

Artificial Intelligence in Occupational Medicine: Upholding Disciplinary Integrity and Editorial Standards

The rapid advancement of Artificial Intelligence (AI) and large language models has generated increasing interest in their potential uses in occupational medicine [1]. Like other journals, we are receiving numerous submissions that assess AI systems using occupational health datasets [2]. This trend has raised important questions about the scope of occupational medicine research and the appropriate role of AI evaluation studies in a specialized occupational health journal. In this editorial, which builds on a recent one [3], we aim to further clarify the criteria we use to evaluate these submissions, distinguish between suitable and unsuitable AI research in our field, and offer guidance for future authors.

At the core of our editorial decisions is a fundamental principle: a manuscript can be technically sound and methodologically rigorous, but if the research question is not primarily an occupational health question, it falls outside the scope of occupational medicine. This distinction is not just an editorial preference; it reflects occupational medicine's identity as a clinical and public health field focused on preventing occupational diseases and safeguarding worker health.

To elucidate, consider two types of research questions that may appear superficially similar. One might ask: "Can AI language models generate reliable and readable responses to questions about an occupational disease?" This is primarily an AI question, focusing on the technical performance of AI systems—their architecture, training, and output quality. The subject matter, being occupational health, does not by itself render it an occupational medicine question and this why the paper has been rejected. By contrast, a question such as: "Can AI assist the standardized radiographic classification of pneumoconioses in occupational health surveillance programs, and what are its limitations and challenges for implementation?" directly addresses a specific occupational health task. It contemplates integrating technology into occupational health practice and recognizes the essential role of occupational health professionals in decision-making, and this is why the manuscript has been published [4].

The distinction is not sharp but is also far from semantic. Occupational medicine is concerned with worker health outcomes, disease prevention, clinical decision-making, and the protection of workers' health. While AI tools may eventually support occupational health practice, assessing whether an AI system can perform a task is categorically different from demonstrating whether its use improves worker health or should be adopted in practice. The former is a technological evaluation; the latter is an occupational health inquiry.

Occupational medicine research typically addresses questions such as: Does this intervention prevent occupational disease? Does this diagnostic approach improve early detection of work-related illnesses? Does this surveillance method more effectively protect worker health than the current practice? Does this clinical management strategy improve outcomes for exposed workers? How can occupational health professionals enhance the protection of worker welfare? These questions are intrinsically about health outcomes and require examination of clinical workflows, integration of new tools, and occupational health practices.

Conversely, AI evaluation questions generally ask: Can this AI system complete a task? How does it compare to other AI models? What are the accuracy, reliability, or readability metrics of AI-generated outputs? How do different AI architectures perform on standard datasets? These questions focus on AI capabilities and limitations—legitimate and essential in AI research—but do not in themselves represent occupational medicine research.

Challenges emerge when AI evaluation studies are presented as occupational medicine research, merely using occupational health data without meaningful engagement with occupational health practice, outcomes, or professionals' needs. In such instances, occupational health serves as a background rather than the research focus. *La Medicina del Lavoro* will consider manuscripts involving AI if they satisfy the following editorial criteria.

(i) First, the research question must fundamentally be an occupational health question. It should address a recognized occupational health task, clinical decision, or surveillance need central to occupational medicine practice. The task must be specific to occupational health—such as diagnostic classification based on standardized occupational frameworks, hazard assessment using occupational health criteria, or disease surveillance employing occupational epidemiology methods—not generic medical applications incidentally involving occupational data. The study should be firmly situated within the context of occupational health practice and professional needs, rather than merely using data as a convenient AI testing ground.

(ii) Second, the study must demonstrate a direct, rather than theoretical, connection to occupational health practice. The AI application should address a specific, standardized task utilized in occupational medicine—for example, the International Labour Organization classification of pneumoconiosis radiographs, a globally recognized occupational health diagnostic framework. Manuscripts should discuss integration into occupational health workflows and recognize occupational health professionals as primary users or beneficiaries. Studies limited to evaluating AI responses to general questions about occupational diseases, without linking them to specific tasks or practice contexts, do not fulfill this requirement.

(iii) Third, authors must explicitly acknowledge the limitations of AI systems for occupational health applications and clearly articulate that AI is a supportive tool—not a substitute for clinical judgment or professional responsibility [5]. Manuscripts should recognize current occupational health practice and regulatory frameworks, distinguish technical performance metrics from clinical utility, and indicate whether the study is exploratory or pilot in nature. Authors must clarify that AI is not proposed for autonomous clinical decision-making and discuss how occupational physicians would integrate such tools into practice. Affirming the central role of clinical judgment and professional responsibility is essential.

(iv) Fourth, methodological rigor is imperative. Studies must use standardized, validated reference datasets or classifications (e.g., the NIOSH B Reader dataset), employ appropriate statistical analyses, including confidence intervals and significance testing, and compare AI applications to existing standards or established occupational health practices. Transparent reporting of model architectures, training data, and prompting strategies is especially important given AI's technical complexity, which can obscure methodological limitations if not carefully addressed [6].

(v) Fifth, while pilot studies are acceptable, manuscripts should convincingly propose pathways by which the AI application could enhance occupational health outcomes or explain why it might not. The study should describe the occupational health issues being addressed, detail how implementation would happen in occupational settings, identify barriers and facilitators to adoption, and compare the AI approach to current standard practices. Pilot studies should place their findings within a broader research agenda aimed at advancing occupational health practice.

To illustrate the application of these criteria, consider a recent study that evaluated generative multimodal AI models in applying the ILO classification to pneumoconiosis radiographs [4]. This work directly addresses a standardized occupational health task central to surveillance programs worldwide, uses a recognized diagnostic framework, situates the study within occupational health practice, and explicitly acknowledges AI limitations and the supportive rather than autonomous role of AI tools [5]. The methodology employed standardized datasets with certified classifications and appropriate statistical analysis. The manuscript discussed potential occupational health benefits, such as standardization and efficiency improvements, while appropriately framing the work as exploratory and emphasizing the need for prospective validation. This study meets our criteria for publication due to its clear disciplinary relevance, methodological rigor, and focus

on worker health outcomes, though AI assessments are based on an outdated imaging technique applied to a disease that is fortunately in sharp decline. An interesting aspect of the study is that the comparison among four AI software tools is not intended to produce performance rankings, but rather to confirm that none of them achieves a level of accuracy sufficient to replace a NIOSH B Reader, who remains necessary to assume diagnostic responsibility.

Authors considering submissions involving AI are urged to carefully evaluate whether their research question is fundamentally an occupational health inquiry rather than an AI performance evaluation. The research should demonstrate direct relevance to occupational health practice and address specific tasks or decisions routinely made by occupational health professionals. Explicit acknowledgment of AI system limitations in occupational health contexts and clear articulation of the role of professional judgment in clinical decision-making are required. Methodological rigor must be maintained, and manuscripts should discuss potential occupational health outcomes or clinical utility, even for exploratory studies.

Manuscripts that primarily use occupational health data as AI test datasets, without addressing specific occupational health questions or demonstrating direct relevance to practice, will be desk-rejected. We encourage authors to consult the journal's scope and these criteria prior to submission. *La Medicina del Lavoro* welcomes rigorous, discipline-grounded research on AI applications in occupational health. We recognize AI's potential to meaningfully support occupational health practice, but we uphold clear editorial standards to ensure that published research advances occupational medicine and worker protection rather than repurposing occupational data solely for AI evaluation.

The distinction between "AI evaluation using occupational health data" and "occupational medicine research involving AI" is not a semantic nuance. It reflects the fundamental mission of our discipline and journal: advancing occupational health, preventing occupational disease, and protecting worker welfare. Occupational medicine is inherently a clinical and public health field that must remain grounded in occupational health questions, professional practice, and worker protection. AI is a powerful tool in the service of this mission—not an end in itself. Maintaining these boundaries will help guide authors and reviewers toward contributions that meaningfully advance occupational medicine amidst rapid technological evolution, ensuring the journal remains a dedicated venue for occupational health research rather than a general AI technology forum.

ANTONIO MUTTI

REFERENCES

1. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. *Nat Med*. 2023;29(8):1930–1940. Doi: 10.1038/s41591-023-02448-8
2. Chaudhry ZS, Choudhury A. Clinical Applications of Artificial Intelligence in Occupational Health: A Systematic Literature Review. *J Occup Environ Med*. 2024;66(12):943–955. Doi: 10.1097/JOM.0000000000003212
3. Mutti A. Where Are We Going by Applying AI Tools? Are We Going to Lose Our Jobs?. *Med Lav*. 2025;116(6):18255.
4. Baldassarre A, Padovan M, Palla A et al. Generative multimodal AI-based ILO Classification of Radiographs of Pneumoconioses. *Med Lav*. 2026;117(2):18371. Doi: 10.23749/mdl.2026.18371
5. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25(1):44–56. Doi: 10.1038/s41591-018-0300-7
6. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;385:e078378. Doi: 10.1136/bmj-2023-078378