National
Engineering
Policy Centre

Royal Academy
of Engineering

MARCH 2023

Autonomous systems

A workshop on cross-cutting governance

Executive Summary

Autonomous systems make decisions for themselves in complex environments. Regulations and standards will play an important role in governing autonomous systems. Technical standards are emerging to enable engineers and developers to embed ethical and safety principles into the design of autonomous systems across different sectors. The National Engineering Policy Centre held a workshop to explore the role for these cross-cutting standards to understand the barriers to adoption, identify the actions required, and ensure safe, ethical development and deployment of autonomous systems.

Call to action

Community

Better Regulation Executive should work with UK Regulator Network to encourage greater cross-sector collaboration on artificial intelligence (AI), machine learning (ML), and autonomous systems to build a community of understanding to tackle common challenges.

Regulator upskilling

There is a need for CPD courses for regulators to better understand existing and emerging standards for AI, ML, and autonomous systems in order to adopt them. Language across standards should be made consistent to make it easier for users to effectively understand and interpret between standards produced by different bodies. This may require standardised terminology and collaboration to build unified understanding. There is a potential role here for the emerging Institute for Regulators, working with organisations including the NEPC and its partners.

Principles and new standards

Standards bodies and regulators should work together to identify and develop usable standards beyond transparency, verification, and failsafe design. This might include principles such as design practice, principles of operational context, human interaction and security.

Industry uptake

Regulators, Professional Engineering Institutions, Catapults and public procurement bodies should promote the adoption of standards that encourage safe and ethical development of autonomous systems.

Context

Since 2019, the National Engineering Policy Centre has been exploring the safety and ethics of autonomous systems to understand the risks and benefits associated with this technology across different sectors. The project seeks to understand how autonomous systems can be ethically designed, developed, and deployed to ensure benefits are widely distributed with special attention to disadvantaged communities. Through our work, it has been argued that regulation and standards will play an important role.

On 28 April 2022, the Academy hosted a cross-sector workshop on the role of international technical standards in regulating autonomous systems, bringing together a mix of regulatory and technical expertise. It convened regulators including from the Health and Safety Executive, the Office for Nuclear Regulation and the Maritime and Coastal Agency, as well as wider expertise from standards bodies, industry, SMEs, Catapults, and academia. This workshop aimed to explore the role for cross cutting standards, understand the barriers to adoption, and identify the actions required to build a community who can collaborate to overcome common issues and ensure the safe and ethical development and deployment of autonomous systems. This echoes the Alan Turing Institute's recent call for a joined-up approach for coordination, knowledge sharing, and resource pooling for regulatory bodies facing the challenge of "AI readiness".[1]

Autonomous systems make decisions, and take actions, often in complex and unpredictable environments. These systems are typically designed to be non-deterministic where the same input can result in multiple different outcomes and this, together with the unpredictability of deployment environments, often make it impossible to predict each outcome with certainty. There are key principles relevant to the development of autonomous systems that can help assure the safety and ethical development of these systems which can be supported by technical standards. A wide range of standards exist, or are under development, that are being produced by national standards bodies, industry, and international organisations.

The standards highlighted in the workshop were chosen for their principle-focus rather than application specificity. The workshop deliberately focused on these horizontal standards, relevant across different sectors, to understand where they can help and where challenges remain. The discussions aimed to inform an action plan for further standards development and to encourage industry uptake, as well as for regulator upskilling and collaboration.

The three principles discussed were **transparency**, **failsafes** and **verification**, presented in the context of emerging Institute of Electrical and Electronics Engineers (IEEE) standards: IEEE P7001-2021

Transparency of autonomous systems; IEEE P7009 *Failsafe design of autonomous systems* and IEEE P2817 *Guide for verification of autonomous systems*. These are generic, umbrella standards, intended to apply to all autonomous systems both physical and software based. These generic, umbrella standards are intended to apply to both physical and software based autonomous systems. These were selected as exemplar principles with an important impact on autonomous systems development. There is, however, no expectation that these provide a complete set of principles.[2,3] Other principle-based standards indeed exist, and IEEE's *Ethically aligned design* report sets out 8 general ethical principles for autonomous systems.

Transparency is critical for understanding how a system operates, why certain decisions are made, and where it went wrong to understand the failures of autonomous systems. In complex, realistic environments, uncertainty and failure of systems is inevitable, hence **failsafe** mechanisms are an essential principle to build into mitigation strategies. It is also crucial to provide evidence of reliability and confidence in both the system and its decision making, by **verifying** that a whole system meets its design specification.

Following the presentations on emerging standards, Andrew White from UK's Office for Nuclear Regulation discussed some of the challenges relating to the regulation of autonomous systems and the role of standards such as these in addressing these challenges. With this context, attendees discussed how the standards could be applied and identified gaps.

Principles for autonomous systems

Transparency IEEE P7001

Alan Winfield

Transparency assumes that the basis of a particular autonomous or intelligent system (A/IS) decision or action should always be discoverable. This is important not only in understanding failures but in building confidence and trust when autonomous systems operate alongside humans. Transparency of behaviour more generally is also important for stronger levels of verification (see below). The P7001 standard was developed to set out measurable levels of transparency so that the level can be specified prior to development and then assessed for compliance. It is intended to be used by designers, manufacturers, operators, and maintainers of autonomous systems. However, transparency often means something different to different stakeholders who will require different information, relayed in plain language. The P7001 standard covers transparency for expert stakeholders (safety certification engineers, accident investigators, lawyers or expert witnesses) as well as non-expert stakeholders (users, wider society).[4]

Failsafe Design – IEEE P7009

Ken Wallace

IEEE's P7009 standard for *Fail-safe design for autonomous and semi-autonomous systems* is being developed to establish a baseline for the development, implementation and use of fail-safe mechanisms in these complex systems. It describes some of the key requirements and properties of these systems and provides tools to implement fail-safe mechanisms and methods to measure and certify the ability to fail safely. The standard will inform the design, testing, and analysis of the failsafe mechanisms as well as the organisational safety processes, should a system fail. These mechanisms are essential as autonomous systems can fail, often without a human being on hand to recover. There is a need to help mitigate risk of harm to people, society or the environment. It is intended that this standard is adapted for different sectors so they can define what is 'safe enough' in each specific context. For example, the safety requirements for a self-driving car on a public road may be different to an autonomous robot in a nuclear facility.[5]

Verifiability – IEEE P2817

Signe Redfield

The development of IEEE's P2817 *Guide for the verification of autonomous systems* will enable users to define an appropriate multistep verification process for autonomous systems based on the available tools, levels of transparency, and good practice. The guide provides resources on formal methods to provide strong evidence (mathematical proof) for the systems; simulation to understand the behaviours in specific scenarios; stochastic methods for probabilistic estimates of system behaviour; real world testing for higher risk scenarios; and runtime verification to ensures the system remains within predicted boundaries. The guide helps developers avoid common pitfalls in the collection, analysis, and inbuilt assumptions underlying the evidence that the integrated system meets the design specification. It focuses on the functionality and decision-making processes within an autonomous system rather than the outcome.

Regulator challenges and the role for standards

Andrew White, ONR

The Office for Nuclear Regulation (ONR) has an 'outcomes focused' approach to regulation, meaning they do not mandate or encourage licensees to adopt a specific standard or provide certification themselves. They do, however, require the licensee to provide explicit evidence that demonstrates the system is safe..

For non-AI software systems, the ONR requires a safety case that demonstrates assurance that a system's risks have been reduced so far as is reasonably practicable (SFAIRP) and it is expected to be primarily deterministic. AI systems are different and pose a challenge as they are typically designed to be non-deterministic, and as a result, evidence cannot always be provided for how the system will function in every possible outcome. It is accepted that AI systems will fail, but how it fails varies and relatively unknown, which creates a further challenge in assuring safety. Despite this, the ONR views AI and autonomous systems as worthwhile technologies to consider due to their potential benefits. There are few established standards relating to the safety of such systems. The ONR takes a view that regulators should act as enablers and not prohibitors of innovation. They define what the acceptable level of risk is, creating regulatory sandboxes for innovation that help identify the questions which need to be answered together.

Discussion

Value

There was agreement that high level principles are needed as they capture the key issues arising from autonomous systems. Standards are helpful as they provide practical ways to assess the system and promote consistency. They are also not prescriptive, which allows for the consideration of specific context and encourages conversations about what is safe enough.

Autonomous systems create similar ethical challenges (avoidance of harm, fairness, transparency) across sectors. However, different sectors do have different needs depending on the context of how autonomous systems will be developed and deployed. Some cross-cutting principles such as failsafes will be more familiar in safety critical domains like space or nuclear, but not all sectors will have this same starting point. So, while high-level principles are of value, it was agreed that sector-specific standards would be needed in addition.

Awareness and understanding

Regulator and developer awareness and understanding of emerging standards needs to be built up. There was agreement that regulators and developers lack the resources and time to learn about standards, which means they do not necessarily know what standards already exist or how best to apply them. With that lack of understanding, it was felt that both confidence and trust to implement and rely on a collection of technical standards was also missing. Regulators also lack understanding of AI, ML, and autonomous systems and are unable to keep up with technological developments. This means there is a tendency to consider autonomous systems as traditional systems, which is a challenge given the reasons above.

Sharing good practice

Greater sharing of information and best practice is important to encourage the adoption of standards. In particular, regulators would benefit from upskilling in techniques and the key components of strong verification. As standard implementation is not formally enforced, it would be useful to encourage the use and adoption of standards within industry.

Cross-sector collaboration

There is value in cross-sector conversation as well as collaboration between all parties (regulators, innovators, standards developers, insurers, the legal profession) to understand and tackle challenges in ensuring the safety and effectiveness of autonomous systems. Currently there is no mandate for cross-sector collaboration between regulators and it would be useful to encourage this. Often there are only a few individuals with expertise in autonomous systems in each organisation and such connectivity would help to make best use of these scarce skills. Cross-sector collaboration may also help to navigate international regulatory differences by sharing an understanding of where the overarching principles remain the same which can help build confidence in safety processes and encourage transferable learning.

Embedding ethical considerations

Adopting principle-based standards can encourage developers to consider the ethics of autonomous systems on a greater scale. There is sometimes a tension between the commercialisation of a product, ethical practice, and beneficial outcomes. Ethical risk assessments are an emerging governance tool to help organisations work through the ethical implications of the systems they are developing or adopting but uptake has been limited. Adoption of ethical standards may be encouraged by increasing ethical consumerism, where the products or services a consumer chooses are those that cause the lease social or environmental damage.

Maturity of standards

The speed at which technology can develop poses a challenge as it is often faster than the development of both regulation and standards. Few mature standards for autonomous systems exist and adoption of emerging standards need to be encouraged through mechanisms such as regulation and procurement, for example by including the requirement to meet certain standards in procurement specification.

Increasing uptake

Uptake can be a challenge where standards are not mandated by regulators. ONR explained they do not encourage licensees to adopt a particular standard whereas some sectors will only adopt ISO standards. Therefore, if there is limited knowledge of the range of standards that exist, developers will be less inclined to adopt good practice.

Clarity

Language used in standards poses a challenge as it is either inconsistent or too complex, resulting in difficulty of, or varying, interpretation of standards. Language across standards should be made consistent with a need for common terminology. It would also be important to consider how language and approaches across sectors differ.

Missing principles/standards

Regulators and developers agreed that transparency, verification, and failsafe design are important cross-cutting principles. However, other principles such as design practice, operational contexts, human interaction (outside of human factors or machine learning explainability), and security would also be valuable. Additionally, regulators would benefit from measures or forms of risk analysis as many still rely on predictable hazard analysis. Due to uncertainty, these are likely to be unpredictably wrong for non-trivial autonomous system applications.

Call to action

Community

Better Regulation Executive should work with the UK Regulator Network to encourage greater cross-sector collaboration on AI, ML, and autonomous systems to build a community to understand and tackle common challenges.

Regulator upskilling

There is a need for CPD courses for regulators to better understand existing and emerging standards for AI, ML, and autonomous systems in order to adopt them. Language across standards should be made consistent to make it easier for users to effectively understand and interpret between standards produced by different bodies. This may require standardised terminology and collaboration to build unified understanding. There is a potential role here for the emerging Institute for Regulators.

Principles and new standards

Standards bodies and regulators should work together to identify and develop usable standards beyond transparency, verification, and failsafe design. This might include principles such as design practice, principles of operational context, human interaction and security.

Industry uptake

Regulators, Professional Engineering Institutions, Catapults, and public procurement bodies should promote the adoption of standards that encourage safe and ethical development of autonomous systems.

Acknowledgements

The National Engineering Policy Centre would like to thank the following for their contributions:

Members of the Safety and Ethics of Autonomous Systems Working Group

Professor Michael Fisher FBCS FIET
Professor Nick Jennings CB FREng FIET FBCS (Chair)
Professor Muffy Calder DBE FREng FRSE
Dr Chris Elliot MBE FREng FRAeS
Professor John McDermid OBE FREng
Paul Newman FREng
Andrew Chadwick
Lambert Dopping-Hepenstal FREng FIET FRAeS
Professor Marina Jirokta FBCS
Gordon Meadow FIMarEST
Dr Robert Merrall FIAgRE
Dr Sylvain Jamais FIMechE

Royal Academy of Engineering Staff

Arizona Rodriguez, Policy Advisor
Dr Alexandra Smyth, Senior Policy Advisor
Dr Natasha McCarthy, Associate Director, Policy
Brittany Hsieh, Senior Policy Advisor

References

1 Common Regulatory Capacity for AI (2022), Aitken, M. et al, The Alan Turing Institute. https://doi.org/10.5281/zenodo.6838946

2 “Trustworthy AI”. Raja Chatila, Virginia Dignum, Michael Fisher, Fosca Giannotti, Katharina Morik, Stuart Russell, Karen Yeung. In Reflections on Artificial Intelligence for Humanity, 2021. https://doi.org/10.1007/978-3-030-69128-8_2

3 “Principles for the Development and Assurance of Autonomous Systems for Safe Use in Hazardous Environments”, Matt Luckcuck, Michael Fisher, Louise Dennis, Steve Frost, Andy White, Doug Styles. 2021. https://doi.org/10.5281/zenodo.5012322

4 “IEEE P7001: A Proposed Standard on Transparency”. Alan Winfield, Serena Booth, Louise Dennis, Takashi Egawa, Helen Hastie, Naomi Jacobs, Roderick Muttram, Joanna Olszewska, Fahimeh Rajabiyazdi, Andreas Theodorou, Mark Underwood, Robert Wortham, Eleanor Watson. Frontiers Robotics AI 8, 2021. https://doi.org/10.3389/frobt.2021.665729

5 “Evolution of the IEEE P7009 Standard: Towards Fail-Safe Design of Autonomous Systems”. Marie Farrell, Matt Luckcuck, Laura Pullum, Michael Fisher, Ali Hessami, Danit Gal, Zvikomborero Murahwi, Ken Wallace. In Proc. ISSRE Workshops, 2021. https://doi.org/10.1109/ISSREW53611.2021.00109

National
Engineering
Policy Centre

Royal Academy
of Engineering

THE ROYAL ACADEMY OF ENGINEERING

The Royal Academy of Engineering is harnessing the power of engineering to build a sustainable society and an inclusive economy that works for everyone.

In collaboration with our Fellows and partners, we're growing talent and developing skills for the future, driving innovation and building global partnerships, and influencing policy and engaging the public.

Together we're working to tackle the greatest challenges of our age.

NATIONAL ENGINEERING POLICY CENTRE

We are a unified voice for 43 professional engineering organisations, representing 450,000 engineers, a partnership led by the Royal Academy of Engineering.

We give policymakers a single route to advice from across the engineering profession.

We inform and respond to policy issues of national importance, for the benefit of society.

Royal Academy of Engineering
Prince Philip House
3 Carlton House Terrace
London SW1Y 5DG

Tel 020 7766 0600 www.raeng.org.uk
@RAEngNews

Registered charity number 293074