Developing Human Factors guidance for introducing Artificial Intelligence into the energy sector

Robert Becker & Stirling Tyler

User Centric Design Ltd

SUMMARY

The project team updated guidance for the Energy Institute on how the introduction of Artificial Intelligence (AI) and highly automated systems will affect Human and Organisational Factors (HOF) within the energy industry. The project team conducted a review of several papers on HOF and AI to develop new guidance specific to the energy sector.

The new guidance focuses on different areas of human performance that AI and highly automated systems will influence. It also details the steps that should be taken during design and implementation of highly automated systems and AI using check sheets and screening tools.

KEYWORDS

Artificial Intelligence, Trust, Ethics, Human-Machine Teaming, Human Performance

Introduction

The Energy Institute is the professional body for the energy sector worldwide. In 2024, they requested that their guidance document, "Guidance on human and organisational aspects of implementing new technologies," be updated.

It was requested that the update should focus on the human and organisational factors surrounding AI and highly automated systems, with information from existing publications such as the CIEHF's white paper *Human Factors in Highly Automated Systems* (McLeod, 2022) being used to inform the guidance updates. While this technology could augment human performance in the energy sector, they pose unique challenges that warrant new guidance.

Method

The project team reviewed the original documentation and noted areas for updating with content about AI and highly automated systems insights. The team then read and reviewed relevant publications on HOF, highly automated systems and AI (see references for a complete list). Insights gained from these papers were collated and considered from the perspective of people working in the energy sector.

The project team then conducted workshops with Energy Institute stakeholders to present research findings and proposed changes to be implemented into the guidance document. These workshops included international representatives from different energy companies who gave feedback on which topics they would like to see and what would be most beneficial.

Following the workshops, the project team updated the guidance document to include content about AI and highly automated systems and their influence on human and organisational factors within the energy industry. Alongside the updated guidance, the project team created check sheets and HOF specialist input screening tools for readers. The check sheets guide readers in considering specific human performance factors related to introducing AI or highly automated systems. The HOF specialist input screening tool can determine if HOF specialists are required during AI/highly automated system development and implementation.

Themes

The main topics captured in the new guidance are explained in this section.

Transparency

Transparency refers to the ability of AI and highly automated systems to explain how they work to operators to foster trust, improve Situational Awareness (SA) and allow for effective human-machine teaming.

Transparency is important for human operators to build accurate mental models of the system they are working with, reducing the risk of over-reliance and 'automation surprises' (Sarter et al., 1997) - where operators are confused or surprised by system behaviour due to it being incongruous with their mental model.

Succinctly explaining how these systems work to human operators poses a significant challenge. Systems such as deep neural networks, which mimic human brain functions, make it hard to explain and predict how a model will react to different stimuli over time - these are systems that learn and adapt their behaviour continuously, meaning that behaviour becomes less predictable as time goes on.

To tackle these challenges, highly automated systems and AI need to be designed in a way so that they offer easy-to-understand explanations of their behaviour and confidence in their outputs. Human Factors intervention during design and validation of these systems can help to ensure that the Human-Machine Interface (HMI) is transparent and usable - e.g., by including users in the design process and carrying out heuristic reviews and performance testing during system validation.

Employee Resistance

Organisations should consider the significant risk of potentially demotivating staff, which could lead to decreased work efficiency (e.g., due to boredom). Fears surrounding job security, changes to workload and requirements for additional training may spur employee resistance to implementing such systems.

It is important to clearly communicate the intent associated with implementing these systems and regularly collect and understand employee attitudes and perspectives. This can help organisations develop implementation and communications strategies that allay any undue employee concerns and encourage acceptance of AI or highly automated systems.

Trust & Confidence

Operators' trust and confidence in AI and highly automated systems are closely tied to transparency, as operators should ideally understand how a system works in order to invest the appropriate amount of confidence or trust in it.

For systems with inherent uncertainty, such as predictive models, operators need an appropriate level of confidence to avoid over-reliance or unnecessary scepticism. Over-trust can lead to complacency and over-reliance on a system, even when incorrect. Under-trust can lead to system underuse, increasing operator workload as they take on tasks the system was designed to handle.

Systems should be designed to provide clear explanations of their processes and decisions, thus prompting the operator to invest the appropriate amount of trust in the system output. For example, if a system responsible for load prediction explains to operators that it is using a training data set from several years ago, the operator may put less confidence in the predictions the system makes and adjust accordingly.

Fictional Scenario: *The latest AI Grid Load Prediction System ('GLoPS') forecasts that a city will consume 2670 kWh over the next 24 hours. It informs Stanley, the human operator, that this prediction is based on 2024 average energy statistics. However, Stanley knows a new events arena is opening the following day (which GLoPS has not accounted for). Rather than dismissing the AI's output entirely, Stanley adjusts the predicted energy load to reflect increased demand. While Stanley has not dismissed the system's output, he has not over-relied on it either - thus showing appropriate confidence in its output.*

Ethics

AI and highly automated systems can develop biases that affect how they process information and create outputs, negatively impacting operators and consumers. One example is algorithmic bias, where the underlying reasoning used by the system inherently disadvantages a specific group of people.

For instance, Obermeyer et al. (2019) found that an algorithm used in American healthcare prioritised white patients over black patients for additional care. The algorithm assumed that patients who spent more on healthcare were sicker and required more support. However, in the United States (US), those who access healthcare more frequently tend to have the financial means to do so. As economic status and race are closely linked in the US, the system, therefore, unintentionally favoured wealthier white patients over poorer black patients.

Similarly, in the energy industry, investment in energy infrastructure has often favoured affluent or urban areas. When AI models use historical performance data, such as outage frequencies or maintenance records, they may inadvertently perpetuate these balances.

Adopting guidance such as the EU Commission's "ethics guidelines for trustworthy AI" into the design of AI systems can help shape organisational policy to promote ethical AI usage. Where possible, these guidelines should be translated into design principles to ensure that AI systems operate ethically. For example, if an AI model is designed to explicitly label how it has arrived at a conclusion, including providing sources (Endsley, 2023) and allowing users to query its logic (Sujan et al., 2021) ensures that a degree of transparency has been embedded into the system at the design level.

Risk Reduction

Energy data and grid behaviour are constantly changing with the increasing use of renewable energy sources. Predictive models must be robust against model drift—a situation where outdated data leads to inaccurate predictions—which is especially dangerous when managing real-time energy distribution.

Unlike traditional automation, where outputs are consistent and predictable, highly automated or AI systems introduce an element of uncertainty (IxDF, 2024) by changing their outputs based on prior experience and contextual awareness. Users should be made aware of this through training and be instructed to exercise discretion when using these systems to ensure an appropriate level of trust is adopted (see Trust & Confidence).

User errors may stem from over-reliance on inaccurate system outputs and committing to inappropriate actions as a result. This, combined with skill fade and startle responses from long periods of underload, can result in delayed and impaired human reactions during system failures (see Failure Conditions).

Risk assessments must consider the transparency of the highly automated or AI system, i.e. how effectively it can communicate its data sources and logic to the user - standards such as IEEE 7001-2021 provide actionable guidance on introducing suitably transparent autonomous systems (IEEE, 2021).

The ongoing maintenance needs of the algorithm must also be considered, for example, periodic updates to an AI model's database, to ensure that it remains valid during its use in live operations and that the risk of model drift is mitigated.

Workload

Highly automated or AI systems can significantly enhance operator capabilities and responsibilities by performing tasks and subtasks that traditionally require extensive knowledge or experience. This effectively lowers the threshold for skill, experience and knowledge required to carry out a specific task. The immediate benefit seen here is the de-concentration of workload from specialist operators to the broader team, removing process bottlenecks and improving efficiency.

However, providing operators with new tasks, even if supplemented with automation or AI, will affect their overall workload and change their role profile. The system should be designed so that users understand how to use the technology properly, spot inaccuracy and bias, and know when to refer to a suitably qualified human operator for support.

Failing to address these needs may lead to increased stress and fatigue, as well as reduced job satisfaction and overall wellbeing. As mentioned in the Transparency section, the system's design must accommodate the user by providing a clear, intuitive interface that is unambiguous in its presentation of data and addresses key user needs for the task being conducted; this is best achieved by including the end-user in the design process.

Failure Conditions

If AI and highly automated systems are used to manage critical energy infrastructure, then sudden failure could lead to widespread disruption and outages. AI and highly automated systems in the energy sector should ideally be designed to degrade gracefully, maintaining partial functionality and allowing for a controlled shutdown rather than a sudden and complete failure. For example, if a server fails, the system can redistribute its load to other servers and suspend low-priority processes. This should be paired with a comprehensive warnings, cautions and alerts philosophy to ensure users are notified of reduced system performance, allowing them to intervene accordingly.

Furthermore, organisations should identify and substantiate human-based safety claims that capture what the operator is required to do in failure conditions and the extent to which their input will mitigate the risk posed by system failure. Risks associated with automating operator tasks include skill fade and under-load – these factors combined will reduce the speed and efficacy of human intervention in the event of system failure.

Methods such as Failure Modes and Effects Analysis (FMEA) could also be used to determine how automation or AI may fail, the consequences of such failure and what demands would, therefore, be made of human operators under such conditions.

Human-Machine Teaming

Highly automated systems and AI can be used to augment existing teams, increase efficiency, and reduce workload. This may motivate organisations to reduce team sizes and augment operator performance with highly automated systems or AI. This approach must be planned carefully – as it may perpetuate a negative perception of highly automated or AI systems as 'stealing' jobs and generating new tasks and workload drivers for existing staff (see Employee Resistance). Furthermore, new roles required for the ongoing maintenance of highly automated or AI systems may increase the size of existing teams to include specialist operators, e.g. AI specialists, data scientists and ethics experts.

AI is more dynamic, adaptive and flexible than traditional computing. In some instances, suitably advanced models may appear to fill the role of a team member rather than a tool or work aid. Depending on the maturity and intended use of the AI model, the impact on team composition and dynamics may range from providing task augmentation to fully realised human-machine teams.

The energy sector relies on interdisciplinary teams that include roles such as maintenance engineers, IT specialists, and control room operators. AI and highly automated systems must be designed to be intuitive across these disciplines, facilitating clear communication and ensuring that all team members can access and understand the system they are working with. To accurately account for the risk posed by such changes, consideration should be given to SA across a team. For a team to have 'good' SA, there needs to be consideration for how non-human elements of the team (i.e. AI) can communicate with human elements and vice-versa. Task analysis can be used to derive the SA needs of each team member.

For example, if human operators neglect to update an AI model with the latest weather data, it may fail to produce accurate energy production predictions for a wind farm. Failing to update the system with relevant data and the failure of the system to communicate its operating parameters and the data it is using may result in degraded productivity, team performance and safety.

Allocation of function and the oversight between human operators and the system should be clearly defined and signposted to all members of a human-machine team. In many cases, AI may act as a system to give a second opinion to human operators or augment task performance. While human operators will still retain overall control and responsibility for their tasks, introducing AI may augment performance, reduce task completion time, and boost productivity and safety.

Check sheets and HOF Specialist Input Screening Tool

The new guidance also includes check sheets and a screening tool for readers to use.

Check Sheets

The check sheets can be used as step-by-step guides for any organisation considering the introduction of AI or highly-automated systems. Each sheet includes a prompt (e.g. *Have the effects of highly automated or AI enabled systems on team composition been identified?*); a list of actions and considerations (e.g. *Identify new roles required for operating the system*), and a list of applicable tools or techniques to use (e.g. *Task Analysis; User Workshops)*. There are multiple sheets which cover the topics discussed in this paper.

HOF Specialist Input Screening Tool

The HOF Specialist Input Screening Tool acts as a decision-making framework to determine when specialised Human Factors expertise is required.

The tool consists of a guide word column (e.g. *System Design*); an accompanying question (e.g. *Does the system require the design of a user interface for interacting with automated and AI systems?*), a tickbox conditions column (*Applies/Does Not Apply/Not Sure*) and an implications column indicating what capabilities and benefits HOF professionals can provide (e.g. *Human factors professionals can help ensure interfaces are designed that are intuitive, user-friendly, and support efficient human performance*).

Conclusions

AI and highly automated systems have the potential to bring significant benefits to the energy sector by providing human operators with a powerful, adaptable tool to aid decision-making and automate existing tasks. This paper has highlighted significant HOF challenges that need to be addressed by human factors professionals during the design, validation and implementation of this technology.

A common theme to many of the challenges posed by AI and highly automated systems is the concept of transparency. If the highly automated system or AI cannot clearly explain how it has arrived at a conclusion, then all other issues highlighted in this paper are amplified. Without transparency, operators may not be able to invest appropriate confidence in the system (leading to over- or under- trust); fail to spot algorithmic bias; struggle to adapt to human-machine teams; resist using the technology or be unable to cope with the system in the event of a critical failure.

Established methods used by Human Factors professionals, such as FMEA; heuristic reviews and user-centred design philosophy can be used to ensure that AI and highly automated systems are designed to be transparent, useful and safe. These methods were highlighted in the check sheets and HOF Specialist Input Screening Tool, providing readers with guidance when implementing AI or highly automated systems in their organisation.

References

Endsley, M.R (2023) HFE perspectives on Artificial Intelligence and the future of work. Available at https://www.youtube.com/watch?v=Uu32ky_Z4yE. Accessed June 2024.

European Commission website. Ethics guidelines for trustworthy AI. Available at https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai. Accessed June 2024.

Flemisch, F., Heesen, M., Hesse, T., Kelsch, J., Schieben, A., & Beller, J. (2011). Towards a dynamic balance between humans and automation: authority, ability, responsibility and control in shared and cooperative control situations. Cognition, Technology & Work, 14(1), 3–18

Institute of Electrical and Electronics Engineers (IEEE) 7001-2021, IEEE Standard for Transparency of Autonomous Systems. Available at https://standards.ieee.org/ieee/7001/6929/. Accessed June 2024.

IxDF website. AI Challenges and How You Can Overcome Them: How to Design for Trust. Available at https://www.interaction-design.org/literature/article/ai-challenges-and-how-you-can-overcome-them-how-to-design-for-trust. Accessed June 2024.

McLeod, R. (2022) Human Factors in Highly Automated Systems. Chartered Institute of Ergonomics and Human Factors. Available at https://ergonomics.org.uk/resource/human-factors-in-highly-automated-systems-white-paper.html. Accessed June 2024.

Microsoft website. Artificial Intelligence (AI) vs. Machine Learning (ML). Available at https://azure.microsoft.com/en-gb/resources/cloud-computing-dictionary/artificial-intelligence-vs-machine-learning. Accessed June 2024.

National Health Service AI Lab website, Moving from trust to appropriate confidence. Available at https://digital-transformation.hee.nhs.uk/building-a-digital-workforce/dart-ed/horizon-scanning/understanding-healthcare-workers-confidence-in-ai/executive-summary-and-report-overview/moving-from-trust-to-appropriate-confidence. Accessed June 2024.

Nielsen Norman Group website. CARE: Structure for Crafting AI Prompts. Available at https://www.nngroup.com/articles/careful-prompts/. Accessed June 2024.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453.

Sarter, N. B., Woods, D. D., & Billings, C. (1997). Automation surprises. Handbook of Human Factors and Ergonomics. Vol. 2, 1926-1943.

Stanton, N. A., Stewart, R., Harris, D., Houghton, R. J., Baber, C., McMaster, R., Salmon, P., Hoyle, G., Walker, G., Young, M. S., Linsell, M., Dymott, R., & Green, D. (2006). Distributed situation awareness in dynamic systems: theoretical development and application of an ergonomics methodology. Ergonomics, 49(12–13), 1288–1311.

Sujan, M. (2021) Human Factors and Ergonomics in Healthcare AI. Available at https://ergonomics.org.uk/resource/human-factors-in-healthcare-ai.html. Accessed June 2024.