nsc National Safety Council | MSD Solutions Lab an nsc program

Emerging Issues

Utility of Artificial Intelligence for Risk Reduction of Musculoskeletal Disorders

TABLE OF CONTENTS

Introduction

Work-related musculoskeletal disorders (MSDs) remain a significant workplace problem, even as we stand on the threshold of Industry 5.0. Unlike Industry 4.0, which focuses mainly on automation (e.g., using technologies like the Internet of Things, big data, cloud connectivity, and analytics powered by artificial intelligence (AI) and machine learning), Industry 5.0 emphasizes collaboration between workers and machines. For example, by integrating the emerging technologies of Industry 4.0, workplaces can either automate work processes to eliminate specific tasks or use 5.0 technologies (e.g., algorithm-based intelligence and collaborative robots) to reduce ergonomic risks associated with repetitive and hazardous activities. This collaboration between humans and machines enhances worker capabilities and elevates overall working conditions.

In response to the surge of interest regarding emerging technologies within the workplace, the MSD Solutions Lab intends to curate insights covering a range of pertinent emerging issues in the domain of MSDs. We aim to highlight the significance of AI and the challenges and opportunities for its integration into workplace health and safety practices.

What is Artificial Intelligence?

Artificial Intelligence (AI) is a field within computer science. AI uses high-powered computing machines to create systems capable of learning, reasoning and adapting autonomously to their behavior or strategies based on varying data or conditions, thus replicating as well as augmenting human intelligence (frontiere.io, n.d.; IBM, n.d.). In simple terms, AI can carry the tasks done by workers, complement the work that workers do and excel in tasks that go beyond what workers can do (Einola & Khoreva, 2023; Manyika & Sneader, 2018). Evaluating the performance of AI-powered models (e.g., using metrics such as accuracy, sensitivity and specificity) is essential to assess the impact of their algorithms.

Machine Learning is a subfield of AI and computer science that centers on employing algorithms and data to replicate the human learning process to enhance its precision over time (IBM, n.d.).

Deep Learning is a sub-field of machine learning and employs a customizable neural network allowing machines to handle large amounts of data and make precise decisions without human assistance (IBM, n.d.). More data yields better predictions, thus reducing uncertainty and improving prediction accuracy.

Neural Networks, also known as artificial or simulated neural networks, are a sub-field of deep learning and form the foundation of deep learning algorithms (IBM, n.d.). They are called neural networks because they imitate how our brain works, with interconnected units that communicate like biological neurons.

Computer Vision is a field of AI that enables computers and systems to derive meaningful information from digital images, videos and other visual inputs and take actions or make recommendations based on that information (IBM, n.d.). If AI enables computers to think, computer vision allows them to see, observe and understand. Neural networks and deep learning make computer vision more capable of replicating human vision. However, computer vision trains machines to perform these functions with cameras, data and algorithms.

Natural Language Processing is a field of computer science and AI. It focuses on enabling computers to comprehend and interpret human language, both written and spoken. It merges computational linguistics with statistical, machine learning and deep learning methods, allowing computers to process text and voice data to grasp the meaning, including the intended message and emotion, often in real time (IBM, n.d.). Some examples include voice-guided global positioning systems, virtual assistants, speech-to-text tools and chatbots for customer support. Figure 1 below represents a schematic representation of AI and other relevant fields.

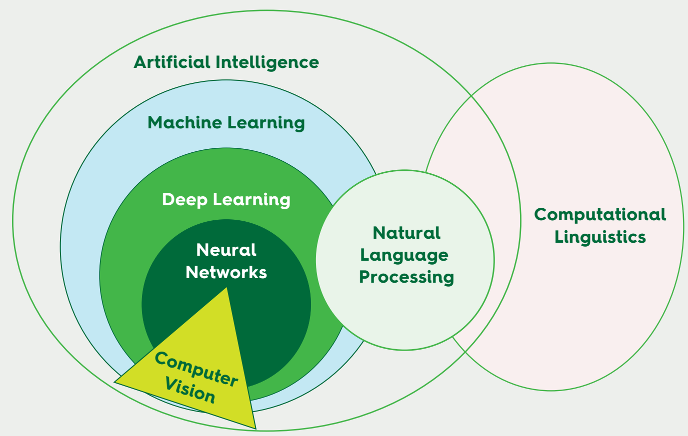

Figure 1 *Artificial Intelligence and Other Sub-Fields*

Predictive Analytics and predictive modeling are other terms discussed in AI. Predictive analytics utilizes statistical methods such as automated machine learning, deep learning, data mining and AI to build models that extract insights, detect patterns and forecast outcomes for organizations (HEAVY.AI, n.d.). Alternatively, predictive modeling, a key component of predictive analytics, uses data mining and statistics to create models that analyze historical data, uncover patterns and estimate the likelihood of future outcomes. This process involves data collection, model creation, making predictions and updating the model with new data.

AI's Relevance to Identifying Risk and Developing Solutions for MSD Prevention

Building an MSD solutions program involves identifying and addressing existing MSD hazards and risks, preventing potential new MSD hazards and risks, as well as medical management. So, how can AI address these challenges?

A. Identifying Existing Hazards and Preventing Potential New Hazards

AI can be used to identify when, where and why particular work (e.g., jobs or specific tasks) and work environments elicit risk factors for MSDs (Chan et al., 2022). This approach includes surveillance for injury prevention and control to determine when and where risk assessments can be performed and where ergonomic interventions can be developed. These scenarios can be further classified as:

1. **Identifying risk factors for MSDs (e.g., using machine learning and natural language processing).** AI can be used to analyze data from various sources, such as injury reports and workers' compensation claims, and identify patterns that may indicate MSD risk. This information can be used to develop risk assessment models and to predict the risk of MSDs for specific tasks.

It is well known that manual coding of injury data from various injury reports is labor-intensive and time-consuming. However, machine learning algorithms can automatically categorize injury narrative streams into distinct causation groups with a high degree of accuracy (Lehto et al., 2009; Marucci-Wellman et al., 2011). By learning from previously coded narratives, algorithms can code thousands of claims within minutes to hours, with high precision (Bertke et al., 2012). In fact, the widely utilized Liberty Mutual Workplace Safety Index uses a hybrid approach of manual data review, direct coding, and a combination of natural language preprocessing and machine learning to translate workers' compensation claims into the standard Bureau of Labor Statistics categories.

2. **Monitoring and analyzing data from wearable sensors to model the injury risk (e.g., using machine learning).** The 2022 Gartner report "Hype Cycle for Frontline Worker Technologies," emphasized the integral role of AI in mobile and remote enabling technologies (e.g., wearables) to reshape the workflows of frontline workers. Wearable sensors can be used to collect data on employee movements and postures. These data can be used to train an artificial neural network to model the risk of MSDs, which can predict the risk of MSDs for individual workers or specific tasks.

Matijevich et al. (2021) applied machine learning techniques to distinct subsets of theoretical and actual wearable sensor data. They systematically measured how the number, types and placement of sensors impacted algorithm accuracy for estimating low-back load during various material handling tasks. They reported that specific algorithms offer a viable, automated and highly precise approach for continuously monitoring lumbar moments with wearables across various material handling tasks. Bustos et al. (2023) developed a four-level prediction model for physical fatigue using physiological signals (e.g., heart rate, breathing rate and temperature responses) and machine learning. They further validated their model, demonstrating 82% accuracy, thus supporting the use of machine learning to monitor fatigue continuously and help prevent overexertion and related injuries.

3 **Predicting work environment risk factors to identify jobs and specific MSD risks to improve workstation design and work environments (e.g., using neural networks).** AI can be used to analyze workstation layouts and identify potential MSD risks. This information can be used to improve workstation design and reduce the risk of MSDs.

Using artificial neural network models, Perez and Nussbaum (2008) predicted starting and ending body positions during manual lifting tasks to enhance the design and evaluation of such activities. These models, trained on subsets of existing posture data, demonstrated consistent prediction accuracy across diverse scenarios, indicating their applicability to novel workplace scenarios. The models also offered insights into potential lifting strategies based on individuals' body characteristics and strength, making them valuable for anticipating postures in manual material handling tasks. Similarly, Zurada et al. (1997) developed an artificial neural network-based diagnostic system to classify jobs into low- and high-risk for potential MSDs based on lifting task attributes and workplace design.

4 **MSD risk assessments and interpreting real-time camera images to alert workers if their postures and movements are potentially hazardous (e.g., using computer vision).** Vision-based algorithms can detect and track worker tasks, specific body movements, behaviors and postures in real time through image processing, pattern recognition and machine learning. This information can be used to identify workers performing tasks that may be hazardous to their health, alert workers to risks and potential hazards, and provide feedback to help them improve their posture and movements. This prevention-through-digitization approach facilitates subsequent (re)design of the tasks and/or worker education and training, with a final goal of MSD risk reduction.

Traditionally, surface marker-based video analysis is utilized for evaluating biomechanical risk in material handling tasks, which is time-consuming and intrusive. However, computer vision techniques can extract features related to a worker's shape, size and location from images or videos of the worker's tasks without intruding on worker activities (Li et al., 2020). This information can then be used to assess MSD risk before and after the job redesign. Regarding accuracy, Mehrizi et al. (2018) demonstrated that computer vision can assess 3-D joint kinematics during lifting tasks and yield comparable results to surface marker-based systems.

B. Addressing Medical Management

In addition to identifying and addressing MSD risks when building an MSD solutions program, it is recommended to incorporate MSD medical management as part of injury prevention. An MSD medical management program aims to prevent pain, impairment and disability, and to ensure the safe return to work for all workers who experience an MSD. A health care provider (e.g., an occupational medicine physician) bridges the gap between employers, workers and the health care system. Therefore, an effective MSD medical management program requires collaboration between the employer, the employee and health care provider(s). So, how can AI help in the medical management of MSDs?

1. **Early identification and intervention (e.g., using machine learning).** AI can be used to identify employees at risk of developing MSDs. This information can provide early intervention and prevent MSDs.

   Leijon et al. (2006) employed cluster analysis, a machine learning technique, on a dataset of individuals who had not received recent MSD-related medical care. They identified three "risk" groups and five "healthy" groups concerning MSD exposure. Among the risk groups, two were primarily comprised of women affected by psychosocial or domestic challenges, while the third group, consisting primarily of men in male-dominated occupations, faced combined physical and psychosocial challenges. In contrast, the healthy clusters revealed moderate work demands and balanced education and gender distributions. The results highlight gender-specific work-life factors linked to MSD risk, which suggests machine learning techniques can be applied to develop targeted intervention strategies.

2. **Diagnosis and treatment (e.g., using deep learning).** AI can accurately assess MSD prognosis (Qiu et al., 2023), and the insights gained from AI assessments can inform patient treatments. Using predictive analytics, AI-driven systems can also assist health care providers in developing patient treatments (Tabata, 2022; Yang, 2022). However, to our knowledge, no peer-reviewed published research exists on tailored medical treatments for workers based on AI-powered algorithms.

   Regarding diagnosis, AI-driven techniques have improved medical imaging processes (from acquisition to diagnostic accuracy) to advance health care providers' clinical decision-making ability (Gyftopoulos et al., 2019). By integrating data extracted from a patient's medical history, including symptoms, laboratory outcomes, physical examination findings and prior imaging results, AI can automate image interpretation and help pinpoint the optimal imaging procedure specific to a patient's medical condition (Lakhani et al., 2018). Such insights from diagnostic imaging could assist health care providers in risk prediction and automated clinical decision-making, thus developing proactive treatment recommendations tailored to the specific MSD.

   Due to the complex structure of spinal discs and vertebrae, diagnosing disc herniation using magnetic resonance imaging (MRI) is challenging. Using an automated deep learning technique, Šušteršič et al. (2022) demonstrated strong segmentation and classification accuracy for axial and sagittal views of the spine. Compared to existing methods, this AI-driven approach excels in multiclassification tasks, offering a potentially fully automated decision support system for timely, accurate disc herniation diagnosis. Similarly, Ghosh et al. (2011) used an automated lumbar herniation diagnosis system to classify MRI spinal imaging data into "degenerative" or "normal." The system applied machine learning algorithms to automate the labeling of lumbar intervertebral discs and identified intensity and texture features from each disc's region of interest, enhancing radiologists' decision-making confidence and accuracy. Their validation studies indicated strong performance (92% or more) of the AI model in terms of accuracy, specificity and sensitivity metrics.

3. **Facilitating return to work (e.g., using machine learning).** AI can track the progress of employees recovering from MSDs. This information can help employees return to work safely and effectively. For example, machine learning algorithms can personalize recovery plans using injury diagnosis, history and progress data. These recovery plans can adapt in real time, optimizing outcomes via accurate exercises, guidance, and real-time monitoring for strength and mobility restoration. Since virtual reality replicates work environments for workers to practice skills in a virtual space safely, AI can customize training and monitor progress. Presenting gradual challenges via work simulation helps build the injured worker's confidence, facilitate re-entry to the workplace and reduce re-injury risks (Stockdale, 2023).

Ng et al. (2023) used AI to examine more than 90,000 multi-modality work injury data points. AI afforded much higher prediction accuracy than possible by solely human-driven processes, outperforming case managers' prediction errors by approximately 30%. In addition, this AI-driven approach explained the steps from the date of injury to the case closure date, thus better informing case managers. An interactive dashboard was also designed to comprehend these machine learning-powered outcomes for patients and therapists. This study suggests that AI-driven predictive analytics can empower rehabilitation professionals, including case managers, to better predict and thus improve injured workers' recovery outcomes and return to work transitions.

Systems Thinking Through AI for Risk Assessment of MSDs

While sections A and B above look at the individual level in terms of MSD risk assessments and mitigations using AI, a shift to systems thinking, focuses on the intricate interplay of various factors such as workers, processes, equipment, environment, culture and management strategies across all system levels, which is essential (Gianatti, 2023; Goh, et al., 2014; McCormack et al., 2021; McCormack et al., 2023). For example, typical risk assessment methods for MSDs focus on task analysis and specific risk factors such as force, posture, repetition, task duration, working heights, tools and equipment (McCormack et al., 2021). However, these methods often neglect broader work system conditions contributing to risks, including equipment standards, company policies, legislative compliance and effectiveness. Therefore, a systems thinking approach to risk assessment considers the entire system where injuries occur, addressing task-related factors and systemic influences. This approach leads to more appropriate and effective risk controls by dealing with broader factors affecting tasks, ultimately reducing the likelihood of MSDs.

As per McCormack et al. (2023), systems thinking-based risk assessment methods are better at identifying risk across whole work systems. To this effect, Gianatti (2023) states that AI in occupational safety and health risk assessments could elevate systems thinking. This shift can focus on systemic rather than individual factors by efficiently processing complex work data, thus exposing connections and distinctive patterns through predictive insights and data-driven strategies. For instance, AI might reveal the placement of a tool in an assembly line is impacting workers' reach, prompting the design of workflow optimization for the whole work process. It can distinguish ergonomic threats across employee groups, tailoring training and ensuring uniform safety standards. AI can also uncover subtle workplace dynamics, flagging systemic challenges such as changes in safety priorities or resistant behaviors. To fully leverage the benefits of AI, Gianatti (2023) opines that integrating AI technologies with a systems thinking approach enhances the identification, understanding and management of ergonomic risks. Since systems thinking is intertwined with safety culture, climate, resilience engineering, macroergonomics, sociotechnical systems and safety management systems, embracing systems thinking is paramount (Goh et al., 2014).

In summary, AI can analyze work-, work environment- and worker-related data from various sources quickly, reliably and in real time. More importantly, every AI-powered example provided in the current document also highlights the importance of the performance metrics of these algorithms (e.g., accuracy, sensitivity and specificity).

Challenges and Opportunities

AI runs on computer algorithms to reproduce or augment human intelligence. As discussed in the previous section, AI-powered automation effectively addresses workplace MSD issues. However, the life cycle of AI technology adheres to a structured sequence which includes the following stages: (a) ideation, (b) data acquisition, (c) data analysis, (d) engineering design and algorithm choice for model construction, (e) refinement and validation, (f) testing, (g) model deployment and (h) model administration (Rowell & Sebro, 2022). Thus, the challenges associated with implementing AI technology are complex. Meeting these challenges also necessitates cooperation among stakeholders, technology developers, workers and employees. **By tackling concerns related to accuracy, privacy, integration, data management and cost, emerging technologies such as AI can efficiently gather workplace data, contributing to enhanced occupational health and safety and the reduction of MSDs**.

The Council's recently published roadmap for piloting and implementing technologies serves as a valuable resource for employers and can be applied specifically in the context of MSDs (Guasta et al., 2022). However, the policy and regulation issues surrounding the complex nature of AI must be addressed before it can be used as a comprehensive technology for MSD risk reduction. Figure 2 below shows some of the challenges and opportunities of AI in the workplace in the risk reduction of MSDs. The following issues are also summarized as **Actionable Insights** in the Appendix.

1 Technology Issues:

When computers first became a part of everyday life, "garbage in, garbage out" became a common catchphrase implying wrong data will lead to wrong answers. This idea also applies to AI, but it is even more challenging since AI feeds on big data and is increasingly impacted by the black box analogy [i.e., since the input is hidden, it is not clear what the model is trained on or how outputs come about (Bathaee, 2018)].

Challenge:
Many commercial AI manufacturers/developers do not or cannot reveal their proprietary models' inner logistics, especially their validation. Consequently, potentially biased data is prone to discriminatory outcomes.

For example, there is a growing concern that algorithms perpetuate ethnic and gender disparities through the individuals creating them or the data they are trained on (Obermeyer et al., 2019; Jain et al., 2023). One study demonstrated that commercial health prediction algorithms, widely used in health care systems, exhibit significant racial bias leading to unequal access to care (Obermeyer et al., 2019). The authors reported that Black patients with the same risk score as White patients were actually more sick. This bias resulted from the algorithm's focus on predicting costs, not illness. Interestingly, the manufacturer of this algorithm independently verified the study findings. Subsequently, the researchers and the AI manufacturer collaborated to modify the algorithm, decreasing racial bias in results by 84%. Such a positive outcome implies the importance of transparency in content creation, sharing of code and data, and consistent reporting standards for increased trust among health care systems, health care providers, employers, health insurance companies, data scientists, AI developers/engineers and the worker/patient community.

Other AI-relevant issues increasingly reported include: non-deterministic systems [i.e., given the same input to an AI system, the output can vary (Cooper et al., 2022)]; hallucinations [i.e., a confident response by an AI model that can be incorrect or unreasonable and does not seem to be justified by its trained data (Ji et al., 2022)]; and the more recent issue of emergent abilities [the possibility that beyond a certain threshold of complexity, AI models do unpredictable things (Schaeffer et al., 2023)]. The review of these issues as well as the importance of weak AI and strong AI is beyond the scope of the current document. However, any of these issues can significantly impact the end user's confidence in AI systems.

In medical diagnosis and treatment, there is a scarcity of large, relevant and valid data sets of images as well as a lack of standardization in data capture and image processing, thus leading to ineffective training data for the AI models (Kurmis & Ianunizo, 2022). Such deficiencies could lead health care providers (e.g., radiologists) to make errors in diagnosing and categorizing images, thereby resulting in poor clinical decision-making. In other words, the generalizability of image-guided interventions and the external validity of proprietary AI systems employed at a workplace becomes questionable.

Opportunity:
Close collaboration between the health care provider, health care systems, employer, health insurance companies, AI manufacturers/developers and data scientists may be required to avoid such disastrous errors.

Challenge:
This makes administrating and implementing AI systems very expensive.

2 Worker Privacy Issues:

Using AI-derived algorithms for workplace surveillance (through data collection and continuous monitoring) generates big data. There is growing concern that AI-driven workplace monitoring and surveillance can result in the misuse of data or introduction of bias into algorithms, impacting the workers' privacy and trust in technologies and the employer (Bernhardt et al., 2023; Fukumura et al., 2021; Howard, 2022; Jetha et al., 2023; Reinhold et al., 2022; Vignola et al., 2023; Zickuhr, 2021). For example, Fukumura et al. (2021) studied office workers' perspectives on AI in the workplace. Based on focus group discussions, the authors reported employees' concerns about privacy and data misuse, with the acceptability of AI linked to the purpose of the data being collected. Participants welcomed data collection for health and wellbeing but not 'Big Brother' surveillance or the employer making assumptions about how employees allocate their time.

Suggestions included data de-identification, transparency in employee monitoring practices, favoring research over commercial data use, and keeping data from being accessed by external companies. Some advocated minimal data storage for immediate AI use only that deletes information afterward. Some suggested a written contract to safeguard data. Participants believed personal data control and secure offline export would boost acceptability. Furthermore, some participants suggested that AI should avoid monitoring or assessing productivity, especially due to the challenges associated with measuring performance across various tasks and work settings. However, some expected positive outcomes of using AI data for behavior motivation and competition are similar to existing technologies like smartphone apps and smartwatches (Fukumura et al., 2021).

Monitoring workers may also result in the Hawthorne (or so-called "observer") effect, such that when workers are aware their behavior is being observed, they may modify their actions and adhere to proper workplace practices (e.g., using appropriate lift equipment to handle an object or a patient, compliance with hand hygiene). For example, Starovoytova (2017) evaluated the time study of operating a screen-printing machine and reported that machine operators noticeably altered their behavior when they were aware their work was being measured.

On the contrary, most recently, Thiel et al. (2023) demonstrated that employees under surveillance were significantly more prone to rule-breaking behaviors such as engaging in unauthorized breaks, disregarding instructions, causing damage to workplace property, stealing office equipment and intentionally working at a reduced pace. The authors found that employee surveillance leads to a subconscious perception of reduced personal responsibility, increasing the likelihood of rule-breaking tendencies. This can be minimized by employers being transparent about the purpose of monitoring.

Workplace surveillance also disproportionately affects marginalized workers by increasing economic disparities and hindering their ability to dispute these increasingly invasive practices (Bernhardt et al., 2023, Parkes, 2023; Vignola et al., 2023; Zickuhr, 2021). For example, per Zickuhr (2021), workplace surveillance through algorithmic management practices encompass detrimental scheduling and timekeeping demands, such as erratic schedules, split shifts and restrictive definitions of “work time” during a shift. These practices can result in economic consequences like wage theft or job losses, as well as health and safety risks, including the pressure to achieve unrealistic production targets and a deficiency of meaningful breaks.

In the most recent report from the Institute for Public Policy Research, Parkes (2023) reported that employees in non-unionized, low-autonomy and low-skilled positions are more prone to being subjected to surveillance, with individuals aged 16 to 29 being the most prevalent demographic in such jobs. Also, non-unionized women and Black women face a 52% higher likelihood of experiencing workplace surveillance.

Challenges:
Some of the surveillance examples suggested in the report were: monitoring employee physical movements in workplaces like warehouses, including tracking bathroom breaks, conversations and even physiological tracking; utilizing webcam technology to observe screen time, concentration and facial expressions to gauge mood, both in office and remote settings; and implementing vehicle monitoring and dash cameras to track workers engaged in driving-related tasks.

Research further suggests the importance of establishing policies and regulations to safeguard workers so employees can contest any decisions resulting from data collection and analysis (e.g., Bernhardt et al., 2023 and Parks, 2023). To this effect, Jacobs et al. (2019) suggested that to obtain worker buy-in for technology acceptance, employers should: prioritize workplace safety with data collection during work hours; cultivate a positive safety culture; demonstrate strong evidence of technology effectiveness (e.g., through use cases); and involve and inform employees in selecting and implementing the technology.

Opportunity:
When addressing MSD medical and disability management, protecting workers’ privacy, consent, confidentiality and data security are important for both the health care provider and the employer.

A substantial volume of clinical imaging data from a patient population is essential to develop accurate AI-driven diagnosis and treatment algorithms that can help health care providers.

For example, Rowell and Sebro (2022) found that to develop a readily available AI system in medicine, it is important to: (a) assess patients’ need for data generation; (b) offer diagnostic inputs as the reference standard; (c) establish a system for data curation and storage; (d) evaluate how its use will be reimbursed for patient health care expenses from health insurance companies and patients; and (e) enlist an AI company, insurance provider, health care system or another third party to recruit programmers, acquire data and fund, advance and bring the AI system to market. Overall, these findings suggest the necessity of obtaining informed consent from workers/patients and underscore the ethics involved in the use of anonymized worker/patient data. Although the research findings reported in Section B are promising, because there are no validation studies of vendor-provided AI medical management systems that exist within the workplace, their trustworthiness as an equitable technology is to be determined.

3 Policy and Regulation Issues:

Safeguarding workplace data involves defining policies for data collection and storage, maintaining oversight for continued data accuracy, classifying who has access to the data and for what purpose, investing in stringent data security software, encrypting data files with user authentication, masking personal identifying information, and continuously monitoring and upgrading AI-powered models to deter any cybersecurity issues. For example, a regulatory standard can protect workers from electronic surveillance through strengthened algorithm design evaluations, such as an annual Algorithm Impact Assessment, thus enhancing confidence in data protection and user privacy (Patel et al., 2022).

To this effect, the White House Office of Science and Technology Policy (2022) outlines five principles for the responsible design and use of automated systems in the AI era. These five principles are: design safe and effective systems; protect people from algorithmic discrimination; provide data privacy protection and agency over data usage; make people aware of automated system use and explanation of its impact on consequential outcomes; the ability to opt out of automated processes when needed; and access to human assistance for prompt issue resolution. Furthermore, in 2022, the U.S. House Energy and Commerce Committee endorsed the American Data Privacy and Protection Act, aiming to establish national rules and safeguards for personal data held by businesses, including a focus on addressing algorithmic discrimination (Patel et al., 2022).

Opportunity:
The "Blueprint for an AI Bill of Rights" is a framework designed to safeguard individuals while harnessing technology. Informed by a wide range of experts and the public, this blueprint is accompanied by From Principles to Practice, a handbook offering practical steps to implement these principles in technology design whenever they affect the public's rights, opportunities or essential access.

More efforts on AI-related policies and procedures are increasingly evident to address ethical, legal and societal concerns. A report from the U.S. National Academy of Medicine (Matheny et al., 2020) suggests being mindful of AI-related issues such as data accessibility, quality and standardization for AI training, ethical data collection practices (e.g., informed consent, inclusivity), transparency in AI, human override capabilities, health care professional training, best practices and adaptable regulatory approaches. To this effect, the U.S. Algorithmic Accountability Act of 2022, although rejected by U.S. Congress in 2023, mandates that companies evaluate the impact of their automated systems, enhance transparency regarding the use of these systems and enable consumers to make informed decisions about critical automated processes. Similarly, the AI Risk Management Framework, developed by the U.S. National Institute of Standards and Technology (NIST), is intended for voluntary use to improve the ability to incorporate trustworthiness into designing, developing, using and evaluating AI products, services and systems.

NIST also developed a playbook, which provides voluntary suggested actions (govern, map, measure and manage) to achieve the outcomes in the Framework. Similarly, the European Union (EU)'s General Data Protection Regulation applies broadly to personal data, and its principles must be followed by AI developers and users handling personal data. Privacy and ethics are prominent in this regulation due to the use of personal data in numerous AI applications. Extending this regulation further, in 2023, the EU proposed the Artificial Intelligence Act, the world's first comprehensive AI law. The proposed act categorizes AI systems by the potential harm to individuals' health, safety or fundamental rights posed by AI technologies (i.e., unacceptable, high, limited and minimal risk), and imposes specific development and usage requirements.

Similarly, in Canada, the Public Awareness Working Group was established in 2020 to increase public awareness, build trust in AI and provide a balanced understanding of AI technology's potential applications and associated risks. This is where resources such as the NIST playbook, and the EU's toolkit might immensely help various stakeholders. The Human Factors and Ergonomics Society of the US, also instituted a policy statement, Artificial Intelligence Guardrails for Human Use, that are needed to protect people from significant harm from these systems, and to support people's ability to engage with AI usefully and effectively. The most recent press release of an executive order on Safe, Secure and Trustworthy Artificial Intelligence by the Biden Administration further reiterates the importance of supporting workers amidst AI's impact on jobs and workplaces. As per the Fact Sheet, the order directs organizations to create guidelines for AI implementation to ensure worker wellbeing, fairness and safety; generate a report on AI's labor-market effects and explore ways to aid workers in the face of job disruptions.

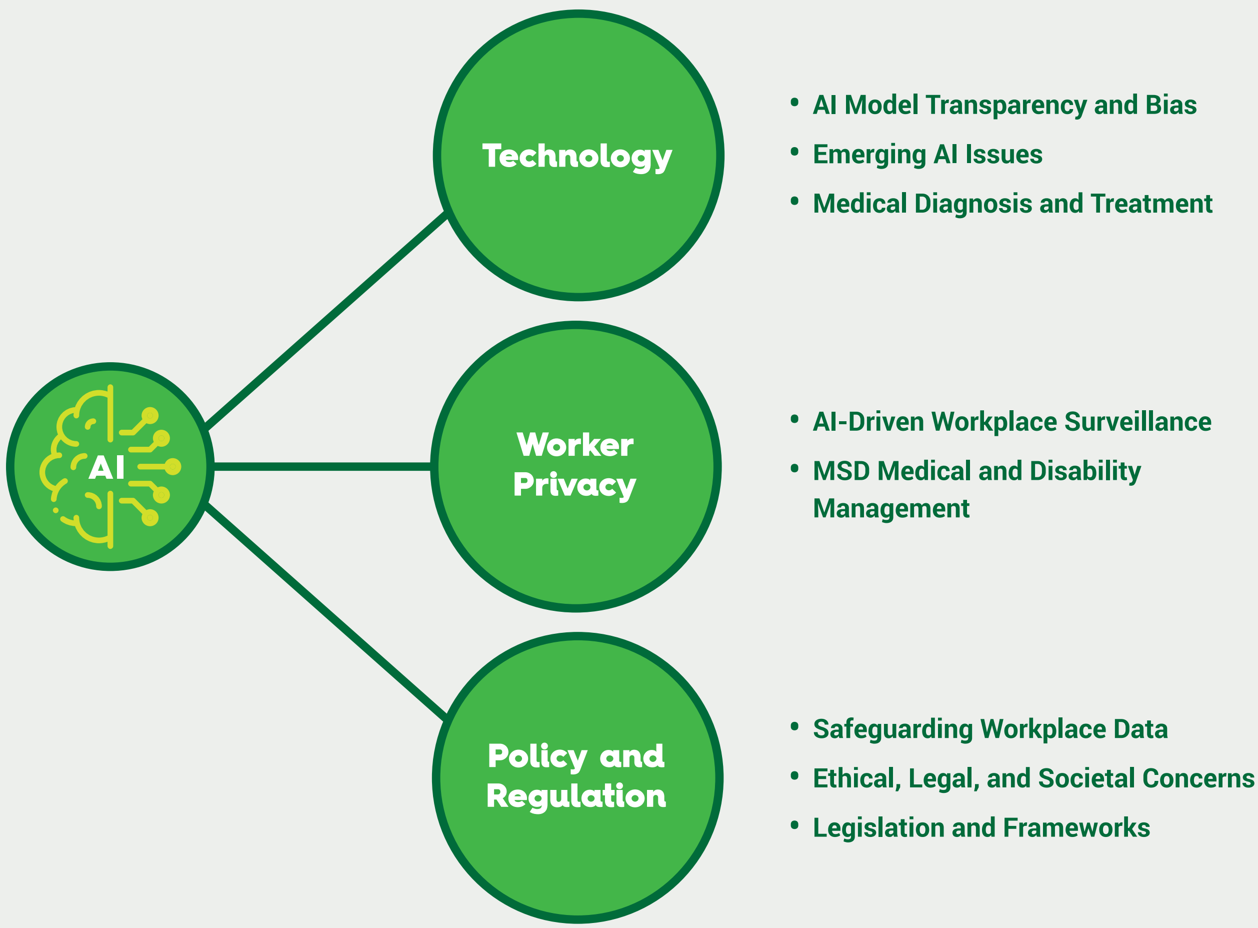

Figure 2 *Challenges and Opportunities of AI in the Workplace Regarding the Prevention of MSDs*

Conclusions

It is impossible to ignore the presence of AI in our daily working lives. **AI-powered algorithms play a crucial role in identifying, modeling and mitigating MSD risk by risk factor identification, wearable monitoring, work environment assessment, and real-time visual risk assessment and hazard alerts**. Additionally, AI could help in MSD medical management by identifying employees with pre-mature risk of MSD development. Still, there is limited work-related MSD research on tailored diagnoses and treatments (i.e., specific to medical and disability management) based on AI algorithms. AI can also track employee recovery progress, facilitating a safe return to work by personalizing recovery plans and offering real-time monitoring.

However, it is essential to note that because these data-driven algorithms are often proprietary and developed by various entities (e.g., industries and third parties), their validity and generalizability may be uncertain. As a result, this document does not include industry-specific use cases. Overall, AI rapidly analyzes data related to work, workers and the work environment, with AI-related performance metrics (e.g., accuracy, sensitivity, and specificity) crucial in its application.

Every research example provided in this document suggests the importance of human factors. In other words, from the development, execution and utility of algorithm-driven technologies to the privacy of the end-user (i.e., the worker) performing the work in varied work environments, one cannot ignore the role of human-AI collaboration. For example, the widely utilized Liberty Mutual Workplace Safety Index applies a hybrid approach of manual review and AI-driven algorithms to translate internal claims since using AI alone would not give an accurate outcome. Given that Industry 5.0 emphasizes cooperation between workers and machines, human-AI collaboration has the potential to further enhance worker capabilities and improve overall working conditions.

To this effect, an emerging field of human-centered AI is creating AI systems that could amplify and augment human capabilities (instead of replacing them) while operating transparently, delivering equitable outcomes, and respecting human privacy (Geyer et al., 2022; Xu, 2019). The EU report (European Agency for Safety and Health at Work, 2022) conveys a similar message by advocating for a human-centered perspective in AI-based worker management. Per the report, AI-based worker management should prioritize trustworthiness, transparency, empowerment and understandability in its AI design, implementation and management. This approach requires worker consultation, participation, and equitable access to information while maintaining human control over AI-based worker management to support, not replace, workers. Key steps include fostering open dialogue, providing worker training, involving workers in system development and raising awareness among stakeholders. The human-centered AI approach also involves establishing a robust ethical framework for AI-based worker management development, implementation and use, all while complying with relevant legal provisions.

In summary, **when implementing AI for MSD risk reduction, it is essential to ensure the AI systems are developed, deployed and used responsibly**. This involves considering the potential impact on workers and taking measures to mitigate any risks associated with AI technology. Ethical considerations ensure AI systems are used in ways that respect worker privacy and consent, do not discriminate against any group and prioritize the wellbeing of workers. **More importantly, AI models used in the workplace should be explainable. This means decisions made by AI systems should be transparent and understandable to both employees and employers**. Finally, while AI has the potential to assist in MSD prevention, it is imperative to approach its implementation with responsibility, ethics and transparency. This ensures AI technologies are effective, trustworthy and respectful of the wellbeing and rights of workers.

Authors and Affiliation

Rammohan Maikala, PhD, ASP, **Paige DeBaylo**, PhD, **Nadia Ullah Morin**, **Sarah Williams Ischer**, CIH, CSP
MSD Solutions Lab, National Safety Council, Itasca, IL, USA

NOTE: Correspondence should be directed to msdsolutionslab@nsc.org

Acknowledgments

Special thanks to the MSD Solutions Lab Advisory Council, who gave insights on this paper.

Funding

This research was supported by Amazon.com, Inc. Its contents and results reflect the authors' contributions and are solely their responsibility and do not necessarily represent the views of Amazon, or the official position of the National Safety Council.

References

Bathaee, Y. (2018). The artificial intelligence black box and the failure of intent and causation. *Harvard Journal of Law & Technology*, 31(2), 889–938. https://jolt.law.harvard.edu/assets/articlePDFs/v31/The-Artificial-Intelligence-Black-Box-and-the-Failure-of-Intent-and-Causation-Yavar-Bathaee.pdf

Bernhardt, A., Kresge, L., & Suleiman, R. (2023). The data-driven workplace and the case for worker technology rights. *Industrial and Labor Relations Review*, 76(1), 3–2. https://doi.org/10.1177/00197939221131558

Bertke, S. J., Meyers, A. R., Wurzelbacher, S. J., Bell, J., Lampl, M. L., & Robins, D. (2012). Development and evaluation of a Naïve Bayesian model for coding causation of workers' compensation claims. *Journal of Safety Research*, 43, 327–332. https://doi.org/10.1016/j.jsr.2012.10.012

Bustos, D., Cardoso, R., Carvalho, D. D., Guedes, J., Vaz, M., Torres Costa, J., Santos Baptista, J., & Fernandes, R. J. (2023). Exploring the applicability of physiological monitoring to manage physical fatigue in firefighters. *Sensors*, 23(11), 5127. https://doi.org/10.3390/s2311512

Chan, V. C. H., Ross, G. B., Clouthier, A. L., Fischer, S. L., & Graham, R. B. (2022). The role of machine learning in the primary prevention of work-related musculoskeletal disorders: A scoping review. *Applied Ergonomics*, 98, 103574. https://doi.org/10.1016/j.apergo.2021.103574

Cooper, AF., Frankle, J., De Sa, C., (2022). Non-determinism and the lawlessness of machine learning code. *Proceedings of the 2022 Symposium on Computer Science and Law*, 1–8, https://doi.org/10.1145/3511265.3550446

Einola, K., & Khoreva, V. (2023). Best friend or broken tool? Exploring the co-existence of humans and artificial intelligence in the workplace ecosystem. *Human Resource Management*, 62(1), 117-135. https://doi.org/10.1002/hrm.22147

European Agency for Safety and Health at Work, Christenko, A., Jankauskaitė, V., Paliokaitė, A., et al. (2022). Artificial intelligence for worker management – An overview – Report. Publications Office of the European Union. https://data.europa.eu/doi/10.2802/76354

frontiere.io. (n.d.). What is Artificial Intelligence and why it is useful for businesses. https://frontiere.io/insights/artificial-intelligence-for-businesses/

Fukumura, Y. E., Gray, J. M., Lucas, G. M., Becerik-Gerber, B., & Roll, S. C. (2021). Worker Perspectives on Incorporating Artificial Intelligence into Office Workspaces: Implications for the Future of Office Work. *International journal of environmental research and public health*, 18(4), 1690. https://doi.org/10.3390/ijerph18041690

Gartner (2022). Hype Cycle for Frontline Worker Technologies, 2022, Analyst(s): Leif-Olof Wallin, Rafael Benitez, G00770126, 8 June 2022. Gartner, Inc.

Geyer W, Weisz, J., Pinhanez, C. S., & Daly, E. (2022). What is human centered-AI? https://research.ibm.com/blog/what-is-human-centered-ai

Ghosh, S., Alomari, R. S., Chaudhary, V., & Dhillon, G. (2011). Computer-aided diagnosis for lumbar MRI using heterogeneous classifiers. *IEEE International Symposium on Biomedical Imaging: from nano to macro*, p. 1179–82. https://doi.org/10.1109/ISBI.2011.5872612

Gianatti TL (2023, October). Using AI to take a systems thinking approach to ergonomic risk assessments (2023, October). *EHSToday*. https://www.ehstoday.com/safety-technology/article/21274620/using-ai-to-take-a-systems-thinking-approach-to-ergonomic-risk-assessments/

Goh, Y. M., Love, P., & Dekker, S. (2014). Editorial for special issue – 'Systems thinking in workplace safety and health'. Accident Analysis & Prevention, 68, 1-4. https://doi.org/10.1016/j.aap.2014.02.005

Guasta, L., Lin, M-L, & Whitcomb, E. (2022). Safety technology pilot and implementation roadmap: making innovation accessible [White paper]. National Safety Council. https://www.nsc.org/workplace/safety-topics/work-to-zero/resources/reports

Gyftopoulos, S., Lin, D., Knoll, F., Doshi, A. M., Rodrigues, T. C., & Recht, M. P. (2019). Artificial intelligence in musculoskeletal imaging: Current status and future directions. *American Journal of Roentgenology*, 213, 506–513. https://doi.org/10.2214/AJR.19.21117

HEAVY.AI. (n.d.). Predictive analytics. https://www.heavy.ai/technical-glossary/predictive-analytics

Howard, J. (2022). Algorithms and the future of work. *American Journal of Industrial Medicine*, 65(12), 943-952. https://doi.org/10.1002/ajim.23429

IBM. (n.d.). What is artificial intelligence (AI)? https://www.ibm.com/topics/artificial-intelligence

IBM. (n.d.). What are neural networks? https://www.ibm.com/topics/neural-networks

IBM. (n.d.). What is natural language processing (NLP https://www.ibm.com/topics/natural-language-processing

IBM. (n.d.). What is computer vision? https://www.ibm.com/topics/computer-vision

IBM. (n.d.). What is human-centered AI? https://research.ibm.com/blog/what-is-human-centered-ai

Jacobs, J. V., Hettinger, L. J., Huang, Y.-H., Jeffries, S., Lesch, M. F., Simmons, L. A., Verma, S. K., & Willetts, J. L. (2019). Employee acceptance of wearable technology in the workplace. *Applied Ergonomics*, 78, 148-156. https://doi.org/10.1016/j.apergo.2019.03.003.

Jain, A., Brooks, J. R., Alford C. C., Chang, C. S., Mueller, N. M., Umscheid, C. A., & Bierman, A. S. (2023). Awareness of racial and ethnic bias and potential solutions to address bias with use of health care algorithms. *JAMA Health Forum*, 4(6), e231197. https://doi.org/10.1001/jamahealthforum.2023.1197

Jetha, A., Bakhtari, H., Rosella, L. C., Gignac, M. A. M., Biswas, A., Shahidi, F. V., Smith, B. T., Smith, M. J., Mustard, C., Khan, N., Arrandale, V. H., Loewen, P. J., Zuberi, D., Dennerlein, J. T., Bonaccio, S., Wu, N., Irvin, E., & Smith, P. M. (2023). Artificial intelligence and the work-health interface: A research agenda for a technologically transforming world of work. *American Journal of Industrial Medicine*, https://doi.org/10.1002/ajim.23517

Ji Z, Lee N, Frieske R, Yu, T., Su, D., Xu, Y., ... & Fung, P. (2022). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), 1-38.

Kurmis, A.P. & Ianunzio, J.R. (2022). Artificial intelligence in orthopedic surgery: evolution, current state and future directions. *Arthroplasty*, 4, 9. https://doi.org/10.1186/s42836-022-00112-z

Lakhani, P., Prater, A. B., Hutson, R. K., Andriole, K. P., Dreyer, K. J., Morey, J., Prevedello, L. M., Clark, T. J., Geis, J. R., Itri, J. N., & Hawkins, C. M. (2018). Machine learning in radiology: applications beyond image interpretation. *Journal of the American College of Radiology*, 15, 350–359. https://doi.org/10.1016/j.jacr.2017.09.044

Lehto, M., Marucci-Wellman, H., & Corns, H. (2009). Bayesian methods: a useful tool for classifying injury narratives into cause groups. *Injury Prevention*, 15, 259-265. http://dx.doi.org/10.1136/ip.2008.021337

Leijon, O., Härenstam, A., Waldenström, K., Alderling, M., & Vingård, E. (2006). Target groups for prevention of neck/shoulder and low back disorders: an exploratory cluster analysis of working and living conditions. *Work*, 27, 189–204.

Li, L., Martin, T., & Xu, X. (2020). A novel vision-based real time method for evaluating postural risk factors associated with musculoskeletal disorders. *Applied Ergonomics*, 87, 103138. https://doi.org/10.1016/j.apergo.2020.103138

McCormack, P., Read, G. J. M., Hulme, A., Lane, B. R., McLean, S., & Salmon, P. M. (2023). Using systems thinking-based risk assessment methods to assess hazardous manual tasks: A comparison of Net-HARMS, EAST-BL, FRAM, and STPA. *Ergonomics*, 66(5), 609-626. https://doi.org/10.1080/00140139.2022.2105959

McCormack, P., Read, G. J. M., Goode, N., & Salmon, P. M. (2021). Do hazardous manual handling task risk assessment methods align with systems thinking? *Safety Science*, 140, 105316. https://doi.org/10.1016/j.ssci.2021.105316

Manyika, J., & Sneader, K. (2018). AI, Automation, and the Future of Work: Ten Things to Solve for. McKinsey Global Institute.https://www.mckinsey.com/featured-insights/future-of-work/ai-automation-and-the-future-of-work-ten-things-to-solve-for#

Marucci-Wellman, H., Lehto, M., & Corns, H. A. (2011). A combined Fuzzy and Naïve Bayesian strategy can be used to assign event codes to injury narratives. *Injury Prevention*, 17, 407-414. http://dx.doi.org/10.1136/ip.2010.030593

Matheny, M., & Israni, S. T. & Ahmed, M. & Whicher, D. (Eds.). (2019). Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril. Washington, DC: National Academy of Medicine.

Matijevich, E. S., Volgyesi, P., & Zelik, K. E. (2021). A promising wearable solution for the practical and accurate monitoring of low back loading in manual material handling. *Sensors*, 21, 340. http://dx.doi.org/10.3390/s21020340

Mehrizi, R., Peng, X., Xu, X., Zhang, S., Metaxas, D., & Li, K., 2018. A computer vision based method for 3D posture estimation of symmetrical lifting. *Journal of Biomechanics*, 69, 40–46. https://doi.org/10.1016/j.jbiomech.2018.01.012

Ng, P. H. F., Chen, P. Q., Sin, Z. P. T., Lai, S. H. S., & Cheng, A. S. K. (2023). Smart Work Injury Management (SWIM) System: A machine learning approach for the prediction of sick leave and rehabilitation plan. *Bioengineering,* 10, 172. https://doi.org/10.3390/bioengineering10020172

Obermeyer, Z., Powers, B., Vogeli, C., Mullainathan, S (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447-453. https://doi.org/10.1126/science.aax2342

Patel, N.K., Shinohara, T.K., Rosa, J.M., Harrington, B.J., Kourinian, A., & Waltzman, H.W. (2022). The American Data Privacy and Protection Act: Is Federal Regulation of AI Finally on the Horizon? https://www.mayerbrown.com/en/perspectives-events/publications/2022/10/the-american-data-privacy-and-protection-act-is-federal-regulation-of-ai-finally-on-the-horizon

Perez, M. A., & Nussbaum, M. A. (2008). A neural network model for predicting postures during non-repetitive manual materials handling tasks. *Ergonomics*, 51, 1549–1564. https://doi.org/10.1080/00140130802220570

Qiu, F., Li, J., Zhang, R. & Legerlotz, K. (2023). Use of artificial neural networks in the prognosis of musculoskeletal diseases—a scoping review. *BMC Musculoskeletal Disorders*, 24, 86. https://doi.org/10.1186/s12891-023-06195-2

Reinhold, K., Järvis, M., Christenko, A., Jankauskaitė, V., Paliokaitė, A., & Riedmann, A. (2022). Artificial Intelligence for Worker Management: Risks and Opportunities; European Agency for Safety and Health at Work, https://osha.europa.eu/en/publications/artificial-intelligence-worker-management-risks-and-opportunities

Rowell, C. & Sebro, R. (2022) Who will get paid for artificial intelligence in medicine? *Radiology: Artificial Intelligence*, 4, 5, e220054. https://doi.org/10.1148/ryai.220054

Schaeffer, R., Miranda, B., & Koyejo, S. (2023). Are emergent abilities of large language models a mirage? ArXiv. /abs/2304.15004 https://doi.org/10.48550/arXiv.2304.15004

Starovoytova, D. (2017). Time-study of Rotary-Screen-Printing Operation. Industrial Engineering Letters, 7, 24-35.

Stockdale, C. (2023). Leveraging AI for successful return-to-work programs: empowering individuals on the path to recovery. https://www.linkedin.com/pulse/leveraging-ai-successful-return-to-work-programs-path-cameron/

Šušteršič, T., Ranković, V., Milovanović, V., Kovačević, V., Rasulić, L., & Filipović, N. (2022). A deep learning model for automatic detection and classification of disc herniation in magnetic resonance images. *IEEE Journal of Biomedical and Health Informatics*, 26, 12, 6036-6046. https://10.1109/JBHI.2022.3209585

Tabata, R. C. (2022). How AI could predict medical conditions and revive the healthcare system? https://www.forbes.com/sites/forbestechcouncil/2022/01/25/how-ai-could-predict-medical-conditions-and-revive-the-healthcare-system/?sh=5f12aeda6c47

Thiel, C.T., Bush, J.T., Bonner, J., Welsh, D.T., & Garud, N. (2023). Stripped of agency: The paradoxical effect of employee monitoring on deviance. *Journal of Management*, 49, 709-740. https://doi.org/10.1177/01492063211053224

Vignola, E.F., Baron, S., Abreu Plasencia, E., Hussein, M., & Cohen, N. (2023). Workers' health under algorithmic management: emerging findings and urgent research questions. *International Journal of Environmental Research and Public Health*, 10, 20(2),1239. https://doi.org/10.3390/ijerph20021239

White House Office of Science and Technology Policy (2022, October). Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People. The White House. https://www.whitehouse.gov/ostp/ai-bill-of-rights/

Xu, W. (2019). Toward human-centered AI: a perspective from human-computer interaction. *Interactions*, 26, 4, 42–46. https://doi.org/10.1145/3328485

Yang, C. C. (2022). Explainable artificial intelligence for predictive modeling in healthcare. *Journal of Healthcare Informatics Research*, 6, 228–239. https://doi.org/10.1007/s41666-022-00114-1

Zickuhr, K. (2021). Workplace surveillance is becoming the new normal for U.S. workers. United States of America. https://policycommons.net/artifacts/4580248/workplace-surveillance-is-becoming-the-new-normal-for-us/5403845/

Zurada, J., Karwowski, W., & Marras, W. S. (1997). A neural network-based system for classification of industrial jobs with respect to risk of low back disorders due to workplace design. *Applied Ergonomics*, 28, 49–58. https://doi.org/10.1016/S0003-6870(96)00034-8

Appendix

1 Technology Issues:

AI Model Transparency and Bias:

- Prioritize transparency in AI model development, including sharing code and data
- Encourage consistent reporting standards to build trust among various stakeholders, including health care systems, providers, employers and workers/patients
- Collaborate with AI manufacturers to address and mitigate biases in algorithms
- Emphasize the importance of validating AI models to ensure fair and accurate outcomes
- Advocate for transparency and validation to reduce ethnic and gender disparities perpetuated by algorithms

Emerging AI Issues:

- Acknowledge the challenges posed by non-deterministic AI systems, hallucinations and emergent abilities
- Consider the impact of these issues on end-users' confidence in AI systems

Challenges in Medical Diagnosis and Treatment:

- Recognize the scarcity of large, relevant and valid image datasets for medical AI applications
- Address the lack of standardization in data capture and image processing in the health care sector
- Highlight the potential for errors in diagnosis and clinical decision-making due to ineffective training data
- Stress the importance of collaboration between health care providers, systems, employers, insurers, AI developers and data scientists to prevent errors
- Acknowledge that administering and implementing AI systems in health care can be costly, but essential for worker/patient safety and effective care

2 Actionable Insights – Worker/Patient:

AI-Driven Workplace Surveillance:

- Be cautious about using AI-derived algorithms for workplace surveillance due to the potential misuse of data and the introduction of bias
- Recognize that such surveillance can impact workers' privacy and trust in both the technologies and the employer
- Acknowledge the potential disproportionate effects on workers
- Advocate for establishing policies and regulations to safeguard workers' rights allowing them to dispute invasive data collection and analysis
- Be mindful of the Hawthorne effect, which causes behavior modification when workers know they are observed – in contrast, surveillance may also increase rule-breaking behaviors, such as unauthorized breaks, purposefully working slowly and property damage

- Workers' acceptability of data collection hinges on its purpose, favoring health and wellbeing monitoring over 'Big Brother' surveillance – be transparent, establish written contracts, inform when and where data is being collected, and grant personal data access and control to enhance acceptability
- To gain worker buy-in for technology acceptance, prioritize workplace safety during data collection, foster a positive safety culture, provide evidence of technology effectiveness, and involve and inform employees in technology selection and implementation
- With worker acceptance, leverage AI data to yield positive outcomes in terms of productivity

MSD Medical and Disability Management:

- Protect the worker's privacy, consent, confidentiality and data security when using AI systems in health care
- Assess worker/patients' data needs, offer diagnostic inputs as reference standards, establish data curation and storage systems, and ensure secure reimbursement for health care expenses
- Collaborate with AI companies, insurance providers, health care systems or third parties to recruit programmers, acquire data and advance AI systems for accurate diagnosis and treatment
- Emphasize the importance of obtaining informed consent from workers/patients when collecting and using their data
- Highlight the ethical considerations associated with using anonymized worker/patient data in AI systems
- Exercise caution when considering vendor-provided AI systems within the workplace, as their trustworthiness and equity may not be validated through studies

3 Actionable Insights – Policy and Regulation

Safeguarding Workplace Data:

- Get buy-in for worker monitoring – the nature, extent and rationale should be explained clearly
- Define clear policies for data collection and storage in the workplace
- Establish ongoing oversight to ensure data accuracy and quality
- Determine who has access to data and for what specific purposes, restricting access as necessary
- Invest in robust data security software to protect sensitive information
- Encrypt data files and implement user authentication protocols
- Mask personal identifying information to enhance privacy
- Continuously monitor and upgrade AI-powered models to prevent cybersecurity issues

Ethical, Legal and Societal Concerns:

- Prioritize data accessibility, quality and standardization for AI training
- Promote ethical data collection practices, including obtaining informed consent and ensuring inclusivity
- Emphasize transparency in AI operations
- Recognize the importance of human override capabilities
- Invest in training to use AI effectively
- Implement best practices in AI development and use

- Support adaptable regulatory approaches to AI oversight

Legislation and Frameworks:

- Be mindful of five key principles for responsible AI system design and usage as per the White House Office of Science and Technology Policy: ensure safe system design, prevent algorithmic discrimination and protect individuals, safeguard data privacy and grant control over data usage, promote transparency and awareness of AI system usage and its impact, and user opt-out with human assistance when necessary
- Consider the AI Risk Management Framework developed by the U.S. National Institute of Standards and Technology (NIST) for improving trustworthiness in AI products and services
- Explore NIST's playbook, which offers suggested actions (govern, map, measure and manage) to achieve trustworthiness outcomes in AI systems
- Familiarize with the European Union's General Data Protection Regulation for handling personal data in AI applications and the proposed EU Artificial Intelligence Act, which categorizes AI systems by potential harm and imposes development and usage requirements
- Engage with initiatives like Canada's Public Awareness Working Group to build trust in AI and understand its applications and risks