AI risks: An organisational practice approach to trustworthiness

AI RISKS: AN ORGANISATIONAL PRACTICE APPROACH TO TRUSTWORTHINESS

Christian Koch[1], May Shayboun[2] and Dimosthenis Kifokeris[3]

[1&2] *School of Business, Innovation and Sustainability, Halmstad University, Kristian IV's väg, Halmstad, Sweden, 30118, Sweden*

[1] *Department of Technology and Innovation, University of Southern Denmark, Campusvej 55, 5230 Odense, Denmark*

[3] *Building Design, Architecture and Civil Engineering, Chalmers University of Technology, Sven Hultins Gata 6, Gothenburg, 412 96, Sweden*

Artificial intelligence (AI) is in need for a framework that balances the opportunities it represents with its risks. But while there is a broad consensus on this, and public regulative initiatives are taken; there is far less knowledge about how these dilemmas/opportunities/risks are played out in practice. The interest into ethics in organisation driven by a discourse on "Trustworthy AI"; makes us wonder whether an ethical approach to AI in organisation is purposeful; or needs modification. We investigate this by viewing the development and use of AI as structuration of practices. The empirical material is our own development of an AI system. Using studies of ethics in moral engineering design; AI is a question of structuration processers with unintended consequences. It is a "slide" from ethics of virtue to ethics of benefit as corroborated by engineers/designers referring ethical dilemmas to managers and politicians. The EU framework of Trustworthy AI for designing and using more accountable AI systems - considering ethics; human autonomy; harm prevention; fairness etc., conflicts with contemporary construction organisations. We propose an extension of the EU guidelines.

Keywords: Artificial Intelligence; accident prevention; contemporary organisation; ethics; explainable AI; EU guidelines; trustworthy AI

INTRODUCTION

There have been calls for studies of the embedding of artificial intelligence (AI) in organisational practice for quite some time (Orlikowski, 2016; Andrejevic, 2020; Hafermalz and Huysman, 2021). This is raising questions on the way users are configured (Woolgar 1990) and how organisational knowledge is changed (Hafermalz and Huysman, 2021). This is also involving a growing concern over downsides of implementing AI solutions (Whittaker *et al.,* 2018; Crawford *et al.,* 2019; FLI, 2023). These concerns were significantly sharpened in the spring of 2023, when the tech giants (Google, Microsoft, Apple, and Metaverse) entered a race of introducing advanced AI-based chatbots - while others are asking for a pause in development until more robust governance of AI has been developed (FLI, 2023). AI legislation was

[2] shayboun@chalmers.se

Koch, C, Shayboun, M and Kifokeris, D (2024) AI Risks: An Organisational Practice Approach to Trustworthiness *In:* Thomson, C and Neilson, C J (Eds) *Proceedings of the 40th Annual ARCOM Conference,* 2-4 September 2024, London, UK, Association of Researchers in Construction Management, 129-138.

passed by the EU in early 2024. A particular concern considers imagined future work contexts where users are "counselled", or receiving "recommendations" by systems, while the users are not equipped to interpret properly the machine "aid" and therefore face the risk of running into black boxing, bias, and faulty counsels (FLI, 2023). Therefore, the concept of "explainability" has gathered particular interest in this context, even by legislative bodies (European Commission, 2019, 2021, 2024; HLEG, 2019). We contextualise these concerns in the domain of health and safety (H&S) and the adjacent practices, while studying a multinational building contractor operating in Sweden. In this contractor, a large database of registered occupational accidents provides the basis for a possible future AI solution, aimed at supporting the prevention work in the contractors' building projects. The registration activity is carried out by safety engineers, middle managers, and H&S representatives. Construction projects are commonly recognised as a risky context and involve a continued stream of occupational accidents (Swedish Work Environment Authority, 2021). Monitoring, reporting, and preventing occupational accidents are therefore well-established organisational practices. Thus, this context of use and organisational practices hosts an excellent field for illustratively studying how AI tools might be embedded in specific work processes, routines, and knowledge practices in the future. What we allow ourselves to label contemporary organisation, is an assumption of homogenisation of organisations, giving them a series of common characteristics as ideal types for analytical discussion (Weber, 1949).

Taking the normative concepts of "Trustworthy AI" including explainability (European Commission, 2019, 2020, 2021) as a point of departure, we raise the questions of how our organisational context looks, who the users or our future AI system are, what other actors have an important role to play, and how the concepts of explainability - with its systems-in-use and process-of-development aspects - intersect with the organisational context. We investigate this by viewing the development and use of AI as structuration of practices (Giddens 1984). The empirical material is our own development of an AI system, and we modestly see this as an option of exploring the raised AI issues. Using studies of moral in engineering design ethics (Buser and Koch, 2012; Munck, 2008), AI is a question of structuration processers with unintended consequences (Siebken Scultz, 2012).

"AI" covers a host of different concepts and algorithms and is largely overlapping with Machine Learning (ML). AI is used as a broad term covering this diversity here. Both technologies are general purpose technologies (Lipsey *et al.,* 2005), which implies an even broader diversity as application begins to flourish in the future. It is also important to note that AI, as a digitalisation technology, may in many ways prolong issues raised with previous technologies. Moreover, its development is placed in parallel with contemporary developments of organisations (Ales *et al.,* 2018; Zuboff, 1988). Basically, we see strong contradictions and tensions between the concepts of AI explainability and the status of contemporary organisations. The apparent paradoxes between the two generate our framework of understanding.

Even if heavily multifaceted, we ask how we can understand AI in the context of contemporary organisational development - especially regarding the ethical aspect - and what the role of digitalisation as part of that development is. To address this in a surmountable manner, we have thus chosen to focus on one contemporary organisation in Scandinavia - to prepare for the even more specific study of the possible role of AI in monitoring, reporting and preventing occupational accidents (such as stumbling, falling from heights, intersecting with equipment) during the business practices of a multinational contractor. As with many other companies, this contractor organises its occupational safety work within its HR organisation and from there intersects with the construction project.

The contribution of this study is the understanding and placement of a construction sector activity in the context of contemporary organisations, as well as the one of ethical issues of handling data. To position this, we initiate our framework of understanding by screening the contemporary research literature on using AI for accident prevention in 2023. The result is that only one of 11 screened articles briefly touches on ethical issues, whereas the remaining 10 do not even mention the issue. As such, along with its contributions, our study commences the discourse on filling out this gap. The development of contemporary organisations responds to our main aims and falls in two categories: The criteria for systems in use, and the design process criteria - a distinction that is highly relevant here, as we look on an ongoing system's design intended for future use. The criteria for "systems in use" are the principles of respect for human autonomy, prevention of harm, fairness, and explicability (HLEG, 2019; European Commission, 2021). The "systems design process" criteria include fair design and governance, context awareness, and a multidisciplinary approach (HLEG, 2019). We present our empirical material and then discuss our imaginary use practices of the future system - both regarding the development and the sought system-in-use processes.

Our framework of understanding informs our inquiries of how our organisational context looks, who the users or our future AI system are, what other actors have an important role to play, and how the concept of explainability intersects with the case context. Within the overall framework of structuration our understanding is developed by carrying out a descriptive screening of the contemporary research literature on using AI for accident prevention. Then, it derives elements from analysing the explainability criteria of AI, and their contradiction with contemporary organisations.

According to Giddens (1984) everyday routines link up with the overall societal structures through processes of structuration. Along with Stones (2007) we allow ourselves to use structuration in a micro-sociological manner. In doing so the understanding of AI design is becoming an issue of practices that lead to intended as well as unintended consequences and structuration (Giddens, 1984; Shatzki, 2017). Unintended consequences (Siebken Schultz 2012) when developing AI might be as important as the intended design. AI designers do not automatically involve the possibility to reflect on and identify ethical questions. Either they do not recognise the moral challenges embedded in their decisions or they see their own role as neutral and pass the tasks of engaging with moral debate and making decisions on to managers or politicians (Munch, 2005). We call the latter mecha-nism "referral." It has been pointed out that developers, when facing many external constraints, deal with few ethical issues, arguing that especially in the case of "low-level normal design," relevant decisions are already embedded in technical norms and codes (Buser and Koch, 2012).

In the recent research on AI in accident prevention in 2023 (here represented by 11 articles), only one study touches briefly on ethical issues in two dimensions, model and data (Zarei *et al.,* 2023). The remaining 10 articles (all from 2023) do not even mention the issue in any dimension (such as Alkaissy et al.; Lu et al.; Luo et al.; Nowobilski and Hoła; Rafindadi et al.; Sadeghi et al.; Tian et al.; Zermane et al.;

Wang et al.; and Wang and El-Gohary). Both European and US references for AI trustworthiness explicability, or ethics, are not considered in the eleven articles, despite these issues having been highly articulated previously (HLEG, 2019). Moreover, it is becoming clear that there is a global split in treating ethical issues transparently, where European and/or US standards are not widely used. Contemporary organisation (and its academic study) has experienced a practice turn, highlighting the importance of mundane everyday routines (Schatzki, 2017) and an increased appreciation of organisational learning and knowledge. This occurs in a context of continuous pressure for higher productivity, innovation for new products and services, and an increasingly stronger managerial control (Zuboff, 2015).

The principles for systems in use are the following (1-4): (1) Respect for human autonomy: HLEG (2019) states that "Humans interacting with AI systems must be able to keep full and effective self-determination… and be able to partake in the democratic process. AI systems should not unjustifiably subordinate, coerce, deceive, manipulate, condition or herd humans. Instead, they should be designed to augment, complement and empower human cognitive, social and cultural skills". (2) Prevention of harm: According to HLEG (2019), AI systems should protect "mental and physical integrity", and the systems and their environment must be "safe and secure". Moreover, HLEG (2019) states that "Vulnerable persons should receive greater attention and be included in the development, deployment and use of AI systems". HLEG (2019) also demands attention to "situations where AI systems can cause or exacerbate adverse impacts due to asymmetries of power or information, such as between employers and employees…". (3) Fairness: This implies a commitment to ensuring equal and just distribution of both benefits and costs, and that individuals and groups are free from unfair bias, discrimination, and stigmatisation (HLEG, 2019). Equal opportunity in terms of access to education, goods, services, and technology, should also be fostered (HLEG, 2019). (4) Explicability: This means that processes need to be transparent, the capabilities and purpose of AI systems openly communicated, and decisions made explainable to those affected (HLEG, 2019). There should be opportunities to "duly contest" proposed decisions (HLEG, 2019). An explanation as to why a model has generated a particular output (recommendation) and what combination of input factors contributed to that is not always possible (i.e., "black box" algorithms) (HLEG, 2019). In those circumstances, other explicability measures (e.g., traceability, auditability, and transparent communication on the system) may be required so the system respects fundamental rights (HLEG, 2019).

How this works in a contemporary organisation

1. Respect for human autonomy has for almost 80 years (since World War II) been contradicted with organisational means such as the division of labour, separation of planning and execution of work, and ever tighter digital control (Ales *et al.,* 2018). However, this has also been accompanied by institutional representations of employees through workers' councils, trade unions, and law-supported co-determination. Nonetheless, labour has been exposed to more and more control (Zuboff, 1988) even if notions of smart work promise a co-existence of advanced systems and skilled professionals (Ales *et al.,* 2018). Therefore, human autonomy in organisations needs to be understood as relative autonomy. This is underpinned by the phenomenon of organisational professionals carrying out their practices under managerial governance, division of labour, and with only a certain level of discretion.

2. Prevention of harm at work has moved its focus from physical to psychological harm. Stress and burnout have become widespread. However, in our context (the building industry), risks of physical harm still prevail and are - despite continuous effort - hardly prevented or reduced (Swedish Work Environments Authority, 2021).
3. Fairness: It is commonplace that management divides and rules and is exercising a hegemony where employee groups are treated differently according to their estimated value creation. Companies are striving for short-sighted profit goals. Large corporations are less long-term oriented than they were before the 2007/2008 financial crisis (FCLT Global, 2019).
4. Explicability in present organisations is closely related to the underdeveloped notion of fairness described above. Knowledge and skills are distributed in the organisation according to other concerns than explicability - as rather needed skills to solve problems in a job. Personal development is often contained into limited areas, preventing explicability to fully be exploited.

According to HLEG (2019) and the European Commission (2019, 2020, 2021) the design of explainable AI - then leading to trustworthy AI - should be characterised by fairness and governance, context awareness, and a multidisciplinary approach. In more detail, HLEG (2019) offers seven guiding principles on realising this:

1. Human agency and oversight, including the respect and facilitation of fundamental human rights.
2. Technical robustness and safety, including resilience to attack and security, contingency plans and general safety, accuracy, reliability, and reproducibility.
3. Privacy and data governance, including respect for privacy, quality and integrity of data, and access to data.
4. Transparency, including traceability, explainability, and communication (see also Bedu and Fritzsche, 2022).
5. Diversity, non-discrimination and fairness, avoidance of unfair bias, accessibility and universal design, and stakeholder participation.
6. Societal and environmental wellbeing, including sustainability, environmental friendliness, and considering the impacts on society and democracy.
7. Accountability, including auditability, minimisation and reporting of negative impact, trade-offs, and redress.

METHOD

The overall approach to address the stated research questions is an interdisciplinary combination of interpretive organisational sociology, occupational accident research, and information systems research. In turn the method is designed to answer the questions of how to study our organisational context, how to identify the users or our future AI system and other actors related to the system, and how to commence analysing the intersections between the concepts of explainability and the case context. The development of the framework of understanding consisted of two steps. First, to position the argument in studies pertaining to the use AI in accident research, we screened the latest contemporary research literature on using AI for accident prevention. Eleven articles published in 2023 were selected, to sense whether the latest literature differs from previously reviewed studies and to focus on recent developments in what is interpreted as a fast-developing field. Literature was selected on the basis of larger searches in the Scopus, Science Direct and Emerald databases, thus covering important journals in the related fields, such as Safety Science and

Automation in Construction (Alkaissy *et al.,* 2023; Lu *et al.,* 2023; Luo *et al.,* 2023; Nowobilski and Hoła 2023; Rafindadi *et al.,* 2023; Sadeghi *et al.,* 2023; Tian *et al.,* 2023; Wang *et al.,* 2023; Wang and El-Gohary, 2023; Zarei *et al.,* 2023; Zermane *et al.,* 2023). Secondly, we draw on recent ethical guidelines from EU (HLEG, 2019; European Commission 2019, 2020, 2021) to develop our framework (see also Hagendorff, 2021). The empirical material was gathered in the context of a large Swedish contractor. This company and its building projects were chosen for a quite pragmatic reason, i.e., it is a convenience sample. Nevertheless, we posit it represent a typical Swedish contractor when it comes to accidents prevention and reporting. We see this approach as exploratory and illustrative. Twelve semi-structured interviews were carried out with respondents selected within the studied company. The interview respondents' positions were safety engineers (4), safety representatives (4), site manager (1), site supervisor (1), safety manager (1), and safety strategist (1). The interviews lasted between 60 and 90 minutes, they were recorded, and then transcribed. Documents related to accident prevention, reporting and analysing, were studied as well. The analysis utilised the framework of understanding and was progressing according to these recurrent themes in the single interview and document - and then iteratively consolidated those across interviews and documents. The gathering and analysis of empirical material builds on ongoing research and should be understood as exploratory. It is thus positioned towards understanding the possible practices that will emerge using AI in accident reporting, analysis, and prevention in the future. It is here a clear limitation that there is not an AI system implemented in the context, neither were informants trained in AI.

The context is a construction project-based organisation (i.e., a large Swedish contractor) where open project competitions are commonplace for obtaining customer orders and production is one-of-a-kind and situated on a specific building site. The study focuses on accident monitoring and prevention in this setting. The organisation is structured in horizontal and vertical elements interacting with each other - e.g., a horizontal structure is the H&S group. The H&S work commences in the project bid group, where the organisation of H&S in a project is planned. This includes two safety representatives and one or more safety engineers. The safety representatives are trained in design and production safety, and collaborate with the on-site H&S, Quality, and Environment (HES) manager. The horizontal organisation of H&S is in turn connected to vertical organisational levels, including a H&S strategic level that oversees all processes, tools and programs supporting H&S.

There are three topics selected for this analysis: context, explainability, and design process. Those three are interlinked in a H&S perspective - where the context and underlying assumptions of the key players in the H&S organisation set the foundation for what is mentioned in accident reports (including causes of accidents such as stumbling). For researchers and data analysts, understanding the organisation's underlying assumptions - such as those carried by safety engineers - is crucial in the explanation and usage of data through AI. Although H&S has a financial and commercial aspect for a contracting business, it is essentially a social issue. Measuring performance of an AI model is often done by the business dimensions of profitability and growth, as well as the testing and validation of the respective technical model. Current AI practices encourage the analyst and the organisation to define measurable key performance indicators that can quantify how well the AI model pays off - almost in the way an investment is evaluated. In our case study however, we found that defining the added value and the system's function in measuring success was more suitable, because safety improvement cannot be only measured in financial terms. Therefore, context understanding is highly important, because such functional measurements do not fit all social and organisational contexts. Instead, they are highly customisable and require consequence analysis, inclusiveness, and (most importantly) systematic handling of who and/or what the data represents. We thus witness here how structuration of the use processes might be mapped into an AI system with different values than this context. It is important to note that understanding the context and translating that into the design of AI systems is not necessarily a linear process. Our case has shown certain preferences of safety managers, site engineers and safety representatives for what is needed to push the efficiency of their safety planning processes. The interviews also showed that there are dominant perspectives about accident causes, pointing to human error and workers' behaviour - and rarely to, e.g., time pressure. Therefore, listening to the priorities of the end-users of an AI-based system does not guarantee its functionality, nor that it will work in the exact way the organisational or H&S group sees fit. Unintended structuration may equally occur among the users. Moreover, data limitations are a huge hurdle when it comes to designing a meaningful AI-based system. In our case, we found that explainability could be established by an examination of the context, the data, and the interplay between them. The role of the organisation in this AI project was to ensure the quality and usefulness of the data which is used to train our AI (machine learning) model. This required both horizontal and vertical elements of the organisation. The horizontal element dictated the model's purpose, added value, assumptions, limitations, and usability. The vertical element - especially the management levels - plays the key role in monitoring and analysing the data content, quality, and usefulness, thinking strategically of potential use cases, evaluating, preparing, and performing changes on what and how the data is being collected, and proposing the best course of action for improving it. This role ultimately provides the foundation for an AI system's design. By exploring the accident data, we found that it was difficult to understand, and it needed lengthy collection and cleaning processes. Importantly, it was not possible to collect and understand the reporting system and the data content without the support from the safety management. We found that using the reporting system over the years has not given detailed enough accident descriptions. Usually, AI projects are equipped with an algorithm on the achievement of desirable values on quantitative metrics, such as high accuracy and low error margins. However, we argue that algorithms should be selected based on the ML model's purpose and characteristics (Shayboun, 2022).

DISCUSSION

When fundamental ethical criteria are mobilised in an organisational practice context, it is revealed that fundamental rights (derived from ethics) have been disconnected or even abolished in contemporary (capitalist) organisations. Therefore, a strong push toward a practical moral accepting this decalage must be developed by the users. The developers' practical morality draws on a pragmatic ethic of benefit where the "good" and "decent" lie with the impact of actions, typically whether projects are a success, and the client is satisfied (Munch 2005). As our reviewed articles do little to develop an ethical stance, the discussion here compares with the EU discourse. Particularly, human autonomy in organisations needs to be understood as relative autonomy. Prevention of harm needs to be relativised by the level of acuteness of work environment problems, such as stress and burnout. Fairness can be contrasted by unequal pay discriminating against gender or minority groups. This indicates that the

EU's trustworthy AI guidelines might be few steps behind before implementing the principles of systems-in-use and design of systems for trustworthy and fair AI applications. Thus, concerns about the technology's explainability, bias, and faulty counsels are not only related to the technology's design, but the context of organisational management control. Far from all aspects of present systems-in-use discourse are the norms of trustworthy AI. The systems design process criteria include fair design and governance, context awareness, and a multidisciplinary approach, yet those are contrasted by a dominance of mass-produced software - which is developed in a manner disconnected from context and on a generic process understanding rather than a contextual practice. This pushes the organisation to adapt to the value added by the software instead of the software being sensitive to local contexts. This might create the risk of further enforcing the status of contemporary organisations focusing on financial profits. In the contractor studied, our empirical material shows how our discussion manifest in practice. Safety has been officially included in management policy and has been assigned an articulated great interest - as preventing harm to employees is postulated as an arguably important goal. However, it is unclear how this one is balanced compared to the pressure of productivity and turnover. Moreover, the organisation did not hesitate to mobilise local knowledge and learning amid apparent shared assumptions about workers' behaviour, which might be in some cases unfair - a phenomenon also seen elsewhere in construction and beyond. This context complexity not only shows that there are possibilities of unfair design, but also that the contractor organisation has contradictory objectives. The organisation could be aware or not of such complexity, because top management is somewhat engaged in improving the status quo. The contractor uses the data for yearly analysis of major accidents, which supports the generation of explicit knowledge. This practice partly supports fairness, especially in raising the voices of the directly involved in daily tasks related to production safety. Nevertheless, it is conditional for this explicit knowledge to surface that a vertical organisational communication is transparent in both a bottom-up and top-down manner. The EU guidelines juxtaposed with contemporary organisations show that the organisation should first establish awareness of its own status of human autonomy, prevention of harm, fairness, and explicability, and then reflect on such aspects in systems-in-use and design of systems. Our empirical material sheds light on how the context can be integrated within the AI design process by involving and asking the relevant actors the right questions. However, designing a multidisciplinary solution based on theorists, data analysts, managers, an industry reference group, and the direct potential users, is challenging. One way around this challenge is to study the data, realise trustworthy AI-based solutions, and set goal-oriented design boundaries in collaboration with experts outside and inside the organisation. Ultimately, an H&S system needs to be used and validated within the context. Part of our future design is to use algorithms which yield a balance between accurate and explainable results, while relying on the AI literature in selecting, combining, and testing algorithms on internal and external data and with users.

CONCLUSION

In this paper we took issue with explainability, trustworthiness and transparency of AI, using a contextual approach of structuration of practice. Therefore, we asked how our organisational context looks, who the users or our future AI system are, what other actors have an important role to play, and how the concepts of explainability would intersect with the case context. What we have learned so far from our project on developing and applying an AI (machine learning) solution for occupational accident causation and prevention in a Swedish construction contractor, is that the first and foremost condition for AI explainability is context understanding. This includes users, processes, organisation structure, and dynamics. Secondly, the interconnection between data content, representation and organisational group needs to be established. We hope that this interaction will balance the requirements for a realistic and effective design of an AI model and will address this in our future research.

REFERENCES

Ales, E and Solinas, G (2018) *Working in Digital and Smart Organisations*, Cham: Palgrave McMillan.

Alkaissy, M and Feng, H (2023) Enhancing construction safety: Machine learning-based classification of injury types, *Safety Science*, **162**, 106102.

Andrejevic, M (2020) Data Civics: A Response to the 'Ethical Turn', *Television and New Media*, **21**(6), 562-567.

Bedu, P and Fritzsche, A (2022) Can we trust AI? An empirical investigation of trust requirements and guide to successful AI adoption, *Journal of Enterprise Information Management*, **35**(2), 530-549.

Buser, M and Koch, C (2012) Eyes wide shut? Non referring, Loyalty and Practical Moral in Engineering Education, In: S Hyldgaard and Y An (Eds.) *Engineering, Development and Philosophy: American, Chinese and European Perspectives*, Cham: Springer, 253-267

Crawford, K and Whittaker, M (2019) *Ai Now 2019 Report*, New York: AI Now Institute.

European Commission (2019) *Ethics Guidelines for Trustworthy AI*, Independent High-Level Expert Group on Artificial Intelligence, European Commission, Brussels: European Commission.

European Commission (2020) *White Paper on AI: A European Approach to Excellence and Trust*, Brussels: European Commission.

European Commission (2021) *Proposal for a Regulation of the European Parliament and of the Council Laying Down Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, COM (2021) 206 Final 2021/0106 (COD)*, Brussels: European Commission (Law passed January 2024).

FCLT Global (2019) *2019 Year in Review*, Boston: FCLT.

FLI (Future of Life Institute) (2023) *Pause Giant Ai Experiments: An Open Letter*, Available from: https://futureoflife.org/open-letter/pause-giant-ai-experiments/ [Accessed 21 July 2024].

Giddens A (1984) *The Constitution of Society: Outline of the Theory of Structuration*, Berkeley: University of California Press.

Hafermalz, E and Huysman, M (2021) Please explain: Key questions for explainable AI research from an organisational perspective, *Morals + Machines*, **1**(2), 10-23.

Hagendorff, T (2021) Linking human and machine behaviour: A new approach to evaluate training data quality for beneficial machine learning, *Minds and Machines*, **31**, 563-593.

HLEG (2019) *Ethics Guidelines for Trustworthy AI*, Brussels: European Commission.

Lipsey, R, Carlaw K I and Bekhar C T (2005) *Economic Transformations: General Purpose Technologies and Long-Term Economic Growth*, Oxford: Oxford University Press.

Lu, Y, Deng, Y, Wu, G and Li, C (2023) Using cased based reasoning for automated safety risk management in construction industry, *Safety Science*, **163**, 106113.

Luo, X and Liu, Q (2023) Application of machine learning technology for occupational accident severity prediction in the case of construction collapse accidents, *Safety Science*, **163**, 106138.

Munch, B (2005) *Moral Og Videnproduktion Fragmenter Af Praktisk Moral I Den Tekniske Rådgivning*, Lyngby: BYG.DTU (Moral and Knowledge production).

Nowobilski, T and Hoła, B (2023) Methodology based on causes of accidents for forecasting the effects of falls from scaffoldings using the construction industry in Poland as an example, *Safety Science*, **157**, 105945.

Orlikowski, W J (2016) Digital work: A research agenda, *In*: B Czarniawska (Ed.), A Research Agenda for Management and Organisational Studies, Northampton: E Elgar, 88-96.

Rafindadi, A D and Alarifi, M (2023) Data mining of the essential causes of different types of fatal construction accidents, *Helyon*, **9**, 13389.

Sadeghi, H, Zhang, X and Mohandes, S R (2023) Developing an ensemble risk analysis framework for improving the safety of tower crane operations under coupled fuzzy-based environment, **Safety Science**, **158**, 105957.

Schatzki, T (2017) Persons and practices, *Management Theory and Practice*, **7**(1), 26-53.

Shayboun, M (2022) *Toward Accident Prevention Through Machine Learning Analysis of Accident Reports*, Licentiate thesis, Gothenburg: Chalmers University of Technology.

Siebken Schultz, C (2012) *The Production of Defects in a Structuration Perspective- a Study of Reactive and Proactive Problem-Solving Practices*, PhD thesis, Gothenburg: Chalmers University of Technology.

Stones, R (2005) *Structuration Theory*, London: Red Globe press.

Swedish Work Environment Authority (2021) *Lethal Occupational Accidents*, Stockholm: Swedish Work Environment Authority.

Tian, D, Li, M, Shen, Y and Han, S (2023) Intelligent mining of safety hazard information from construction documents using semantic similarity and information entropy, *Engineering Applications of Artificial Intelligence*, **119**, 105742.

Wang, Z and El-Gohary, N (2023) Deep learning-based relation extraction and knowledge graph-based representation of construction safety requirements, *Automation in Construction*, **147**, 104696.

Wang, Z, Wang, B, Ren, M and Gao, D (2023) A new hazard event classification model via deep learning and multifractal, *Computers in Industry*, **147**, 103875.

Whittaker, M and Schwartz, O (2018) *Ai Now 2018 Report*, New York: AI Now Institute.

Weber, M (1949) *Max Weber on the Methodology of the Social Sciences*, Glencoe: Free Press.

Woolgar S (1990) Configuring the user, *Sociological Review*, **38**(1).

Zarei, E, Khan, F, Abbassi, R (2023) How to account artificial intelligence in human factor analysis of complex systems? *Process Safety and Environmental Protection*, **171**, 736-50.

Zermane, A, Tohir, M and Yusoff, H M (2023) Predicting fatal fall from heights accidents using random forest classification machine learning model, Safety Science, 159.

Zuboff, S (1988) *In the Age of the Smart Machine: the Future of Work and Power*, New York: Basic Books.