Research Briefing:
Ethical Use of AI in the Workplace – the AI WHS Scorecard

This research was funded through partnership with the Centre for Work Health and Safety. We acknowledge the Centre for Work Health and Safety who, as a research partner, oversaw the work, reviewed the manuscript, provided project management, engagement and technical expertise.

Ethical Use of Artificial Intelligence in the Workplace – the AI WHS Scorecard

About the research

1. This research develops a risk assessment tool to assist businesses in identifying and assessing Work, Health and Safety (WHS) risks related to the use of Artificial Intelligence (AI) in the workplace: the AI WHS Scorecard. It enables AI adoption to be clearly integrated into governance and risk systems.
2. Integrated within the AI WHS Scorecard are the principles of the ethical use of AI from the Australian Government endorsed AI Ethics Principles, together with Safe Work Australia's WHS concepts of hazards and risks.
3. More and more Australian workers are experiencing the introduction of Artificial Intelligence (AI) in the workplace. AI can change the work environment significantly affecting role design, task allocation, time management, organisational structure, and communication.
4. The AI WHS Scorecard, together with an accompanying Protocol explaining its context and recommending how it may be used, supports the understanding of potential risks and hazards to workers, and supplies auditable resources for assessing and mitigating WHS risks in using AI in the workplace when making decisions during the process of introducing AI in the workplace. This helps organisations adopt AI with an ethical WHS focus, taking a strategic approach to anticipate the impacts of AI on the workplace beyond the intended process or product change.

Key findings

5. The emergent WHS risks related to AI use in the workplace were found most commonly to be harms from AI use which impacted workers psychologically - psychosocial hazards[i]. However, workers' physical safety and health might still be impacted if the use of AI influences the intensification of workflows or surveillance in the workplace, causing workers to accelerate their pace of work and thus creating new hazards. These potential risks and hazards to workers related to AI use in the workplace were identified from existing literature.
6. Consultations through interviews and workshops with AI experts and organisations that had recently introduced, or were in the process of introducing AI in the workplace, gauged awareness of and compiled concerns about AI effects on WHS. These highlighted expectations that AI would partially automate tedious and repetitive tasks; therefore, impacted employees would have to adapt to new workflows and learn how best to integrate AI solutions into their daily routines. AI would also be used for work augmentation. That is, employees would be able to improve the quality of their work owing to features and functionalities provided by AI.
7. Potentially far-reaching organisational implications of AI were acknowledged, resulting in new data-sharing arrangements, new job descriptions and the creation of new positions. AI was thought especially likely to cause deep changes to how organisations schedule or allocate workloads for their employees. AI was then positioned alongside legal obligations as using AI to schedule employee workloads needs the outcomes to be within the confines of the enterprise bargaining agreements. As well, it means AI

capabilities are starting to take over from traditional managerial tasks, and the consultations highlighted concerns that AI tools might create barriers between workers and managers. This may then challenge WHS, which requires effective communication between workers and managers.

8. High levels of concern were demonstrated, as well as strong interests in better understanding the potential effect that use of AI in the workplace may have on workers' health and safety. Even though there was common consensus on the importance of understanding the impacts of AI on workers and managing any associated potential risks, the lack of resources to take actions was also acknowledged.
9. The AI WHS Scorecard was found to improve current WHS management practices by influencing awareness of, and practices around managing, WHS risks arising from the use of AI in workplaces. The benefits were found in the deliberate, risk-aware approach which was seen to condition organisational representatives to be sensitive to the potential of AI-driven risks affecting the workplace. Evidence for these aspects came from consultations with representatives from organisations using or planning to use AI who had been asked to consider WHS management practices, and the utility of the proposed risk assessment tool, the AI WHS Scorecard; as well as consultations with WHS inspectors to collect feedback on the scorecard from a WHS practitioners' perspective.

Background

10. This research collected evidence by completing a literature review and conducting a series of consultations with AI experts, WHS professionals, regulators and policymakers, representatives from organisations adopting or having adopted AI, and others with knowledge in the field.
11. In developing the risk assessment tool, the research incorporated feedback from the consultations and made continuous refinements for improvement.
12. The AI WHS Scorecard incorporates the Australian Government endorsed AI Ethics Principles, which are used to identify and understand potential WHS risks of AI. It adopts the AI Canvas, which identifies the stages through which organisations transition as they conceive, develop, and use AI. Within these dimensions, AI-related WHS risks are described and linked to specific hazards and risks that Safe Work Australia has defined as part of Principles of Good Work Design.
13. The AI WHS Scorecard, is equipped with a risk rating matrix. The approach taken by the AI WHS Scorecard is to formulate the potential consequences of AI use for WHS from both the perspective of workers and that of an organisation and its ability to perform its core service. This design draws on how increased stress levels of workers in an organisation can lead to diminished organisational performance. The magnitude of effects on workers and organisations is measured using a five-point scaled rating. The combination of the consequence and likelihood scales results in a gradient of low to high risk levels.

Authors

Dr Andreas Cebulla, Australian Industrial Transformation Institute, Flinders University

Dr Genevieve Knight, SA Centre for Economic Studies, University of Adelaide

Dr Zygmunt Szpak, Australian Institute for Machine Learning, University of Adelaide

Dr Catherine Howell, University of Adelaide

Dr Sazzad Hussain, Centre for Work Health and Safety, NSW Department of Customer Service

Project link

https://www.centreforwhs.nsw.gov.au/knowledge-hub/ethical-use-of-artificial-intelligence-in-the-workplace-final-report

[i] **Psychosocial hazards** are factors in the design or management of work that increase the risk of work-related stress which can then lead to psychological or physical harm. These are sometimes also called psychological hazards, work-related stressors or organisational factors.