Seminar: How to Make the Most of AI: A Workshop for Safety Engineers



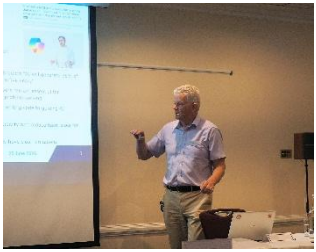
The "How to Make the Most of AI: A Workshop for Safety Engineers" seminar was held at the Hilton Euston hotel in London, on 25th June 2026[4] and was hosted by Mike Parsons. The speakers discussed a range of techniques and insights into how safety engineers might benefit from the use of generative Artificial Intelligence (AI) / Large Language Models (LLMs), examining everything from prompt engineering to building applications and custom tools from AI code generators.

The four speakers presented interesting and lively sessions, all of which generated great discussion. (Note that AI is used broadly to include all aspects such as LLMs and Machine Learning). Topics included:

- Alan Simpson (Ebeni): **Beyond the Hype: LLMs in Safety-Critical Engineering - Principles, pitfalls and the changing role of the engineer** – what AI will mean for safety engineers, how we can ensure that we continue to use AI safely and appropriately.
- Rudi (Redford-Brown) Hennessy (NHS England) **Prompting LLM to support hazard assessment** – the use of AI in safety-critical engineering within the healthcare sector, expectations and consequences
- Paul Hampton (Consultant and SCSC): **The Unreliable Guru** – using AI to rapidly generate code exploring some real-world examples
- Ben Fulford – (BMT): **Putting AI to Work: From Better Prompts to Building Your Own Tools** – practical techniques for using AI, how to get started, prompt engineering and tool building

[4] This was an incredibly hot day and we were very grateful to all those who attended in person!

Alan Simpson – Beyond the Hype: LLMs in Safety-Critical Engineering - Principles, pitfalls and the changing role of the engineer



Alan began by defusing some of the more extreme mainstream fears: namely, reassuring the audience that AI is not going to take over the jobs of safety engineers! The audience contributed by providing use cases for how safety engineers are already using AI, and Alan raised some important points about these, including the potential for untrustworthy outputs. He pointed out that the safety engineering challenge is how to justify with sufficient confidence, that the combination of AI tool, human oversight and process controls is appropriate for the consequences of being wrong.

AI tools, Alan reminded us, can fail in very different ways to traditional tools, and the real challenge is in justifying the use of the AI tools with sufficient confidence for the environment of use. He warned us of some of the risks of generative AI, including hallucination risk, context misunderstanding, risk of over-reliance / de-skilling and the potential for malicious compromise. He also explored the trust paradox: the greatest risk from AI may be the gradual erosion of scrutiny which we apply to its answers. Alan wrapped up with a discussion of how we might reframe the question using standards and the outputs from the SCSC AI Working Group (https://scsc.uk/saias), building on ISO 42001. The takeaway message was that AI – when it comes to safety engineering – is not wholly good, bad (or even ugly!) It is a tool that offers significant benefit but must be used only when we have sufficient confidence and sound engineering judgement in its capability.

Rudi (Redford-Brown) Hennessy (NHS England) Prompting LLM to support hazard assessment – Practical Gains, Real Risks



Next up was Rudi, who took us through an introduction of what AI could usefully be used for within healthcare, including drafting, summarising, research, process automation and ideation, or helping us brainstorm. Within safety specifically, she identified AI to be genuinely useful in hazard log drafting, standards research and incident analysis – all very complex fields, where many safety engineers would appreciate AI serving as a "first pass"

Rudi next took us through a practical workflow for AI assisted hazard identification, which prompted much audience interest and a number of questions. Many of these were answered by her next topic, a discussion of the AI risks to consider and precisely what these might mean in terms of healthcare. This was reinforced by her emphasis that AI is a supplementary tool and its primary value lies in performing the more routine tasks and "giving us time to think". AI, she emphasised, is not a replacement for clinical safety expertise, but it can help healthcare safety engineers to focus their efforts on the areas that most need human judgement and assessment.

Paul Hampton (Consultant and SCSC): The Unreliable Guru

Paul considered the use of AI specifically for generating code and software engineering and asked us: can AI be the software "guru", consolidating creativity, art, and effective use of language to produce code? He let us into the secret that much of the technical entertainment material for SSS'25 and SSS'26 was produced with human-AI complementary effort. The example quiz question requiring the audience to identify the safety concept referred to in an AI-remixed version of ABBA's Waterloo proved very popular!



Paul also identified AI was very helpful in checking code that he had written, noting that when he asked it questions about whether code might fail in one particular way (failing to exit) not only did it confirm this, but identified other concerns. This was, he considered, an example of Guru-level coding in terms of the suggestions that it provided. Nevertheless, the "nearly there" rabbit hole – when you get almost what you want, but not quite! – remains a significant concern. AI, he reminded us, can often produce answers that are almost correct, but crucially, not quite. A better classification of it might be that it is an "unreliable guru", in that it is functional for well-defined problems but not for open problems. He then demonstrated the power of AI Integrated Development Environments by showing how Anysphere's Cursor application generated a fully functioning Boeing 737-800 Take-off Speed Calculator within 25 minutes. He finished with some salutary reminders that AI might be better, faster and more accurate than human developers but – again echoing the theme of the day – could not be used as a replacement in safety-critical systems engineering (yet!) Ultimately, accountability must remain with human engineers, who must understand, justify and take ownership of every AI-generated component before it is incorporated into a deployed system.

Ben Fulford - Putting AI to Work: From Better Prompts to Building Your Own Tools – Getting More from AI

The last of our talks was from Ben Fulford, presenting a practical route to AI integration for safety engineers including prompt engineering and tool building. He set out the four rungs of the ladder: prompting well, context engineering, orchestration, pay-off and then recapped the BMT prompt framework for effective LLM prompting. He emphasised the importance of pressure-testing the output, and also of providing context: give the AI your standards, your safety case and your hazard log. This provides the necessary *context engineering*, the second rung on the ladder.



Moving onto the third rung – orchestration – Ben invited us to consider this as creating a board of advisors who disagree. This involves running several models, each given a different job, and intended to challenge each other.

This should involve human-in-the-loop contribution, but this does not necessarily have to mean a human reviewing the AI code! Instead, Ben suggested that we consider the use (by a human) of a second model to perform the verification, checking generated code against the specification. Finally, rung 4 – the payoff – where Ben illustrated this by reference to a narrative game he's been involved in creating, where the safety engineer is the hero! Certainly, an attractive thought, at least judging by the audience reaction. Ben wrapped up with some concrete and practical suggestions for steps we might take to start using the prompting framework and techniques he had introduced.

Workshops

The afternoon saw interactive workshops, with Alan, Rudi, Paul and Ben all leading sessions on how to incorporate and begin using AI. The first set of workshops involved the topics described below:

Example hazard analysis of a fictitious Autonomous Rail Inspection Drone



This workshop kicked off with an interesting use of ChatGPT: one participant had copied the seminar description into ChatGPT to ask what might be a valuable output from the task. Immediately the participants realised it had folded much of the written information about the workshops together – an example of why human insight is vital!

Working together with the AI, the participants were able to unpick the task and begin creating the hazard analysis by using their own experience of similar artefacts. Questions raised were how to define hazards to the AI, and how to establish the level of hazard and the depth of the analysis required. The AI-human collaboration performed well, with the team identifying 45 gaps in the system description that it was given.

2 out of 3 Voting Logic Module with Dynamic Sensor Health Tracking and Rate-of-Change Limiting



This workshop dived right into addressing the problem, with the facilitator demonstrating how quickly AI can generate a fully functioning 2 out of 3 Voting Logic Module from a one-page requirements specification. Participants explored hazards in depth, including those related to the rate of change – and the requirement for the code to flag significant changes in a sensor's value as indicative of an error – as well as considering erroneous rates of change, and the requirement for the AI to monitor this. One issue raised was that due to the need to manage the context window size, every new AI agent was required to begin from first principles: it had no knowledge of previous agent chats, and therefore couldn't use previous context (which has both advantages and disadvantages!) The AI was reasonably effective at assessing the adequacy of its code, but appeared to require significant prompting to refine the solution.

GUARDIAN: a preliminary hazard list for a safety programme



This group began similarly to the others, by feeding the workshop topic into AI and asking it to produce hazard examples for a draft system. Participants also discussed the potential for AI use in this example to have associated hazards including data leakage and privacy compromise.

An excellent example of how human expertise is essential for collaboration! One point that was particularly noted was the benefit of having (human) coding expertise within the room. Participants also began to identify a preliminary hazard list with the help of AI, and multiple participants identified that the use of two models was beneficial so that the same task could be run through both to provide assurance. The team also expressly asked the AI to act as a red team to challenge and interrogate the output from the team.

Hazard examples for Ambient voice technology / AI discharge summaries



This group focused on the healthcare domain, and immediately noted from their first prompt – asking AI for hazard examples – that the system clearly lacked system context.

Human context was also identified as significant, including the relevance of patient behaviour. Participants then assessed the adequacy of an AI-produced hazard log (after identifying that this had originated in an NHS England hazard log!)

Participants also found that when they provided context, including relevant standards, it produced a more comprehensive and defensible hazard log without the need for amending or refining the prompt. Participants also experimented with providing the AI with the specific system under investigation, and identified that prompt engineering was of significant benefit with more context provided.

Panel discussion and wrap up

The workshops concluded with a panel discussion with the speakers, and audience questions. The topics were wide ranging, including more case studies of how people were using AI or potential future use cases (the dreaded "thousand-word email" example came up here!) The conclusion overall was that AI and humans make great collaborative partners but that for safety critical work the AI should always be that: a partner and not a leader.

Report by Catherine Menon