

Shaping the development and use of Artificial Intelligence: how human factors and ergonomics expertise can become more pertinent

Gudela Grote

To cite this article: Gudela Grote (2023) Shaping the development and use of Artificial Intelligence: how human factors and ergonomics expertise can become more pertinent, *Ergonomics*, 66:11, 1702-1710, DOI: [10.1080/00140139.2023.2278408](https://doi.org/10.1080/00140139.2023.2278408)

To link to this article: <https://doi.org/10.1080/00140139.2023.2278408>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 02 Jan 2024.



Submit your article to this journal [↗](#)



Article views: 820



View related articles [↗](#)



View Crossmark data [↗](#)

Shaping the development and use of Artificial Intelligence: how human factors and ergonomics expertise can become more pertinent

Gudela Grote

Management, Technology, and Economics, ETH Zürich, Zürich, Switzerland

ABSTRACT

New developments in Artificial Intelligence (AI) are extensively discussed in public media and scholarly publications. While in many academic disciplines debates on the challenges and opportunities of Artificial Intelligence (AI) and how to best address them have been launched, the human factors and ergonomics (HFE) community has been strangely quiet. I discuss three main areas in which HFE could and should significantly contribute to the socially and economically viable development and use of AI: decisions on automation versus augmentation of human work; alignment of control and accountability for AI outcomes; counteracting power imbalances among AI stakeholders. I then outline actions that the HFE community could undertake to improve their involvement in AI development and use, foremost translating ethical into design principles, strengthening the macro-turn in HFE, broadening the HFE design mindset, and taking advantage of new interdisciplinary research opportunities.

Practitioner summary: HFE expertise could and should significantly contribute to the socially and economically viable development and use of AI. Translating ethical into design principles, opening up to broader multi-stakeholder perspectives, and engaging in interdisciplinary collaboration within a design science framework are discussed as measures to achieve that.

ARTICLE HISTORY

Received 10 May 2023
Accepted 30 October 2023

KEYWORDS

Artificial Intelligence; HFE expertise; power; accountability; design science



Introduction

While in most disciplines the last few years have seen an intense scholarly debate on the challenges and opportunities of Artificial Intelligence (AI) and how to best address them, the human factors and ergonomics (HFE) community for the most part has been strangely quiet (Salmon, Carden, and Hancock 2021, Salmon et al. 2023). For instance, examining the articles published in the journal *Ergonomics* over the last few years, one finds few papers that discuss AI in relation to its emerging qualities as autonomous, self-learning systems, used in wide-ranging contexts from judicial decision-making to matching drivers to ride-hailing customers.

Even in Hancock's (2019) much discussed article on self-driving cars, the focus is more on classical issues to be addressed in any automated system, such as allocating functions between humans and technical systems and handling problems arising from humans becoming supervisory controllers, rather than on novel challenges resulting from the

inherent intransparency of systems using deep neural networks or the continuous adaptations of such systems during their use. Reasons for this reticence may lie, among others, in the deep-seated tendency of HFE researchers to be rather timid about making strong statements and recommendations on any emerging issue due to their concern for fulfilling highest standards in their methods and in the empirical evidence they accumulate (Dul et al. 2012; Hancock, Nourbakhsh, and Stewart 2019; Salmon et al. 2022). However, HFE perspectives and knowledge need to become a much larger part of the current discourse to help steer developments towards economically and socially viable use of AI.

In the following, I first outline specific challenges brought about by increasingly autonomous, self-learning systems. I also indicate how HFE could and should significantly contribute to addressing these challenges. Subsequently, I propose some measures the HFE community might take to become more relevant for

CONTACT Gudela Grote  ggrote@ethz.ch  Management, Technology, and Economics, ETH Zürich, Weinbergstrasse 56/58, 8092 Zürich, Switzerland.

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

shaping AI development and use, especially in light of its extensive expertise in managing risks of new technologies. By helping to address AI risks more effectively, HFE can and should play a significant role in using AI to find equitable solutions to many pressing societal problems, such as affordable healthcare and sustainable use of natural resources (Chui et al. 2018).

Main topics in the current debates on AI

Over the last few years, new developments in AI have been a topic in public media and in scholarly publications almost daily. AI-based systems that ‘generate outputs such as predictions, recommendations, or decisions (...) (and) operate with varying levels of autonomy’ (NIST. 2023, 1) have been proclaimed to significantly reduce the need for human labour even in highly skilled occupations (Frey and Osborne 2017), to shift control over workers from managers to algorithms (Kellogg, Valentine, and Christin 2020), and to revolutionise education (Kasneci et al. 2023). While many of these predictions may overrate the impact of AI, there is no doubt that AI has already shaped and will continue to shape how we live and work. Three topics appear particularly relevant for how our future with AI will unfold, which I will outline in the following to then also highlight the pertinent HFE knowledge for each. None of these topics is new as such, but in line with other authors I would argue that the growing ability of AI systems to autonomously learn and adapt adds new perspectives to these topics that need to be addressed in at least partially new ways (Berente et al. 2021; Salmon et al. 2023; Slota et al. 2023).

Will AI automate and/or augment human work?

Ever since Frey and Osborne (2017) published their prediction of up to 50% of jobs being lost due to emerging technologies including AI, the future of work has been high on the agenda both in the public and in academia. Subsequent analyses have shown that these predictions most likely are exaggerated, especially because technology usually only affects certain tasks within jobs and not whole jobs (Arntz, Gregory, and Zierahn 2016). The general tenet remains, though, that the current landscape of occupations will be dramatically changed (Parker and Grote 2022). In order to manage these changes well, a crucial question to answer is—as in all previous waves of technology development—whether AI is used to automate and/or augment human work. However, with the growing adaptive and learning capabilities of

AI systems, the range of tasks that could be automated has substantially increased and the possibilities for humans to interact with these systems has decreased due to the inherent opacity of self-learning systems (Castelvecchi 2016). These developments have led Hancock (2019, 483) to state that ‘the slow, steady and apparently inexorable extinction of any human contribution may be the natural sequella of automation myopia’.

These concerns notwithstanding, as Hancock himself and others have stressed, HFE expertise is dearly needed to help technology developers and organisations using these technologies make sensible decisions about the future role of humans in AI-supported environments (Bentley et al. 2021; Hancock 2019). In the management literature, the question of automation versus augmentation has only recently been taken up, often with little reference to the decades of HFE research on levels of automation and function allocation (e.g. Murray, Rhymer, and Sirmon 2021; Raisch and Krakowski 2021). Research on automation in HFE is often couched in generic terms that can be applied to any system with a high level of automation, thereby possibly not grabbing the attention of AI researchers (e.g. Chartered Institute of Ergonomics & Human Factors 2022). Moreover, HFE discussions of automation tend to focus on the immediate interaction between humans and technology in primary work processes. To render the tremendous HFE knowledge on how to best allocate functions between humans and technology more useful for the development and use of AI systems, multi-level frameworks are required to model and make decisions on higher-level augmentation of humans through potentially fully automated systems at lower levels. Methods such as EAST may be a starting point for such developments (Stanton and Harvey 2017).

Will AI become uncontrollable and with what effects for accountability?

Probably the most hotly debated topic concerns control and accountability for AI-based decisions and their outcomes and wider impact. The moratorium on generative AI development recently proposed by key technology developers themselves speaks to the intensity of these worries (Tracy 2023). Self-learning systems such as ChatGPT create the much discussed ‘black-box problem’ (Castelvecchi 2016) referring to the fact that these systems are opaque and unpredictable even for their developers as they autonomously change with any new data available. Thus, they are

the first technology that can endogenously adapt and improve through its use, raising the fundamental question of who is in control of these systems and who is to be held accountable for the consequences of their use (Salmon, Carden, and Hancock 2021; McLean et al. 2023). Moreover, the emerging capabilities of AI seemingly get these systems closer to what has been termed artificial general intelligence (AGI), that is highly autonomous systems that perform cognitive tasks as well as or better than humans. These developments may create qualitatively new forms of automation which challenge human control over their safe and ethical use in even more fundamental ways, as they supercede humans' intellectual capabilities (Salmon et al. 2023).

To date, from a legal standpoint it still has to be humans who are accountable. Bovens (2007, 447) has provided a widely accepted definition of accountability as the 'relationship between an actor and a forum, in which the actor has an obligation to explain and to justify his or her conduct, the forum can pose questions and pass judgement, and the actor may face consequences'. From this definition, it is clear that accountability should always be aligned with control, otherwise actors are held to account for actions they had no influence over. Control implies not only influence, but also transparency and predictability so that influence can be applied most efficiently and effectively. These considerations are at the root of intense efforts to render AI explainable (Kim and Doshi-Velez 2021; Rudin 2019).

Current discussions on what explainable AI entails and how transparent and predictable AI can and should be and for whom bring back memories of early work on mental models in HFE (Wilson and Rutherford 1989). Concepts proposed for explainable AI also echo other prominent HFE concepts, such as human-centred automation (Billings 1991), supervisory control (Sheridan 1987), and situation awareness (Endsley 1995). However, the relevant literatures are rarely considered and these principles seem to be reinvented in the plethora of publications on explainable AI. A laudable step towards changing this state of affairs is recent work by Suján, Pool, and Salmon (2022) which outlines relevant HFE principles for the design and use of AI in healthcare, but also points out that current HFE knowledge is not sufficient to answer all questions related to aligning control and accountability among the many stakeholders involved in developing and using advanced AI systems. New research efforts, in close cooperation with other relevant disciplines from computer science to ethics, will be needed to

substantiate regulatory requirements and practice recommendations for giving human actors sufficient control to also bear the responsibility for how AI is developed and used.

Will AI render the powerful even more powerful?

Emerging technologies always raise the question of whether and how power imbalances between employers and workers, between regulators and regulatees, or between business and public interests are affected. For instance, the internet and the decentralised forms of communication and coordination and the broad access to knowledge it permits have been considered a democratising force by many (Clegg, Courpasson, and Phillips 2006). Regarding AI, though, there appears to be a general apprehension that power imbalances will worsen. Algorithmic control of workers (Kellogg, Valentine, and Christin 2020; Möhlmann et al. 2021), the difficulties involved in independent oversight desired for effective AI governance (Shneiderman 2016), the growing power of organisations that own large data sets and the resources for training complex models on those data (Faraj, Pachidi, and Sayegh 2018), and financial pressures from venture capitalists on AI startups (Bessen, Impink, and Seamans 2023) are sources of this apprehension.

HFE with its general focus on rather narrow and specific applications of technology may seem to have not much to offer to alleviate these apprehensions. But in fact the contrary is the case. Especially the immense knowledge on participatory system design (e.g. Carroll 1996; Hignett, Wilson, and Morris 2005; Mumford 2000) and on shaping regulation for new technologies (Kirwan, Hale, and Hopkins 2002; Bieder and Bourrier 2013) could and should be brought to the table. HFE methods help to manage the complexities of goal conflicts among different stakeholders in pursuit of the goal to create the best possible systems for users (Waterson 2014). Compared to much of the work in the social sciences that aims at describing and explaining the impact of technology on work and organisations, the fundamental interest of HFE to design systems that foster human well-being and performance also helps to directly tackle power imbalances rather than just lamenting them (Norros 2014). This design orientation aligns well with current efforts to define principles for AI governance that are hoped to strengthen the position of weaker stakeholders such as end-users or regulatory bodies (e.g. NIST. 2023). Thus, it is essential to bring more HFE expertise to these efforts (Salmon et al. 2023).

How the HFE community may become more pertinent in shaping AI

Answering to the challenges raised by the possibility of automation beyond human intellectual capabilities, by widening gaps between human control and accountability in AI-supported work processes, and by growing power imbalances between different stakeholders involved in developing and using AI can rely on a wide range of HFE concepts, such as levels of automation, function allocation, mental models, situation awareness, human-centred automation, and participatory ergonomics, as outlined above. However, these concepts have had little impact on on-going AI developments to date (Salmon et al. 2023). HFE research and practice not keeping pace with technological development, insufficient involvement in companies' strategic decision-making, resistance to acknowledging the primacy of economic rationales for firms, and the complexity and impracticality of HFE methods have been mentioned as reasons for this lack of impact (Hancock, Nourbakhsh, and Stewart 2019; Norman and Euchner 2023; Waterson 2019). None of these reasons is new—as Emmenegger and Norman (2019, 513) have formulated it: 'The field of HF/E has great potential. The problem is that it has long had great potential: It is time to change'. The strong worries about the risks AI entails and the motivation to manage these risks well so that AI really becomes the solution to many societal problems as AI proponents proclaim (Chui et al. 2018) is hoped to create sufficient momentum for the HFE community to claim their seat at the many tables where the role of AI for our future is debated (Oswald et al. 2022). In the following, I outline a few concrete measures that may help to do just that.

Translate ethical into design principles

Ethics has been at the forefront of many discussions on AI, due to presumed imminent threats posed by AGI and the increasing public exposure to AI risks, from fake news to faulty chatbots. The main thrust for managing these risks has been directed to regulation, to date mostly in the form of recommendations rather than firm standards, such as the 'Ethics Guidelines for Trustworthy AI' published by the High Level Expert Group on AI (2019). For ethical principles to have an impact on actual development and use of AI, they need to be translated into design principles. Otherwise, they run the risk of turning into fig leaves for unabated techno-centric innovation. This creates a great opportunity for HFE. Accordingly, HFE researchers and practitioners should try their best to get into

the relevant committees at national and international levels. They could help specify what, for instance, 'human agency and oversight' (High Level Expert Group on AI 2019) entails in terms of concrete design choices regarding appropriate levels of automation, and what the resulting requirements for system reliability and transparency and for supervisory control are. Such detailed regulations would help to prevent loss of life linked, for instance, to Tesla's 'autopilot' (Siddiqui and Merrill 2023).

Generally, regulation has been considered the most powerful lever to influence the course AI developments will take (Hwang, Kesselheim, and Vokinger 2019; NIST. 2023; Shneiderman 2016). To effectively contribute to current regulation efforts, HFE experts may have to be more courageous and trust their own knowledge more than in the past. Committing type II errors by rejecting the evidence for new knowledge as presumed chance results can be worse than committing type I errors by unduly accepting it. The speed of technological development requires for all who want to have a say in it to be fast as well. Attempts to halt developments, such as the recent moratorium on generative AI, are bound to fail (Tracy 2023). However, speaking up to change the course of technological development is not necessarily easy either, as the power struggle between Tesla, the National Highway Traffic Safety Administration, and human factors professor Missy Cummings has shown (Ross 2023).

One specific debate where ethical and design principles need to be carefully balanced concerns whether learning algorithms should only be updated periodically, rather than being left to learn incessantly from new data, in order to facilitate explainability and regulatory oversight (Babic et al. 2019). 'Freezing' algorithms seems sensible so that users and regulators can be provided with sound information on which input data were used and in what ways to create the output users see. A moral dilemma may arise, though, because leaving an algorithm to continuously learn from new data may result in better decisions that in extreme cases might make the difference between life and death, as in acute medical care. Involving HFE experts in efforts to provide a regulatory framework for this crucial issue is essential as they can help to evaluate the risks for different options of dealing with autonomous adaptations of AI systems.

Strengthen the macro-turn in HFE

Given that AI is discussed at all levels from human-AI interaction to national and international AI governance,

it is paramount that the discipline of HFE broadens its remit and brings its extensive knowledge to bear on these discussions. HFE has much to offer to enrich concepts and methods for explainable AI, for decisions on automation and augmentation through AI, and for fostering stakeholder dialogue, to name a few examples. To not get involved more, is an enormous opportunity lost.

For decades, the HFE community has been discussing the necessity to address larger societal questions and to expand its theories, models, and methods to include larger work systems and processes reaching far beyond single individuals' interaction with technology (Bentley et al. 2021; Moray 1995; Thatcher et al. 2018). Accordingly, HFE should not be defined as a scientific field focused on the design of human-machine systems only (e.g. Dempsey, Wogalter, and Hancock 2000), but on the design of larger socio-technical systems within organisational and possibly societal contexts (e.g. IEA. 2000). To date, HFE knowledge nevertheless has been absorbed most easily in the development of specific human-technology interfaces compared to decisions on function allocation between humans and technology or on job and organisational design (Challenger, Clegg, and Shepherd 2013; Sauer, Sonderegger, and Schmutz 2020). Many authors have discussed this void and how to best fill it, urging to get HFE involved much earlier in technology development and in the strategic decisions that even precede any development efforts (Davis et al. 2014; Dul and Neumann 2009; Emmenegger and Norman 2019; Grote 2014).

For such a change to happen, Dul et al. (2012) have argued that better marketing and higher standards for HFE knowledge are needed. Additionally, Grote (2014) and Challenger, Clegg, and Shepherd (2013) have pointed to strategic opportunities for HFE by translating their concepts into the language of risk management as an overarching concern for all stakeholders involved in developing, using, and overseeing technology. By using Rasmussen's (1997) risk management framework for detailing impacts of socio-technical choices at multiple levels from technology to society, Brady and Naikar (2022) have illustrated the utility of such an approach for decisions related to human-automation collaboration in military aviation.

A more macro perspective that includes stakeholders at different levels of organisations and stakeholder interaction across organisations can cast a new light on human-centred automation and the alignment of human control and accountability as its core requirement (Boos et al. 2013). Beyond the well-known

problem that accountability often devolves to individual users, for instance the human occupants of a self-driving car, even if they have the least understanding of the system and little or no control over its functioning, misalignments for other stakeholders have to also be considered. For instance, organisations commissioning AI systems are usually accountable for providing high-quality training data, but they have insufficient knowledge of the complexities involved in validating, updating, and licencing AI-based technology to really understand the quality requirements (Hwang, Kesselheim, and Vokinger 2019). Technology developers, on the other hand, do not always realise the difficulties in providing good training data, especially for more complex applications such as medical diagnosis or hiring decisions, where adequate differentiations and inappropriate biases are not always easily distinguished (Teodorescu et al. 2021). HFE experts should be aware of these varying expectations regarding who can and should be in control, and who can and should be held accountable when problems arise in order to support appropriate alignment of control and accountability for all stakeholders across the AI life cycle.

By taking on such a larger perspective, HFE also becomes relevant for fields it to date has not played much of a role in, for instance personnel selection. With the advent of AI for automatically filtering job candidates, human resource managers have struggled to use the technology in ways that preclude biased decision-making, usually without the involvement of HFE experts (Tippins, Oswald, and McPhail 2021; van den Broek, Sergeeva, and Huysman 2021). By creating a bridge between human resource professionals and technology developers, HFE experts could have been and still can be very beneficial for translating both the intricacies of social perception and decision-making and requirements for building AI systems.

Broaden the HFE design mindset

Embracing early and fuller involvement in technology development, in the strategic decisions leading up to such development, and in the regulatory processes accompanying it requires for HFE researchers and practitioners to rethink their roles vis à vis different stakeholders and to develop broader design mindsets. The IEA. (2000) definition of HFE is a perfect starting point for this endeavour. It reaches well beyond the design of human-technology interaction for single systems by stressing the larger socio-technical and societal systems within which technology is developed

and used. It also points to the broad range of stakeholders from designers and users to international regulatory bodies that are relevant for influencing, taking and enacting decisions on how emerging technologies should be and are employed. Reflecting these roles in the specific context of AI and developing an action plan of how these roles can be filled by different actors in the HFE community would be a good starting point.

An HFE design mindset corresponding to these new roles has to combine the currently prevalent physical and cognitive ergonomics with organisational ergonomics. In order to get involved with the many stakeholders that influence how AI is developed and used, HFE experts should be ready to consider larger 'socio-technical landscapes' (Slota et al. 2023) rather than just socio-technical systems within organisations in their design considerations. Knowledge on participatory design is crucial in this respect, not only in terms of user involvement, but with respect to instigating stakeholder dialogue across professional, organisational, and institutional boundaries (Davis et al. 2014). Design thinking (Brown 2008) has become a popular concept in this regard that may inform new HFE approaches to participatory design (Norros 2014). HFE is needed not only for its knowledge of what it takes to design effective and safe socio-technical systems, but also of the social processes required for a design team to succeed in creating such systems. This dual role stemming from content and process knowledge has to be reflected in the mindset with which HFE experts approach design projects.

Take advantage of new interdisciplinary research opportunities

Current discussions on concepts such as automation versus augmentation and explainable AI will greatly benefit from systematically integrating existing HFE knowledge. However, HFE research itself has to advance to be able to answer new questions involved in aligning control and accountability for autonomously learning systems across multiple stakeholders and large power divides (Sujan, Pool, and Salmon 2022). As organisation and management sciences have started to acknowledge the necessity to become more design science oriented and foster knowledge on 'how things ought to be' (Simon 1996, 4) if they want to be relevant for decisions on AI development and use, this may open new doors for collaboration with HFE (Clegg et al. 2017; Goes 2014; Parker and Grote 2022).

One research domain for such new collaborations could be explainable AI. Many long-standing HFE concepts such as mental models, situation awareness, and trust in technology need to be updated in relation to AI (Sanneman and Shah 2022). Moreover, larger socio-technical systems and different application contexts have to be considered, especially with respect to explanation requirements for different stakeholder groups. Langer et al. (2021) list the many possible underlying desires for explanation, such as system acceptance, fairness, or privacy, reflecting the perspectives of technology developers, users, regulators, or people affected by the outcomes of AI-based decisions respectively. Just these few examples already show that developing explainable AI cannot be achieved by one discipline alone and Langer et al. (2021) describe in more detail what questions should be answered through interdisciplinary collaboration: for instance, how explanation leads to understanding and how this process may be different for different stakeholder groups due to their professional background or personality, or how explainability requirements should be combined to cater for the needs of different professional groups that interact with and through the AI system.

Hancock, Nourbakhsh, and Stewart (2019, 7690) talk about fluency required for different stakeholders to be able to even discuss requirements for AI system, which can be considered a meta-layer of explainable AI, as 'designers, fabricators, manufacturers, and vendors of these emerging systems need to explicate their products in a way that can be understood by legislators and the public alike'. This fluency is important for making sensible decisions about explainability of AI at the operational level, for instance regarding explainability-accuracy trade-offs. The assumption that more complex and thereby also more opaque systems are more accurate, has been challenged by pointing to cases where simpler models were just as accurate and by highlighting the business interests involved in selling complex models to customers who cannot verify accuracy claims (Rudin 2019). Accordingly, suggested approaches for explainable AI are not only to increase transparency about the data used to train models and about modelling choices, or to provide post hoc explanations of outcomes, but also to deliberately use simpler, more easily comprehensible models, (Kim and Doshi-Velez 2021). The principles for responsible algorithmic systems published by the European and US Technology Policy Committees of the ACM (2022, 5) state that 'the most desirable operational system setup is rarely the one with maximum accuracy',

emphasising yet again that defining and developing explainable AI is a multi-stakeholder interdisciplinary endeavour to which HFE has much to contribute.

Conclusion

My overall contention has been that to date the HFE community has not succeeded in bringing its wealth of knowledge on how to best design socio-technical systems to bear on the directions AI development has taken and will take. I would even argue that compared to discussion of other kinds of automated systems, the influence of HFE has been particularly insignificant in the domain of AI, where just about every discipline but HFE has been very verbose. In the burgeoning literature on AI, HFE design principles have been reinvented and sometimes distorted with little participation by HFE scholars themselves. To shape the future of AI technologies in economically and socially viable ways, HFE knowledge is desperately needed as is interdisciplinary collaboration within a design science framework to expand this knowledge. I hope that I have been able to contribute to ongoing discussions on how HFE knowledge may be rendered more powerful and collaboration with other scientific communities more fruitful to foster design-oriented research and practice for the effective and safe use of AI.

Ethics statement

No ethics approval was sought as the work is purely conceptual and did not involve any data collection.

Disclosure statement

No potential conflict of interest was reported by the author.

Funding

The author(s) reported there is no funding associated with the work featured in this article.

References

- Arntz, M., T. Gregory, and U. Zierahn. 2016. "The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis." OECD Social, Employment and Migration Working Papers, No. 189.
- Babic, Boris, Sara Gerke, Theodoros Evgeniou, and I Glenn Cohen. 2019. "Algorithms on Regulatory Lockdown in Medicine." *Science (New York, N.Y.)* 366 (6470): 1202–1204. doi:10.1126/science.aay9547.
- Bentley, T., N. Green, D. Tappin, and R. Haslam. 2021. "State of Science: The Future of Work—Ergonomics and Human Factors Contributions to the Field." *Ergonomics* 64 (4): 427–439. doi:10.1080/00140139.2020.1841308.
- Berente, N., B. Gu, J. Recker, and R. Santhanam. 2021. "Managing Artificial Intelligence." *MIS Quarterly* 45: 1433–1450.
- Bessen, J., S. M. Impink, and R. Seamans. 2023. "The Role of Ethical Principles in AI Startups." *SSRN Electronic Journal* <https://ssrn.com/abstract=4378280> or doi:10.2139/ssrn.4378280.
- Bieder, C., and M. Bourrier, Eds. 2013. *Trapping Safety into Rules: How Desirable and Avoidable is Proceduralization of Safety?* Farnham: Ashgate.
- Billings, C. E. 1991. "Human-Centered Aircraft Automation: A Concept and Guidelines." Retrieved from NASA, United States. <https://ntrs.nasa.gov/search.jsp?R=1991022821>.
- Boos, Daniel, Hannes Guenter, Gudela Grote, and Katharina Kinder. 2013. "Controllable Accountabilities. The Internet of Things and Its Challenges for Organisations." *Behaviour & Information Technology* 32 (5): 449–467. doi:10.1080/0144929X.2012.674157.
- Bovens, M. 2007. "Analysing and Assessing Accountability: A Conceptual Framework." *European Law Journal* 13 (4): 447–468. doi:10.1111/j.1468-0386.2007.00378.x.
- Brady, A., and N. Naikar. 2022. "Development of Rasmussen's Risk Management Framework for Analysing Multi-Level Sociotechnical Influences in the Design of Envisioned Work Systems." *Ergonomics* 65 (3): 485–518. doi:10.1080/00140139.2021.2005823.
- Brown, T. 2008. "Design Thinking." *Harvard Business Review* 86 (6): 84–92, 141.
- Carroll, J. M. 1996. "Encountering Others: Reciprocal Openings in Participatory Design and User-Centred Design." *Human-Computer Interaction* 11 (3): 285–290. doi:10.1207/s15327051hci1103_5.
- Castelvecchi, D. 2016. "The Blackbox of AI." *Nature* 538 (7623): 20–23. doi:10.1038/538020a.
- Challenger, R., C.W. Clegg, and C. Shepherd. 2013. "Function Allocation in Complex Systems: Reframing an Old Problem." *Ergonomics* 56 (7): 1051–1069. doi:10.1080/00140139.2013.790482.
- Chartered Institute of Ergonomics & Human Factors. 2022. "Human Factors in Highly Automated Systems." White Paper. <https://ergonomics.org.uk/resource/human-factors-in-highly-automated-systems-white-paper.html>
- Chui, M., M. Harryson, J. Manyika, R. Roberts, R. Chung, A. van Heteren, and P. Nel. 2018. *Notes from the AI frontier - Applying AI for Social Good*. McKinsey Global Institute.
- Clegg, S. R., D. Courpasson, and N. Phillips. 2006. *Power and Organizations. Foundations for Organizational Science*. London: Sage.
- Clegg, C. W., M. A. Robinson, M. C. Davis, L. E. Bolton, R. L. Pieniazek, and A. McKay. 2017. "Applying Organizational Psychology as a Design Science: A Method for Predicting Malfunctions in Socio-Technical Systems (PreMiSTS)." *Design Science* 3: e6. doi:10.1017/dsj.2017.4.
- Davis, M. C., R. Challenger, D.N. Jayewardene, and C.W. Clegg. 2014. "Advancing Socio-Technical Systems Thinking: A Call for Bravery." *Applied Ergonomics* 45 (2): 171–180. doi:10.1016/j.apergo.2013.02.009.
- Dempsey, Patrick G., Michael S. Wogalter, and Peter A. Hancock. 2000. "What's in a Name? Using Terms from Definitions to Examine the Fundamental Foundation of

- Human Factors and Ergonomics Science." *Theoretical Issues in Ergonomics Science* 1 (1): 3–10. doi:10.1080/146392200308426.
- Dul, J., R. Bruder, P. Buckle, P. Carayon, P. Falzon, W.S. Marras, J. R. Wilson, and B. van der Doelen. 2012. "A Strategy for Human Factors/Ergonomics: Developing the Discipline and Profession." *Ergonomics* 55 (4): 377–395. doi:10.1080/00140139.2012.661087.
- Dul, J., and P. W. Neumann. 2009. "Ergonomics Contributions to Company Strategies." *Applied Ergonomics* 40 (4): 745–752. doi:10.1016/j.apergo.2008.07.001.
- Emmenegger, C., and D. Norman. 2019. "The Challenges of Automation in the Automobile." *Ergonomics* 62 (4): 512–513. doi:10.1080/00140139.2019.1563336.
- Endsley, M. R. 1995. "Towards a Theory of Situation Awareness in Dynamic Systems." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37 (1): 32–64. doi:10.1518/001872095779049543.
- European and US Technology Policy Committees of the ACM. 2022. "Statement on Principles for Responsible Algorithmic Systems." <https://www.acm.org/binaries/content/assets/public-policy/final-joint-ai-statement-update.pdf>
- Faraj, S., S. Pachidi, and K. Sayegh. 2018. "Working and Organizing in the Age of the Learning Algorithm." *Information and Organization* 28 (1): 62–70. doi:10.1016/j.infoandorg.2018.02.005.
- Frey, C.B., and M.A. Osborne. 2017. "The Future of Employment: How Susceptible Are Jobs to Computerisation?" *Technological Forecasting and Social Change* 114: 254–280. doi:10.1016/j.techfore.2016.08.019.
- Goes, P. G. 2014. "Editor's Comments: Design Science Research in Top Information Systems Journals." *MIS Quarterly* 38 (1): III–VIII.
- Grote, G. 2014. "Adding a Strategic Edge to Human Factors/Ergonomics: Principles for the Management of Uncertainty as Cornerstones for System Design." *Applied Ergonomics* 45 (1): 33–39. doi:10.1016/j.apergo.2013.03.020.
- Hancock, P. A. 2019. "Some Pitfalls in the Promises of Automated and Autonomous Vehicles." *Ergonomics* 62 (4): 479–495. doi:10.1080/00140139.2018.1498136.
- Hancock, P. A., I. Nourbakhsh, and J. Stewart. 2019. "On the Future of Transportation in an Era of Automated and Autonomous Vehicles." *Proceedings of the National Academy of Sciences of the United States of America* 116 (16): 7684–7691. doi:10.1073/pnas.1805770115.
- High-Level Expert Group on Artificial Intelligence. 2019. *Ethics Guidelines for Trustworthy AI*. Brussels: European Commission.
- Hignett, S., J. R. Wilson, and W. Morris. 2005. "Finding Ergonomic Solutions—Participatory Approaches." *Occupational Medicine (Oxford, England)* 55 (3): 200–207. doi:10.1093/occmed/kqi084.
- Hwang, T. J., A. S. Kesselheim, and K. N. Vokinger. 2019. "Lifecycle Regulation of Artificial Intelligence- and Machine Learning-Based Software Devices in Medicine." *JAMA* 322 (23): 2285–2286. doi:10.1001/jama.2019.16842.
- IEA. 2000. "What is Ergonomics." <https://iea.cc/about/what-is-ergonomics/>
- Kasneći, Enkelejda, Kathrin Sessler, Stefan Kűchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Gűnnemann, Eyke Hűllermeier, Stephan Krusche, Gitta Kutyniok, Tilman Michaeli, Claudia Nerdel, Jűrgen Pfeffer, Oleksandra Poquet, Michael Sailer, Albrecht Schmidt, Tina Seidel, Matthias Stadler, Jochen Weller, Jochen Kuhn, and Gjergji Kasneći. 2023. "ChatGPT for Good? On Opportunities and Challenges for Large Language Models for Education." *Learning and Individual Differences* 103: 102274. doi:10.1016/j.lindif.2023.102274.
- Kellogg, K. C., M. Valentine, and A. Christin. 2020. "Algorithms at Work: The New Contested Terrain of Control." *Academy of Management Annals* 14 (1): 366–410. doi:10.5465/annals.2018.0174.
- Kim, B., and F. Doshi-Velez. 2021. "Machine Learning Techniques for Accountability." *AI Magazine* 42 (1): 47–52. doi:10.1002/j.2371-9621.2021.tb00010.x.
- Kirwan, B., A. R. Hale, and A. Hopkins, Eds. 2002. *Changing Regulation: Controlling Hazards in Society*. Oxford: Pergamon.
- Langer, M., D. Oster, T. Speith, H. Hermanns, L. Kűstner, E. Schmidt, A. Sasing, and K. Baum. 2021. "What Do We Want from Explainable Artificial Intelligence (XAI)? - A Stakeholder Perspective on XAI and a Conceptual Model Guiding Interdisciplinary XAI Research." *Artificial Intelligence* 296: 103473. doi:10.1016/j.artint.2021.103473.
- McLean, S., G. J. M. Read, J. Thompson, C. Baber, N. A. Stanton, and P. M. Salmon. 2023. "The Risks Associated with Artificial General Intelligence: A Systematic Review." *Journal of Experimental & Theoretical Artificial Intelligence* 35 (5): 649–663. doi:10.1080/0952813X.2021.1964003.
- Műhlmann, M., L. Zalmanson, O. Henfridsson, and R. W. Gregory. 2021. "Algorithmic Management of Work on Online Labor Platforms: When Matching Meets Control." *MIS Quarterly* 45 (4): 1999–2022. doi:10.25300/MISQ/2021/15333.
- Moray, N. 1995. "Ergonomics and the Global Problems of the Twenty-First Century." *Ergonomics* 38 (8): 1691–1707. doi:10.1080/00140139508925220.
- Mumford, E. 2000. "A Socio-Technical Approach to Systems Design." *Requirements Engineering* 5 (2): 125–133. doi:10.1007/PL00010345.
- Murray, A., J. Rhymer, and D. G. Sirmon. 2021. "Human and Technology: Forms of Conjoined Agency in Organizations." *Academy of Management Review* 46 (3): 552–571. doi:10.5465/amr.2019.0186.
- NIST. 2023. *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. doi:10.6028/NIST.AI.100-1.
- Norman, D., and J. Euchner. 2023. "Design for a Better World." *Research-Technology Management* 66 (3): 11–18. doi:10.1080/08956308.2023.2183015.
- Norros, L. 2014. "Developing Human Factors/Ergonomics as a Design Discipline." *Applied Ergonomics* 45 (1): 61–71. doi:10.1016/j.apergo.2013.04.024.
- Oswald, F. L., M. R. Endsley, J. Chen, E. K. Chiou, M. H. Draper, N. J. McNeese, and E. M. Roth. 2022. "The National Academies Board on Human-Systems Integration (BOHSI) Panel: Human-AI Teaming: Research Frontiers." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 66 (1): 130–134. doi:10.1177/1071181322661007.
- Parker, S. K., and G. Grote. 2022. "Automation, Algorithms, and beyond: Why Work Design Matters More than Ever in a Digital World." *Applied Psychology* 71 (4): 1171–1204. doi:10.1111/apps.12241.
- Raisch, S., and S. Krakowski. 2021. "Artificial Intelligence and Management: The Automation-Augmentation Paradox." *Academy of Management Review* 46 (1): 192–210. doi:10.5465/amr.2018.0072.

- Rasmussen, J. 1997. "Risk Management in a Dynamic Society: A Modelling Problem." *Safety Science* 27 (2–3): 183–213. doi:10.1016/S0925-7535(97)00052-0.
- Ross, P. E. 2023. "A Former Pilot on Why Autonomous Vehicles Are so risky - Five Questions for Missy Cummings." *IEEE Spectrum* 60 (6): 21–21, June. doi:10.1109/MSPEC.2023.10147081.
- Rudin, C. 2019. "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models instead." *Nature Machine Intelligence* 1 (5): 206–215. doi:10.1038/s42256-019-0048-x.
- Salmon, P. M., C. Baber, C. Burns, T. Carden, N. Cooke, M. Cummings, P. Hancock, S. McLean, G. J. M. Read, and N. A. Stanton. 2023. "Managing the Risks of Artificial General Intelligence: A Human Factors and Ergonomics Perspective." *Human Factors and Ergonomics in Manufacturing & Service Industries* 33 (5): 366–378. doi:10.1002/hfm.20996.
- Salmon, P. M., T. Carden, and P. Hancock. 2021. "Putting the Humanity into Inhuman Systems: How Human Factors and Ergonomics Can Be Used to Manage the Risks Associated with Artificial General Intelligence." *Human Factors and Ergonomics in Manufacturing & Service Industries* 31 (2): 223–236. doi:10.1002/hfm.20883.
- Salmon, P. M., G. M. Read, G. H. Walker, N. J. Stevens, A. Hulme, S. McLean, and N. A. Stanton. 2022. "Methodological Issues in Systems Human Factors and Ergonomics: Perspectives on the Research-Practice Gap, Reliability and Validity, and Prediction." *Human Factors and Ergonomics in Manufacturing & Service Industries* 32 (1): 6–19. doi:10.1002/hfm.20873.
- Sanneman, Lindsay, and Julie A. Shah. 2022. "The Situation Awareness Framework for Explainable AI (SAFE-AI) and Human Factors Considerations for XAI Systems." *International Journal of Human-Computer Interaction* 38 (18–20): 1772–1788. doi:10.1080/10447318.2022.2081282.
- Sauer, J., A. Sonderegger, and S. Schmutz. 2020. "Usability, User Experience and Accessibility: Towards an Integrative Model." *Ergonomics* 63 (10): 1207–1220. doi:10.1080/00140139.2020.1774080.
- Sheridan, T. B. 1987. "Supervisory Control." In *Handbook of Human Factors*, edited by G. Salvendy, 1243–1268. New York: Wiley.
- Shneiderman, B. 2016. "The Dangers of Fault, Biased, or Malicious Algorithms Requires Independent Oversight." *Proceedings of the National Academy of Sciences of the United States of America* 113 (48): 13538–13540. doi:10.1073/pnas.1618211113.
- Siddiqui, F., and J. B. Merrill. 2023. "17 Fatalities, 736 Crashes: The Shocking Toll of Tesla's Autopilot." *The Washington Post*, June 10.
- Simon, H. A. 1996. *The Sciences of the Artificial*. 3rd ed. Cambridge, MA: MIT Press.
- Slota, S. C., K. R. Fleischmann, S. Greenberg, N. Verma, B. Cummings, L. Li, and C. Shenefiel. 2023. "Many Hands Make Many Fingers to Point: Challenges in Creating Accountable AI." *Ai & Society* 38 (4): 1287–1299. doi:10.1007/s00146-021-01302-0.
- Stanton, N. A., and C. Harvey. 2017. "Beyond Human Error Taxonomies in Assessment of Risk in Sociotechnical Systems: A New Paradigm with the EAST 'Broken-Links' Approach." *Ergonomics* 60 (2): 221–233. doi:10.1080/00140139.2016.1232841.
- Sujan, M., R. Pool, and P. Salmon. 2022. "Eight Human Factors and Ergonomics Principles for Healthcare Artificial Intelligence." *BMJ Health & Care Informatics* 29 (1): e100516. doi:10.1136/bmjhci-2021-100516.
- Teodorescu, M. H. M., L. Morse, Y. Awwad, and G. C. Kane. 2021. "Failures of Fairness in Automation Require a Deeper Understanding of Human-ML Augmentation." *MIS Quarterly* 45 (3): 1483–1500. doi:10.25300/MISQ/2021/16535.
- Thatcher, A., P. Waterson, A. Todd, and N. Moray. 2018. "State of Science: Ergonomics and Global Issues." *Ergonomics* 61 (2): 197–213. doi:10.1080/00140139.2017.1398845.
- Tippins, N. T., F. L. Oswald, and S. M. McPhail. 2021. "Scientific, Legal, and Ethical Concerns about AI-Based Personnel Selection Tools: A Call to Action." *Personnel Assessment and Decisions* 7 (2): 1–22. doi:10.25035/pad.2021.02.001.
- Tracy, R. 2023. "Biden Administration Weighs Possible Rules for AI Tools like ChatGPT." *The Wall Street Journal*, April 11.
- van den Broek, E., A. Sergeeva, and M. Huysman. 2021. "When the Machine Meets the Expert: An Ethnography of Developing AI for Hiring." *MIS Quarterly* 45 (3): 1557–1580. doi:10.25300/MISQ/2021/16559.
- Waterson, P. 2014. "Health Information Technology and Sociotechnical Systems: A Progress Report on Recent Developments within the UK National Health Service (NHS)." *Applied Ergonomics* 45 (2): 150–161. doi:10.1016/j.apergo.2013.07.004.
- Waterson, P. 2019. "Autonomous Vehicles and Human Factors/ergonomics - A Challenge but Not a Threat." *Ergonomics* 62 (4): 509–511. doi:10.1080/00140139.2019.1563335.
- Wilson, J. R., and A. Rutherford. 1989. "Mental Models: Theory and Application in Human Factors." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 31 (6): 617–634. doi:10.1177/001872088903100601.