AI Regulation: Managing risk and safety in engineering

A BCS insight paper in partnership with the National Engineering Policy Centre.

On behalf of the UK Government's Office for AI, BCS, the Chartered Institute for IT (BCS) worked with the National Engineering Policy Centre to bring together experts in AI from across the UK engineering sector. This roundtable discussion focused on the principles of safety and risk and how the regulatory environment outlined in the UK Government's White Paper could work in practice and where it might need to be amended. This was the second of two roundtables BCS convened in response to the AI White Paper Consultation. This document builds on BCS's existing positioning and insights on AI including the latest paper: 'Helping AI Grow Up Without Pressing Pause' and our call for the UK to "lead the way in setting professional and technical standards in AI roles, supported by a robust code of conduct, international collaboration and fully resourced regulation.

Key points:

- Premature AI deployment without human oversight is of critical concern – Iterative, virtuous cycles of feedback with sectors and including communities are needed to inform public policy development and implementation
- Certification of engineers and developers will build public trust and transparency in AI operations
- At-scale, cross-sector education on AI and emerging technology is needed for effective AI adoption
- "There's an obsession with implementing autonomy without understanding it" - poor understanding of AI at senior levels across public policy and governance is a significant risk factor in making decisions critical to public welfare
- AI safety and risk should be defined across different contexts
- Preventative guidelines should be informed by an assessment of sector-related critical risks, including societal and technological aspects and worst-case scenarios
- AI system application in contexts it is not trained for can be dangerous,

understanding limitations and potential biases is essential to safety
- Learn from failure with iterative practice guidance on safety and robustness principles curated with industry and professional bodies

The roundtable involved representatives from various organisations, and discussions revolved around the principles of the AI White Paper and their impact. Fairness, inclusivity, transparency, and accountability were necessary for AI governance and regulation.

Roundtable attendees:

- Adam Leon Smith, Chair of the BCS Fellows Technical Advisory Group (FTAG)
- Dr Natasha McCarthy, Associate Director, National Engineering Policy Centre at the Royal Academy of Engineering
- Rashik Parmar, CEO at BCS, The Chartered Institute for IT
- Professor Alan Bundy, School of Informatics at the University of Edinburgh
- Professor Ibrahim Habli, Deputy Head of the Computer Science Department at the University of York
- Dr Chris Elliott, Systems Engineer at Engineering X
- Dr Matthew Forshaw, Reader in Data Science at Newcastle University
- Gordon Meadow, CEO at SeaBot Maritime
- Prof Austin Tate, Professor of Knowledge-Based Systems at the University of Edinburgh
- Dr Caitlin Bentley, Lecturer in AI Education at Kings College London
- Dr Robert Merrall, Independent Consultant
- Dr Carolyn Ten Holter, Research Responsible Technology Institute - University of Oxford
- Dr Anthony Cohn, Professor of Automated Reasoning at the University of Leeds
- Gabriella Commatteo, Senior Policy Advisor at the Office for AI
- Rebecca Anselmetti, Head of International AI Policy and Tools, AI Regulation at the Office for AI
- Professor James Davenport, Hebron & Medlock Professor of Information Technology at the University of Bath
- Andrew Chadwick, Technology and Innovation Lead - Aviation at Connected Spaces Catapult

Insights:

Participants were asked to reflect on the White Paper's principles and their impact on the organisations they represent. Below is a synthesis of the main points from the panel.

Question 1: What does AI safety mean in the context of your use?

According to Professor Austin Tate from the School of Informatics, University of Edinburgh people want to run before they can walk regarding AI implementation, and the desire to use and deploy autonomous systems without human oversight is a cause for concern. He argued that there's a need for 'bigger reflection in policy before it goes down to the individual sector regulators', particularly regarding putting humans in the commanding position concerning AI decision-making. Dr Chris Elliott, Systems Engineer at Engineering X, concurred giving examples of how trusting autonomous systems prematurely can lead to harmful inefficiency

in medicine.

Professor Ibrahim Habli who is Deputy Head of the Computer Science Department at the University of York encouraged participants to take a step back and ask what is meant by AI Safety. He said safety is not absolute, and immediate and long–term risks must be considered before implementing any safety measures. There was a consensus across the panel that the question of safety needed greater specificity and inquiry.

Rashik Parmar, CEO of BCS, said that there are many ways to interpret safety:

- machine-machine,
- machine/human,
- human/human,
- organisation/organisation

Each of those changes the context of safety and the impact of implementing AI, he said, adding ethics, accountability, and inclusiveness must also be considered less abstractly.

In response to comments made above, Adam Leon Smith, Chair of BCS FTAG said the focus should be on regulating AI systems in the context they're used, rather than AI itself. He said AI shouldn't be relied upon in a safety-critical context unless it's making the overall system safer in its own right. He also raised concerns about the sufficient human oversight needed to prevent harm.

When asked to think about creating a skilled workforce equipped to deal with these challenges, Gordon Meadow, CEO at SeaBot Maritime, suggested having a set standard of use cases – a description of how a user interacts with a system or product – outlining the competencies required and delineating accountability.

Dr Robert Merrall, an independent consultant who works with Innovate UK, referenced the agricultural sector, which is currently experiencing a significant labour shortage, resulting in calls for the rapid implementation of automation. He added what seems to be driving safety measures for the 'cobots' – collaborative robots – is what the insurance industry will accept rather than any objective measure.

Rounding off this part of the session, Professor Habli urged the panellists not to focus too heavily on regulation: having a certificate doesn't always translate to competence.

Question 2: What key risks should be addressed, and how can these be measured and mitigated?

The panellists said there was a need for consideration and discussion on whether they were discussing societal or technological risks, as they're not always distinct.

Dr Elliott said professionals need to imagine worst-case scenarios and then use those to shape preventative guidelines. He added the discussion should focus on hazards and how to prevent them because risk prevention is difficult when the system exists in a black

box which doesn't tell you how the risk materialised.

Dr Natasha McCarthy, Associate Director at the National Engineering Policy Centre, said that long-term dependence on computer systems that are energy intense should give a reason for pause and reconsideration, given the impact it could have on the Government's drive for Net Zero.

Professor Alan Bundy of the School of Informatics at the University of Edinburgh said that the general public – including elected officials – had very little understanding that because AI systems can be spectacular in one specific area they are designed to excel in, this doesn't mean they will produce the same results in other contexts. He said applying AI to a system in which it's not trained to excel then blindly trusting the results is dangerous.

Gordon added there is an obsession with implementing autonomy without understanding it.

Question 3: How should safety and robustness principles be implemented in practice?

The panel said that robustness wasn't a helpful term in this context. Adam framed the discussion as more about quality, resilience, and the security of a system performing under extreme inputs.

Professor Habli reiterated that there's no need for more principles but instead practice guidance delivered through professional bodies.

Dr Elliott emphasised the need to learn from failure and seek out cause rather than blame. He added that aviation is the safest mode of travel because each crash has been thoroughly investigated and lessons learnt: similar principles are needed in this context and the equivalent of the aviation industry's Flight Data Recorder or "black box" – the kind that Marina Jirotka and Alan Winfield are working on within robotics.

Adam also suggested that sandboxes could be harnessed in assessing the skills required for using AI systems. Another consideration was having transparency in supply chains and an independent auditor.

Rashik said AI is leading to the commodification of knowledge. Breaking AI operations down into knowledge, skills, and abilities will enable us to address challenges at a smaller scale, increasing our chances of successfully creating quality AI systems fit for purpose.

Question 4: How widespread is knowledge of safety techniques among developers? How could we measure use and knowledge? (E.g. How has regulation impacted (or not) knowledge and use of tools for trustworthy AI)

Several participants commented that knowledge of safety techniques should be higher across several sectors, noting a lack of skills in this area among developers.

Professor Habli also emphasised the need to establish a safety-by-design culture, referencing the aviation industry's success due to embracing safety as a culture.

Professor Bundy sign-posted a large UKRI-funded project on trustworthy autonomous systems researching best practices.

Concluding statements

As the session drew to a close, panellists were asked what could be done to increase trust. Dr Carolyn Ten Holter said certification of engineers and developers and that an increase in professionalism and professional standards would help the industry improve. Dr McCarthy said that professionals must first understand the nature of trust to build it. Transparency – helping people understand what's being done and why – will help build trust. Complementing this, several participants also noted how the provision of assurances is key for insurance requirements across several industries.