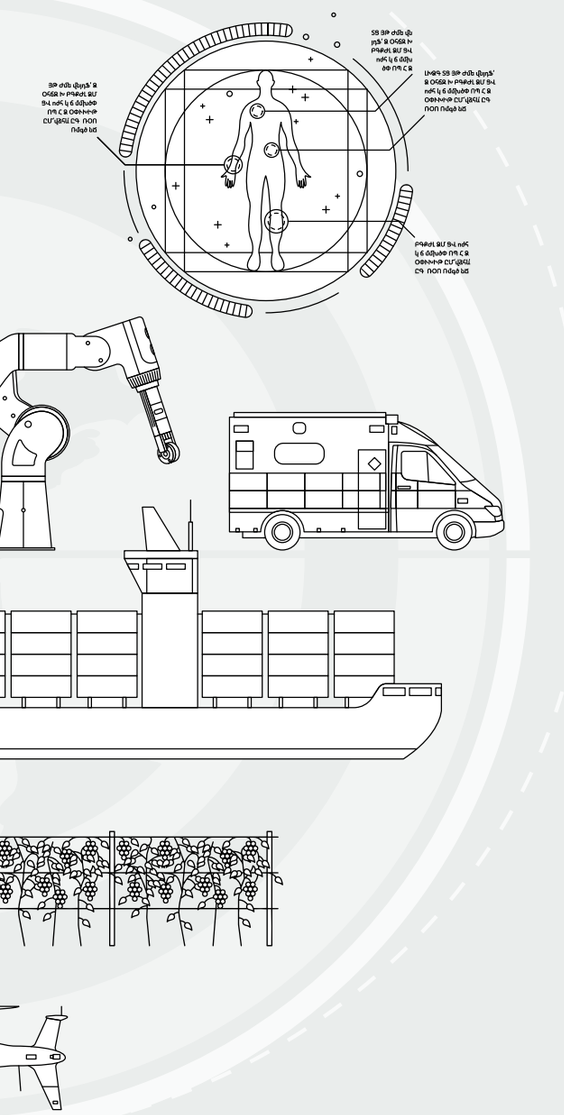
ASSURING AUTONOMY
INTERNATIONAL PROGRAMME

IMPACT REPORT



International experts

How can we be sure an autonomous system, like a self-driving car, assistive robot or drone, is safe and will act as expected? That's exactly the problem being solved by the Assuring Autonomy International Programme (AAIP).

In 2018, Lloyd's Register Foundation and the University of York established the AAIP. In the five years since, we have translated our influential research into guidance that has been accessed across the globe and referenced in new standards.

Through collaborative real-world demonstrator projects, we have discovered best practices and developed methodologies that can be used to give confidence in the safety of autonomous systems in numerous domains.

We understand the skills that are needed by those who develop, assess, or use autonomous systems. Our training and education programmes meet these needs and are influencing colleagues in safety-critical sectors.

We have created an international community of academics, clinicians, developers, engineers, regulators, and students. This diversity in skills, experience, and backgrounds has led us to novel and robust advances in the field.

The scale and complexity of assuring the safety of robotics and autonomous systems have increased since we began, with new challenges emerging and wider societal impacts to consider. Yet we are making an impact that cuts through the noise and have established ourselves as the go-to experts in the safety of autonomous systems. And there is more to come.

Read on to find out how our work is having an impact on people and practices.

This report is based on a review of emerging impact carried out by independent evaluators Paul Rhodes (Paul Rhodes Consulting) and Alan Graver (Skyblue Research Ltd), between September and November 2022, that is available on our website. Paul and Alan conducted 13 in-depth case studies, met with members of the AAIP team, reviewed existing publications, sampled survey data from the community, and attended AAIP workshops. We would like to thank members of the #AssuringAutonomy community who contributed to this independent research.

We're improving safety practices

We have produced the only systematic and structured guidance on the safety assurance of autonomous systems and it is now being adopted and adapted by regulators and industry

By enhancing design and assurance capabilities in industry, the AAIP is paving the way for the safer introduction of autonomous systems.

We are the first organisation to introduce methodical, freely accessible, expert guidance on how to assure the safety of autonomous systems - AMLAS (Assurance of Machine Learning for use in Autonomous Systems) and SACE (Safety Assurance of autonomous systems in Complex Environments).

> *"We are developing an AI flight controller for drones and safety assurance is a key aspect of our product roadmap. Without a way to assure the safety of our neural networks, there is no way to certify them and bring them to market. Despite the interest from government and industry, there is currently no established process for certifying and assuring an AI component. AMLAS fills this gap by providing us with a framework to integrate safety assurance into our development process and build a compelling argument for our safety case."* **Dr Matthew Carr, Co-Founder and CEO, Luffy AI**

Our AMLAS guidance was published in February 2021. It has been accessed across the globe by functional safety leads, software engineers, CEOs, and clinical scientists in domains as diverse as aerospace, rail, metrology, and healthcare. In August 2022 we followed AMLAS with SACE and it has already enabled us to gain visibility in additional countries including Algeria, Mozambique, and Uruguay.

Technical content

Our AMLAS guidance is valued for both its structure and rigorous technical content. This unique methodology enables users to create stronger safety arguments.

> *"The AMLAS methodology really helped to shape our thinking and approach to the assurance of ML/AI. The lifecycle structure mirrors our development phases and it is well presented and easy to read."* **Simon Diemert, Systems/Software Engineer, Critical Systems Labs Inc**

AMLAS provides a strong basis for challenging and testing safety cases. Users are creating supplementary guidance specific to their context and adapting it for ML in operation.

> *"AMLAS is really good, well-written and specific. In RAILS [Responsible AI for Long-term Trustworthy Autonomous Systems project] 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**

Impact

Our guidance is influencing organisations across the globe. It is being embedded in safety processes, adapted for specific contexts, and referenced in new standards.

> *"The AMLAS guidance, though not directly referenced, had some influence on the thinking that is used in the Dstl's Assurance of AI and Autonomous Systems Biscuit Book published in 2021.* **Senior Principal Scientist, Dstl**

> *"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**

It's not just our guidance that influences standards. The impact of our research collaborations also helps to shape our partners' thinking for future standards.

> *"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**

NAME: Shakir Laher
JOB TITLE: Safety Engineer
ORGANISATION: NHS Digital

Context:

The UK's National Health Service (NHS) has 1.5 million interactions with patients every day. NHS Digital has a remit to ensure that the technology used to treat these patients is integrated, deployed, and used safely. As part of that work, Shakir develops software and leads research projects in AI/ML assurance focused on safety. "We began collaborating with AAIP to help us understand our gaps and limitations and to take relevant guidance and implement that in the healthcare domain to assess its applicability and 'give back to the state-of-the-art.'"

Collaboration:

Shakir and colleagues at NHS Digital lead a project group with the British Standards Institution and Human Factors Everywhere. This AAIP-funded research project, Safety Assurance FRamework for Machine Learning in the Healthcare Domain (SAFR), is creating resources to help manufacturers and others to meet the regulatory requirements for their ML healthcare tools.

> *"A specific guidance we have reviewed is AMLAS. It was primarily written for manufacturers of ML, so we have embarked on extending it for 'adopters ' (e.g. clinicians) of ML so that they can assure safety of the technologies in their clinical pathways. This approach has been inspired by AMLAS and by working together. We engaged in a process of reviewing the draft AMLAS guidance through a series of workshops with two manufacturers who brought invaluable insights based on their hands-on use of the technology – this process overall really helped remove bias and it was focused on understanding how AMLAS converged with manufacturers' regulatory practice."*

Impact:

Consequently, Shakir recommends AMLAS for healthcare. "...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," he said, "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."

> *"[Our partnership] spurred us on to more work, to extend AMLAS and include some extra stages. If AMLAS is going to make up part of the ML safety assurance process at national level then manufacturers and adopters will be able to use that and our additional guidance with confidence."*

1200+
downloads of our guidance

Colleagues in
29
countries and
24
domains downloaded our guidance

'BS 30440
Validation framework for the use of AI within healthcare – Specification' references AMLAS

We're leading an international community of change

Our multidisciplinary community benefits every collaboration

The community we have developed is of huge value to us and our collaborators. It enables us to influence change within those organisations and acts as an enabler to our wider impact.

The strengths of our community are numerous and our unique approach offers many advantages to those we work with.

> *"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...so that safety is 'live' for all, not just something that is produced at the start of a project and put in a ring binder to pull down when you think you need it."* **Professor Tom Lawton, Critical Care Consultant, Bradford Teaching Hospitals NHS Foundation Trust**

Multidisciplinary

Not just a buzzword. Our community comprises academics, clinicians, computer scientists, ethicists, lawyers, ML developers, regulators, safety engineers, students, and others. This truly broad group ensures we steer away from groupthink and silos.

> *"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 that we've worked on together has benefited hugely from this diversity."* **Dr Xingyu Zhao, Lecturer in AI, University of Liverpool**

> *"The AAIP has got scale, a range of different demonstrators [projects] 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**

> *"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**

> *"AAIP provided not just funding, but gave us confidence, networks, insights and expertise across disciplines we might not otherwise have accessed as easily including IT, engineering and human factors."* **Dr Nigel Rees, Head of Research and Innovation, Welsh Ambulance Services NHS Trust (WAST)**

Impact

The multidisciplinary collaborations we make possible lead to new research outcomes. They create the conditions for complex challenges to be tackled, and knowledge and connections to be extended. The results are more credible and compelling and more widely shared and potentially applied than would otherwise be possible.

They also create the space for 'failure', learning from what does not work as well as what does.

> *"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 develop 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**

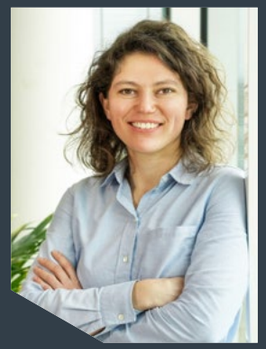
> *"The programme allows ideas to be tested and workshopped. We determine if the idea is viable, but crucially, then we have the critical mass to actually do something with it. I see a very broad, interdisciplinary consensus that can have more of an impact and achieve greater levels of acceptance."* **Professor Dr Simon Burton, Research Division Director, Fraunhofer IKS**

> *"We have been really pleased with the collaboration. There was strong communication and integration of ideas through our weekly researcher meetings. Without that, the outputs would have been nowhere near as interesting."* **Dr James Law, Director of Innovation and Knowledge Exchange, Sheffield Robotics**

NAME: Lydia Gauerhof

JOB TITLE: Research Engineer

ORGANISATION: Corporate Research at Robert Bosch GmbH



Context:

Lydia's work at Bosch is focused on the safety assurance of AI in automated driving. She is passionate about bringing together industry and research. Lydia was working with AAIP Fellow, Professor Dr Simon Burton, on ways to safely introduce ML to different applications in the automotive domain, when she was introduced to AAIP. "There was no ISO standard for AI," she says, "and we both recognised the need for consensus between academia and industry."

Collaboration:

Lydia first began working with AAIP as part of the review of the draft AMLAS guidance. "I was glad to be part of the 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 from medical devices."

We then welcomed Lydia as an AAIP Fellow, providing a more direct connection to the AAIP team. "The York team had a good understanding of ML components. It was firstly useful to have this sanity check that 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."

Impact:

Lydia was one of the experts who reviewed the AMLAS guidance and helped to develop an example on safety requirements for pedestrians at a crossing. The collaboration has also enabled us to share the AMLAS guidance widely within Bosch.

To share the guidance, Bosch convened two workshops facilitated by AAIP. The workshops were convincing: "Universities focus more on methods, Bosch focuses more on products," says Lydia. "Of course, methods are needed for development and here is where the value of universities comes in through their cross-domain expertise."

The guidance was then shared across different departments. Lydia highlights the function AMLAS can play within the automotive industry. AMLAS provides a shared resource and approach that can be shared across project teams. It can be used as a credible basis for discussions about assuring safety. The alternative is to develop a bespoke safety case, which typically is only well understood by those who created it.

> *"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?'"*

Research collaborators based in
16
countries

17
Programme Fellows

2000+
citations of 178 papers

We're educating people with the new skills they need

Our unique educational offer is improving awareness, understanding, and knowledge of safety assurance

As our research and innovation activities have matured, we have developed and deliver educational materials of increasing depth to wider audiences.

Our formal education focuses on the key audience of safety engineers and developers. We developed and deliver the only MSc module and CPD module focused on the safety assurance of robotics and autonomous systems. We have supported the upskilling of practitioners in multiple domains.

> *"Much of education is about confidence. Even though you may not know everything, if you have confidence, you go ahead and you try things and maybe you'll find the solution. Without that confidence, you may not get to those solutions. Taking the module has made our lives easier now, as we know what people are talking about in terms of the safety of autonomous systems. It allows us to make more informed decisions."*
> **Romas Puisa, Product Safety Engineer, Thales**

Informing assurance

Through demonstrator projects and research, our partners develop their skills and knowledge in their area of focus.

> *"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 demonstrator [project] has shown what could be possible from an assurance perspective and developed a framework that can be compared with other projects to help build consensus. 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."* **Nicholas Hall, HM Principal Specialist Inspector (Advanced Automation and Cyber Security), Health & Safety Executive**

Impact

The impact of our training extends beyond the person who attends the course or CPD training to those who they talk with about the processes and concepts they have studied. The healthcare domain is an early adopter of AI and autonomy and consequently our impact is very strong and clearly demonstrated.

> *"Safety has been improved at WAST and we are due to present our findings to the Board shortly...[the impact of the demonstrator 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)**

We have improved the skills, knowledge, and understanding of those who work in this safety-critical sector. We are also influencing the guidance made available to developers and users of AI in healthcare, with AMLAS referenced as part of the forthcoming 'BS 30440 Validation framework for the use of AI within healthcare – Specification' standard. Guidance to support the use of AMLAS in healthcare will also be published on the NHS Digital website in 2023, supporting the use of our leading methodology in this sector.

NAME: Sean White
JOB TITLE: Safety Engineering Manager
ORGANISATION: NHS Digital

Context:

"In the last four to five years AI as a concept has exploded across society. Its benefits are known, but there is also a great deal of misunderstanding about it – a belief that the technology is running wild, changing its decisions and being unpredictable in its behaviours. We need to correct this perception and assure society that it is safe to exploit this technology. I also realised that as a team we needed to develop and grow our knowledge and experience with this technology to support our national safety assurance capability."

Collaboration:

Together we delivered a one-day pilot course to 100 NHS Digital delegates exploring what AI is and how it can be used and assured from a healthcare perspective. "This was targeted at people such as clinicians, safety engineers, developers, project managers and people allied to the role of 'clinical safety officer' as defined in the national standards. Feedback was strong, so we made a commitment with the AAIP to develop a deeper course.

> *"This was purposely delivered to a small community of 25 external delegates including clinical safety officers, clinicians and developers. The course was fully subscribed, and participants gained a greater understanding of how to effectively assure AI for use in a care pathway...We ran this course twice and feedback from learners was that they had a stronger understanding of the state of the art; it corrected misperceptions about AI; and it encouraged them to think differently about how to assure its safe deployment."*

Impact:

Our partnership has enabled us to support the needs of NHS Digital. "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."

> *"We are committed to run the course again and will schedule dates for 2023 and beyond...I would like to see understanding of AI grow and skills develop in effective assurance, the AAIP gives us the resources and knowledge to achieve this."*

1
unique MSc module focused on the safety of RAS

500+
people trained on AMLAS

150
health professionals trained

We're influencing the wider landscape

We're bringing safety to the forefront and are recognised as the experts

Interest, research, and funding in the area of safe and trusted autonomous systems and AI has increased hugely in the five years since AAIP began.

Internationally there is now an Institute for Assured Autonomy at Johns Hopkins University in the USA, a European Training Network for Safe Autonomous Systems, the Trusted Autonomous Systems Defence Cooperative Research Centre in Australia, and in the UK the UKRI Trustworthy Autonomous Systems (TAS) Hub was established in 2020.

We value our collaborations with these organisations and programmes and know that to move forwards we need to bring the right people together at the right time. The unique experience and expertise that we bring to the field are evident in our influence.

International experts

It is clear that we are viewed as the world experts in the safety of autonomous systems. This is because of our knowledge and experience, our ability to bring the necessary disciplines together, to lead the way, and to do things differently to ensure they're done right.

> *"There's been a lot of interest in safety of robotic systems and the approach that AAIP have taken, and the expertise that the York team specifically bring to the space is quite unique...York's team are the international experts in safety assurance, with a very different skillset, methods and approaches."* **Dr James Law, Director of Innovation and Knowledge Exchange, Sheffield Robotics**

> *"The AAIP team at York are doing world-leading research and are at the forefront of the field that they're working in."* **Rachel Horne, Assurance of Autonomy Activity Lead, Trusted Autonomous Systems**

A one-of-a-kind facility

In 2022, the University of York opened the Institute for Safe Autonomy. It is funded through Research England's flagship capital investment scheme, the UK Research Partnership Investment Fund (UKRPIF) and co-investment from our partners. AAIP is a founding partner and helped secure additional funding from an international network of collaborators and private donors.

The Institute is the new home for the AAIP team and our research. Providing world-class lab facilities and space to welcome our partners, the Institute is a living lab where we test, validate, and demonstrate our research and methodologies.

Impact

AAIP has brought a safety focus to the robotics and autonomous systems landscape that had been missing. Our expertise is filling the gap, building on more than 30 years of work at York on the safety of complex systems.

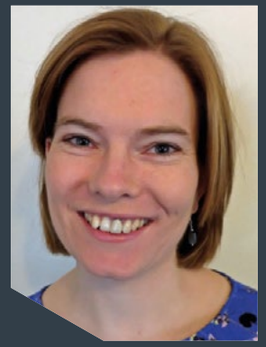
We have influenced funding decisions, leading to increased investment in the safety assurance of autonomous systems in the UK. Our work has brought clarity and structure that was previously lacking through our cross-sector perspective.

> *"There are different paradigms for safety in different sectors. Each has their own paradigm, and they are reaching the limits of those. Structures have become calcified in these industries – particularly in the language and setting of industry standards. The Programme can then disrupt those staid structures. 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**

NAME: **Dr Helen Niblock**

JOB TITLE: **Head of Regional Engagement (NE, Yorks & Humber)**

ORGANISATION: **Engineering and Physical Sciences Research Council (EPSRC)**



Context:

The EPSRC's vision is for the UK to be recognised as the place where the most creative researchers can deliver world-leading engineering and physical sciences research. AI is one of the largest growth areas for EPSRC funding: their AI Technologies portfolio increased from £56M in 2016 to £128M in 2022 and their AI training portfolio increased from £31M to £102M in the same timeframe.

Connection:

Helen works across the region to understand the research and innovation taking place, identify synergies, and make connections. As part of this, she connects with the AAIP team. EPSRC lead the UKRI-funded £33M Trustworthy Autonomous Systems (TAS) programme, which launched in 2020. Its vision is to enable the development of socially beneficial autonomous systems that are both trustworthy in principle and trusted in practice by the public, government, and industry.

Impact:

York's influence is particularly apparent in the TAS programme. "York influenced the topics of the nodes within the 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."

Helen also recognises the wider influence as a result of AAIP's research publications: "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."

> *"There was no [safety assurance] framework in place and if York hadn't been doing that work, I don't know who would have done it."*

£35M
Institute for Safe Autonomy established

£41.6M
investment leveraged by AAIP

£14M
funding leveraged by demonstrators

Lessons learned and looking ahead

We're still discovering

And can recognise the scale and complexity of what's to come

The most tangible way that we have supported safer industrial practices is through our guidance. Over the past five years, we have refined this through our work with demonstrators and research projects.

We have learned and improved our understanding of best practice as well as current and upcoming challenges. Our original Body of Knowledge has been transformed into an online resource with practical, systematic, and clear guidance.

We want to extend our guidance further, adapting it to different domains to help with the unique challenges that distinct operating contexts and regulations bring and to ensure that it keeps pace with new technologies. Furthermore, now that RAS are closer than ever to market, our work on pre-deployment assurance needs to be complemented with guidance for assuring safety in operation.

We have learned that our cross-domain work is not only imperative if we are to avoid reinventing techniques and wasting precious time, but it is our calling card – bringing to bear learning from one sector on another. We know that we must continue to work in multiple industries with a variety of partners to realise the greatest benefit.

Our multidisciplinary community is the golden thread that runs through our work. It leads to new research outcomes, additional funding to support our work, and faster progress. We have learnt that embedding AAIP team members in projects, even in a small way, allows us to influence work outside our immediate scope. As such, we will continue to nurture and grow this community.

With existing collaborators in 16 countries, we have been successful with the intended global focus of our work, but we will now look at how we expand our international community further, to China and sub-Saharan Africa, for example.

Our impact on regulations and standards is advancing steadily, though these are habitually slow and methodical in their development. Our projects now regularly involve regulatory organisations from the start, which adds significant value to the experience, learning, and outcomes achieved.

Our MSc module, 'Advanced Topics in Safety', makes a unique and important contribution with its focus on safety assurance. Rolling it out at scale is a challenge we need to tackle. To do this, we will learn from our successful partnership with NHS Digital, where we have developed customised professional education and training aligned to their wider organisational goals around AI/ML skills and competence. We recognise that education is needed at different levels, and that rolling out professional development for varied learners will be key to our future impact.

We have established a strong foundation on which to build the future of AAIP. We are aware that many challenges remain and are already identifying where our attention should be focused. For the next phase, we are developing a broad portfolio of activities to enable us to deliver additional long-term impact. We will need more stability and new partners as well as continuing current successful collaborations. Furthermore, we will work with our international community to transition to a new AAIP that extends our research and impact.

The last word *(for now)*

We've created a tipping point

We will build on the impact we've already had through further research, education and training, and through direct support to industry and regulators

We're delighted with our impact 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. Standards can take decades to evolve, yet we have started to make an impact even in that area, particularly through work with the BSI.

There is more to do. We have new guidance to publish on the safety of understanding and decision-making, and on the societal acceptability of autonomous systems. And we must turn our attention to the safe deployment of the technology - how can users be confident in the safe operation of the system, and how can they be managed safely in service?

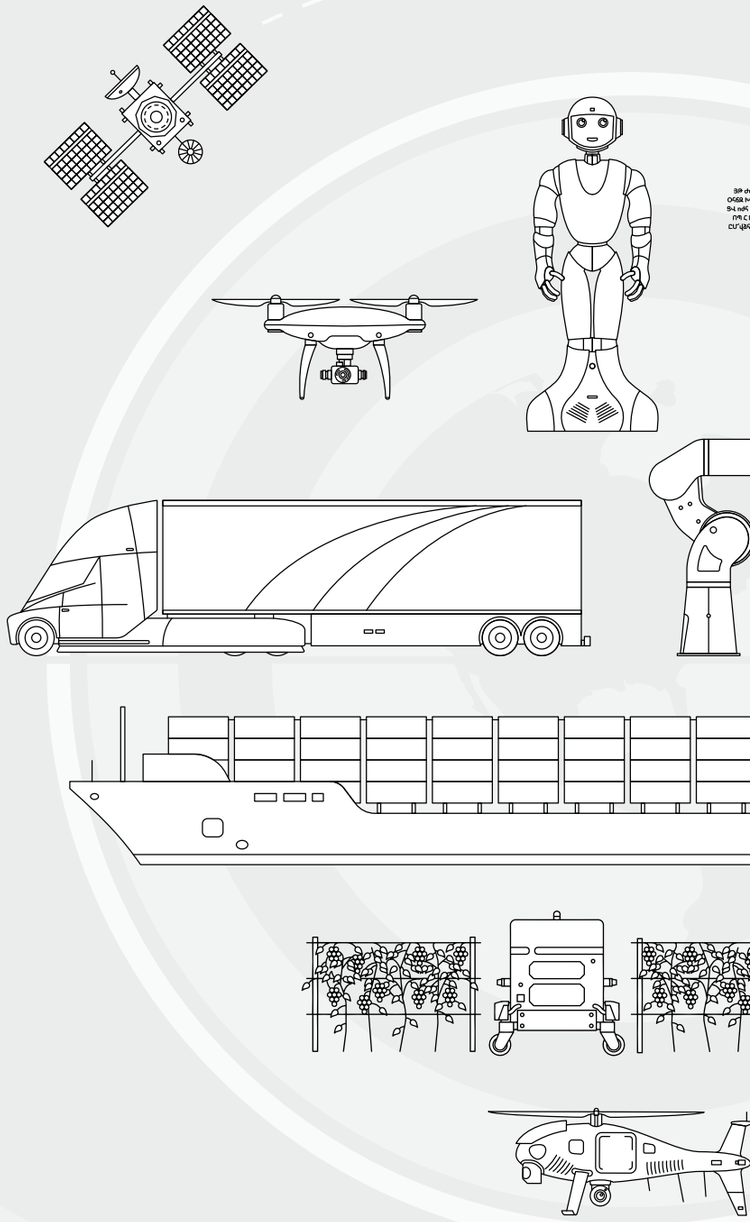
To do this, we will build on the international recognition of our expertise to bring safety even further to the forefront. Our training and education offer will expand to reach more people and give them the confidence to assure the safety of the autonomous systems in their organisation. And we will continue working collaboratively across disciplines - bringing together the right people at the right time.

We extend our gratitude to those who already support us and who have worked with us. Together we are making a measurable difference in the safety of autonomous systems. We hope you are proud of what we have achieved together.

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.

Professor John McDermid
PROGRAMME DIRECTOR

Dr Ana MacIntosh
PROGRAMME MANAGER



UNIVERSITY of York