Recent Topics in Human Factors
— Focusing on Safety and Human Error Research in Engineering Fields —

Hiroshi Shimoda (Graduate School of Energy Science, Kyoto University)

1. Introduction

Human Factors (also called Ergonomics) is an interdisciplinary field concerned with the fit between human cognitive and physical characteristics and artifacts, technology, organizations, and environments; in a broader sense, it can also be viewed as a perspective that pursues a better relationship between human society and science and technology. In recent years, the rapid development of technologies such as generative AI (Artificial Intelligence) and other AI technologies, collaborative robots, and digital twins has been reshaping the research landscape of human factors as well. A review paper co-authored by 20 leading human factors researchers [1] identifies six grand challenges: (1) the evolution of societal thinking, (2) the future of human work in Industry 5.0, (3) climate change and sustainability, (4) the future of education and training, (5) the future of personalized health, and (6) life, technology, and the metaverse. On the other hand, it has also been pointed out that human factors research has failed to keep pace with the speed of the AI revolution and that its presence is fading [2].

In the following, this article organizes recent research topics within the field of human factors, focusing in particular on research related to safety and human error in engineering fields.

2. AI-Based Detection, Prediction, and Mitigation of Human Error

In engineering fields where safety is paramount — such as nuclear power, aviation, and chemical plants — research is actively being conducted to model human error using AI and apply it to prediction and detection. A review paper by Gursel et al. [3] classifies related studies, from the perspective of how AI is used to understand, categorize, and mitigate human error, into four types of modeling: (1) descriptive, (2) predictive, (3) prescriptive, and (4) generative. Meanwhile, a review paper by Xiao et al. [4] discusses how research on Human Reliability Analysis (HRA), AI, and human cognitive models can contribute to mitigating human error going forward; it argues that combining cognitive models with AI-based risk information can improve the accuracy of human error prediction, while also pointing out the need for richer datasets, more transparent algorithms, and rigorous validation. In aviation as well, efforts are progressing to have AI analyze conversations between captains and first officers in the cockpit to detect early signs of fatigue and cognitive overload; rather than replacing human judgment with AI, such systems should be used strictly to support human decision-making [5].

3. Trust Between Humans and Automation/AI

As AI and automation are increasingly introduced into machine control, both overtrust and automation bias, as well as the non-use of automation due to undertrust, pose safety risks. For this reason, trust calibration — aligning the degree to which humans trust automation or AI with the actual reliability of that automation or AI — is extremely important and is being actively studied. A review paper by Mehrotra et al. [6] on research into human–AI collaboration identifies trust calibration as a central research theme going forward. As an example of research in this area, Wen et al. [7] propose a quantitative framework that models human-side trust as a function of automation reliability, showing that distrust is the safer default in high-risk situations. In addition, a simulation study on automated driving [8] shows that increasing explainability reduces trust error by 42.5%, significantly suppressing both overtrust and distrust. Related to this, research is also progressing on designing Explainable AI (XAI) tailored to human cognitive characteristics; it has been shown that providing explanations adapted to users' cognitive and emotional needs is effective in building trust and reducing trust bias [9]. Furthermore, with the spread of AI agents that autonomously carry out tasks, the design of human oversight authority and the timing of intervention has become a new point of discussion [10].

4. Safety and Resilience

In engineering fields where safety is critical — such as nuclear power, aviation, and chemical plants — attention is shifting from the traditional approach of investigating causes after accidents and devising countermeasures (Safety-I) toward an approach that analyzes and reinforces cases that went well (Safety-II). What matters here is the perspective of "resilience." Resilience engineering aims at safety management that assumes system variability, centered on four core capabilities: anticipation, monitoring, response, and learning [11]. For example, an empirical study conducted at petrochemical plant construction sites [12] shows that workers with higher cognitive resilience exhibit fewer unsafe behaviors.

Turning to organizational factors such as safety culture, WHO's Global Patient Safety Report 2024 [13] indicates that only 25% of countries overall have managed to cultivate a patient safety culture, and only 23% have adopted a human factors approach, suggesting that raising safety culture internationally remains an ongoing challenge. In the aviation sector, the International Civil Aviation Organization (ICAO) has identified inadequate ergonomic design, shift work, and punitive organizational cultures as causes of latent failures, and has presented organizational interventions based on human factors [14].

5. Human Reliability Analysis (HRA)

In the nuclear field, one study [15] analyzed the factors behind the psychological stress of nuclear power plant operators using Grounded Theory, a qualitative research method, and identified seven factors as sources of psychological stress: (1) interface management load, (2) task complexity, (3) load during alarm occurrences, (4) task load during transient/emergency conditions, (5) shift work load, (6) load when responding to errors, and (7) load during communication. Another study [16] statistically analyzed 137 human error reports that occurred between 2010 and 2017, constructing a causal model linking patterns of execution errors, diagnostic errors, and cognitive errors with contributing factors such as safety culture, training, procedures, human interface, competence, and mental state.

The U.S. Nuclear Regulatory Commission (NRC) is developing guidance on the effects of degraded digital instrumentation and control systems on operator performance [17], and expert opinions are also being gathered and examined regarding human reliability analysis for next-generation advanced reactors [18]. The OECD Nuclear Energy Agency (OECD/NEA) likewise designates human and organizational factors as a continuing priority area of activity [19].

6. Summary and Future Outlook

Taking an overview of the above, the current pillars of research related to safety and human error in engineering fields can be broadly organized into the following three areas.

(1) Trust and oversight in human–AI collaboration, and error detection
AI-based detection and prediction of human error, and the associated trust calibration and explainable-AI design methodology.

(2) The shift toward Safety-II and resilience
A shift in the safety-management paradigm from a preventive to an adaptive approach, driven by resilience engineering and safety-culture research.

(3) Advancement of human reliability analysis in high-reliability systems such as nuclear power and aviation
Understanding and reducing operator stress, fatigue, and cognitive load amid ongoing digitalization and automation.

The author is engaged in safety research in the engineering field and is not well versed in patient safety; however, I understand that the medical field has also, in recent years, been drawing on the achievements of safety research from engineering — for example, through macro-ergonomic approaches such as SEIPS (Systems Engineering Initiative for Patient

Safety) [20], which describes the three elements of work system, process, and outcome along with their feedback loops. At the same time, the usability of electronic health records (EHRs) and information overload are issues that are being particularly addressed in the medical field, where independent research is progressing. While engineering and medical fields differ greatly in the number, complexity, and underlying assumptions of the targets of safety, they also share no small number of commonalities. Going forward, it is hoped that research, development, and practice oriented toward safety will move beyond disciplinary boundaries.

References

[1] Karwowski, W., et al.: Grand challenges for human factors and ergonomics, Theoretical Issues in Ergonomics Science, Vol. 26, No. 4, pp.361-456 (2025). https://doi.org/10.1080/1463922X.2024.2431336

[2] de Winter, J.C.F., Eisma, Y.B.: Ergonomics and human factors: still fading—and why we need to embrace the AI revolution, Ergonomics, Vol. 69, No. 3, pp.418-424 (2025). https://doi.org/10.1080/00140139.2025.2509352

[3] Gursel, E., et al.: The role of AI in detecting and mitigating human errors in safety-critical industries: A review, Reliability Engineering & System Safety, Vol. 256 (2025). https://doi.org/10.1016/j.ress.2024.110682

[4] Xiao, X., et al.: A Comprehensive Review of Human Error in Risk-Informed Decision Making: Integrating Human Reliability Assessment, Artificial Intelligence, and Human Performance Models, arXiv (2025). http://www.arxiv.org/abs/2507.01017 (*Preprint, Not reviewed*)

[5] TechJournal.uk: Artificial Intelligence and Human Factors Unite to Strengthen Aerospace Safety, https://www.techjournal.uk/p/artificial-intelligence-and-human (Access on August 21st, 2026).

[6] Mehrotra, S., et al.: A Systematic Review on Fostering Appropriate Trust in Human-AI Interaction: Trends, Opportunities and Challenges, ACM Journal on Responsible Computing, Vol. 1, No. 4, pp.1-45 (2024). https://doi.org/10.1145/3696449

[7] Wen, H., Mounir, A.: Mathematical Modeling of Trust Calibration for Human–Automation Safety, WCX SAE World Congress Experience, Detroit, Michigan, United States, April 14 (2026). https://doi.org/10.4271/2026-01-0530

[8] Hani, S., et al.: When to Trust the Machine: A Simulation Framework for Human–AI Collaboration, Intelligent Human Systems Integration (IHSI2026), pp.362–372 (2026) https://doi.org/10.54941/ahfe1007095

[9] Myakala, P.K., Jonnalagadda, A.K., Bura, C.: The Human Factor in Explainable AI Frameworks for User Trust and Cognitive Alignment, International Advanced Research

Journal in Science, Engineering and Technology, Vol. 12, No. 1, pp.80-93 (2025). https://doi.org/10.17148/IARJSET.2025.12110

[10] Ugo-Enyinnah, C., Iwobi, I.: Between autonomy and oversight: Trust calibration and human controllability in agentic AI systems, Global Journal of Engineering and Technology Advances, Vol. 27, No. 3, pp.27-36 (2026). https://doi.org/10.30574/gjeta.2026.27.3.0137

[11] Valente, M., Ricci, F., Cozzani, V.: A systematic review of Resilience Engineering applications to Natech accidents in the chemical and process industry, Reliability Engineering & System Safety, Vol. 255 (2024). https://doi.org/10.1016/j.ress.2024.110670

[12] Larestani, A., et al.: Mitigating Cognitive Failures and Unsafe Acts Through Cognitive Resilience: A Study in the Petrochemical Construction Sector of Asaluyeh, Iran, Indian Journal of Occupational and Environmental Medicine, Vol, 29, No. 2, pp.160-166 (2025). https://doi.org/10.4103/ijoem.ijoem_138_24

[13] WHO: Progress on patient safety across health systems around the world, https://www.who.int/news/item/23-05-2025-progress-on-patient-safety-across-health-systems-around-the-world (accessed on August 21st, 2026).

[14] ICAO: Human factors in civil aviation safety oversight, Working Paper, A42-WP/455, https://www.icao.int/sites/default/files/Meetings/a42/Documents/WP/wp_455_en.pdf, (Accessed on August 21st, 2026).

[15] Wang, Y., et al.: Identifying NPP operator stressors using grounded theory. Scientific Reports, Vol. 15, 30618 (2025). https://doi.org/10.1038/s41598-025-16193-0

[16] Wang, Y., et al.: Human Error Analysis and Mitigation Strategies in NPPs, Social Science Research Network(SSRN), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5093930 (2025). (*Preprint, Not reviewed*)

[17] NRC: "Managing the Effects of Degraded Digital I&C Conditions on Operator Performance" (NUREG/CR-7264) (2025). https://www.nrc.gov/reading-rm/doc-collections/nuregs/contract/cr7264/index

[18] Arigi, A.M.: Expert Perceptions on Human Reliability Analysis for Advanced Reactors, European Safety and Reliability Conference (ESREL 2026), pp.492-499 (2026). https://doi.org/10.3850/ESREL2026061419_esrel26-p26592-cd

[19] OECD/NEA: Human and organisational factors, https://www.oecd-nea.org/jcms/pl_26323/human-and-organisational-factors (Accessed on August 21st, 2026).

[20] Carayon. P., et al.: SEIPS 3.0: Human-centered design of the patient journey for

patient safety, Applied Ergonomics, Vol. 84, 103033 (2020).
https://doi.org/10.1016/j.apergo.2019.103033