UKRI
**Verifiability Node**

UKRI Trustworthy Autonomous Systems Node on

# Verifiability

ANNUAL REPORT 2020-2021

UKRI Verifiability Node



# Foreword

**Welcome to the first annual report of the UKRI Trustworthy Autonomous Systems Node on Verifiability! The Node work started off in November 2020 and in this report we cover the Node activities till the end of November 2021.**

We had an exciting first year: the research associate posts were filled, common case studies were identified, and several project meetings as well as community-building and engagement activities were organised. We engaged very actively with all Trustworthy Autonomous Systems Nodes and the Hub, and planned research activities with them. Although we are very early in the life-time of the project, we already have exciting results.

The management of the Node moved from the University of Leicester to King's College London. This provided us with additional investment from our new host, as well as an opportunity to engage more closely with the Hub, which is also co-hosted at King's College London, and plan several joint activities with them. With the gained momentum, we expect the Node to move vigorously ahead and achieve all its promised outcomes.

In the meantime, you will find many interesting pointers in our first annual report as to how to engage with us and enrich our community for Trustworthy Autonomous Systems. We are eagerly looking forward to collaborating with you and together building a strong and inclusive community interested in Trustworthy Autonomous Systems.

**Mohammad Reza Mousavi**

Principal Investigator of the Verifiability Node
and Professor of Software Engineering

King's College London

King's College London | UKRI Engineering and Physical Sciences Research Council

UKRI Verifiability Node



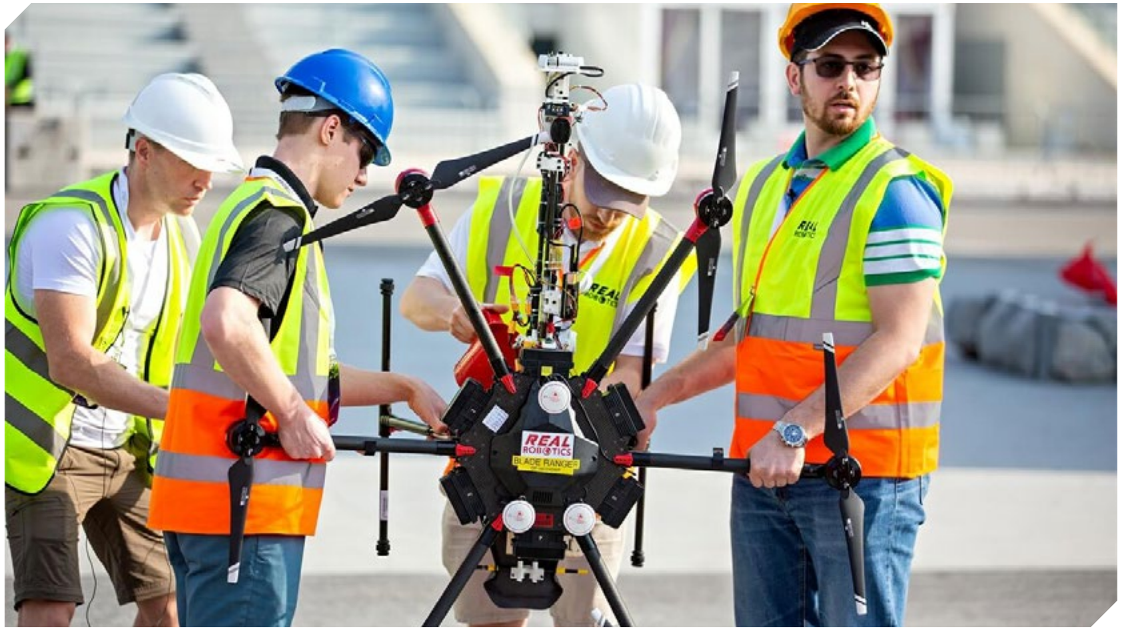
# Introduction

The vision of our Node is to carry out the foundational research to enable that in about 10 years we will have a verified autonomy store. Autonomous systems and components for autonomy in such a store go through rigorous and extensive verification processes upon submission and throughout their evolution. Having passed submission checks, components and systems are made available in a package providing the software, models for design, for compatible platforms and environments, properties, and verification evidence. The store also provides automated facilities for verification of updates to models (to include new algorithms, platforms, and environments) and components (to cater for adaptive and evolving behaviours and for changes or extend functionality) and for incorporating new verification evidence such as deployment test results. Verification covers components and their variability and evolution, their interoperability, and system-level properties for component compositions.

A holistic approach to the verification of autonomous systems is currently missing and is the key enabler to our long-term vision. Once made available, it enables a coherent and meaningful judgement about the trustworthiness of autonomy, and connects the otherwise separate disciplines used in the design and verification of autonomous systems.

A single universal modelling language, verification tool or technique is not feasible or desirable. However, we must be able to verify different aspects of these systems and how they operate together to enable trust. Our goal is hence to develop a unifying framework that will integrate and coalesce rigorous verification techniques of autonomous systems to quickly and easily verify complex autonomous systems. We achieve this outcome through a world-class program comprising foundational research as well as community-building and engagement.

In the remainder of this report, we first review the basic building blocks of our research, and the community-building and engagement program. Subsequently, we review the main objectives, initial results, and future plans for the various work packages that collaborate in delivering this program. We conclude our annual report with an overview of our first year outputs, comprising research papers, policy papers and standards, keynote addresses, major events, and posters.



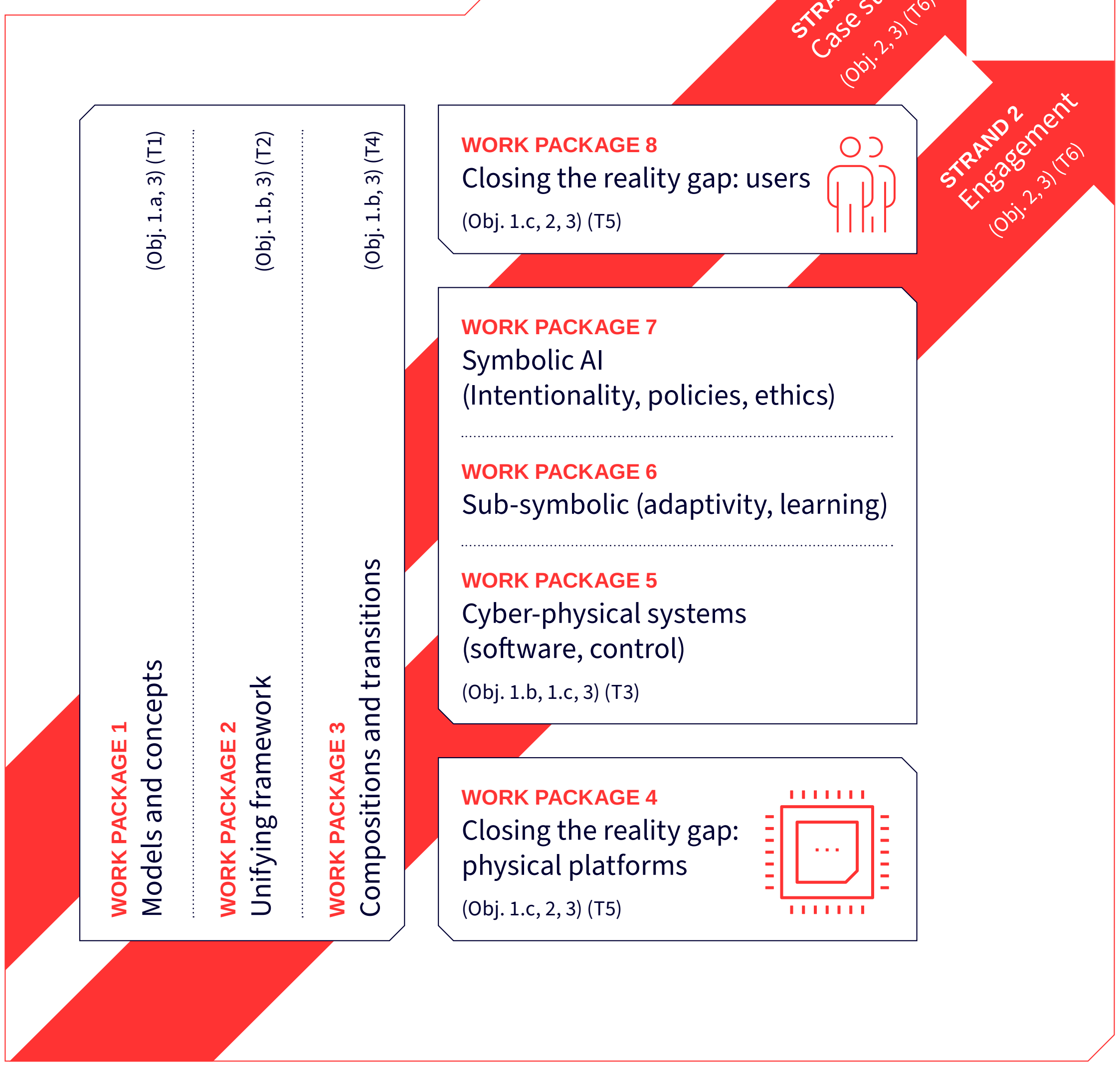
*Firefighting drone*

# Research

Our research themes are realised by eight work-packages, and two cross-cutting strands, as shown in Figure 1. Work Packages 1 and 2 will build the syntactic and semantic foundation for our unifying framework while Work Package 3 will mechanise the unifying framework, and provide translations and refinements across various sub-domains. The domain-specific aspects of verification are developed in Work Packages 4 to 8, which also have as a common goal connecting their concepts and verification tools and results through the unifying framework.

Our research program is designed in three phases: in phase 1, the foundational work on models and concepts (Work Package 1) commence with the start of the project and will continue for 30 months. In phase 2, the unifying semantic framework and the algorithms and tools required for its mechanisation (in Work Packages 2 and 3) are developed; this phase has started in month 6 and will continue till the end of the project. Finally, phase 3 starts after the first year of the project and building upon the unifying framework, delivers the holistic verification of autonomous systems, by exploiting and connecting the domain-specific aspects of autonomous systems (Work Packages 4 to 8). In this report, we review the initial results of all three phases: although the main body of work in phase 3 starts after the first year, the co-investigators of the project have already started looking into the approach for this phase and report on their initial results and future plans.

**Figure 1:** A Schematic View of the Verifiability Node Research Program



UKRI Verifiability Node



# Community-building and engagement

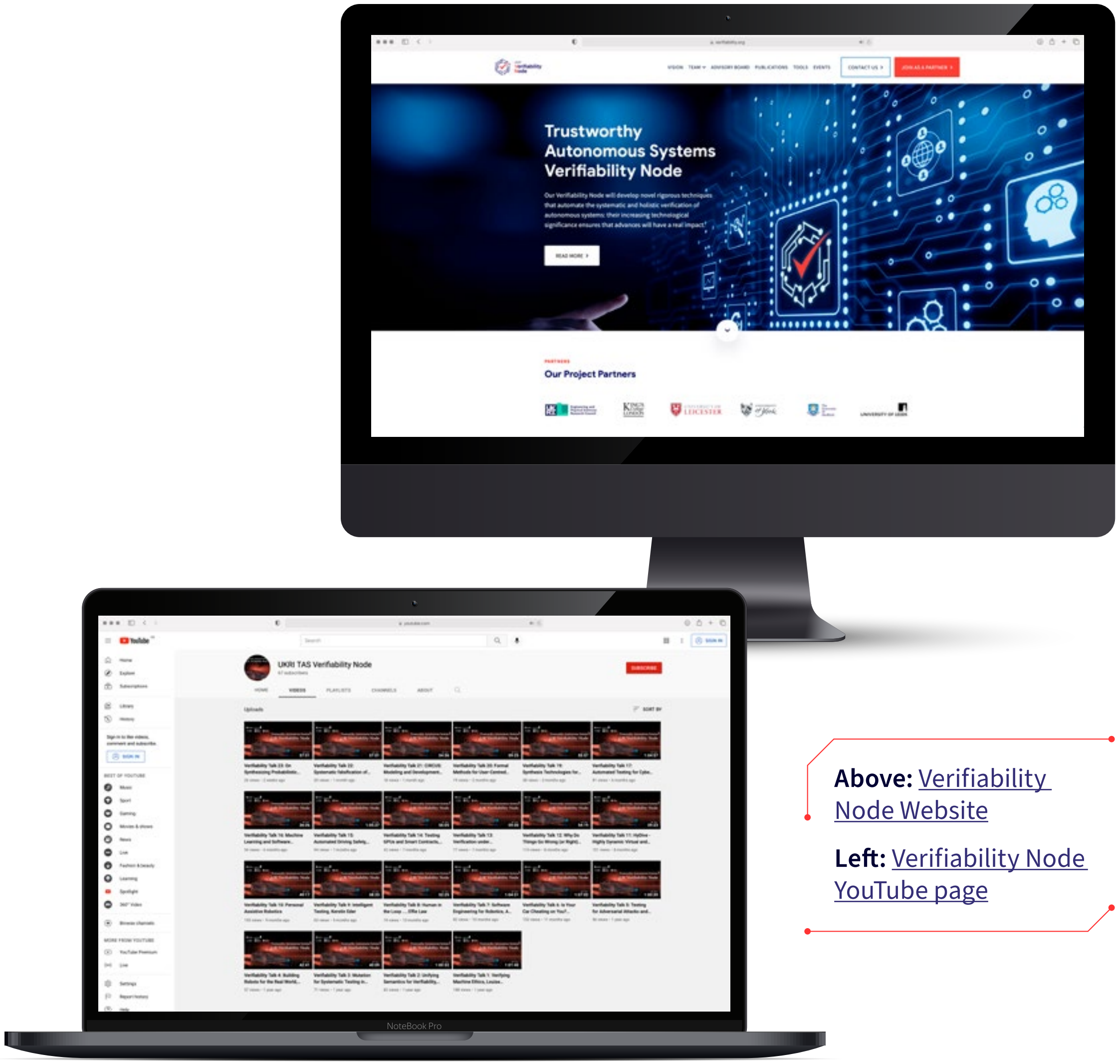
Our Node proposal has included an ambitious plan on building a strong and inclusive community interested in Verifiability for Autonomous Systems. To realise these ambitious plans, we organised several community building and public engagement activities during the first year.

To reach out to the community and disseminate our community-building and engagement activities we have adopted various media and resources. We have a professional website (https://verifiability.org), which went live right from the start of the project to make our results accessible to all stakeholders and interested parties. We have actively used social media (Twitter handle @tas_verif and a LinkedIn page), and maintained and expanded our public mailing list (verifiability@googlegroups.com). Our public mailing list currently has close to 150 members. Finally, we have disseminated our events through our Youtube channel.

## Community building

Our community building activities are aimed at engaging with researchers and industry participants to present and discuss state-of-the-art research. The outcome of such activities has been a wider reach of the Node's research and knowledge exchange.

We organise the Verifiability Talk: a bi-weekly seminar, in which distinguished experts on the subject of testing, validation, and verification of autonomous systems present their research to members of our community. Thus far, we have held 24 instances of the Talk. We have edited and published the recordings on our Youtube channel, where we have had close to 2,000 views so far.




**Above:** Verifiability Node Website

**Left:** Verifiability Node YouTube page

# Community-building and engagement continued

Within the TAS community, our Node has been actively engaged in the TAS Hub all-hands meeting by presenting four posters and research talks, as well as co-organising two workshops (on Uncertainty and Functionality), and a mentoring session (on Research Funding Proposals). We have also organised mutual meetings with all other Nodes during the first year. These have led to subsequent collaborative workshops with the Resilience Node (on specifying and reasoning about uncertainty) and the Security Node.

We have presented our work at TAS events, including a poster on mechanised assurance at the TAS All Hands meeting, and presentations on the case study at the Verifiability Node WP-Leads Meeting and at the Verifiability Node PDRA meeting. We have presented our work on uncertainty and probability in a lecture series at Aarhus University in Denmark and at summer schools at Southwest University and Northwest Polytechnic University in China. We have contributed to the organisation of a TAS workshop on uncertainty. We have organised a workshop on Applicable Formal Methods at the Formal Methods Symposium.

The all-hands meetings within the Node has helped us refine work collaboration plans and discuss future steps. For instance, we dedicated one of our all-hands meetings to case studies, where we discussed how each work package can contribute and what are the collaboration points based on two concrete case studies: a firefighting drone and a dressing-assistant robot.

To complement these meetings, PDRAs retreats have been planned and one such retreat has been organised. This first retreat occurred in September 2021 with subsequent retreats planned every four months. The goal of such meetings is to outline concrete steps and action points and engage in further cooperation with respect to the case studies. As part of future plans, we have made arrangements for concrete international collaboration with researchers from Arizona State University. We aim to arrange travel between the two organisations to conduct research on case studies related to automotive scenarios.

## Public engagement

Our public engagement activities are aimed at a wider community that comprises not only academia and industry researchers but also members of the general public. We have engaged in two podcasts: Living with AI, where we discuss exploring trust in driverless cars and the first instance of TAS Conversations, where we debate the complexity of integrating autonomous vehicles safely into our society. We have also acquired equipment that will help us present our research to the general public, such as components for vehicle simulation and physical autonomous cars. Our goal is to use this equipment to participate in events (e.g., with the Royal Society and British Science Museum) whose aim is to build public trust in autonomous systems.

## Policy engagement

In the context of policy engagement, we have produced two policy papers in response to the U.K. government requests. More specifically, we have responded to consultations about connected and automated mobility (CAM) in the UK and data protection. In the former, we give evidence about the impact of CAM in the UK's future, the country's preparedness and how it fares compared to other countries. In the latter, we emphasise the importance of research in explainable AI and also in facilitation and transparency of data sharing. We have also taken part in workshops co-organised by UKRI on Ethical Funding for AI as well as a Workshop on a Funding Scheme at the Intersection of Quantum Computing and ICT.

With our expertise in verification and autonomous systems, we are also involved in a range of standards activities, both at the national and international levels. At the national level, Fisher is a member of the British Standards Institution AMT/10 committee on Robotics, which feeds into the ISO TC299 international standards activity on Robotics. Particularly relevant is involvement in the updating and revision of the BS8611 "Guide to the ethical design and application of robots and robotic systems", due to be published in 2022. Internationally, Dennis and Fisher are both involved in the "IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems" and specifically the new international standards being developed on "Transparency of Autonomous Systems" (IEEE P7001; Dennis) and "Failsafe Design of Autonomous Systems" (IEEE P7009; Fisher).

# Models and concepts

Ana Cavalcanti and Matt Windsor (University of York)

## Overview

**This work package concerns the linguistic concepts for the specification of properties and description of autonomous systems. Here, we aim to study and specify designs and properties to enable various static and dynamic verification techniques. To enable this, we seek notations that are appealing to practitioners and mathematical models that can be used to support and justify (compositional) verification.**

## Year 1 Activities

### Requirements

Requirements based on local and node case studies and essential features, in particular the Firefighting UAV and Assistive-dressing robot.

### Notations

Preliminary version of RoboCert: a language to describe properties of mobile and autonomous robots that are integrated in the RoboStar framework, based initially on both natural language and UML sequence diagrams (and related notations).

Preliminary version also of RoboWorld, the associated notation for operational requirements, and a rule-based language to capture legal and ethical requirements in collaboration with the Resilience Node. A unified semantics will bring all these notations together.

### Formalisation

Metamodel, semantics and tool being developed, using Eclipse.

### Case studies

Initial work on the Firefighting UAV, with WP8 (University of Leeds). We have a partial model of the control software, including the robotic platform, complete outer controller specifications, planning, and some of the spraying and aiming. We also have a physical model of the robotic platform, and automatically generated simulations. Work on the assistive-dressing robot is in collaboration with the Resilience node. Other examples have been used to evaluate RoboWorld.

## Plans for Year 2

Produce a taxonomy of concepts and techniques to be captured by RoboCert.

Version 0 of languages, followed by publications informed by case studies.

Ambition to cover more notations (possibly collaborating with new partners in Brazil on activity diagrams), modelling languages (e.g. RoboSim), and formalisms (PRISM for probabilistic properties, Isabelle/UTP, and so on).

Upkeep of metamodel, semantics, and tool. Initial work on constraints in the Eclipse ecosystem, and possibly also as OCL.

### Case studies

Finishing the case studies to a publishable standard. Use them to foster collaboration with other work packages and nodes.

*Assisted care robotic arm*

# Unifying framework

Jim Woodcock and Thomas Wright (University of York)

## Overview

**The goal of this work package is to link together diverse computational paradigms and semantics at different levels of abstraction. We plan to mechanise our unifying framework and build sound toolchains that are open and extensible. We will use the framework further to link techniques from other research nodes.**

## Year 1 Activities

### Unifying Framework

We have started work on the formalisation of heterogeneous semantics. This work is informed by the ongoing work in Work Package 1 on Language and Verification Concepts. We started foundational research on heterogeneous semantics. Our initial work is in formalising the semantics of the process algebra underlying the RoboStar tools. We are adapting the process algebra Tock CSP to express it in our Unifying Theories of Programming (UTP) formalism framework and implement this in the theorem proving framework Isabelle/UTP.

We are collaborating with the Resilience Node on techniques for describing uncertainty in modelling autonomous systems. We have devised a simple probabilistic programming language with Bayesian learning. We have applied this to simple case studies, including a robotic problem: localisation with noisy sensors. We are applying our ideas to modelling uncertainty in an incubator and modelling the digital twin for a thermal incubator system. The incubator is an insulated box fitted with a heat-bed with a software system for communication, a controller, and simulation models. Our next step is to mechanise this semantics in Isabelle/UTP and after that, to combine it with our timed reactive model specified above. Our work feeds into the next steps on Refinement and from there to Work Package 3 on refinement relations and transformations.

### Case study: Robot-Assisted Dressing

We started work on the safety assurance of the robot-assisted dressing case study using a range of tools and techniques developed by the York RoboStar group. We have been working in three areas.

**(i) Structured Assurance:** We are developing assurance arguments using the Isabelle/SACM framework. We are carrying out hazard analysis, documenting a set of potential hazards based on the structured hazard analysis whilst defining SACM terminology capturing the methodology of hazard analysis.

**(ii) Control Flow Modelling:** We have constructed the initial model of the dressing robot controller in RoboChart, including the core dressing flow and processes to halt dressing in response to emergency stop requests and environmental conditions.

**(iii) Physical Modelling:** We have constructed the initial physical model of the dressing robot arm using the RoboSim physical modelling tool, generating an SDF model capturing the arm's kinematics.

## Plans for Year 2

Our future work will further develop these models to design the refinement relations. We will extend the RoboChart model to add details of interactions to capture the key hazards of the system, enabling verification of hazard avoidance using model checking. We will extend the physical model and produce a platform mapping, linking the physical model and the controller, enabling combined analysis of the robot as a cyber-physical system.

1. Complete UTP semantics for ✓-Tock CSP. Mechanise in Isabelle/UTP. Begin work on extension to probabilistic semantics. Start work on Refinement. Explore UTP semantics for agent-based systems components of heterogeneous semantics.
2. Robot-Assisted Dressing Case Study Complete RoboChart models: control system, robotic platform, environment. Complete analysis of the model. Continue the assurance argument.
3. Uncertainty and Probability Mechanise the semantics for the Bayesian probabilistic programming language in Isabelle/UTP. Mechanise the proofs of the small case studies. Tackle larger case study: Fire-fighting UAV.

# Compositions and translations

Robert Hierons and Yasmeen Rafiq (University of Sheffield)

## Overview

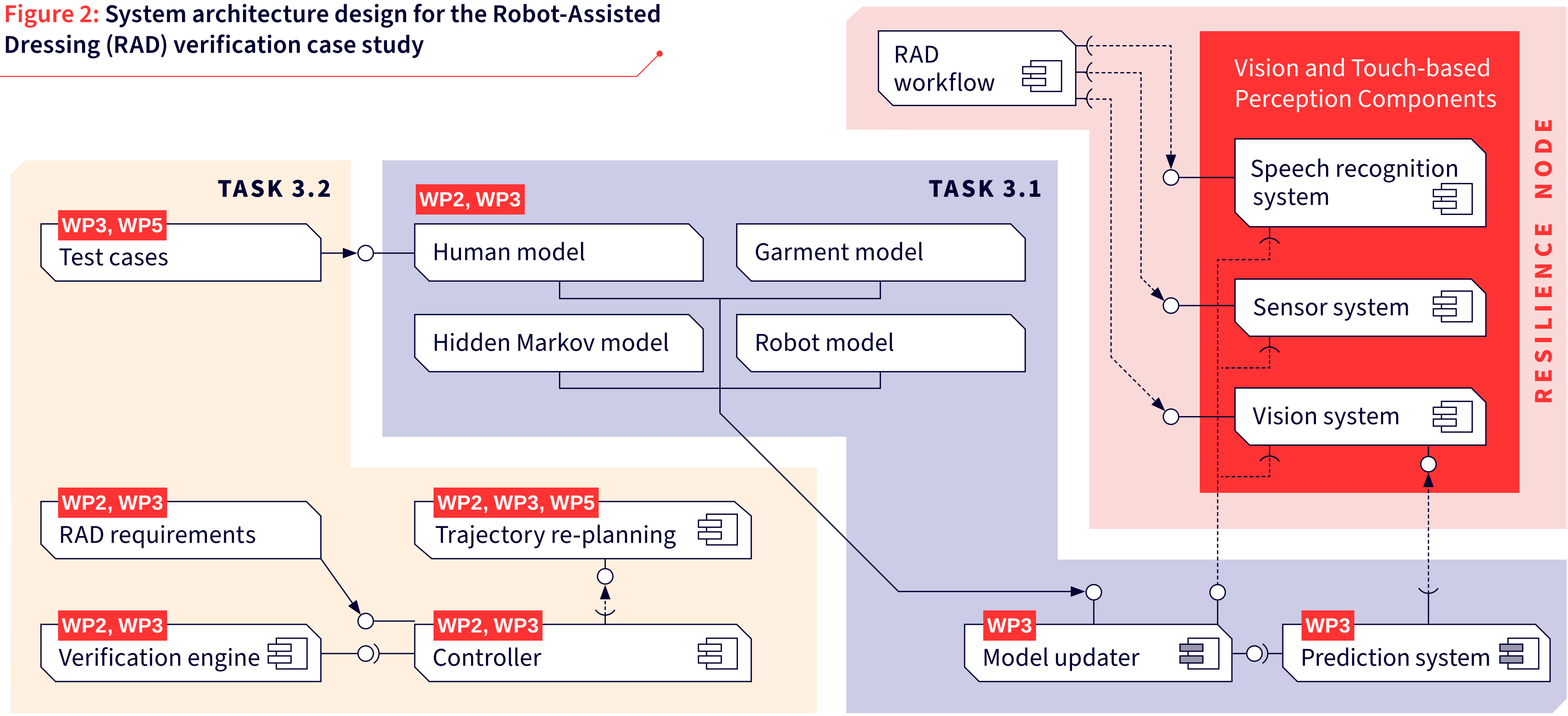
**The objective of this work package is to compose techniques meant for the verification of autonomous systems and develop guidelines to improve model verifiability and compositionality. These goals are to be achieved by means of model translation and transformation (horizontally across domains and vertically across layers of abstraction) whilst properties are preserved.**

## Year 1 Activities

Most of the work to date has been on the automated dressing case study, with the aim to produce models and verification conditions that can inform the development of integration approaches. The initial work involved the development of an approximate model of the robot (that applies the dressing) in the form of a probabilistic state machine (a Hidden Markov Model).

The other component to be reasoned about is the human and so we have recently explored the problem of producing a model of the human arm and describing how the arm can move. Once there are accurate models of the human arm, the sleeve, and the robotic arm, it should then be possible to, for example, verify that certain safety properties hold.

**Figure 2:** **System architecture design for the Robot-Assisted Dressing (RAD) verification case study**



# Compositions and translations continued

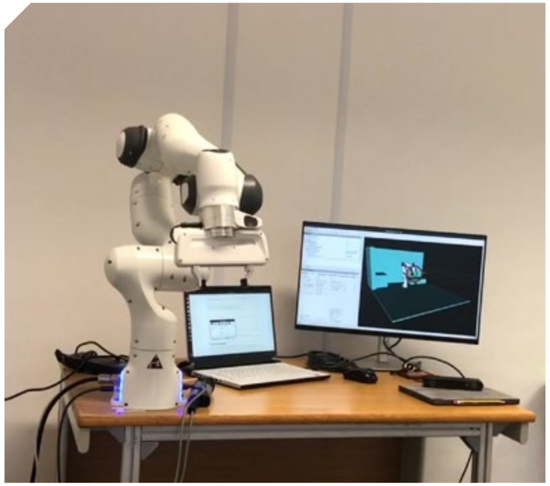
Robert Hierons and Yasmeen Rafiq (University of Sheffield)

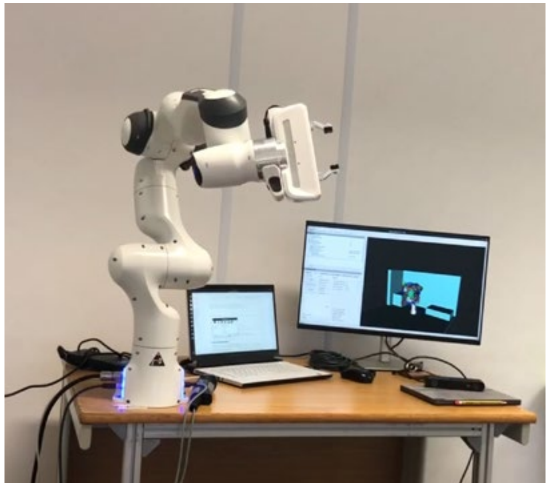
In addition to the development of initial models, there have also been collaborative meetings with members with Work Package 2 and (separately) members of the Resilience Node, who are also using this case study.

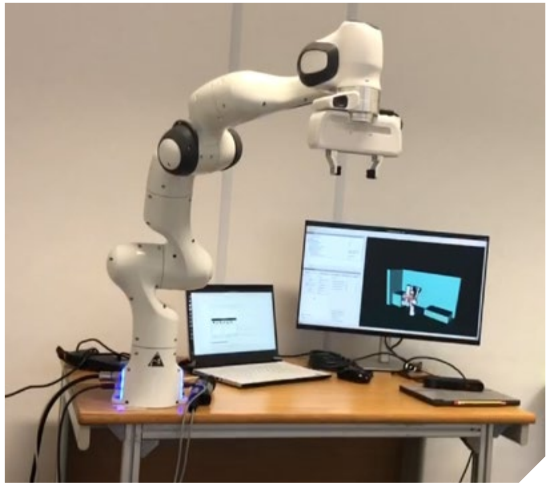
## Plans for Year 2

The main objectives for the coming year are to:

- Complete the model of the human arm and develop and implement techniques for predicting the position of the arm.
- Develop an approach that takes into account the uncertainty resulting from the sleeve obscuring the view of the arm.
- Work with members of the Resilience Node to develop an integrated model of the robotic system and human arm, with sufficient detail for this to be simulated.
- Work with Work Package 2 to explore how the models of the human arm and robot can be mapped to the semantic domain used and how this can be used to integrate verification.
- Work with Work Package 5 to verify safety properties of the (simulation of) the automated dressing system. Ideally, this work will also explore the problem of determining how 'close' the system can get to thresholds. For example, there is a specified maximum force F that the robotic arm can apply to the arm, and one might wish to not only know whether the system satisfies the requirement but also how close the applied force can be to the maximum F.
- Start collaboration with Work Package 8 to incorporate models of human behaviour.







*Assisted-care robotic arm*

# Closing the reality gap: physical platforms

Rob Richardson, Bilal Kaddouh, and Chengxu Zhou (University of Leeds)

## Overview

**This work package focuses on capturing the physical and computational properties of real autonomous systems. Its main challenge is to determine how to implement real-world environments and digital twins and also how to reconstruct the digital models into operational robots.**

## Year 1 Activities

Deconstruction of autonomous systems. Elements of this task have already started on the Firefighting UAV as they were required to feed into the work of Work Package 1 and Work Package 3. Even though this is meant to start in January 2022 the early start has given us a good insight into what is relevant to the verification framework and how to approach the question of deconstruction. This will be continued and ramped up once the PDRA starts.

### Case studies

A number of case studies have been collected from members of the consortium after a wide consultation. These case studies were presented to the consortium covering a wide range of autonomous systems examples. This is important so that members of the consortium can work on the same challenges in order for their results to be compatible and comparable with each other. Some of these have been taken on-board by members of the consortium to build models around them in order to experiment with methods and processes of verifying their safety. We have been directly involved in supporting the work being carried out in York with the modelling of the firefighting UAV in RoboSim and RoboChart. We have also been involved with the wider consortium particularly KCL (formerly Leicester) on the specific details of the implementation and modelling of the autonomy used on the firefighting UAV.

## Plans for Year 2

A number of questions are still open which we are working on answering to enable the work of consortium members to move forwards. We are also looking at providing a more mature scenario of multi-UAV applications in support of our colleagues in Manchester. At the moment we are holding bi-weekly meetings to go through the assumptions made during the modelling of the firefighting UAV case study and ensure the representation is accurate and realistic. We are also working on the physical modelling of the system. In the future we will be supporting the simulation of the firefighting UAV as well as the flight experiments of that case study and other case studies that will be picked up in order to evaluate the reality gap between the modelled and the real experiment. This will also be useful in order to evolve the models in order to be accurate enough to verify the safety of the autonomy used by the robots.

The work conducted so far has engaged with both Work Packages 1 and 5 in the details of the development of the Firefighting UAV. We plan to engage further with the rest of the node as well as other nodes and hubs in the future.

*Firefighting drone*

# Cyber-Physical systems

Mohammad Reza Mousavi, Hugo Araujo, and Jan Oliver Ringert (King's College London)
José Miguel Rojas and Uraz Turker (University of Leicester)

## Overview

**The goal of this work package is to cater for variability and uncertainty in the verification of cyber-physical systems (CPS). Additionally, we aim to bridge the abstraction layers in verification between discrete vs. continuous, synchronous – vs. asynchronous communications, stochastic vs. nondeterministic behaviour. Finally, we aim to exploit and analyse causal relationships in CPS in order to provide explanations of verification results.**

## Year 1 Activities

We carried out a systematic literature review on testing, verification, and validation of autonomous systems. We have successfully disseminated our review throughout the consortium and considered the feedback. We have been using the compiled knowledge obtained from the review to help conduct our research tasks, which are detailed below.

### Variability and Uncertainty

We have extended an existing model learning algorithm to consider new criteria in order to generate less false positives and negatives from occurring learning family models from product configurations. The new algorithm is also adaptive, with configurable weights being learned to cater for the subject domain. The results of this work are due to be published in the upcoming year.

### Compositional Conformance and Adequacy

We have proposed a methodology for filtering randomly generated test cases in order to make fault detection more efficient (Entekhabi et al. ICTSS 2021). For that, we make use of a DSL to specify test adequacy and filter the test suite based on relevant criteria. The strategy has been empirically evaluated on autonomous agents and implemented in QuickCheck. This work has been conducted in cooperation with Work Package 2.

As joint work with Work Package 6, we promote the use of AI to increase efficiency of testing strategies (Türker et al. ASE 2021). We propose the use of reinforcement learning techniques to generate synchronising sequences that lead the SUT towards a test initial state. This has great implications on the cost of executing tests derived from model-based testing techniques, as shown in the experimental results.

We have developed a runtime monitoring algorithm that searches for anomalies in the state space of the system (Dimitrova et al. LMCS 2022). We inflict perturbations on the input sequences and investigate their effects on the output to search for unreasonable behaviour. We plan to connect the results of this work with our model learning algorithm to make testing more efficient.

## Plans for Year 2

### Case study

We have engaged with the other work packages (Work Packages 1 and 4) on designing initial test models for the firefighting drone case study. We expect much collaboration and joint research emerging in the second year of the project.

### Causality and explanation in CPS Verification

We plan to devise a formal theory for analysing causality in cyber-physical systems. Existing theories only cater for discrete systems and are not equipped to deal with the intricacy of continuous (autonomous) systems. Based on our theory, we will develop analysis techniques that can be used to uncover the causes for counterexamples of failures resulting from verification techniques, and applying our strategy to a model of connected vehicles. This work has been conducted in collaboration with the Governance Node (Hana Chockler). This framework will be integrated with the results of the above-mentioned tasks and eventually incorporated into the Verifiability common framework.

# Sub-symbolic AI

Gavin Brown (University of Manchester) and Ivan Tyukin and Wendy Otieno (University of Leicester)

## Overview

**The goal of this work package is to provide proven defence in the design of AI components against physically viable adversarial attacks. For pre-deployed learning (AI) components, we would like to use the developed techniques to repair them without affecting their operation. All these techniques are meant to be applied in the presence of dynamic interactions with an adversarial environment in a cause-effect loop.**

## Year 1 Activities

To date, the focus of the work has been primarily on determining conditions specifying when verification of sub-symbolic AI models is possible using only input-output observations. We have shown that major faults can be successfully hidden within large-scale high-dimensional models. This provided input into our subsequent set of results, as it is now very clear that there is a subclass of problems which cannot be verified using off the shelf data approaches. Examples of algorithms hiding such faults, their corresponding tight probabilities of success, and formal statements along with the proofs have been provided.

This work, in addition to showing, for the first time, how challenging it is to verify modern AI models using classical input-output observations without looking into the model's latent spaces, also shows why backdoor attacks and vulnerabilities are practically possible in these systems.

In order to bring attention of the community to these challenges we published a SIAM News article, which in the first two weeks after publication attracted more than 3000 views. In October, a part of the work was presented to All Party Parliamentary Group on AI at an evidence meeting concerning national security: Regulation of AI-driven live facial recognition technologies (LFR).

The other direction of work has been concerned with the challenge of "repairing" malfunctioning non-symbolic AI models. A new theorem, the law of high dimensions, has been formulated and proven. This research led to the idea of exploring tools from explainable AI and counterfactual analysis for supervised cluster of errors.

## Plans for Year 2

1. Exploitation of explainable AI / counterfactual methods to build better AI correctors (bringing causal context into the picture) – phase 2. We will assess how new tools improve AI correction in various use cases ranging from automotive applications (jointly with Toyota) to healthcare.
2. Develop a statistical test for predicting AI instabilities of a given relevant type. This work will be based on our earlier results on AI instabilities and will enable testing for specific fault scenarios avoiding exhaustive exploration of models' inputs.
3. Work on case studies: provide examples of verifiable tests so that they could be checked within the scope of these case studies.
4. Since first deliverables have already been created, we will coordinate with relevant work package leads to ensure that they are informed and, as appropriate, guided by our findings.

# Symbolic AI

Michael Fisher, Louise Dennis, Joe Collenette [50% from July 2020] and Fatma Faruq [50% from July 2020] (University of Manchester)

## Overview

**The focus of this work package is to capture the required high-level behaviour of an autonomous system regarding rules, regulations, and ethics. Furthermore, the work package aims to provide tools and techniques to verify that Symbolic AI components (typically, agents making these key, high-level decisions) follow required behaviours, and adopt appropriate principles and priorities even when there is no prescribed behaviour available (or it is ambiguous) or there is significant uncertainty over the behaviour of sub-components.**

## Year 1 Activities

### Capturing Rules and Regulation

We have initiated activities regarding the representation and verification of human-level rules and regulations. For example within autonomous vehicles, this potentially requires enhanced logics and refined verification techniques. Our work in this direction has involved formalisation and verification of the “rules of the road” (especially junction regulations), and has led to two research outputs (Alves, Dennis and Fisher. ERS 2021 and JSAN 2021)

### Formalising and Verifying Ethics

Critical decisions, particularly where a range of ethical issues, formalisms and priorities are concerned must be described and assessed in any truly autonomous system. Ongoing work on this led to high-level ethical formalisation and verification in (Dennis et al. AAAI 2021) (Cardoso et al. EMAS 2021) and (Dennis and del Olmo. CME 2021).

### Verification and Trustworthiness for Autonomy

Capturing and clarifying issues of trustworthiness and verifiability (hence, regulation or certifiability) across autonomous systems is important. Consequently, we have refined the notion of "trustworthiness" from just being based on reliability (as is common in cyber physical systems) to being crucially dependant on "beneficiality". A system is "beneficial" if we believe it is acting to our benefit and has appropriate goals and intentions to ensure this. We also provide systems architectures, with associated verification techniques, that can be used as the basis for verifiably trustworthy systems.

### Runtime Verification

We have showcased the technology we have developed to carry out “runtime verification”, i.e. dynamic verification as the autonomous system is running, and are using this across an increasing variety of autonomous (especially robotic) systems. The tool provides runtime verification for Robot Operating System (ROS) nodes and so can be deployed on a range of robots and vehicles.

## Plans for Year 2

Our plans are to develop and extend work on:

- Wider, more complex, “rules of the road”, and the representation and verification of these within a decision-making agent for a “driverless car”;
- Human-Agent Teamwork, targeting relevant autonomous teamwork case studies, and their verification, as well as explainability through dialogue, such as the work being developed in (Dennis and Oren. AAMAS 2021);
- Runtime verification, specifically recognising “out of envelope” systems behaviour;
- Collaboration with the Security Node on verification focused by security analysis;
- Collaboration with a Responsibility project on the formalisation of responsibility;
- Continued work across standards, e.g. (Winfield et al. FRAI 2021), policy and regulators.

# Closing the reality gap: users

Effie Lai-Chong Law (Durham University) and Genovefa Kefalidou (University of Leicester)

## Overview

**The goal of this work package is to produce verification models of user behaviour in terms of human multisensory mental state in real-time, uncertain, and unconstrained scenarios. As part of its research goals, it aims to develop techniques to iteratively improve the user model by incorporating verification results and validate transparency in autonomous systems by examining human-system interaction patterns. Work Package 8 starts in Year 2 of the Node. A PDRA will be recruited to implement the Work Package tasks.**

## Year 1 Activities

### Exploration of HCI Models and Modelling Tools

Two promising modelling tools IVY (Jose Campos) and CIRCUS (Philippe Palanque) have been explored. Campos and Palanque presented their tools to the Node through the seminar talks. Currently, we are discussing further work on it.

A presentation on Work Package 8 was given in February 2021 to the Trust Node coordinated by Prof. Helen Hastie. An ECR in the Trust Node participated in the Dagstuhl Seminar 21381 *Conversational Agent as Trustworthy Autonomous System* (Trust-CA, 19-24 September 2021), which was organised by Law and other conversational AI experts.

## Plans for Year 2

We plan to systematically review the existing HCI modelling tools, including IVY and CIRCUS, to identify the most robust and viable ones for defining user models in the context of collaborative task analysis. To concretise the work, Conversational Agent (CA) will be used as a case to start with and subsequently generalise it to other use contexts with the goal of producing a domain-agnostic collaborative analysis framework. Findings of this review task will be fed forward to analysing interaction patterns between humans and autonomous systems, especially the understandability and perceptibility of multisensory signals, including voice, gaze, gesture, gait and other psycho-physiological data. User-based studies on defining and measuring the transparency in human-agent interaction will be designed and conducted. We plan a further workshop with the Trust Node in March/April 2022 when more findings of s and other Work Packages (e.g. Work Package 7) will be available.

### Conversational Agent for Case Study

The follow-up work based on Dagstuhl Seminar Trust-CA (please see outputs) will contribute to Work Package 8. Social interaction patterns between CAs and human users are relevant. CAs are increasingly deployed in a range of essential services in everyday life, including healthcare, finance, education and transport. Many of the participants of the Seminar are researching trust in CA different application domains. A workshop will be organized to discuss the modelling of trust in CA.

# Outputs

## Publications

### Research papers

Alves, G. V., Dennis, L., & Fisher, M (2021). An Agent-based Architecture with Support for Ethical Decisions on a Road Traffic Scenario. In *Proc. IROS Workshop on Building and Evaluating Ethical Robotic Systems* (ERS), 2021.

Alves, G. V., Dennis, L., & Fisher, M. (2021). A Double-level Model Checking Approach for an Agent-based Autonomous Vehicle and Road Junction Regulations. *Journal of Sensor and Actuator Networks, 10*(3).

Biewer, S., Dimitrova, R., Fries, M., Gazda, M., Heinze, T., Hermanns, H., & Mousavi, M. R. (2020). Conformance Relations and Hyperproperties for Doping Detection in Time and Space. *arXiv preprint arXiv:2012.03910*.

Cardoso, R. C., Ferrando, A., Dennis, L. A., & Fisher, M. (2021). Implementing Ethical Governors in BDI. In *Proc. 9th International Workshop on Engineering Multi-agent Systems (EMAS)*.

Chatila, R., Dignum, V., Fisher, M., Giannotti, F., Morik, K., Russell, S., & Yeung, K. (2021). Trustworthy AI. In *Reflections on Artificial Intelligence for Humanity* (pp. 13-39). Springer.

Damasceno, C. D. N., Mousavi, M. R., & da Silva Simao, A. (2021). Learning by sampling: learning behavioral family models from software product lines. *Empirical Software Engineering*, 26(1), 1-46.

Dennis, L. A., Bentzen, M. M., Lindner, F., & Fisher, M. (2021). Verifiable Machine Ethics in Changing Contexts. In *Proc. 35th AAAI Conference on Artificial Intelligence (AAAI)*.

Dennis, L. A., & Oren, N. (2021). Explaining BDI Agent Behaviour through Dialogue. In *Proc. 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. International Foundation for Autonomous Agents and Multiagent Systems.

Dennis, L. A., & del Olmo, C. P. A (2021). Defeasible Logic Implementation of Ethical Reasoning. In *Proc. First International Workshop on Computational Machine Ethics (CME)*.

Entekhabi, S., Mostowski, W., Mousavi, M., & Aarts, T. (2021). Locality-based Test Selection for Autonomous Agents. In *Proceedings of the 33rd IFIP International Conference on Testing Software and Systems (ICTSS 2021)*. Springer.

Fisher, M., Mascardi, V., Rozier, K. Y., Schlingloff, B. H., Winikoff, M., & Yorke-Smith, N. (2021). Towards a Framework for Certification of Reliable Autonomous Systems. *Autonomous Agents and Multi-Agent Systems 35*(1), pp1-65.

Fisher, M., Ferrando, A., & Cardoso, R. C. (2021). Increasing Confidence in Autonomous Systems. In *Proc. 5th ACM International Workshop on Verification and mOnitoring at Runtime EXecution* (VORtEX), pp1-4.

Foster, S., Hur, C. K., & Woodcock, J. (2021). Formally Verified Simulations of State-Rich Processes using Interaction Trees in Isabelle/HOL. *arXiv preprint arXiv:2105.05133*.

Følstad, A., Araujo, T., Law, E. L. C., Brandtzaeg, P. B., Papadopoulos, S., Reis, L., ... & Luger, E. (2021). Future directions for chatbot research: an interdisciplinary research agenda. *Computing*, 1-28.

Kulik, T., Dongol, B., Larsen, P. G., Macedo, H. D., Schneider, S., Tran-Jørgensen, P. W. V., & Woodcock, J. (2021). A Survey of Practical Formal Methods for Security. *arXiv preprint arXiv:2109.01362*.

Law, E. L-C., Følstad, A., Grudin, J., & Schuller, B. (under prep.). Dagstuhl Report on *Conversational Agent as Trustworthy Autonomous System*.

Thule, C., van de Pol, J., & Woodcock, J. (2021). Verification of Co-simulation Algorithms Subject to Algebraic Loops and Adaptive Steps. In *Formal Methods for Industrial Critical Systems: 26th International Conference, FMICS 2021, Paris, France, August 24-26, 2021: Proceedings* (Vol. 12863, p. 3). Springer Nature.

Tsvarkaleva, M., & Dennis, L. A. (2021, September). No Free Lunch: Overcoming Reward Gaming in AI Safety Gridworlds. In *Proc. International Conference on Computer Safety, Reliability, and Security* (pp. 226-238). Springer.

Türker, U., Mousavi, M., Hierons, R., & Tyukin, I. (2021). Efficient state synchronisation in model-based testing through reinforcement learning. In *Proceedings of the 36th IEEE/ACM International Conference on Automated Software Engineering* (ASE 2021).

# Outputs continued

## Publications (continued)

Tyukin, I. Y., Higham, D. J., Woldegeorgis, E., & Gorban, A. N. (2021). The Feasibility and Inevitability of Stealth Attacks. *arXiv preprint arXiv:2106.13997.*

Winfield, A. F., Booth, S., Dennis, L. A., Egawa, T., Hastie, H., Jacobs, N., ... & Watson, E. (2021). IEEE P7001: a proposed standard on transparency. *Frontiers in Robotics and AI*, 225.

Ye, K., Foster, S., & Woodcock, J. (2021, November). Automated Reasoning for Probabilistic Sequential Programs with Theorem Proving. In *Proc. International Conference on Relational and Algebraic Methods in Computer Science* (pp. 465-482). Springer, Cham.

J. Baxter, P. Ribeiro, and A. L. C. Cavalcanti. Sound reasoning in tock-CSP. *Acta Informatica*, 2021. online April 2021. [ bib | doi ]

S. Foster, K. Ye, A. L. C. Cavalcanti, and J. C. P. Woodcock. Automated verification of reactive and concurrent programs by calculation. *Journal of Logical and Algebraic Methods in Programming*, 121:100681, 2021. [ bib | doi ]

M. Zhang, D. Du, A. C. A. Sampaio, A. L. C. Cavalcanti, M. Conserva Filho, and M. Zhang. Transforming RoboSim Models into UPPAAL. In *15th International Symposium on Theoretical Aspects of Software Engineering*, pages 71--78. IEEE, 2021. [ bib ]

A. L. C. Cavalcanti. RoboStar Modelling Stack: Tackling the Reality Gap. In *1st International Workshop on Verification of Autonomous & Robotic Systems*, VARS 2021. Association for Computing Machinery, 2021. [ bib | doi | http ]

A. L. C. Cavalcanti, B. Dongol, R. Hierons, J. Timmis, and J. C. P. Woodcock, editors. *Software Engineering for Robotics*. Springer International Publishing, 2021. [ bib | doi | http]

A. L. C. Cavalcanti, W. Barnett, J. Baxter, G. Carvalho, M. C. Filho, A. Miyazawa, P. Ribeiro, and A. C. A. Sampaio. *RoboStar Technology: A Roboticist's Toolbox for Combined Proof, Simulation, and Testing*, pages 249--293. Springer International Publishing, 2021. [ bib | doi | pdf]

### Policy papers, standards

Chapman, A., et al. "Data: a new direction: A call for evidence from the Department for Digital, Culture, Media & Sport". (2021). [doi]

Winfield, A. F. T., et al. IEEE P7001: A Proposed Standard on Transparency. (2021) [http]

Ramchurn, Sarvapali, et al. "The future of connected and automated mobility in the UK: call for evidence." (2021). [doi]

Tyukin, Ivan. All Party Parliamentary Group on AI evidence meeting concerning national security: Regulation of AI-driven live facial recognition technologies (LFR), (2021).

Law, Effie, et al. IEEE SA P1228 – Standard for Software Safety (Dissemination) (2021) [http]

## Events

Louise Dennis: IJCAI 2021 tutorial on "Machine Ethics" (January, 2021)

Michael Fisher: CPS-IoT Week Workshop on Verification of Autonomous and Robotic Systems (May, 2021)

Michael Fisher: IROS'21 Workshop on Quality and Reliability Assessment of Robotic Software Architectures and Components (October, 2021)

Dagstuhl Seminar 21381 – Conversational Agent as Trustworthy Autonomous System (Trust-CA), 19-24 September 2021

# Outputs continued

## Keynotes / Invited talks

Louise Dennis, *Verifying Machine Ethics*, FMAS workshop, December 2020.

Louise Dennis. *Explicitly Ethical Agent Reasoning*, EMAS workshop, May 2021.

Michael Fisher, *Verifying Autonomous Systems – We need help from you!* CPS-IoT Week Workshop on Machine Learning in Control (LEAC), May 2021.

Michael Fisher, Increasing Confidence in Autonomous Systems, VORtEX workshop, July 2021.

Michael Fisher, Verifying Autonomous Systems, ORCA/SOLITUDE Workshop on Safety Assurance for Deep Learning in Underwater Robotics, September 2021.

Mohammad Reza Mousavi, Model-Based Testing and Model-Learning for Variability-Intensive Systems, Keynote at the 17th Workshop on Advances in Model Based Testing (A-MOST 2021), April 2021.

Mohammad Reza Mousavi, Learning About the Change: An Adaptive Approach to Automata Learning. Keynote at the 15th International Conference on Tests and Proofs (TAP 2021), June 2021.

Mohammad Reza Mousavi, Catch Me If You Can: Doping Detection in Cyber-Physical Systems, Keynote at the 33rd IFIP International Conference on Testing Software and Systems (ICTSS 2021), November 2021.

Mohammad Reza Mousavi, Trust in Autonomous Systems through Verifiability, IEEE Technical Committee on Verification of Autonomous Systems, December 2021.

Ana Cavalcanti, Invited talk at IROS2021 Workshop on Quality and Reliability Assessment of Robotic Software Architectures and Components.

Ana Cavalcanti, Invited talk at IROS2021 Workshop on Standardised Software Frameworks for Robotics in Nuclear

Ana Cavalcanti, Summer School organised by Zhimming Liu at the School of Software, Northwest Polytechnical University, Xi'an, China.

Jim Woodcock, Summer School organised by Zhimming Liu at the School of Software, Northwest Polytechnical University, Xi'an, China.

Ana Cavalcanti, Invited talk at ABZ on RoboStar technology

Ana Cavalcanti, invited talk at the workshop on Verification of Autonomous & Robotic Systems (VARS) as part of the CPS-IoT Week 2021.

Ana Cavalcanti, seminar on Software Engineering for Robotics as part of the women+@DCS group at The Department of Computer Science, University of Sheffield.

I.Tyukin, A.N. Gorban. The mathematics of learning from high-dimensional low-sample data with small neural networks, December 1, 2021. Workshop on Complex and simple models of multidimensional data : from graphs to neural networks.

I.Tyukin. The challenge of trustworthy data-driven Artificial Intelligence: boundaries, limitations, and opportunities, October 18–19, 2021 [keynote], 2-nd EAI International Conference on IoT and Big Data Technologies for HealthCare.

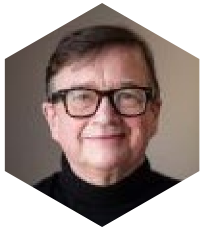
I.Tyukin. Breaking into Deep Learning models, October 5, 2021, Isaac Newton Institute. Mathematics of Deep Learning.

I.Tyukin. Breaking into a Deep Learning box, April 10 – 11, 2021. ICERM Safety and Security of Deep Learning

# Advisory board

**Dejanira Araiza Illan**
Assistant Principal Engineer in Robotic Applications at Johnson & Johnson.

**Raja Chatila**
Professor Emeritus, Institute of Intelligent Systems and Robotics.



**Sebastian Conran**
CEO of Consequential Robotics.

**Danit Gal**
Associate Fellow at The Leverhulme Centre for the Future of Intelligence

**Koen Hindriks**
Professor Artificial Intelligence, Social AI group, Vrije Universiteit Amsterdam.

**Joost Noppen**
Principal Researcher Software, BT Applied Research.

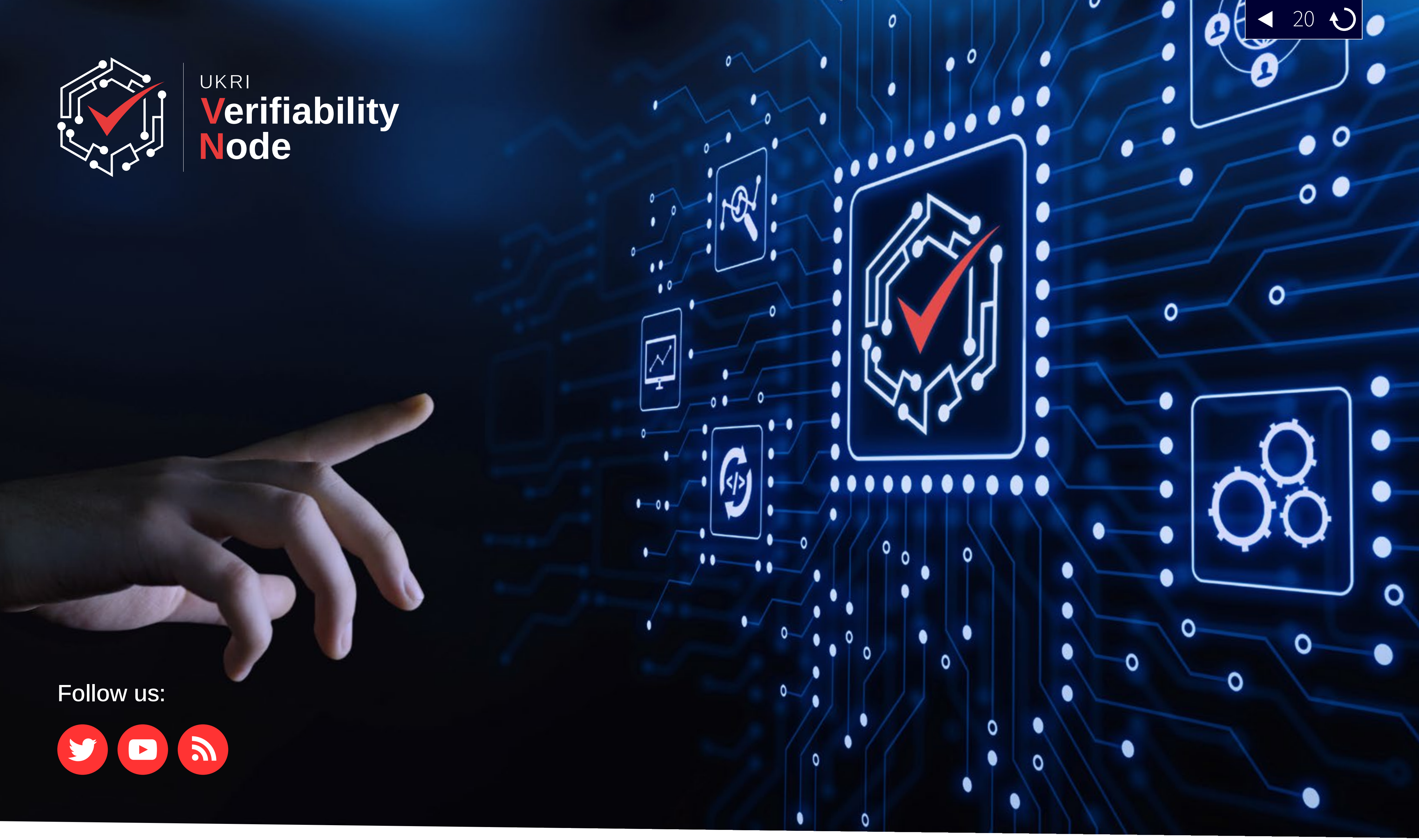
**Colin O'Halloran**
Technical Director of D-RisQ Ltd.

**Kristin Yvonne Rozier**
Associate Professor, Department of Aerospace Engineering.

**Rich Walker**
Managing Director, The Shadow Robot Company.

**Henry Tse**
Connected Places Catapult, the UK's innovation accelerator for cities, transport, and places.

**Thierry Lecomte**
R&D Project Director at CLEARSY, French SME specialised in safety critical systems.



UKRI
**Verifiability Node**

Follow us:

## Our Project Partners

UKRI Engineering and Physical Sciences Research Council
Durham University
King's College London
University of Leeds
University of Leicester
Manchester 1824 The University of Manchester
The University Of Sheffield.
University of York