DOI:10.5937/IIZS26514L

GENERATIVE ARTIFICIAL INTELLIGENCE AS A SUPPORT TOOL FOR HSE PROFESSIONALS IN THE ASSESSMENT OF SIMULATED INDUSTRIAL WORK SITUATIONS: A PILOT EXPLORATORY STUDY

Research paper

Ana Lončar[1*], Mladenka Novaković Bežanović[2], Maja Petrović[3]
[1,2,3] University of Novi Sad, Faculty of Technical Sciences, Novi Sad, Serbia

Abstract: This pilot exploratory study investigates the use of generative artificial intelligence (AI) as support for HSE professionals in the assessment of simulated industrial work situations. Five professionals with industrial HSE experience assessed six scenarios each, resulting in 30 assessments, of which 15 were completed independently and 15 with standardized AI-generated support. The assessments were evaluated in terms of recognition of key HSE aspects, identification of additional information requirements, adequacy and specificity of control measures, unsupported assumptions, final decisions, and confidence in the assessment. Descriptive results showed slightly higher average values with AI support for recognition of key HSE aspects (85.0% vs. 80.0%), identification of additional information requirements (60.0% vs. 53.3%), and adequacy of control measures (75.0% vs. 70.0%). The specificity and applicability of proposed measures remained unchanged, while confidence was slightly higher in the AI-supported condition. Participants also supplemented AI-generated content with additional professional considerations. The findings suggest that generative AI may provide useful complementary support in HSE assessment, while professional experience and expert judgment remain essential for evaluating and contextualizing AI-generated recommendations.
Key words: generative artificial intelligence, HSE, occupational safety, risk assessment, professional judgment, decision support

INTRODUCTION

Occupational health and safety (OHS) risk assessment is a fundamental component of preventive safety management, requiring hazard identification, recognition of relevant risk factors, and selection of appropriate control measures. At the same time, industrial work environments are continuously changing due to technological development, new work processes, and increasingly complex interactions between people, equipment, and organizational systems. The International Labour Organization therefore emphasizes the importance of research addressing emerging occupational risks and the impact of new technologies on safety and health at work (International Labour Organization, 2023).

Recent developments in generative artificial intelligence (AI), particularly large language models (LLMs), have created new opportunities for supporting professional analysis and decision-making. In safety-related applications, LLMs may assist in identifying potential hazards, proposing risk-control measures, generating alternative perspectives, and structuring complex information. Previous research suggests that generative AI has potential to support selected risk assessment and safety analysis tasks, including the generation of possible failure modes and risk-mitigation options, as well as the development of initial safety-analysis models and alternative perspectives (Collier et al., 2025; Sujan et al., 2025). However, these studies also indicate important limitations. AI-generated outputs may be generic, incomplete, inconsistent, or occasionally incorrect, making expert review particularly important in safety-critical contexts (Collier et al., 2025; Oviedo-Trespalacios et al., 2023).

Consequently, generative AI in safety assessment may be more appropriately considered as a decision-support tool rather than a replacement for professional expertise. Sujan et al. (2025) showed that LLMs can provide useful starting points and alternative perspectives in safety analysis, while emphasizing that domain expertise remains necessary to identify gaps and validate AI-generated outputs. This issue is particularly relevant in industrial HSE practice, where professional judgment is influenced not only by formal knowledge but also by practical experience and understanding of actual work conditions. However, empirical evidence on the

[1*]Corresponding author, email: loncar.anaa05@gmail.com, ORCID: 0009-0002-1463-5624
[2]majadjogo@uns.ac.rs, ORCID: 0000-0003-3423-6559
[3]email:mladenkanovakovic@uns.ac.rs, ORCID: 0000-0001-5036-6682

effectiveness of AI-supported approaches in occupational safety practice remains limited. A recent systematic review concluded that the field is still at an early stage and that further research is required before widespread use of AI in OHS practice can be supported by strong evidence (Jetha et al., 2025).

Therefore, this study explores the use of generative AI as a support tool for experienced HSE professionals when assessing simulated industrial work situations. A pilot within-subject design was applied in which five HSE professionals each assessed six industrial scenarios, completing three assessments independently and three with standardized AI-generated support, resulting in a total of 30 assessments. The study compares recognition of key HSE aspects, identification of additional information requirements, adequacy and specificity of control measures, final safety decisions, and confidence in the assessment. In AI-supported tasks, perceived AI usefulness and participants' reactions to AI-generated content were also examined. The objective was to explore whether and how standardized AI-generated support may influence selected characteristics of professional HSE assessment, while retaining professional judgment as the basis for the final assessment.

MATERIAL AND METHODS

Participants and study design

A pilot exploratory study was conducted with five professionals working in occupational health, safety, and environmental (HSE) roles in industrial settings. Participants were selected purposively to include professionals with practical experience in industrial HSE contexts. Purposive sampling is commonly applied when participants are selected on the basis of specific knowledge or experience relevant to the research question (Palinkas et al., 2015). Their HSE experience ranged from 2 to 25 years, while total industrial experience ranged from 10 to 31 years. Four participants reported no previous professional use of generative artificial intelligence, while one participant reported rare professional use. Participation was voluntary, and the responses were analyzed without personal identifiers.

A within-subject design was applied. Each participant assessed six simulated industrial work situations covering typical HSE challenges: maintenance of a process pump, opening of a process pipeline, hot work in a process area, simultaneous activities in the same work zone, chemical transfer, and confined-space entry. Each participant completed three scenarios independently and three scenarios with standardized generative AI support. The allocation of AI-supported and independent scenarios was varied across participants to reduce the potential influence of scenario-specific characteristics on the comparison between the two conditions.

Procedure

For each scenario, participants were asked to identify relevant hazards and risk factors, specify additional information required before making a final HSE assessment, propose appropriate control measures, identify up to three priority measures, and provide a final recommendation regarding whether the activity could proceed. Participants also rated their confidence in their assessment on a five-point scale.

For AI-supported scenarios, participants were provided with standardized AI-generated analyses prepared using the free version of ChatGPT (OpenAI) prior to data collection. A predefined prompt structure was used to generate the analysis for each scenario. The AI-generated analyses were prepared once prior to data collection and were used without substantive modification. They remained unchanged throughout participant assessment. Free interaction with the AI system was not allowed; participants assessing the same scenario with AI support received the same predefined AI-generated content. Participants were additionally asked to rate the usefulness of the AI analysis on a five-point scale and indicate whether they accepted, supplemented, modified, or rejected parts of the AI-generated content. The final HSE assessment remained the responsibility of the participant.

Coding and data analysis

Responses were evaluated by a single HSE expert using a predefined coding framework. The coding framework was established prior to the analysis and applied to all 30 assessments using the same predefined criteria. Each complete expert response to one scenario represented one unit of analysis, resulting in 30 assessments in total.
The analysis included recognition of key HSE aspects (HRC), identification of additional information requirements (IGR), adequacy of control measures (CA), specificity and applicability of proposed measures (MSS), unsupported assumptions (UA), and self-reported confidence.
For each scenario, four reference elements were predefined for each of three assessment dimensions: key HSE aspects (HRC), additional information requirements (IGR), and adequacy of control measures (CA). HRC, IGR, and CA scores represented the percentage of these four reference elements identified or adequately addressed in each participant response. MSS was rated on a three-level scale from 0 to 2, while UA represented the number of unsupported assumptions identified in the response.
For AI-supported assessments, additional indicators included perceived AI usefulness, participants' reaction to AI-generated content, the number of expert-added relevant elements not included in the AI analysis, and the number of relevant elements overlapping with the AI-generated support.
Due to the exploratory nature of the study and the small sample size, the analysis was descriptive. Pilot studies are primarily intended to explore procedures, patterns, and feasibility rather than to provide statistically generalizable conclusions (Thabane et al., 2010). Descriptive summaries of assessments performed with and without AI support were compared using mean values and frequency distributions. No inferential statistical testing was performed, and no claims of statistical generalizability were made.

RESULTS AND DISCUSSION

A total of 30 HSE assessments were analyzed, including 15 assessments performed independently and 15 assessments performed with standardized generative AI support. The comparison focused on recognition of key HSE aspects, identification of additional information requirements, adequacy and specificity of control measures, unsupported assumptions, and participants' confidence in their assessments

Table 1. *Comparison of HSE assessment indicators with and without AI support*

Indicator	Without AI	With AI	Difference
Recognition of key HSE aspects	80.0%	85.0%	+5.0 pp
Identification of additional information requirements	53.3%	60.0%	+6.7 pp
Adequacy of control measures	70.0%	75.0%	+5.0 pp
Specificity and applicability of proposed measures (0–2)	1.73	1.73	0.00
Unsupported assumptions (count)	0	0	0
Confidence in assessment (1–5)	4.27	4.47	+0.20

The descriptive results indicate slightly higher values in the AI-supported condition for recognition of key HSE aspects, identification of additional information requirements, and adequacy of control measures. The largest difference was observed for identification of additional information requirements, increasing from 53.3% without AI support to 60.0% with AI support. Recognition of key HSE aspects increased from 80.0% to 85.0%, while adequacy of control measures increased from 70.0% to 75.0%. No difference was observed in the specificity and applicability of proposed measures, with an average MSS score of 1.73 in both

conditions. No unsupported assumptions were identified in either condition. Participants also reported slightly higher confidence in their assessments when AI support was available, with an average score of 4.47 compared with 4.27 without AI support.

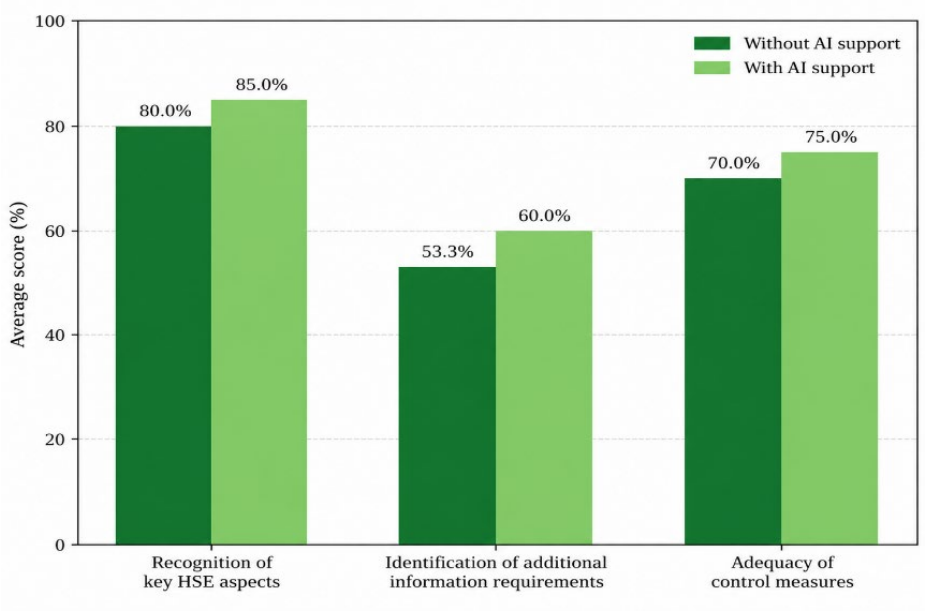


***Fig. 1**. Comparison of HSE assessment indicators with and without AI support*

In addition to the comparison between the two assessment conditions, participants' interaction with the AI-generated support was analyzed separately. Across the 15 AI-supported assessments, the mean perceived usefulness of AI was 2.60 on a five-point scale. In seven assessments, participants reported that they did not modify or supplement the AI-generated content, while in five assessments the AI output was supplemented with additional professional considerations. In three assessments, participants indicated that the AI content had been changed, supplemented, or rejected, but the exact type of reaction could not be reliably determined from the completed form.

***Table 2.** Participants' interaction with AI-generated support*

Indicator	Result
Mean perceived AI usefulness (1–5)	2.60
AI content not modified/supplemented	7/15
AI content supplemented	5/15
Reaction could not be classified reliably	3/15
Mean number of expert-added relevant elements per AI-supported assessment	1.87
Mean number of relevant elements overlapping with AI support per AI-supported assessment	3.73

The results indicate that the AI-generated analysis was not simply accepted as a final assessment in all cases. Participants frequently added professional considerations that were not included in the standardized AI support, with an average of 1.87 additional relevant elements per AI-supported assessment. At the same time, an average of 3.73 relevant elements overlapped with the AI-generated content. The results indicate that participants did not rely exclusively on the AI-generated analysis in all AI-supported assessments. Participants added an average of 1.87 relevant elements that were not included in the standardized AI support, while an average of 3.73 relevant elements overlapped with the AI-generated content.

This pattern suggests that AI-generated support was used alongside participants' own professional judgment.

The observed descriptive differences suggest that generative AI may provide support in selected components of HSE assessment. Slightly higher average values were observed in the AI-supported condition for recognition of key HSE aspects, identification of additional information requirements, and adequacy of control measures. However, these differences were relatively small, and no difference was observed in the specificity and applicability of proposed measures. This pattern is consistent with Collier et al. (2025), who reported that large language models performed relatively well in generating potential hazards and risk-mitigation ideas, while also noting that their outputs could be generic and did not reflect the depth of knowledge of experienced safety professionals.

Notably, the mean perceived usefulness of the AI-generated support was 2.60 on a five-point scale, despite the slightly higher descriptive values observed for several assessment indicators. This suggests that differences in selected assessment indicators do not necessarily correspond to professionals' subjective perception of AI usefulness.

The AI-specific results also highlight the role of professional judgment in interpreting and supplementing AI-generated content. Similar conclusions were reported by Sujan et al. (2025), who found that large language models can provide useful starting points and alternative perspectives in safety analysis, while human expertise remains essential for validating and interpreting AI-generated outputs. This is particularly important in safety-critical applications, since previous research has demonstrated that generative AI may also produce incomplete, incorrect, or potentially harmful safety-related information (Oviedo-Trespalacios et al., 2023).

The study has several limitations. The small purposively selected sample and the use of simulated industrial scenarios limit the generalizability of the findings. In addition, all responses were evaluated by a single HSE expert using a predefined researcher-developed coding framework. Although predefined criteria were applied consistently across all assessments, the absence of independent coding may have introduced interpretative bias. The results should therefore be considered descriptive and exploratory. Future studies should include larger samples, additional industrial scenarios, and independent coders, and should further evaluate the reliability of the assessment framework.

CONCLUSION

This pilot study explored the use of generative artificial intelligence as support for HSE professionals in the assessment of simulated industrial work situations. Descriptive results showed slightly higher average values in the AI-supported condition for recognition of key HSE aspects, identification of additional information requirements, adequacy of control measures, and confidence in the assessment. No difference was observed in the specificity and applicability of proposed measures.

The findings also indicate that participants used their own professional judgment alongside the AI-generated support, supplementing the AI output in several assessments with additional considerations based on their knowledge and experience. The results suggest that generative AI may have potential as a complementary decision-support tool in HSE assessment, while professional judgment remains essential for evaluating and contextualizing AI-generated content.

Due to the small purposively selected sample of five professionals, the findings should be interpreted as exploratory. Future research should include larger samples, additional industrial scenarios, independent coders, and further evaluation of how professional experience influences the use and validation of AI-generated HSE recommendations.

ACKNOWLEDGEMENT

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Education and Culture Executive Agency (EACEA). Neither the European Union nor the

granting authority can be held responsible for them. Grant agreement number 101179757. The authors would like to thank the HSE professionals who participated in the pilot study and contributed their time and professional experience.

REFERENCES

Collier, Z. A., Gruss, R. J., & Abrahams, A. S. (2025). How good are large language models at product risk assessment? *Risk Analysis, 45*(4), 766–789. https://doi.org/10.1111/risa.14351

International Labour Organization. (2023). *Global strategy on occupational safety and health 2024–30 and plan of action for its implementation* (GB.349/INS/8).

Jetha, A., Bakhtari, H., Irvin, E., Biswas, A., Smith, M. J., Mustard, C., Arrandale, V. H., Dennerlein, J. T., & Smith, P. M. (2025). Do occupational health and safety tools that utilize artificial intelligence have a measurable impact on worker injury or illness? Findings from a systematic review. *Systematic Reviews, 14*, 146. https://doi.org/10.1186/s13643-025-02869-1

Oviedo-Trespalacios, O., Peden, A. E., Cole-Hunter, T., Costantini, A., Haghani, M., Rod, J. E., Kelly, S., Torkamaan, H., Tariq, A., Newton, J. D. A., Gallagher, T., Steinert, S., Filtness, A. J., & Reniers, G. (2023). The risks of using ChatGPT to obtain common safety-related information and advice. *Safety Science, 167*, 106244. https://doi.org/10.1016/j.ssci.2023.106244

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research, 42*(5), 533–544. https://doi.org/10.1007/s10488-013-0528-y

Sujan, M., Slater, D., & Crumpton, E. (2025). How can large language models assist with a FRAM analysis? *Safety Science, 181*, 106695. https://doi.org/10.1016/j.ssci.2024.106695

Thabane, L., Ma, J., Chu, R., Cheng, J., Ismaila, A., Rios, L. P., Robson, R., Thabane, M., Giangregorio, L., & Goldsmith, C. H. (2010). A tutorial on pilot studies: The what, why and how. *BMC Medical Research Methodology, 10*, 1. https://doi.org/10.1186/1471-2288-10-1