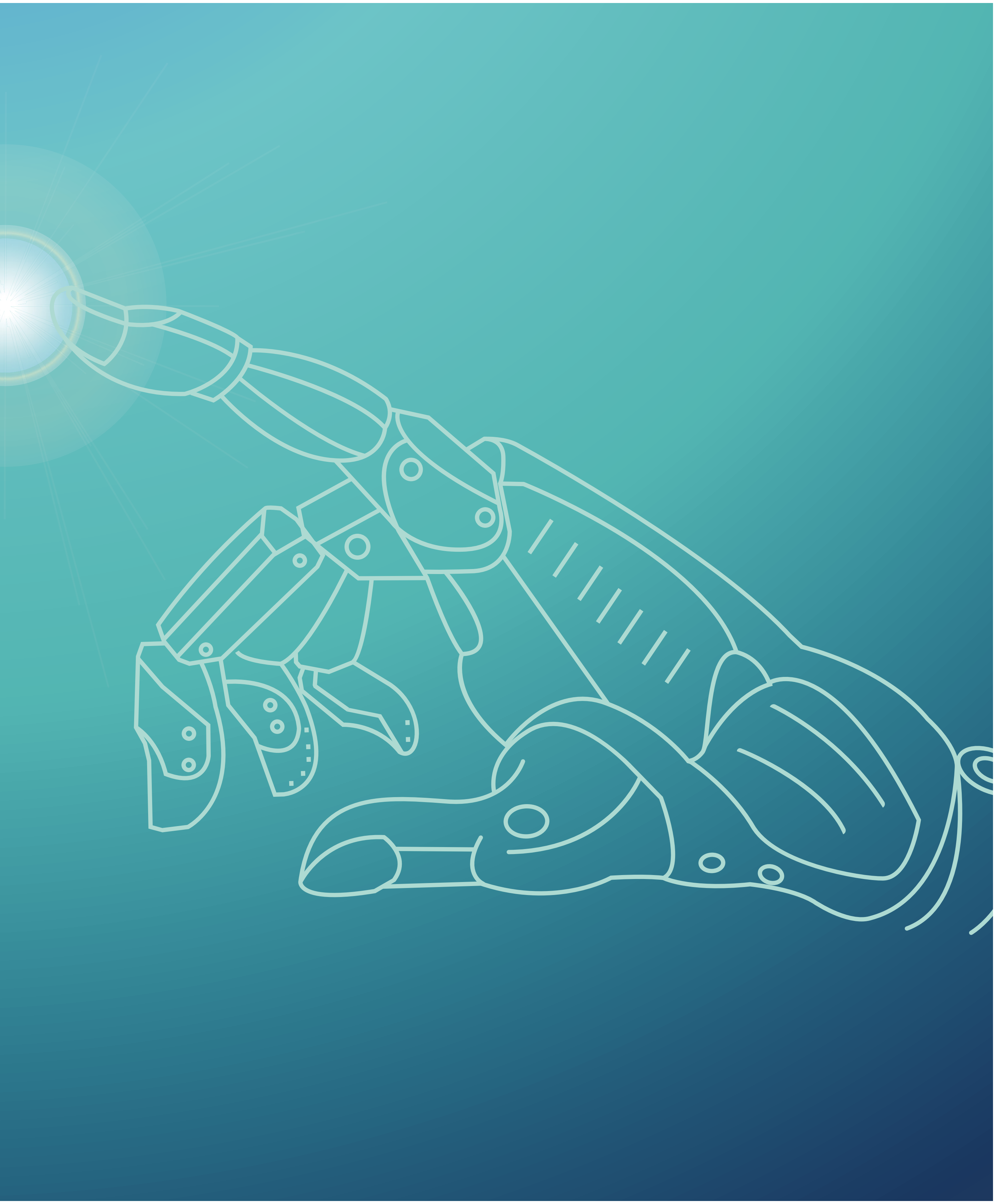
ASSURING AUTONOMY

INTERNATIONAL PROGRAMME

A YEAR IN REVIEW 2018



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

Key

- Funders
- Demonstrator projects
- Collaborative links
- Programme Fellows

Welcome

2018 was a high point of the current technological revolution. A driverless car was part of the procession at London's Lord Mayor's Show, the UK harvested its first hands-free hectare, and AI was used to detect serious eye conditions.

We are starting to be able to visualise just what promise robotics and autonomous systems (RAS) hold.

The Assuring Autonomy International Programme is ensuring that this innovation continues and grows. Our work is helping the safety community, regulators, technology developers and others to understand, achieve and assure the safety of critical autonomous systems.

Our first year has been about building strong foundations: a team of experts at York; real-world demonstrator projects underway across the globe; basic research tackling the underlying challenges; and creating a structure for the Body of Knowledge. We have also been supporting an international community to work together; undertaking an in-depth needs analysis to understand what training and education is required; and opening the dialogue with the public so that we can start to understand their questions, reservations and enthusiasm.

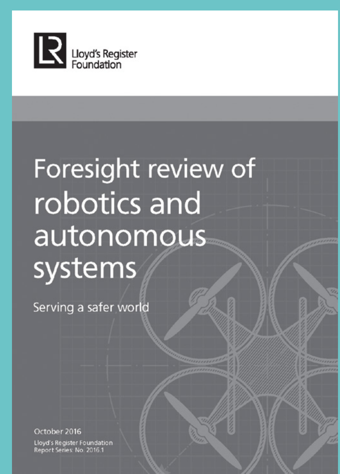
We've learnt a lot too. The focus of the Programme on assurance and regulation is exactly right – but the challenges are even greater than we anticipated. Although there are some very impressive new commercial prototypes, it has become clear that the step from successful demo to "prime time" is very significant. Some systems developers understand this well but there are many who are over-optimistic, thus there is a real need for our training and education programme to span from senior decision makers to the engineers designing and assessing these systems.

Our initial work and interaction with the demonstrators has given us a better understanding of the research landscape, and this will influence our future plans, including calls for new demonstrators.

As we move through 2019 the Programme and its work continue to grow, just as the challenge itself increases with further developments in technology. We will welcome more specialists to the team, fund additional demonstrator projects, present research papers, and host and attend workshops and conferences to talk with others about how best to address the global challenges of assuring RAS. The Body of Knowledge will be tested and brought to life with the first outputs from the research we are supporting, and we will launch our education and training programme.

We invite you to be a part of this future: collaboration is the key.

Professor John McDermid OBE FREng
Programme Director



www.lrfoundation.org.uk

Technology has the power to transform lives. RAS could change the way we live and work in ways that we are only just imagining. We must ensure that their introduction is done safely, or we risk putting people in harm's way or jeopardising the benefits that RAS can bring to us all.

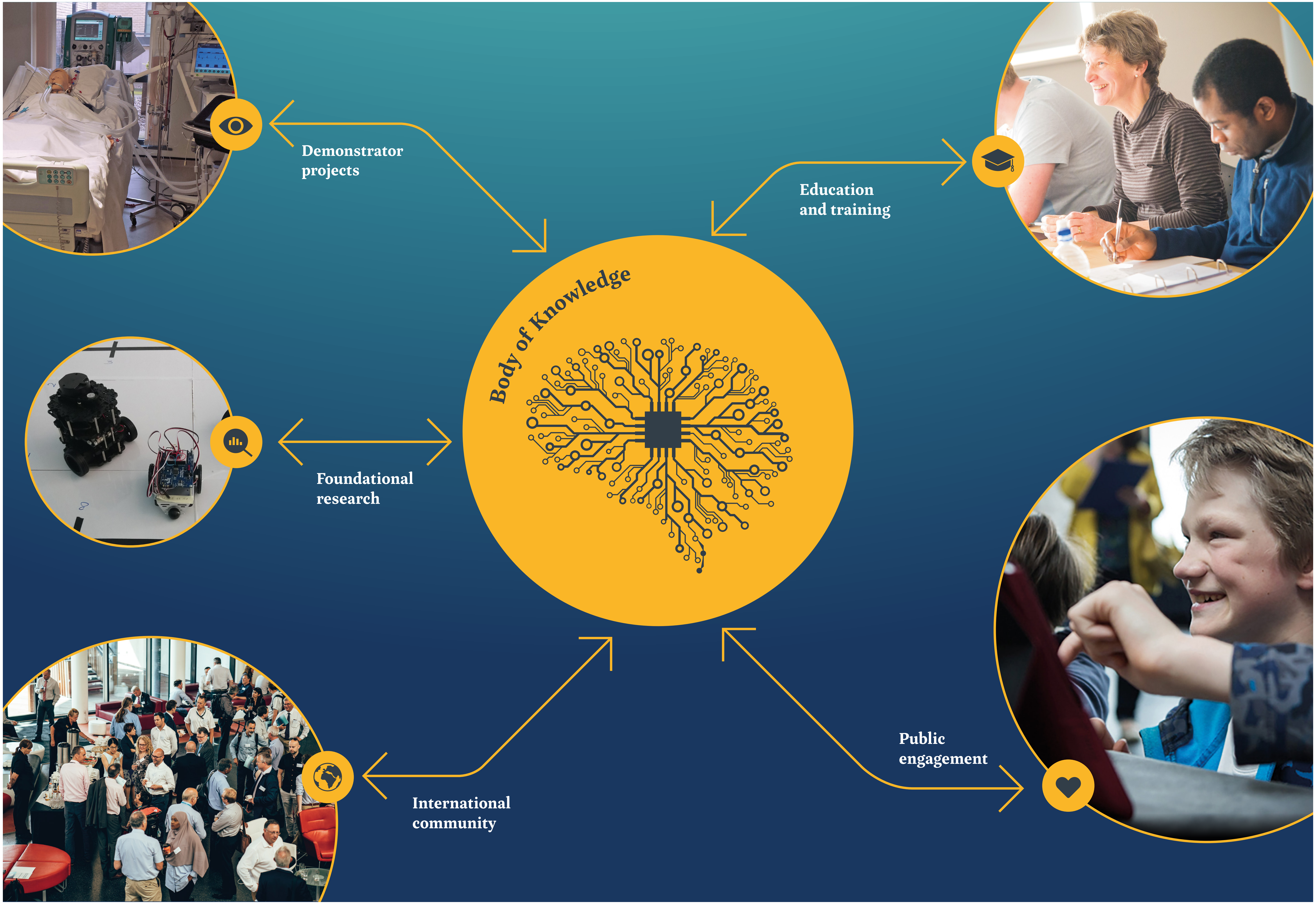
The University of York is our natural partner for this Lloyd's Register Foundation-funded programme to address the gaps we identified in our Foresight Review of Robotics and Autonomous Systems. The research that started last year at York, and through demonstrator projects, will be the cornerstone of ensuring that we all benefit from the safe introduction and adoption of RAS.

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

This is a unique and exciting project for the University of York. Developing assurance frameworks and standards for complex safety-critical systems is one of the many areas where York is world leading in its research.

I'm also particularly grateful for our partnership with Lloyd's Register Foundation whose generous support has enabled the innovative and ground breaking work detailed throughout this annual report.

Professor Saul Tendler
Acting Vice-Chancellor and President,
University of York


Demonstrator projects
Education and training
Body of Knowledge
Foundational research
International community
Public engagement

Demonstrator Projects

Technology continues to develop rapidly. Central to the work needed to assure these novel robotics and autonomous systems (RAS) are collaborative real-world use cases. This bottom-up approach of taking newly developed technology in genuine scenarios complements the foundational research taking place at the University of York.

The demonstrator projects are addressing one or more Critical Barriers to Assurance and Regulation (C-BARs) we have identified (see page 13 for more information).

Our partners are testing new approaches to assuring and regulating their innovative technology. They are identifying what works, what changes to regulatory frameworks are required, and what we need to know in order to safely introduce and adopt these systems.

The outputs from each, whilst developed in a specific sector, will be relevant across a range of domains, and will contribute to the Body of Knowledge to benefit the entire community.

In 2018 we funded five demonstrator projects with partners from across the world. This first £2.2 million worth of research is leading the way; helping to identify what is needed for us to work together to solve the challenges in assuring RAS.

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

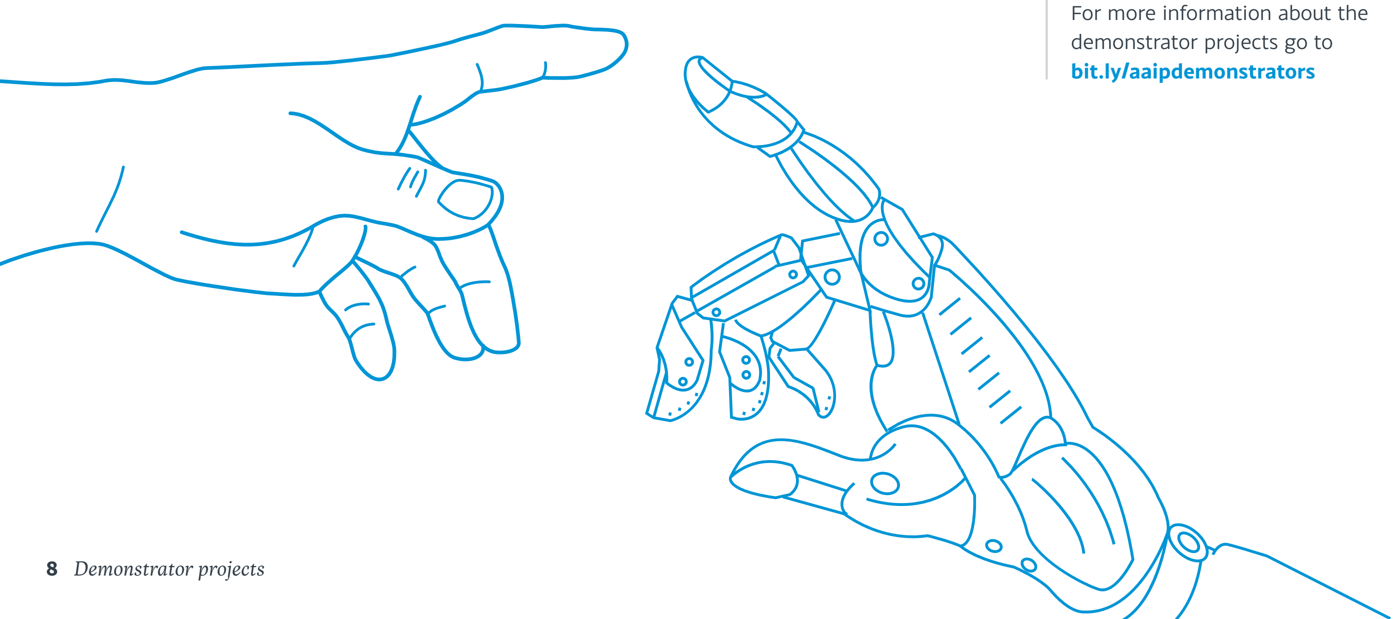


Photo credit Dr. Nick Reynolds, Royal Derby Hospital

Safety assurance of autonomous intravenous medication management systems – requirements and strategies (SAM)

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

Introduction The complexity of intravenous (IV) medication preparation, administration and management leads to frequent mistakes. Highly automated technology, and machines that can make decisions independently of healthcare professionals, could help to provide the answer through personalised treatments and reduced errors.

Research overview The project is looking at what kinds of safety assurance are needed by the different stakeholders, how well current safety assurance methods can evaluate highly automated IV medication technology, and whether these methods can form the basis for assurance strategies that will satisfy the needs of each stakeholder.

Progress The project team has been working with an intensive care patient who requires blood sugar level control through IV insulin. Four use scenarios at different levels of autonomy have been identified; from current status to the use of an autonomous infusion device, which dynamically adjusts the insulin delivery based on an analysis of the patient's physiological factors.

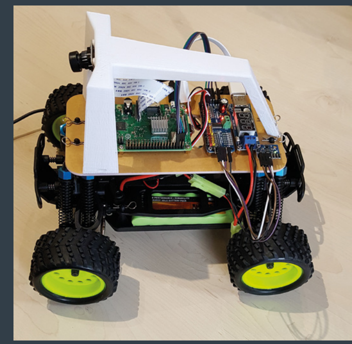
Next steps The identified scenarios will be used to explore stakeholder perceptions about safety assurance and regulation. The team will undertake risk analyses of each use scenario, and will develop a safety assurance strategy for each.

C-BARs being addressed

- Risk acceptance
- Handover
- Incident and accident investigation

Project team

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



Donkey autonomous car being used in some of the initial experiments

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

How can we improve the assurance of autonomous vehicles by adapting current engineering processes and technical analysis of first generation autonomous systems?

Introduction The first generation of autonomous systems are being introduced despite little change to regulation and standards. This project aims to improve the assurance of these initial autonomous systems, and identify how existing approaches for assurance need to change to address current and future systems.

Research overview
The project aims to: (1) identify and assess current autonomous systems engineering approaches, and identify assurance gaps; (2) assess the feasibility of deploying current state-of-the-art static analysis, verification and testing techniques; (3) investigate how to address identified assurance gaps with new analysis approaches based on verification of machine learning, using simulation and test strategies, and an in-depth analysis of defence; and (4) provide recommendations to regulatory and policy organisations and standards developers.

Progress The project started with a joint UK-Japan workshop. The team has started to identify assurance gaps in an experimental vehicle, to set up test facilities, and to develop experimental and theoretical approaches to static analysis and dynamic assurance. The project partners have been active in the standardisation area both internationally and in the UK and Japan.

Next steps The team will be developing focused experiments on static analysis, data collection, in-depth analysis and requirements modelling, alongside continuing work on standards.

C-BARs being addressed

- Verification
- Role of simulation

Project team

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

The sit-to-stand modular robot in the Bristol Robotics Laboratory assistive robotics test bed at UWE

Assistive robots in healthcare

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

Introduction As populations across the world get older we need to find ways to continue to maintain high quality health and social care services in ways that enable people to lead healthy and fulfilling lives. RAS offer an opportunity to provide personalised and cost effective support for a range of care-related tasks including physical and social assistance and physiotherapy.

Research overview
This project is investigating and evaluating the safety and regulatory requirements of close-proximity physical human-robot interaction in a range of care environments, from a human-centred perspective.

Progress Practical use cases are being developed in consultation with a range of stakeholders, including potential end-users, occupational and physiotherapists, carers and regulators, that focus on the use of a modular ceiling-based robotic system that can provide physical and cognitive assistance to frail, older adults.

Next steps The project team will run a series of experiments to facilitate safety analysis of the system, taking into account reasonably foreseeable potential hazards posed either by the person or the environment. Following this, a comprehensive set of the system's safety requirements will be defined with a detailed specification of safe-guarding hardware and software functionalities and control architectures.

C-BARs being addressed

- Human-robot interaction
- Incident and accident investigation
- Bounding behaviour
- Handover
- Cross-domain usage
- Risk acceptance
- Validation

Project team

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

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 Collaborative robotics can transform the role of operators of flexible machining systems, allowing them to take on more value-added tasks, leaving repetitive or assistive tasks to robots within the same shared workspace (see image below). When hybrid tasks are allocated dynamically and/or layouts are changed during operation, how can safety of operators be assured?

Research overview
This project is studying the safety-related human-robot behaviour (movements, layout occupation, contacts, near misses, etc) in a prototype machining production setup. The allocation and handover of tasks to machine and human varies over the duration of the process. The objective is to sample quantitative safety-critical data in a real production environment to identify trends and hazard precursors.

Progress The project team has identified a manufacturing company and site in which to set up an MCM flexible manufacturing system and robot applications. They have established how to introduce robot–assisted activities (e.g. risk assessment for ISO/TS 15066 compliant collaborative modes, design of tools and sensor equipment), the expected impact on operators' behaviour and tasks, and the safety training that will be required.

Next steps The team will identify, map out and analyse the safety-critical behaviours in the newly set up collaborative manufacturing space. They have established environmental recording equipment for tracing (anonymous) human actions related to movements and intended tasks of the robot system, part of which is determined and loaded at runtime.

C-BARs being addressed

- Handover
- Human-robot interaction

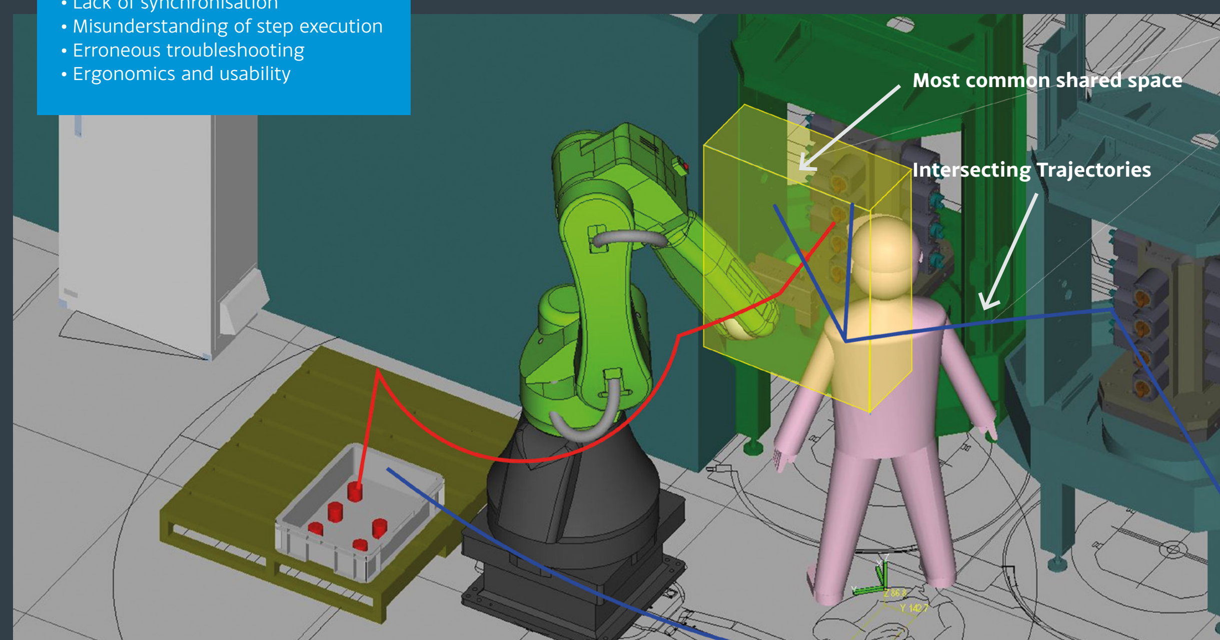
Project team

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

Hazards and hazardous situations

- Impact hazards
- Crushing, entanglement
- Probability of errors
- Lack of synchronisation
- Misunderstanding of step execution
- Erroneous troubleshooting
- Ergonomics and usability

Human (blue traces) and robot (red traces) share tasks in the same fence-less workspace. Most frequent common occupancy is in front of a machine pallet. Multiple patterns and types of access to the area occur: external tracing is used to log the distribution of occupancy, synchronisation of operator, deviations from intended use, patterns in idle time. [Safety note: workspace is supervised by IEC 61496-1 ESPEs, FANUC DCS2.0 safety functions, ISO/TS 15066 collaborative modes]



Quarry site automation

Safety assurance of cooperating construction equipment in semi-automated sites (SUCCESS)

What assurance methodology and standards are needed to enable us to assure the safety of autonomous machines that are working collaboratively as a system of systems?

Introduction
Quarry sites require a series of repetitive and sometimes dangerous tasks. A system of cooperating construction machines offers an opportunity to take humans out of harm's way whilst providing a cost-effective and more environmentally-friendly process for construction.

Research overview
This project is investigating safety assurance and certification of cooperative functions in a quarry environment, and is clarifying what additions or modifications of current safety standards are needed to cope with such functions.

Progress The project team is currently investigating hazard analysis techniques.

Next steps Hazard analysis techniques will be adapted as appropriate for the quarry scenario and safety assurance modelled using contract formalisms and extensions.

C-BARs being addressed

- Handover
- Human-robot interaction
- Adaptation
- System of systems

Project team

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

Body of Knowledge

The full structure can be found by folding out these pages

The Assuring Autonomy International Programme is developing a Body of Knowledge (BoK) intended, in time, to become the definitive reference source on assurance and regulation of robotics and autonomous systems (RAS).

The Body of Knowledge will be cross-domain, cross-technology and cross-application. It will cover all aspects of assurance and regulation, and it will curate and present accessible information which is useful to a range of stakeholders.

The scope of the Programme, and therefore of the BoK, is huge. We have been liaising with stakeholders to identify what the BoK must include, the level of information needed, and what is extraneous: knowing what to leave out is as important as knowing what to include.

Specification

Some of the knowledge we include in the BoK will inevitably be specific in nature. However the guiding principles are that the resource should:

- Be as general as possible
- Use established assurance approaches wherever possible

We intend to indicate where:

- Traditional methods can be directly applied
- Traditional methods must be extended or adapted for application to RAS
- New approaches are required

Scope

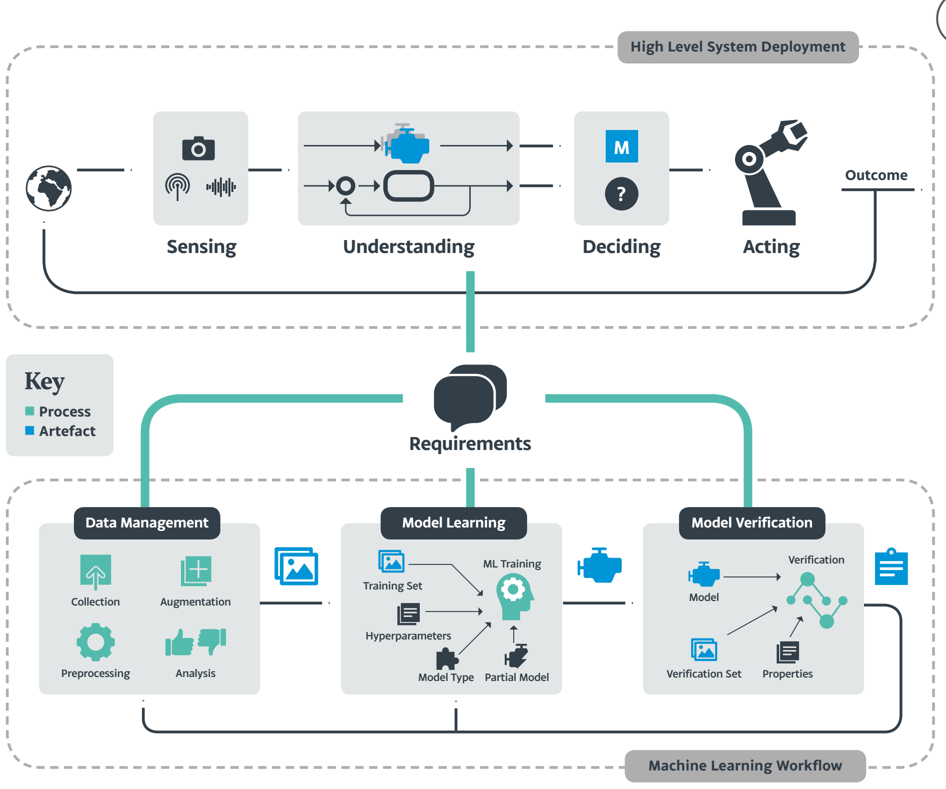
Four main areas of assurance and regulation have been identified through discussions:

- **Defining required behaviour** – defining what it means for a RAS to be 'safe'
- **Implementing a RAS to provide the required behaviour** – demonstrating the sufficiency of the implementation
- **Understanding and controlling deviations from required behaviour** – identifying and controlling sources of deviation
- **Gaining approval for operation of RAS** – regulating operation in the specified environment

In each of these areas a set of assurance objectives has been defined. Each objective will then be populated with a description of appropriate means for demonstrating the achievement of the objectives.

Structure

The structure of the BoK has been influenced by a variety of approaches. The first is the SUDA model of an autonomous agent (shown below and in a more detailed format on page 19). This has been used to structure many of the technical aspects of the BoK.



The SUDA model and explanation of the machine learning workflow

Secondly, a number of key areas of the BoK have been identified as Critical Barriers to Assurance and Regulation (C-BARs). The C-BARs are those obstacles that could result in:

A safe system not being deployed
- resulting in low rates of technology adoption despite safe operation

An unsafe system being deployed
- potentially leading to accidents and incidents

Find out more about the C-BARs at **bit.ly/aaipcbars**

An invitation to you

We have spoken to a range of stakeholders: industry, regulators, researchers and others. They have helped us to outline the specification and scope for the BoK, and to create the structure.

In order to make the BoK a successful and useful resource it will be constantly developing as the state-of-the-art matures and technologies evolve. As we start to curate and populate the BoK now is the time to give us feedback on its structure – will it help you to do your job, how will you use it, is there anything missing, or something that's not needed? Please take a look at the structure on the fold out pages.

You can also view a full document with an outline of each area of assurance consideration at **bit.ly/aaipbokstructure**

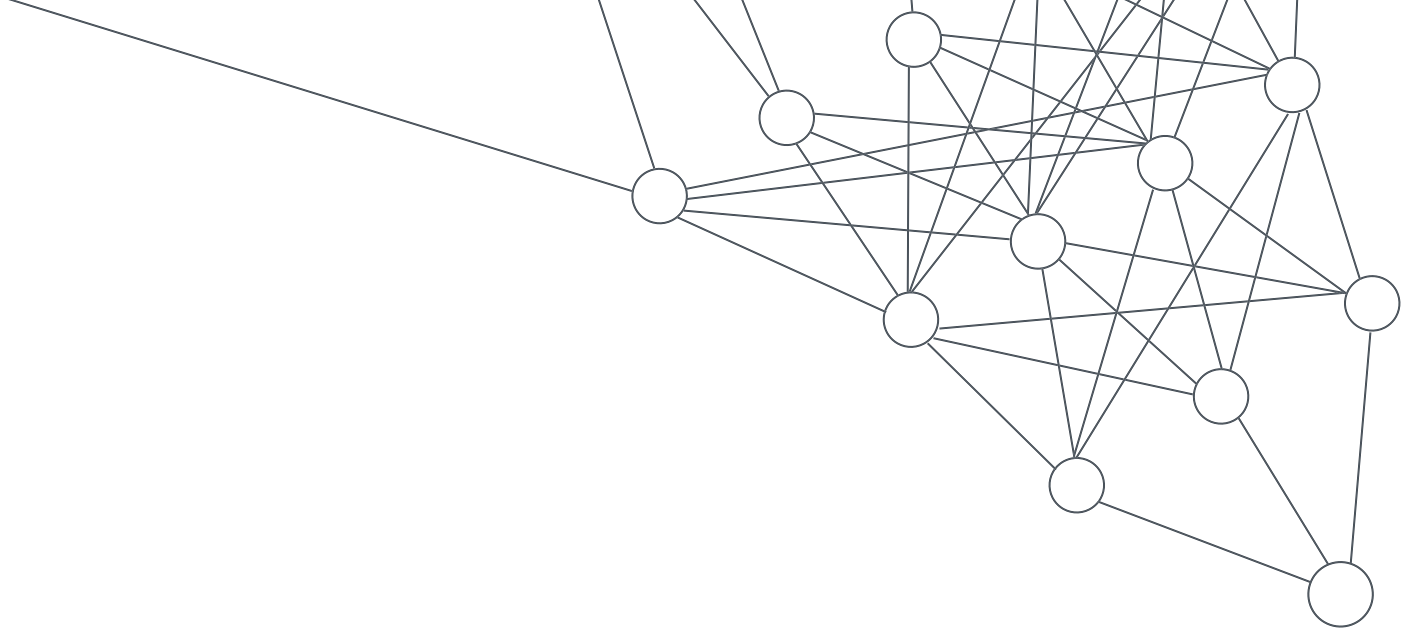
Then let us know what you think at **assuring-autonomy@york.ac.uk**
Thank you.

Following discussion with a variety of stakeholders the structure below has been established for the Body of Knowledge.

Each heading represents an assurance consideration on which guidance will be provided. This guidance will consist of:

- A definition of assurance objectives relating to that area of consideration (e.g. things that must be demonstrated when putting a RAS into operation)
- A contextual description providing further information and rationale for the assurance objectives
- Details on approaches for demonstration that the assurance objectives are met. This will often define multiple alternative strategies for demonstration. These alternative strategies may reflect alternatives in the state-of-practice, or reflect different approaches required in different domains

References to information in the public domain will be provided where appropriate.



Education and Training

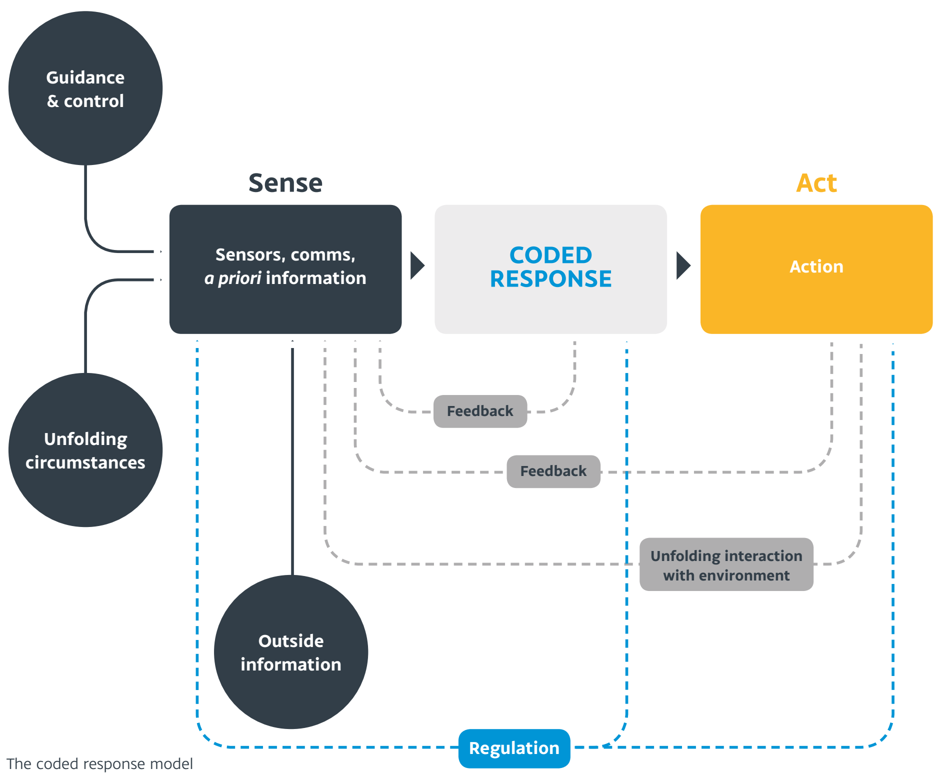
Safety Critical Systems Engineering has been an education discipline since the 1990s.

The University of York pioneered the teaching of this subject in the 1990s and still runs a successful MSc in this area. Even with the advent of robotics and autonomous systems (RAS) the core educational requirement for someone working with these systems remains: we still need a safety case, we need to know how to monitor the system, and to re-evaluate its safety during the system's life. But the systems themselves, the environments in which they operate, and the set of people who interact with them, is changing. As a result, the emphasis of the required education and associated training needs to evolve.

The starting point

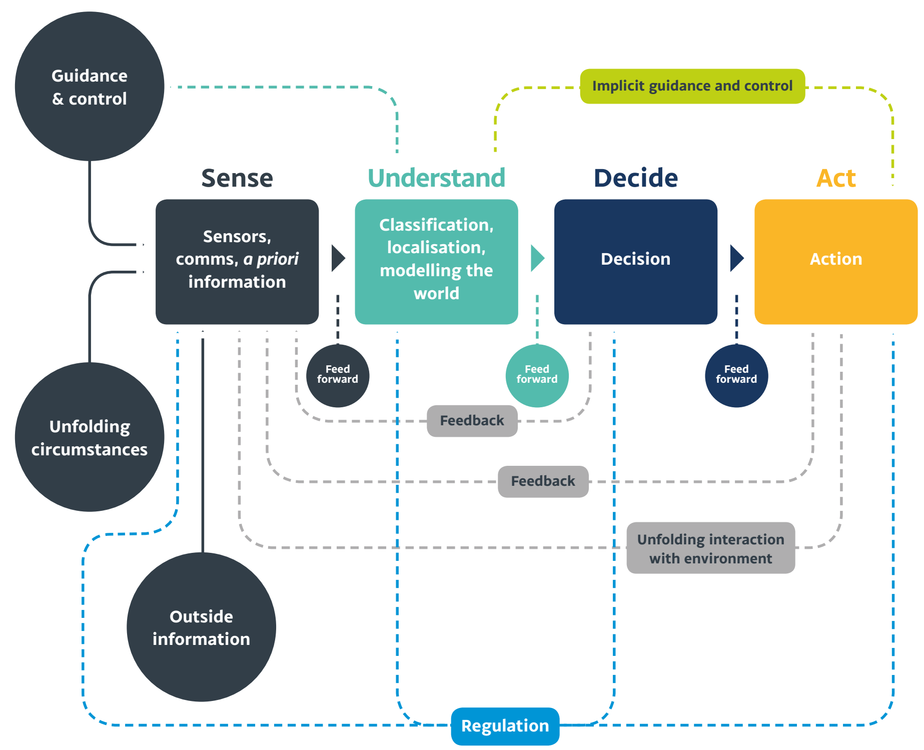
As human engineers, developers and operators, we have historically been expected to be able to understand the complex systems we work with. As a result we would either directly tell the machine what to do or would do so indirectly by coding in responses to particular inputs. Thus, we made the decisions about what actions to take in response to a situation and the form the actions took.

Education and training programmes taught us the Knowledge, Skills and Behaviours (KSBs) we needed to be able to do this. Further, we had sufficient knowledge and understanding to assure ourselves that the actions undertaken by the system were acceptably safe. Once again education and training provided the KSBs to undertake this task. The challenge for training and education now is that we are no longer in that world.



The coded response model

The new challenge



Adapted from Boyd's OODA loop (https://en.wikipedia.org/wiki/John_Boyd_(military_strategist)#The_OODA_Loop)

In the world of RAS the understanding and decision making parts of this SUDA model are more and more being provided by the machine rather than coded specifically by developers. Human interaction with the system is more one of guidance than direct command.

This has the potential to significantly change the KSBs needed by the stakeholders. It also increases the groups of stakeholders who need to interact with RAS (perhaps with cobots or within a connected system of systems) and therefore need to be educated and trained.

The strategy needed

We need to provide the right education and training to the right stakeholders so that RAS can safely and reliably be introduced across all domains. We are developing and undertaking a wide training needs analysis to understand who the new stakeholders are and what they, and existing stakeholders, need to know. Furthermore, the needs analysis will give us an insight into the consistency of views on who needs which KSBs.

We are now:

- Producing a map of the KSBs across RAS stakeholders relating to the assurance of RAS
- Developing level 7 risk and safety apprenticeships so that newcomers to the industry have the skills they need
- Creating CPD courses for those already working with safety critical systems to ensure their skills develop to match the results of the needs analysis
- Collaborating with the robotics industry to develop training packages for non-traditional stakeholders
- Taking the lessons being learned in demonstrator projects to use in course materials
- Investigating the use of innovative education resources, such as augmented reality and simulation

For more information go to:
bit.ly/aaipeducation

Foundational Research

The University of York has a strong history in pioneering research in the safety of complex systems. We are applying this expertise to address some of the key technical issues that underpin the challenge of assuring the safety of robotics and autonomous systems (RAS).

Our initial focus is on assuring the safety of artificial intelligence (AI), and on managing the risks associated with systems that learn while they are operating. These are core issues. Assurance of machine learning is perhaps the biggest technical challenge in dealing with RAS. In addition, classical safety assessment analyses systems prior to operation, knowing that the analysis will stay valid – but where systems learn in operation this is no longer true, challenging the foundation of our technical and regulatory processes. Hence this work is central to the Programme.

Dynamic risk assurance

When systems learn during operation how can we create appropriate safety cases and how can we explain what the machine has done?

Introduction RAS are expected to learn and evolve in order to deal with circumstances that cannot be predicted prior to their deployment. This means that the safety evidence and the overall safety case also have to evolve in order to reflect the nature of the actual risk, and explain and justify the means for managing this dynamic risk.

Research overview There are two main areas of focus. The first is exploring mechanisms for incorporating dynamism in safety cases, particularly for systems that use machine learning. The second is the notion of explainability and the extent to which this could be provided dynamically.

Progress The challenge of dynamism in safety cases is being addressed empirically, by focusing on real-world applications. One critical application, amongst others of interest, is healthcare. A safety case for a deep learning system for diagnosis and referral in retina disease (based on a study published in Nature Medicine by Google DeepMind) is currently being developed. This will: (1) create a complete and self-contained exemplar safety case for a real machine learning system; and (2) identify and analyse the assurance factors that are dynamic and expected to evolve (e.g. changes in clinical practice and scanning technology and frequency of retraining).

The first step to explore the core issue of explainability is a systematic literature review to appraise the current evidence on explainability of dynamic machine learning decisions and evaluate how this could support a RAS safety case.

Published papers The ethical dimension of explainability (e.g. the ability of a developer to explain the behaviour of the machine learning technology) has been explored in "The Moral Responsibility Gap and the Increasing Autonomy of Systems", published and presented at the International Workshop on Artificial Intelligence Safety Engineering (WAISE) in September 2018. It was shortlisted for the best paper award. Reinforcement learning and safety assurance has been examined in "What is Acceptably Safe for Reinforcement Learning?", which was also published and presented at WAISE 2018.

Next steps The outcomes of the study into the safety case for the deep learning system are expected to be reported at the 17th Conference on Artificial Intelligence in Medicine in June 2019.

Safety of AI techniques

How can we support safety assessments of RAS by improving the safety and transparency of the artificial intelligence (AI) and machine learning (ML) they use?

Introduction RAS increasingly use AI and ML to deliver useful applications in many areas of economy and society. This poses major challenges for safety assessment due to: the potential for inaccurate models being generated by ML; the lack of visibility of what AI has learnt prior to deployment; and the potential changes in behaviour from learning in operation.

Research overview The project is focused on developing methods and guidelines for three areas: (1) verifying and ensuring the safety of machine-learnt RAS components; (2) restricting AI to ensure the learnt behaviour is safe; and (3) augmenting AI techniques with the ability to reveal information about the learnt behaviour.

Progress In collaboration with the dynamic risk assurance project team and external partners, the first end-to-end methodology for developing trustworthy RAS control software and dynamic assurance cases was introduced. This supports the dynamic creation of assurance cases through the integration of assurance evidence generated both at development time and during operation. It enables the safe adoption of a new class of RAS – those for which only partial assurance cases can be produced during development, but for which the missing assurance evidence can be obtained through additional verification performed at run time.

Another key project result has been the delivery of an ML-based method that greatly improves the accuracy of an important type of probabilistic models that underpin decision making under uncertainty in systems such as RAS. Starting from a high-level description of a system, the method uses ML techniques to fit statistical models to observation data obtained from the real system. This fully automated method significantly reduces analysis errors for real-world systems, lowering the risk of invalid engineering decisions.

Key publications The new end-to-end methodology established by the project and partners was published in the November 2018 issue of IEEE Transactions in Software Engineering. The ML-based method was published as a preprint in IEEE Transactions in Software Engineering.

International community The project team contributed to establishing, and now participates in, an international study group for the development of the IEEE Guidelines for Verification of Autonomous Systems.

Next steps Further development of the results and methodologies.

For more information go to:
bit.ly/aaipresearch

International Community

The challenge of assuring RAS is one that goes beyond national borders. To best support the safe adoption of these technologies, our Programme must be global in its outlook, and work with partners internationally.

In our first year, we have worked to develop links with organisations on four continents (see map on pages 2-3), and expect to consolidate and strengthen these links, as well as extend our network, in future years.

Europe

The Programme is supporting five demonstrator projects in Europe, and 2019 will see the start of our involvement in both a Marie Sklodowska-Curie European Training Network on safer autonomous systems, and an EC Horizon2020 project on RAS for infrastructure inspection, both awarded in 2018. We have engaged with numerous regulators and policy-makers, and expect to progress these relationships further in 2019. In November 2018, our Director was invited to speak to the All-Party Parliamentary Group on AI. A final comment from the audience was that there needed to be much more investment in assurance and regulation of RAS.

Asia

In 2018, we participated in reciprocal visits to a number of research groups in Singapore, including the Lloyd's Register Foundation Institute for the Public Understanding of Risk, and we continue to develop partnerships here. One of our demonstrator projects is in partnership with automotive and academic collaborators in Japan, which will provide some of the first content for the Body of Knowledge.

Australasia

Following a visit to Australia in 2018, we hosted a return visit from The Autonomy, Agency and Assurance Innovation Institute (3A Institute), and also expect to support some work on a future demonstrator, having built good working links with researchers there.

North America

We have strengthened existing links to US organisations including the Naval Research Lab, with cooperation on verification techniques. Our Programme Manager was invited to speak on a panel with senior US and UK policy makers in Washington DC in December 2018.

Our international work will continue throughout the life of the Programme, and our Body of Knowledge and Programme Fellows will be important channels for this.

"Due to the scale of the task, it is vital that the research remains focused on open questions directly relevant to industrial applications. I see my contribution as providing direct insight into the key challenges the industry faces as well as the practicality of solutions developed within the research programme."

Programme Fellow
Dr Simon Burton

Organisation
Robert Bosch GmbH

Job title
Chief Expert – Vehicle Computer Software, Safety and Security

Areas of Fellowship research
System-level assurance cases, highly automated driving, safety of machine learning, and ethical and legal expectations on autonomous systems

Expected contributions to the Body of Knowledge Holistic assurance strategies for the safety of highly automated driving, and assurance cases for machine learning.

Increasing levels of autonomy in systems such as passenger vehicles and robotics require us to design systems that can cope with a huge number of environmental conditions that cannot all be predicted during the development phase. This requires a paradigm shift in the way we address the safety of these systems. We must develop the systems such that they remain demonstrably safe under all possible conditions. At the same time, the technologies required to implement this level of autonomy, such as machine learning, present yet more challenges for safety assurance. Existing safety standards do not transfer well to these classes of systems, we therefore need to develop new assurance strategies and argue the safety of the systems from "first principles". This in turn will require significant breakthroughs in a number of key research areas. As a representative of industry, participation in the AAIP provides me with valuable access to cutting edge research that I can transfer back into my day-to-day work.

You can read Simon's blog post, *Zen and the art of safety assurance for machine learning in autonomous driving*, here: **bit.ly/aaipzenblog**

Programme Fellow Rob Ashmore

Organisation Defence Science and Technology Laboratory (Dstl)

Job title Dstl Fellow

Areas of Fellowship research My main research interest involves finding ways of providing confidence in an Artificial Intelligence (AI) component that has been developed using Machine Learning (ML) techniques. I'm also interested in understanding what it means for a training set to be, in some sense, "good". It seems to me that this has to cover at least two aspects. Firstly, how well the data covers the problem domain; equivalently, this relates to how well the data covers the requirements associated with the AI / ML component. Secondly, how completely the data exercises the internal structure of the component; this is similar to the notion of, for example, branch coverage for traditional safety-related software.

Expected contributions to the Body of Knowledge I hope that the various conversations I've had during my time with the AAIP have contributed to both the structure and content of the Body of Knowledge. More specifically, I am supporting two AAIP colleagues, both from the University of York, in producing an academic paper that establishes ways of providing assurance for (and, consequently, providing confidence in the outputs of) the ML workflow.

"By bringing together a range of contributors, including academia, charities, government and industry, the AAIP is well positioned to address an urgent problem from a pragmatic perspective, without sacrificing a sound theoretical basis. I very much wanted to be a part of AAIP, to work alongside some very talented people and to try and make my own small contribution to overcoming the challenge we face. Being able to make this important work more accessible to Dstl was another key motivation."

Public Engagement

The benefits to society from the introduction of robotics and autonomous systems (RAS) could be huge. However, we must explore public perception, understanding and acceptance of these systems and their risks if we are to fully and successfully introduce them to market.

In parallel with other groups, the Programme is starting that conversation. Through blog posts, articles, and events we are talking about RAS; the potential, the dangers, and how we can work to assure their safety. We are asking the public what they already know, and what they need to know to be able to accept their introduction.

This work is currently at an early stage but will grow over time. We need to continue to engage with the public as the technology develops and the world we are living in evolves.

Festival of Ideas 2018

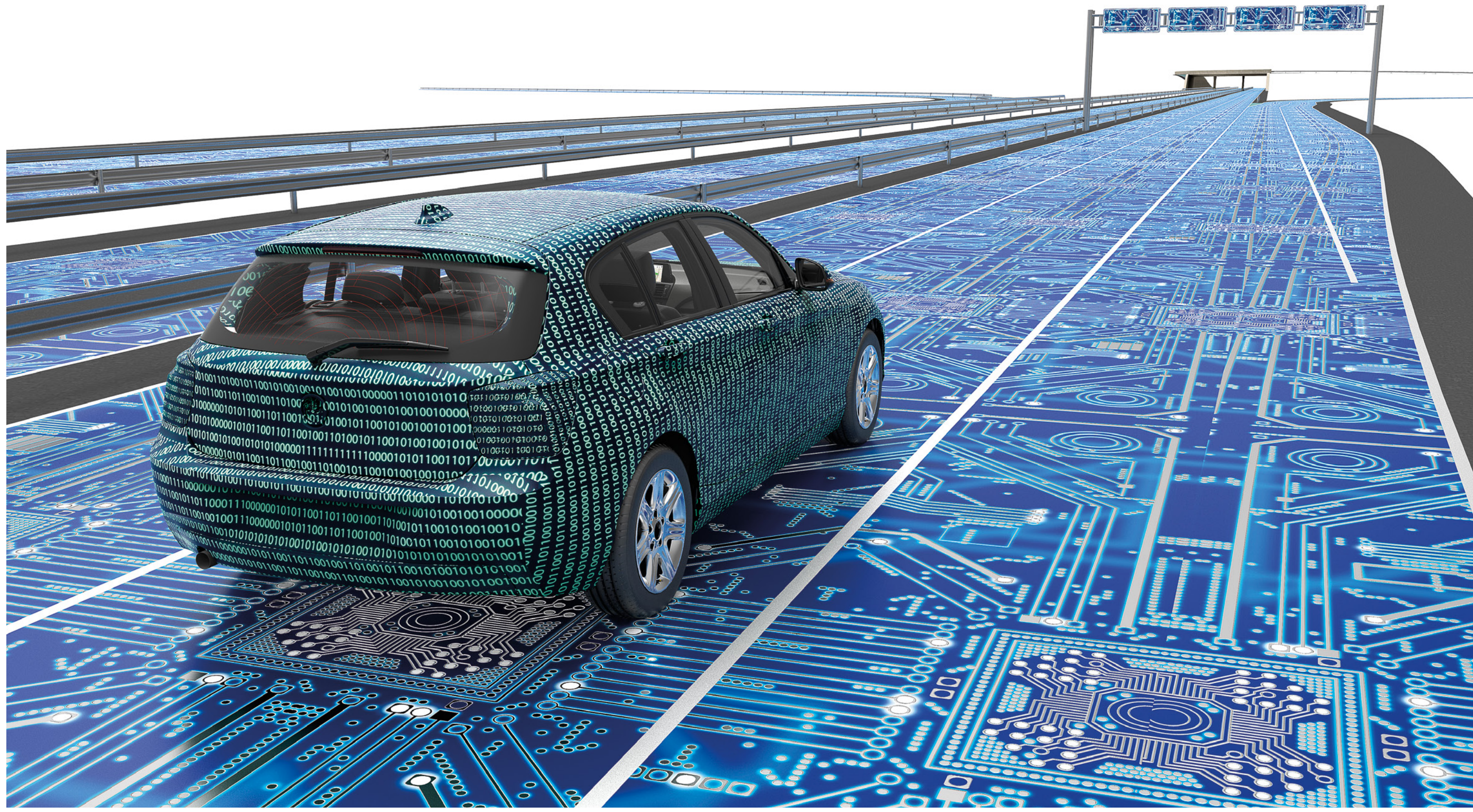
Artificial Intelligence: Promises and Perils

We held our first major public engagement event at York's Festival of Ideas in June 2018. Around 500 people, two thirds of whom were between the ages of 25 and 64, engaged with interactive demonstrations, heard about the latest research, questioned experts during panel discussions, or took part in research into public opinions.

During the exhibition, attendees could talk with a chatbot, test whether they could distinguish between a human and a machine using AI, see robots in action then program their own, and play noughts and crosses against a computer and watch its "thought processes" as it decided what move to make.

Panel discussions during the day included:

- **What is Artificial Intelligence (AI)?:** Professor Alan Winfield and Sir Malcolm Grant introduced audience members to current AI, robotics and autonomous systems and discussed what the future may hold.
- **Driverless Vehicles:** University of York academics were joined by the University of Hertfordshire, Rolls-Royce and Thatcham Research to discuss how we can live safely with autonomous cars, crew-less tankers and parcel delivery by drones.
- **Artificial Intelligence for Health:** Academics and industry engineers debated whether the potential for AI to revolutionise healthcare is overhyped or whether it could transform the NHS.
- **The Future of Work:** Industry and academic experts discussed the implications of increasing the number of autonomous systems in the workforce.



Festival of Ideas 2018

Would you ride in a driverless car?

Colleagues from the Department of Philosophy at the University of York helped the Programme to engage further with some of those who attended the Festival of Ideas AI focus day.

They ran an interactive decision tree activity with around 100 people, trying to establish the rationale underlying people's intuitive reactions to driverless cars, and whether their opinions remained the same following critical reflection.

Initially 65 participants responded that they would ride in a driverless car and 32 said they would not. Their answers led them through different lines of questioning, with two main counter-questions:

- Whether any safety concerns would be offset by a high statistical expectation of safety; and
- Whether greater control of the car, understanding of the car, or 'explainability' of its decisions would substantially change people's views on autonomous driving.

The final question was whether, following the questioning, they would still ride in a driverless car ('yes', 'probably', 'unlikely', 'no', or 'unsure').

There was found to be a greater movement and graduation of opinion amongst those who initially started off with a negative opinion.

This shows that effective public engagement can influence acceptance of autonomous systems. This understanding will influence our future public engagement work.

The Future

We are looking forward to the year ahead. With strong foundations in place, we will be able to focus on key issues in solving the challenges of assuring and regulating robotics and autonomous systems (RAS).

Partnerships and collaboration will be a key strategic enabler in meeting our goals. The foundational research we are carrying out in York tackles some of the central, cross-domain problems, but we need links to others, both to validate these ideas in practice, and to draw in results from other research programmes. 2018 saw the start of five pivotal demonstrator projects, and in 2019 we expect to expand the number of demonstrators and to identify more critical work that needs our support across the globe.

We are looking for mutual support from industrial partners to focus more deeply on narrower, domain-specific challenges in order to accelerate their technology developments and feed into the Body of Knowledge. We want to work with a wider set of regulators and policy makers to make the safe adoption of these technologies a reality. And finally, our research and teaching will become increasingly global, and we will forge deep relationships with academic partners, old and new, near and far.

We now have a structure in place for the Body of Knowledge and during 2019, and beyond, this will be populated with curated, practical information from our research, demonstrator projects, and contributions from the international community.

We have already seen a strong interest in the training and education provision we are developing. 2019 will see the first delivery of these new courses, addressing the needs of the full spectrum of stakeholders from senior leaders to professionals developing and assuring RAS.

Through partnerships, Programme Fellows, funding calls and more, we are looking forward to expanding the international community we are already working with. The challenges we face not only extend across technologies and domains, they span geographical borders too. Working together will help us to ensure that assurance and regulation keeps apace with the leaps being made in technology development. The future is now. The future is assured.

Work with us

The challenge of assuring the safety of robotics and autonomous systems is cross-domain, cross-technology and cross-application. We must work together to solve it. We would love to talk to you about how we can work together.

+44 (0) 1904 325345
assuring-autonomy@york.ac.uk
@AAIP_York
www.linkedin.com/company/assuring-autonomy
https://medium.com/@AssuringAutonomy