UKRI Trustworthy Autonomous Systems Hub

RESILIENCE

DEVELOPING RESILIENT AUTONOMOUS SYSTEMS WE CAN COUNT ON

Reliability and predictability are important in our lives. We like to be sure that when we switch on our smartphone, we can make calls and access our apps, or when we turn on the ignition, our car will start and we can drive to our destination. We rely on our machines to behave as we expect them to. We need to trust them – and for this we need them to be resilient.

Resilience is crucial to the development of ever-smarter technology and plays an essential role in ensuring that our autonomous systems are reliable and trustworthy. These complex issues are among those being addressed by the UKRI Trustworthy Autonomous Systems (TAS) Programme – a £33m multi-disciplinary research programme funded as part of the UKRI Strategic Priorities Fund. Resilience makes up one of the six TAS Nodes – focused research projects examining individual aspects of trust in autonomous systems (AS) through new research.

What does resilience in Autonomous Systems look like?

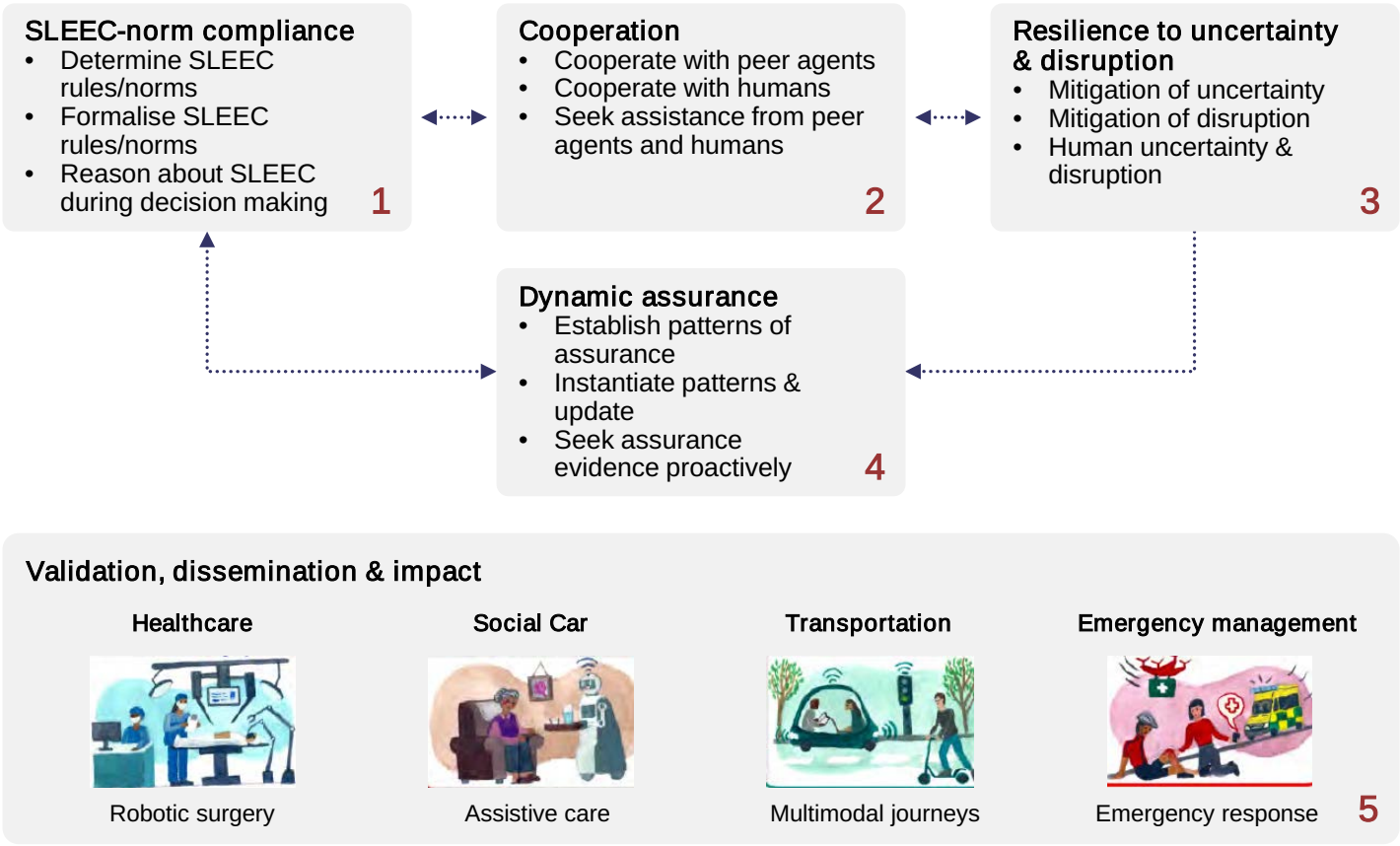
When it comes to autonomous systems, resilience is about ensuring that our machines can operate effectively, consistently and appropriately within our day-to-day lives. However, the 'real world' throws up challenges and situations that can be very difficult to predict and control. According to Professor Radu Calinescu from the University of York: "Resilience plays an essential role in ensuring that AS can be trustworthy. It is about mitigating the uncertainties and disruptions which AS encounter when deployed in real-world environments. This requires sophisticated resilience mechanisms."

The way in which the autonomous systems interact with humans and society as a whole – the socio-technical aspects – is the focus of much of the resilience work currently taking place. What are our autonomous systems required to do and what are they not allowed to do? In order for us to trust our machines, it is essential that they work in harmony with us. They must operate in ways that don't violate the norms of our society, our rules or our values. A benchmark known as SLEEC[1] - which stands for Social, Legal, Ethical, Empathetic and Cultural - encapsulates this ideology, and is based on the principles that these factors should be incorporated into the design, operation and governance of autonomous systems.

However, the reality is far more complex. Our norms and values can be subtle and nuanced. Professor Ana Cavalcanti from the University of York recounts an analogy that highlights some of these uncertainties: "If you are from the US, you know you can be arrested for trying to cross the road (jaywalking), but then if you come to the UK, you could be waiting at the roadside for a long time unnecessarily - much in the same way a robot can be brought to a halt as it doesn't understand its environment and the SLEEC norms that apply."

So how can we go about creating autonomous systems that operate within the SLEEC principles? How can we ensure that our robots always 'do the right thing'? Work is currently being carried out, both internationally and within the TAS Programme, to develop standards on the ethics of autonomous systems. We need to understand the limitations of our AS, how are they used and where. We need to understand how they behave in unpredictable environments. We also need to learn more about how they can become more resilient through collaboration with other autonomous systems.

1 Townsend, B., Paterson, C., Arvind, T. T., Nemirovsky, G., Calinescu, R., Cavalcanti, A. L. C., Habli, I., & Thomas, A. P. (2022). From Pluralistic Normative Principles To Autonomous-agent Rules. URL: https://cutt.ly/SLEEC-rule-elicitation



TAS Node in Resilience Research Strands

Accounting for human input

One of the critical areas of research, however, is how autonomous systems interact with humans. They may need some of our input, a lot or none at all. They can work alongside humans or operate around us. They have the potential to operate entirely independently.

The TAS Programme is looking at how to mitigate the difficulties and unpredictable situations that arise from this. The aim is to develop autonomous systems which learn from us about how to deal with uncertainty and disruption. Humans are capable of identifying aspects of uncertainty that impact on our goals and objectives. We need our autonomous systems to do the same; to observe trends and patterns and to anticipate disruption so that they can identify potential problems, reduce uncertainty and mitigate it. This is an important area in the development of our AS – but is not at all simple, as Dr Mark Chattington from Thales UK explains:

"To build something that is trustworthy we deal with issues at the concept level which may then get magnified throughout the design cycle. So, there are certain misunderstandings or misconceptions or unknown quantities during development, and these get magnified. This is one of the biggest struggles from the industry perspective."

Additionally, there is the element of human error. Thales have been carrying out research into the subject area of prediction and control. Amongst the findings was the discovery that an autonomous system can be working perfectly for a period of time, but then starts to exhibit undesirable behaviour due to interactions with a biased human user. The autonomous system simply re-enforces these actions, which causes the robot to fail the resilience test. The research highlighted that new tools and techniques are needed to determine resilience of the socio-technical aspects of humans and robots operating together.

Other risks need to be assessed, such as accounting for the cognitive and physical impairment of humans and the potential physical and psychological harms that can occur. How can autonomous systems communicate information to humans to avoid misunderstandings and bad outcomes?

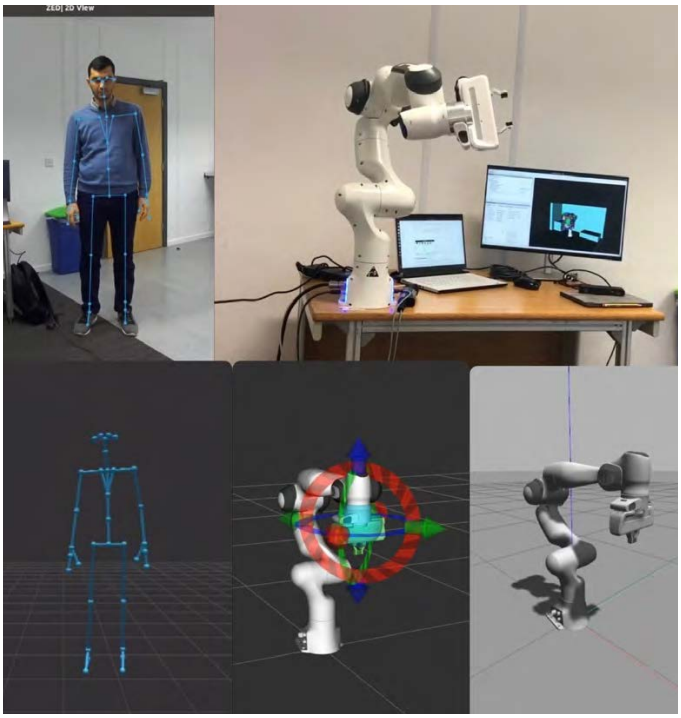
And what happens when things do go wrong? How can we be confident these systems will work and find ways to mitigate the situation as we hope? If a machine can't be made to fix itself, the answer may be to pass the responsibility to the users, which brings up other regulatory and safety considerations.

THE IMPACTS OF RESILIENT AUTONOMOUS SYSTEMS

Research being carried out has the potential to have a real impact our day-to-day lives. An area that demonstrates this well, is the growing need for personal care. Studies are being carried out in the field of occupational therapy, observing the use of an assisted dressing machine for the mobility-impaired.

Researchers have been collecting data around how the robot interacts with patients. The ongoing study, involving seven patients and therapists, will provide information that will help shape the design of future autonomous systems.

Information being gathered includes motion data and human reactions to differing types of disruption. By understanding human intentions and analysing possible hazards, the aim is to ultimately teach the robots to 'reason' - understand how to predict problems and failures based on human and environmental observations. As well as for personal care, this has the potential to open up a world of opportunity for the medical and surgical fields.



University of Sheffield, Resilience case study assisted-dressing robotic arm, Franka Emika, (top right) including ROS/Gazebo simulator (bottom), with haptic feedback and AI-based motion tracking (left) (from https://www.resilience.tas.ac.uk/annual-report)

However, it's not just the physical interactions between humans and autonomous systems that need to be taken into account. There are also should be the social and ethical complexities to consider.
In the assistive dressing example above, even basic interactions between one AS and one end-user are hard to establish and to build into rules and norms. There are further questions around the meaning of privacy or fairness, which pose important challenges to consider.

MAKING A DIFFERENCE IN THE REAL WORLD

Among the many 'real world' situations where resilient robots would be beneficial, several examples bring the nuances and complexities around the SLEEC[1] principles into sharp focus.

Emergency management during natural disasters, such as fires, floods, hurricanes, or earthquakes is one such case. Artificial Intelligence and autonomous systems can greatly assist human decision-making in these disaster zones and extreme environments.

However, complexities arise from there being multiple users of the robots, a range of other people involved in the relief efforts, plus unpredictable human reactions in an emergency. For example, how would a migrant at sea react to a drone (Unmanned Aerial Vehicle or UAV) above them? Or in a flood management situation, where there are different people, organisations and sources of information being used to formulate the emergency response?

Professor Mohammad Mousavi from King's College London gives this example of the many challenges: "Machine learning may detect an object which is apparently coming towards the UAV. What is useful to do in this situation? Having awareness of what is useful to the user in various unanticipated scenarios is a big challenge."

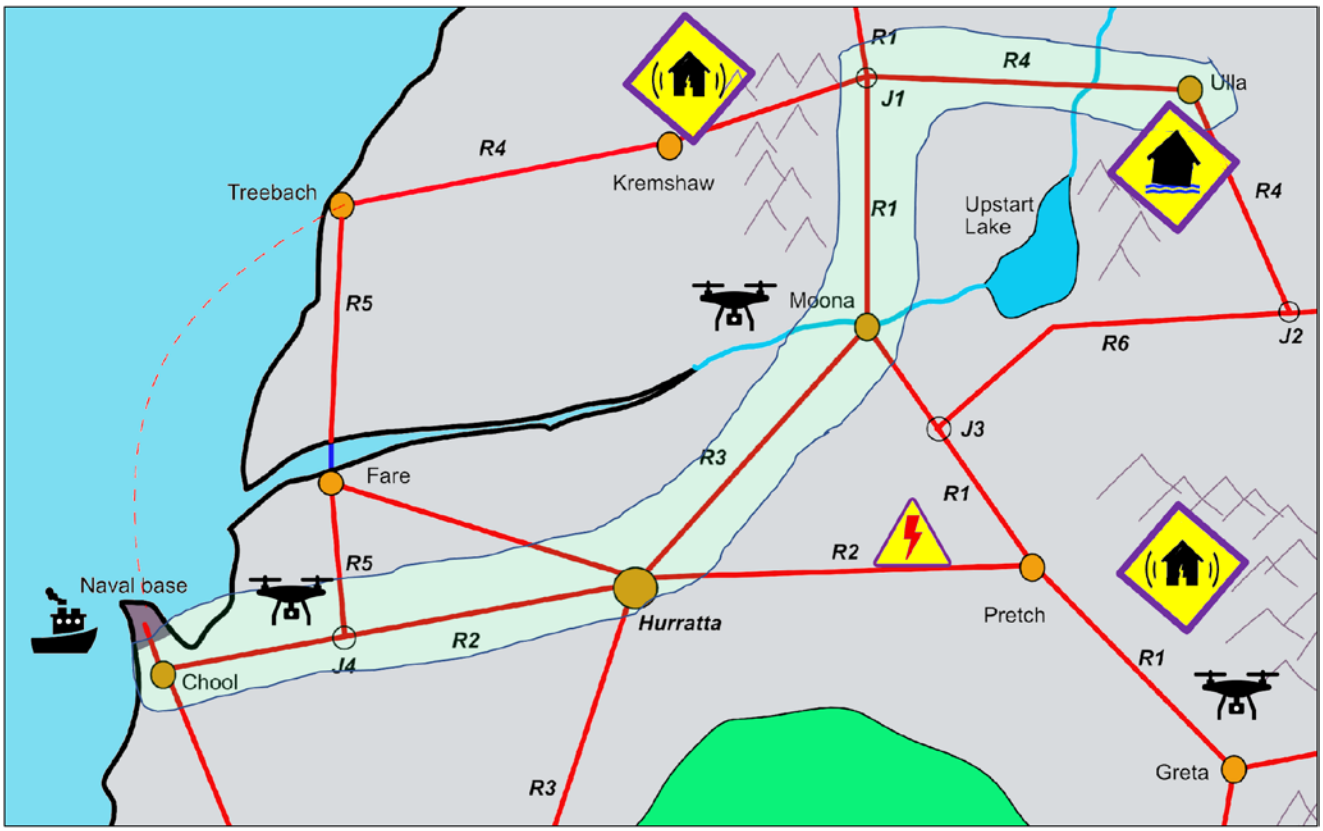
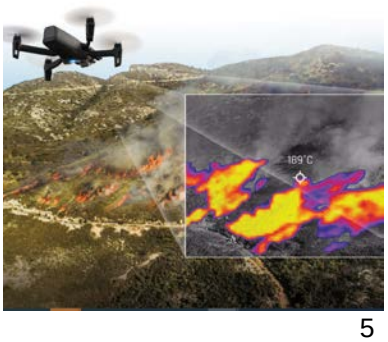


Illustration of a UAV-assisted emergency evacuation system (adapted from Paterson et al, 2019)

There is significant untapped potential for autonomous systems to be used in disaster response situations, but equally there are many variables to try to overcome.

For instance, is there a conflict between regulations and potential loss of life? This is a critical question that is currently being studied as part of ongoing research. There are collaboration issues too. It would be useful if the autonomous system identifies when it needs to delegate to a human, but ensuring ongoing cooperation between the system and the human operator is important too. To robustly consider resilience, these interactions need to be captured.



ONGOING RESEARCH AND REQUIREMENTS

Research into developing robots we can count on is continuing at a pace. Resilient autonomous systems have the potential to impact all parts of society –from space exploration to helping disabled people take part in activities that are currently inaccessible to them.

Within the TAS Programme itself, research is continuing into issues impacting resilience. Key progress has been made on the development of a SLEEC[1] Framework. Work on uncertainty, disruption, uncertainty reduction and disruption prediction has been carried out. Professor Radu Calinescu underlines its importance: “This is essential for resilience because you might not be able to mitigate all the uncertainty thrown at you and so need to obtain additional information about the environment in which you operate, which is how humans make decisions too.”

Other important areas of research include communication: working with different sectors to develop a common language that is understood by all users of resilient autonomous systems. An initial version is currently being prepared for publication, but it is still early days. Professor Ana Cavalcanti explains the impact this would have: My vision is that a diverse team of people — say a sociologist, lawyer and philosopher — can use a common language as they describe the capabilities and restrictions of the AS. There will be a library of concerns and tools to identify any conflicts and resolve them. There will also be tools that the engineer can use to inform their design.”

Is the future autonomous?

So, with respect to resilience, what is the ultimate goal? Can we realistically design resilient robots that complement and enhance all aspects of our lives? Will we really be able to fully rely on them when we need them? Professor Radu Calinescu believes this goal could be achieved: “My vision is of interoperable AS that self-organise into resilient, safe and beneficial 'systems of systems’ capable of working alongside and for humans. For example, AS that assist with planning of evacuation routes after a natural disaster co-operating with AS that provide supplies and medical assistance to those evacuating.”

However, to create truly resilient autonomous systems, we cannot rely solely on design. A key objective needs to be operational resilience. We need to ensure our autonomous systems proactively cooperate with other machines and humans, and form teams to tackle challenging problems. We also need a way to communicate the SLEEC requirements, both to developers and the robots themselves, so as they can incorporate these into their decision-making processes.

If we can achieve these goals, it could pave the way towards the creation of autonomous systems that carry out a wealth of socially-beneficial tasks in real-world environments. Professor Mohammad Mousavi adds: “It would be wonderful if we had an autonomous vehicle (AV) which could be trusted and for which we had evidence of safety and usefulness. For example, a SLEEC AV that could do the school run for me.”

However, although research and development will enable us to create systems that are more and more autonomous, we need to proceed with a certain amount of caution. From society’s perspective, what are we trying to achieve through automation?

Effectively, we need to weigh up the value of adding autonomy versus the risk it poses. With so much uncertainty, for example from the environment, the human user and the autonomous system itself (especially machine learning systems), we cannot assume it that is always the solution to everything; but recognise that in some environments, such as disaster management, that the risks are justified by the added value.

REFERENCES

Chance, G., Jevtic, A., Caleb-Solly, P. and Dogramadzi, S., 2017. A quantitative analysis of dressing dynamics for robotic dressing assistance. Frontiers, 4(13), p.1.

Chance, G., Camilleri, A., Winstone, B., Caleb-Solly, P. and Dogramadzi, S., 2016, June. An assistive robot to support dressing-strategies for planning and error handling. In 2016 6th IEEE International Conference on Biomedical Robotics and Biomechatronics (BioRob) (pp. 774-780).

Paterson, C., Calinescu, R., Manandhar, S. and Wang, D., 2019. Using unstructured data to improve the continuous planning of critical processes involving humans. In 14th IEEE/ACM International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS) (pp. 25-31).

About the Trustworthy Autonomous Systems (TAS) Hub

The TAS Hub sits at the centre of the £33M Trustworthy Autonomous Systems Programme, funded by the UKRI Strategic Priorities Fund. Its role is to coordinate and work with six research nodes to establish a collaborative platform for the UK to enable the development of socially beneficial autonomous systems that are both trustworthy in principle and trusted in practice by individuals, society and government. For more information please visit the website: https://www.tas.ac.uk/.