

SCSC Seminar: How Safety Culture has to Change with AI



The “How Safety Culture Has to Change with AI” seminar was held at the Hilton Euston hotel in London, on 19th June 2025 and hosted by Mike Parsons. The speakers presented their work and insights into the ways in which AI impacts existing safety culture, as well as methods to integrate AI and human expertise, address concerns around accountability and recognise the importance of human factors.

The six speakers presented interesting and lively sessions on development and frameworks for safe AI systems, covering a range of topics such as:

- Michael Wright (Wright Human Factors Ltd): **AI: Emergent Safety Culture Issues** – how AI impacts human factors, how accidents can be attributable to both human and AI failures
- Ben Fulford (BMT): **Merging Human Expertise with AI for Better Safety Management** – how can we make the output of generative AI more accurate and effective, integrating risks and mitigation strategies into our processes
- Paul Leach (RSSB) – **Human Factors Principles for Artificial Intelligence Systems in Rail** – a psychologist’s perspective on rail safety, how human factors can inform AI integration
- Adam Johns (Marsh McLennan) – **Redefining Accountability: Just Culture in the Age of AI** – maintaining psychological safety of workers, the cultural and organisational challenges of integrating AI into socio-technical systems
- Kathy Syfret (RAF) – **AI and Aircraft Maintenance** – AI’s role in aircraft maintenance, and the importance of understanding organisational and human factors
- Paul Traub (Paul Traub Associates) – **Automation Policy and Human Performance in Autonomous Weapon Application** – levels of autonomy, risks associated with AI and human control

Michael Wright – AI: Emergent Safety Culture Issues

Michael warned us all that there were currently more questions around AI integration and culture than there were answers, which proved to be a recurring theme throughout the day! Taking us through some preliminary definitions including automation vs AI, Michael provided a number of potential AI applications including both collaborative and cooperative systems. He noted that while humans perform badly at “monitoring roles”, this could be an area where AI could excel, and illustrated this with some examples of accidents resulting from human disengagement with monitoring.

Michael then moved on to discuss how accountability within safety culture can be affected by AI, providing us with some more enlightening illustrations of how over-reliance on AI can lead to serious consequences. Not only do we experience loss of ownership of decision making, but we risk de-skilling, loss of trust and the potential for some truly unsettling incorrect AI learning and adaptation. Michael concluded by showing us an AI generated essay on AI and safety culture from Microsoft Copilot, which was – he and the audience estimated – probably about 70% right. Certainly “pretty good”, but as Michael warned us, the difficulty is in validating which parts are in that 70%!



Ben Fulford – Merging Human Expertise with AI for Better Safety Management



Next up was Ben Fulford, discussing the risks on the opposite side of the question: the risks of not using AI. Ben showed us that AI adds most value where both the user’s job experience and the job complexity are either both high – for example in safety-critical systems assessment – or both low. He illustrated this with a compelling example describing how better outcomes from a System-Theoretic Process Analysis (STPA) were achieved by using a combination of experts and generative AI, as opposed to either working alone. However, he did caveat this with the observation that ChatGPT is not trying to tell its users the truth, it is trying to sound plausible. Definitely something to keep in mind there!

Ben also presented BMT’s work on a prompt creation framework to get the best out of generative AI, requiring that each query is provided with context, the relevant role the quest is applicable for, the question itself, the required format, the required tone and a template to show the generative AI what a good answer would look like. Ben presented the BMT Co-pilot tool which uses both the prompt creation framework and a knowledge graph to generate accurate and effective answers. He finished with a timely quote which sums up much of the work presented today: “it’s not about AI replacing expertise, it’s about merging expertise with AI”.

Paul Leach – Human Factors Principles for Artificial Intelligence Systems in Rail

Paul Leach then took over to provide us with a psychologist’s perspective on rail safety, serving as a very good illustration for the multi-disciplinary nature of safety engineering! Paul emphasised the need to identify human-centred design principles relevant to AI and consider the socio-technical aspects of human-AI interaction, noting that we should not jump headfirst into using AI technology just because it is “new and shiny”: a relevant reminder for us all.



Paul showed us some recent research on identifying a set of human factors principles that could be used to inform the design, implementation and operation of AI systems based on literature review, expert interview and repeated internal review and refinement. These principles included recognising that AI should support rather than sideline human strengths, working towards creating effective human-AI teams where the AI is a team-mate rather than a tool, and ground-up recognition of ethical challenges and potential mitigations. Along with this, Paul emphasised the importance of recognising “work as done” vs “work as imagined, or prescribed”. In other words, when we discuss AI integration into safety culture, we need to consider failures of the system rather than only concentrating on those situations where the system performs ideally!

Adam Johns -Redefining Accountability: Just Culture in the Age of AI



After lunch Adam Johns picked back up on the theme of “questions, not answers” by asking us all to define what we thought the primary purpose of a just culture was. Adam followed up on this by reminding us all of the importance of definitions, including – especially salutary given the day’s topic – the necessity of recognising that safety culture is about how to make trade-offs between safety and other competing goals. He emphasised that culture is about humans and that accountability should not be regarded as a

synonym for blame. This is particularly important when considering how to foster an effective safety culture even without the complexity of AI integration.

As with several of the other presenters, Adam had coaxed AI into generating a sample image: in this case a beautiful image of a tree as an ecological metaphor for aviation safety (although the less said about the words that the AI had provided to go along with the image, the better: a humorous and very realistic example of the importance of monitoring AI output!). The importance of language is a crucial point here: when we think of AI using environmental metaphors this opens our minds to new ways of describing both its risks and its benefits. Adam finished with the important reminder that we must fully understand a system before intervening in it, particularly where this intervention can make things both worse and better at the same time, as AI is particularly wont to do!

Kathy Syfret (RAF) – AI and Aircraft Maintenance



Kathy Syfret was next, bringing us right back to practical elements with a very enlightening look at how AI might be used for aircraft maintenance. Kathy emphasised the importance of understanding both organisational and human factors, particularly how culture is made up of a kaleidoscope of ideas, customs and social behaviour. Referencing the Haddon-Cave report³, which will no doubt be familiar to everyone reading, Kathy showed us how culture can be changed in order to improve safety, identifying that such a shift will be important to ensure successful integration of AI.

Kathy then moved on to provide us with some examples of how AI might be used in practice for aircraft maintenance, looking at predictive maintenance and reliability-centred maintenance as particular areas where AI could be integrated. Supply chain management was another, which she noted tended to be overlooked in favour of communications: while it's certainly true that "if you have no comms then you have no bombs", the same could be said if you indeed *have* no bombs due to a supply chain failure! Kathy also referenced the importance of adaptive safety culture when integrating AI, and the need to address both complexity of organisations and complacency in our existing processes.

Paul Traub – Automation Policy and Human Performance in Autonomous Weapons Application

Paul Traub demonstrated technology in action by dialling in to give the final talk of the session, returning to the question which had cropped up again and again during the day: is AI the panacea for human performance limitations? The answer, he warned us, is "not necessarily": there are risks from automation too. For example, AI usage has been linked to loss of situational awareness, where the operators of a system do not understand the extent of automation in that system, nor how best to manipulate this. As automation occurs, Paul demonstrated, the operator becomes more "out of the loop", thus further increasing the chances for error.

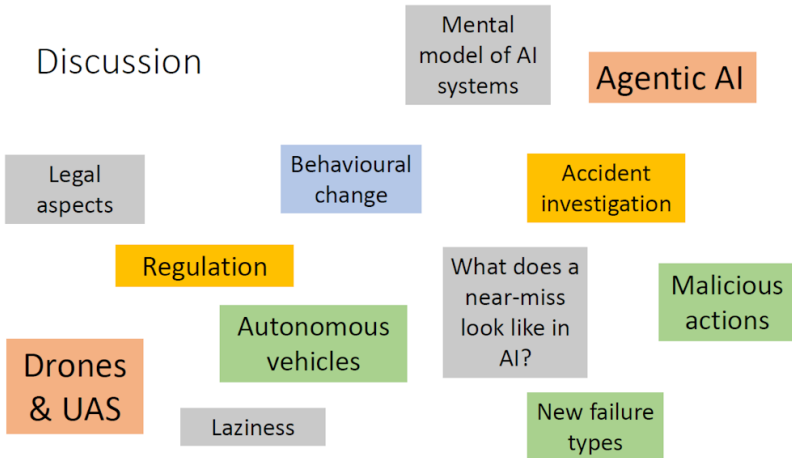


Paul next discussed how a robust automation policy can reduce risks of automation, and that this should start at the ground up with assessing how automation is to be integrated into design. He illustrated how the implementation of automation should include a task and error analysis, and emphasised the importance of identifying and understanding a system's critical tasks and potential points of human failure. He finished with another important reminder that problems in AI must not just be simply reported, but that the report must be acted on: this is not always something which is sufficiently highlighted in safety culture, but is going to become increasingly important as more systems integrate AI.

³ "The Nimrod Review: an independent review into the broader issues surrounding the loss of the RAF Nimrod MR2 aircraft XV230 in Afghanistan in 2006 report", See <https://www.gov.uk/government/publications/the-nimrod-review>

Mike Parsons – Wrap Up

Mike Parsons brought the threads of the day together in a final general discussion, noting some important areas we hadn't yet had time to discuss. These were wide-ranging, including regulation, drones, autonomous vehicles, malicious action and what a near-miss might look like in AI. Other participants also raised concerns such as the maturity of AI: is it really mature enough to be used in safety-critical systems? Audience members also contributed, highlighting the "human" aspects of emotional resistance to AI and the importance of avoiding anthropomorphism when we use AI, a timely reminder to everyone who has ever become irritated by ChatGPT's hallucinations!



The audience were in general agreement that AI integration into safety-critical systems is as much a human question as it is a technological question. Safety culture is, inevitably, a by-product of human culture, and safe AI integration will require an adaptation of the ways we think, interact and relate to systems. All in all, this was a very interesting and lively day, with some fantastic insights into AI frameworks and the challenges that must be overcome.

Report by Catherine Menon.

Catherine Menon is a principal lecturer at the University of Hertfordshire. Her research focuses on the intersection of trust, safety and ethics in public-facing autonomous systems. She has a PhD in category theory and has been involved in the development of regulatory standards and guidance for safety, security and ethics assurance within autonomous systems.