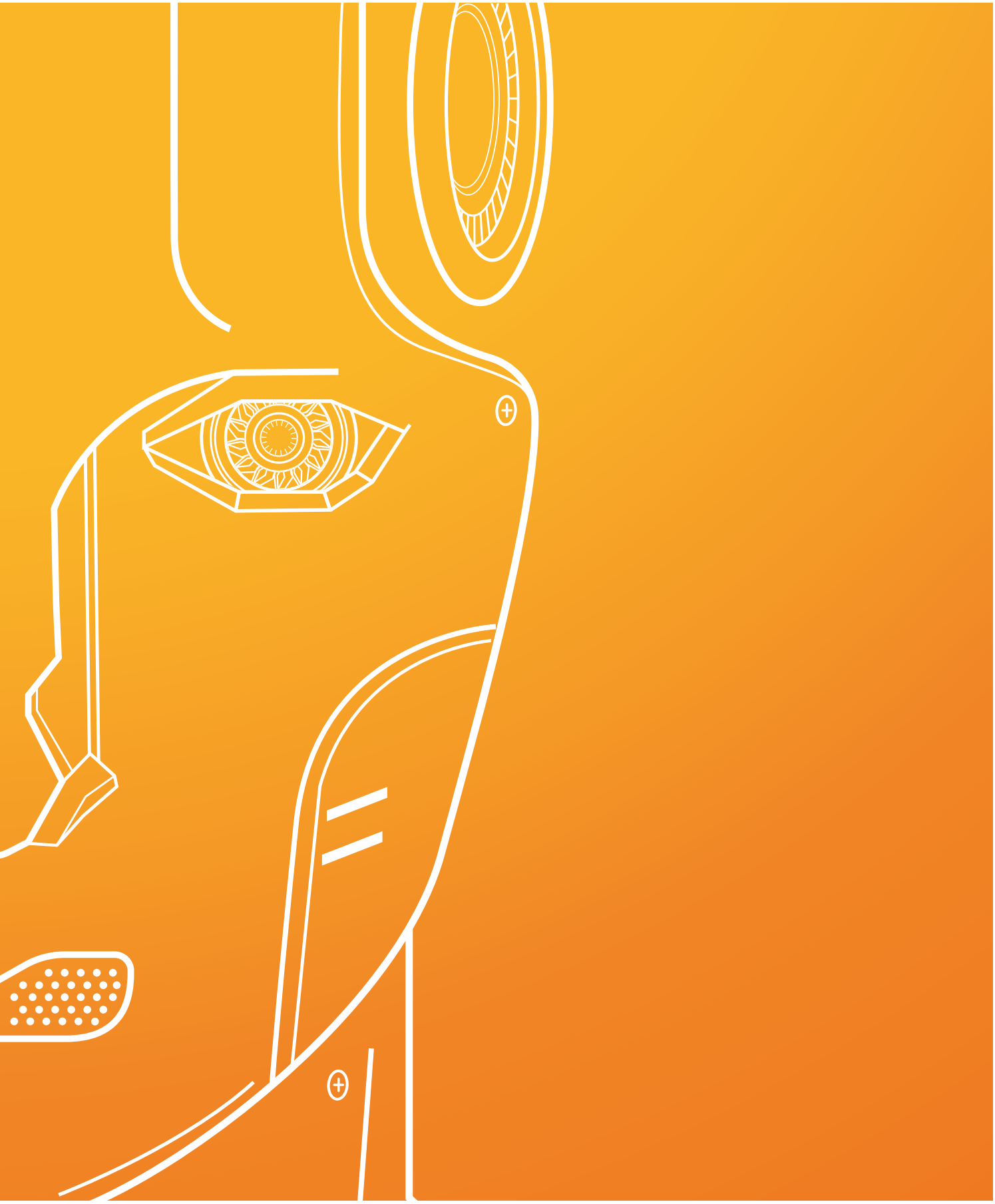
# ASSURING AUTONOMY
INTERNATIONAL PROGRAMME

## A YEAR IN REVIEW 2019



# Addressing global challenges in assuring the safety of robotics and autonomous systems.

## Key

- Funders
- Demonstrator projects
- Collaborative links
- Programme Fellows

# Welcome

**2019 has shown just what an enormous challenge we face in assuring autonomy. The promise of robotics and autonomous systems continues to grow and enthuse us, but the issues are complex.**

The challenges are not insurmountable, however, and we made great progress in 2019, both in York and through partnerships across the world.

We have continued to grow the York team, welcoming new Research Associates to further our foundational research, and build on our existing expertise. We also invested an additional £1.2million of funding in new demonstrator projects, each addressing the challenges of assuring autonomous systems in their domain, while considering how we foster public trust in the technologies. Some demonstrator projects that began in 2018 are coming to an end and we look forward to sharing their results with you in 2020.

Programme Fellows have been a fantastic asset and we have welcomed ten new colleagues from six different countries to the scheme. Fellows bring new perspectives to our research, and provide insight that enhances and complements the work that we are doing.

Collaboration is key to the success of the Programme and we have worked across the globe, through demonstrator projects, Programme Fellows, research, and membership of boards and networks, to join forces where needed.

A real triumph of our partnerships is their multi-disciplinary nature. We brought together lawyers, ethicists, philosophers, social scientists, clinicians, and others: the safe introduction and adoption of robotics and autonomous systems is a complicated issue and we must work across disciplines to solve it. This collaboration recently resulted in a paper in Elsevier's prestigious Artificial Intelligence journal, with more to come in 2020.

Over the year we established York as the leading centre of excellence for the safety assurance of robotics and autonomous systems. Our leadership in this area was strengthened with the announcement of significant Government funding through the UK Research Partnership and Innovation Fund. This funding, alongside corporate and philanthropic investment, will enable us to build a dedicated facility on the University of York campus and develop a £35million initiative that will bring together our work with other complementary expertise from the University, including robotics, advanced communications, and quantum technologies.

We are using these achievements to build more momentum as we move into 2020, and we invite you to join us on the journey.

**Professor John McDermid OBE FREng**
Programme Director



Lloyd's Register Foundation was established to engineer a safer world. Our new strategy sets out the biggest safety challenges the world faces, and outlines that we need to think globally, and collaboratively, to solve them. The safety of autonomous systems is one such challenge and the Assuring Autonomy International Programme is collaborating on a global scale to ensure that society benefits from all that such technology can offer.

The team is leading and supporting partnerships across the world to develop practical guidance that is based on evidence of what works, and what doesn't. This is of great importance to the developers and safety engineers who are progressing this technology, the regulators who must be confident in their assessment of the technology, and of course the public and end users who will benefit from their safe introduction. We will all feel the impact of autonomous systems in our lifetime, and we must not underestimate the importance of having confidence in their safety.

**Professor Richard Clegg FREng**
Chief Executive
Lloyd's Register Foundation

As a University with a global outlook, we understand the importance of working across traditional disciplines to address the world's most pressing challenges. The work of the Assuring Autonomy International Programme is a fantastic example of computer scientists and engineers working with philosophers, lawyers, clinicians, industry experts and others to develop novel ways to assure the safety of robotics and autonomous systems.

This challenge is multi-faceted, and the only way to address it is with a multi-disciplinary approach, underpinned by research excellence. Working with Lloyd's Register Foundation has enabled us to establish this world-leading Programme, and helped us to unlock further Government funding to build a new and unique facility on the University of York campus, dedicated to the advancement of safe autonomous systems.

**Professor Charlie Jeffery CBE**
Vice-Chancellor and President
University of York

# A year in numbers

## Research

**10**

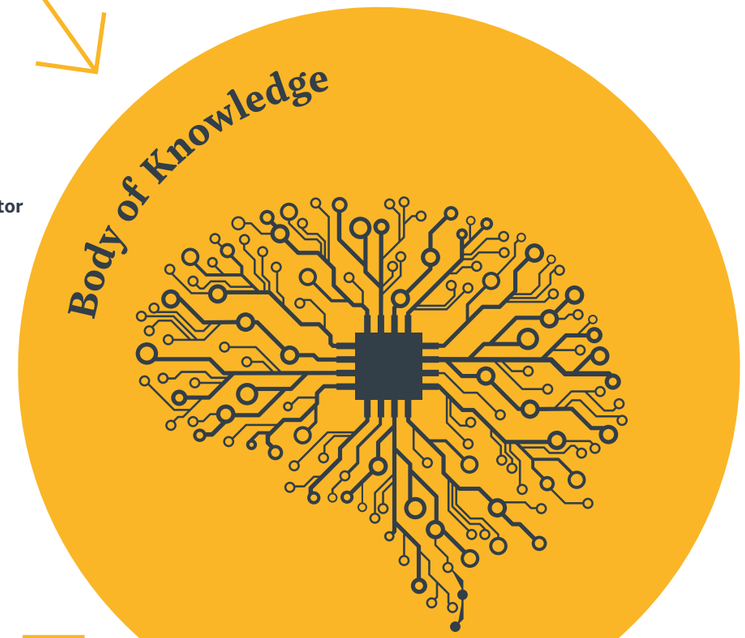
active demonstrator projects


- Demonstrator projects
- York team

**£30.7M**

funding leveraged

## Body of Knowledge



## Public engagement

**500**

people at Festival of Ideas AI day


- Industry
- Local
- National
- Regional

## International community

Worked across **5** continents


- UK
- Germany
- Norway
- Belgium
- Brazil
- Australia

Presented in

**20**

cities

## Education and training


- Academic
- Aerospace
- Automotive
- Communications
- Cross-domain
- Defence
- Education
- Energy
- Healthcare
- Manufacturing
- Maritime
- Professional services

AAIP demonstrator and research workshop (York, October 2019)

# International Community

**This year has seen us expand our links with the international community working on the safe introduction and adoption of robotics and autonomous systems (RAS). We know that collaboration is essential: the challenges are too great for one organisation to solve alone.**

The community is growing as we come to realise that we must collaborate, not only across geographic borders and across different domains and technologies, but also across disciplines. The solutions to assuring the safety of RAS must be found in partnership with lawyers, ethicists, social scientists, and others, else we risk creating solutions that don't meet our varied needs.

Through our Programme Fellowship scheme, our research, demonstrator projects and other networks we have expanded our reach into five continents (see map on pages 2-3) over the course of 2019.



## Europe

We have supported eight demonstrator projects in Europe (see pages 20-27), totalling more than £2million across domains including automotive and healthcare. We have welcomed four new PhD students to York as part of our involvement in the Marie Sklodowska-Curie European Network for Safer Autonomous Systems, and a Research Associate as part of the European Union's Horizon 2020-funded Robotics for Inspection and Maintenance project.

We have been involved in numerous healthcare domain initiatives, including presenting at MedInfo on the safety of AI for clinical decision support, with colleagues from Europe, the USA and Australia. We also co-hosted a digital health safety conference with NHS Digital at the University of York, bringing together more than 100 clinicians and technologists to discuss the safe introduction of AI-based digital health technologies.

We were delighted to be part of a debate on ethics and safety at a European Parliament event organised by the White Rose University Consortium, which brought together European Members of Parliament with leading academics from the universities of York, Sheffield and Leeds, to discuss how we address the challenges of increasing human-robot interaction in many areas of society. New partnerships developed after Professor John McDermid led the development of a white paper on ethical issues for RAS with other members of the UK-RAS Network.

In the automotive domain we co-hosted a workshop with the Centre for Connected and Autonomous Vehicles to discuss automated vehicles and ethics. The workshop brought together experts from across the UK, and will shape future policy in this area.

European Societal Challenges: The Future of Human-Robot Interaction (Brussels, February 2019)

## Asia

The TIGARS demonstrator project (see page 22) is nearing its end. With its partners in Japan the team made great progress in the assurance of autonomous vehicles, specifically in the areas of resilience and safety requirements, validation and verification of machine learning, and security-informed safety.

We were invited to Macau in August to present a novel framework for safety assurance of autonomous systems, as part of the AI Safety workshop at the International Joint Conference on Artificial Intelligence.

In November we were back in Macau, and delivered a workshop on assuring safety for physically assistive robots at the 2019 International Conference on Intelligent Robots and Systems with colleagues from Bristol Robotics Laboratory. Dr Richard Hawkins gave a presentation on the practical guidance on safety assurance available through the Body of Knowledge.

We were also pleased to participate in a gathering of leading thinkers in AI and autonomy in Beijing. The meeting, organised by the Minderoo Foundation and the Centre for International Security and Strategy at Tsinghua University, considered how we ensure that these technologies are leveraged for the benefit of all of society.


Assuring Safety for Physically Assistive Robots Workshop (Macau, November 2019)

## Australasia

Our partnerships in Australasia have expanded significantly in the past year, particularly in the healthcare and automotive domains. We were delighted to welcome a new Programme Fellow, Dr Farah Magrabi, from Macquarie University in Sydney. Farah has worked with another Fellow from the UK, Dr Mark Sujan, and others from York and beyond to examine health apps and how to assure their quality, safety and security. Their work was published in BMJ Health and Care Informatics journal. We are also extending this collaboration with Macquarie University by joining the Australian Alliance for Artificial Intelligence in Healthcare, working with their international membership on safety, quality and ethics.

In 2019 we also funded a new demonstrator project (ATM – see page 26) with the Australian National University. This project is looking at how to design an automatic testing mechanism for the autonomous software that controls a RAS.

## North America

A joint UK-USA demonstrator project has strengthened existing partnerships with the University of Virginia and Carnegie Mellon University (see page 24), developing our work on training and verification of machine learning in the context of shared autonomous driving.

We presented at numerous conferences in the USA over the year, with one of our Research Associates, Nikita Johnson, winning the George Peters Award at the International System Safety Society's 37th Conference for her excellent contributions to two of their conferences.

## South America

We are delighted to have appointed our first Programme Fellow from South America, Dr Genaína Rodrigues from the University of Brasilia. She will be working with the team in York on a new approach for the accurate verification of autonomous systems at design and run-time. This partnership will help us to open up new areas for collaboration in South America.

More information about working with us: **bit.ly/aaipworkwithus**

# Public Engagement

**"You have to make it as safe as possible, but how safe can you truly make it?". This was a question posed by a member of the public we spoke with during the year, highlighting that the issue of safety is not just one that academics, the Government, and technology developers are concerned with.**

To ensure that robotics and autonomous systems are acceptable to society, we need to explore public opinions, and understand their expectations, worries and enthusiasm.

Over the year we undertook informal interviews with the public, held a day of discussion and demonstrations as part of the York Festival of Ideas, and took part in a range of media interviews, including with the BBC as part of their Tech Tent series. We also worked with Missions Publiques on their International Citizens' Dialogue on Driverless Mobility.

Festival of Ideas (York, June 2019)

## Thought provoking vox pops

*In the summer we went out on the street and asked the public what they think about the introduction of RAS, what they are excited about, what concerns they have, and what they would need in order to feel that an autonomous system was safe.*

"It's very very positive. Thinking in terms of [autonomous] vehicles, I'm a lifelong pedestrian. It will enable me to actually drive for once."

"If a doctor made a decision on your treatment, you'd want to know why or how they did it, so I think the same thing goes [for a robot]."

"People, myself included, may be cautious to begin with but very quickly new technology becomes common technology and we put so much faith in it and we forget how we did without it originally."

"Can accidents still happen? I don't know. I think there are lots of questions to be answered."

"I think it's crucially important to have some sort of black box technology in robots because things will go wrong."

"In the future we'll look back on this time and it will be an industrial revolution."

"I think I'd have to know what kind of safety precautions were in place because I don't know if I think the technology is...advanced enough."

"It needs some kind of standard, some institution or organisation that put a standard on it."

"For me to actually get in and participate with one of these vehicles and actually to feel safe in it, I would want a test, for my own knowledge, a bit like a driving test."

## Family friendly activities

As part of York's Festival of Ideas we held a day of discussions, hands-on activities and driverless cars in action. We welcomed more than 500 people of all ages to the event, who were able to talk to Pepper the robot, find out how artificial intelligence (AI) works with an interactive game, and make their own robot. The panel discussions offered opportunities to talk with experts about the future of AI, whether AI should be feared or welcomed, how AI could be used to improve health and social care, and driverless vehicles.

Festival of Ideas (York, June 2019)

In one session we asked the audience to tell us what they thought were unacceptable uses of AI. The overwhelming answer was related to autonomous life and death decisions, with the audience agreeing the necessity for a human in the loop. Other popular answers included creative tasks such as dancing, art or music. The audience also suggested that AI should not be used for cooking, for making legal or financial decisions, or for teaching or childcare, and these discussions are sure to continue for many years.

### 2020 and beyond

We'll be undertaking more in-depth conversations with the public to understand their opinions and concerns about the introduction of RAS in 2020 and will use this to shape our engagement with partners in government and regulatory bodies.

View the vox pops:
**bit.ly/aaipvoxpops**

# Body of Knowledge

**The Body of Knowledge is an openly accessible resource for the community, capturing the experience of that community in usable, practical guidance.**

The structure we published in 2018 has been validated by stakeholders, and this structure is now being populated with guidance and information about what has worked well, or not so well, in a broad range of domains.

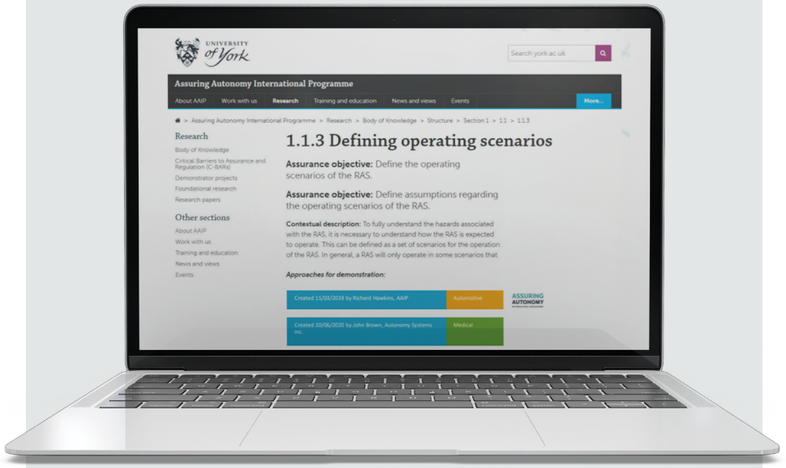
### A resource for safety professionals

To make the Body of Knowledge a usable resource for all domains and technologies each assurance objective will have different approaches for demonstration from a variety of domains. You will be able to search the resource for the particular area of safety assurance you are working on, and read about others' experience of doing that same piece of work.

### A practical philosophy

The Body of Knowledge is not intended to be a standard: it is a practical guide, giving safety engineers and others a quality resource on safety assurance to refer to when developing autonomous systems.

The Body of Knowledge will complement work that is being done to develop standards and guidelines – it is more guidance-led and less prescriptive.




An example of what you'll find in the Body of Knowledge for each assurance consideration. Different approaches may be provided for each assurance consideration. Some guidance may be domain specific.

### Community contributions

The demonstrator projects we are funding have already started providing guidance for the Body of Knowledge, along with guidance from Programme Fellows and members of the York team. The resource will be continually added to and updated.

We would also like this resource to grow from other members of the community working on the assurance of RAS – the more people and organisations that input and get involved, the more useful it will be to us all.

To discuss with us how you can contribute your experiences and guidance to the Body of Knowledge please contact us **assuring-autonomy@york.ac.uk**

View guidance in the Body of Knowledge: **bit.ly/aaipbokguidance**



## How to use the Body of Knowledge

The guidance in the Body of Knowledge is practical and based on the experience of members of the community. Here we have outlined exactly what you'll find and how to use the guidance.

# Foundational research

**Our research this year has centred on demonstrating, with sufficient confidence, that machine learning (ML) components can perform their tasks safely (i.e. with the risk of human harm as low as is reasonably practicable).**

We have seen some remarkable examples of ML components exceeding human performance, particularly in the field of clinical diagnosis. The current state-of-the art for ML focuses on this continual improvement of the performance and efficiency of the ML models: in order to allow the benefits of ML to be fully realised in safety critical domains the focus must be moved to safety.

To do this we have developed a safety assurance process: AMLAS (Assurance of Machine Learning for use in Autonomous Systems). This process is for the engineering of ML components in which the assurance evidence can be generated at each stage of the ML lifecycle. It gives the first systematic, documented approach to safety assurance of ML components, with the aim of giving others the confidence they need to use, certify or regulate the component or the system it is part of.

**Find more detail on the AMLAS process by folding out these pages**

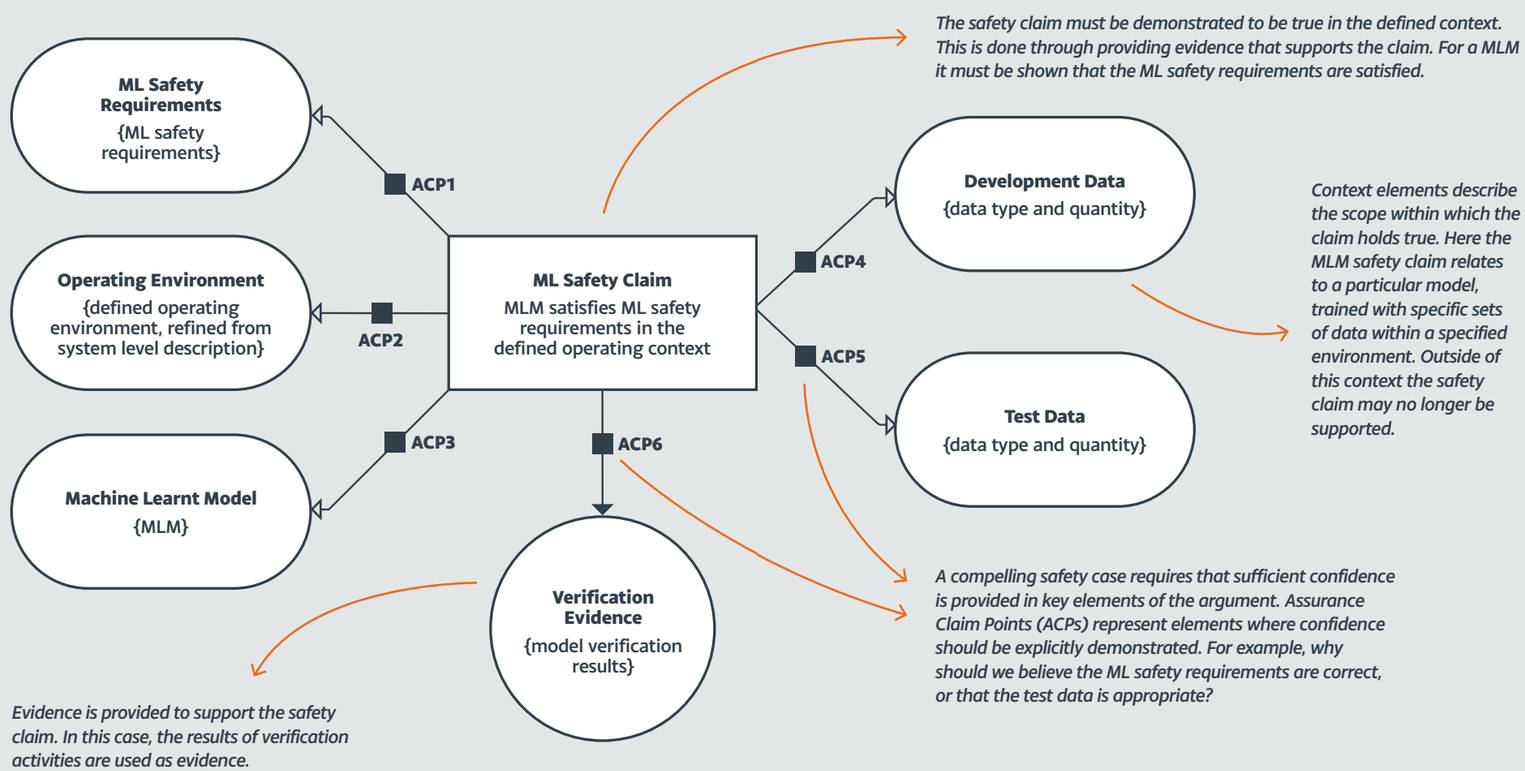
### Patterns into process

The development of AMLAS started with the team's work on assurance argument patterns that guide the structure of the confidence arguments required for every claim made about the safety of the component (i.e. assurance claim point).

A safety case is a structured argument supported by evidence. The figure below shows a pattern for the structure of a safety case relating to a Machine Learnt Model (MLM).

Established practice for developing a MLM does not systematically and explicitly provide the information needed in order to instantiate the patterns and provide the evidence.

Our AMLAS process has been developed in order to do this. You can find more detail on the AMLAS process and how it is used to create safety arguments by folding out these pages.




Example assurance pattern for ML

### Presented papers

More details have been presented in our work over the year, with further publications planned for 2020.

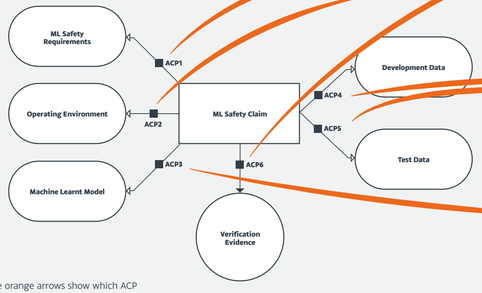
- *Assuring the Machine Learning Lifecycle: Desiderata, Methods, and Challenges* – Ashmore, R., Calinescu, R., Paterson, C. (preprint arXiv.org, May 2019)
- *Confidence Arguments for Evidence of Performance in Machine Learning for Highly Automated Driving Functions* – Burton, S., Gauerhof, L., Hawkins, R., Habli, I., Sethy, B. (WAISE workshop, September 2019)
- *A Pattern for Arguing the Assurance of Machine Learning in Medical Diagnosis Systems* – Picardi, C., Hawkins, R., Paterson, C., Habli, I. (38th International Conference on Computer Safety, Reliability and Security, September 2019)
- *Assurance Argument Patterns and Processes for Machine Learning in Safety-Related Systems* – Picardi, C., Paterson, C., Hawkins, R., Calinescu, R., Habli, I. (AAAI's Workshop on Artificial Intelligence Safety, February 2020)

Find links to all of our published papers: **bit.ly/aaipresearchpapers**

# Assurance of Machine Learning for use in Autonomous Systems

## Pattern

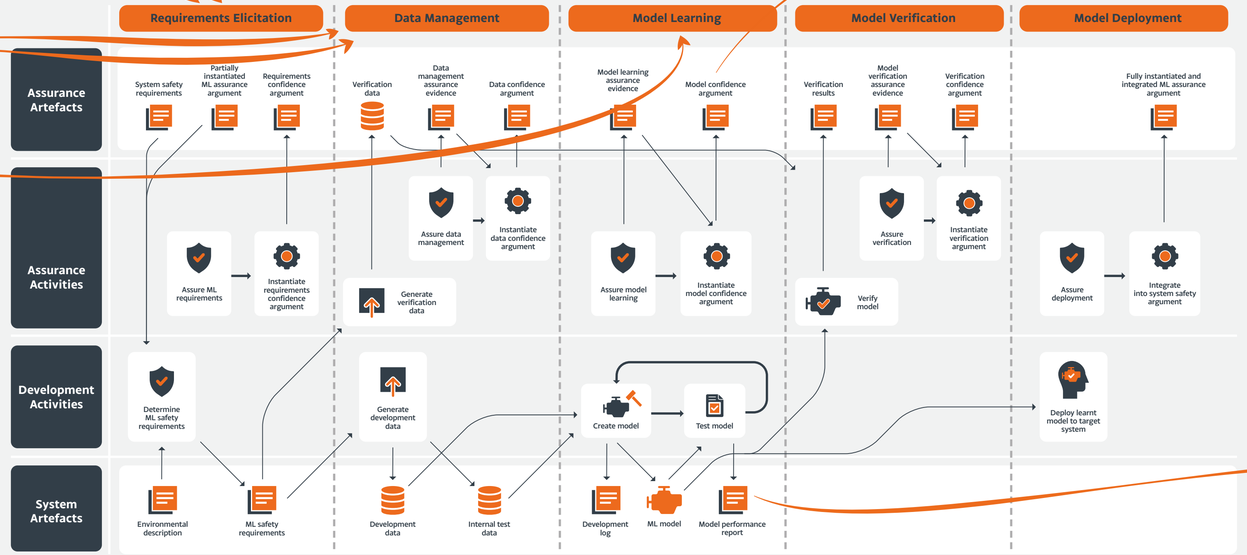
The confidence arguments for the ACPs shown on the pattern diagram are developed using the AMLAS process (shown on the right).




The orange arrows show which ACP relates to which phase of the machine learning lifecycle in the process.

An example of the outcome of this, for ACP3, is shown on the right hand page.

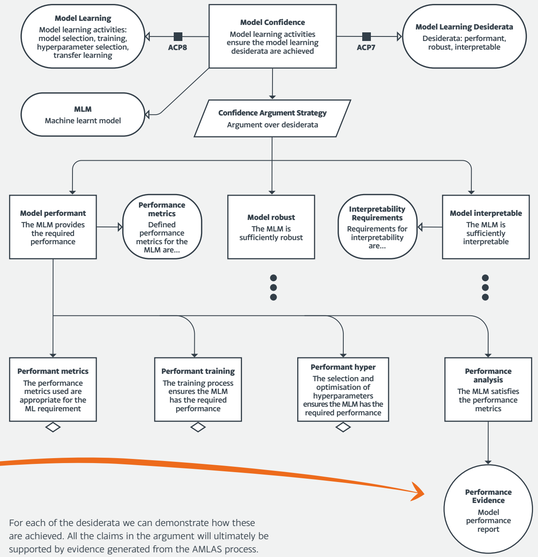
## The AMLAS process




## An example

Our patterns give the structure of the confidence arguments, based around the process activities and sets of identified desiderata (desired properties). The diagram below shows the model confidence argument that would be used to support assurance claim point 3. For model learning, the desiderata we identified were:

- **Performant** – the MLM should provide the required performance
- **Robust** – the MLM should perform well in circumstances where the inputs encountered at run-time are different to those present in the training data
- **Interpretable** – the MLM should be able to produce artefacts that support the analysis of its output, and thus any decision based on it.



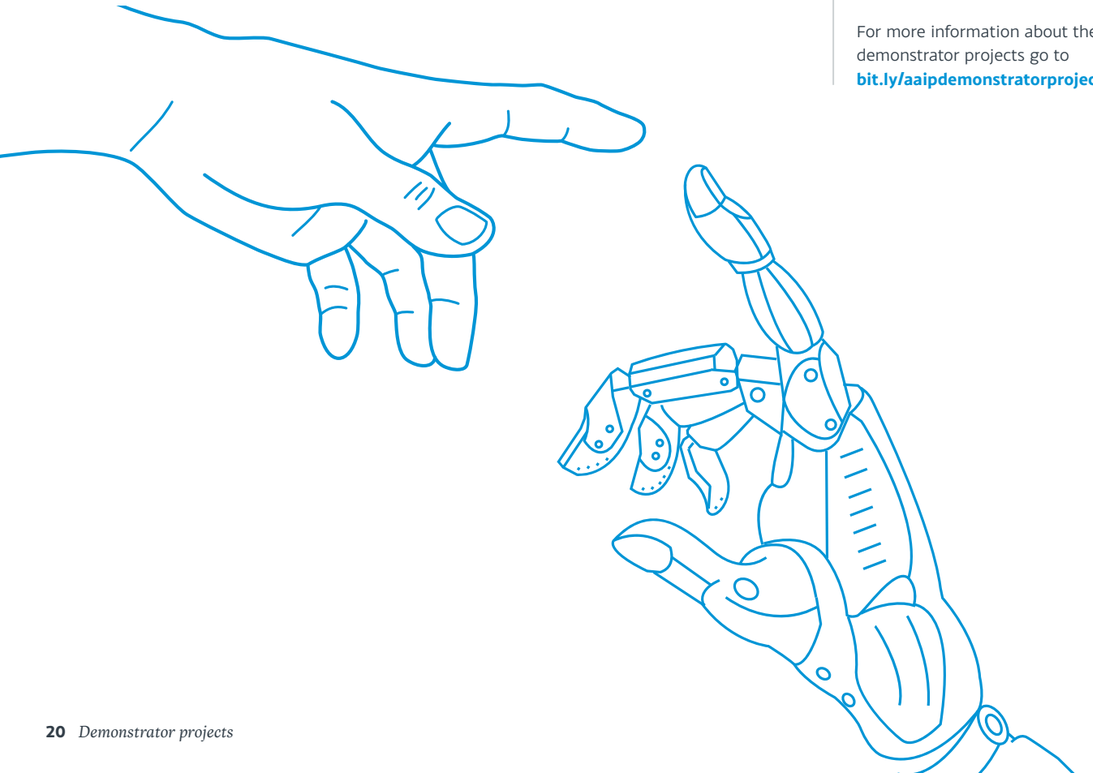

For each of the desiderata we can demonstrate how these are achieved. All the claims in the argument will ultimately be supported by evidence generated from the AMLAS process.

# Demonstrator Projects

**Our dual approach of foundational research at York and domain-specific research through our demonstrator projects continues to work well to help us solve some of the challenges outlined in the Body of Knowledge.**

In 2019 we invested a further £1.2million in demonstrator projects that will advance the safety assurance of robotics and autonomous systems (RAS), while building public trust in such technology. The project teams are using their research to develop guidance for the Body of Knowledge so that others in the community can use this to assure the safety of their own technology. As we move into 2020 we look forward to sharing this guidance with you.

For more information about the demonstrator projects go to **bit.ly/aaipdemonstratorprojects**





The SUCCESS team with the Volvo simulator

## Safety Assurance of Cooperating Construction Equipment in Semi-Automated Sites (SUCCESS)

*Enabling the benefits of component-based design of autonomous machines by assuring their safety when they work together as a system-of-systems (SoS).*

**Introduction** This project is looking at SoS in the autonomous earth moving machinery domain, with a focus on investigating and adapting hazard analysis techniques, safety assurance through contracts, and confidence measures.

**Progress** The team has been investigating requirements, scenarios, and hazard analysis. Requirements work has identified the current workflow at the quarry site, modes of operation, safety aspects, and scenarios of collaboration/ cooperation between machines. Elaboration of these scenarios has specifically looked at remote take-over and implementing collaborative functions.

Hazard analysis work has found that existing methods are suitable for static or individual systems, but lack the ability to deal with the emergent or disruptive behaviours associated with SoS. The team has applied Petri nets to create an extension to System Theoretic Process Analysis (STPA), to model the states of the involved systems and analyse if state changes could lead to critical situations. They found Petri nets useful to identify critical state-related situations in SoS, but that the nets can become very complex.

The team has also investigated combining Hazard and Operability (HAZOP) and Fault Tree Analysis (FTA) for identifying and mitigating hazards in their SoS.

**Next steps** The team will be doing more work on simulation to support hazard analysis and providing guidance for the Body of Knowledge.

**Project team**

- Mälardalen University, Sweden
- Volvo Construction Equipment
- Safety Integrity AB

**Started in 2018**

The experimental work cell

## Safety of reconfigurable collaborative robots for flexible manufacturing systems (RECOLL)

*Can collaborative robots safely increase manufacturing productivity while transforming the role of the human operator?*

**Introduction** The team has set up an experimental work cell in the machining department of a manufacturing company in Northern Italy, where MCM flexible manufacturing systems and robots are in use. This site enables the team to record the behaviour of human workers alongside their interaction with the robot system.

**Progress** The preparation of the work cell allowed the team to consider all intended uses by the operators. From this they formalised a model of the collaborative task, including steps, notable positions, expected occupancy of the cell, etc.

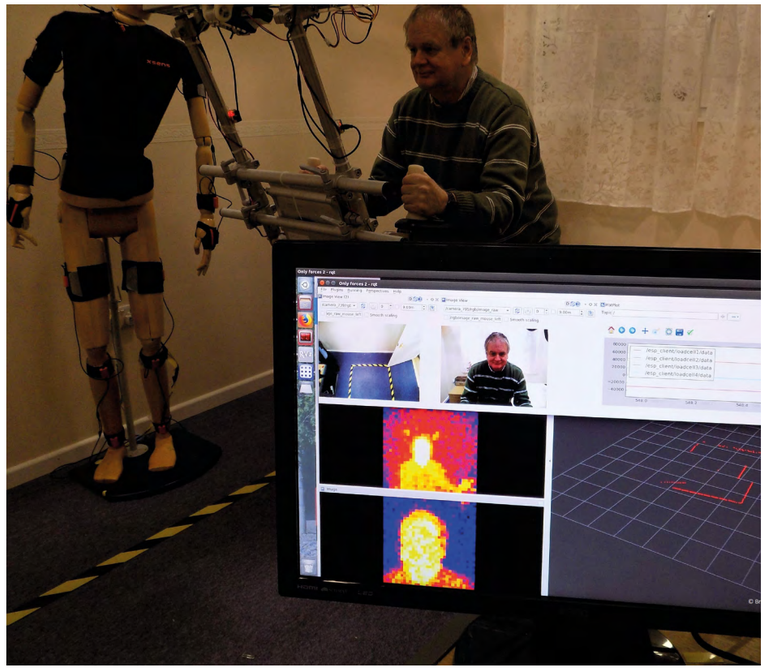
The team is now using sensors and a software tool that tracks human movements to record data from the collaborative task. They will then identify the principal components of the regular elements of the human-robot collaboration (e.g. rate of task sharing, distances, contacts) and the potential deviations.

**Next steps** The team will analyse the data from the work cell in order to determine the probability of a hazard and its correlation with an incorrect task execution. This will enable them to design a work cell with the capability of recognising situations that can lead to undesired risks – something that will be generalisable to other domains where collaborative robots are in use.

**Project team**

- Machining Centers Manufacturing (MCM) SpA
- University of York
- National Research Council of Italy, Institute for Intelligent Industrial Systems and Technologies for Advanced Manufacturing

**Started in 2018**



A range of sensors (e.g. IR and RGB-D cameras, laser scanner, load cells) provide data for developing AI algorithms to ensure safe operation of the CHIRON system in different test scenarios © Bristol Robotics Lab, UWE

# Assistive robots in healthcare

*How can robots and artificial intelligence be used safely to improve quality of life and increase independent living in an ageing population?*

**Introduction** This project uses the CHIRON overhead robotic system that provides physical and cognitive assistance to frail older adults. It is investigating and evaluating the safety and regulatory requirements of this type of close human-robot interaction.

**Progress** The team has completed a review of existing standards and guidelines for an assistive robot, finding that the current assistive robotics standard and hazard analyses do not take into consideration the issues relating to vulnerable users with accessibility needs. They also fail to address issues relating to situations where users are in direct physical contact with the assistive robot, and using it for physical support, while it is moving.

Workshops with care providers have helped to determine issues associated with supporting people with a range of impairments. The outcomes from these have been translated into use cases and scenarios. In combination with Environmental Survey Hazard Analysis the outcomes are being transformed into a set of functional and safety requirements.

**Next steps** The next stage is a task identification phase to develop a task requirements specification. Then each task selected for design and development will be developed, verified and validated to comply with the stated requirements, and hence in turn with industry standards. This will involve the acquisition and installation of sensor hardware and software components, to enable experiments with users.

**Project team**

- Bristol Robotics Laboratory, University of West of England
- Designability Charity Ltd

**Started in 2018**



# Towards identifying and closing gaps in assurance of autonomous road vehicles (TIGARS)

*Investigating the assurance gaps and challenges for first generation autonomous systems and researching techniques and engineering processes for addressing them*

**Introduction** The TIGARS team has been continuing their research to address some of the important assurance challenges of autonomous road vehicles that they have identified this year.

**Progress** The project has included experimental trials with a TEV (TIGARS Experimental Vehicle) and investigating resilience and safety requirements, validation and verification of machine learning, security-informed safety, and defence in depth and diversity.

Work in the international standardisation area has continued, and the team has held regulatory workshops in the UK and Japan.

The project has now entered its final phase: concluding the research studies, and reporting technical results, guidance and recommendations.

The team has completed a series of TIGARS Topic Notes on key areas:

- Resilience and safety requirements
- Open systems perspective
- Static analysis and formal verification
- Testing and simulation
- Defence in depth and diversity
- Experimentation
- Security-informed safety
- Standards and guidelines
- Assurance framework

**Next steps** The project will support the sharing of techniques and strategies through the TIGARS Topic Notes, the Body of Knowledge, and guidance publications for the UK Centre for the Protection of National Infrastructure. The partners will use the results of the project for their evaluation and development of RAS.

**Project team**

- Adelard LLP
- City, University of London
- Kanagawa University
- Nagoya University
- Witz Corporation

**Started in 2018**

The TIGARS team

Photo credit Dr. Nick Reynolds, Royal Derby Hospital

# Safety Assurance of Autonomous Intravenous Medication Management Systems (SAM)

*What if we could improve patient outcomes by using automated systems to administer just the right amount of medication at the right time to patients in intensive care units?*

**Introduction** This research is focused on the clinical scenario of an intensive care patient who requires blood sugar level control through intravenous insulin administration. Four use scenarios at different levels of automation and autonomy have been identified, ranging from the current use scenario to that of an autonomous infusion device.

**Progress** The identified use scenarios have been used to explore stakeholder perceptions about risk during interviews with patients, regulatory bodies, technology developers, and hospital staff. The team is also undertaking risk analyses of the scenarios using bow-tie analysis, human reliability analysis, and Functional Resonance Analysis Method (FRAM).

To support the development of a safety assurance strategy, the team is using the Safety Modelling, Assurance and Reporting Toolset (SMART) software tool, enabling them to define, express and analyse all aspects of a safety case. The work so far has identified several human factors challenges that need to be considered when using autonomous technology within clinical systems. Examples include handover between the system and clinician, the impact on the relationship between clinicians and patients, overreliance on autonomous systems, and supervision and monitoring of the technology.

**Next steps** The team will undertake further stakeholder interviews, and an additional study of how autonomous infusion devices might affect performance variability. They are also working with the Chartered Institute of Ergonomics and Human Factors to produce guidance for the consideration of human factors in the development and use of AI in healthcare.

**Project team**

- Human Reliability Associates Ltd
- NHS Digital
- University Hospitals of Derby and Burton NHS Foundation Trust

**Started in 2018**

# CSI: Cobot - Confident safety integration for cobots

*Removing the cage and curtains: how can we assure the safety of cobots to support increased human-robot collaboration in manufacturing?*

**Introduction** The collaborative robotics market is expected to exceed $9.0billion by 2025. However, safety and trust issues are hindering deployment in collaborative manufacturing processes: international standards for safe human-robot collaboration are in their infancy (ISO/TS 15066), difficult to enforce in practice, and so manufacturers are falling back on non-collaborative segregation-based safety methods including caging and light curtains. This substantially reduces the benefit of collaborative robotic systems.

**Research overview** This project will demonstrate how novel safety techniques can be applied to build confidence in the deployment of uncaged collaborative robot systems operating in spaces shared with human users. Existing collaborative processes provided by the project's industrial partners will act as case studies and demonstrators. These vary in complexity, but they provide a tractable safety problem whilst providing a good representation of current industry applications and needs.

To make the issue tractable, research will address specific safety elements related to the key issues identified by partners:

- Volumetric sensing
- Security
- System testing
- Safe operation-mode switching

This will result in evidence to support the assurance of general collaborative robot systems to enable further deployment of collaborative robots in manufacturing. Regulators will be involved to advise on compliance with evolving standards.

**Project team**

- University of Sheffield
- University of York

**Started in 2019**



# Safe-SCAD - Safety of shared control in autonomous driving

*How do humans and machines safely share control of an autonomous car?*

**Introduction** Ensuring and assuring the safety of shared control in autonomous driving is very challenging due to the uncertainties associated with measuring the level of situational awareness of safety drivers while not in control of the vehicle, and with the mapping of such measures to control hand-back times and likelihood of success.

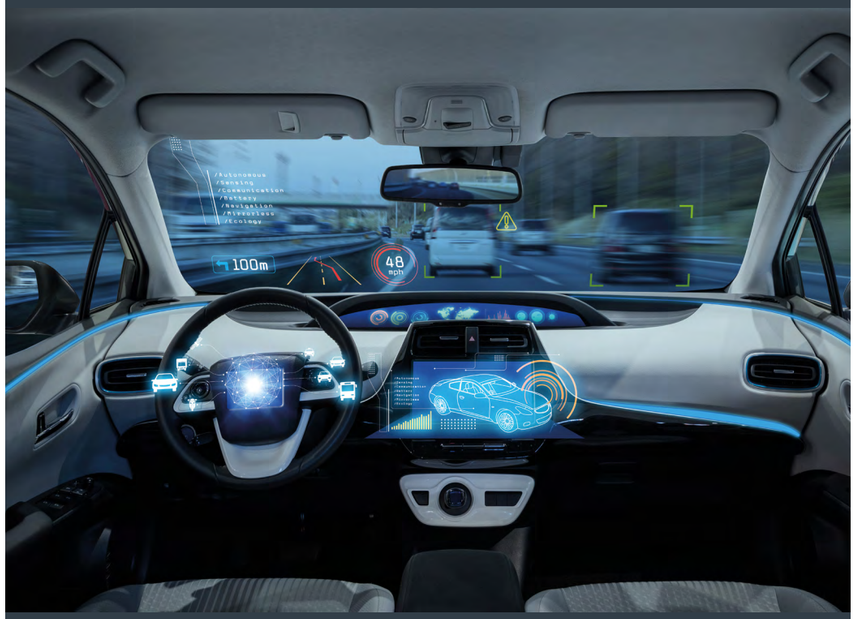
**Research overview** This project will extend, adapt and integrate the team's recent research and the latest advances from human behaviour and cognitive modelling, verification of deep neural networks, and automated controller synthesis to tackle these challenges. An advanced semi-autonomous driving simulator will be used to deliver methods for ensuring and assuring the safety of shared control in autonomous driving. The project will make significant and generalisable contributions in the areas of:

- Shared autonomy
- Training and verification of machine learning
- Monitoring of autonomous systems by human operators

**Project team**

- University of Virginia
- Carnegie Mellon University
- University of York

**Started in 2019**




# SAX: Sense-Assess-Explain - Building trust in autonomous vehicles in challenging real-world driving scenarios

*How can we develop autonomous vehicles that can explain the decisions they take?*


Off-road trials with a specifically adapted ORI Landrover

**Introduction** Understanding the decisions taken by an autonomous machine is key to building public trust in robotics and autonomous systems (RAS). This project will design, develop, and demonstrate fundamental AI technologies in real-world applications to address this issue of explainability.

**Research overview** The aim of the project is to build robots, or autonomous vehicles, that can:

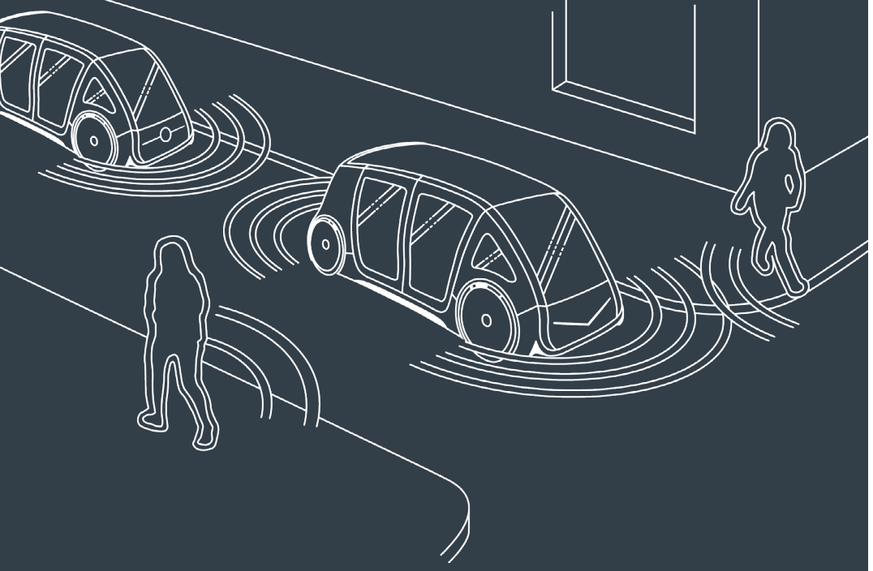
- Sense and fully understand their environment
- Assess their own capabilities
- Provide causal explanations for their decisions

In on-road and off-road driving scenarios, the project team will study the requirements of explanations for key stakeholders (users, system developers, regulators). These requirements will inform the development of the algorithms that will generate the causal explanations. The work will focus on scenarios in which the performance of traditional sensors (e.g cameras) significantly degrades or completely fails (e.g. in harsh weather conditions). The project will develop methods that can assess the performance of perception systems and adapt to environmental changes by switching to another sensor model or a different sensor modality. For the latter, alternative sensing devices will be investigated (including radar and acoustic sensors), which can guarantee robust perception in situations when traditional sensors fail.

**Project team**

- University of Oxford

**Started in 2019**



# Automatic Testing Mechanism: Towards an NCAP-like rating for robotics and autonomous systems

*What if we could develop an automatic rating system to assess the safety of RAS?*

**Introduction** System developers want to be able to thoroughly and quickly test their new software for robotics and autonomous systems (RAS). Regulators need to be able to quickly assess the safety of software and software updates that control RAS. This project will design an automatic testing mechanism for the autonomous software that controls the system, including artificial intelligence and machine learning based controllers of RAS.

**Research overview** The testing mechanism will generate scenarios that are likely to cause dangerous or undesirable behaviour in the RAS software being tested, by applying recent advances in decision making under uncertainty.

In the longer term, this project will form the core component of an automatic rating system to assess the safety of RAS – akin to the NCAP (New Car Assessment Programme) rating for cars.

**Project team**

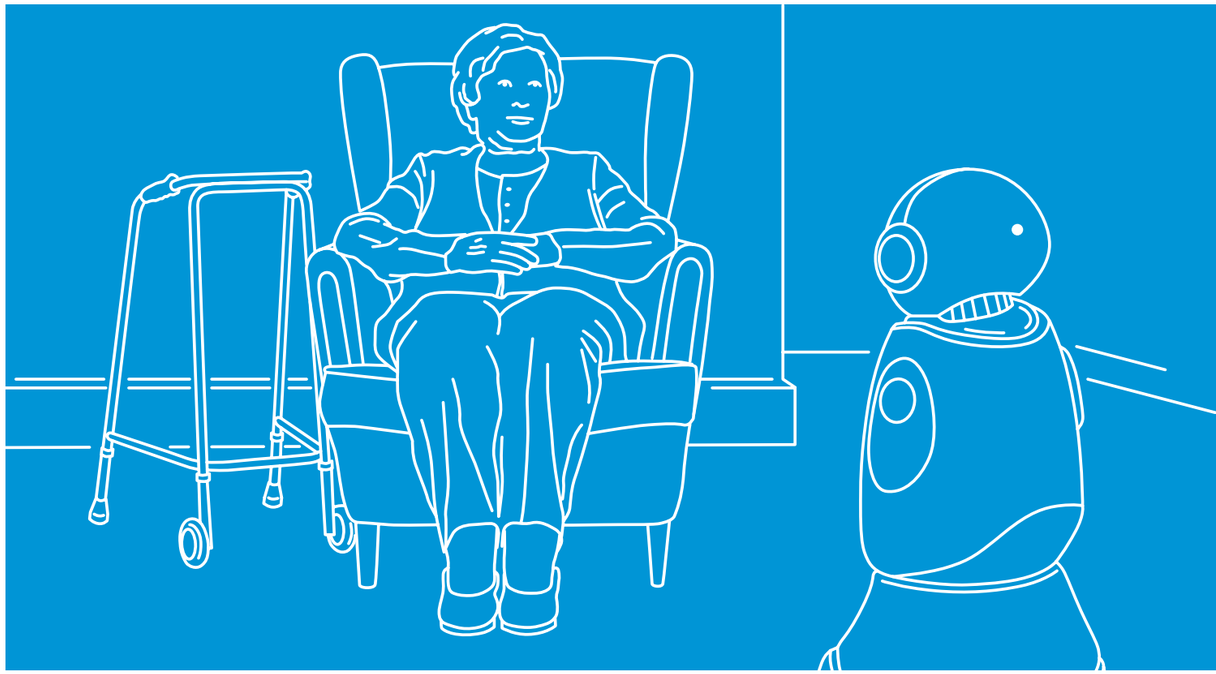
- Australian National University

**Started in 2019**



# Social acceptability: balancing social credibility and effective performance of safety functions

*How do we achieve social credibility of assistive robots in the home?*



**Introduction** Assistive robots offer significant benefits to an aging population. They can undertake care-giver support such as carrying food items, reminding the user to take medication, and alerting the user if the oven has been left on, alongside companionship and social functions.

In order to be accepted by users, however, the robot must have social credibility: it must behave in a way that is empathic and socially interactive in order to perform its safety functions effectively.

**Introductory work**
The team undertook a literature review which identified that the social effects of assistive robots are not typically factored into hazard analysis. In addition, there is often very little consideration of the ways in which the social performance of an assistive robot is affected by safety features (e.g. automatic stops, avoidance of physical contact).

**Experimental work** The team designed an experiment to validate their hypothesised link between social credibility and safety. 15 participants worked with a robot which violated social norms (e.g. it spoke abruptly or was too close to the human when speaking), and 15 worked with one which complied with social norms (e.g. it spoke politely and remained an appropriate distance from the human while interacting).

While sitting at a table and completing cognitive tasks (e.g. Sudoku puzzles) participants were interrupted by their robot four times, and they decided whether or not to perform an action in response. The robot interrupted with information that:

1. The oven in the kitchen was left on
2. The power sockets in the kitchen were on
3. Some of the power sockets in the kitchen were still on
4. A Pepper robot in a different room was overheating while charging

The team observed participants via camera feeds and smart sensors and made objective measurements of responses to the interruptions (e.g. movement, action, time taken). Participants also completed a post-experiment questionnaire.

**Results** The team was able to identify a number of trends from the collected data:

- The robot which complied with social norms was considered much more socially credible
- Participants were more likely to respond to a socially credible robot
- Participants were more likely to take action to mitigate the hazard when interrupted by a socially credible robot

The identified trends provide some indication of how safety assurance might be affected by an autonomous system's social behaviours in this domain. The most notable impact is on a user's willingness to accept the robot's assessment of hazards and the extent to which the user considers it necessary to cross-check these against their own experience.

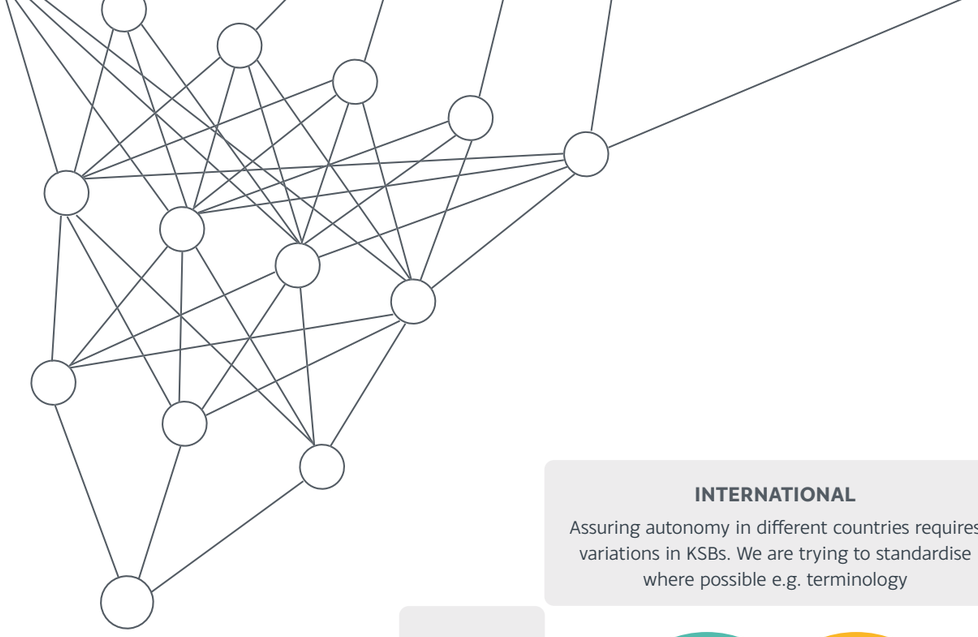
This behaviour of disbelief was seen most clearly with the second power socket warning, with the results demonstrating that participants are more likely to accept a socially credible robot's assessment of a situation, even where this directly contradicts their own experience (as they had turned off the sockets). This general effect can be taken to indicate that when it comes to assessment of safety-critical situations, users are more likely to believe a robot that they consider socially intelligent instead of one lacking social competency.

**Project team**

- Robot House, University of Hertfordshire

**Started in 2019**

# Education and Training



**People make systems safe. This is just as true with robotics and autonomous systems (RAS) as it is with more traditional systems; but what the people need to know to make the system safe is changed by the technology.**

Our education and training addresses three groups of people: those developing the technology, those working with the technology (e.g. operators), those assuring and regulating the technology.

### Evolving knowledge, skills and behaviours

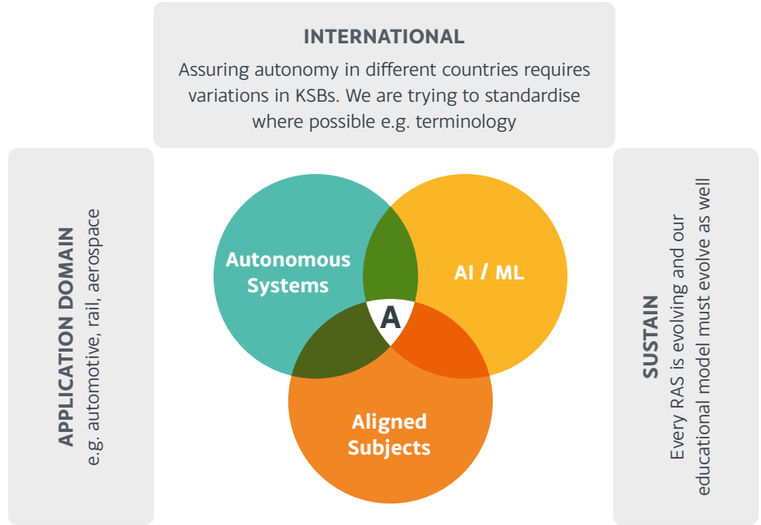
Working in safety assurance of robotics and autonomous systems (RAS) requires different knowledge, skills, and behaviours (KSBs), to working with more traditional complex systems: increasingly open environments, machine learning and new stakeholders have disrupted the educational model for teaching safety critical systems engineering.

Over the past year we have been working with stakeholders, including demonstrator project teams, to understand the competences required of those assuring and regulating RAS. These competences require knowledge, skills and appropriate behaviours. We are continuing to work with them to understand how the KSBs are changing and evolving according to domain and technology, and how they need to progress over time.

### A new model of education

During 2019 we have been refining our educational model. If we consider the areas of autonomous systems, artificial intelligence/machine learning, and aligned subjects such as security and ethical issues (see educational model opposite), the training and education possibilities are very large. To have the greatest effect we are focusing our work around area A of the educational model, starting with safety professionals and regulators.

Safety-Critical Systems Symposium (Bristol, February 2019)




The AAIP educational model (to be presented at Safety-Critical Systems Symposium 2020)

### A story of four parts

In order to reach the various stakeholders who have different needs and varying levels of existing safety education and experience, we need (and will offer) four different types of education:

- **Academic education** – masters and level 7 apprenticeships
- **Industrial education** – onsite at companies and online
- **Research dissemination** – conferences, journal papers and workshops
- **Informal dissemination** – digital (news articles, blogs, etc) and non-digital (demonstrations, events etc)

### 2020 and beyond

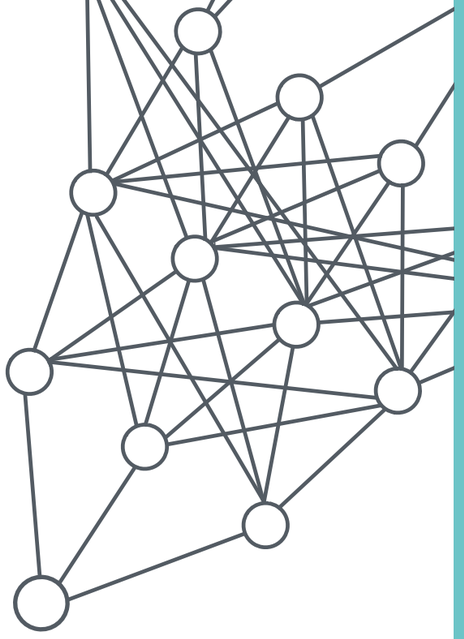
As we develop our academic education offer, we are already running industrial workshops and training. We currently offer a one-day or three-day course, tailored to the needs of individual companies, including a case study based on that organisation.

Please contact us on **assuring-autonomy@york.ac.uk** to find out more and to discuss further if you would like us to support your training programme.

**Workshop on autonomy and AI**

As part of the Safety-Critical Systems Symposium in February 2019 we ran an over-subscribed workshop for safety engineers and others. It offered an introduction to the assurance of robotics and autonomous systems, artificial intelligence, and machine learning. Using real-world case studies it covered the steps to assurance delegates need to consider when employing autonomous systems in their domain.

# The Future



**We want 2020 to be a year of influence. We will bring together the results of our foundational research and demonstrator projects to engage with the community and help establish guidelines for regulation and industry best practice.**

The guidance that comes from our research and demonstrator projects populates the Body of Knowledge and is useful for safety engineers and technology developers to give them practical advice on ways to assure the safety of their technology. It will also help shape conversations with those regulating the technology.

Regulators will be a major focus of our work in 2020. They must be confident that the systems presented to them are safe. Through a series of workshops we will help regulators to understand the questions that need to be answered to give them justified confidence in a system's safety. Understanding what challenges remain will also help us to direct future demonstrator projects.

We will also be planning for our future as we break ground on the new facility at the University of York that will bring together the work of the Programme with York's expertise in robotics and advanced communications. The facility will include specialist indoor and outdoor design and testing facilities to help us move closer to safe autonomy. When the facility opens in 2021 we will use it to engage further with industry, regulators and the public.

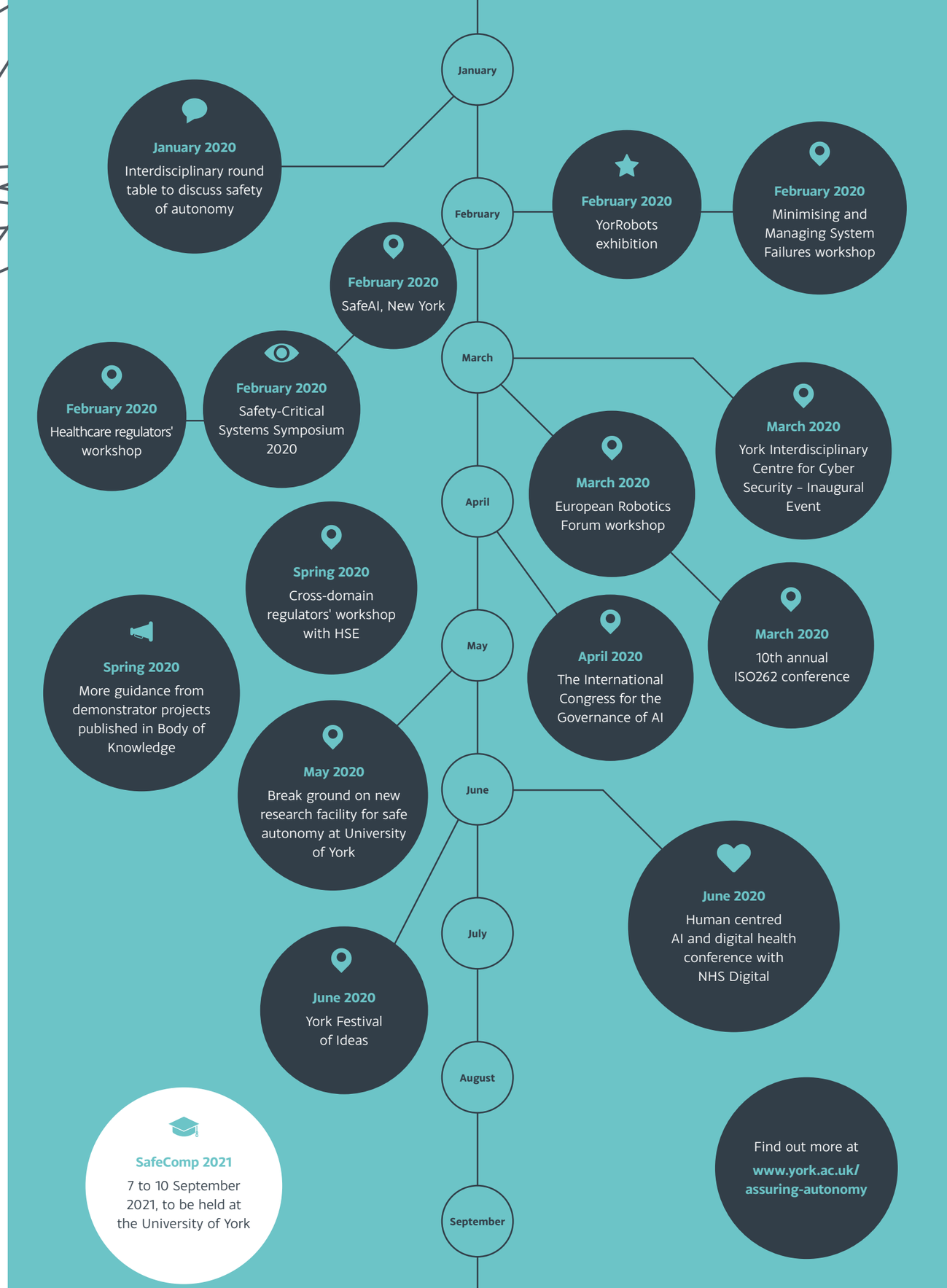
Next year is set to be another busy and exciting one. As one member of the public who we spoke with said, "In the future we'll look back on this time and it will be an industrial revolution." We're ensuring that the revolution is a safe one.

**Work with us**

The challenge of assuring the safety of robotics and autonomous systems requires further funding, collaborative research, technology case studies, evolving regulations, and more – we'd love to hear from you if you would like to collaborate with us for the greater good.


**+44 (0) 1904 325345**
**assuring-autonomy@york.ac.uk**

**@AAIP_York**
**www.linkedin.com/company/assuring-autonomy**
**https://medium.com/@AssuringAutonomy**

UNIVERSITY of York

Lloyd's Register Foundation