related to protecting sensitive data and ensuring result accuracy and relevance.

Augmenting System-Theoretic Process Analysis (STPA) [14]: LLMs have also been explored to assist in STPA, particularly in identifying potential unsafe control actions (UCAs) and causal scenario. The study evaluated different interaction schemes between human experts and ChatGPT-4, considering varied complexity of system control loop descriptions and the role of prompt engineering. Two case studies involving automatic emergency braking and electricity demand-side management systems were analysed. Results indicated that a recurring duplex collaboration approach, which involves continuous bi-directional interactions with human feedback, yielded the most comprehensive set of UCAs. Interestingly, variations in system description complexity and prompt engineering had slight impact, with even the simplest prompting approach outperforming expert analyses. Despite these promising results, the study emphasized the importance of prompt engineering, human oversight, and a deep understanding of the system under analysis to ensure the accuracy and usefulness of AI-generated insights. Challenges included issues with LLM trustworthiness and the lack of standardization.

Augmenting Functional Resonance Analysis Method (FRAM) [15]: The use of LLMs in FRAM analysis was investigated for safety-critical scenarios in healthcare and aviation. ChatGPT-3.5 and Bard (now Gemini) were employed to generate initial models through multiple prompts. LLMs provided reasonable initial models and valuable insights, potentially enhancing the comprehensiveness and accuracy of assessments. However, the study stressed the critical role of human expertise in validating LLM outputs and developing effective prompting strategies, noting the necessity of responsible use given the limitations of LLMs.

Hazard Extraction for Risk Trend Analysis [16]: A framework for risk trend analysis incorporating LLMs for hazard extraction. BERTopic, a modelling technique leveraging pre-trained word embeddings from LLMs, was used to interpret hazards based on input descriptions of problems, events, and other information. These identified hazards were then analysed using FMEA and risk matrices, within the framework applied to improve wildfire response. The semi-automatic process effectively identified hazards and performed risk analysis, suggesting LLMs' potential to support data-driven risk analysis in wildfire response scenarios. Despite its benefits, manual interpretation and labelling were necessary to ensure the quality of hazard extraction.

Extraction of Emergency Plan Information [17]: A study explored the use of smaller language models (~10 billion parameter) for extracting information from power system emergency plans to assist in emergency response. The method successfully extracted key information, proving to be more accurate than traditional machine learning approaches without requiring sample data training. The study highlighted the difficulties associated with knowledge retrieval, and content correlation matching in traditional unstructured emergency plans. However, it also noted the need for iterative prompt learning to ensure the accuracy and relevance of extracted information.

These studies present the diverse applications and approaches for utilizing LLMs in risk analysis, providing a foundation for discussions on the potential opportunities and challenges in other use cases.

3.2 Illustrative LLMs Queries Related to Risk Analysis

To further explore LLM capabilities in risk analysis, we conducted an exploratory assessment using two state-of-the-art models, GPT-4 and Gemini Advanced. This approach aligns with [18] (for future transportation scenarios query) and [19] (examination of hazards associated with new highway operations) as part of broader studies.

Two zero-shot prompts related to human reliability analysis (HRA) were formulated to evaluate the LLMs' baseline understanding of specific risk analysis concepts, particularly focusing on Performance Shaping Factors (PSFs), which are crucial in HRA for influencing human performance and error likelihood. Full responses are provided in the Appendix.

The first prompt explored dependencies between four specified PSFs:

"In Human Reliability Analysis, we consider Performance Shaping Factors (PSFs). Please describe the dependencies between the following four PSFs and how to address them in an HRA: Available time, Stress/Stressors, Complexity, Experience/Training."

Both LLMs provided coherent and complete responses, demonstrating an understanding of the interplay between these PSFs. Gemini Advanced offered a comprehensive overview, including definitions and general considerations, while ChatGPT-4 provided a more focused and concise analysis.

The second prompt addressed the definition and rating of PSFs:

"In Human Reliability Analysis, Performance Shaping Factors are often rated on a scale from -3 to +3. Please suggest definitions for each of the seven rating levels (-3 to +3) and provide an example for each."

Again, both LLMs displayed a sound description of the topic. Gemini Advanced also emphasized the importance of context, PSF relationships, and the dynamic nature of PSFs, while ChatGPT-4 response was clearer with more concise definitions.

These illustrative zero-shot prompts revealed that both models possess a foundational understanding of HRA concepts, even without explicit task-specific information or training. The responses highlighted their ability to process information, identify relationships, and generate insights that align with expert judgement.

4 DISCUSSION

The findings from limited scoping literature review and illustrative trial assessment of LLM's capabilities show some potential for integrating these models into various aspects of risk analysis. This section synthesizes findings as (1) strengths and potential applications of LLMs in risk analysis, (2) noticed challenges and limitations, and (3) some possible related future directions and research needs.

4.1 Strengths and Potential Applications of LLMs in Risk Analysis

The reviewed studies show that LLMs can be applied to diverse risk analysis tasks across multiple domains. Key areas where LLMs show some promise include:

Information Summarization and Extraction: LLMs are capable at synthesizing diverse information sources, as demonstrated in the medical domain [11]. This capability could be valuable in risk analysis, where analysts often need to process a lot of heterogeneous data.

Semi-Automated Hazard Identification and Risk analysis: Studies including UAV [12], FMEA [13], STPA [14], FRAM [15], and highlight LLMs' potential for semi-automate risk analysis and the identification of failure modes, hazards, or unsafe control actions. This could enhance the efficiency and comprehensiveness of risk analysis processes.

Knowledge Management and Retrieval: LLMs show potential in extracting relevant current information, as seen in the study on wildfire fire risk [16] and emergency plan information extraction [17]. This capability could streamline the process of accessing and utilizing historical risk data and expert knowledge in real-time scenarios.

Enhanced Risk Analysis: Illustrative trial usage in this work showed that largest foundational LLMs possess a knowledge of RA concepts and capability to reflect on specific questions. This suggests potential for LLMs to support or augment expert analysis in different RA tasks.

Scenario Generation: The potential of AI-assisted scenario generation for strategic planning [18] suggests that LLMs could be valuable in developing risk scenarios for new systems or emerging threats.

The strengths of LLMs in these applications is from their ability to process and analyze large volumes of unstructured data, identify patterns and relationships, and generate insights that align with expert knowledge [19]. This can lead to improved efficiency, scalability, and potentially more comprehensive risk analysis.

4.2 Challenges and Limitations of LLMs Important for use in Risk Analysis

Despite their potential, several challenges and limitations need to be addressed for the effective integration of LLMs in risk analysis:

Data Quality and Bias: LLMs are susceptible to potential and unknown biases present in their training data, which can lead to inaccurate or misleading results. Ensuring the quality and representativeness of training data is important for risk analysis.

Contextual Understanding: LLMs may struggle to capture the context of specific risk scenarios fully and properly, potentially leading to irrelevant, incomplete, or even wrong outputs. This limitation underscores the continued importance of human expertise in guiding and interpreting LLM results.

Explainability and Transparency: The "black box" nature of LLMs poses challenges for explainability and transparency in risk analysis. Developing techniques to enhance the interpretability of LLM outputs is crucial for building trust and ensuring regulatory compliance.

Validation and Verification: Establishing robust methods for validating and verifying LLM outputs in risk analysis contexts is important. This includes developing strategies to mitigate the risk of "hallucinations" or factually incorrect information generated by LLMs.

Uncertainty Quantification: The ability of LLMs to adequately capture and present uncertainty, a critical aspect of risk analysis, is a concern.

4.3 Future Directions and Research Needs Related to use of LLMs in RA

To realize the potential of LLMs in risk analysis, several areas require further investigation:

Advanced Prompting, Fine-Tuning, and other techniques: Exploring sophisticated prompting techniques, domain-specific fine-tuning, and using other advanced techniques to enhance LLM performance on specialized risk analysis tasks.

Domain-Specific Applications: Exploring the use of LLMs across various risk analysis domains and methods to identify best practices and identify specific limitations.

Hybrid Approaches: Investigating ways to effectively combine LLM capabilities with traditional risk analysis methods and human expertise to achieve best results.

Ethical Considerations: Addressing ethical concerns related to the use of LLMs in risk analysis, including issues of bias, accountability, and potential misuse.

Standardization and Guidelines: Developing industry and regulatory standards and best practices for the successful integration of LLMs into risk analysis processes.

Uncertainty and Confidence: Developing methods for quantifying and communicating the uncertainty and justifications associated with LLM-generated risk analysis.

While LLMs show promise in enhancing various aspects of risk analysis, their effective integration requires consideration of their limitations and potential pitfalls. As research in this field progresses, a balanced approach that leverages the strengths of LLMs while maintaining human oversight will likely yield the most valuable outcomes for risk analysis practices.

5 CONCLUSION

This paper has explored the potential of Large Language Models (LLMs) in risk analysis through a targeted literature review and exploratory assessment of state-of-the-art LLM capabilities. The findings reveal that LLMs offer potential for applications across various aspects of risk analysis, including information summarization, hazard identification, knowledge management, scenario generation, and semi-automated modelling support. The strengths of LLMs lie in their ability to process vast amounts of unstructured data, identify patterns, and generate complementary insights aligned with expert knowledge, potentially leading to more complete and efficient risk analysis.

However, several challenges must be addressed for effective LLM integration in risk analysis. These include ensuring data quality, mitigating biases, assuring contextual understanding, including uncertainty quantification, enhancing explainability, and establishing validation processes. Further research is needed in domain-specific applications, hybrid approaches, ethical considerations, standardization, best use (prompting) techniques, and uncertainty evaluation.

In conclusion, while LLMs show promise for enhancing risk analysis processes, their effective integration requires a balanced approach that is using their strengths while maintaining human oversight and expertise. As the potential evolves, continued research and collaboration between risk analysis experts and AI researchers will be important in developing responsible and innovative applications of LLMs in risk analysis. This work provides a basis for future efforts, highlighting both opportunities and challenges associated with integrating LLMs into risk analysis practices. Moving forward, it will be essential to maintain a critical perspective, ensuring that LLM use enhances rather than compromises the rigor and quality of risk analysis.

REFERENCES

1. Zarei, E., F. Khan, and R. Abbassi, *How to account AI in human factor analysis of complex systems?* Process safety and environmental protection, 2023. **171**: p. 736-750.
2. Tamascelli, N., et al., *Artificial Intelligence for Safety and Reliability: A Descriptive, Bibliometric and Interpretative Review on Machine Learning*. Journal of Loss Prevention in the Process Industries, 2024: p. 105343.
3. Sethu, M., et al., *Application of artificial intelligence in detection and mitigation of human factor errors in NPPs: A review*. Nuclear Technology, 2023. **209**(3): p. 276-294.
4. Wang, C.J., D. Mandelli, and J. Cogliati, *Technical Language Processing of Nuclear Power Plants Equipment Reliability Data*. Energies, 2024. **17**(7): p. 1785.
5. Raiaan, M.A.K., et al., *A Review on Large Language Models: Architectures, Applications, Taxonomies, Open Issues and Challenges*. Ieee Access, 2024. **12**: p. 26839-26874.
6. Stribling, D., et al., *The model student: GPT-4 performance on graduate biomedical science exams*. Sci Rep, 2024. **14**(1): p. 5670.
7. Strachan, J.W.A., et al., *Testing theory of mind in large language models and humans*. Nat Hum Behav, 2024: p. 1-11.
8. Sun, L., et al., *Trustllm: Trustworthiness in large language models*. arXiv preprint arXiv:2401.05561, 2024.
9. Kojima, T., et al., *Large Language Models are Zero-Shot Reasoners*. Advances in Neural Information Processing Systems 35, Neurips 2022, 2022. **35**: p. 22199-22213.
10. Minaee, S., et al., *Large language models: A survey*. arXiv preprint arXiv:2402.06196, 2024.
11. Van Veen, D., et al., *Adapted large language models can outperform medical experts in clinical text summarization*. Nature medicine, 2024. **30**(4): p. 1134-1142.
12. Brito, M.P., M. Stevenson, and C. Bravo, *Subjective machines: Probabilistic risk assessment based on deep learning of soft information*. Risk Anal, 2023. **43**(3): p. 516-529.
13. El Hassani, I., et al., *Integrating large language models for improved failure mode and effects analysis: a framework and case study*. Proc. of the Design Society, 2024. **4**: p. 2019-2028.
14. Qi, Y., et al., *Safety analysis in the era of large language models: a case study of STPA using ChatGPT*. arXiv preprint arXiv:2304.01246, 2023.
15. Sujan, M., D. Slater, and E. Crumpton, *How Can Large Language Models Do a Fram Analysis?* Available at SSRN 4736249 (preprint), 2024.
16. Andrade, S.R. and H.S. Walsh, *Machine learning framework for Hazard Extraction and Analysis of Trends (HEAT) in wildfire response*. Safety Science, 2023. **167**: p. 106252.
17. Liu, Z., et al. *Method for extracting power emergency plan information based on LLM prompt learning*. in *International Conference on Frontiers of Applied Optics and Computer Engineering (AOCE 2024)*. 2024. SPIE.
18. Spaniol, M.J. and N.J. Rowland, *AI-assisted scenario generation for strategic planning*. Futures & Foresight Science, 2023. **5**(2): p. e148.
19. Stodle, K., et al., *Artificial intelligence for risk analysis-A risk characterization perspective on advances, opportunities, and limitations*. Risk Anal, 2024.

APPENDIX

1. Performance Shaping Factors and Dependencies

Prompt: *You are human reliability analysis expert. Describe dependencies between selected performance shaping factors (PSF) and suggest how to treat them in the analysis. Consider the following four PSFs: Available time, Stress/Stressors, Complexity, and Experience/Training.*

★ Gemini Advanced Response:

Absolutely! Let's dive into the dependencies between these Performance Shaping Factors (PSFs) in the context of Human Reliability Analysis (HRA).

Understanding the PSFs

- **Available Time:** The time available to complete a task relative to the time required.
- **Stress/Stressors:** The physiological or psychological demands placed on an individual, including workload, time pressure, environmental conditions, or emotional factors.
- **Complexity:** The inherent difficulty of a task, related to the number of steps, decision points, information to process, or unfamiliarity.
- **Experience/Training:** The level of knowledge, skills, and practice an individual has with the task, tools, or equipment involved.

Dependencies Between PSFs

1. **Available Time & Stress:** These have a strong, inverse relationship. Reduced time often increases stress, as individuals feel pressured to complete the task quickly. High stress can impair decision-making and lead to errors.
2. **Available Time & Complexity:** Less time available for a complex task can exacerbate difficulties. Rushing through complex procedures can lead to overlooking critical steps or misinterpreting information.
3. **Stress & Complexity:** Increased stress can make complex tasks feel even more overwhelming. Stress can hinder the ability to focus, analyze, and problem-solve, which are crucial for handling complexity.
4. **Experience/Training & Complexity:** Experience and training can mitigate the negative effects of complexity. Well-trained individuals are better equipped to handle complex situations due to their knowledge and developed skills.
5. **Experience/Training & Stress:** Experience can reduce stress in familiar situations. Knowing what to expect and having practiced procedures can build confidence and reduce anxiety.

Treating Dependencies in HRA

1. **Qualitative Assessment:**
 - Start by acknowledging and documenting the dependencies. Understand how changes in one PSF can affect others.
 - Use qualitative techniques like structured interviews, expert elicitation, or task analysis to gather insights on how these PSFs interact in specific scenarios.
2. **Quantitative Modeling:**
 - Incorporate dependencies into quantitative HRA models (e.g., THERP, SPAR-H).
 - This might involve adjusting performance probabilities based on the combined effects of multiple PSFs.
 - Some advanced models like Bayesian networks can explicitly represent dependencies and their influence on human error probability (HEP).
3. **Sensitivity Analysis:**
 - Conduct sensitivity analyses to see how changes in one PSF affect the overall HEP. This can help identify which dependencies are most critical.
 - For example, you might vary the available time in a scenario and observe how it impacts stress levels and, consequently, error rates.

Important Considerations

- **Individual Differences:** Recognize that the impact of PSFs can vary between individuals.
- **Context Matters:** The strength of dependencies can depend on the specific task, environment, and individual involved.

Let me know if you'd like a more detailed discussion on a specific aspect of this analysis or a particular HRA method.

ChatGPT 4o Response:

Performance Shaping Factors (PSFs) are critical in Human Reliability Analysis (HRA) as they influence human performance and the likelihood of human error. Here, we analyze the dependencies between four specific PSFs: Available Time, Stress/Stressors, Complexity, and Experience/Training.

1. Available Time - Available time refers to the duration a person has to perform a task. The relationship between available time and human performance is non-linear:

- **Dependency on Stress/Stressors:** Limited available time can increase stress levels, as the individual might perceive time pressure, which can negatively impact performance. Conversely, excessive time may reduce stress but could lead to complacency.
- **Dependency on Complexity:** For highly complex tasks, insufficient available time can drastically reduce performance due to the cognitive load and intricate decision-making required. However, for simpler tasks, the impact might be less severe.
- **Dependency on Experience/Training:** Experienced and well-trained individuals are typically better at managing time constraints and may perform adequately even with limited available time, whereas less experienced individuals might struggle.

2. Stress/Stressors - Stress can be physical, mental, or emotional and significantly affects human performance:

- **Dependency on Available Time:** As noted, limited time increases stress. Effective time management and realistic time allocation can mitigate this stress.
- **Dependency on Complexity:** Complex tasks inherently generate more stress due to higher cognitive demands and uncertainty. Simplifying tasks can reduce this stress.
- **Dependency on Experience/Training:** Experienced individuals often have better coping mechanisms and stress management techniques, reducing the impact of stressors on their performance.

3. Complexity - Complexity involves the intricacy of tasks, including the number of steps, interactions, and the cognitive load required:

- **Dependency on Available Time:** Complex tasks require more time to perform accurately. Limited time can lead to errors due to rushed or incomplete task execution.
- **Dependency on Stress/Stressors:** Higher complexity typically increases stress, as individuals must process more information and make more decisions under pressure.
- **Dependency on Experience/Training:** Training and experience can reduce the perceived complexity of a task, as experienced individuals can rely on heuristics and established patterns to navigate complex scenarios more effectively.

4. Experience/Training - Experience and training represent the knowledge and skills an individual has acquired over time:

- **Dependency on Available Time:** With adequate experience and training, individuals can perform tasks more efficiently, often needing less time to achieve the same results compared to less experienced individuals.
- **Dependency on Stress/Stressors:** Trained and experienced individuals can manage stress better, often having developed coping strategies and resilience.
- **Dependency on Complexity:** Training can simplify complex tasks by providing individuals with the tools and techniques to break down and manage the complexity effectively.
- **Treatment of Dependencies in Analysis**

To account for these dependencies in HRA, consider the following approaches:

1. **Quantitative Modeling:** Use models such as the Human Error Assessment and Reduction Technique (HEART) or the Technique for Human Error Rate Prediction (THERP) that can incorporate the influence of multiple PSFs. Adjust the error probabilities based on the interdependencies observed.
2. **Scenario-Based Analysis:** Develop realistic scenarios that capture the interactions between PSFs. For example, simulate high-stress environments with limited available time for complex tasks and assess performance outcomes for individuals with varying levels of experience and training.
3. **Sensitivity Analysis:** Conduct sensitivity analyses to determine how changes in one PSF impact the overall performance and error probabilities. This helps identify critical PSFs and their interdependencies.
4. **Training and Simulation:** Use training programs and simulations to mimic real-world conditions, including the interplay of PSFs. This helps in understanding how these factors interact and affect performance in practice.
5. **Data Collection and Empirical Studies:** Collect empirical data on human performance under varying conditions of PSFs. Use this data to refine models and validate assumptions about the dependencies between PSFs.

By understanding and appropriately modeling these dependencies, HRA can more accurately predict human error probabilities and enhance the safety and reliability of complex systems such as power plants.

2. Rating Performance Shaping Factors

Prompt: *You are human reliability analysis expert. Suggest definition for performance shaping factor (PSF) rating for each of seven levels (i.e.: -3 very negative, -2 moderately negative, -1 somewhat negative, 0 neutral, +1 somewhat positive, +2 moderately positive, +3 very positive). Also provide few examples for each level.*

✦ **Gemini Advanced Response:** Absolutely! Here's a breakdown of Performance Shaping Factor (PSF) ratings with definitions and examples relevant to Human Reliability Analysis (HRA):
 Understanding PSFs: Performance Shaping Factors are any conditions or influences that can either help or hinder human performance in a specific task or environment. They can be external (environmental, organizational) or internal (psychological, physiological).

Table 2 presents definitions and examples.

Important Considerations

- **Context is Key:** The impact of a PSF can vary depending on the specific task, individual, and situation.
- **Combinations:** The combined effect of multiple PSFs can be more than the sum of their parts (either positively or negatively).
- **Dynamic:** PSFs can change over time, so it's important to continually assess and reassess them.

Let me know if you'd like more examples or a deeper dive into any specific aspect of PSFs in HRA.

Table 1: Gemini Advanced suggestions for rating performance shaping factor definitions and examples

Rating	Definition	Examples
-3	This factor significantly impairs performance, making errors or failure much more likely.	Extreme time pressure, a critical system alarm malfunctioning, lack of necessary training, extreme fatigue, high levels of stress, extreme environmental conditions (e.g., excessive heat, cold, noise)
-2	This factor noticeably hinders performance and increases the probability of errors.	Moderate time pressure, distractions, outdated procedures, unfamiliar equipment, moderate fatigue, conflicting goals, moderate environmental stressors (e.g., some noise, low lighting)
-1	This factor has a slight negative impact on performance, making errors a bit more likely.	Minor distractions, minor inconveniences in the workspace, mild fatigue, slightly ambiguous procedures, minor changes to routine
0	This factor has no significant effect on performance, neither helping nor hindering it.	Standard lighting, comfortable temperature, well-maintained equipment, familiar tasks, adequate time, routine work
+1	This factor slightly improves performance and makes errors a bit less likely.	Clear instructions, positive feedback, supportive team environment, recent training refresher, minor ergonomic improvements
+2	This factor noticeably enhances performance and reduces the probability of errors.	High motivation, well-designed procedures, advanced tools and equipment, expert knowledge, strong teamwork and communication, excellent physical and mental state
+3	This factor greatly improves performance, making errors or failure much less likely.	Extensive training and experience, highly motivating rewards, optimal environmental conditions, redundant systems in place (to catch any errors), clear and unambiguous instructions, highly skilled team with a proven track record

ChatGPT 4o Response: Table 2 presents definitions and examples).

These definitions and examples provide a framework for understanding how various factors can influence human performance, ranging from highly detrimental to extremely beneficial conditions.

Table 2: ChatGPT 4o suggestions for rating performance shaping factor definitions and examples

Rating	Definition	Examples
-3	Conditions that severely degrade human performance, creating a high likelihood of errors or failures.	<p>Extremely high workload: Operators are overloaded with tasks, leading to cognitive overload and potential mistakes.</p> <p>Severe environmental conditions: Extremely noisy or hot environments that severely affect concentration and physical endurance.</p> <p>Inadequate training: Lack of essential training or experience to perform tasks correctly and safely.</p>
-2	Conditions that moderately degrade human performance, increasing the probability of errors.	<p>High workload: Operators are under significant pressure, with multiple tasks to juggle simultaneously.</p> <p>Distracting environment: Moderate levels of noise or interruptions that hinder focus and task performance.</p> <p>Insufficient training: Operators have basic training but lack depth in understanding complex procedures.</p>
-1	Conditions that slightly degrade human performance, causing occasional errors.	<p>Increased workload: Tasks are manageable but require sustained attention and can lead to fatigue.</p> <p>Mild distractions: Background noise or minor interruptions that occasionally disrupt focus.</p> <p>Basic training: Operators have fundamental knowledge but may struggle with less familiar or complex tasks.</p>
0	Conditions that neither enhance nor degrade human performance; performance is at baseline levels.	<p>Normal workload: Operators handle tasks at a typical pace without undue stress or pressure.</p> <p>Standard environment: Working conditions are typical and do not impose additional challenges.</p> <p>Adequate training: Operators are well-trained and competent in their tasks.</p>
+1	Conditions that slightly enhance human performance, reducing the likelihood of errors.	<p>Reduced workload: Tasks are slightly less demanding, allowing for better focus and less fatigue.</p> <p>Supportive environment: Mildly enhanced conditions such as reduced noise or better lighting.</p> <p>Refresher training: Operators have received recent training updates, reinforcing their skills.</p>
+2	Conditions that moderately enhance human performance, significantly reducing the probability of errors.	<p>Light workload: Tasks are well within the operator's capacity, allowing for efficient performance.</p> <p>Optimized environment: Quiet, comfortable, and well-organized workspace that promotes focus.</p> <p>Advanced training: Operators have undergone thorough training, including simulations and advanced scenarios.</p>
+3	Conditions that significantly enhance human performance, making errors highly unlikely.	<p>Minimal workload: Tasks are very easy to manage, providing ample time for thorough checks and high-quality performance.</p> <p>Ideal environment: Perfect working conditions, including ergonomic design and optimal lighting.</p> <p>Expert training: Operators are experts in their field with extensive training and experience, capable of handling even unexpected situations flawlessly.</p>