UKRI Trustworthy Autonomous Systems Hub

VERIFIABILITY

CHECKS, BALANCES AND GUARANTEES – HOW VERIFICATION IS CRITICAL TO THE CREATION OF TRUSTWORTHY TECHNOLOGY

Autonomous systems are predicted to play a central role in our day-to-day lives in the future, with technology such as drones, driverless cars and assistive care robotics already making a huge impact. It is an exciting prospect, but not without its concerns. Without humans to operate them, how can we be sure that our autonomous systems are going to function in unconstrained and complex environments as we need them to? To fully welcome them into our lives, we need to be confident in the safety and reliability of their decision-making in the presence of uncertainty and while comprising artificial intelligence components. We need assurances.

Determining trust is central to the work of the Trustworthy Autonomous Systems (TAS) Programme – a £33m multi-disciplinary research programme funded as part of the UKRI Strategic Priorities Fund. The UKRI TAS Programme comprises six Nodes, which are separate research projects, each examining individual aspects of trust in autonomous systems. Among these is the TAS Verifiability Node, which is exploring the tools and techniques we can use to assess our autonomous systems and give us confidence in their abilities.

There are many challenges to be overcome in this context. Is it really possible to guarantee that our smart technology is consistently reliable, safe and secure? What is the best approach to verification? What checks and balances can we put in place to ensure our autonomous systems meet expectations? How best can we guarantee their operation and decision-making throughout their life cycle?

As with many areas of autonomous systems (AS), artificial intelligence (AI) and machine learning (ML), verification is not straightforward.

The complexities of assurance

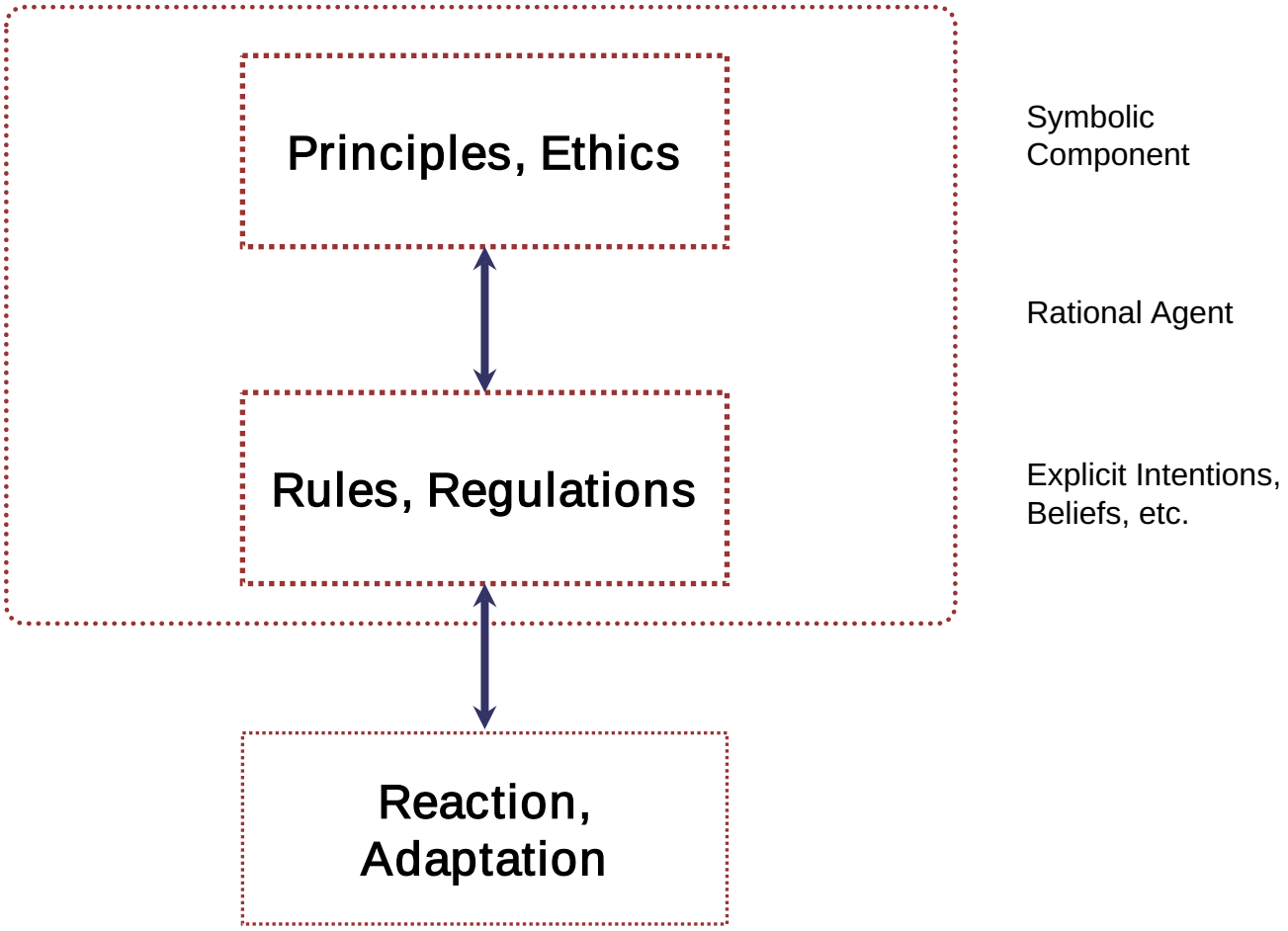
One of the main challenges of verification is what's known as heterogeneity; the fact that autonomous systems all have differing software, hardware and application environments. A one-size-fits-all approach does not work. Verifying AS involves many different areas of expertise. What do we verify in terms of a system's properties? How can we ensure these systems are fault-tolerant and adaptive? How can we efficiently upgrade them? And how do we interpret results and know if we have succeeded?

A big challenge is making the step from cyber physical systems to autonomous systems. How can we apply verification techniques used in traditional computer science to AI-based systems? Software is now making decisions that humans used to make. We need to be confident that we can trust these decisions, especially when they are mission-critical and in unpredictable environments. We need to assure safety. How do we do this?

According to Hamid Asgari from Thales UK, who are involved in verification research, the best approach is a holistic one: "Incorporation of AI into AS will need new processes and techniques from the start, across the whole lifecycle from context, through design, development, testing, integration, runtime evaluation and verification."

There is also the human factor. How much input - if any - will we have in any scenario, and what impact will this have? Ethical concerns also need to be taken into account during the verification process, which poses interesting challenges. Many real-life situations are nuanced, and we make decisions based on circumstances, multiple sources of information and social conventions. Would an autonomous system act in the same way as a human in a complex situation - and how can this be verified? As an example, would an autonomous vehicle waiting at a road junction make the same decision we would? Would it follow the highway code at all costs? How would we verify its decision?

In the face of these complexities, it becomes clear that we need to look at verification in a broad context across the entire AS hierarchy. In this way, we can try to understand not just what decision was made, but why it was made.



The Hierarchy of Autonomous System decision making from low and mid-level autonomy to high level human in the loop decisions (extract from white paper (Fisher, 2021).

VERIFICATION TECHNIQUES

The various types of software, hardware and uses for our technology means that, in reality, there is no single technique to verify any one system.

Different approaches to verification are also required throughout a system's lifecycle, from the design stage right through to the online verification of the deployed system. Once in operation, there is a need to verify the decisions that autonomous systems make in uncertain and changing environments. In addition, the verification technique used depends on the assurances required. For example, if an absolute guarantee for a critical system is required, mathematical proof may be needed. For other systems, a statistical guarantee may be sufficient for the specification.

Model-based testing is currently the most widely-used technique for verification, but due to the heterogeneous nature of AS, finding models from different disciplines to enable verification is a big challenge. How can we build and specify verification test cases, create robust simulations and representative synthetic environments? There are challenges such as requiring good specification in the first place - something that's even harder for AI systems. Formal modelling expertise and the ability to abstract away from unimportant details.

Dr Son Hoang from the University of Southampton believes the answer lies in controlled simplification: “Model Abstraction is key for successful verification. If the model is too abstract you cannot verify properly, if too concrete it would be impossible to verify practically.”

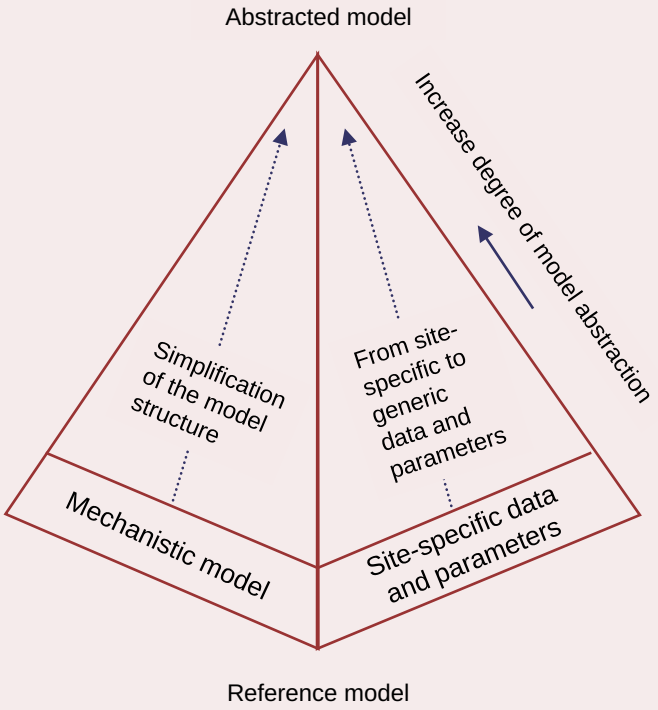


Diagram illustrating model abstraction. The base of the pyramid represents the most detailed model, often the reference model, as you move up the pyramid the model becomes more abstracted both in terms of the model structure and dataset (Schneider et al., 2010)

Verification in action

It is fair to say that verification is a continuous process that is constantly evolving. However, encouraging progress is being made.

Researchers at the TAS Verifiability Node are currently collaborating with other TAS Nodes and industry partners to gather data which will influence future tools and techniques.

Academic research projects are underway that are proving to be important case studies for verifiability. One such study focuses on a fire-fighting drone. This project at the University of Leeds, is looking at the use of trustworthy autonomous systems in emergency situations. The focus is on the use of AI to help tackle fires in high-rise buildings using a simple computer vision-based ‘search, position hold and extinguish’ implementation method. TAS researchers have been working with the project teams on modelling, with simulations and flight experiments expected in the future.

University of Leeds Fire-fighting drone Verifiability Case Study (from https://bit.ly/Verifiability)

This work will help pave the way for other potential applications in an emergency response situation; for example, supporting refugees within camps or search-and-rescue operations at sea. These are unpredictable scenarios where verification is needed to provide assurances and affirm the safety of the system. Professor Radu Calinescu from the University of York explains:
“If the presence of the robot is unknown to the human being supported, the interaction cannot be planned in advance, so ensuring safety, security and compliance with legal and ethical norms all need to be verified.”

Potential uses of verified autonomous systems are not, however, limited to the emergency services. Another field with great potential is healthcare.

TAS researchers across multiple Nodes have been working on a case study with the University of Sheffield involving an assistive dressing robot. Using AI, the robot would help a patient put on a jacket using a single or bi-manual robotic system with sensors, touch-based and speech-based feedback, computer vision and prediction of the patient's position and movement. This is a significant case study for verification research, as it involves the use of real robots which the TAS researchers can model. Work is underway to provide safety assurances using a range of tools and techniques. Modelling of human and environmental interactions and robotic behaviour is also taking place.

However, despite having real-life examples to work with, verification is by no means straightforward.

Robotic-assisted care for people with physical impairments needs to factor in complex interactions between the robot and the patient. For domestic, social and healthcare robots, the issues go beyond safety and security to include privacy, ethics, transparency, explainability and regulatory barriers. This case study perfectly illustrates the importance of verification for wider acceptance of technology in society. It is hoped the findings will provide vital information on these aspects and help 'stress test' the decisions that the robots need to make.

University of Sheffield assisted-care robotic arm Verifiability case study (from https://bit.ly/Verifiability)

Next-level robot teams

Another exciting area that will push the capabilities of verification is swarm robotics – multiple robots acting together as a system. With swarms, there are complexities at all levels; from the design, deployment and formal verification of the properties, to the behaviour of the robots working together. Re-creating true-to-life outdoor scenarios in a simulated environment is challenging. There are regulatory standards for individual robots, but none currently exist for robotic swarms.

It is still early days for this emerging technology, but various studies are already underway.

Thales are working on a project to verify the behaviour of a fleet of drones, including verifying some of the hardware that will be running the operating system and software. They are also looking at verifying the interaction between the drones and the actions of the fleet. This is not something that can be done in the software alone, as one drone acting alone cannot fulfil the mission; it depends on the interaction of all the drones. Formally verifying this kind of scenario is challenging.

CAN WE GUARANTEE THE FUTURE?

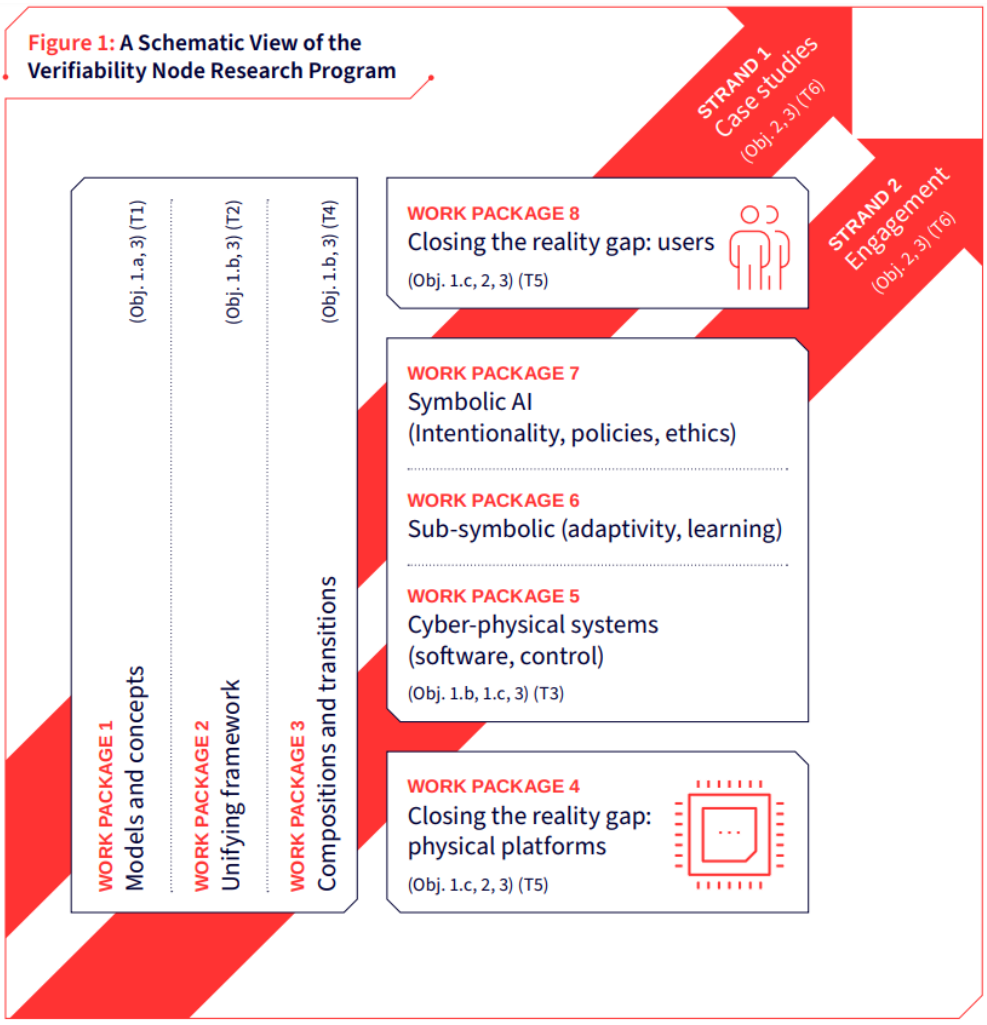
In recent years we have all begun to accept more and more intelligent technology into our everyday lives, from helpful to life-dependant: smart speakers, chatbots, delivery drones and driverless trains to robot-assisted surgeries.

But as advances in AI and AS gather pace, verification of these emerging systems will be crucial to their integration into society. Trusted testing and reliability is key to this - but achieving it is complex. Significant strides are, however, being made and collaboration is at the heart of this.

The TAS Verifiability Node has been casting its net wide, reaching out to AS researchers in the UK and internationally, building a community of people involved in verification to share expertise and information. They have also been taking part in government consultations and engaging with the wider public to expand knowledge.

As for verifiability itself, the vision is a holistic one – a unifying framework where domain experts from different disciplines, from human aspects to roboticists, can interact with the same verification framework using their different domain-specific languages and notations. It needs to be a framework that is understandable for all and accessible for any type of industry.

According to Professor Mohammad Mousavi from King's College London it is about creating "one tool that connects all these different aspects and that comes up with the holistic verification result – not only for the initial design but also throughout the lifetime of the system. This common verification framework will provide continuous trustworthiness guarantees."



A schematic of the Verifiability Node Research Program (from https://bit.ly/Verifiability)

This highly-integrated framework would also address another key issue – that of the cost. It would help reduce the financial, computational and energy implications of verification, even with the increase in complexity that is likely to be required in the coming ten years.

According to Professor Jim Woodcock of the University of York, the impact would be significant: “The cost of verification will decrease 100-fold for the same level of trustworthiness, or better, and it will be scalable. This depends on understanding problems, notations, proof techniques, mechanisation technology, model checking but also theorem proving.”

As for the future and our acceptance of ever-more autonomous systems, researchers believe we need to look closely at what we really want to achieve. Professor Kerstin Eder from the University of Bristol explains: “Rather than push limits of machine learning and autonomous systems, we need to reassess the way we design and engineer these systems. For safety-critical systems we need to fundamentally understand what we can or cannot do and be clear about that.”

In effect, it is about accepting trade-offs that will create trust: using safety-critical technology in situations when it is required and pushing the limits of AS, AI and ML in cases where it is not.

If advances in technology depend on public confidence, it critical that we build that trust through assurances so its future can be guaranteed.