

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



## KEY

- Teal pin: Funders
- Orange pin: Demonstrator Projects
- Green pin: Collaborative links
- Pink pin: Programme Fellows

## CONTENTS

Welcome	4
A Year in Numbers	6
International Community	8
Public Engagement	10
Impact	12
Education and Training	18
Guidance	20
Demonstrator Projects	22
The Future	26

# WELCOME



**Collaboration has always been at the heart of our assuring autonomy work. The challenges of assuring the safety of autonomous systems are vast, with new ones emerging as we edge closer to wide deployment. We must work together to ensure their safe introduction and operation.**

**O**ur multidisciplinary approach runs through all of our work and is celebrated by our collaborators. We bring the right people together at the right time. In doing so, we help to bring safety to the forefront of technology development.

In 2022 we grew our international community. New partners have helped us engage with the public, influence upcoming standards, and find new funding to answer some of the emerging challenges. We've also worked with partners to validate and refine our guidance.

We moved into our new home, the Institute for Safe Autonomy at the University of York, and have been delighted to welcome existing and new partners to this fantastic facility. We have also brought additional members of the team on board, including new research associates as well as several postgraduate researchers and interns.

Our work with regulators has progressed, with bespoke training for two regulatory organisations, and more planned for next year. In December we brought together over 40 staff from 11 different regulatory and standards organisations for the first in a series of workshops that will

help us identify common challenges of assuring and regulating AI- and ML-based systems. By identifying these areas of overlap we will enable regulators to make greater progress by working together. This work will be developed at follow-on workshops in 2023.

This year's achievements, and those of previous years, have been part of an evaluation by independent consultants, considering our impact to date. Our community, our expertise, our multidisciplinary approach, and our influence on safety practices and skills are elements we're especially proud of. We've also set the conditions for more change and look forward to seeing this impact in the future.

So, as we move out of our busiest year yet and into the final year of our current Programme, we are exploring new ventures and considering how we best tackle the challenges associated with the safe operation of autonomous systems and the wider societal impacts of their introduction. We've delivered some notable achievements this year and we're excited by what is to come.

**Professor John McDermid OBE FREng**  
Programme Director



Our world is rapidly changing and technological innovation is a big part of this transformation. Assuring the safety of such advances is a core part of Lloyd's Register Foundation's mission to engineer a safer world.

The World Risk Poll 2022 identified widespread concern about the use of artificial intelligence (AI). 65% of the global population reported that they would not feel safe in an autonomous vehicle, with just cause.

If the public is to trust and accept autonomous systems, they need confidence in their safety. Our work with the University of York to fund and support the Assuring Autonomy International Programme (AAIP) provides the most robust and relevant tools and methods for developers, regulators, policymakers, and the public to gain this confidence. By bringing together diverse perspectives and expertise, the team advances best practice and makes it accessible to all, ensuring it can be adopted with immediate impact, whether in robotics, autonomous systems or AI.

We are delighted to be part of AAIP's success so far and look forward to seeing more impact and influence in the future.

**Dr Ruth Boumphrey**  
Chief Executive  
Lloyd's Register Foundation



Autonomous systems have already begun to shape our world: domains including automotive, aviation, healthcare, manufacturing, and maritime are being changed by their introduction. The AAIP's research focuses on changing lives for the better by leading the safe development, introduction, and regulation of these technologies. It advances both curiosity-driven and action-oriented research, and in doing so contributes in key ways to the University of York's mission to be a university for public good.

AAIP brings together a community with wide-ranging experience and backgrounds for a collective aim. This is both a core strength of the programme and also at the heart of research at York.

The AAIP has built partnerships across domains and works closely with stakeholders around the world. Over the last year, this community has come together and produced some remarkable achievements. They have influenced standards, improved safety practices, adopted and adapted AAIP guidance, and increased knowledge and awareness of safety assurance methods.

This collaborative approach is helping to solve the challenges of assuring the safety of autonomous systems and ensuring they are acceptable to the public: something that will benefit us all.

**Professor Matthias Ruth**  
Pro-Vice-Chancellor for Research  
University of York

# A YEAR IN NUMBERS

## Research

13

demonstrator projects



## Education and training

50+

trained in healthcare and maritime

## Engagement

15

regulatory and standards organisations engaged

Guidance

• 1300+ downloads of SACE and AMLAS in 29 countries



## Funding leveraged

£41.5M

## International community



- Iran - 1
- UK - 9
- Germany - 4
- Australia - 1
- UAE - 1
- Belgium - 1
- Brazil - 1





AAIP team and collaborators at our workshop, November 2022

# INTERNATIONAL COMMUNITY

The community we have built is an essential part of the success of the Programme. Truly interdisciplinary collaboration with colleagues from varied backgrounds and experiences ensures that our guidance on safety assurance works for all stakeholders.

**W**e have translated our leading research into accessible guidance on how to assure the safety of autonomous systems.

A diverse range of stakeholders across the globe have accessed it in the last year.

## UK

We were delighted to host a number of workshops and conferences this year, including July's *Future health*:

*how can we assure the safety of AI in healthcare?* in partnership with NHS Digital. With presentations from academics, clinicians, manufacturers, regulators, and safety engineers, a range of perspectives on the safe introduction of AI and autonomous systems in healthcare were presented. The conference also helped to bring delegates up to date with technology developments in healthcare, how the safety of these systems is being assured, and how ethics and governance are being considered.

We brought together around 40 delegates in November for a workshop to identify existing and upcoming challenges to the safety of autonomy. Considering both domain-specific and cross-sector challenges, new projects have been identified that will help us tackle these challenges together.

Working with the Health and Safety Executive and Department for Transport, we held the first in a series of workshops with regulators in December to help identify common challenges of assuring AI/ML-based systems, particularly identifying areas where regulators may be able to make greater progress by working together.

Our public engagement work developed further this year with a new exhibition at the National Railway Museum. We worked with the museum, part of the Science Museum Group, to develop and create a new exhibit to help their visitors to better understand how driverless technologies work and how our research is helping to assure they are safe.

## Europe

Our joint demonstrator project with PAL Robotics is investigating how

we can safely use service robots to support those with mild cognitive and motor impairments to live in their own home. The York team visited Barcelona in November to finalise and evaluate the integration of the end-to-end autonomous robotic assistive-care solution developed by the ALMI project team, using a TIAGo robot within a testbed in the PAL Robotics labs.

Dr Richard Hawkins was part of the DREAMS workshop organising committee in September. The conference was co-located with the 18th European Dependable Computing Conference in Zaragoza, Spain. It was organised as part of the ICON project "Layers of Protection Architecture for Autonomous Systems" (LOPAAS), a collaboration between Fraunhofer IKS, Fraunhofer IESE, and the AAIP.

## Australasia

Our collaboration with the Trusted Autonomous Systems Defence Cooperative Research Centre in Australia has continued. We were delighted to welcome Rachel Horne, their Assurance of Autonomy Activity Lead to the

UK in the summer and Professor John McDermid presented at their Autonomous Vessel Forum in September.

## North America

The team behind the Safe-SCAD demonstrator project published DeepDECS, a new method for assuring the safety of autonomous systems with deep-learning perception components. DeepDECS uses a suite of deep neural networks (DNN) verification methods to quantify the random uncertainty that the use of DNN perception components introduces.

## South America

It was a pleasure to welcome Programme Fellow Dr Genaina Rodrigues to York in October. Genaina has worked on a solution for the online adaptation and evolution of the goals of autonomous systems, to support their decision making in applications affected by disruptions. A joint research paper describing this solution is under review.



Delegates at Future health: how can we assure the safety of AI in healthcare? conference, July 2022





# PUBLIC ENGAGEMENT

Ultimately it is the public who will be impacted by autonomous systems. Unless people accept the systems and are comfortable with them, then the technology will never be successfully deployed.

Our engagement with members of the public to help them understand these systems and understand whether they can trust them is really important. This year we worked with the National Railway Museum, part of the Science Museum Group, to develop and create an exhibit that would help their visitors to better understand how driverless technologies work and how our research is helping to assure they are safe.

## Visual demonstration

Our exhibit was the first to be shown in the Museum's new Innovation Platform and welcomed over 19,000 visitors. The star of the exhibition was an interactive display visually explaining how machine learning is used in hazard detection and categorisation.

Earlier in the year, our team had

Dr Richard Hawkins leads a guided tour of the exhibit as part of York Festival of Ideas 2022

added cameras to the Museum's road train to film the route from York Minster to the museum. The exhibit allowed visitors to watch the footage in its original form and then press a button to add overlays to the film.

The first overlay showed how the YOLO (You Only Look Once) algorithm detects and recognises various objects in the film, for example, people, cars, and buses. Visitors could press another button to see how DeepSORT is used to track the objects identified by giving them an ID and tracking them throughout the film. The final overlay was a hazard detection layer that colour-coded the identified objects with a percentage chance of them being a hazard.

The wider exhibition included cultural references to autonomous systems, historical objects, and

items that are part of the wider context of the introduction of self-driving vehicles. These included a film with AAIP Research Associate, Dr Zoe Porter, considering the ethical questions associated with the introduction of autonomous systems.

## York Festival of Ideas

In June 2022 we held a York Festival of Ideas day at the Museum, offering tours of the exhibition and an opportunity to ask the team questions about the introduction of the technology.

*"I definitely learned about how past experiences and now are really related."*

*"I think that really provokes your thinking, with that [Stephenson's Rocket] at the side of the actual exhibition, to think about the way that this is all going... and that actually, we've had these problems in the past and we've had these ethical dilemmas in the past."*

*"I learned a little bit about some of the percentages and the way it picks things up and the hazard detection was really good."*

## Thoughts on the future

We also asked visitors their thoughts on the introduction of autonomous systems and found that they were generally very positive.

*"I think it's actually a very exciting development and I'm actually looking forward to it."*

*"I do think it's inevitable we're moving in that direction, I think fairly quickly to be honest."*

*"I think it's exciting. It's a new thing that could help with different ways of the streets working and how it all comes together."*



A film with Dr Zoe Porter was part of the National Railway Museum exhibition

# IMPACT

In October 2016, Lloyd's Register Foundation (LRF) published its 'Foresight review of robotics and autonomous systems (RAS)'. The report identified that the biggest obstacle to gaining the benefits of RAS was their assurance and regulation. The AAIP was established to tackle these challenges.

Open to find out what others say about AAIP



“ York influenced the topics of the nodes within the [Trusted Autonomous Systems] programme leading to an open competition as to who got funding. [Members of the AAIP team] were involved in conversations with EPSRC during the development of the TAS business case and this involved the topics of the nodes. These conversations influenced the strategy and priorities of the programme.”

**Dr Helen Niblock, Head of Regional Engagement (NE, Yorkshire and Humber) – Engineering and Physical Sciences Research Council (EPSRC)**

“ It’s such an incredibly complex area. Industrial partners and regulators need confidence, which is really the big issue to overcome if cobots are to be deployed. Whilst Sheffield has expertise in robotics and manufacturing, we needed York’s expertise in safety assurance to tackle this challenge.”

**James Law, Senior Innovation Fellow – University of Sheffield**

“ We worked with a philosopher, Zoe Porter, to co-write a paper that set out the ethical dilemmas in a qualified, expert manner. It created a bridge between the technical world and the moral and ethical one. Engineers cannot be blind to these topics, but they can’t solve the ethical dilemmas. We hit a glass ceiling, from an engineering perspective, trying to answer the question ‘how safe is safe enough?’, unless we take a wider interdisciplinary perspective.”

**Professor Dr. Simon Burton, Research Division Director – Fraunhofer IKS**

“ Without the AAIP, being in the network and the visibility that created would have been missed. The Programme has been great at facilitating our papers, and enabling opportunities to present at conferences.”

**Dr Lars Kunze, Departmental Lecturer in Robotics – University of Oxford**

“ [The impact of the demonstrator project is] much bigger than the adaptation of a tool for early recognition of cardiac arrest. It has given us the keys to unlock the potential for future AI.”

**Dr Nigel Rees, Head of Research and Innovation – Welsh Ambulance Services NHS Trust (WAST)**

“ The AAIP team at York are doing world leading research and are at the forefront of the field that they’re working in, and that we’re also seeking to work in. We don’t have any projects specifically looking at how to actually do assurance. For example, what should those assurance methodologies look like? And certainly not to the level of detail that the AAIP have produced. For us, it’s a no brainer to collaborate with such an organisation, especially one that’s also funded for the public good.”

**Rachel Horne, Assurance of Autonomy Activity Lead – Trusted Autonomous Systems Defence Cooperative Research Centre**

“ I was glad to be part of the AMLAS review team. It really helped me to gain a better and broader understanding as AAIP has strong cross-domain knowledge, not just in the automotive sector, but also learning, e.g. from medical devices.”

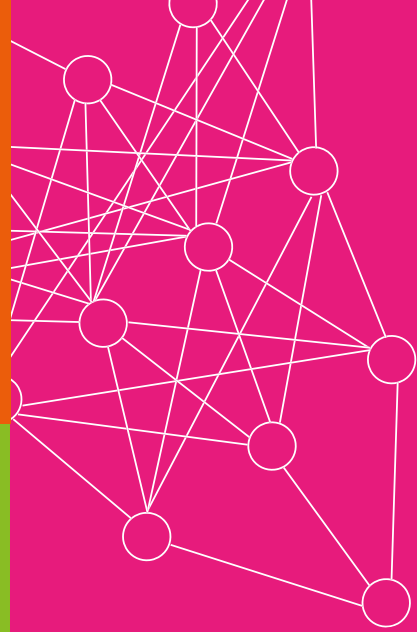
**Lydia Gauerhof, Research Engineer – Corporate Research at Robert Bosch GmbH**

“ On a personal level, being involved in the work of the AAIP has helped me build links, improve my competence to understand a complex robot system using machine learning that I had not been exposed to before, and appreciate the different interactions involved and what can be ringfenced. ...The impacts for the wider organisation are through my work with policy colleagues helping develop policy around AI, but also through the cross-sector discussions I’m involved with. AI is such a cross-cutting technology that we get enquiries from manufacturing, transport, construction, health and other departments.”

**Nicholas Hall, HM Principal Specialist Inspector (Advanced Automation and Cyber Security) – Health & Safety Executive**

“ Consequently, we published a review paper with a recommendation that AMLAS was fit for purpose as a safety assurance methodology when applied to healthcare ML technologies, although development of healthcare specific supplementary guidance would benefit those implementing the methodology. That supplementary guidance has since been created and will be published in late 2022.”

**Shakir Laher, AI/ML Research Associate /Safety Engineer – NHS Digital**



“ The course was fully subscribed, and participants gained a greater understanding of how to effectively assure AI for use in a care pathway. The delegates were ‘close to the use of the technology or product’ and would be able to influence and impact others around them...Using AMLAS in the training helped learners appreciate the need to be systematic in their approach to assurance.”

**Sean White, Safety Engineering Manager – NHS Digital**

“ The Programme brought us together. It is such a huge topic, with so many actors involved and so much literature being published. With a wider net to throw we can get a good view of what’s going on. As a Programme Fellow, I can learn from what’s going on in UK industries.”

**Professor Dr. Simon Burton, Research Division Director – Fraunhofer IKS**

“ Without the AAIP engagement I would probably have arrived at a similar position in my thinking, however, I think the AAIP has accelerated my development along this path, and it has been achieved in spite of competing demands. The AAIP has got scale, a range of different demonstrators across sectors and cross-learning is really beneficial to regulators so accessing the AAIP’s body of knowledge is valuable too.”

**Nicholas Hall, HM Principal Specialist Inspector (Advanced Automation and Cyber Security) – Health & Safety Executive**

“ AMLAS is really good, well-written and specific. In RAILS we will consider what happens if there are changes in the system once it’s been deployed.”

**Dr Lars Kunze, Departmental Lecturer in Robotics – University of Oxford**

“ Collaboration is essential. That doesn’t mean we have to work in the same way though, in fact, a vital role of the AAIP has been to create a space where people with different opinions have been able to gather together and discuss their approaches. This mitigates against silos and group think which is where the mistakes happen.”

**Professor Tom Lawton, Critical Care Consultant – Bradford Teaching Hospitals NHS Foundation Trust**

“ ... the approach that AAIP have taken, and the expertise that the York team specifically bring to the space is quite unique. The other work I am aware of is largely being done by experts in robotics, and experts in robotics for manufacturing. They understand the regulations and how to build robot processes. What they don’t have is the expertise in safety assurance frameworks, which is completely different. York’s team are the international experts in safety assurance, with a very different skillset, methods and approaches.”

**James Law, Senior Innovation Fellow – University of Sheffield**

“ The demonstrator project enabled us to produce a report highlighting regulatory and legal challenges that need to be dealt with to enable the operation of autonomous and remotely controlled ships in UK waters. This report has been well received and been cited by the Maritime and Coastguard Agency and also by various academic works...Working with the AAIP not only provided a suitable platform for us to present our work but also enabled us to work closely with experts from other disciplines thus appreciating different aspects of the debate better.”

**Professor Baris Soyer, Professor of Commercial and Maritime Law – Swansea University**

“ [Working with AAIP has] ...consolidated the information, and misperceptions we had. It’s provided a clearer path for what we need to do for the next few years and consolidated our thinking towards tangible outputs.”

**Shakir Laher, AI/ML Research Associate /Safety Engineer – NHS Digital**

“ York’s work on safety and policy have made a massive contribution to the wider landscape. The Body of Knowledge has been a big piece of work which has been very useful to academics and professionals as well.”

**Dr Helen Niblock, Head of Regional Engagement (NE, Yorkshire and Humber) – EPSRC**

“ The AAIP has helped us to work with specialists from different organisations...This brought into focus other aspects of assurance, peculiar to the use of AI, that we may have overlooked.”

**Sean White, Safety Engineering Manager – NHS Digital**

“ My fellowship collaboration helped me understand the problems better and discuss them with the field experts to find appropriate solutions. Further, I could expand my network by knowing and connecting to outstanding people in self-adaptive systems and safe autonomy. I strongly recommend this kind of fellowship to young researchers and industry experts.”

**Mehran Alidoost Nia, Researcher – University of Tehran**

“ As a standards writer you draw on lots of influences, so the AAIP experience may contribute to work I’m involved with around the new Industrial Robots safety standards ISO10218 parts 1 and 2 and the new Autonomous Mobile Robots standard which I’ll be working on from next year.”

**Nicholas Hall, HM Principal Specialist Inspector (Advanced Automation and Cyber Security) – Health & Safety Executive**

“ The York team had a good understanding of machine learning components. It was very useful to have this sanity check to ensure we had the same understanding; then we could extend our respective knowledge. We could also talk about similar problems we’d encountered in different domains. We could do this safely, and without sharing commercial information.”

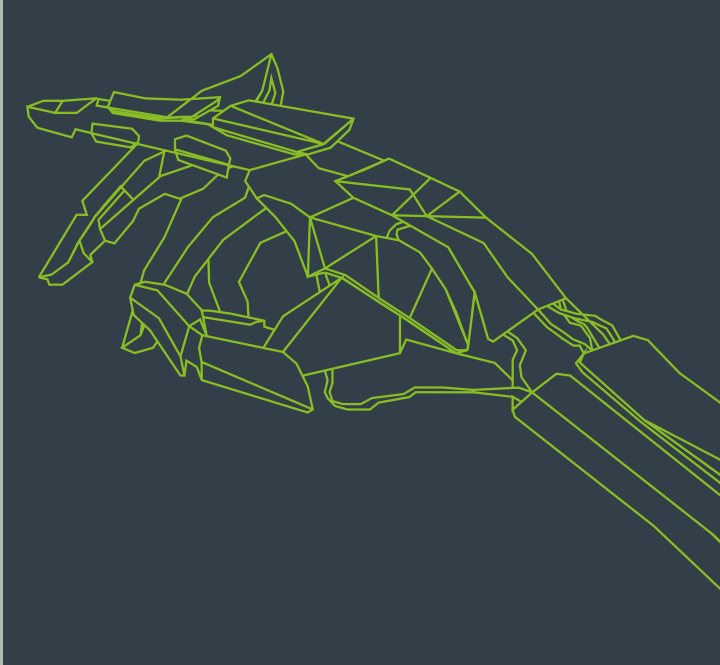
**Lydia Gauerhof, Research Engineer – Corporate Research at Robert Bosch GmbH**

“ The biggest benefit of working with the AAIP team is the collaboration across disciplines. Working with experts with diverse skills and experiences has led me to new ways of thinking. The research we’ve worked on together has benefited hugely from this diversity. I’m using this knowledge to co-design and co-evaluate functional, social, legal, ethical, empathetic and cultural properties of AI systems.”

**Xingyu Zhao, Lecturer in AI – University of Liverpool**

“ There is now a huge network of industry and academics – it would be great to maintain this community long term. The AAIP approach of projects and communities brings people together.”

**Dr Lars Kunze, Departmental Lecturer in Robotics – University of Oxford**



“ Working with the AAIP team resulted in novel methodologies for the safety assurance of shared control in autonomous driving. It also enabled me to develop my own knowledge of the state-of-the-art in this area. I’m now taking this forward in a new project that aims to developing principled approaches and tools for assuring and demonstrating accountability of safety-critical autonomous systems with respect to laws and regulations.”

**Dr Lu Feng, Assistant Professor – University of Virginia**

“ The British Standard 30440 aims to define a validation framework for the use of AI within healthcare. Through working with the AAIP we have ensured that human factors and ergonomics have been considered and appropriate clauses out in place. Specific references to usability, consistency, training, explanation, control as well as automation bias and over-reliance were deemed to be critical components of the standard in ensuring safe and effective AI models to be developed and deployed.”

**Haider Husain, COO – Healthinnova Limited & Panel Chair for BS 30440**

“ Where the credit is due to the AAIP is that they tell you ‘how to do’ not just ‘what you need to do.’ The expertise in York, the level of detail and depth is incomparable to anywhere else. If the AAIP wasn’t there you would immediately miss the intelligence – the high level interactions and intellectual conversations that can help you with your chain of thought and put you straight.”

**Shakir Laher, AI/ML Research Associate / Safety Engineer – NHS Digital**

“ The Programme can then disrupt those staid structures [in industry]. It’s something that the AAIP is really well placed to do because it’s got this broad view and looks at the concept of autonomy itself, rather than one particular industry, and the problems associated with autonomy rather than road regulations for example.”

**Professor Dr. Simon Burton, Research Division Director – Fraunhofer IKS**

“ This connection with AAIP has clearly resulted in more funding and more collaborative work. It is now truly collaborative.”

**Dr Lars Kunze, Departmental Lecturer in Robotics – University of Oxford**

“ AAIP put me in contact with other researchers focused on autonomy, which had complementary perspectives to mine (legal, policy, etc.) that I both enjoyed hearing about, and which gave me a better understanding of various stakeholders I meet in my own work. I think this has helped during my work on other research proposals, making it easier to integrate other organizations in research ‘narratives’.”

**Dr Fredrik Asplund, Assistant Professor – KTH Royal Institute of Technology**

“ AAIP provided us enormous opportunities to work colleagues who share common research interests in robotics, AI, safety and autonomy. With the support from AAIP, the research team at UCL MechEng can further expand its expertise in marine engineering and further push the boundary of marine autonomy.”

**Dr Yuanchang Liu, Lecturer – UCL**

“ York has the pedigree and is seen as having the right level of expertise, so the trust is there that in York there was an organisation to address these issues. The second aspect is the execution. Reputation alone doesn’t make a successful project...The other thing that worked really well was the concept of the Programme Fellows, because it allowed people like me to come from industry and feel somehow intimately involved as part of the team. Basically, they’ve turned the usual model on its head. For me, I’d normally be paying the university to come and do work for me. It was a completely different type of approach and it’s been successful.”

**Professor Dr. Simon Burton, Research Division Director – Fraunhofer IKS**

“ The researchers from York ...brought a huge amount expertise in the safety part of autonomy, and methods for verification and testing.”

**Dr Lars Kunze, Departmental Lecturer in Robotics – University of Oxford**

“ The SUCCESS [demonstrator project] team benefitted immensely from the technical interactions with the world-leading safety researchers at the University of York as well as with researchers from other demonstrator projects. The SUCCESS project resulted in increased strategic collaborations with Volvo CE, initiation of three follow-up projects and made contributions in extending the state of the art through eight publications and a doctoral thesis related to it.”

**Professor Sasikumar Punnekkat – Professor – Mälardalen University, Sweden**

“ Safety has been improved at WAST and we are due to present our findings to the Board shortly.”

**Dr Nigel Rees, Head of Research and Innovation – Welsh Ambulance Services NHS Trust (WAST)**

“ AMLAS is important as it provides the know how to develop a safety argument. The structure really works; it is very clear for people who come from an AI rather than a safety background. There will be special details that are outside of the guidance which are domain specific. We can use the guidance as a sanity check, asking for example, ‘did AMLAS use the same argument?’”

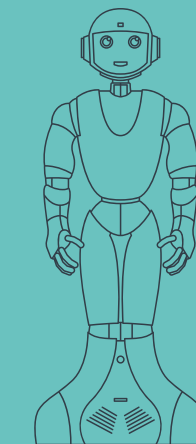
**Lydia Gauerhof, Research Engineer – Corporate Research at Robert Bosch GmbH**

“ The CPD we’ve developed together has added value to our national portfolio, broadened the scope and depth of our training content, ensured a holistic approach (not purely the technology perspective), incorporated human factors considerations, and furthered us towards a position where more people can deploy products safely that meet healthcare needs.”

**Sean White, Safety Engineering Manager – NHS Digital**

“ As a result of this work, we are getting closer to the position where some of these technologies will be seen in use. Hopefully within the next two years they will be tested in the real world, and some of the approaches being espoused by the AAIP come in earlier than that... The AAIP provides a link to everyone and everything you might need – the AI person, the safety engineer, the lawyer, the philosopher – all the right expertise at the right moments.”

**Professor Tom Lawton, Critical Care Consultant – Bradford Teaching Hospitals NHS Foundation Trust**







This year we worked with independent evaluators to review our impact to date. We're delighted with what they found. Over the last five years there have been some remarkable advances in the development of autonomous systems. While safety assurance methodologies and standards have not quite kept pace, AAIP has impacted and influenced through our guidance, research, training and education, and international community, and has created conditions for more impact in the future.

The review identified us as the experts in the safety assurance of autonomous systems. We have impacted companies, people, policymakers, the public, researchers, regulators, standards, and the wider landscape.

#### Companies

We work with small and large companies helping to increase knowledge of safety assurance and influence safety practices. In Bosch, for example, we have run workshops on machine learning safety and disseminated our AMLAS guidance. With Luffy AI we are supporting their work to adopt AMLAS into the development of their autonomous air vehicles.

#### People

Our work is improving awareness, understanding, and knowledge of safety assurance and making it accessible to all. We do this through our training and education programme, demonstrator projects, research collaborations, and the

community we have established across the globe. This multidisciplinary community benefits every collaboration through its broad range of skills and experience.

#### Policy makers

Ongoing work with organisations including the UK's Department for Transport enable us to influence new policies around the introduction of autonomous systems. A new partnership with the Centre for Science and Policy at the University of Cambridge is accelerating this work by organising a series of targeted briefings and fellowships for staff within the civil service.

#### Public

Through lectures, events, and exhibitions we engage the public to help them understand the technology behind autonomous systems and our work to assure their safety.

#### Regulators

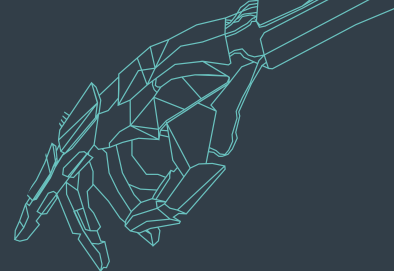
In the UK, healthcare is at the forefront of the deployment of autonomous and advisory systems and we have made great progress through our long-term partnership with NHS Digital. We have supported the training of the workforce and the adoption of AMLAS through the development of supporting guidance.

#### Researchers

We collaborate with researchers across the globe on projects in numerous domains, particularly through Programme Fellowships and demonstrator projects. Through these collaborations we adopt a practical approach to developing structured guidance, methods, and processes for assuring the safe deployment of autonomous systems.

#### Standards

The BSI standard BS 30440 Validation framework for the use of AI within healthcare directly references the AMLAS guidance and is one example



of how our work is starting to influence new standards. Our impact report also highlighted how working with us has influenced a standards developer within the Health and Safety Executive in the UK.

#### Wider landscape

Since 2018 the focus on the safety and trustworthiness of AI and autonomous systems has grown. Funding from councils, industry, and grantmakers has increased. New networks and programmes have started. Facilities such as the Institute for Safe Autonomy at the University of York have opened. We have leveraged funding, developed partnerships, and influenced funding decisions that have supported this growth, and through our community have enabled more.

#### The future

We are delighted with our progress so far. We have learnt, refined, validated, and now started to embed our expertise across domains and organisations. In doing so we have created the right conditions to make long-term changes.

There is more to do and we are looking to the future. We will build on the international recognition of our expertise to bring safety even further to the forefront. Our training and education offer will expand to reach more people and give them the confidence to assure the safety of the autonomous systems in their organisation. We will continue to work collaboratively across disciplines – bringing together the right people at the right time.

Through this work we will continue to make a significant difference to the safety of autonomous systems.

Read more about our impact:



# EDUCATION AND TRAINING

Alongside our MSc module in the safety of autonomous systems, we have focused on two safety-critical domains this year with our bespoke CPD – healthcare and maritime.

#### Academic education

In April this year, we ran our Advanced Topics in Safety module for the second time. We welcomed 12 students from a range of domains, broadening their systems engineering knowledge with an insight into the challenges to safety processes and product safety brought about by the introduction of autonomous systems.

For the first time, the course was run in person in York, with self-study material giving students an overview of the current state-of-the-art in system safety to ground later discussions. The course then covered the safety and security assurance challenges posed by robotics and autonomous systems (RAS), software and machine learning (ML), safety

assurance in complex environments, assurance of ML, human factors, and risk acceptance.

“Completing the Advanced Topics in Safety CPD module and learning in detail about SACE/AMLAS has proved incredibly useful in developing systems that aren't completely managed by more conventional system safety approaches. Working in a modern R&D environment I've found the content of the course essential, especially for identifying the limits of more classical safety approaches, as well as taking view of possible future safety challenges, their causes and possible ways to mitigate them.”

Phillip Mulvana, Lead Technologist – Safety Engineering

#### A focus on maritime

A key focus this year has been working with regulators. As part of this we ran a bespoke training course for the Maritime and Coastguard Agency (MCA).

This began with a blended synchronous and asynchronous course to bring MCA staff up to



The panel discussion at Future health – how can we assure the safety of AI in healthcare, July 2022

speed in current industrial system safety engineering practice in domains that employ highly complex computer based systems, especially control systems. This increased the understanding of the gaps between these practices and current practice in the maritime domain, supporting improvement of regulation and oversight as more safety-critical technology is introduced into the domain.

The learning outcomes of the course were to ensure that participants could:

- identify safety risks and challenges from a wide range of sources associated with complex systems and services
- use consistent and clear terminology in communications about safety engineering and safety management issues as well as translate terms across vocabularies used in different industrial domains
- recognise key classes of safety analyses techniques, classify individual techniques into their respective class and select the class of techniques applicable for given systems

- appreciate the importance of system safety assurance processes and documentation as a necessary part of maritime safety-critical systems development and operation

Weeks one and three of the course consisted of a set of self-learning activities via recorded presentations, directed reading, and optional exercises. The second week of the course was structured as an intensive set of online facilitated plenaries and group case studies where attendees could interact with each other and University of York staff to explore the concepts and practices behind system safety engineering.

This was followed by a two-day in-person workshop for a selection of staff who were addressing the assurance of complex software based systems and would be asked to address the emerging autonomous capability of ships and the use of ML systems.

This workshop was split into four parts, addressing software, data, autonomy, and machine learning. Each topic had a discussion on a maritime example associated with

it and discussions around the future impact on regulations, guidance and legislation in the maritime domain.

#### Training the healthcare workforce

Our partnership with NHS Digital continued this year. In the summer we ran a joint conference at the University of York, *Future health: how can we assure the safety of AI in healthcare?* This event presented different perspectives on the safe introduction of autonomous systems, including academics, clinicians, human factors experts, patients, and safety engineers.

We also ran our bespoke CPD for 25 delegates, including clinicians, clinical safety officers, and developers. The course gives delegates a greater understanding of how to effectively assure AI for use in a care pathway.

We can develop a bespoke training course for your organisation:



“The CPD we've developed together has added value to our national portfolio, broadened the scope and depth of our training content, ensured a holistic approach (not purely the technology perspective), incorporated human factors considerations and furthered us towards a position where more people can deploy products safely that meet healthcare needs. Historically, our training has focused on safety management processes, the AAIP-supported CPD we've developed has enabled us to extend the scope and explain this in the context of AI technology.”

Sean White, Safety Engineering Manager – NHS Digital



# GUIDANCE

Our guidance enables us to support safer industrial practices across different domains. This year we have developed and validated more of the guidance that is translated from our peer-reviewed research and launched [assuringautonomy.com](https://assuringautonomy.com) to make this guidance accessible to all.

## Complex environments

An autonomous system, such as a self-driving car or clinical support tool, does not operate in isolation: it is part of a complex environment. An autonomous vehicle will interact with other cars, traffic lights, and street infrastructure, as well as with humans. An AI tool in healthcare will become part of a complex healthcare pathway that could involve numerous

clinical staff and care options. Accordingly, assuring the system's safety cannot be done in isolation.

We published our guidance on the Safety Assurance of autonomous systems in Complex Environments (SACE). It is the first methodology that takes the autonomous system and its environment and defines a safety process that leads to the creation of a safety case for the system.

“The autonomous system, its environment, and the interactions that take place between actors in the environment must all be part of the assurance process, in order to demonstrate that the system is safe. Our SACE guidance enables engineers and assessors to take a holistic view of the system within its environment. They can follow the activities in the eight stages that make up the SACE process. This will lead to the creation of artefacts that can be combined to create a compelling safety case for the system.”

**Dr Richard Hawkins, Senior Research Fellow, AAIP**

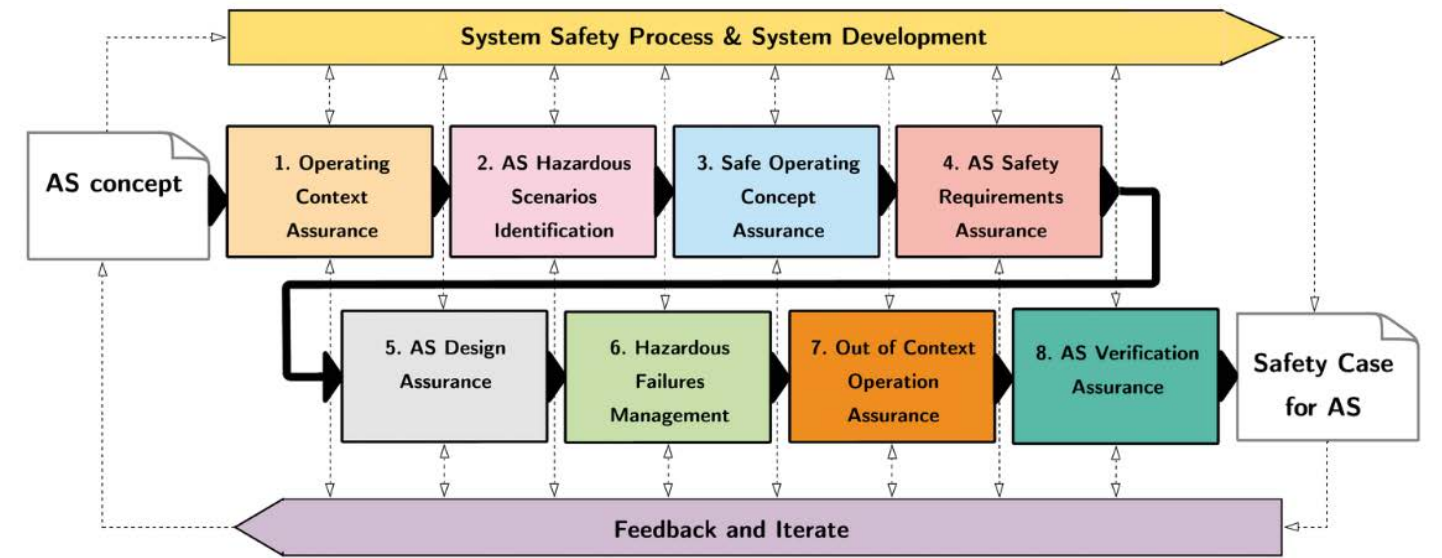
## Understanding

Our research in the Safety Assurance of Understanding in autonomous Systems (SAUS) has allowed us to analyse and identify the principal failure mechanism of the understanding components of autonomous systems. We have developed an adapted HAZOP (HAZard and OPERability study) to support their evaluation. A parallel programme of work, in collaboration with Fraunhofer IKS, is making good progress towards evaluating and making recommendations for practical sensor systems.

## Decision-making

New methods supporting a wide range of autonomous decision-making aspects have been delivered by the Safety Assurance of Decision making in Autonomous systems (SADA) research team. These include methods for:

- reducing epistemic uncertainty
- mitigating the risks introduced by the classification errors of deep neural networks and real-time object detection
- adaptive planning for assistive-care robots
- safety controller synthesis for collaborative robots



Overview of the SACE methodology

Multidisciplinary research has begun into how humans make decisions, to identify attributes of human decision-making that could transfer to the design and safety assurance of autonomous system decision-making.

## Societal acceptability

The SOCIetal Acceptability of autonomous systems (SOCA) research team has been developing a methodology and worked examples for a principles-based ethical assurance argument pattern for AI and autonomous systems to inform SOCA guidance. Significant further funding from EPSRC has been secured to support this work, with a focus on establishing moral responsibility and legal liability for the outputs and impact of autonomous systems.

## Machine learning

Our Assurance of Machine Learning for use in Autonomous Systems (AMLAS) guidance was the first of

our methodologies to be published, in 2021. It has now been validated through use cases in healthcare and space. As part of a Trustworthy Autonomous Systems Hub-funded project, the use of AMLAS is now being considered for safety assurance during deployment of systems using ML.

## Practical and accessible

The guidance we develop is based on our leading research into the safety of autonomous systems. It reflects the collaborative, evidence-based approach we take. Our guidance is accessible to all and has been downloaded over 1,300 times by colleagues in 29 countries and 24 different domains.

Use and download our expert guidance:



“As a result of working with AAIP and using AMLAS, we have been able to understand the process of deploying and assuring the use of machine learning components in small satellite missions. We have applied AMLAS to a simulated wildfire monitoring mission in the first instance and are now leveraging it in ongoing work on other autonomous space technologies.”

**Murray Ireland, Head of Autonomous Systems – Craft Prospect**



# DEMONSTRATOR PROJECTS

We have committed more than £5.5M to demonstrator projects since 2018. These projects have helped us to better understand the challenges of assuring the safety of autonomous systems and to develop, validate, and implement guidance on how to assure their safety.

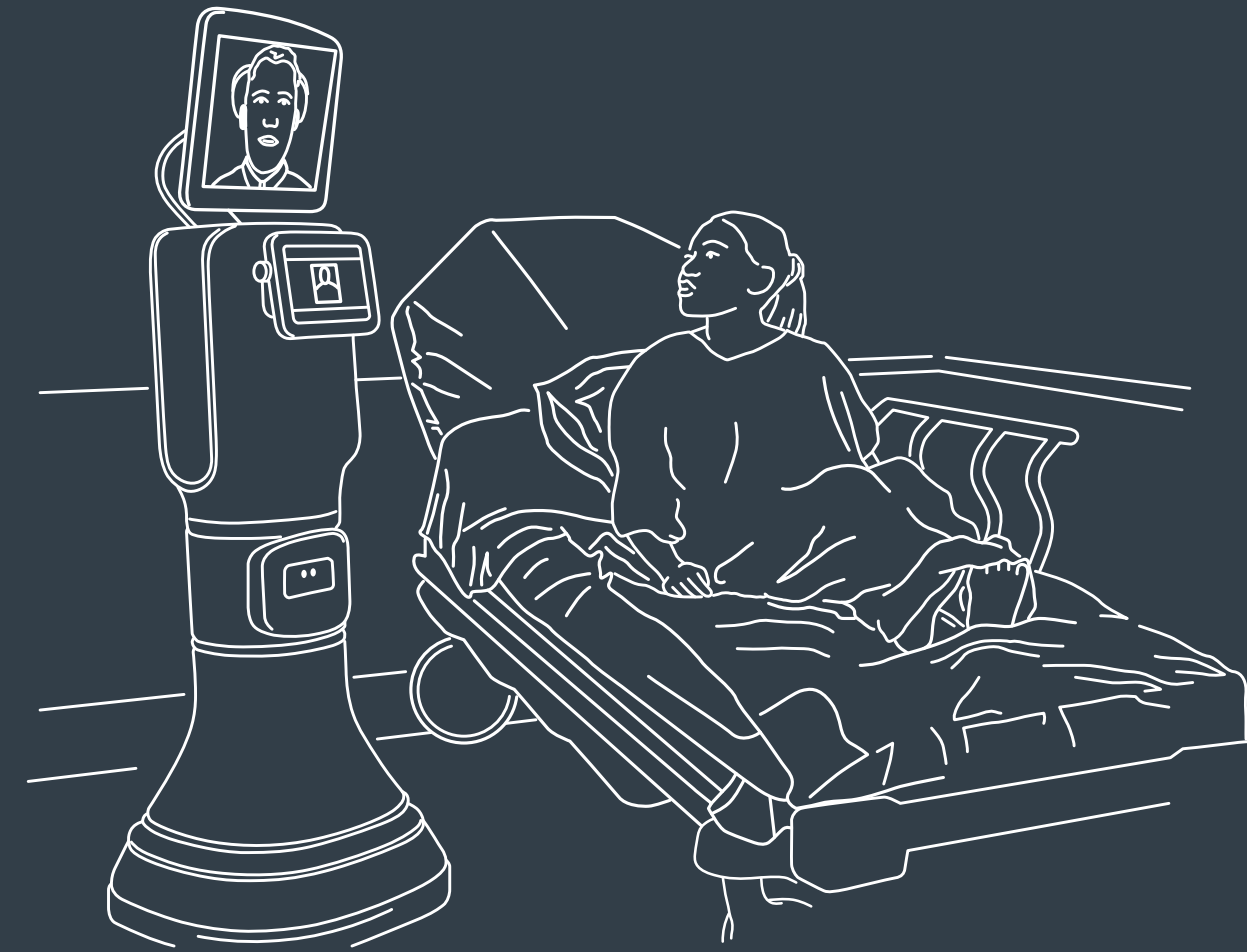
Over the past year, we have focused on projects that have enabled us to further learn, refine, and validate our leading approach to the assurance of autonomous systems. We are working with teams in numerous safety-critical domains to adapt and extend our guidance to ensure its suitability for users in that sector.



## Safety Assurance Framework for machine learning in the healthcare domain (SAFR)

### Using AMLAS to develop a safety assurance framework for the use of machine learning in healthcare

Machine learning (ML) technologies are being developed and deployed into health and social care settings to support a range of clinical areas,



including imaging and decision support. These technologies offer great benefits in terms of supporting and augmenting clinical management and decision-making tasks. Their potential impact on patient safety, however, must be managed in line with the requirements of prevailing regulations.

The SAFR project is developing a framework to support digital health technology manufacturers and organisations using their technology to assure their ML-based healthcare technology and meet their regulatory requirements.

The work is underpinned by AAIP's AMLAS guidance, helping to validate it for use in the healthcare domain and establish whether additional healthcare-specific guidance is needed.

This year, the project team reviewed AMLAS through a series of workshops with two digital healthcare manufacturers. This consultation helped the team to understand if AMLAS converges or diverges from their current safety assurance practices, whether there are gaps or limitations in the structure of the guidance, and if it is fit for purpose when applied to the healthcare domain.

Their review paper was published in September 2022, with a recommendation that AMLAS is fit for purpose as a safety assurance methodology when applied to healthcare ML technologies. The team recommended that healthcare-specific supplementary guidance on using AMLAS would be beneficial. This guidance extends AMLAS so that users and adopters of the

technology (e.g. clinicians) can assure the safety of the technologies in their clinical pathways. It will be published in early 2023.

The team's work has also influenced a new standard being issued by the British Standards Institution in January 2023 that references AMLAS: BS 30440 Validation framework for the use of AI within healthcare.

#### Project partners

- NHS Digital
- British Standards Institution
- Human Factors Everywhere

Read the AMLAS review paper:







## Safety of the AI Clinician

### Can we develop a safer decision support system using AMLAS?

In 2018, the AI Clinician was published by Imperial College London. This was the first decision support system (DSS) that used a reinforcement learning (RL) agent to suggest doses of intravenous fluids and vasopressors to treat sepsis. With reward as the key factor driving the AI Clinician's behaviour, it was trained by rewarding it every time the patient had a positive outcome.

The next step was to assess and refine the safety of the system. Assessing the safety of DSS is a complex and challenging task that requires collaboration between clinicians, technical experts, and safety and human factors experts. This demonstrator project enabled this multidisciplinary interaction to advance the safety assurance of the AI Clinician system.

When applying AMLAS to the AI Clinician, the team identified the fact that there were no hard encoded rules in the AI agent that would prevent it from taking obviously dangerous decisions as a potential safety hazard. Avoiding such behaviours does not make a system 100% safe in every situation possible, but it helps improve its overall safety profile.

Consequently, the team identified a set of scenarios and treatment options that would be unquestionably dangerous for the patient. In particular, they focused on underdosing and overdosing fluids or vasopressors to the extreme. The AI agent was given additional penalties if it suggested any of these treatment options and the model was retrained and its new safety profile assessed.

The updated agent demonstrated fewer unsafe recommendations than both human clinicians and the original AI while maintaining great performance in terms of expected patient survival.

#### Project partners

- Imperial College London
- University of York
- NHS Digital

## Wizard of Oz prototyping for automated decision-making tools in air traffic control (WIZARD)

### How can we assure the safety of automated decision-making tools in the complex air traffic control environment?

Air traffic control (ATC) is a complex socio-technical environment. Introducing automated decision-making could improve the efficiency and effectiveness of the sector. But existing methods to assure the safety of such tools require simulation in dedicated physical environments with skilled personnel, specialised facilities, and equipment that are resource and time intensive to set up and deliver.

This project is researching "Wizard of Oz" prototyping, validation, and testing, to see if they can be used to develop and test a new automated decision-making prototype tool. The project started during 2022 and is focused on the use case of an automated to-do list tool on the controller's

workstation for non-complex tasks. Through this work, the project will provide a use-case and validation for the first three steps of our guidance on the Safety Assurance of autonomous systems in Complex Environments (SACE), which will be used to evolve the methodology.

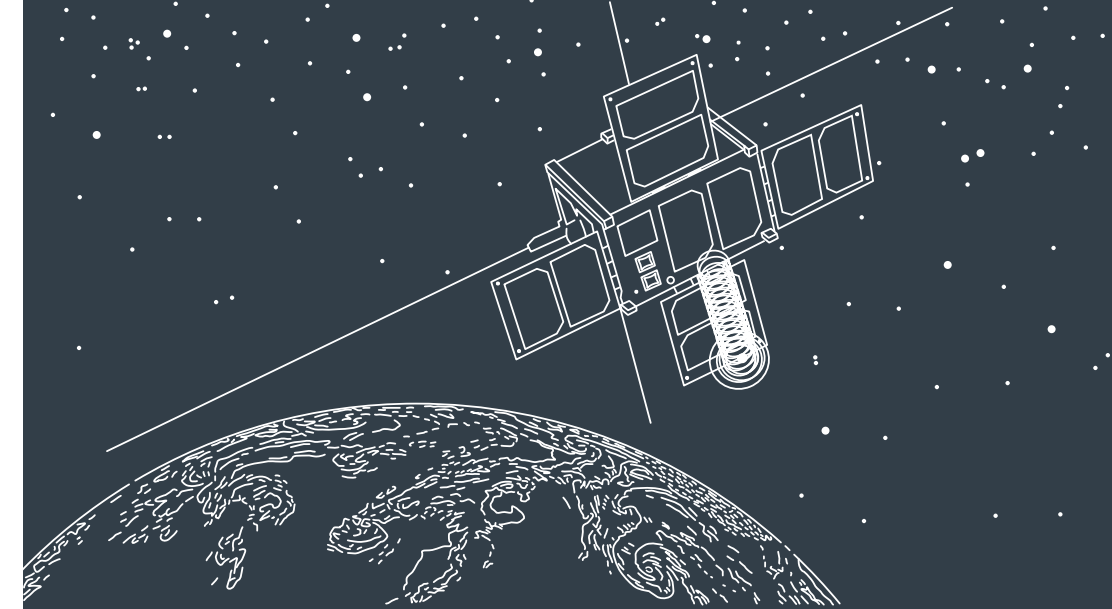
The project team have interviewed air traffic controllers (ATCOs) to understand their current working practices. The findings have been translated into user requirements that are informing the development of an interactive prototype of the automated decision-making tool.

To support this development, the team held workshops to explore the types of scenarios a user would encounter with such a tool, and how they might respond. This helped them gain an understanding of what the tool should look like and what it could be capable of doing.

In the next phase of the project, the team will gather feedback on how ATCOs may respond and interact with such a tool, using paper and digital prototypes.

#### Project partners

- NATS
- University of York



## Autonomous Capabilities and Trusted Intelligent Operations in Space (ACTIONS)

### Assuring the safety of a machine learnt component onboard a satellite used in fire detection

The introduction of onboard autonomy to satellite missions is an increasing area of interest. One of the key concerns of these missions is the level of trust that can be placed in ML components versus ground-based human operators. This project, which concluded in 2022, focused on a demonstration scenario of satellites used for active fire detection carried out autonomously by an onboard ML component.

The application needed to generate and send a fire detection alert to emergency response services on the ground, with confidence that the data generated was accurate, truthful, and timely. Data products were also created for downstream commercial applications, supporting the recovery of areas affected by wildfires.

The demonstrator included elements of model-based systems engineering, onboard flight software, simulation, and hardware-in-the-loop testing for specific ML algorithms and other components. Assuring the safety of the ML component was done using the AAIP's AMLAS process.

The six steps of AMLAS were followed, each generating part of the safety argument for the ML component. Each fragment was connected together to provide the complete safety argument and evidence for the ML safety case. This ML safety case was then integrated

as part of the overall safety case for the wildfire alert system. It is thought that this work represents the first fully developed safety case for an ML component containing explicit argument and evidence as to the safety of the ML.

The team found that implementation of the AMLAS framework led to a higher quality end product, though note that its rigorous implementation also bears a financial cost. Consequently, levels of assurance of commercial applications that are not safety-critical can be considered as a trade-off between rigour and cost and will vary from application to application. The assurance artefacts generated when following AMLAS were found valuable for communication with customers and partners to help build trust in the ML component.

#### Project partners

- Craft Prospect
- University of York
- Global Surface Intelligence

Read the first fully developed safety case for an ML component:





# THE FUTURE

**We are making an impact in a fast-paced field and have established ourselves as the go-to experts in the safety of autonomous systems. Over the coming year, the final of our current Programme, we will continue our broad portfolio of activity and our collaborations across the globe.**

**W**e will continue to publish more unique guidance, specifically with the publication of our Safety Of Deployed Autonomous systems (SODA) methodology that will assure safety of autonomous systems in operation. In addition, we will work with collaborators in different domains to adapt our guidance according to their distinct operating contexts and regulations. Our cross-domain approach enables us to bring learning from one sector to another.

Students from numerous domains will be welcomed to our MSc module on the safety assurance of autonomous systems. We will deliver bespoke training for organisations in safety critical domains including healthcare and automotive, and will

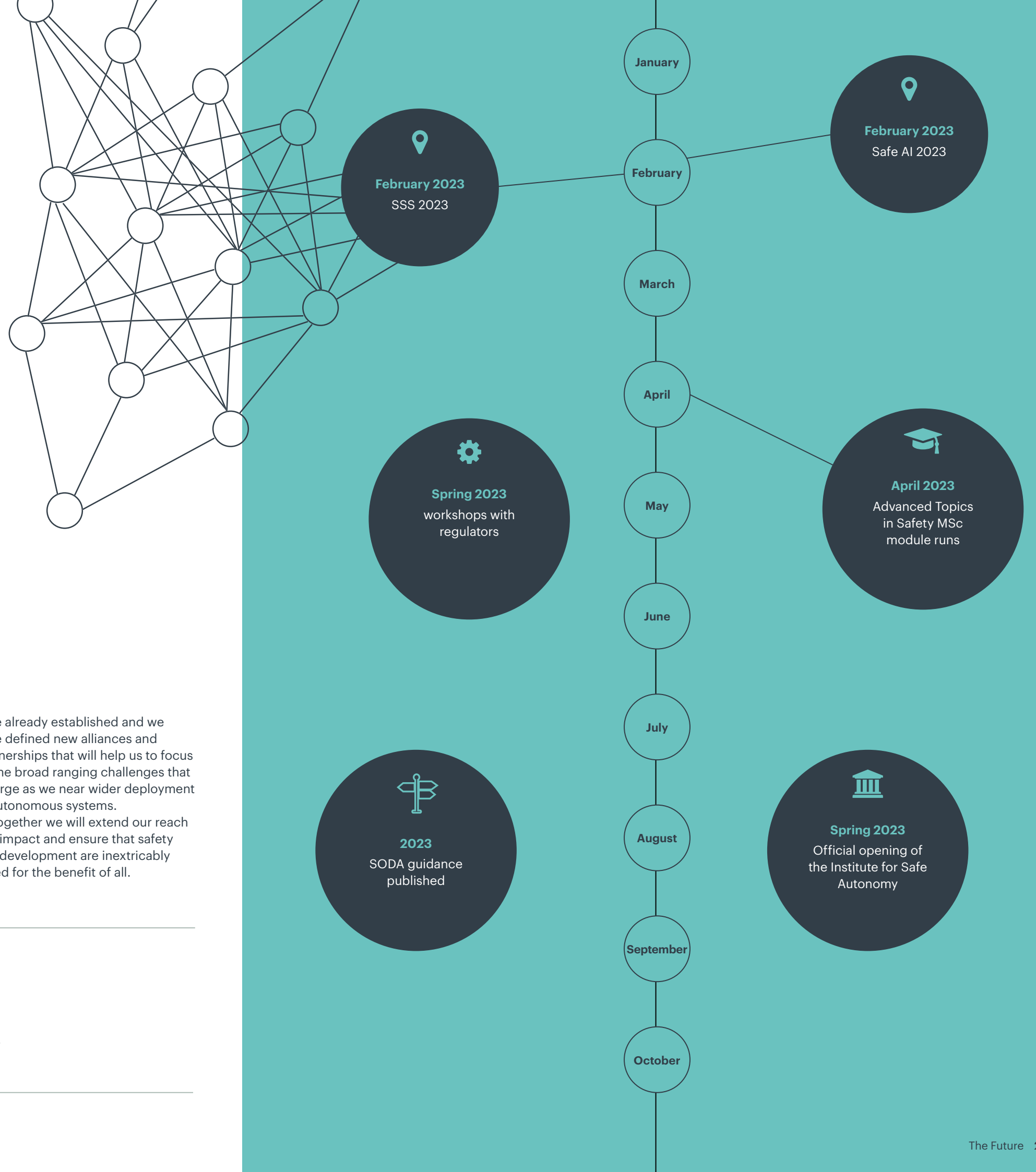
disseminate our research through a range of international workshops and conferences.

Our multidisciplinary community is the golden thread that runs through our work. Last year we identified further areas of research and innovation that we will pursue together. The facilities in our new home, the Institute for Safe Autonomy at the University of York, will help us to work with partners to develop and validate this new research and guidance.

Towards the end of the year we intend to evolve from the Assuring Autonomy International Programme to a new Centre for Assuring Autonomy based in the Institute for Safe Autonomy. We will continue to work with the significant and successful community we

have already established and we have defined new alliances and partnerships that will help us to focus on the broad ranging challenges that emerge as we near wider deployment of autonomous systems.

Together we will extend our reach and impact and ensure that safety and development are inextricably linked for the benefit of all.



## Work with us

If our impact to date has inspired you and you would like to support us as we build on our research to tackle emerging challenges and wider societal issues, we would love to talk with you.

+44 (0) 1904 325345  
[assuring-autonomy@york.ac.uk](mailto:assuring-autonomy@york.ac.uk)

@AAIP\_York  
[linkedin.com/company/assuring-autonomy](https://www.linkedin.com/company/assuring-autonomy)  
[assuringautonomy.medium.com](https://assuringautonomy.medium.com)



