Ergonomics– rok 2023, ročník 66

Číslo 11

Automation, Artificial Intelligence and Robotics: Emerging issues and debates in Human Factors and Ergonomics – Part 1; Guest Editors: Gemma J. M. Read and Patrick Waterson

Editorial - Gemma J. M. Read & Patrick Waterson. *Automation, artificial intelligence and robotics: emerging issues and debates in human factors and ergonomics*. Pages: 1653-1655.

Mica R. Endsley. *Ironies of artificial intelligence*. Pages: 1656-1668.

ainbridge's *Ironies of Automation* was a prescient description of automation related challenges for human performance that have characterised much of the 40 years since its publication. Today a new wave of automation based on artificial intelligence (AI) is being introduced across a wide variety of domains and applications. Not only are Bainbridge's original warnings still pertinent for AI, but AI's very nature and focus on cognitive tasks has introduced many new challenges for people who interact with it. Five ironies of AI are presented including difficulties with understanding AI and forming adaptations, opaqueness in AI limitations and biases that can drive human decision biases, and difficulties in understanding the AI reliability, despite the fact that AI remains insufficiently intelligent for many of its intended applications. Future directions are provided to create more human-centered AI applications that can address these challenges. **Practitioner summary:** Artificial Intelligence (AI) creates many new challenges for human interaction. Five ironies of AI are discussed that limit its ultimate success, and future directions are provided to create more human-centered AI applications that can address these challenges.

- **Keywords:** Automation, artificial intelligence, human-centered AI, situation awareness, bias

Neelam Naikar, Ashleigh Brady, Glennn Moy & Hing-Wah Kwok. *Designing human-AI systems for complex settings: ideas from distributed, joint, and self-organising perspectives of sociotechnical systems and cognitive work analysis*. Pages: 1669-1694.

Real-world events like the COVID-19 pandemic and wildfires in Australia, Europe, and America remind us that the demands of complex operational settings are met by multiple, distributed teams interwoven with a large array of artefacts and networked technologies, including automation. Yet, current models of human-automation interaction, including those intended for human-machine teaming or collaboration, tend to be dyadic in nature, assuming individual humans interacting with individual machines. Given the opportunities and challenges of emerging artificial intelligence (AI) technologies, and the growing interest of many organisations in utilising these technologies in complex operations, we suggest turning to contemporary perspectives of sociotechnical systems for a way forward. We show how ideas of distributed cognition, joint cognitive systems, and self-organisation lead to specific concepts for designing human-AI systems, and propose that design frameworks informed by contemporary views of complex work performance are needed. We discuss cognitive work analysis as an example. **Practitioner Summary:** Emerging developments in AI will pose challenges for the design of human-machine systems. Contemporary perspectives of sociotechnical systems, namely distributed cognition, joint cognitive systems, and self-organisation, have design implications that are unaccommodated by traditional methods. Cognitive work analysis may provide a way forward.

Mark S. Young & Neville A. Stanton. *To automate or not to automate: advocating the 'cliff-edge' principle*. Pages: 1695-1701.

We reflect briefly on the last forty years or so of ergonomics and human factors research in automation, observing that many of the issues being discussed today are the same as all those decades ago. In this paper, we explicate one of the key arguments regarding the application of automation in complex safety-critical domains, which proposes restraining the capabilities of automation technology until it is able to fully and completely take over the task at hand. We call this the 'cliff-edge' principle of automation design. Instead, we espouse a use for the technology in a more problem-driven, human-centred way. These are not entirely new ideas and such a philosophy is already gaining traction in ergonomics and human factors. The point is that in a given system, tasks should be controlled either by human or by automation; anything in between only causes problems for system performance. **Practitioner summary:** Human factors problems with automation have been with us for over forty years, and have changed little in that time. This brief review shows a groundswell of opinion that points to what we call the cliff-edge automation principle – restraining the full capabilities of technology until it is ready to fully and completely take over the task. This approach improves human performance in the system by keeping the person in the loop and in control. Researchers and practitioners in ergonomics and human factors should continue to push this message to the designers and manufacturers of automated systems.

Gudela Grote. *Shaping the development and use of Artificial Intelligence: how human factors and ergonomics expertise can become more pertinent*. Pages: 1702-1710.

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.

- **Keywords:** Artificial Intelligence, HFE expertise, power, accountability, design science

P. A. Hancock. *Are humans still necessary?* Pages: 1711-1718.

Our long accepted and historically-persistent human narrative almost exclusively places us at the motivational centre of events. The wellspring of this anthropocentric fable arises from the unitary and bounded nature of personal consciousness. Such immediate conscious experience frames the heroic vision we have told to, and subsequently sold to ourselves. But need this centrality necessarily be a given? The following work challenges this, oft unquestioned, foundational assumption, especially in light of developments in automated, autonomous, and artificially-intelligent systems. For, in these latter technologies, human contributions are becoming ever more peripheral and arguably unnecessary. The removal of the human operator from the inner loops of momentary control has progressed to now an ever more remote function as some form of supervisory monitor. The natural progression of that line of evolution is the eventual excision of humans from access to any form of control loop at all. This may even include system maintenance and then, prospectively, even initial design. The present argument features a 'unit of analysis' provocation which explores the proposition that socially, and even ergonomically, the human individual no longer occupies priority or any degree of pre-eminent centrality. Rather, we are witnessing a transitional phase of development in which socio-technical collectives are evolving as the principle sources of what, may well be profoundly unhuman motivation. These developing proclivities occupy our landscape of technological innovations that daily act to magnify, rather than diminish, such progressive inhumanities. Where this leaves a science focused on work as a human-centred enterprise serves to occupy the culminating consideration of the present discourse. **Practitioners Summary:** Understanding the changes in discretionary, as compared to obligatory, roles of human users and operators in systems is central to Ergonomic practice. Envisioning this path of potential progress, and then witnessing and impacting its actual realisation, permits practitioners to optimise their professional and personal strategies as they deal with this next critical step in the relationship between humans and technology.

- **Keywords:** Automation, autonomous systems, human users, obligatory necessity

Jan Maarten Schraagen. *Responsible use of AI in military systems: prospects and challenges*. Pages: 1719-1729.

Artificial Intelligence (AI) holds great potential for the military domain but is also seen as prone to data bias and lacking transparency and explainability. In order to advance the trustworthiness of AI-enabled systems, a dynamic approach to the development, deployment and use of AI systems is required. This approach, when incorporating ethical principles such as lawfulness, traceability, reliability and bias mitigation, is called 'Responsible AI'. This article describes the challenges of using AI responsibly in the military domain from a human factors and ergonomics perspective. Many of the ironies of automation originally described by Bainbridge still apply in the field of AI, but there are also some unique challenges and requirements that need to be considered, such as a larger emphasis on ethical risk analyses and validation and verification up-front, as well as moral situation awareness during deployment and use of AI in military systems. **Practitioners Summary:** 'Responsible AI' is a relatively novel transdisciplinary field incorporating ethical principles in the development and use of AI in military systems. I describe the prospects and challenges with Responsible AI from a human factors and ergonomics perspective. There is in particular a need for new methods for testing and evaluation, validation and verification, explainability and transparency of AI, as well as for new ways of Human-AI Teaming.

- **Keywords:** Artificial Intelligence, military systems, ethics, human-machine teaming, explainability, transparency, testing and evaluation, validation and verification

Yueying Chu & Peng Liu. *Automation complacency on the road*. Pages: 1730-1749.

Given that automation complacency, a hitherto controversial concept, is already used to blame and punish human drivers in current accident investigations and courts, it is essential to map complacency research in driving automation and determine whether current research can support its legitimate usage in these practical fields. Here, we reviewed its *status quo* in the domain and conducted a thematic analysis. We then discussed five fundamental challenges that might undermine its scientific legitimation: conceptual confusion exists in whether it is an individual versus systems problem; uncertainties exist in current evidence of complacency; valid measures specific to complacency are lacking; short-term laboratory experiments cannot address the long-term nature of complacency and thus their findings may lack external validity; and no effective interventions directly target complacency prevention. The Human Factors/Ergonomics community has a responsibility to minimise its usage and defend human drivers who rely on automation that is far from perfect. **Practitioner summary:** Human drivers are accused of complacency and overreliance on driving automation in accident investigations and courts. Our review work shows that current academic research in the driving automation domain cannot support its legitimate usage in these practical fields. Its misuse will create a new form of consumer harms.

- **Keywords:** Traffic crash, driving automation, complacency, overreliance, responsibility attribution

S. McLean, B. J. King, J. Thompson, T. Carden, N. A. Stanton, C. Baber, G. J. M. Read & P. M. Salmon. *Forecasting emergent risks in advanced AI systems: an analysis of a future road transport management system*. Pages: 1750-1767.

Artificial Intelligence (AI) is being increasingly implemented within road transport systems worldwide. Next generation of AI, Artificial General Intelligence (AGI) is

imminent, and is anticipated to be more powerful than current AI. AGI systems will have a broad range of abilities and be able to perform multiple cognitive tasks akin to humans that will likely produce many expected benefits, but also potential risks. This study applied the EAST Broken Links approach to forecast the functioning of an AGI system tasked with managing a road transport system and identify potential risks. In total, 363 risks were identified that could have adverse impacts on the stated goals of safety, efficiency, environmental sustainability, and economic performance of the road system. Further, risks beyond the stated goals were identified; removal from human control, mismanaging public relations, and self-preservation. A diverse set of systemic controls will be required when designing, implementing, and operating future advanced technologies. **Practitioner summary:** This study demonstrated the utility of HFE methods for formally considering risks associated with the design, implementation, and operation of future technologies. This study has implications for AGI research, design, and development to ensure safe and ethical AGI implementation.

- **Keywords:** Artificial general intelligence, artificial intelligence, risk, safety, road transport, AI alignment

Teresa Zayas-Cabán, Rupa S. Valdez & Anita Samarth. *Automation in health care: the need for an ergonomics-based approach*. Pages: 1768-1781.

Healthcare quality and efficiency challenges degrade outcomes and burden multiple stakeholders. Workforce shortage, burnout, and complexity of workflows necessitate effective support for patients and providers. There is interest in employing automation, or the use of 'computer[s] [to] carry out… functions that the human operator would normally perform', in health care to improve delivery of services. However, unique aspects of health care require analysis of workflows across several domains and an understanding of the ways work system factors interact to shape those workflows. Ergonomics has identified key work system issues relevant to effective automation in other industries. Understanding these issues in health care can direct opportunities for the effective use of automation in health care. This article illustrates work system considerations using two example workflows; discusses how those considerations may inform solution design, implementation, and use; and provides future directions to advance the essential role of ergonomics in healthcare automation. **Practitioner Summary:** This article highlights the essential role of ergonomics in the effective design, implementation, and use of automation in health care. By discussing unique considerations for automation in health care and through two illustrative examples, we demonstrate the importance of an ergonomics approach for developing automated healthcare solutions.

- **Keywords:** Automation, health care, health information technology, workflow, work system

Sage Kelly, Sherrie-Anne Kaye, Katherine M. White & Oscar Oviedo-Trespalacios. *Clearing the way for participatory data stewardship in artificial intelligence development: a mixed methods approach*. Pages: 1782-1799.

Participatory data stewardship (PDS) empowers individuals to shape and govern their data via responsible collection and use. As artificial intelligence (AI) requires massive amounts of data, research must assess what factors predict consumers' willingness to provide their data to AI. This mixed-methods study applied the extended Technology Acceptance Model (TAM) with additional predictors of trust and subjective norms. Participants' data donation profile was also measured to assess the influence of individuals' social duty, understanding of the purpose and guilt. Participants ($N = 322$)

completed an experimental survey. Individuals were willing to provide data to AI via PDS when they believed it was their social duty, understood the purpose and trusted AI. However, the TAM may not be a complete model for assessing user willingness. This study establishes that individuals value the importance of trusting and comprehending the broader societal impact of AI when providing their data to AI. **Practitioner summary:** To build responsible and representative AI, individuals are needed to participate in data stewardship. The factors driving willingness to participate in such methods were studied via an online survey. Trust, social duty and understanding the purpose significantly predicted willingness to provide data to AI via participatory data stewardship.