UKRI
Verifiability
Node

UKRI Trustworthy Autonomous Systems Node on

Verifiability

SECOND ANNUAL REPORT
2022-2023

Contents

UKRI Verifiability Node

Foreword

Welcome to the second edition of the Verifiability Node annual report. Last year was an exciting year for Verifiability. I would like to summarise the year in two words: “maturing” and “integration”. Our foundational results matured further and many results were published in the flagship venues of the diverse set of disciplines involved in the Node. Also we set as our priority to use our common case studies as vehicles to integrate the various developed techniques across the Node.
This is clearly reflected in the report before you. We connect everything that has happened in the Node already in the introduction, through our case studies. We also report on several in-depth results

In addition, the Node has gained much visibility and as a result was involved in many public and policy engagement venues. You will also see a glimpse of these in this edition of the report.

All in all, I like to think that we have delivered on much of the promise of the Node so far and I am very much looking forward to the final year of the Node where the results will bear fruit and will be further integrated into our Verifiability platform.

I hope you enjoy reading through our second report and we will get the opportunity of collaborating with you during the forthcoming final year of the project and beyond!

Mohammad Reza Mousavi

Professor of Software Engineering
Department of Informatics
King’s College London

Principal Investigator of the UKRI Trustworthy Autonomous Systems Node in Verifiability

King's College London

UKRI Engineering and Physical Sciences Research Council

Introduction

Research

We recall the structure of the workplan for the Verifiability Node in Figure 1. We have established a firm semantic foundation in work packages 1 to 3 encompassing aspects such as probabilistic modelling of uncertainty and modelling sub-symbolic and symbolic AI. This sets the scene for several work packages to connect their results through the common semantic framework and the modelling abstractions provided by these work packages.

The initial integration that has taken place during the second year of the project is presented in Figure 2. This figure shows the workflow of the models and techniques used across the Node's work-packages. Here, we show how we cover the heterogenous nature of autonomous systems and provide multiple avenues for checking system properties, leading to a holistic judgement about the system.

Starting from the system requirements and system properties (which are artefacts typically written in natural language), a formal design model (using the RoboStar* framework) can be built based on the specified design. The system controller (RoboChart) is built based on the requirement specifications, assurance cases, and human agent models (Circus). From this point, formal verification techniques (such as proofs in Isabelle/UTP and model checking in FDR and nuXMV) are being used to offer mathematical assurances about the system. Furthermore, discrete unit testing and hybrid system testing mechanisms are being developed for RoboChart models.

In order to capture the physical aspects inherent to robotics, the RoboChart controller is extended to take into consideration the kinematics and dynamics of the system under test, of the other agents, and also the environment; the combination of the (discrete) controller behaviour and the (continuous) physical aspects are captured in RoboSim models. Such models can be used to automatically generate SDF models, which, in conjunction with simulation code, are fed to popular robot simulators/middleware (e.g., CoppeliaSim and ROS). Simulation takes a vital role in the testing of robots due to practicality. With simulation, one can verify the continuous aspects of the system's correctness and apply corrections before the physical system is used. For instance, in this Node, we consider system-wide conformance/mutation testing (using Matlab/Simulink), runtime monitoring (ROS Monitoring), and adversarial attacks checks. In summary, we offer evidence of verification based on multiple verification paradigms; in turn, the obtained evidence is directly applied into the development of the physical system.

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

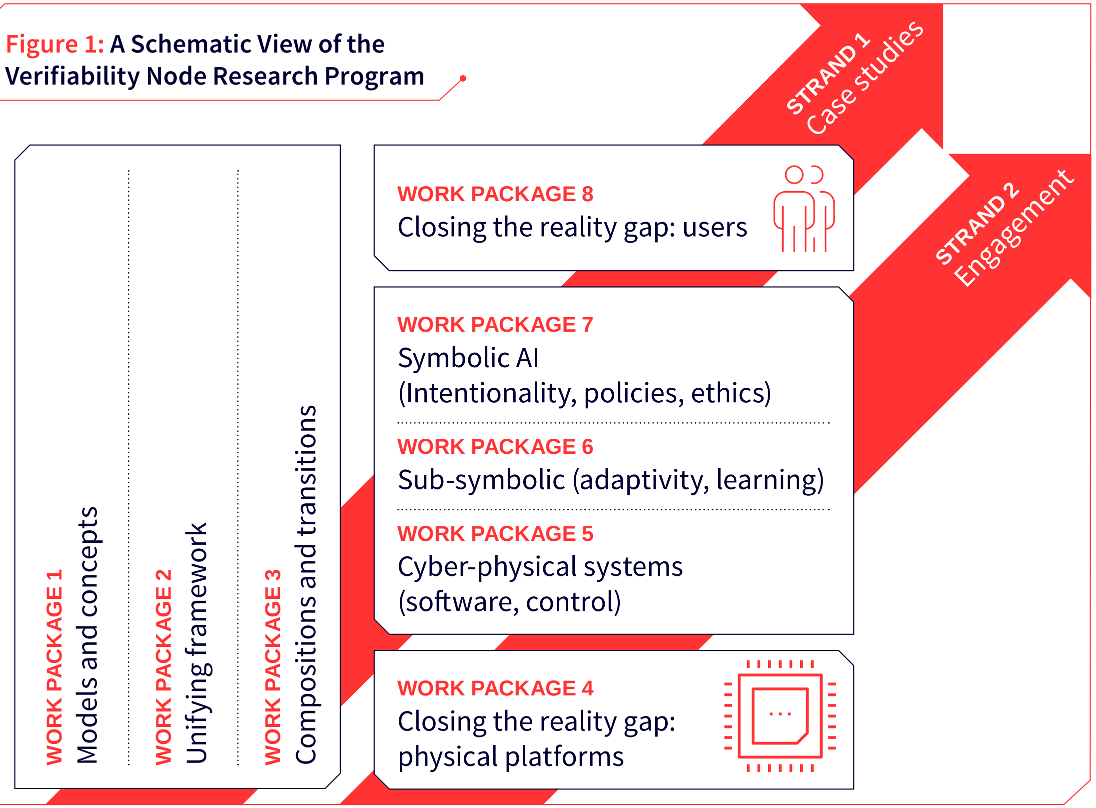
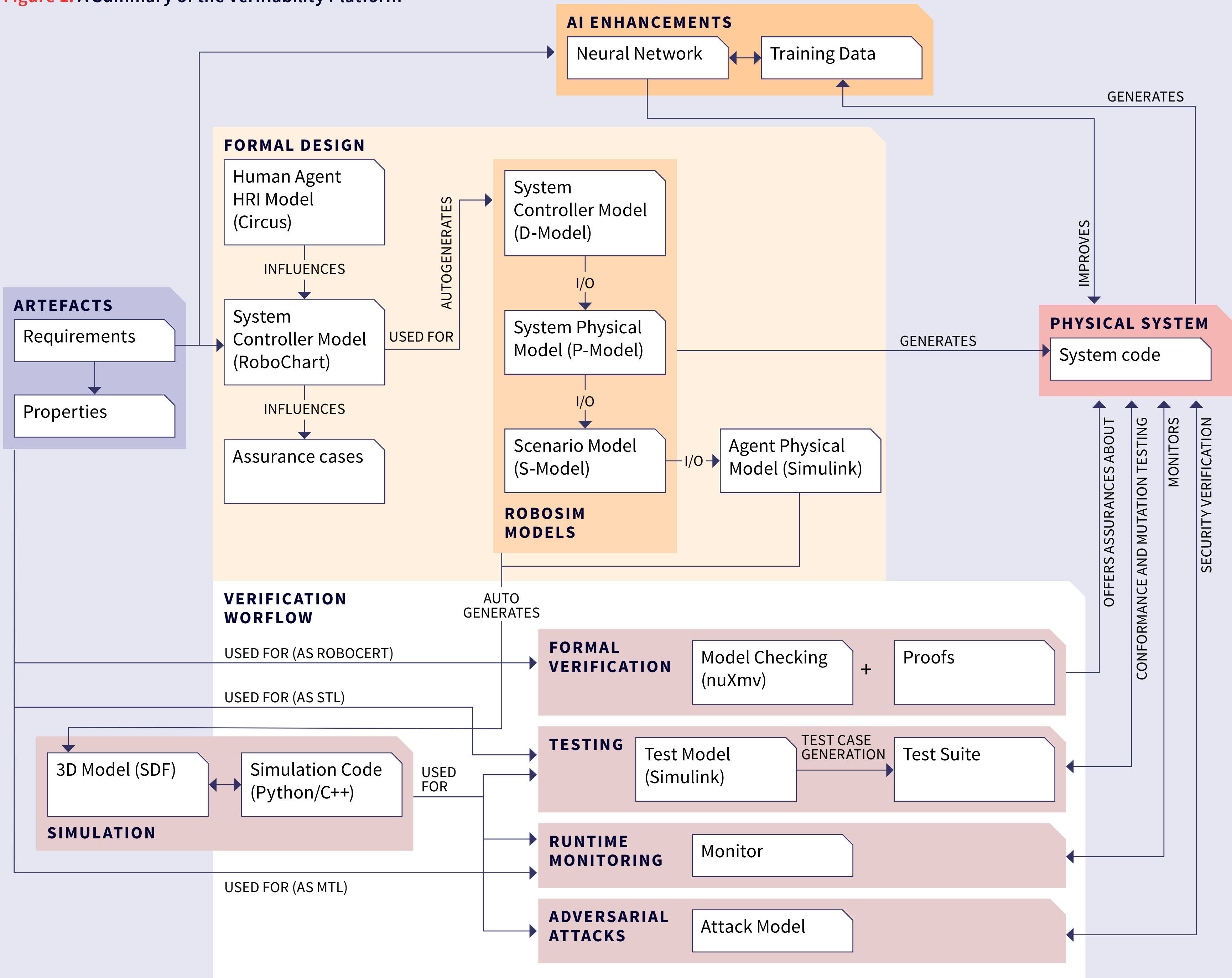


Figure 1: A Summary of the Verifiability Platform



Introduction continued

Community Building and Public Engagement

In the past year, the Verifiability Node has been actively involved in various community-building and public-engagement activities, including reaching out to schools, publishing popular science articles in journals and magazines, as well as organising events for community building and engagement.

Regarding public engagement, Mousavi was featured in an interview with City Transport and Innovation (CiTTi) Magazine, entitled “Sleep at the Wheel) (see Figure 1). There he made a case for distinguishing the different use-cases for autonomy on the road and the level of autonomy needed to make positive societal, economical, and environmental impact.

Figure 3: **“Asleep at the wheel?” Interview with Mousavi, CiTTi Magazine, November 2022**

Asleep at the wheel?

The UK government wants driverless cars on public roads by 2025 following the release of a £100m plan that prioritises safety through new laws. But how likely is this? **Ursula O’Sullivan-Dale** reports…

UKRI Verifiability Node

Contents
Foreword
Introduction
Common Case Studies
WORK PACKAGE 1: Models and Concepts
WORK PACKAGE 2: Unifying Framework
WORK PACKAGE 3: Translations and Refinement
WORK PACKAGE 4: Reality Gap
WORK PACKAGE 5: Cyber-Physical Systems
WORK PACKAGE 6: Sub-symbolic AI
WORK PACKAGE 7: Symbolic AI
WORK PACKAGE 8: Closing the Reality Gaps: Users
Key Outputs
Advisory Board

Introduction continued

Community Building and Public Engagement

Dennis and colleagues initiated a series of Lego Robot programming activities at schools (see Figure 4) and museums. These events were attended by a total of about a thousand young people and prompted engagement from several teenagers who took part in solving problems and programming tasks.

The Verifiability Node CoIs published a joint position paper on trustworthy autonomous systems in the IEEE Computer Magazine.

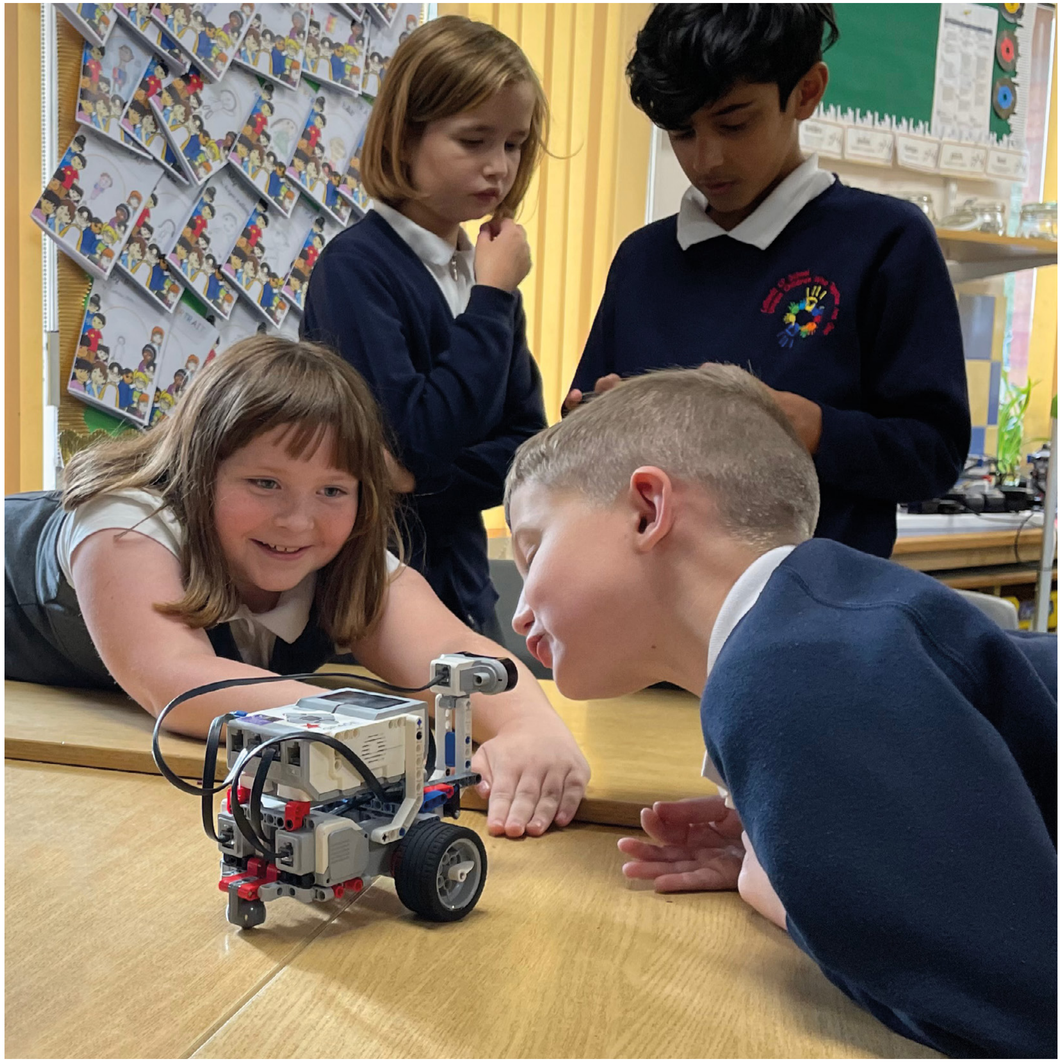
We continued the tradition of Verifiability Talks and organised 55 instances to date. Many of the talks are also featured on our YouTube channel with more than 120 subscribers and more than 3500 views of the featured talks.

For example, in October 2022, we hosted a delegation of PhD students and their supervisor from the Wallenberg AI, Autonomous Systems and Software Program (WASP) in Sweden. The one-day visit was filled with lively discussions on autonomous systems and concluded with screening "The First" and thought-provoking discussion with Luca Viagnò.

As promised in our proposal, we will be organising the 18th edition of the International Summer School on Training And Research On Testing.

We maintain two public Github repositories making our models of the two common case studies publicly available.

Figure 4: Lego Robot Programming Activities



Introduction continued

Policy Engagement

Regarding policy making, we have been involved in the IEEE P7001 proposed standard on Transparency, and the IEEE SA P1228 Standard for Software Safety.

We have been engaging with the transport industry, academic institutions and the Government (particularly with the Department for Transport, DfT) for the past year regarding policy making for autonomous vehicles. In particular, we engaged in an activity led by WMG Warwick and commissioned by DfT to formulate a Cross-Domain Safety Assurance Framework in Transport, which resulted in a report (available here) and a fascinating Panel discussion at the report launching.

We also organised a workshop on Maritime Autonomy, in collaboration with the Connected People and Places Catapult, the TAS Hub, and the Institute for Safe Autonomy. The workshop attracted participants from industry and academia and the report was shared with the participants.

Dennis and Mousavi participated in the ITF Roundtable on Policy and Regulation for Autonomous Vehicles held at OECD in Paris. The Roundtable comprised policy makers, academics, and representatives of public and private organisations. The report of this roundtable is to appear and be shared with the ITF member states soon.

The Verifiability Node has continuously contributed throughout on the aspects of Test Environments for establishing safety assurance and on Test Scenarios and Safety evidence.

Communication for distilling trust and safety perceptions to ASs in a cross-domain transport context.

This is an engagement that is planned to continue in the next period as well. Through our Hub Liaison's Officer, Menendez, we engaged actively with the Hub events, presented our results and participated in the Hub Sandpits. The Node engaged in the TAS-Hub Policy Lab focused on social inclusion and autonomous vehicles, to identify and address the challenges associated with the adoption of autonomous systems. This resulted in a publication on Policy for social inclusion and autonomous vehicles.

We continue to work with TAS Hub on RRI and TAS EDI Framework to ensure inclusivity and responsibility in conducting research with Autonomous Systems (ASs); this is an ongoing engagement, thanks to Kefalidou as the co-chair of the TAS Hub EDI Group.

Common Case Studies

Summary

Two main case studies are running at the moment; the UAV Firefighter and the Robot Assisted Dressing. The overall objective for the case studies is to provide a focal point for the efforts of the consortium members to develop and to test and demonstrate their developed processes on a real autonomous system.

We have already made significant strides towards developing a state-of-the-art UAV for use in firefighting scenarios, while the robotic assistive dressing project aims to create a novel robot-assisted dressing system to help people with disabilities or mobility issues. In both cases, the consortium partners have worked closely together to pool their expertise and resources and achieve their goals. The progress made thus far has resulted in a joint paper on the firefighting UAV and promising developments in the robotic assisted dressing project. We are excited to see what other breakthroughs will emerge from these fruitful collaborations. Both cases studies have also been tackled in work in collaboration with the Resilience Node.

Firefighting UAV

The collaboration on the Firefighting UAV case study has produced a RobotChart Model that is being developed and refined as well as a physical model of the platform in RoboSim, and an account of operational requirements in RoboWorld. It has driven the process of developing verifiable test cases with experimental validations. The verification procedure is looking at 3 levels: Component test, System level tests and Runtime verification. Combining the three layers will provide an insight into the ability to verify the autonomous system in an efficient process.

This process is being supported by the development of a simulation environment in ROS which allows the team to conduct the tests and verify the test cases in a safe and rapid environment.

In parallel, the UAV hardware has been developed and tested particularly on the fire detection and targeting subsystems. Field experiments have been conducted to evaluate the performance of the implemented system before fully integrating it into the UAV's autonomous system. Further development and implementation work is being carried out ensuring the models and implemented systems match in order to allow for a successful verification activity.

Future work related to this case study is described in the plans of the work packages further into the report.

Robot Assisted Dressing case study

The Robot Assisted Dressing case study has involved adapting different sources of domain knowledge on assisted dressing to support the development and assurance of an automated robotic assistant. A series of dressing trials have been carried out in collaboration with the Resilience Node to explore the techniques used by Occupational Therapists to manually support stroke patients with limited mobility in their left arm with dressing and to investigate the impact of external distractions. We were able to collect a rich dataset of various dressing scenarios as an outcome from the human-human trials. The dataset is used as input to a Convolutional Neural Network to reconstruct 3D patient's pose for the various dressing trajectories and to update the multibody model of the human subject, expressed as a URDF file. A VICON system was also used as the ground truth. Furthermore, we have developed formal hazard and requirements documents based on a prior structured hazard identification process.

These activities have then supported the modelling and verification activities of the case study, which span the physical components of the dressing scenario as well as the control software which directs the dressing process. The control software of the dressing robot has been captured as a detailed model using the RoboChart framework, allowing us to verify a wide set of safety specifications based on the hazard analysis results and captured in the RoboCert language. We are also to apply runtime monitoring techniques based on the ROS monitoring framework, to verify that the dressing process satisfies crucial safety properties at runtime. Collectively these activities allow us to develop a design for the dressing robot which meets its safety requirements, alongside runtime monitors to ensure that these requirements are maintained in the implementation.

Models and Concepts

WP Lead: Ana Cavalcanti Collaborators: Matt Windsor, Pedro Ribeiro, Alvaro Miyazawa,James Baxter, Holly Hendry, Cade McCall, Ziggy Attala, Arjun Badyal, and Will Barnett. In addition, we have collaborated with colleagues in Brazil, Leeds, and Thales.

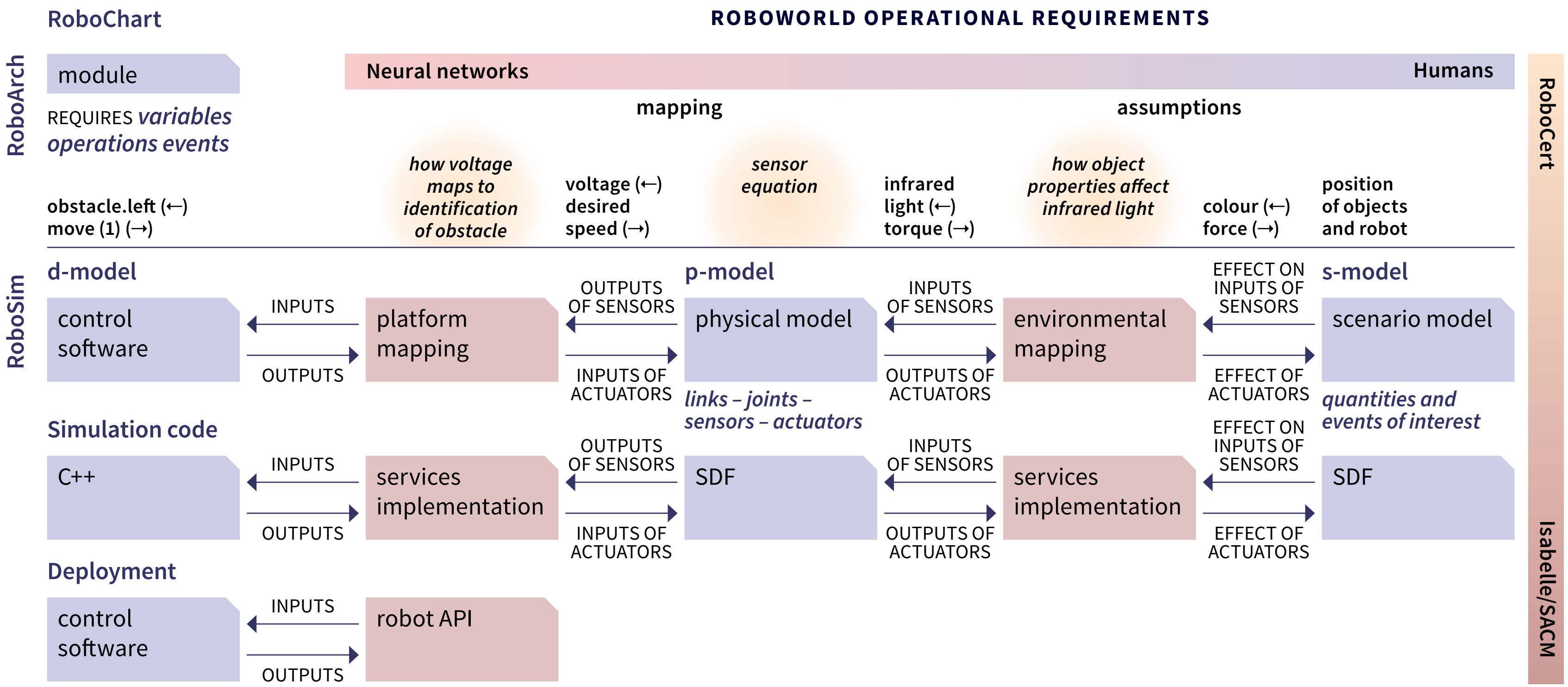
Summary of Overall Objectives

The objective of this work package is to identify linguistic concepts for design and verification of autonomous systems. Our aim is to identify relevant existing notations, appealing to practitioners, and enrich them with mathematical semantics to support verification. Here, we take the view of verification as covering activities based on proof (via model checking and theorem proving), simulation and testing.

Planned Activities for Year 2

- Finalisation of the definition and mechanisation of RoboCert, a diagrammatic domain-specific notation for definition of properties, and submission of associated conference and journal paper.
- Definition of a model-to-model transformation to connect design and simulation models (as captured in RoboChart and RoboSim).
- Design of a controlled natural language to capture operational requirements, RoboWorld.
- Preliminary work on verification in the presence of neural networks.
- Development of notational support to co-verification using simulation and proof.
- Development of notational support to describe architectural patterns for mobile and autonomous robotics systems.

Achievement and Result Highlights



Models and Concepts continued

RoboCert

RoboCert is now fully defined (metamodel, well-formedness conditions, semantics) and has a plug-in that works in the context of RoboTool, the tool that supports modelling and verification. One of our joint case studies with the Resilience Node, the Assisted Dressing Robot, is being tackled with the help of RoboCert.

From RoboChart to RoboSim

We have now devised and implemented a model-to-model technique to translate design models (RoboChart) to simulation models (RoboSim) to support verification by simulation. This work has been applied to both our case studies, the firefighter UAV, and the Assisted Dressing Robot. Another case study, a fruit picker, has also been addressed. A journal paper is in preparation. To complement this work, we have also devised a second model-to-text transformation to generate code for the RoboSim models. This technique is partially implemented and is being applied to our case study and to a new case study on swarm robotics.

RoboWorld

We have concluded the design of RoboWorld, including its metamodel, concrete grammar, semantics, and tool support. We have applied to a simplified version of the firefighter UAV. A journal paper has been submitted.

Human-in-the-Loop Modelling

In collaboration with WP1, we have carried out interviews with colleagues in industry to ascertain the needs for a notation specifically targeted at capturing timed behaviours of humans interacting with robots. A preliminary version of the notation is available, and further study of requirements has used a search-and-rescue robot put forward by our industrial partner.

Neural Networks in RoboChart

We have extended our design notation and its semantics to cater for components implemented using neural networks. The semantics is predicative, using a process algebra called Circus. We are now working on a proof strategy to support system-level verification in the presence of neural networks. With the given semantics, we are exploring a combination of theorem proving with support of an existing tool tailored for neural networks, namely, Marabou. A conference paper is in preparation.

Co-verification: simulation and proof

Physics engines used in simulation are mostly black boxes when it comes to the equations and algorithms that they use to simulate physical behaviour. This prevents the provision of consistent verification evidence based on simulation and proof, as we cannot guarantee that the models used in proof are compatible with those used in simulation. We are developing a framework to support use of physical RoboSim models for co-verification based on simulation and proof. We are considering the simple example of a robotic arm. To illustrate how our framework may be adapted for existing software, we have adopted Drake's physics engine (multibody plant and scene graph), because it supports greater modularity in implementations, compared to others used in robotics, and there is a thorough account of the equations it implements. These features make Drake more configurable and closer to achieving transparency. A conference paper is in preparation.

RoboArch

We have designed and formalised RoboArch, a notation to describe common architectural patterns in robotics. It supports the widely used layered pattern and the description of layer patterns. It is an extensible notation with semantics given via RoboChart models.

Internode work: we have also worked with the Resilience Nodes colleagues to tackle the social, legal, ethical, empathetic, and cultural aspects of our firefighter example. We have published our results in a short tool paper, and a journal publication is under preparation.

Planned Activities for the Final Year

In the final year, we will focus on case studies to consolidate and demonstrate the impact of our results, creating also a roadmap for future work.

- Significant case study on the use of RoboCert.
- Significant application of the model-to-model transformation technique for RoboChart and RoboSim.
- Significant case study on the use of RoboWorld.
- Significant case study on verification of a system with humans in the loop, in partnership with Thales.
- Conclusion of the strategy for verification in the presence of neural networks.
- Case study on co-verification using simulation and proof.
- Significant case study on agricultural robotics.

WORK PACKAGE 2:

Unifying Framework

WP lead: Jim Woodcock Members: Thomas Wright

SEMANTICS

✓-Tock CSP
+
Circus
↓
Tock Circus (WP2)

Unifying Framework

LANGUAGES	VERIFICATION TECHNIQUES
RoboChart	Model Checking
RoboCert (WP1)	Runtime Verification (WP7)
Isabelle/SACM	Mutation Testing (WP3)

Robot Assisted Dressing Case Study

Summary of Overall Objectives

The overall objective of WP2 is to define a rigorous *unifying semantic framework,* supporting the heterogeneous nature of autonomous systems, and providing a basis for relating and interconnecting various modelling techniques and tools. As a secondary focus, this work package has contributed towards a Verifiability Node case study, modelling an assisted dressing robot, under development within the TAS Reliability node, which aims to help stroke patients getting dressed. This case study serves to explore the applicability of the RoboStar framework to a novel healthcare robotics application, and apply a range of verification notions and techniques to assuring its safety.

Planned Activities for Year 2

- UTP semantics for ✓-Tock CSP: As its foundation RoboChart builds upon a combination of Tock CSP as a theory of timed reactive processes, and Circus, which extends CSP with facilities for modelling the mutable state of components. Baxter, Ribeiro, and Cavalcanti introduced ✓-Tock CSP (4), an extended semantics for CSP to support modelling deadlines and timeouts, both of which play important roles in specifications of timed robotic systems, and hence provide a richer backend for RoboChart modelling of robotic systems. We propose developing a new “Tock Circus” UTP semantics which integrates both ✓-Tock CSP processes and Circus-style state within our unifying semantics. Additionally, building upon the work of Foster et al (5), we propose developing a reactive contract theory for this semantics, enabling systematic techniques for reasoning about ✓-Tock Circus processes, and a mechanisation of these techniques within the Isabelle/UTP framework enabling formal and automated reasoning.
- Robot Assisted Dressing Case Study: We aimed to model the control software of the assisted dressing robot within RoboChart, whilst using a combination of specification and verification techniques to document and verify the safety requirements of the robot based on a prior hazard analysis (96) for the case study. Additionally, we aimed to build a 3D model of the Franka-Emika robotic platform underlying the case study, utilising the RoboStar framework’s “RoboSim Physical Model” support (10).

Unifying Framework continued

Achievement and Result Highlights

Tock Circus Semantics and Reactive Contract Theory

We have developed the Tock Circus theory which encompasses many of the core operators of ✓-Tock CSP within the UTP framework and mechanised this theory within Isabelle/UTP. This is implemented as a design contract theory, providing a number of healthiness conditions which make it possible to express processes as a timed extension of *reactive contracts* (8), separating specifications into triples consisting of *preconditions, postconditions,* and *preconditions,* providing reasoning rules for combining these contracts, and developing a combination of tactics and reasoning rules for proof automation. Additionally, we have developed a translation from this theory back to the existing refusal-trace–based semantics of ✓-Tock CSP to provide the consistency of many of the definitions of the new semantics with ✓-Tock CSP.

Robot Assisted Dressing Case Study

For the case study we have developed a detailed RoboChart model for the dressing controller alongside a RoboSim physical model of the Franka-Emika robotic arm. This provides one of the largest case studies of the RoboChart modelling tool, and has aided the development and testing of translation between RoboChart models and RoboSim simulations of robotic systems. We have also collaborated with WP1 to apply the RoboCert specification language to capture the safety requirements of the system, and verified many of these requirements using the FDR4 model checker. We are also collaborating with WP3 to apply mutation testing techniques to the RoboChart model to generate additional test cases, and with WP7 to apply runtime monitoring techniques to assure the safety of the robot at runtime. Finally, we have encoded the hazard analysis results for the case study within the Isabelle/SACM framework, providing structured documentation of the assumptions underlying our verification approach.

Verification of heterogeneous autonomous systems

We have carried out some additional external collaborations to apply novel combinations of verification techniques to the heterogeneous autonomous systems which are the focus of this work package. Firstly, we developed a method for verifying the safety of self-adaptive digital twin systems based on Signal Temporal Logic (STL) specifications in collaboration with Claudio Gomes et al. Secondly, in collaboration with Paulius Stankaitis, we implemented and evaluated a range of techniques for verifying the dynamics of black box Functional Mockup Interface (FMI) systems, in which we have limited information about the internal dynamics

Planned Activities for the Final Year

In the final year, we plan to conclude and up the work on the assisted dressing case study and the Tock Circus semantics, and to write each of these up as high-quality conference papers. Additionally, we intend to continue our collaborations with WP1, WP3, and WP7.

As part of this we are working towards leading the following publications:

- Unifying theories of timed reactive contracts: A paper detailing the Tock Circus UTP semantics.
- RoboChart modelling of an Assisted Dressing Robot controller. A paper detailing the assisted dressing robot controller.

And plan to contribute towards the following publications in collaboration with other work packages:

- Windsor M, Cavalcanti A, Wright T, Woodcock J. RoboCert: Property Specification for Robotics. **[In Preparation for International Journal on Software and Systems Modeling (SoSyM)].**

We will present our use of RoboCert in the Robot Assisted Dressing case study, as part of WP1's journal paper on the RoboCert specification language

- Yasmeen Rafiq, Thomas Wright, Jim Woodcock, and Rob Hierons. Mutation Testing of an Assisted Dressing Robot.

Translations and Refinement

WP Lead: Rob Hierons   Collaborators: Yasmeen B. Rafiq

Summary of Overall Objectives

The objective of this work package is to integrate verification techniques. This is being done via a case study, developed in collaboration with the Resilience Node, and so much of the initial work has been around this case study.

Planned Activities for Year 2

- *Extending the Assisted Dressing Case Study:* The first challenge was to complete the model of the human arm and develop and implement techniques for predicting the position of the arm. We also aimed to develop an approach that takes into account the uncertainty resulting from the sleeve obscuring the view of the arm.
- *Integrate models of the arm and robotic system:* We intended to 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. This was then to be used, 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.
- *Define safety properties:* 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.
- *Incorporating human behaviour into the case study:* Start collaboration with Work Package 8 to incorporate models of human behaviour.

Achievement and Result Highlights

Developments within the assisted dressing case study

The work carried out includes user studies that included seven individuals and several different scenarios. For each, we have both video, which represents the type of information that a robotic system might have, and also the locations of markers through the trials (representing ground truth). We have found that a neural network can process the visual data to provide a good approximation to the ground truth.

Complete Testing

We have devised a complete test generation technique that takes as input a set of refusal traces that forms the semantics of a design model. The decision to use refusal traces rather than, say, traces or failures was motivated by several factors. First, refusal traces provide the richest semantics that is consistent with a realistic testing scenario. Second, it has been observed that refusal traces provide the expressiveness required for systems with discrete time.

Mutation Testing framework

We have an initial testing framework and tool chain for the use of mutation testing.

Mapping reactive models to cyclic models

Models in languages such as RoboChart are reactive but simulations and deployed systems to test are cyclic. This introduces challenges since a test case (trace or refusal trace) generated from a reactive model does not directly correspond to a possible behaviour of a system under test. We have defined a mapping from forbidden traces of a RoboChart model, which potentially correspond to test cases, to traces of a cyclic model. Interestingly, we have found that some forbidden traces of a reactive model get mapped to traces that are not forbidden in the cyclic paradigm. We are currently devising a test suite generation technique that produces, from a reactive model, a test suite that is complete for a cyclic implementation. This work has identified some practical issues that may lead to testability properties for a reactive model. For example, there are conditions under which one can apply transformations to the reactive (RoboChart) model in order to simplify test generation.

Translations and Refinement

Mapping the semantics of a RoboChart model to a canonical state-machine

The aim here was to build upon the model-independence work of Wen-ling Huang and Jan Peleska, in which a model is mapped to an LTS that represents the semantics (set of traces) of that model. The motivation for this previous work is that testing is then based on the semantics of the model and not its structure, potentially leading to the generation of complete test suites (with guaranteed fault detection capability) from this semantics. For RoboChart models, we need more than traces: we need to consider refusal traces that include discrete time. We have extended the work of Huang and Peleska to refusal traces, with a paper being submitted, and are in the process of adding discrete time. This work builds on results regarding the structure of healthy languages of refusal trace.

Funding has been secured via an Integrator Project. This has brought together researchers from several institutions, from both within the Node and outside the Node. It also includes collaborators from the Resilience Node. This project will allow us to further explore human factors, such as emotion, and also to verify neural networks.

Planned Activities for the Final Year

There are several planned lines of work. The following are some of the main activities but there should also be causality-based work.

Model transformations: model-independent testing. We have results (a complete testing technique and underlying theory) for untimed models that do not have internal/unobservable actions. We are close to completing the extension to allow internal actions and will then include discrete time (to be consistent with the tock CSP semantics of RoboChart models).

Model transformations from reactive to cyclic models. We have identified the challenges and have an initial mapping from traces of reactive models to corresponding traces of cyclic models. We will produce a theory that allows one to map complete test suites for a reactive model to complete test suites for a cyclic model. We will then address the issue of how one might actually implement such an approach and expect this to lead to notions of testability.

Mapping between RoboChart and Simulink models. We have corresponding models in these two languages for the assisted dressing case study. We will define a corresponding mapping and aim to generalise this. The fire-fighting case study should then provide an opportunity to assess how well this generalises and lead to improvements.

Mutation testing. We will use a mutation-based approach to generate tests for the assisted dressing case study. We will use a tool-chain based on the three tools: the Wodel mutation tool, RoboTool, and the FDR model-checker. This should (semi-)automate the generation of test cases from the assisted dressing models.

Search-based testing. We will use search to find test cases that stress the assisted dressing model. The aim here will be to find test cases that break requirements properties identified in a hazards analysis and also to explore alternative designs.

Verification of AI. We will have at least two neural networks that we can test. First, a CNN has been developed to track the movements of the patient, in the assisted dressing study, in order to provide input into planning. Second, partners in the Integrator project are developing a neural network solution that aims to identify reactions such as pain or fear in a patient, so that the robotic system can react appropriately. Both need to be verified.

Reality Gap

WP Lead: Rob Richardson Collaborators: Bilal Kaddouh, Lenka Mudrich, Shival Dubey

Summary of Overall Objectives

The objective of this work package is to capture the physical and computational properties of real autonomous systems via deconstructing the autonomous system into modelled components and then reconstructing it using verified component units. The work is applied on an implemented real world environment and digital model of it to explore and overcome the reality gap. Most of the work in this work package is conducted via the Firefighting UAV case study.

Planned Activities for Year 2

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.

Achievement and Result Highlights

Testing the Robochart model

The RobotChart model that has been developed in York started to be tested for verification in Leeds. This required a number of translation and matching activities in order to ensure the implementation of the UAV systems are instrumented to support testing using the reactive models defined in RoboChart. This meant that the test cases and scenarios can be generated from RobotChart and tested in the simulated environment in ROS. The generation of the unit test codes has been automated in an iterative process starting from sample test codes and arriving into the ability to automatically generate test cases and their matching test code implementation to run on the tested system.

System-level verification

Similar work has been carried out on the systems level with the system test cases generated to cover unit tests. These test cases are being prepared to run on an experimental set-up preceded by simulation runs.

Runtime-verification integration

The tests developed on both component and systems level are being considered for runtime verification. This work is looking at using the scenarios and test cases to identify the expected behaviour of the system and keep monitoring this behaviour in real-time while the system is in operation. This will allow the system to detect any anomaly during the execution of the simulation or the flight experiment. This is being developed and is close to being tested within the simulation environment.

Simulation development

A ROS simulation model for the Firefighting UAV is being developed. The current model requires the hardware of the UAV to be plugged in and serves as a hardware in the loop simulation for the flight control system, however we are working on developing a simulation that can run fully in ROS with no need to connect the hardware. This is based on the open source autopilot software of PX4. The interfaces with the flight control system will be kept the same and therefore making the simulation platform agnostic.

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.

- Property and scenario specification: we worked with most other work packages to specify system-level properties and scenario specification and generation to provide a holistic verification.
- Physical system modelling: we are developing realistic and accurate models of our case study and embedding them into simulation models and ensuring safety of the autonomy used by robots.

Reality Gap

Figure 1: Sample fire detection and tracking in action

Hardware Testing

The hardware systems of the UAV are being tested with the automatic aim and spraying control undergoing a number of experimental tests to evaluate its performance.

Planned Activities for the Final Year

The development of the visual navigation implementation to match the RobotChart model.
This activity is needed to ensure both models match so that verification exercises applied on the RobotChart model are applicable to the real robot.

The expansion of the onboard mission management system to allow for runtime monitoring and the conduct of instrumented verification flight tests.
This is an activity that supports the integration of the work conducted at the university of Manchester as part of WP7.

Development of instrumented real test environment
Development of instrumented real test environment involving the installation of sensors and other equipment to measure key test parameters. This test environment will be used to conduct controlled experiments on real robots and to compare their behaviour with digital twins of the same robots and environments.

Identify the reality gap
Attempting to identify the reality gap between the real robots and the digital twins by analysing the data collected from the experiments. This will involve comparing the behaviour of the real robots with their digital twins and identifying any differences or discrepancies.

Share data
Data collected from the simulations and experiments will be shared with the consortium, to ensure that all partners have access to the latest information and can collaborate effectively.

Cyber-Physical Systems

WP Lead: Mohammad Reza Mousavi   Collaborators: Hugo Araujo, Uraz Turker, Hector Menendez

Summary of Overall objectives

The objective of this work package is to cater for variability and uncertainty in the verification of cyber-physical systems (CPS). Furthermore, we would like to analyse causal relationships in CPS, providing causal explanations of verification results.

Planned Activities for Year 2

- Case Study: We plan to engage with the other work packages (Work Packages 1 and 4) on designing initial test models for the firefighting drone case study. We plan 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.

Achievement Results and Highlights

Causality for Cyber-Physical Systems

Most of the focus of the second year was on causality and explanation in CPS Verification. We developed a theory of causality for cyber-physical systems that is submitted to a journal and is undergoing revision. The theory has been mechanised into a tool and is publicly available alongside the models used in its evaluation. We attained the following results: 1) we extended an existing theory of actual causality to cope with cyber-physical systems where we obtain causal models from hybrid automata 2) we developed a process to apply the theory of causality and developed and integrated the algorithms in our tool (HyConf), and 3) we applied the developed technique to a series of case studies and benchmarks to evaluate its effectiveness.

The strategy works as follows. Given evidence of failures found in the testing phase, we inspect which elements of the system are causing the fault. More precisely, we build a causal model for continuous systems that is then explored using search-based metaheuristics in order to find causes for failures observed during a verification phase. Furthermore, our notion of causality takes time into consideration as one cannot ignore the history of a continuous system execution. This way, we aimed to identify not only to which components the fault should be attributed, but also provide the intervals of time when the causal behaviour has occurred in the respective components. We mechanised our causal analysis using the Matlab/Simulink framework, which is a commonly used environment for modelling and analysis of control systems, thus increasing the accessibility of our strategy to the CPSs community. Furthermore, we have conducted controlled experiments against complex systems, in which, we have shown how it can be used to determine causes for a set of induced failures. Finally, we analysed our algorithms performance against a series of benchmark models. The results indicate that increasing the complexity of the subject systems increases the time taken to perform causal analysis linearly. However, increasing the maximum number of possible variable interactions in a single cause has a much greater impact on the overall efficiency of the strategy and the time taken increases non-linearly.

Kaspar Causally Explains

Still on the topic of causality, we applied causal reasoning to furnish a robot, called Kaspar, with the ability to provide causal explanations. Kaspar is used for robot-assisted learning for children with autism spectrum disorder (ASD). Children with ASD often struggle with their Visual Perspective Taking (VPT) skills, that is, one's ability to understand that other people have a different point of view. Thus, in order to improve their social interaction skills, Kaspar has been proven to be an essential tool. Based on a series of experiments (retrospective, pilot, and main studies), we developed a set of explanations that were designed to be automatically given to the children by Kaspar during their interactions. The interactions occur via controlled games in which a task is given by Kaspar and, in case of non-compliance, an explanation relating to the mistake is given.

Cyber-Physical Systems

The process to decide which explanations to provide is established on causal analyses; in order to provide the causal explanations, a causal model for the games is built and, based on mathematical proofs related to the Halpern-Pearl causality theory, the correct cause for the mistake is computed. The JAVA implementation for this has been integrated into Kaspar and this work has been published at a conference on social robotics. The experiment involving Kaspar and children with ASD has also been conducted but remains unpublished. The conclusion is that children performed better when they were given causal explanations as they performed more correct actions and reduced the number of mistakes. Furthermore, our results indicate that, once the children had been exposed to the explanations, they learned the correct course of action, which means that they've levelled up their VPT skills and retained them.

Systematic Review on Verifying RAS

We performed a systematic review of the interventions for testing robotics and autonomous systems in order to answer the following research questions: (i) What are the types of models used for testing RAS? (ii) Which efficiency and effectiveness measures were introduced or used to evaluate RAS testing interventions? (iii) What are the interventions supported by (publicly available) tools in this domain? and (iv) Which interventions have evidence of applicability to large-scale and industrial systems? To this end, we started off by performing a pilot study on a seed of 26 papers. Using this pilot study, we designed and validated a search query, designed rigorous inclusion and exclusion criteria and developed an adaptation of the SERP-Test taxonomy. Subsequently, we went through two phases of search, validation and coding, in total going through a total of 10,534 papers. We finally coded the set of 192 included papers and analysed them to answer our research questions. We have provided a summary of the findings of the review with regards to our research questions and some suggestions for researchers and practitioners.

We have found that there is a wealth of formal and informal models used for testing RAS. In particular, there is a sizable literature on using generic property specification languages (such as linear temporal logic) and qualitative modelling languages, such as variants of state machines, UML diagrams, Petri nets and process algebras. There is a clear gap in quantitative modelling languages that can capture the complex and heterogeneous nature of RAS. There is also a lack of domain-specific languages that can capture domain knowledge for various sub-domains of RAS. Furthermore, we observed a gap in rigorous and widely accepted metrics to measure effectiveness and efficiency, and adequacy of testing interventions. Similar to the previous items, those measures used in the literature are very generic and do not pertain to the domain specific aspects of RAS. Hence, there is a gap and a research opportunity for defining and evaluating rigorous (domain-specific) measures for efficiency, effectiveness, and adequacy for RAS testing interventions. As another observation, there is also a considerable number of interventions that rely on public tools to implement or evaluate their interventions. However, there are very few which make their proposed / evaluated interventions available for public use in terms of publicly available tools. There is hence a considerable gap in providing data-sets and public tools for further development of the field. And finally, we have concluded that there are less than a handful of testing interventions that have been evaluated in an industrial context. There have been some other interventions that used some real robots or autonomous systems, but in an academic context. This signifies the importance of future co-production between academia and industry in industrial evaluation of testing interventions for RAS.

Test Model Enrichment process for CPSs

We have developed a study on test model effectiveness for RAS, which is still unpublished. In this study, we specify how a test model can be ‘informed’ by a design, but not generated from it, by a process of stepwise enrichment, where more details are added based on the behaviour of the design. The process starts with an abstract model that comprises a start and a goal state informed by the design model. The iterative enrichment adds constraints on the dynamics. The process includes steps to take an independently drafted model to a close translation of the design model.

Moreover, we quantify the ‘effectiveness’ of a test model in terms of three metrics: i) its fault detection capabilities, i.e., the higher the number of detected faults, the better the test model, ii) its precision, i.e., the lower the number of false negatives, which are the fail verdicts that should actually be pass, the better the test model and iii) its sensitivity, i.e., the larger the difference between a system

Cyber-Physical Systems

behaviour and its specified properties, the better the test case. We hypothesise that test models that are highly detailed and, hence, costly to develop, offer little impact with respect to fault detection capabilities compared to more abstract models. On the other hand, if test models that are too abstract, they may comprise wrong assumptions about the system and it may behave in a way that is not possible, thus generating invalid test cases. The final point indicates that since larger differences are observed, then, more subtle faults should be found. We are in the process of assessing our hypotheses via controlled experiments.

Instantaneous Trust in RAS

In conjunction with the Leap of Faith project (TAS Hub), we are co-designing a conceptual model to improve our understanding of the factors that influence instantaneous trust (i.e., trust that occurs between parties that have little to no interaction history). One of the goals is to conduct user studies to validate our instantaneous trust conceptual model and understand the factors that play a role in it, such as evidence of verification (guarantees), the authority behind the system (guarantors), and personal bias (such as personal history and the robot's appearance). So far, we have published a paper in which we reflect on our experience in the conceptualization and implementation of Responsible Research and Innovation (RRI) in this project. In the paper, we report the narratives captured during a series of focused discussions involving the project team and industry partners, using RRI Prompts and Practice Cards (PPC) as a tool to guide the discussion. Overall, we highlight the importance of incorporating RRI principles into exploratory research projects to ensure a human-centred and ethical approach. We found out that being a 'responsible' researcher requires both ethical awareness and EDI responsiveness.

Planned Activities for the Final Year

- Further develop and publish the results regarding the quality of test models and test-model refinement for cyber-physical and autonomous system; and
- Publish the results of experiments using causal explanations in Kaspar on the educational experience of children in the autism spectrum disorder.

Sub-symbolic AI

WP Lead: Ivan Tyukin Collaborators: Wendy Otieno

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. A part of the work was presented to the 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 was concerning 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. The work is currently ongoing. We expect to complete the first exploratory phase of this work by the end of the year.

Additional output, supporting work of other work packages and connecting AI to testing and verification, include developing a novel reinforcement learning technique to model-based testing (in collaboration with WP5).

Summary of Overall Objectives

The work package aims at the development of new theory and algorithms for AI verification without the need to generate adversarial cases. Expected deliverables are:

D1) theorems for AI verification without the need to generate adversarial examples;

D2) verifiable specifications;

D3) algorithms and prototype tools models, producing quantitative tight error bounds.

Planned Activities for Year 2

- Proven robustness for sub-symbolic AI: Following up on the activities carried out last year, we aim to establish proven bounds of robustness and vulnerability for sub-symbolic AI .
- Using sub-symbolic AI for verification: in collaboration wit WP4, we plan to develop reinforcement-learning-based verification techniques that work efficiently for complex systems .
- Developing a framework for verifying ethical aspects of sub-symbolic AI: in collaboration with WP5, we plan to develop a framework for capturing, specifying, and testing ethical aspects of sub-symbolic AI.

Achievement Results and Highlights

Robustness and Adversarial Attacks

To date, the focus of the work was primarily on determining conditions specifying when verification of sub-symbolic AI models is possible using only input-output observations.

It has been proven that unfortunately, major faults can be successfully hidden within large-scale high-dimensional models. It is now very clear that there is a subclass of problems which cannot be verified using “hold out” data approaches. Examples of algorithms hiding such faults, their corresponding tight probabilities of success and formal statements along with the proofs have been provided.

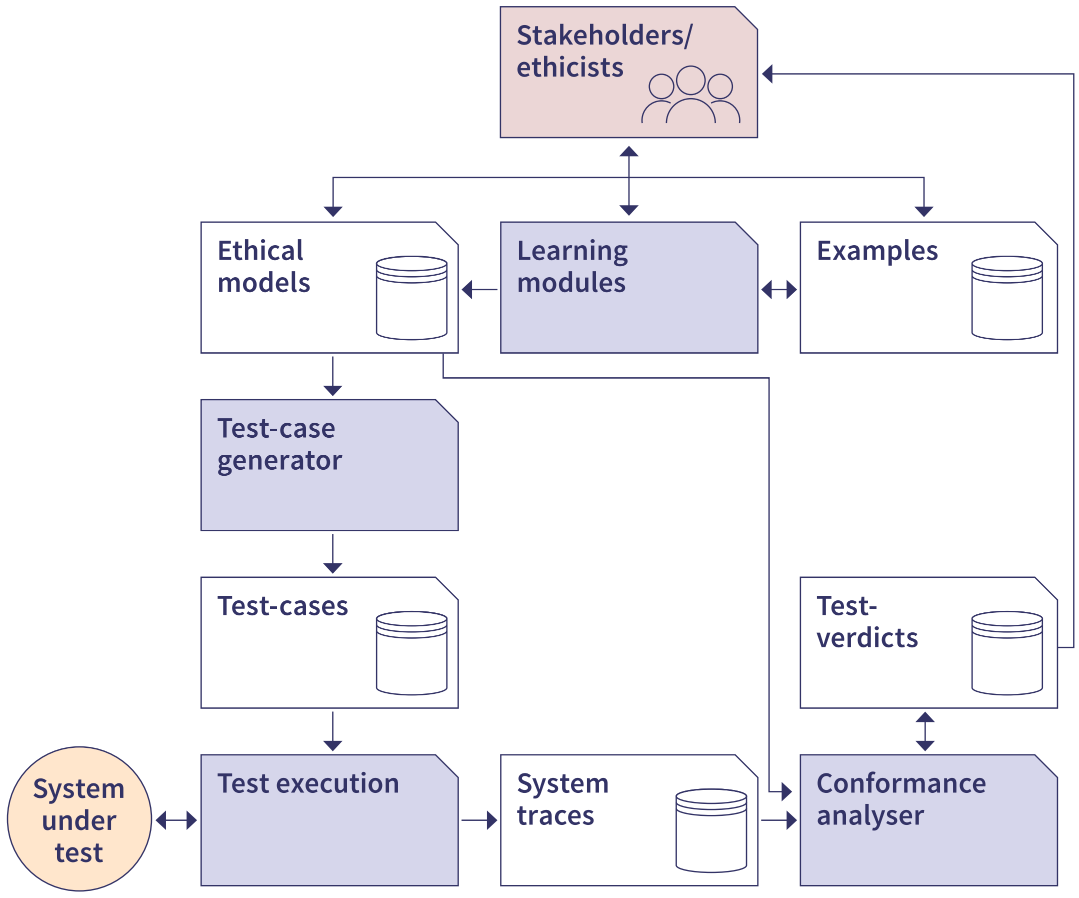
This work, in addition to showing, for the first time, how challenging it is to verify modern AI models using classical input-output observations without looking into the model’s latent spaces, also shows why backdoor attacks and vulnerabilities are practically possible in these systems.

Testing Ethical Decision Making

In collaboration with WP5, we developed an architecture for testing moral and ethical decision making in AI-enabled systems. The architecture is depicted in Figure 5. The architecture is based on an iterative engagement with different stakeholders to understand and formalise their ethical concerns in a formal model and use it for generating test cases in the corresponding architecture.

Sub-symbolic AI

Figure 5: General architecture for learning and testing ethical decision making in AI-enabled systems



In the next year we will study a dual setting. The following directions will be explored

1. Exploitation of explainable AI / counterfactual methods to build better AI correctors (bringing causal context into the picture). 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. It will enable testing for specific fault scenarios avoiding exhaustive exploration of models' inputs.
3. Realising our conceptual architecture for testing ethical concerns in sub-symbolic AI.
4. Work on case studies: provide examples of verifiable tests so that they could be checked within the scope of these case studies.

Planned work for the final year

Using sub-symbolic AI for verification In collaboration with WP5, we developed a novel reinforcement learning framework to generate test sequences from specifications based on finite-state machines. The results appeared in a conference publication (at ASE 2022) and will be developed further and be prepared for a journal publication.

Symbolic AI

WP Lead: Michael Fisher   Collaborators: Louise Dennis, Fatma Faruq, Maryam Ghaffari Saadat

Summary of Overall Objectives

Work Package 7 is concerned with the modelling and verification of high-level, symbolic reasoning. This is a key component in the **neuro-symbolic/hybrid** architectures that are widely used for autonomous systems. In particular, describing and verifying the symbolic component allows us to tackle issues such as *how* and *why* high-level decisions are made by the autonomous system.

The issue of "why" decisions are made distinguishes Autonomous Systems from Cyber-Physical Systems as, once our systems are autonomous, then we not only need to verify what they will do but why they choose to do it.

Planned Activities for Year 2

The overall objectives lead us to work on to a range of issues in Year 2, specifically:

- High-level reasoning – transparency, ethics and explainability;
- Architectural aspects of autonomous systems – reliability, certification, modularity, contracts, and heterogeneous/ corroborative verification;
- Handling inherent uncertainty, particularly in robotic/cyber-physical systems – dynamic specifications, Runtime Verification, etc;
- Security verification – mixing safety/security, and holistic security analysis;
- High-level reasoning concerning autonomous vehicle rules, e.g "rules of the road"; and
- Runtime verification of both fire-fighting UAV and robot-dressing scenario (with Node partners).

Achievement and Result Highlights

Reasoning and verification in (transparent) high-level symbolic layer(s).
We have continued to tackle ethics, explainability, etc, formalising and verifying multiple different theories of machine ethics, as well as arbiter processes taking into account a range of ethical issues. Key work has also been carried out on *explainability*, especially dialogue. We have also explored the modelling, verification, and use of appropriate (human-level) "rules", such as the "rules of the road" for autonomous road vehicles.

Architectural Aspects.
How autonomous systems can be constructed to be transparent, verifiable by design, flexible, resilient, etc, through both compositional specification of heterogeneous components, and heterogeneous verification. This also helped to shape work towards a general reliability/certification framework.

Handling inherent uncertainty.
Our predictions about most things will be wrong! So, recognising mismatches, especially at runtime, is important. Practically, we have upgraded our Runtime Verification tool for robotic systems to ROS2, but we have produced new work on foundational aspects of RV, coordinating multiple runtime monitors, and recognising **attacks/faults**. We are also working with partners across the Node to provide Runtime Verification for Firefighting UAV and Robot-Assisted Dressing case studies.

Beyond the Node.
We have worked with other TAS entities such as the Security Node the Functionality Node via a public engagement event and a Responsibility Project on Machine Ethics, Explainability, "Rules of the Road". More broadly, we have been strongly involved in standards, regulation, and safety, for example through IEEE, BSI/ISO, etc.

Symbolic AI

Planned Activities for the Final Year

- Runtime Verification: Continue runtime verification work as part of the Node-wide case studies on Firefighting UAV and on Robot Assisted Dressing. Typical properties checked for the former might involve **speed/temperature** (e.g: “don’t fly too close to the fire”) while typical properties for the latter might involve force/movement (e.g.: “robot arm should move slower the closer it gets to the human”).
- Architectural – heterogeneous verification across heterogeneous architectures: Theoretical work on combining different verification techniques on diverse **modules/nodes** across **robotic/systems architectures**.
- RoboChart Verification: New work exploring the translation of RoboChart into NuXMV for subsequent model-checking of RoboChart models.
- Swarm Verification: Exploring mechanisms for (robot) swarm specification and verification, currently via Propositional Temporal Logic and Graph Grammars and both verified via model checking, then subsequently with First-Order Temporal Logic and verification via deduction.
- Logical formalisation of “responsibility”: with a TAS Responsibility project, on the formalisation of autonomous agent “responsibility” and the potential for logical verification in this area.
- Neuro-Symbolic AI: with the TAS Governance Node at Edinburgh, looking at Neuro-Symbolic AI, its principles and development.

Closing the Reality Gaps: Users

WP Lead: Effie Law Collaborators: Ana Cavalcanti, Louise Dennis, Farkhandah Komal, Swaroop Panda, Ben Summerill, Zhongtian Sun

Summary of Overall Objectives

- To produce verification models of user behaviour in terms of human multisensory mental state in real-time, uncertain, and unconstrained scenarios.
- 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.

Planned Activities for Year 2

- Develop a case study on trust in conversational agents (CAs) as trustworthy autonomous systems to demonstrate a full cycle of producing verification models of user behaviour with specific focus setting on the notion of human-in-the-loop.
  – perform Hierarchical Task Analysis (HTA);
  – develop a prototype of CA for older adults on online banking, incorporating NLP techniques and ChatGPT;
  – conduct empirical user studies with older adults in June 2023 to validate user models and refine HTA;
  – Preliminary design of a notation to describe a system with humans in the loop.
- Contribute to the TAS Integrated Project on Robot Assisted Dressing (RAD)
- Apply for new research projects related to TAS.

Achievement and Result Highlights

Case study on a fintech chatbot for older adults

Rationale
Many banks have closed their physical branches and switched to digital banking for all kinds of services. Financial institutions have been leveraging the chatbot technology to provide efficient and personalised customer support. However, digital banking and fintech are generally viewed with scepticism by the elderly population due to their conservative attitudes toward money and technology. The elderly population is at risk of marginalisation if they distrust and disuse technology that is designed to improve their quality of life. To cater specifically to older adults, it is crucial to ensure high usability of the chatbot to invite their trust, provided that the chatbot is trustworthy, safe, and secure from the algorithmic perspective.

The main goal of this case study is to understand how to inculcate trust in a trustworthy fintech chatbot, enabling elderly users to undertake financial transactions with it (transfer money, make online payments, and manage pensions). To this, we use a bimodal (text/speech) chatbot and integrate ChatGPT into this fintech chatbot.

The chatbot is built upon hierarchical task analysis (HTA). It requires natural language processing (NLP), speech recognition, text generation, and Deep Learning (DL) techniques. In addition, ChatGPT is integrated as it can give the chatbot human-like characteristics. Whether such human-likeness can enhance elderly users' trust in a trustworthy chatbot is an empirical question to explore. Presumably, having a casual conversation with the chatbot may satisfy the user need of social relatedness. Nonetheless, it is of paramount importance that the elderly users will not overtrust the banking chatbot. Relevant context-aware reminders and warnings need to be issued to alert elderly users of potential risks. Trust calibration is very challenging.

From the interactive design perspective, the user interfaces of the chatbot should be inclusive to address older adults' characteristics:

- *Larger font sizes:* Increase the font size in the chatbot interface to ensure readability for users with visual impairments or difficulty reading small text.
- *Optimal contrast and background:* Adjust the contrast and background colour of the interface to minimise visual strain and enhance readability.
- *Voice interaction capabilities:* Integrate voice recognition and speech synthesis technology to enable elderly users to interact with the chatbot using speech, reducing the reliance on typing skills and catering to users with mobility limitations.

Closing the Reality Gaps: Users

Integrated project RAD
WP8 has been in collaboration with WP3 for the Integrated Project Robot-Assisted Dressing (RAD; Lead: Sheffield). Specifically, our inputs to RAD include:

- Collection and preliminary analysis of the empirical data from the stroke patients.
- Doing conversational analysis of the focus group results.
- Developing insights and implications from the co-design workshop.
- Helping develop the Human-Robot Interaction (HRI) model.

TAS Hub Pump Priming Round 3 – CA4OA
The proposal submitted to TAS Hub Pump Priming Round 3 has been granted (£195K FEC). The title of the project is "Co-designing Inclusive and Trustworthy Conversational Agents on Basic Services with Older Adults"

- Interactive prototypes of banking chatbots based on the HTA and user models have been developed.
- Contributing to the development of the human-robot interaction model for robot-assisted dressing.
- Developing initial modelling concepts for human-in-the-loop.

Planned Activities for the Final Year

- Produce a methodological framework for investigating older adults' trust in conversational agents (CAs), which tap into the power of NLP and large language models (LLM).
- Develop CA prototypes in different domains, including online banking, healthcare, and entertainment.
- Conduct user-based evaluation studies with older adults to identify their mental models of such CAs and iteratively improve user models by incorporating verification results based on human-system interaction patterns identified.
  - Explore how to Integrate HTA and RoboHuman (York) to formalise human-in-the-loop
  - Finalise the case study on RAD with Sheffield and York on emotion analysis
  - Write up the work on CAs for CHI'24 and other HCI conferences/journals.

Key Outputs

Publications

G. Yaman, C. Burholt, M. Jones, R. Calinescu, and A. L. C. Cavalcanti: "Specification and Validation of Normative Rules for Autonomous Agents." In L. Lambers and S. Uchitel, editors, *Fundamental Approaches to Software Engineering.* Springer, 2023.

J. Baxter, A. L. C. Cavalcanti, M. Gazda, and R. M. Hierons: "Testing Using CSP Models: Time, Inputs, and Outputs." *ACM Transactions in Computational Logic,* 24(2), 2023.

A. L. C. Cavalcanti, A. Miyazawa, U. Schulze, and J. Timmis: *"Bringing RoboStar and RT-Tester together."* In Jan Peleska's Festschrift. Springer, 2023.

A. L. C. Cavalcanti, A. Miyazawa, U. Schulze, and J. Timmis: "Challenges in testing cyclic systems". *In 27TH International Conference on Engineering of Complex Computer Systems.* IEEE, 2023.

K. Ye, A. L. C. Cavalcanti, S. Foster, A. Miyazawa, and J. C. .P. Woodcock: "Probabilistic modelling and verification using RoboChart and PRISM." *Software and Systems Modeling,* 21:667--716, 2022.

W. Barnett, A. L. C. Cavalcanti, and A. Miyazawa: "Architectural Modelling for Robotics: RoboArch and the CorteX example." *Frontiers of Robotics and AI,* 2022.

M. Windsor and A. L. C. Cavalcanti: RoboCert: "Property Specification in Robotics." In A. Riesco and M. Zhang, editors, *International Conference on Formal Engineering Methods,* volume 13478 of Lecture Notes in Computer Science. Springer, 2022.

M. Mousavi, A. L. C. Cavalcanti, M. Fisher, L. Dennis, R. Hierons, B. Kaddouh, E. L.-C. Law, R. Richardson, J. O. Ringert, I. Tyukin, and J. C. P. Woodcock: "Trustworthy Autonomous Systems through Verifiability." *IEEE Software Magazine,* 2022. (To appear).

J. Baxter, P. Ribeiro, and A. L. C. Cavalcanti: "Sound reasoning in tock-CSP." *Acta Informatica,* 59:125--162, 2022.

B. Townsend, C. Paterson, T. T. Arvind, G. Nemirovsky, R. Calinescu, A. L. C. Cavalcanti, I. Habli, and A. Thomas: "From Pluralistic Normative Principles to Autonomous-Agent Rules." *Minds and Machines,* 2022.

T. Wright, C. Gomes, J. Woodcock: "Formally verified self-adaptation of an incubator digital twin." In: *Leveraging Applications of Formal Methods, Verification and Validation. Practice: 11th International Symposium,* ISoLA 2022, Rhodes, Greece, October 22–30, 2022, Proceedings, Part IV 2022 Oct 17 (pp. 89-109). Cham: Springer Nature Switzerland.

S. Bogomolov, C. Gomes, C. Isasa, S. Soudjani, P. Stankaitis, T. Wright: "Data-Driven Reachability Analysis of FMI Models." (Under review).

J. Baxter, A. Cavalcanti, M. Gazda, and R. M. Hierons: "Testing using CSP models: time, inputs, and outputs." *ACM Transactions on Computational Logic,* 24, 2 pp. 1-40, Article No. 17, 2023.

A. Cavalcanti and R. M. Hierons: "Challenges in testing of cyclic systems", *27th International Conference on Engineering of Complex Computer Systems* (ICECCS 2023).

M. Gazda and R. M. Hierons: "Model Independent Refusal Trace Testing" (Under review).

Gazda, M., & Hierons, R. M.:" Removing redundant refusals: Minimal complete test suites for failure trace semantics." *Information and Computation,* p. 291, (2023).

M. Nunez, R. M. Hierons, and R. Lefticaru: "Implementation relations and testing for cyclic systems: adding probabilities." In *Robotics and Autonomous Systems* (to appear).

M.E. Akintunde, M. Brandao, G. Jahangirova, Hector Menendez, M.R. Mousavi, and J. Zhang: "On Testing Ethical Autonomous Decision-Making." *In the Festschrift dedicated to Jan Peleska's 65th Birthday.* LNCS Festschrift series. Springer, 2023.

Key Outputs

H. Araujo, P. Holthaus, M, Sarda Gou, G, Lakatos, G. Galizia, L. Wood, B. Robins, M.R. Mousavi, and F. Amirabdollahian: "Kaspar Causally Explains.", *Proceedings of the 14th International Conference on Social Robotics, Lecture Notes in Computer Science,* Springer, 2022.

M. Sarda Gou, G. Lakatos, P. Holthaus, L. Jai Wood, M.R. Mousavi, B. Robins, and F. Amirabdollahian: "Towards understanding causality – a retrospective study of using explanations in interactions between a humanoid robot and autistic children." In *Proceedings of the 31st IEEE International Conference on Robot & Human Interactive Communication* (RO-MAN 2022), IEEE, 2022.

H. Araujo, M.R. Mousavi, and M. Varshosaz: "Testing, Validation, and Verification of Robotic and Autonomous Systems: A Systematic Review." In *ACM Transactions on Software Engineering and Methods* (ACM TOSEM), 2022.

H. Araujo, M.R. Mousavi, G. Carvalho, and A. Sampaio, H. Chockler: "Causality for Cyber-Physical Systems." (Under review for ACM TOSEM).

S. Weerawardhana, J. Lisinska, M. Akintunde, H. Araujo, G. Kefalidou, E. Nichele, Y. Lu, O. Malpass, A. Roberts and I. Sandu: "Implementing Responsible Research Innovation Prompts and Practice Cards in a project investigating Instantaneous Trust." In *Trustworthy Autonomous Systems Symposium* (TAS'23), 2023.

L. Dennis, M. Fisher: "Verifiable Autonomous Systems—Using Rational Agents to Provide Assurance about Decisions Made by Machines". In *Cambridge University Press,* April 2023

D. Araiza-Illan, M. Fisher, K. Leahy, J.I. Olszewska, & S. Redfield: "Verification of Autonomous Systems [TC Spotlight]." In *IEEE Robotics & Automation Magazine,* 29(1), 99-101. 2022.

L. Dennis, R. Dixon, M. Fisher. "Verifiable Autonomy: From Theory to Applications." In *AI Communications* 35(4):421-431, 2022.

Collenette, J., Dennis, L. A., & Fisher, M.: "Advising Autonomous Cars about the Rules of the Road." In *Electronic Proceedings in Theoretical Computer Science* (EPTCS) 371:62-76, 2022.

R. C. Cardoso, A. Ferrando, and M. Fisher: "Extending Attack-Fault Trees with Runtime Verification." In *Electronic Proceedings in Theoretical Computer Science* (EPTCS) 371:193-207, 2022.

L. Dennis, and N. Oren: "Explaining BDI agent behaviour through dialogue." In *Autonomous Agents and Multi-Agent Systems.* Springer. (2022)

A. Ferrando, R. Cardoso, M. Farrell, M. Luckcuck, F. Papacchini, M. Fisher, V. Mascardi.Bridging the Gap between Single- and Multi-model Predictive Runtime Verification. Formal Methods in System Design 59(1):44-76, 2022.

E.L.C. Law, A. Følstad, A., & N. Van As: "Effects of Humanlikeness and Conversational Breakdown on Trust in Chatbots for Customer Service." In *Proceedings of the Nordic Human-Computer Interaction Conference* (pp. 1-13), Aarhus, Denmark. (2022).

E.L.C. Law, N. Van As, A. Følstad: "Effects of Prior Experience, Gender, and Age on Trust in a Banking Chatbot with(out) Breakdown and Repair." In *Proceedings of IFIP TC13 Conference* INTERACT 2023, York, UK. (2023).

E.L.C. Law, A. Følstad, J. Grudin, & B. Schuller: "Conversational Agent as Trustworthy Autonomous System (Trust-CA)" (Dagstuhl Seminar 21381). In *Dagstuhl Reports* (Vol. 11, No. 8). Schloss Dagstuhl-Leibniz-Zentrum für Informatik. 2022.

M.E. Akintunde, M. Brandao, G. Jahangirova, Hector Menendez, M.R. Mousavi, and J. Zhang.: "On Testing Ethical Autonomous Decision-Making." In *The Festschrift dedicated to Jan Peleska's 65th Birthday.* LNCS Festschrift Series. Springer, 2023.

M. S. Gou, G. Lakatos, P. Holthaus, B. Robins, S. Moros, L. Jai Wood, H. Araujo, C. A. E. deGraft-Hanson, M. R. Mousavi, F. Amirabdollahian: Kaspar Explains: The Effect of Causal Explanations on Visual Perspective Taking Skills in Children with Autism Spectrum Disorder.". In *Proceedings of the 32nd IEEE International Conference on Robot & Human Interactive Communication* (RO-MAN 2023), IEEE, 2023. (To appear).

Key Outputs

Keynotes and workshops

Michael Fisher: Keynote on “Autonomous Systems, AI, and Machine Learning – Why the Differences are Important”, 42nd SGAI International Conference on AI, [www.bcs-sgai.org/ai2022].

Michael Fisher: Panel/consultation: AI Standards Hub workshop on Transparency and Explainability, Alan Turing Institute, Jan 2023.

Ivan Tyukin: Talk on the subject of mathematical quantification and high-level tests of non-symbolic AI / data-driven AI vulnerabilities. “Breaking into a Deep Learning box”, April 10 – 11, 2021. ICERM Safety and Security of Deep Learning.

Louise Dennis: Keynote: “Verifying Autonomous Systems” at 17th International Conference on Integrated Formal Methods, 2022. [ifm22.si.usi.ch/pages/keynotes].

Louise Dennis: Organised Lego Rovers at primary schools.

Jim Woodcock: Talk at Huawei’s engineering conference on advanced software engineering practice.

Mohammad Reza Mousavi: Keynote at the MDENet Annual Symposium 2022 about MDE for cyber-physical systems.

Hector Menendez: Keynote about the malware Arms Race at AsturCon (2023). [https://asturcon.tech/https://asturcon.tech/].

Robert Hierons: Keynote at the 32nd International Symposium on Logic-based Program Synthesis and Transformation (LOPSTR 2022). The title of the keynote was “Systematic Software Testing for Robotics”.

Ana Cavalcanti: Invited talks delivered at a Summer School at ECNU in Shanghai on Trustworthy Systems.

Organised ICRA’23 Workshop on Multidisciplinary Approaches to Co-Creating Trustworthy Autonomous Systems. [https://mactasworkshop.github.io/2023/].

Ana Cavalcanti: Organised Workshop Yorrobots Industry Exhibition [https://www.york.ac.uk/yorrobots/news-events/yorrobots-events/].

Michael Fisher: Organised “Autonomous Systems: A Workshop on Cross-cutting Governance”, National Engineering Policy Centre, April 2022.

Michael FIsher: Organised ICRA’22 Workshop on Verification of Autonomous Systems, May 2022. [robotistry.org/vaswg/ICRA22_Workshop].

Michael FIsher: Organised International Symposium on the Verification of Autonomous Mobile Systems, March 2023 [www.irt-systemx.fr/evenements/vams-is-23].

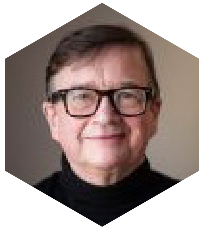
Ana Cavalcanti: Featured as a speaker in the latest Living with AI Podcast of the TAS Hub titled: “Are you Talking to your Autonomous Car? (Maybe you should!)”.

Louise Dennis: Panel/consultation: International Transport Forum Roundtable on AI, Machine Learning and Regulation, OECD, January 2023.

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