**M. W. Johnson***

# AI, DECISION AND HEALTHY WORK

UDK 613.6:004.8
RECEIVED: 2025-10-28
ACCEPTED: 2026-06-15

This work is licensed under a Creative Commons Attribution 4.0 International License

*SUMMARY: This paper explores the architectural novelty of artificial intelligence (AI) and its implications for occupational health as both a field of practice and a domain of organisational design. Unlike previous general-purpose technologies such as databases, AI operates through stochastic modelling of normative patterns, generating predictions rather than retrieving stored information. This fundamental difference makes AI simultaneously powerful and limited: it can codify norms but struggles with specificity, nuance, and contextual variation central to occupational health decision-making. Drawing on management science, systems theory, and social theory, the paper contrasts two key organisational questions: how to use AI within existing structures, and how to reorganise occupational health given that AI exists. It argues that the latter question, aligned with Stafford Beer's systems thinking, is essential for responsible and effective adoption. AI's greatest potential lies not in automation, but in enhancing organisational perception—of human relations, meaning in research, social and spatial dynamics—functioning as a new kind of scientific instrument that amplifies reflexivity and end-to-end "flow." The paper concludes that sustainable integration of AI in occupational health demands socio-technical design, responsible innovation frameworks, and a focus on organisational learning to maintain healthy, adaptive workplaces in an era of technological transformation.*

**Key words:** *Responsible AI, Occupational Health Management, Organisational Design, Artificial Intelligence, Stafford Beer, Perception, End-to-End Design, Organisational Flow*

## INTRODUCTION

### The import of ai's architectural novelty

AI is a general-purpose digital technology, and like other general-purpose technologies such as databases and the web, it is likely to have a lasting effect both on work itself, and on the processes of regulating worker health. Importantly, AI is structurally dissimilar to previous technologies like databases *(LeCun et al. 2015, Kraska et al. 2017).* Its structural dissimilarity can be summarised in terms of its ability to codify patterns of data, and make predictions about previously unseen data based on those codified patterns. Unlike a database, which retrieves data entered into it to match a user request or query, AI processes data it hasn't seen before using what Kraska et al call "learned approximators" *(Kraska et al., 2017),* predicting likely categories to which that data belongs. It is a stochastic model of likelihoods of words, diagnoses, pixels (in the case of art generation), sound waves, etc. Succinctly, we can say that AI is a codification of "normativity" and as we shall discuss, this highlights both the transformative strength of the technology, as well as its fundamental weaknesses.

Occupational health presents an interesting case with regard to both strengths and weakne-

*Dr. Mark William Johnson, (mark.johnson-8@manchester.ac.uk), University of Manchester, Manchester, UK.

sses of AI. In the highly complex domains in which occupational health professionals have to exercise judgement about the specificity of work situations, causal agents, personal circumstances, available and viable controls for exposure, or nuance in interpretation of data, no normative technology (by definition) can be sufficiently specific to be safe and useful in all cases. For example, in the detection of early silicosis, the determination of a diagnostic threshold for referral or non-referral is rarely clear *(Aksoy et al., 2024)*. Patient consultation often remains essential to determine the level of risk and exposure, where an AI threshold of sufficiently high sensitivity to ensure no false negatives would overwhelm the system with unnecessary referrals. The organisational challenge of occupational health demands that scarce expert human resource is effectively deployed to deal with complex challenges. The question is whether and how AI may assist in this organisational challenge.

The effective allocation of scarce expert resource to address the complex challenges of occupational health can only result from having a good model of occupational health services and the demands which they must meet. The central argument of this paper is that good models of complex systems are anticipatory, and for this reason, an anticipatory technology is potentially of great value, but not for the reasons which are often cited in the AI literature (for example, automation and scale). AI's anticipatory capability raises questions about human perception which, as biologists have noted for many decades *(Rosen, 2013)*, is also anticipatory.

The mechanics whereby AI is able to make predictions grounded in training data are currently poorly understood *(Lipton, 2016)*. Unlike conventional approaches to evidence synthesis, AI outputs are not determined by data, but *constrained* by them. The process of constructing AI models requires redundancy within a dataset – that is, many representations of the same thing – and this redundancy can take the form of pixel patterns (for images), word combinations, physiological data, or time series. The training process itself is a statistical algorithm that encodes patterns at different orders of scale, suggesting that behind AI's capabilities are the mathematics of self-reference, recursion and fractality *(Wolfram, 2019).*

As with previous technological revolutions, a distinctiveness in the architecture is usually a herald of transformations to working practice, the organisation of labour and power relations across the field of work and society. While there is much hyperbole and unwarranted expectation about AI (which is likely to produce a readjustment to expectations by markets), AI's architectural novelty means that it is most likely to have knock-on effects on work activities and the health and wellbeing of workers, which may (as in previous industrial transformations) have social and political consequences. While it is too early to predict exactly how these transformations will play out, and notwithstanding the hyberbole, AI will change the way work is done – both in terms of transformation to work itself, and transformation of the ways in which worker health is maintained.

As with previous architecturally novel general purpose tools (from steam engines to aircraft), changes to working practices, internal workflows, and nature of work will almost certainly unfold in the light of this technology. However, no technological innovation from history has yet removed or replaced the role of human beings from the fundamental processes of human adaptation to their biopsychosocial environment - or what might most broadly be called "work". To talk of work is to consider those aspects of human behaviour which must continually adjust to the prevailing challenges of individual and social organisation within continually shifting constraints, of which technology is one dimension. AI is a rapid and destabilising shift in the constraints of living, to which the work of humans will have to adapt. As with all adaptation, there are some aspects of work which will need to meet the survival needs of the biological organism: those things which exist at the bottom of Maslow's hierarchy of needs *(Maslow, 1943)*, the sustenance of which Hanah Arendt distinguished as 'labour' rather than 'work' *(Arendt, 1998)*. For Arendt, work consists of something contributing to lasting value to the human world which potentially outlives individual workers that produce them. Arendt's distinction is useful when replacement of human jobs with AI is speculated upon: there is little clarity as to whether the technology will replace or produce 'work' or 'labour' in Arendt's sense, although the production and management of AI itself by humans is clearly work and not labour.

All work can potentially make workers sick, and the work of ensuring work is healthy and safe is another kind of work. As new means of working produce new ways in which workers can get sick, it equally presents new ways in which people can help maintain wellness among workers. Ultimately it is human decision which creates working environments just as it is human decision which contributes to those factors which aim to keep people well. The nature of decision and its relation to technology is a crucial element to understand the future of maintaining a population of healthy workers in a world in which AI exists.

## AI AND OCCUPATIONAL HEALTH: TWO KEY QUESTIONS

The fundamental questions that decision-makers ask about their organisation or business when faced with a new technology are a critical determining factor in the emergent dynamics that render that technology harmful or effective. Management scientist Stafford Beer presented the choice facing managers succinctly in relation to the use of the computer:

"The question which asks how to use the computer in the enterprise is, in short, the wrong question. A better formulation is to ask how the enterprise should be run given that computers exist. The best version of all is the question asking what, given computers, the enterprise now is" *(Beer, 1995)*

In occupational health, as in many other industries, the "wrong" question can be rephrased for the present context with AI where: (1) "How can I use AI in my present work in occupational health?" is the wrong question, and (2) "Given AI exists in the world, how should we organise ourselves to maintain a population of healthy workers?" is much closer to his "best question".

Answers to question 1 seek to maintain existing working practices and organisational structures, reinforcing them with the new technology. Question 2 seeks a re-evaluation of the purpose of existing organisational structures in a transformed world. This latter question is much more difficult than the former, but, as Beer implies, the first question is the one that is most likely to be asked in organisations, including Occupational health organisations. Its dangers are well-documented in organisational theory *(Beer, 1995, Seddon, 2008, Davies, 2024)* – it risks exacerbating existing organisational problems, with a consequent loss of control and accountability.

To address question 2 demands critical knowledge about the purpose and organisation of occupational health operations, the changing demands on their work, and how they might be better and more effectively organised to achieve their stated purposes. It demands an analysis of requirements for a technical system to meet the desired purposes of occupational health. However, as Beer further remarked, the stated purposes of many organisations are often not what they actually do *(Beer, 1985).* In arguing that "The Purpose Of a System is What It Does" (POSIWID), Beer highlighted that there was no point in claiming the purpose of a system is what is continually fails to do! So if "the effective organisation of occupational health services which maintain a population of healthy workers" is the purpose, the question is then how AI might help achieve this?

Before addressing this question, it is worth saying more about the wrong question but why it remains the likely route that informs AI implementation in occupational health. AI's weakness is that it is a codification of normative expectations *(Schramowski et al., 2021)*. This makes it good at predicting the words summarising a clinical case which follows the standard patterns of other clinical cases of its type, but makes it bad at identifying the specificity and nuance of individual case histories. Consequently, since these nuances are considered important by clinicians, AI's text generation requires checking. Unfortunately AI has little way of highlighting which texts require most checking, and so more work is created in ensuring reliability of all automated results. At the same time, incremental changes to models can produce differences in performance and output, meaning that clinicians cannot assume that version 1 of a tool will produce output equivalent to version 2. In order to accommodate the technology it is likely that some attenuation to the technical practice is introduced by management as bureaucratic constraints, thereby attempting to ensure reliability in the use of tools at the expen-

se of the expression of nuance and variety in the outputs.

A further explanation for why Question 1 is pursued is the rationale that AI can automate human tasks *(Frey et al., 2017),* driving efficiency from increasing scale of operations. Use of AI technologies to write medical notes is defended as a means of increasing scale and throughput of case management. Even well-intentioned advocacy of clinical AI such as that by Topol *(2019),* which envisages the automation of work giving clinicians more time for patients is naive both in terms of what the technology will actually do, and what the organisation will actually do. As Seddon, drawing on Deming, has indicated *(Seddon, 2008, Deming, 1986),* as scale increases, the end-to-end flow of organisational operations is compromised. Seddon is not alone in appealing for "designing for flow", and recent critical work in AI implementation has similarly appealed for end-to-end design and implementation. The key risk of pursuing scale through digitalisation is what Seddon calls "failure demand" – the demand created internally by the failures of the organisation's systems. There is a legitimate concern that AI may also contribute to failure demand, and that this needs to be a consideration in the implementation decisions of management.

Decision-making itself in the context of AI is an indicator of the organisational challenges presented by the technology. The seminal study of organisational decision-making by Cohen and March which led to the development of the "Garbage can model of organisational choice" *(Cohen et al., 1972)* applies particularly strongly to AI. As Cohen and March described, technologies (and particularly powerful general purpose technologies) are pulled out from the "garbage can of solutions" as an "answer" to many organisational problems as a means of demonstrating management decisiveness while evading really difficult challenges that need to be addressed.

The challenge of AI in occupational health relating to the challenge of Question 2 is to conceive of AI as a means to assist in the perception of the real needs of the organisation and its purpose, its operations and its efficiency in maintaining a population of healthy workers. Rather than see the technology as an instrument for existing management practice and operational functions, it considers the technology as a new kind of scientific instrument: something which provides new insight into the organisation of occupational health and human behaviour.

If AI's weakness is its codification of normative expectation, this codification can also be seen as a strength when looking at AI as a new kind of scientific instrument. AI is comprised of the relations between words, based on the fundamental redundancy of communication from which society itself is constituted. As a universal touchstone of social communication in any field, AI objectifies established social theories which present human experience as the result of the interplay of communication dynamics, as in the encyclopaedic social theoretical work of Luhmann *(1995)* and the more recent analytical work of Leydesdorff, which prefigures the mechanisms of AI *(Leydesdorff, 2021).* The alignment between social theory and technological function means that the technology potentially offers new means of perceiving social dynamics because it itself contains a model of the linguistic performances relating to those dynamics. Therefore, the opportunity for the exploitation of AI tools could potentially address the issues raised by question 2.

## THE PERCEPTUAL AFFORDANCES OF GENERATIVE AI

Some evidence can be brought to bear in support of the claim that a scientific use of AI in occupational health can enhance perception and coordination. Early efforts to use AI in occupational health education and research suggest that generative AI can amplify perception of human issues including:

1. Perception of the dynamics of human relationships
2. Perception of meaning in research literature
3. Perception of organisational dynamics
4. Perception of time and space in physical situations.

In each of these cases, AI's predictions are not ungrounded in data, but equally AI is not a deterministic system whose responses can be traced

to logical relationships within datasets. In recognising that real-world data constrains AI outputs rather than determines them, focus is placed on a more dynamic, experimental and iterative relationship between humans and the technology, rather than regarding a single computed output as authoritative. While it is possible to constrain an AI in ways that make it appear more deterministic (for example, by clearly specifying datasets it should draw on), in practice there is a conversational 'dance' with the technology which implicates the anticipatory systems of both humans and machines. While this potentially unsettles conventional approaches to epidemiology in occupational health which privilege traceable deterministic outputs from codified study designs and statistical results, AI's constraint-based technology presents both an epistemological challenge to conventional scientific methods, and new opportunities for scientific inquiry and action.

In Manchester University, we have experimented with virtual patients in occupational health teaching. This illustrates the significance of the conversational dance, where the dialogue between a human (usually a student) and an AI bot pretending to be an occupational health patient presents rich text of the conversation for analysis, and for specific learning points to be highlighted in real-time. This affords the kind of adaptive learning technologies which have long been the goal of applications of AI – that the technology learns about the strengths and weaknesses of the students, and tailors its responses to elicit and improve particular skills as the student engages with the bot. In so doing, it reveals the time series and events of learning and makes available for comment and analysis patterns of interaction over the flow of time.

This approach to learning has a long history, with theories variously called "cognitive apprenticeship" or "situated learning" *(Lave and Wenger, 1991)*. Fundamentally, the AI not only mimics an authentic conversational situation, but renders that situation amenable to analysis by the student or teacher, from which a meta-conversation can be conducted about the interaction itself. With software simulation such as this, AI models are useful as generative representations of normative experience. The capture of dialogical interactions which can provide insight into the behaviours of different human subjects, and the fast analysis of these dialogical interactions. For example, it becomes possible to analyse the formation of personal constructs from of students with the bot, or the revealing of mechanistic understanding of physiology revealed through the interaction, or of personal dynamics.

Without this affordance of the technology, the flow of learning interaction could only be assessed in a more haphazard way, reliant on capturing the teacher's observations and their insight into the learning needs of a particular student. With the AI agent, the machine is capable of enormous variety in the face on responses by the learner, where each response is underpinned by "knowledge" of the occupational health situation which itself is established through the likelihood of words chosen in the context of the overall discussion and the training of the AI. This variety of response encourages learners to experiment with interactions in a way that suggests that there is a balance between the variety of learner's engagements and the variety of the computer's responses: a phenomenon bearing comparison with the cybernetic "law of requisite variety" that a system of complexity x can only be managed by a system of equal or greater complexity, or more succinctly "Only variety can destroy variety" *(Ashby, 1956).*

## THE PERCEPTION OF MEANING IN RESEARCH

If AI is a high-variety machine for managing human complexity, then this affordance of the technology can also be used for the management of the variety in human-generated text of published research outputs and data analysis. AI can identify deep patterns in large bodies of text and rearticulate it in ways which synthesize its latent meaning. The capability stems directly from AI's "normativity" discussed above: if likelihoods of sequences of words can be calculated as a function derived from the analysis of vast amounts of text, then those likelihoods can be used to generate new succinct text which synthesizes all the texts. To say that this means that AI can synthesize "meaning" is to suggest that meaning itself may be a synthesis of multiple utterances leading to the

articulation of succinct summaries. The key to the mechanism whereby this facility arises is the processing of redundancy in large datasets, and while this appears as a kind of miracle, communication theory has long suggested that the nature of meaning might arise from precisely this kind of statistical processing, as the work of Leydesdorff *(2021)* identifies, building on that of Husserl, Schutz and Luhmann.

This technological capability significantly extends some of the key functions of scholars in processing published research discourse. Not only is AI able to synthesize research findings, but its estimation of data, including its production of synthetic health data, can be surprisingly accurate. This is partly because there is a correlation between the prevalence of a health condition, and the extent to which that condition is represented in the discourse, and therefore there is redundancy in the training data reflecting actual population data. Consequently, AI's weighting of word selection can provide an indicator as to normative population results. In addition to this, AI's competence in writing code means that it has multiple routes to data synthesis, including the writing of statistical algorithms such as Montecarlo simulations and agent-based models which dynamically simulate epidemiological scenarios.

In performing synthesis of findings from academic texts, AI provides powerful new affordances in the ability to abstract semantic content from texts without need for metadata encoding. Since almost all approaches to systematic review and meta-analysis rely on the coding of metadata, this semantic affordance means that underlying patterns in research literature can be identified with raw text. That further means that synthetic data can be produced from real data, meaning that data which might have been protected due to privacy concerns can be made publicly available as synthetic data. Moreover, the capacity exists to find new forms of data representation, encoding underlying mechanisms and basing predictions of occupational risk on a foundation of semantically-derived information from the background data.

As a scientific instrument, AI affords the ability to see the underlying structures of meaning which underpin the production of multiple texts in the research discourse. When coupled with computational facilities for checking and guiding AI's selection of words such as Retrieval-Augmented Generation, AI presents possibilities not only for automating processes such as systematic review, but also for automatically coding data collected as health surveillance, or creating bot-like interfaces which provide conversational interfaces with patients mimicking doctor-patient interactions aimed at extracting key information.

## PERCEPTION OF ORGANISATIONAL BEHAVIOUR

The collective establishment of meaning also suggests opportunities for the simulation and modelling of social dynamics. Since organisational culture is a manifestation of the communications within specific social constraints (for example, in a particular industry, or corporate culture), it is possible to generate multiple utterances of different orders of degree for a population. From this simulation, it is further possible to explore the dynamics of social situations within a range of shifting constraints.

This matters in occupational health because the implementation of effective regulation rests both on appropriate analysis of meaning in research data and on the awareness of organisational cullture and the likely effects of regulation. Semantically-derived analysis can create dynamic situations between simulated people in the light of complex mechanisms underpinned by mechanisms identified in the data. This can facilitate the creation of artificial communities with which policy interventions might be tested. With the possibility of simulating organisational behaviour there are also possibilities to present new ways in which organisations can reflect on their own operations: AI can be an assistant to organisational reflexivity. For example, if AI tools can amplify details of individual worker behaviour in lifting and handling, the conversation "I tend not to lift things like that... I do it like this..." can create a discursive environment where individuals and the organisation as a whole can have a conversation about tacit practice that they would not have discussed before.

Just as with the creation of dynamic models in research through the generation of code for agent-based models and statistical simulations, these simulations can be valuable for organisational modelling and prediction. AI is likely to accelerate the production of digital twins with which organisational interventions, such as new regulatory frameworks, new working practices or new technologies can be simulated, refined and optimised. Not only can existing approaches to social modelling be enhanced by AI, such as those drawing on psychological and sociological theory, but the historical foundations of social theories can both be revisited and explored dynamically and operationalised.

Management consultancies have traditionally applied theoretical techniques from the systems sciences in the form of visiting experts to provide advice, guidance and coordination of discussion among staff in a business. The affordances of the technology mean that new possibilities arise for stimulating richer conversations about strategy, operations and knowledge management which can have additional benefits to the health and wellbeing of workers. This is more important as AI itself is likely to become a destabilising force within organisations. While workplaces try to establish how or if they can exploit the technology, fast-changing workflows, roles and responsibilities can affect the underpinning regulation of workplace health. The rise of occupational stress is a likely consequence as individuals and managers attempt to keep up with the technology, and steer viable pathways for their organisations and for their own careers.

A key element in the organisational benefits and risks of AI implementation lie in what has become termed "Responsible innovation in AI". Frameworks such as the UK research council's AREA framework, derived from the work of Stilgoe et al. *(2013)*, which maps "Anticipation, Reflection, Engagement and Action" focusing specifically on the organisational need for greater transparency in the design and deployment of AI, and the organisational engagement of all stakeholders to ensure that AI interventions address end-to-end challenges and embrace the whole system of the corporation. That AI can both be a powerful assistant in this process by making visible some aspects of organisational decision-making, and that it can also be a driver of institutional pathology, presents a key inflection point for decision-makers in ensuring that organisations are optimised not only for efficiency but also for worker health.

In the wake of difficulties in implementation of AI solutions, lack of trust by clinicians, overclaiming of AI effectiveness in the academic literature, and unforeseen consequences and hazards introduced by the technology, there have been an increasing number of initiatives to encourage more holistic approaches to technology development around AI. The challenge of bias in AI outputs has been met with increasing demands for clarity in the reporting of training data and awareness of the risk of bias at the design stage of AI tools. Frameworks such as Tripod-AI *(Collins et al., 2024)* and STARD-AI *(Sounderajah et al., 2025)* seek to address these issues by introducing criteria for the reporting of AI results. These frameworks prefigure demands by regulatory bodies which set standards for approval of AI-based medical devices. The drivers for these initiatives are not only to prevent potential harms that AI might introduce, but also to provide a foundation for trust. At a more operational level, Gartner's TRiSM (AI Trust, Risk, and Security Management) framework *(Habbal et al., 2024)* provides guidelines for managing the dynamic challenges that AI presents, highlighting the need for mitigations to address risks that AI might introduce, alongside the need to ensure secure and reliable operations.

While frameworks and guidelines to address the risks of the technology are useful to try to encourage deeper reflection on the potential effects of an innovation, there are other risks which are inherent not in the technology, but in the management of organisations. Whatever frameworks are in place to guard against risky innovation, there is little that can be done to address the hubris of a senior manager seeking greater efficiency and throughput of occupational health cases by using the technology to automate human labour. As Deming warned decades ago, "It is a common assumption that quality and productivity can be achieved by putting on the screws, and by installing gadgets and new machinery […] Beat horses, and they will run faster – for a while." *(Deming, 1986).* The greatest risk that AI introduces

for occupational health is management failure to understand their organisations and to metaphorically 'beat robot horses' (maybe in the form of AI agents) in pursuit of higher productivity which will not only not arise, but will result in worker sickness and corporate failure. This of course applies to occupational health providers as much as it does to their clients.

## PERCEPTION OF SPACE AND TIME IN OCCUPATIONAL HEALTH

The patterns with which AI is trained can be multimodal: AI can process sound, images, text, time-series, geographical data, spatial data, etc. An emerging area of activity involves integrating different streams of data for capture, analysis and prediction. Augmented reality technologies, for example, are already making this possible through analysis of real-time spatial and image data. AI makes predictions based on aggregating data sources, and as richer data becomes amenable to capture and analysis, the more possibilities arise for the real-time application of the technology to occupational health and hygiene.

The literature on human perception has long emphasised the spatial and enactive aspects of human conscious processes in negotiating and making sense of our environment, notably through the work of Gibson *(2015)*. Gibson's basic concept of 'affordance' which studied how movement, physical shifts in perspective and knowledge relate directly to perception are translatable to wearable technologies which offer a way of monitoring, recording and feeding back on human affordance processes which have previously been latent and difficult to observe.

The combination of AI with virtual reality (VR) simulations (which themselves are getting easier to create because the content of simulations can be rapidly generated through AI), means that the perception of space, and the mimicking of human spatial judgement can greatly assist not only in the analysis of movement and work, but in the design of safer working environments. If perception of space and time can be accurately simulated, then design-based approaches such as 'healthy by design' to improving worker health *(Adaku et al., 2021),* can become feasible in a similar way to the use of simulation in the architectural design of public spaces (Crossrail, n.d.).

The combination of AI, VR and design activity broadens the scope and potential of the integration of all these technologies. The educational use-case of VR has tended to be problematic: simulation in VR spaces (maybe including AI agents) for OH training entails considerable set-up costs for educational establishments, while the selection of appropriate environments will be a decision that will not suit all users. Equally, keeping such resources up-to-date adds further cost and management burden. However, we already see new forms of spatial design work which is performed predominantly in VR *(Strand, 2020),* and it is likely that AI integration in these design process – from recommendations for optimal design interventions to the ability to share and analyse tacit experiences of spatial designs, whether real of virtual – could make processes of healthy design more codified. Furthermore, in the light of recent geopolitical events, we see AI and VR coming together in sophisticated military technologies – for example, in the control of drones. These military use-cases have civilian application in environmental monitoring and occupational hygiene *(Eninger and Johnson, 2015, Fanti et al., 2022).*

Spatial AI, as with the other examples of the use of AI discussed here, presents new opportunities to examine human perception and experience which has been previously tacit and unanalysable. Because AI is a general purpose technology, its integration and recombination with other technologies (not only VR) is highly likely. This carries potential to stimulate organisational reflection and conversation about working practices, work environments and their relationship to worker health and industrial organisation.

## FLOW, SCALE AND RESPONSIBLE INNOVATION FOR ORGANISATIONAL EFFECTIVENESS

To summarise the previous arguments, the choice facing occupational health organisations and workplaces is about the difference between seeking efficiency by increasing scale of operati-

ons (for example, through automation of existing human tasks), and the pursuit of efficiency through end-to-end 'flow' by exploiting AI as a scientific instrument. AI's ability to enhance perception of organisational dynamics, stimulate (and simulate) conversation, break-down barriers between organisational functions and make visible what had previously been invisible or tacit can provide occupational health providers with the means to reinspect how best to maintain a population of healthy workers. AI as a normative technology has potential to enhance perception of human relations, sensemaking, organisational dynamics and spatial awareness in the same way that telescopes and microscopes could make visible craters on the moon and bacteria.

Organisational reflexivity is central to the effectiveness of organisation more generally, and the burden of occupational disease and ill-health on a business is an indicator of sub-optimal organisation and lack of organisational reflexivity. The perceptual affordances of AI to enhance organisational reflexivity, communication, design and implementation could directly help to address these challenges. Pursuit of scale over flow not only risks introducing Seddon's "failure demand", but deeper pathologies of decision-making which lead to management misconceiving their role in maintaining *organisational* viability, but instead to see the purpose of their action in conserving *management* viability, using AI implementation as an instrument to achieve this. The consequent technological complexification of the enterprise goes hand-in-hand with a separation of management from what is managed, with its associated instability of employment. This applies both to the businesses and industries where occupational health challenges are regulated, as much as it does to businesses who are providers of occupational health services.

Pursuit of organisational flow entails whole-system, end-to-end thinking with regard to AI implementation – something which has recently been articulated by Lawrence *(2024).* The drive for "Responsible AI" and frameworks such as AREA suggest that deeper emphasis on reflexivity and engagement with all stakeholders in the organisation is recognised by the AI research community. As Stilgoe and colleagues argue with regard to responsible innovation, "The strength of a framework for responsible innovation lies in the integration and application of such approaches as an institutionally-embedded culture, one in which the total becomes far greater than the sum of its parts." *(Stilgoe et al., 2013)*

The balance between anticipation, reflection and engagement places focus on the requirements for technical systems to improve organisational effectiveness, and consequently maintain occupational health. The spirit and ideas behind this are not new: in the 1980s, computer scientists like Enid Mumford were insisting on the importance of sociotechnical design in an age when computer systems required extensive planning before any kind of intervention could be made *(Mumford, 1985).* Equally, the work on requirements engineering by Joseph Goguen and Marina Jirotka focused on formal specification of systems, but emphasised modelling the social system and taking account social theory in order to ensure organisational needs could be met *(Goguen and Jirotka, 1994).*

It is unclear whether compliance with regulatory frameworks – for example with reporting frameworks like Tripod-AI or STARD-AI, or with operational frameworks like TRiSM – can address the risks and challenges of AI implementation. The underlying problem doesn't necessarily rest with the technology itself, but with approaches to management – particularly those which seek to deploy technology to amplify the scale of operations and increase profits. These commercial drivers for AI adoption are additionally shared with the providers of AI services themselves, which – to date – are not yet profitable, but dependent of massive speculative investment, which at some point will revenue generation beyond anything that is currently being achieved. These commercial and management problems are fundamentally sociopolitical in nature, connecting technology corporations with investors and national governments on the one hand, with the small-scale politics of individual businesses, the ambitions and motivations for professional practice. These sociopolitical dynamics create organizational pressures where even the regulatory measures to support responsible AI can be co-opted into pathological mechanisms. As Seddon has remarked, complian-

ce frameworks and the target setting which can accompany them, can be damaging to productivity because organisations become target-driven rather than demand-led *(Seddon, 2024).*

In the present moment, there is much value in revisiting deeper sociotechnical analysis such as presented by Mumford or Goguen and Varela, as well as the organisational analysis of Beer, Seddon or Deming. Despite the remarkable advances of the 'black box' of modern AI, organisations as workplaces have themselves always been black boxes *(Davies, 2024),* whose requirements have been very challenging to elicit. Yet the path of ensuring that the intervention of AI within the black box of the organisation leads to more effective organisation may entail focusing the use of AI precisely on making explicit the technical requirements of workplaces. Without a clear-sighted analysis of technical requirements, haphazard AI interventions which seek to automate existing operations are only likely to make things more complex.

## CONCLUSION

This paper has situated the challenges facing decision-makers thinking about AI in occupational health within a broader organisational context. That organisational context must take into account both the architectural novelty of AI as well as the pitfalls of management decision-making with regard to technology in general. With regard to these two foci, AI's novelty in codifying normative expectations has direct consequences on decisions about its potential use. While AI's normativity will make it a tempting candidate to automate human tasks which similarly appear to be normative (such as clinical note writing), this would be to address what Stafford Beer calls the "wrong" question.

I have explained why this is the wrong question to address, drawing on literature from management science, systems researcher and (increasingly) AI research. The message from this work is that what is required is whole-system thinking and end-to-end design. The important message about AI is that the technology can assist in these deeper design processes because its normativity provides a new perspective on human perception – particularly, the perception of human relationships, meaning in research, organisational dynamics and spatial awareness.

Having said this, there is a real danger that the institutional logic of occupational health businesses, and the industries they serve, will drive an adoption pattern of AI that increases internally-generated demands on organisations and their workforce (Seddon's "failure demand"). David Graeber noted that "technology has been marshaled, if anything, to figure out ways to make us all work more." *(Graeber, 2018).* Considering Seddon's warning about the risk of "failure demand" it is not difficult to see how Graeber's prediction might be accurate. Equally, the market pressures from big tech corporations to embed AI services in daily technical operations, risk baking-in failure demand in the interests of those corporations rather than worker health. This paper has attempted to lay-out these risks in the context of what is a very powerful new kind of technology.

AI is a very young technology. Its widespread adoption has taken the form of a febrile commercially-driven initiative propped-up with astounding financial speculation. It has yet to prove itself of worth to businesses, and there are mounting concerns that the speculation on its profitability is overestimated. Nevertheless, this is a novel technical architecture which represents a permanent technological inflection point in human history. It is up to our intelligence to use it wisely to maintain a population of healthy workers and a civil society.

## LITERATURE

Adaku, E., Ankrah, N. A., Ndekugri, I. E.: Design for occupational safety and health: A theoretical framework for organisational capability, *Safety Science,* 133, 2021., 105005. doi: https://doi.org/10.1016/j.ssci.2020.105005.

Aksoy, H., Atila, Ü., Arslan, S.: Deep learning-based detection of silicosis from computed tomography images, *Computer Methods and Programs, Biomedicine Update,* 6, 2024., 100166. doi: https://doi.org/10.1016/j.cmpbup.2024.100166.

Arendt, H.: *The Human Condition,* 2nd ed., University of Chicago Press, Chicago, 1998.

Ashby, W. R.: *An Introduction to Cybernetics,* Chapman & Hall, London, 1956.

Beer, S.: *Diagnosing the System: For Organizations,* J. Wiley, Chichester, 1985.

Beer, S.: *Brain of the Firm,* J. Wiley, Chichester, 1995.

Cohen, M. D., March, J. G., Olsen, J. P.: A garbage can model of organizational choice, *Administrative Science Quarterly,* 17, 1972., 1, 1–25. doi: https://doi.org/10.2307/2392088.

Collins, G. S., Moons, K. G. M., Dhiman, P., Riley, R. D., Beam, A. L., Van Calster, B., Ghassemi, M., Liu, X., Reitsma, J. B., van Smeden, M., Boulesteix, A.-L., Camaradou, J. C., Celi, L. A., Denaxas, S., Denniston, A. K., Glocker, B., Golub, R. M., Harvey, H., Heinze, G., Hoffman, M. M.: TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods, *BMJ,* 2024., e078378. doi: https://doi.org/10.1136/bmj-2023-078378.

Crossrail: *Health and Safety Improvements,* accessible at: https://learninglegacy.crossrail.co.uk/learning-legacy-themes/health-and-safety/health-and-safety-improvement/, accessed: 24.06.2026

Davies, D.: *The Unaccountability Machine,* Profile Books, 2024.

Deming, W. E.: *Out of the Crisis,* MIT Press, Cambridge, 1986.

Eninger, R. M., Johnson, R. L.: Unmanned aerial systems in occupational hygiene - learning from allied disciplines, *Annals of Occupational Hygiene,* 59, 2015., 8, 949–958. doi: https://doi.org/10.1093/annhyg/mev041.

Fanti, G., Spinazzè, A., Borghi, F., Rovelli, S., Campagnolo, D., Keller, M., Borghi, A., Cattaneo, A., Cauda, E., Cavallo, D. M.: Evolution and applications of recent sensing technology for occupational risk assessment: A rapid review of the literature, *Sensors,* 22, 2022., 13, 4841. doi: https://doi.org/10.3390/s22134841.

Frey, C. B., Osborne, M. A.: The future of employment: How susceptible are jobs to computerisation?, *Technological Forecasting and Social Change,* 114, 2017., 1, 254–280. doi: https://doi.org/10.1016/j.techfore.2016.08.019.

Gibson, J. J.: *The Ecological Approach to Visual Perception: Classic Edition,* Psychology Press, New York and London, 2015.

Goguen, J., Jirotka, M.: *Requirements Engineering,* Academic Press, London, 1994.

Graeber, D.: *Bullshit Jobs: A Theory,* Penguin, 2018.

Habbal, A., Ali, M. K., Abuzaraida, M. A.: Artificial intelligence trust, risk and security management (AI TRiSM): Frameworks, applications, challenges and future research directions, *Expert Systems with Applications,* 240, 2024., 122442. doi: https://doi.org/10.1016/j.eswa.2023.122442.

Kraska, T., Beutel, A., Chi, E., Dean, J., Polyzotis, N.: The case for learned index structures, *Arhiv,* 2017. doi: https://doi.org/10.48550/arXiv.1712.01208.

Lawrence, N. D.: *The Atomic Human,* Random House, 2024.

LeCun, Y., Bengio, Y., Hinton, G.: Deep learning, *Nature,* 521, 2015., 7553, 436–444. doi: https://doi.org/10.1038/nature14539.

Leydesdorff, L.: *The Evolutionary Dynamics of Discursive Knowledge: Communication-Theoretical Perspectives on an Empirical Philosophy of Science,* Springer, Cham, 2021.

Lipton, Z. C.: The mythos of model interpretability, *Arhiv,* 2016. doi: https://doi.org/10.48550/arXiv.1606.03490.

Maslow, A. H.: A theory of human motivation, *Psychological Review,* 50, 1943., 4, 370–396. doi: https://doi.org/10.1037/h0054346.

Mumford, E.: *Sociotechnical Systems Design: evolving theory and practice,* Manchester Business School and Centre for Business Research, Manchester, 1985.

Rosen, R.: *Anticipatory Systems,* Elsevier, 2013.

Schramowski, P., Turan, C., Andersen, N., Rothkopf, C. A., Kersting, K.: Large pre-trained language models contain human-like biases of what is right and wrong to do, *Arhiv,* 2021. doi: https://doi.org/10.48550/arXiv.2103.11790.

Seddon, J.: *Systems Thinking in the Public Sector,* Triarchy Press, 2008.

Seddon, J.: *Rethinking Regulation: A Manifesto,* Triarchy Press, 2024.

Sounderajah, V., Guni, A., Liu, X., Collins, G. S., Karthikesalingam, A., Markar, S. R., Golub, R. M., Denniston, A. K., Shetty, S., Moher, D., Bossuyt, P. M., Darzi, A., Ashrafian, H.: The STARD-AI reporting guideline for diagnostic accuracy studies using artificial intelligence, *Nature Medicine,* 2025. doi: https://doi.org/10.1038/s41591-025-03953-8.

Stilgoe, J., Owen, R., Macnaghten, P.: Developing a framework for responsible innovation, *Research Policy,* 42, 2013., 9, 1568–1580. doi: https://doi.org/10.1016/j.respol.2013.05.008.

Strand, I.: Virtual reality in design processes, *FormAkademisk – forskningstidsskrift for design og designdidaktikk,* 13, 2020., 6. doi: https://doi.org/10.7577/formakademisk.3874.

Topol, E. J.: Deep Medicine: *How Artificial Intelligence Can Make Healthcare Human Again,* Basic Books, New York, 2019.

Wolfram, S.: *A New Kind of Science, Wolfram Media,* Champaign, Illinois, 2019.

## *UI, ODLUKA I ZDRAV RAD*

*SAŽETAK: U radu se istražuje arhitektonska novost umjetne inteligencije (UI) i njezine implikacije na zdravlje na radu, kao praksa i kao područje dizajna organizacije. Za razliku od prethodnih općenamjenskih tehnologija poput baza podataka, UI djeluje kao stohastičko modeliranje normativnih uzoraka, generirajući predviđanja, a ne prethodno pohranjene podatke. Ova temeljna razlika čini UI u isto vrijeme moćnom i ograničenom: ona može kodificirati norme, ali muči se sa specifičnošću, nijansama i kontekstualnim varijacijama važnim za donošenje odluka o zdravlju na radu. Oslanjanjajući se na upravljačku znanost, teoriju sustava i socijalnu teoriju, članak suprotstavlja dva ključna organizacijska pitanja: kako koristiti UI unutar postojećih struktura, i kako reorganizirati zdravlje na radu s obzirom na prisutnost umjetne inteligencije. Članak tvrdi da je ovo drugo pitanje, u skladu sa Stafford Beerovim razmišljanjem o sustavima, bitno za odgovorno i učinkovito prihvaćanje. Najveći potencijal umjetne inteligencije nije automatizacija već povećavanje organizacijske percepcije – ljudskih odnosa, u istraživanjima, društvenim i prostornim dinamikama – djelovanjem poput nove vrste znanstvenog instrumenta koji povećava refleksivnost i tijek od kraja do kraja. Članak zaključuje da održiva integracija umjetne inteligencije u zdravlje na radu zahtijeva socijalno tehnički dizajn, odgovorne inovacijske okvire i fokus na ogranizacijskom učenju koji će održati zdrava, prilagodljiva radna mjesta u ovom dobu tehnološke transformacije.*

***Ključne riječi:*** *odgovorna UI, upravljanje zdravljem na radu, organizacijski dizajn, umjetna inteligencija, Stafford Beer, percepcija, dizajn od kraja do kraja, organizacijski tijek*

*Stručni rad*
*Primljeno: 28.10.2025.*
*Prihvaćeno: 15.6.2026.*