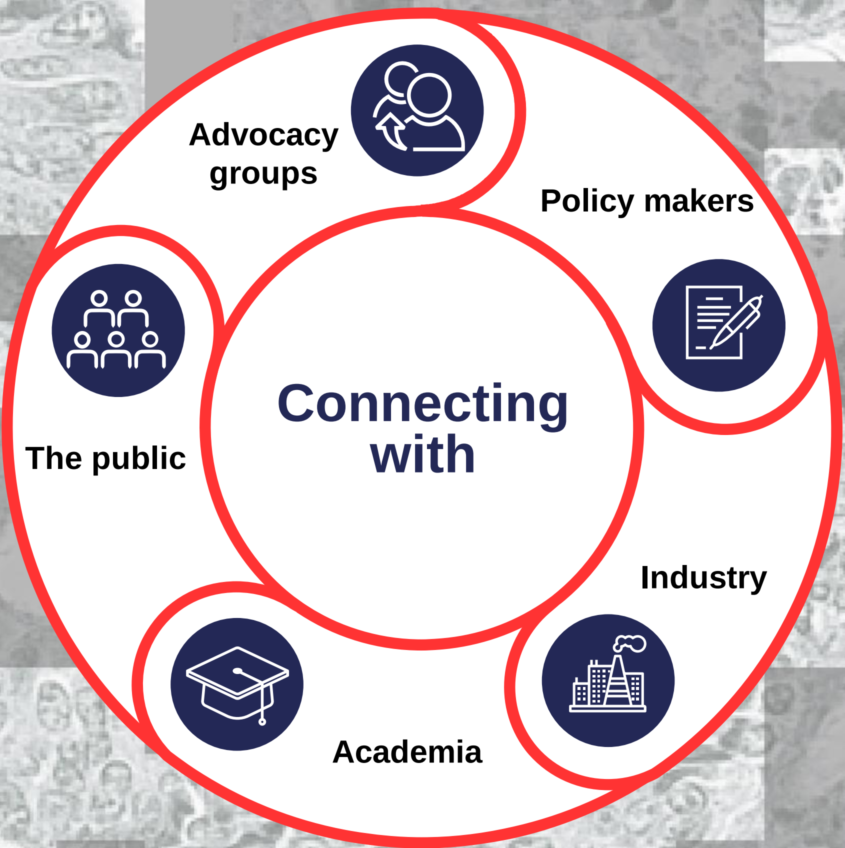
UKRI Trustworthy Autonomous Systems Hub

Celebrating the impact of the UKRI Trustworthy Autonomous Systems (TAS) programme



To capitalise on the opportunities and address the challenges presented by autonomous systems.

Index page:

Cover picture: Epithelial cells in colorectal tissue as identified by an interpretable machine learning algorithm. This is an important clinical task, as colon cancer originates from epithelial cells. From "Model Agnostic Interpretability for Multiple Instance Learning" – Early et al. 2022.

Below: Hazel Sears, Knowledge Exchange and Impact Officer at the University of Nottingham with UBTech Alpha Mini.

The Trustworthy Autonomous Systems Programme: a community building effort

By Professor Sarvapali (Gopal) Ramchurn
Director of the UKRI Trustworthy Autonomous Systems (TAS) Hub from the University of Southampton

Autonomous systems are becoming embedded in almost every aspect of our lives in complex ways, some of which are more visible than others. While such systems hold the potential to help address many of the most fundamental socio-economic challenges on a personal and global level, they can also invoke a fear to many of losing control to machines.

It is a great privilege to lead the UKRI Trustworthy Autonomous Systems (TAS) programme. Over the last four years, we have been granted a timely opportunity to develop the best practices for world-class research that will enable the development of autonomous systems which are both trustworthy in principle and trusted in practice by individuals, society and government.

We champion fundamental principles of Responsible Research and Innovation (RRI) to help ensure systems are socially desirable and undertaken in the public interest. As you will see in this publication, all TAS projects adopt RRI, by investigating emerging socio-technical issues with autonomous systems or developing mitigations for the unintended consequences.

Our engagement with over 150 industry, government, and third-sector organisations demonstrates how the core concepts of trust and responsibility have been grounded in real-world use, rather than hypothesised scenarios.
This, we believe, bodes well for the long-term impact of our research.

Through more than three hundred engagement activities and our extensive social media networks, TAS has created a community which has contributed to the development of academics with relevant skills to drive forward research, having generated 192 publications already in a wide range of sectors, supporting economic growth and helping with some of the challenges facing society.

We estimate that TAS Hub has secured at least £36 million in 'follow on funding' from other research grants and activities and the TAS Nodes have gained around £40 milion.

We have created this publication to help communicate and accelerate the impact of our pioneering research programme.

We hope it will be a starting point for further engagements with stakeholders, who will make use of the work in a responsible and considered way.

We will continue to collaborate and embed many of the lessons learnt in upcoming initiatives, such as Responsible AI UK, the AI Hubs, and AI Centres for Doctoral Training (CDTs).

Professor
Sarvapali (Gopal) Ramchurn

The transformative potential of trusted autonomous systems

The TAS programme has played a central role in shaping government policy on trustworthy autonomous systems, through working with the Department for Culture, Media and Sport, the Department for Transport, the Ministry of Defence, and the Law Commission, among many others.

Pictured: Dr. James Dracott (right) with Professor Steve Benford (left)

TAS stands as a "beacon of innovation for an inclusive future"

Dr James Dracott,
Head of Artificial Intelligence and Robotics at UK Research and Innovation (UKRI) said:

"At UKRI, we are profoundly aware of the transformative potential of AI and autonomous systems in addressing the large-scale, intricate challenges faced globally. From driverless vehicles to advanced patient care, the possibilities are vast, but their success hinges on society's trust and utilisation. The TAS programme, funded through the UKRI Strategic Priorities Fund and led by EPSRC, has been instrumental in nurturing a vibrant community of researchers and stakeholders. This diverse network, cutting across various sectors and disciplines, is united in its commitment to Responsible Research and Innovation, and inclusivity. Through TAS, we have seen how such collaborative efforts can significantly enhance the development, deployment, and governance of autonomous systems, aligning technological advancements with ethical principles and social values."

"Our vision extends beyond technological advancement; through TAS, we have cultivated an ecosystem where future researchers are equipped to create autonomous systems that are not only technically adept, but are also ingrained with trustworthiness and societal benefit. This initiative reflects UKRI's broader mission: to champion research that not only expands knowledge, but also addresses critical societal challenges and contributes to the UK's economic prosperity. As TAS continues to evolve and its impacts mature, it stands as a beacon of innovation, guiding industry and policymakers towards a future where autonomous systems are synonymous with reliability, inclusivity, and public welfare."

Working together to deliver impact across society and the economy

"We've reached an exciting stage of the programme where TAS can showcase broad impacts across society and the economy. Our impacts are transforming Cultural and Creative Industries, Health and Wellbeing, Defence and Security, Energy and Environment, Transport, and Governance, Regulation and Management of autonomous systems. From improving acceptance of autonomy in healthcare through to regulation for safeguarding AI, novel public engagement in research, and increasing participation in creative experiences for people living with disabilities, the TAS programme has shown how autonomous systems can improve lives, create a more inclusive society and contribute to a productive economy."

TAS Impact Director, Professor Michael Boniface, from the University of Southampton

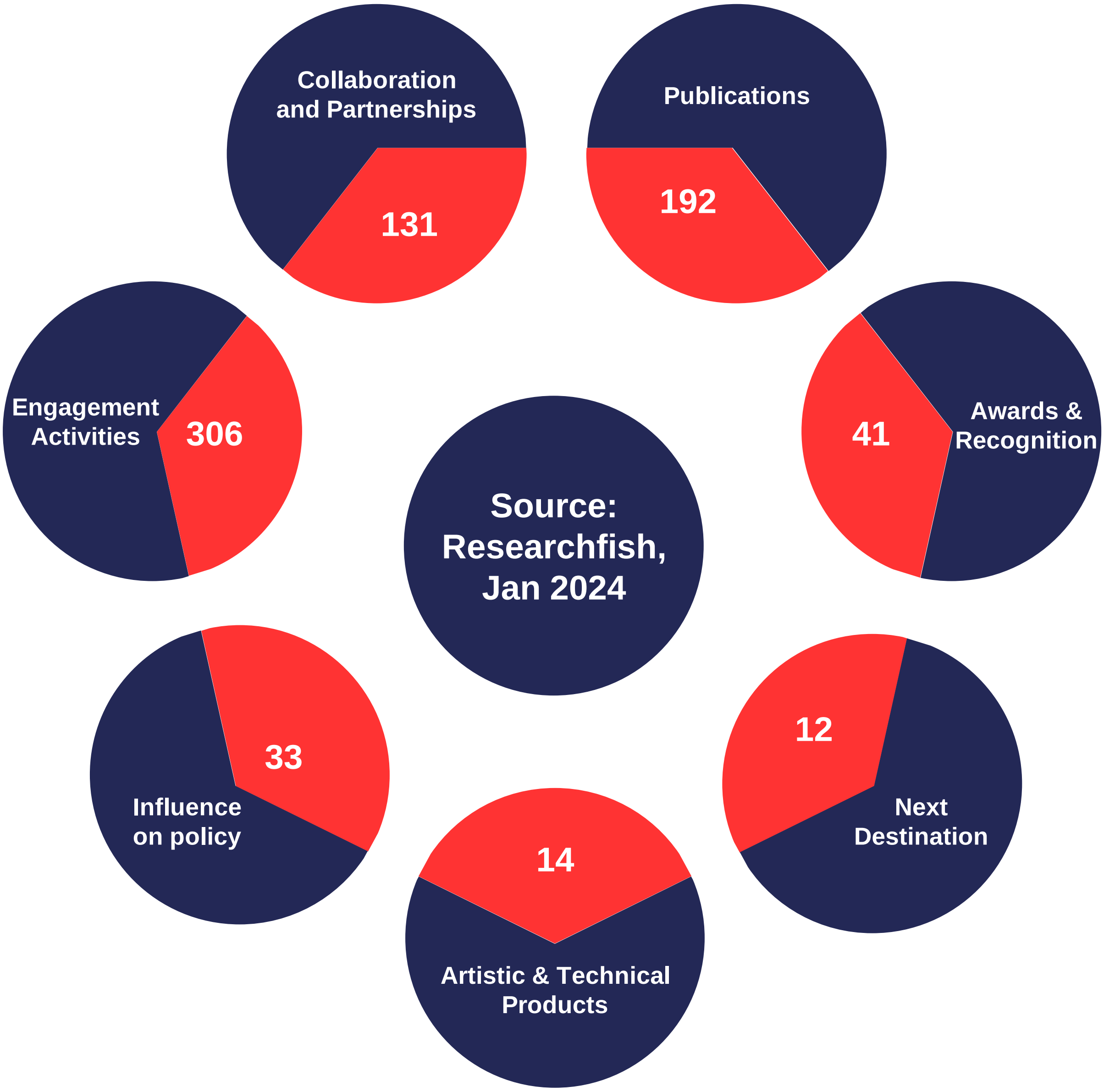
Reaching communities far beyond academia

"TAS has been actively engaging with communities outside of academia since the early days of the programme. Across TAS' span, members have taken part in international industry events, featured on TV and radio, given workshops at schools and universities, featured in galleries and museums, participated in podcasts, and appeared in print media. These are really important routes for sharing the message of responsible research and innovation.

"As well as that, our own podcast series has also reached beyond academia with topical, evidenced and trustworthy content, as well as providing a useful teaching resource for our own undergraduate and postgraduate students.

"Overall, TAS' outreach has been important in helping to break the silos which can occur if we're restricting institutional, geographic or academic disciplines and has been an important channel to share the fruits of our academic endeavour more widely."

Dr Kate Devlin *(pictured below)* TAS Director of Engagement and Reader in Social and Cultural Artificial Intelligence, King's College London, said:



Accelerating the impact from research

"The TAS programme consists of the TAS Hub, six thematic Nodes and four projects focused on Responsibility of TAS, and the Pump Priming projects that the Hub has funded. As part of its Research programmes, the TAS Hub has funded and delivered 76 multidisciplinary projects on diverse topics, spanning more than 20 academic institutions, involving over 170 partners.

"The TAS programme has brought together a diversity of disciplines, project areas and representations from across the UK.

"A key feature of all TAS projects is the commitment to integrating Responsible Research and Innovation (RRI) and Equality, Diversity, and Inclusion (EDI) as part of the research plans. RRI in particular is an impact highlight, with TAS supporting the creation of novel research methods that have been adopted widely by projects and external organisations."

Professor Joel Fischer TAS Research Director, from the University of Nottingham

Building collaborative communities

121
Industry partners

55
Government agencies

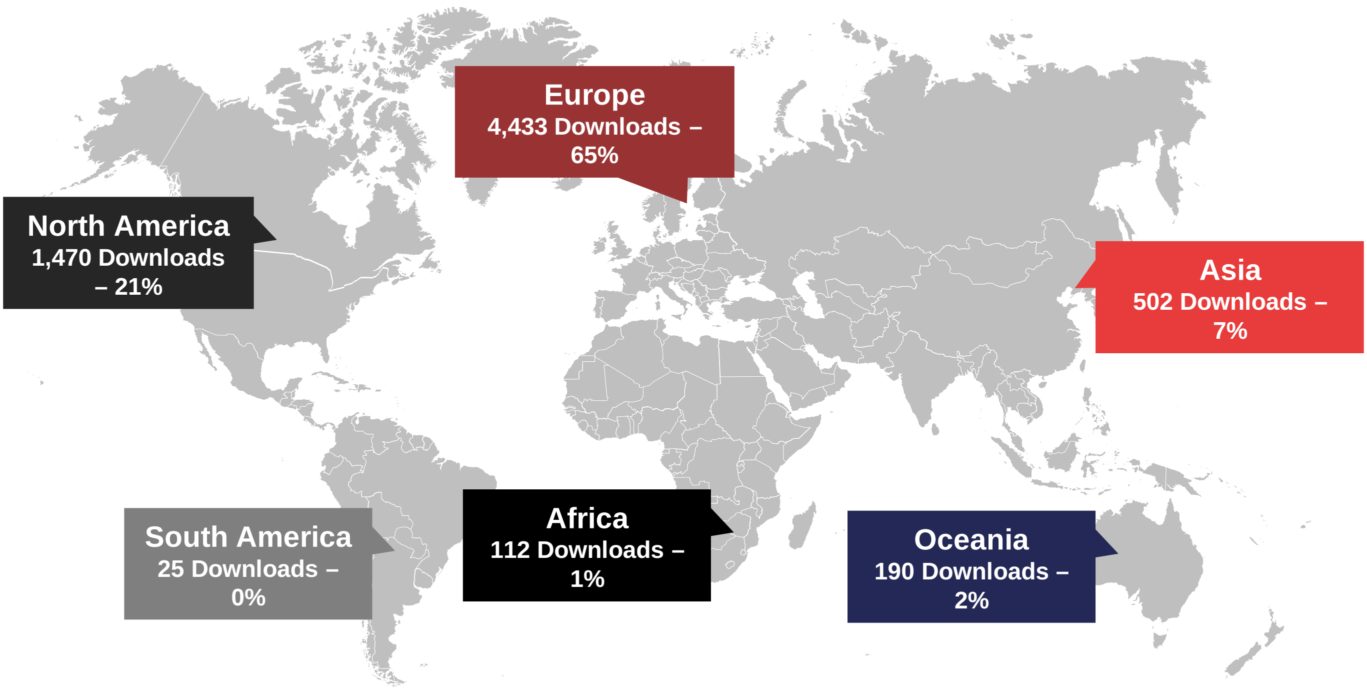
We've engaged

17
NGOs

306
Engagement activities

Source: Researchfish, June 2023 mid-term report

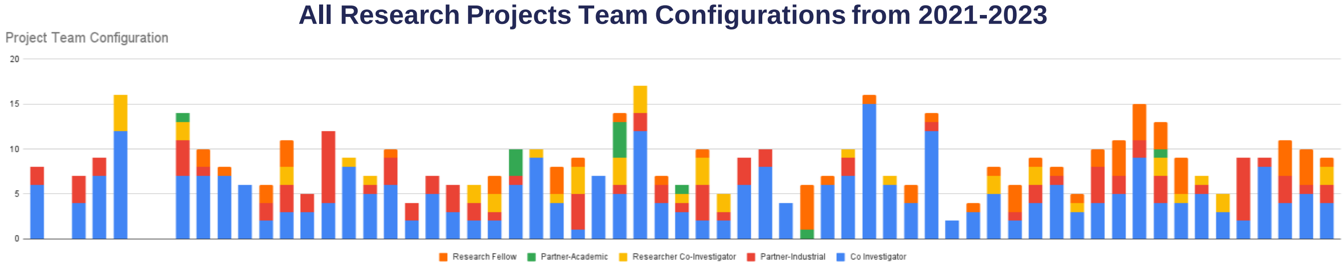
The TAS "Living with AI" Podcast has included already included 39 programmes and been downloaded in six continents



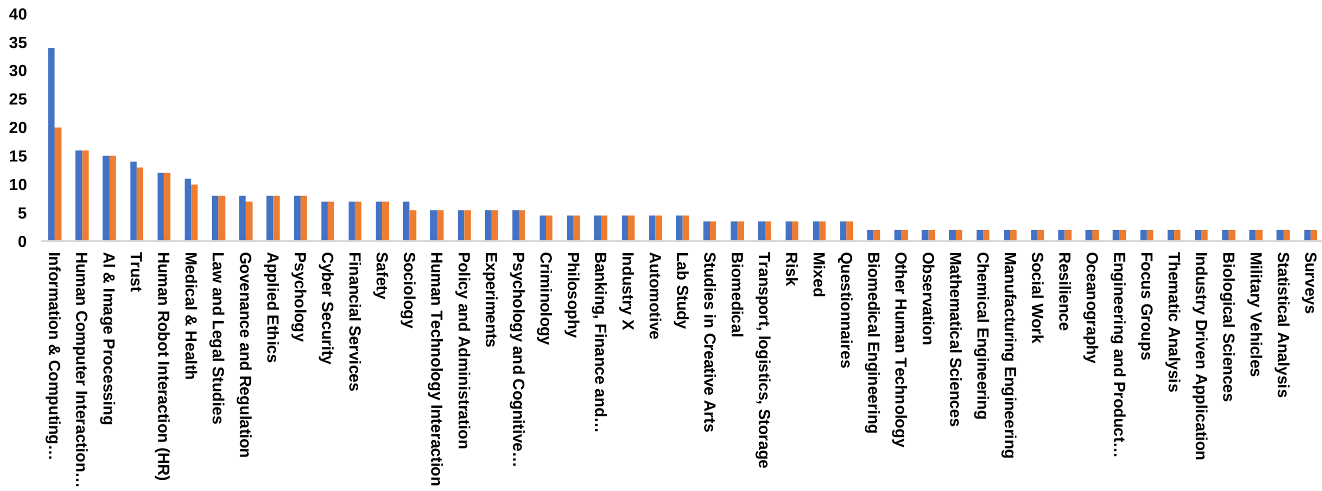
Countries where TAS has engaged

Over 60 disciplines are represented in TAS Hub projects demonstrating the significance of autonomous systems across domains. Priority is given to fund Early Career Researchers (ECRs) as project leads to help foster opportunities for career development, shown below. Many reviewers reported more than one disciplinary expertise and subject areas ranged from Information and Computing Sciences to Criminology and Political Science.

All Research Projects Team Configurations from 2021-2023



Number of reviewers per research area across TAS Research Funding Calls



Collaborating internationally

The TAS programme is one of the largest research programmes in the world, providing academic leadership in responsible and trustworthy autonomous systems. International collaboration has been important to foster global problem solving, global governance and sharing of knowledge and expertise. Novel collaborative approaches were established, leading to funding from the US Department of Defense and Department for Transport. TAS has led global conversations on trust in AI through participation, international events and standard-setting activities.

International conferences and workshops

TAS Hub and Nodes have organised and participated in international conversations on responsible and trustworthy AI and autonomous systems. Key activities included:

1. Organisation of the TAS Symposium at Herriot Watt, with over 150 attendees.
2. Organisation of the International Conference on Robotics and Automation (ICRA) with over 30 attendees.
3. Sponsoring of the international conference on Autonomous Agents and Multi-Agent Systems (AMMAS), involving over 2000 researchers.
4. Organisation of Wallenberg AI, Autonomous Systems and Software Programme.
5. The UT Good Systems Symposium in Austin.

Stacey Ingram Kaleh, Good Systems Network Relationship Manager said: "The TAS team exemplify the ways in which values-driven and values-aligned partners can work together to accelerate and maximize collective impact, and we look forward to continued collaboration and innovation in service of our global community."

Pictured left, the Executive Team of Good Systems Ethical AI at UT Austin. Below TAS conference panel.

TAS Global Sandpit

Following a series of exploratory activities funded by the US-UK GRIP, key universities with aligned programmes were chosen for a significant international research collaboration, which became recognised as the TAS Global Sandpit.

JOHNS HOPKINS
INSTITUTE for
ASSURED AUTONOMY

The Global Sandpit brought together researchers from TAS partner universities - John Hopkins University Institute of Assured Autonomy and University of Texas Good Systems - to co-develop ideas for collaborative research projects, as part of an accelerated four-day process.

Mentors were selected from US and UK institutions. The Sandpit was held at Airlie House in Washington D.C. Groups were formed organically through team-building activities, and this led to the production of seven research proposals and pitches. Two were selected to go forward.

The two projects funded out of the Global sandpit included:	TAME Pain: Trustworthy AssessMEnt of Pain - Listening Between the Lines	MACAIF: Machine-learning Auditing for Clinical Artificial Intelligence Fairness

25 Researchers

04 Days

05 Mentors

07 Project ideas

02 Selected projects

Picture: Jess+ an intelligent digital score system for a mixed ensemble of non-disabled and disabled musicians. Credit BBC R&D.

Cultural and creative industries

Artists exploring the experience of interacting with autonomous systems by Steve Benford, Professor of Collaborative Computing at the University of Nottingham

Thirteen artist awards were given in this area, highlights include:

Music industry
Fairness in selection bias and the public debate on competition in music recommender streaming.

Culture and Heritage
Enhancing museum visits through a social robot to make the experience more accessible and inclusive.

Social Media
Methods, ethics and ecosystem of regulatory inspection and audit of algorithmic systems in social media platforms.

Artists in residence delivered huge value to the TAS ecosystem:

1. **Artists bring creative thinking to TAS that drives emerging technologies in innovative directions**
2. **Public artworks give audiences first-hand experiences of emerging technologies, provoking deep reflections on matters of trust, autonomy, responsibility and inclusion**
3. **Creative industries are an important socioeconomic sector of the UK that need to be addressed on their own terms through artistic methods.**

We initiated cultural ambassadors, Blast Theory, to deliver a landmark artwork called Cat Royale in which a small family of three cats, Ghostbuster, Clover, and Pumpkin, inhabited an AI-driven 'cat utopia,' at the centre of which a robot arm tried to enrich their lives by playing with them. A video installation of the work exhibited at the Brisbane World Science Fair and the Science Gallery, London, engaged thousands of people with the deep questions of trust and autonomy posed by the seemingly simple idea of using AI to create such a utopia. The project also delivered research papers on designing robot worlds, embracing play with robots, and the ethics of multispecies

Cat Royale: a utopia where cats live in harmony with artificial intelligence.

Overcoming barriers caused by disability when creating music by Craig Vear, Professor of Music and Computer Science at the University of Nottingham said:

"Jess+ an intelligent digital score system for shared creativity with a mixed ensemble of non-disabled and disabled musicians. The overarching aim is that the digital score enables disabled musicians to thrive in a live music conversation with other musicians, regardless of the potential barriers of disability. Our findings showed that the implemented user-centred design decisions and trustworthy AI approach led to rich experiences for the musicians which in turn transformed their practice as an inclusive ensemble.

"TAS support amplified the potential of this research project through funding enabling the support of two additional members of the team. They enhanced the development of the project and the significance of the data we collected. Without this support we would not have had such an impact on the musicians involved nor the depth of insights generated by the research process."

Inviting artists to explore the human experience of interacting with autonomous systems

TAS has engaged artists through innovative residencies and projects. The programme has delivered a series of residencies, where artists participated in ongoing TAS projects, with new public artworks on AI-related topics, from engaging disabled dancers with robots to exploring bodily interaction and inclusion, to designing telepresence robots for museums.

Luca Vigano and Ali Hossaini

"The First" film created in partnership with the National Gallery using art created with machine learning and other AI techniques.

Rachel Ramchurn and Roma Patel

Delving into the aesthetic interactions with an autonomous system to help narrow the gap between robotics and interactive arts for children (their work is pictured right).

Dr Richard Ramchurn

Creating 'Autocar' a speculative design which can be used as a way of exploring people's attitudes to and implications of future technologies to investigate how trust works with autonomous vehicles.

Dr Rachel Jacobs

Exploring disrupting ideas of trust in our mirrored reflections with the potential to develop meditative and playful apps that could be embedded in a diagnostic mirror that sits in people's homes.

Picture: Dr Soorati and team demonstrate their human swarm systems at the University of Southampton's Science and Engineering Festival.

Defence and Security

Autonomous machines to protect us By Dr Mohammad Soorati, Assistant Professor of AI at University of Southampton

TAS has addressed the critical challenge of safeguarding against harm and ensuring that autonomous systems remain a force for societal good. This includes applications in the Defence and Security space, working with colleagues at the Royal United Services Institute (RUSI), Defence Science and Technology Laboratory (Dstl), the Ministry of Defence, and a number of companies including Northrup Grumman, BAE Systems, and Thales. TAS also attracted over $750k USD from the US Dept of Defence and Dstl to fund three defence-related projects on Testing, Evaluation, Verification and Validation (TEV&V).

TAS supported research projects that:

- **Enable situational awareness, explaining the state and intention of multiple autonomous systems to a single human operator without increasing the operator's mental workload.**
- **Cyber-security frameworks and algorithms to protect systems involving humans and autonmomous systems.**
- **Built an open-source platform to operate large multi-robot systems that is now accessible to the general public, researchers, industry experts, and policymakers to learn what it means to have control over not only one but multiple autonomous systems.**
- **Contributed to the Defence AI Strategy through the MoD's AI Ethics Panel.**

We aim to protect society through better governance and regulation of autonomous systems applied to the defence and security sector. With our emphasis on policy, through multiple projects, white papers and consultations, we helped shape a better governance of our data and considered compound risks. We supported research in ethical solutions for biometric AI in our law enforcement and a scalable platform for data in forensics.

eXplainable Human-Swarm System-Verification of Trust in Human-Swarm System

Dr Mohammad Soorati
Assistant Professor of AI at University of Southampton said:

“The xHS project developed and evaluated novel interfaces for small human teams to deploy and monitor large swarms of uncrewed aerial vehicles in highly dynamic and uncertain environments. We developed digital twins of such UAVs and their environment to challenge human operators with realistic situations.

We achieved international recognition including:

- **Best demo award at the International Conference on Autonomous Agents and Multiagent Systems (AAMAS) 2023.**
- **Three Conference Papers at the IEEE International Symposium on Robot and Human Book chapter published by Elsevier**
- **Workshop on Human-swarm Teaming**
- **Demonstration at AI UK 2023**
- **Demonstrating Performance Benefits of Human-Swarm Teaming at the International Conference on Agents and Multiagent Systems (AAMAS)**
- **2658 registered interactions on the simulation platform**
- **Workshop on Urban AI with International and Industry Partners**

Further funding: from the UK Defence Science and Technology Laboratory (£250K) to work on FAST-PI: Flexible Autonomy in Human-Swarm Teams Using Real-Time Physiological Information."

Health and wellbeing

Can we trust autonomous systems to look after our health and wellbeing?

By Prokar Dasgupta OBE, FKC Professor of Surgery, King's Health Partners

King's College London, King James IV Professor of Surgery, Royal College of Surgeons of Edinburgh, pictured left in a robotic operating room with the semi-autonomous DaVinci Xi system.

Healthcare projects have been an exciting and inspiring part of the TAS programme and have attracted considerable interest from the public, particularly after COVID-19. Robots are already being used to suture a bowel alongside human surgeons. Similarly, health apps claim to help improve mental and physical wellbeing. However, questions arise as to whether such technologies are sustainable and trustworthy given the complexities of human health and the safety critical and highly regulated nature of clinical practice.

TAS sought to investigate such issues through a range of projects, touching on many parts of human health and wellbeing.

1. **Mental health:**
   Safe Spaces NLP project explored the use of Natural Language Processing (NLP) for classifying behavioural harms within online forum posts (e.g. bullying; drugs & alcohol abuse; gendered harassment; self-harm), especially for young people *(featured on the next page).*

2. **Chronic Disease Management**
   Autonomous systems including AI and Machine Learning are expected to play a major role in supporting the management of chronic diseases. COTADS explored ways to involve patients in the co-design of algorithms for the management of Type 1 Diabetes (T1D). Just because such algorithms pass legal scrutiny, unless they are accepted by patients and clinicians they will not be adopted. The project demonstrated how interactive design tools and explainable AI could be used to engage people living with type 1 Diabetes in data-driven design processes to improve acceptability.

3. **Cancer management**
   The standard of care practice in the UK is to discuss cancer patients within multi-disciplinary teams. TAS has used oesophageal cancers as a case study to explore how data-driven decision support tools can improve treatment decisions. This is an evolving field and the TAS study is expected to lead to further validation in clinicals trials involving larger groups of cancer patients in multiple centres (*featured on the next page*).

4. **Robotics**
   TAS has funded many teams working in this field. DAISY, a robot assisted A&E triage system, has the potential to reduce waiting times in casualty. We found that assisted robotics can help chartered physiotherapists and empower social care professionals through digital technologies tested in a diverse group with lived experience. A cross-cutting approach here has helped to understand the complexity of robot-human interactions, such that some populations with disability are not marginalised.

We even have robotic devices which have been developed and tested for autonomous wound healing. As 2.2 million adults in the UK are affected by this problem, such low-cost wearable devices could prove to be game changers.

Finally, as robotic surgery becomes more semi-autonomous, detailed structured interviews have shown that the attitudes of surgeons, their supporting teams and patients seem to be positive towards this digital future of surgery. TAS has shown that the focus of AI based systems needs to switch from purely technical to socio-technical. The Wisenbaum test looks at the social impact of such technologies today, perhaps it is as important as the Turing test. Society must be able to tell good autonomous systems from bad ones if we are to have widespread application underpinned by public trust.

Addressing *online harm* to support the Online Safety Bill

Dr Tayyaba Azim

Research Fellow at the University of Southampton:

"Our socio-technical AI explored both zero-shot and graph-based NLP algorithms for behaviour classification, using a cyclic socio-technical methodology. Using incremental human feedback for iterative learning and re-ranking, overcoming the limited training data issue and keeping a 'human in the loop'. This study underpinned evidence to the Draft Online Safety Bill being cited in the main report from parliamentary team and follow up workshops with academic researchers and practitioners from government and industry."

Decision-making in oesophageal cancer

Dr Ganesh Vigneswaran

Clinical Lecturer at the University of Southampton:

"Multidisciplinary clinical teams make critical treatment decisions in every oesophageal cancer patient's journey. Such decisions have a profound impact on survival and quality of life. AI offers the potential to standardise decisions reducing unnecessary treatments and shortening hospital stays, leading to improved patient outcomes and saving healthcare costs. One of the main lessons from the project is the importance of interdisciplinary collaboration, combining clinical experience, data management and AI to foster a comprehensive approach to problem-solving."

"Through our regular discussions and engagement with relevant public and patient groups we ensured our research was aligned with patient needs and societal expectations. A critical insight from the project was the role of age in decision-making, pointing to potential subconscious biases and the importance of utilising explainable autonomous systems to learn about biological processes. This is especially important in healthcare, where trust and understanding from both clinicians and patients are paramount. We expect our research to lead to clinical trials, where our decision support tools and practices would be validated for safety, effectiveness and acceptance."

Increasing trust in digital mental health interventions

Dr Emma Palmer-Cooper

Lecturer in Psychology at the University of Southampton:

"Online, digital technologies to improve intervention adherence have been explored. However, some people with severe mental health problems experience higher levels of suspicion.

"We wanted to know if this was similar to people with no history of mental health problems. We found that there is a general sense of mistrust around digital technology from all members of the population, not just those experiencing mental health problems.

"Our findings and expertise are being used by the Met Police Cyber Crime Unit Protect team to support the development of appropriate, tailored materials for victim advice and future crime prevention when experiencing a mental health problem."

Supporting autistic children

Professor Farshid Amirabdollahian
at the University of Hertfordshire said:

"The Kaspar robot (*pictured above*) has been used to work as an educational and social mediator with children with autism spectrum disorder. One aspect of its usage is to enable the robot to automatically generate causal explanations as a way to enrich the interaction scenarios for children and promote trust in the robot.

"We validated our explanations for user satisfaction in an empirical evaluation with general users and then brought Kaspar, furnished with causal explanation, to schools. The results revealed that children improved their visual perspective taking abilities more significantly when the robot provided causal explanations."

Creating the prototype of a robot-assisted A&E triage

Pictured right: Professor Calinescu, Principal investigator of the project demonstrating the DAISY prototype at the University of York's Institute for Safe Autonomy to Chloe Smith MP, Secretary of State for the Department of Science, Innovation and Technology.

Professor Calinescu at the University of York said: "We explored the benefits and limitations of the technology — which may be capable of making a positive difference to patient care and the working lives of medics — through the prototype of a robot-assisted A&E triage solution for reducing patient waiting time and doctor workload.

"Our project co-defined the requirements for this solution with medics (through a series of interviews with healthcare practitioners) and end-users (through a Patient and Public Involvement study) and shared our research internationally through conferences and scientific papers. We are now working towards extending the project to a potential clinical trial."

Picture caption: AgriTrust to test soil quality in challenging environments

Energy and environment

Sustainability-Centric AI: a crucial pillar for trustworthy autonomous systems

By Professor Enrico Gerding, Director of SustAI, the UKRI AI Centre for Doctoral Training in AI for Sustainability at the University of Southampton (pictured right)

In a world facing pressing environmental challenges, the synergy between sustainability and artificial intelligence (AI) is more crucial than ever. Autonomous systems have the potential to revolutionise our approach to climate change, offering solutions to improve transportation, optimise supply chains, reduce waste, and monitor biodiversity. Their integration into our daily lives is inevitable, making it essential to develop systems that enhance efficiency and contribute to the greater good of society. Existing governance measures for ensuring resilience in society are struggling to deal with compound risks such as extreme climate change events during pandemics.

AI for risk and resilience assessment:

Our ARGOS project investigated the applicability of AI methods to support anticipatory planning for climate change and resilience, simulation for policy appraisal in view of compound risks, and coordination mechanisms for resilience-aware decision making. ARGOS explored current use of AI methods and published the following findings:

Planning and risk assessment

50% of risk and/or resilience assessment studies use machine learning techniques, including random forests, artificial neural networks, and support vector machines.

Policy analysis

29% of studies apply AI methods to analyse various policies, measures and interventions related to climate adaptation. These studies use modelling, simulation or participatory approaches to evaluate and compare potential impacts of different policies.

Implementation of resilience policies

Only 6% of studies apply AI methods to support implementation of adaptation and resilience actions, mainly focusing on improving communication and information sharing.

Motivated by the insights and challenges identified through TAS's research, the UKRI Centre for Doctoral Training in AI for Sustainability (SustAI) has emerged as an innovative program at the University of Southampton. This program strategically facilitates real-world impact, fostering collaboration with leading research institutes, industry partners, and state-of-the-art facilities, including the Sustainability and Resilience Institute (SRI) and the Responsible AI (RAI) UK. SustAI transcends the traditional PhD program, offering students a unique opportunity to collaborate with over thirty business and government organisations at the forefront of sustainability. The programme encompasses a comprehensive training curriculum, workshops and cohort-building activities, ensuring that students are well equipped with AI techniques and motivated to drive meaningful change in businesses and society. By placing sustainability at its core, SustAI exemplifies a responsible approach to harnessing the capacities of trustworthy AI for the betterment of society.

Assessing the feasibility and accessibility of personal carbon allowances

Dr Gisela Reyes Cruz Transitional Assistant Professor at the University of Nottingham explained:

"This work responds to the recommendation by the Climate Change Committee to explore avenues for public engagement and demand carbon reduction, focusing on people's lifestyles. This project has sought to explore the viability, regulatory concerns, trustworthiness, and public acceptability of a Citizen Carbon Budget app; a speculative app for testing the assumptions of a policy option on personal carbon allowances. The idea entails that every person has a carbon budget that they can spend each month.

"Daily activities have associated carbon emissions that impact their budget. For example, the carbon emitted by travel will be different, based on the mode of transport and distance travelled - investigating the legal, regulatory, and technological aspects of smart farming will help make recommendation for use of AgriTech to deal with climate change and food insecurity."

Monitoring soil quality in challenging environments *(pictured on page 21)*

Shishir Nagaraja, Professor of Cyber Security at the University of Newcastle said: "We developed a soil-monitoring testbed that serves as an advanced research platform for AgriTech. This included engineering and installing 50 Internet of Things "Squirrel boxes" in a smart farms at the University of Newcastle to monitor soil quality. Two mobile platforms will be deployed on marine environment. Engineering a low-cost monitoring platform that can assess soil and water quality in harsh outdoor environments in the presence of physical and network adversaries, is a significant challenge. The AgriTrust testbed enables the development of resilient monitoring solutions which is key to fighting malicious environmental damage."

"We developed a soil-monitoring testbed that serves as an advanced research platform for AgriTech. This included engineering and installing 50 Internet of Things "Squirrel boxes" in a smart farms at the University of Newcastle to monitor soil quality. Two mobile platforms will be deployed on marine environment. Engineering a low-cost monitoring platform that can monitor soil and water quality in harsh outdoor environments in the presence of physical and network adversaries, is a significant challenge. The AgriTrust testbed enables development of resilient monitoring solutions which is key to fighting malicious environmental damage."

Helping to reduce whale strikes from shipping vessels

Dr Calum Corrie Imrie research Associate in Computer Science at the University of York said:

"Shipping is paramount, however, there is a crisis of whale strikes contributing to the decline of marine life populations. Training crew to spot whales and enact suitable collision mitigation protocol is often difficult and time-consuming. Deep Neural Networks (DNN) are being investigated for the detection of whales, however, DNNs are sensitive to gradual changes in input data, for example, those occurring through whale aggregation, as well containing uncertainty in general. Our feasibility study, PREVAIL, focused on the deployment of verified, explainable DNN, and what this would mean if integrated into an advisory system to reduce whale strikes.

"This study led to creating new connections with experts in DNNs and marine biology. The next step is employing the techniques developed within our group to this problem, including controller synthesis capturing the uncertainty, and validating with a simulator consistent with the modelling performed for whale-ship collisions."

Picture credit: IM@UCL lab and James Tye (Photographer)

Transport

Transporting us safely into the future

By Professor Henry Tse

Industry representative and Chair of the TAS Hub's Strategic Advisory Network

In the ever-evolving landscape of transportation, connected autonomous (road-based) vehicles is a sector poised to be a £650 billion market by 2035. Uncrewed air systems are gaining widespread adoption across commercial applications and smart port implementations. To achieve the benefits promised by such technologies without marginalising parts of society will require a collaborative effort from all key actors and those likely to be impacted (sometimes unintentionally) by such technologies. The TAS Programme has paved the way to support such efforts taking a very inclusive and responsible approach. I am grateful to have been able to support this effort in my role on the Strategic Advisory Network and as part of the Connected Places Catapult.

Situational awareness and trust during shift between autonomy levels in automated vehicles

Professor Bani Anvari, University College London said:

"Understanding the correlation between levels and 'on-demand changes' of Situational Awareness and trust between humans and Autonomous vehicles under failure conditions is essential to engendering the confidence of the public in the technology.

"We built a track record in the field of situational awareness and trust. In addition, the team gave an eight-week soft-robotics workshop at Elutec Academy, participated in the Bloomsbury Festival and the British Science Festival. Films were also shared on YouTube and included the research in an ITV West County documentary."

Professor Gary Burnett

Professor of Digital Creativity now at Loughborough University said:

"Our aim was to better understand how natural language interfaces could provide a more calibrated trust relationship between drivers and automated vehicles. We approached this in the context of near-future level 3 capable vehicles - where frequent handover of control tasks need to be carried out, looking at the effects of wording and gendered voices on the acceptability of voice assistants in future autonomous vehicles."

Trustworthy machines? Responsibility and self-driving cars

by Dr Carolyn Ten Holter
University of Oxford

The RAILS project gathered data from expert interviews and stakeholder workshops to unpick some preconceptions and implications surrounding 'responsibility' for autonomous vehicles. Several themes were evident, but some key points are directly relevant to 'responsibility' questions.

The myth of machine infallibility

The cultural idea that 'humans make mistakes, but machines do not' is a view not generally shared by designers and engineers, who expect that machines can and will fail. However, cultural expectations of infallibility may create significant gaps in understandings of responsibility if designers are working towards a 'viable' model but societal expectations are for a 'perfect' system. This is especially true for self-driving cars, particularly as they are frequently promoted as safer than human-driven vehicles. Against such an expectation of infallibility, the inevitable accidents will undermine long-term societal trust in the technology.

Cost and safety

Linked to the above is the connection between reduction of risk, and cost. Commercial vehicles must balance calculations of 'acceptable' safety against the expense of increased safety. An 'infallibility myth', where a self-driving vehicle must be safer than an equivalent human-driven vehicle, could drive prices so high that self-driving cars would be unaffordable for most people. This would damage one of the potential benefits of self-driving vehicles – increased accessibility for people in later life, disabled, and people who cannot drive themselves – if their target population cannot afford them.

Systemic trust

Understandings of fallibility are not necessarily fatal to levels of public trust. Transparent and trustworthy development systems can still receive high levels of acceptance alongside acknowledgement of failures. This model can be seen in the aviation industry – which receives a high level of public trust. Similar development in the self-driving car sector would require reducing present levels of hype – self-driving cars need to be perceived as products of a rigorous, robust, responsible system of development, with ongoing work on safety. They need to be boring.

Next steps

The RAILS team is continuing to analyse its data, which has also contributed to the project's technical side. The team is also investigating fundamental questions relating to how state-of-the-art autonomous driving models embed core responsible AI principles, such robustness, transparency, sustainability and data drift, among others. This led to the development of quantitative metrics -'responsible AI (RAI) indices' - that stakeholders could use as a responsibility lens for their models. To create evidence for this, the team is setting the Car Learning to Act (CARLA) responsible AI challenge, involving an RAI leaderboard for benchmarking new autonomous driving agents.

The safety and desirability criteria for AI-controlled aerial drones on construction sites

Dr David Bossens said:

“Currently, construction is based on manual labour, drones are employed only non-autonomously and we lack suitable AI methods to deal with tasks on construction sites. We also lack the legal framework.

We:

- **Created a larger interdisciplinary network of peers across institutions working in robotics, AI, sociology, and law, to discuss construction applications and societal implications**
- **Designed a suitable reinforcement learning technique that satisfies safety constraints and performs robustly in construction environments**
- **Constructed a new algorithm which is more scalable, has a smoother learning curve than previous versions in this category which provides a lot more flexibility in various settings and safety by strongly adhering to user-defined constraints, for example staying a safe distance away from humans**
- **Presented scenarios to participants, including monitoring, transport, and other examples of AI controlled drones on construction sites through a workshop report.**

We showed that:

- **Laws are very limited in terms of semi-autonomous systems privacy/general data protection**
- **There are no laws for drones with cameras, unless specifically setting out to spy on neighbours**
- **Most uses would require approval from the civil aviation authority**
- **Laws are spread around different documents so it would be helpful to create a national legal framework**
- **Using drone transport can reduce manual labour and work-related accidents**
- **Using drones to detect security breaches can reduce the labour of security staff.”**

Picture: a drone on a construction site

Professor Sarah Sharples, Chief Scientific Advisor to the Department for Transport, **pictured right, explains:**

“With the Automated Vehicles Bill announced in the Kings Speech in Autumn 2023, the UK’s position as a leader in application of AI and automation technology in transport remains. The work of TAS has provided critical opportunities for the exploration of how to implement technology responsibly. The Transport AI strategy, due for publication in Spring 2024, which provides a framework for industry application of AI in the transport sector has been directly informed by the work of TAS and its members.”

Governance and Regulation

Science-based advice to inform autonomous systems regulation

Dr Jennifer Williams
Assistant professor in Electronics & Computer Science at the University of Southampton (pictured right)

Autonomous systems will proliferate in more or less visible forms over the coming years, it is not enough for stakeholders and the public to be passive consumers of these technologies. It is important to anticipate the potential impacts of such technologies in various applications and prepare regulators, businesses, government, and the public. Research programmes like TAS help bring together the evidence required to inform public policy, business pratices and the public.

As part of this, TAS organised AI Regulation Workshops to explore ideas and distil the most important themes. We have seen that it has the greatest impact when research is performed outside of our proverbial disciplinary silos.

Be it a policy brief based on data-driven research or include co-designing a new autonomous system that meets the needs of Government or Industry. A common problem is best examined from multiple angles by bringing together all the disciplines and interests. We created a space for people to come together, express disagreements and state contrarian viewpoints, ultimately moving the projects forward constructively.

We also funded research projects to explore some aspects of trust and governance in more detail.

Findings from the TAS AI Regulation Workshops

- UK Regulators can benefit from the TAS community developing solutions to their biggest problems
- Case studies and stakeholder communication are equally important for public understanding and acceptance of AI technology
- Cross-sector regulation highlights regulators have different needs, for example AI on products versus services or markets
- AI regulation is developing differentially across sectors
- The term "safety-critical systems" means different things to different regulators in terms of risk of harm.

Photo: Regulators workshop, credit Adam Gassonam

Dr Jennifer Williams
Assistant professor in Electronics & Computer Science, University of Southampton:

"In my research, I explore the theme of safe and trustworthy audio capture for speech technology. This is because we're surrounded by such devices, from smart watches and virtual assistants to voice-enabled vehicles. Speech contains a variety of sensitive information, including personal identity attributes. These traits can be misused, highlighting the need for increased data protection, privacy, and security. Speech technology can also be positively life changing for people with sight, hearing or speaking impairments. So there are a range of viewpoints on trust that need to be explored. My work helps to identify and mitigate risks like voice identity misuse by raising public awareness of these technologies and looking for ways to in-build safety solutions."

Contributing to government policy

Dr Justyna Lisinska, Policy Fellow at the Kings College London Policy Institute said:

“The UK government wanted to hear views on how to make it easier for actors at different stages of the AI life cycle to adhere to the UK government’s proposed cross sectoral AI regulation principles. For that reason, we held a Policy Lab in London with industry, government officials, and academics. This helped to inform the government’s white paper on AI.”

Title	Description
Trusted Internet of Things: at home and in the workplace	As we enter the fourth industrial revolution, the Internet of Things – and how people adopt it – transforms how we live, communicate, and conduct businesses. This is likely to bring issues where government, industry and researchers need to respond. In this policy landscape review, we explore the main policy issues the government should focus on and provide areas for further research.
Trusted Autonomous Systems in Vehicles: a policy landscape review	The UK government has taken a proactive approach towards encouraging the development of autonomous vehicles. Despite the political appetite for autonomous vehicles as a means to solve the UK logistics challenges, development has stalled in recent months. This review looks at the regulatory, ethical, and technological hurdles we need to overcome to address the challenges we’re facing.
Trusted Autonomous Systems in Defence: a policy landscape review	When it comes to autonomy in defence systems, there are huge opportunities to improve efficiency and save costs, but significant issues around trust, ethics and legal challenges remain. We unpack these issues in this report with The Policy Institute at Kings College.
Trusted Autonomous Systems in Healthcare: a policy landscape review	The UK government and the NHS have asserted clear ambitions around the application of artificial intelligence to healthcare. This review collects issues of relevance to policymakers in developing and deploying trustworthy autonomous systems.

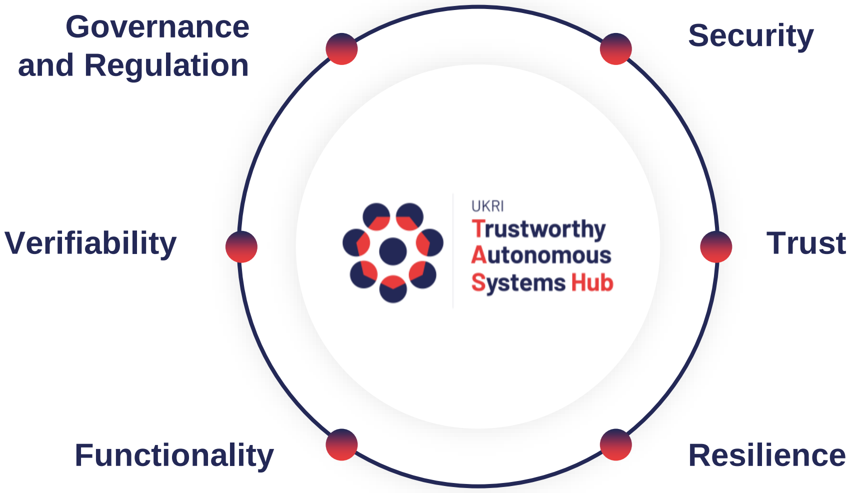
Delivering trustworthy electoral oversight

Kate Dommett, Professor of Digital Politics at the University of Sheffield said:

"Our aim is to understand how electoral regulators and management bodies can benefit from using automated systems to better analyse and administer elections. It is widely accepted that democratic elections should be both free and fair and that citizens believe that elections exhibit these traits. It is also essential that the institutions in charge of managing elections are trusted in the delivery of these goals. Building on our analysis of existing electoral transparency systems, we have used a collaboration between politician scientists and computer scientists to show how electoral transparency and oversight could be materially improved. Working closely with the Electoral Commission and the other regulatory agencies within the UK government we are demonstrating how automation can deliver tangible benefits for government oversight and be honed to promote public trust."

Introducing the six TAS Nodes

The six TAS nodes explored aspects of the design, operation, and governance of autonomous systems in the following areas:



The next six pages summarise their key recommendations.

Picture: A TAS Hub and Node away day

Regulating the Domestication of Autonomous Systems

By Professor Subramanian Ramamoorthy

Principle Investigator of the TAS Governance and Regulation Node at the University of Edinburgh

When we first encounter an autonomous system, it can seem wild, unpredictable, in some instances damaging and potentially hard to control. The TAS Node on Governance and Regulation is developing a framework, tools, and processes to enable the domestication of autonomous systems into everyday use.

Our framework starts by asking the question: What behaviours of an autonomous system should we take responsibility for?

A system is domesticated to a particular use when it is well enough understood that an identified group of people are prepared to be answerable for how it behaves.

A key part of Governance and Regulation of Autonomous systems is specifying the roles and resources that people need to fulfil them. Domestication involves many overlapping processes and varies considerably in scope and scale.

In studies of autonomous vehicles, we see the need for high levels of safety assurance combined with the requirement to draw on a wide diversity of knowledge sources e.g. legal, driving experts and a range of engineering disciplines. We have developed techniques to formalise legal knowledge, incorporate it into reasoning tools and combine this with simulations of naturalistic driving conditions. These are translated from natural language descriptions to close knowledge gaps to enable new roles who can answer for the behaviour of vehicles.

Our approach has been to study real-world cases in depth, run workshops involving regulators, system vendors, users and academics, to identify regulator, vendor and user needs. Based on this, we have evolved and experimented with tools and processes that enable the definition and eventual professionalisation of roles so they can be incorporated into better governance and regulation.

Working out what it takes to enable a group of people to be answerable for the behavior of an autonomous system is a productive starting point for regulation.

Building role often means spanning knowledge gaps so people in a role can be confident in their ability to answer for the behaviour of systems. Tools to identify the relevance of data to a decision and to support reasoning about Autonomous Systems are key outputs.

Autonomous systems are often adopted by large and complex organisations that use many of these systems. Making this possible involves the creation of platforms that package many roles.

Building roles often means spanning knowledge gaps so people in a role can be confident in their ability to answer for the behaviour of systems. Tools to identify the relevance of data to a decision and to support reasoning about Autonomous Systems are key outputs.

Visit:
web.inf.ed.ac.uk/tas

Adaptation opens up opportunities and challenges

By Dr Shane Windsor

Principal investigator of the TAS Functionality Node at the University of Bristol

The ability to adapt offers the potential for autonomous systems to move from conducting well defined tasks in predictable situations, to undertaking complex tasks in changing real-world environments. The TAS Node on Functionality, based at the University of Bristol, has been exploring how creating autonomous systems with the ability to adapt, or evolve their functionality, changes how we specify, design, verify, validate, trust, and regulate these systems.

We have been working closely with our project partners to understand the broader contexts under which autonomous systems need to operate and to develop frameworks for adaptive autonomous systems. Highlights include collaborating with Ocado Technology to develop trustworthy soft grippers, using benchmarking and equipment provided by Ocado Technology at the Bristol Robotics Laboratory (BRL).

At BRL we have also been developing a swarm arena testbed in collaboration with Toshiba Bristol Research and Innovation Laboratory to test swarm technologies. In addition, we have been working in partnership with data scientists from LV=General Insurance to understand ethical considerations and regulations for machine learning algorithms in the real world.

Three main challenges which we have addressed are:

Adaptive systems require a different balance of verification, validation, and monitoring techniques from conventional non-adaptive systems.

The non-explicability and changing behaviour of adaptive systems require appropriate risk mitigation strategies to enable their use.

Different methods of adaptation, such as machine learning, emergent behaviors in swarm robotics, and material adaptation in soft robotics, all offer different challenges for trustworthy design and use. Different approaches can be needed in different use cases.

Visit: tasfunctionality.bristol.ac.uk

Sociotechnical resilience is key to autonomous system trustworthiness

By Professor Calinescu

Principle Investigator of the TAS Resilience Node and the University of York

Autonomous systems are being developed at a staggering pace for use in health and assistive care, transportation, environment protection, and many other domains of great societal importance. These systems are uniquely capable of performing tasks and making decisions in changing environments, and can operate without human intervention for extended periods of time. As such, they have the potential to undertake complex missions that are dangerous, difficult or tedious for human operators.

We:

- **Hosted the UKRI Trustworthy Autonomous Systems workshop in health and social care**
- **Co-organised international conferences on Software Engineering and Formal Methods, Computer Safety, Reliability and Security and Software Architecture Software Engineering for Adaptive and Self-Managing Systems**
- **Edited journal special issues (Springer's Computing Journal, JSSand Elsevier's journal of Systems and Software) on the safe development, deployment and use of autonomous, AI and self-adaptive systems**
- **Contributed to the World Health Organisation's 'regulatory considerations on AI for health'**
- **Pursued applied research on the use of autonomous technologies for the triage of A&E patients, to support young disabled people**
- **We engaged, for forest protection, in discussions and collaborations with industrial, governmental, third sector and academic partners spanning over 50 organizations from the UK and the rest of the world.**

Insights and advances generated by these activities are summarised below:

Sociotechnical resilience is key to autonomous system trustworthiness

To achieve their potential, autonomous systems must continue to provide the required functionality despite the uncertainty and disruptions present in their real-world operating environments. Essentially, they must do this without violating their stakeholders' social, legal, ethical, empathetic and cultural (SLEEC) norms.

Operationalising SLEEC norms for autonomous systems is extremely challenging

SLEEC norms are often framed as high-level principles. Addressing normative concerns in concrete applications of autonomous systems requires the refinement of these normative principles into explicitly formulated SLEEC rules. We devised an end-to-end process for the elicitation and validation of SLEEC rulesets, and for verifying the compliance of autonomous systems with a set of such rules. Our tools automate key stages of this process, allowing the participation of non-technical stakeholders (lawyers, ethicists, sociologists, regulators, etc.) in the co-creation of SLEEC-norm aware autonomous systems.

Autonomous systems encounter regular instances of uncertainty and disruption

Prototype AI-enabled autonomous systems we have been developing for A&E patient triage, assistive care tasks including dressing and meal preparation, and forest health management show that these solutions must cope with frequent and significant uncertainties and disruptions. We devised theory and tools for quantifying and reducing uncertainty, and for resisting (through robustness and fault tolerance), anticipating (through proactive adaptation) and absorbing (through graceful degradation) disruption.

Visit: resilience.tas.ac.uk

Exploring the relationship between opportunity and trust

By Professor Thusha Rajendran

Principal Investigator of the TAS Trust Node at Heriot-Watt University

The TAS Node on Trust has worked with third sector organisations and industry including Age UK (Health and Wellbeing) and Lloyd's Register (Transport). This diversity of stakeholders has helped inform us about the similarities and differences in organisation needs.

The commonality between both organisations is the centrality of the 'human-in-the-loop'. For Age UK it is about the added value that an autonomous system can bring to the lives of older people in managing issues such as informed consent, mistrust and over trust. For organisations such as Lloyd's Register, special importance is placed on having humans take executive control, and allowing a person to have ultimate responsibility.

Our interdisciplinary Team has enabled us to investigate the human as well as the technical factors in understanding the trust between human and robots.

Two human factors that influence trust are:

Prior expectations	Individual differences in propensity to trust

Our prior expectations of what a system can do is a type of bias (which we have systematically manipulated in our experiments). Humans vary in their propensity to trust in the same way in which we are all introverted or extraverted to different degrees. In our Node we are creating a 'gold standard' measure of propensity to trust to share with the community. This will allow researchers to model how much variability in trust is due to individual differences and how it is due to a particular system.

Another key pillar of our strategy is to engage with the public through events and media work. For example, our TV and print media work have allowed us to discuss the increasing roles of autonomous systems in our lives and discuss the various roles that robots will have.

Our Science Festival events have allowed the general public to interact with our robots and ask questions including the ethics of the Wizard of OZ vs Autonomous Systems, what the different types of AI are and how they are impacting our lives. Our major learnings have been the value of user and public engagement and placing interdisciplinary approach as integrated rather than separate entities of a research.

Visit: trust.tas.ac.uk

Should we worry about the security of autonomous systems?

By Professor Neeraj Suri

Principal Investigator of the TAS Security Node and Chair in Cybersecurity at the University of Lancaster

Mobile autonomous systems (AS) such as self-driving vehicles or unmanned drones are becoming commonplace across diverse commercial and military applications.

They increasingly use AI technologies to operate without human oversight and intervention, adapting to the environment to conduct the missions. AS obtain situational awareness by collating sensor inputs to subsequently plan/execute a requisite course of action. Each AS technological block can be compromised. Communications can be disrupted, GPS/sensor inputs spoofed, and AI processes distorted. So how do we ensure that despite security breaches, the AS "adequately" completes the mission and does not result in damage to the system or collateral damage to the environment?

The TAS Node on Security has consulted extensively with academics (Lancaster, Cranfield, CMU-US, AcademiaSinica-Taiwan, RISE-Sweden), practitioners (National Highways, Airbus, BAE Systems, Boeing, Leonardo, TTTech), and policy makers (ISO standards bodies) across socio-technical dimensions to obtain AS mission requirements across a variety of applications. We have explored AS scenarios from a multi-dimensional/disciplinary perspective. We consider the impact of attacks over the specification and design of “safe and secure” AS behaviors (Theme: “Securing AS-Usage”).

We consider the impact of attacks on core AS operations of communication, navigation, routing, control, and AI-based decision making (Theme: “Securing AS-Operations”). Finally, AS often operate in populated urban spaces where a compromised AS can cause severe collateral damage (Theme: “Securing AS-Users”). In case of compromises, how is a response ethically determined which could potentially result in damage to the vehicle, the environment, or to people? How do we accomplish predictable and explainable AI decision-making while taking human-machine interactions into account? How do we certify AS and how is the liability of their actions assessed?

We have explored varied socio-technical solutions with several technical and policy recommendations to AS researchers, practitioners, and standard bodies.

Some key findings being:

Specifications rule!

As we cannot anticipate all attack scenarios or the complexity of attacks, it is crucial to unambiguously specify the “normal” AS operations as well as the boundaries of “degraded” behaviour in case of compromises.

Need for trustworthy AI

AI technologies are inherently fragile where any compromises affecting the accuracy of the data used to train, interpret, and decide a course of action can result in undesirable consequences. AI processes are not fully repeatable and result in variations in their outcomes over different runs. These issues make it exceptionally hard to predict or reproduce their outcomes.

Humans are a first-class consideration in the AS world where their interplay with technology remains mostly uncharted for behavioral and ethical aspects.

Visit: tas-security.lancs.ac.uk

Verifiability for autonomous systems through a holistic and inclusive community

Professor Mohammad Mousavi
Principal Investigator of the TAS Verifiability Node at King's College London

The challenge addressed by the UKRI TAS Node on Verifiability is a holistic and inclusive framework for verification of autonomous systems that allows stakeholders across many different domains to express and verify their concerns and requirements. We have produced more than 60 academic papers, three policy papers, and a number of accessible expositions of the Node activities.

To complement the technical and scientific activities in the Node, we organised and participated in a number of activities, created a large interdisciplinary community, and provided input to both academic, policy, and public engagements. A selection of some of these impactful activities is given below:

We organised 55 instances of the Verifiability Talk series to date, open to the public. Most talks are featured on the Verifiability YouTube Channel, with more than 120 subscribers and more than 3,000 views. We co-authored the report led by WMG Warwick on Cross-Domain Safety Assurance for Automated Transport Systems and participated in the panel discussion at the report launch event.

We delivered a keynote and participated in the ITF Roundtable on Policy and Regulation for Autonomous Vehicles held at the OECD in Paris.

We hosted a delegation of about 60 PhD students and their supervisors from the Wallenberg AI, Autonomous Systems and Software Program (WASP) in Sweden. We also organised a number of mutual visits for faculty members and PhD students within WASP.

Our top messages are:

1

Verifiability of autonomous systems is a multi-faceted concern, requiring involvement from a wide range of disciplines.

2

The UKRI TAS Verifiability Node has created a multi-disciplinary community of researchers, practitioners, and users to address this multi-faceted concern.

3

There is imminent and increasing need for sharing best practices and contributing to policies in verifiability of autonomous systems, to which the TAS Verifiability Node has responded.

verifiability.org

Responsible Research and Innovation (RRI) is embedded in everything we do

By Elvira Perez Vallejos
Professor of Digital Technology for Mental Health, University of Nottingham and Chair of the EDI Committee

TAS has encouraged researchers and innovators to consider the bigger picture, including the long-term impacts of their work and its value to society. RRI highlights important concerns that can often be overlooked or undervalued.

We have used a specialised deck of "high level" RRI Prompts and Practice Cards to support the exploration of Responsible Research and Innovation in research and development projects. They are intended to be lightweight and quick to use. They are based on the AREA framework and can be used to discuss RRI issues and identify strengths, weaknesses and areas for further action.

Several other decks have been developed:

1. Ethics in Research cards
2. Equality, Diversity and Inclusion cards
3. Policy cards

Addressing inequities

We are committed to increasing equality in the accessibility of systems, research engagement including underserved communities, diverse data representation, digital literacy/skills, inclusive design, cultural diversity, and socio-economic opportunities.

Data imbalance is one area which is particularly important in relation to AI. At TAS we recognise and address the inequalities emerging from the current lack of data diversity. We are also concerned about the harmful design decisions arising from the lack of involvement and consideration of people who are affected by or are at the receiving end of the systems.

The AI landscape must resolve a key issue: data is not representative and often contributes to biased outputs. When the data feeding Large Language Models (LLMs), for example, is not diverse and only represent a small proportion of the global population, the outputs from such generative AI cannot be diverse, representative, or generalisable due to its narrow intake, often rooted in discriminatory practices and beliefs.

Efforts to prioritise data equity are very important for TAS and to do that, the TAS community is encouraged to engage and champion the promotion of equality, diversity, and inclusion (EDI) values.

Those researchers and partners funded by TAS must indicate their EDI plans and commitment to both equality and equity, by engaging with diverse groups of stakeholders, ideally social systems and demographics that are traditionally excluded from data collection. For example, ethnic and gender minorities are too often missed from those data points that will feed AI and autonomous systems. At TAS, we look to engage as widely as possible and have worked with young people with disabilities, older adults and people with different health conditions that bring their unique lived experiences and expertise.

Investing in skills at every level

Next generation skills development

Dame Wendy Hall, Regius Professor of Computer Science at the University of Southampton said:

"The UK's National AI Strategy outlines the government's plans to harness the power of AI to increase resilience, productivity, growth and innovation across the private and public sector.

"Achieving this aim requires a highly skilled, diverse and multi-disciplinary workforce able to conceive, design, build and operate Trustworthy Autonomous Systems.

"Success in inspiring and investing in future generations from school children to professors will be one of TAS' lasting achievements. We have also created five policy papers in this area and much more."

Dr Eike Schneiders,

Transitional Assistant Professor at the University of Nottingham

"Joining TAS has given ample opportunities to network, collaborate and develop my own multidisciplinary research projects. I have been encouraged to engage in projects beyond my comfort zone, working with colleagues from the UK and abroad in Computer Science, Psychology, Law, and Medicine, among others. These projects enabled me to explore novel research questions and adopt methodologies that were unfamiliar to me. This significantly contributed to my growth as a researcher, broadening my understanding of various topics in Human-Computer and Robot Interaction. Following this, I was promoted from Research Fellow to Transitional Assistant Professor and with guidance from senior staff, continue to evolve as a researcher."

Professor Steven Meers

Fellow at the Defence Science and Technology Laboratory & co-chair of the TAS Skills Committee

"Investment in skills is the single most important area to drive the adoption of Trustworthy Autonomous Systems in the UK. Safe and responsible AI systems require highly skilled and experienced people to design and operate them. Investment in skills is the single most important area to drive the adoption of Trustworthy Autonomous Systems in the UK. The TAS Skills Committee is addressing this through a number of activities including:

The Doctoral Training Network

Has connected doctoral students at institutions across the UK and has provided a range of training activities, including seminars, workshops, summer schools and student conferences.

.

The Syllabus Lab

Has helped better identify the skills requirements, reviewed international best practice and provided policy recommendations such as "Closing the AI skills gap in the UK". This was done with the Department of Digital, Culture, Media and Sport and is continuing with Responsible AI UK.

The Early Career Researcher Awards

Celebrates outstanding contributions made by PhD students and postdoctoral researchers to any area of Trustworthy Autonomous Systems research."

TAS worked with the Institute of Engineering and Technology IET Faraday® Challenge

07	226	13-15	23
Events	Students	Ages	Schools

Locations included:

Derby, Lanarkshire, London, Scotland, Somerset, Portsmouth and Yorkshire.

94% *Agreed that they learnt new things.*

62% *Are now considering studying or working in engineering.*

93% *Would like to do something like this again.*

Some of the participants said:

"This was awesome, I loved doing this, 10/10 would do this again!"
"It was a really fun day and I have developed skills that I can use at school and at work when I get a job."
"I understand more about what engineering is."
"I have a better idea about what engineers do and the skills they need."

Professor Andy Wright FREng.

Chair of the TAS Hub board and Professional fellow at the University of Southampton said:

"TAS has focused on looking at the crucial issues of trust, certification in autonomous systems in AI. This has been not just from a technical perspective but also considering the wider societal, legal and regulatory aspects we need to understand to ensure take up of these systems. The programme's inclusive approach has covered potential applications for autonomy across a diverse range of sectors from cultural and creative industries to health and wellbeing as well as governance and regulation.

"Over the last four years, with the support of UKRI, TAS has sown the seeds of research and innovation which will allow the UK to grow its capability in areas we will depend upon in future. It has also cultivated an exciting body of work and developed interdisciplinary skills, with international renown, for future practitioners to harvest. We trust readers will make use of this exciting opportunity to collaborate with the TAS community and take forward this important work."

The TAS Executive Management Team are:

Sarvapali (Gopal) Ramchurn
(chair)
Professor of Artificial Intelligence,
University of Southampton

Kate Devlin
Reader in Social and Cultural Artificial Intelligence, Kings College London

Joel Fischer
Professor of Human-Computer Interaction,
University of Nottingham

Michael Boniface
Professional Fellow of Information Technology, University of Southampton

Derek McAuley
Professor of Digital Economy,
University of Nottingham

Elvira Perez Vallejos
Professor of Mental Health,
University of Nottingham

Carmine Ventre
Chair in Computational Finance,
King's College London

Dame Wendy Hall
Regius Professor of Computer Science,
University of Southampton

TAS Board

- **Professor Andy Wright,** Professor, University of Southampton (Chair)
- **Derek McAuley**, Professor of Digital Economy, University of Nottingham
- **Marion Oswald**, Vice Chancellor's Senior Fellow in Law, University of Northumbria
- **Sarvapali (Gopal) Ramchurn**, Professor, University of Southampton
- **Tom Rodden**, Chief Scientific Advisor for the Department for Digital Culture, Media and Sport
- **Cameron Ross**, UKRI TAS Programme Manager
- **Sarah Sharples**, Chief Scientific Advisor for the Department of Transport
- **Jack Stilgoe**, Professor, University of London
- **Carmine Ventre**, Professor, King's College London
- **Adrian Woolard**, Head of Northlab and Connected Studio
- **Previous chairs: Carly Kind and Alvin Wilby**

TAS Research Fellows

- **Michael Akintunde,** Research Fellow, King's College London
- **Tayyaba Azim,** Research Fellow, University of Southampton
- **Pepita Barnard,** Research Fellow, University of Nottingham
- **David Bossens,** now Senior Scientist at Centre for Frontier AI Research, A*STAR Singapore
- **Jediah Clark,** Research Fellow, now at NATS
- **Jeremie Clos,** Research Fellow, University of Nottingham
- **Liz Dowthwaite,** Research Fellow, University of Nottingham
- **Joshua Krook,** now Researcher at University of Antwerp
- **Justyna Lisinska,** Research Fellow, King's College London
- **Peta Masters,** Research Fellow, King's College London
- **Elnaz Shafipour,** Research Fellow, University of Southampton
- **Mohammad Soorati,** now Assistant Professor at University of Southampton
- **Dominic Price,** Research Fellow, University of Nottingham
- **Sachini Weerardhana,** Research Fellow, King's College London
- **Mohammad Naiseh,** now Lecturer at Bournemouth University
- **Jennifer Williams**, now Assistant Professor at University of Southampton
- Victoria Young now at the Ministry of Defence

The TAS Operations Team are:

- **Angela Westley/ Florence Gallien,** Head of Operations, University of Southampton
- **Laura Armstrong,** Projects Manager, University of Southampton
- **Ben Coomber,** Research Development Manager, University of Nottingham
- **Josie Gray,** Transformation Manager, University of Southampton
- **Stacha Hicks,** Social Media Coordinator and Project Administrator, King's College London
- **Claire Wilkins/ Alison Tebbutt**, Skills Managers, University of Southampton
- **Past Members: David Maffin, Helen Shaw and Lou Male**

UK Research and Innovation

UKRI Trustworthy Autonomous Systems Hub

Find out more at
www.tas.ac.uk

@tas_hub

linkedin.com/company/tas-hub

Living With AI Podcast
tas.ac.uk

University of Southampton

King's College London

University of Nottingham
UK | CHINA | MALAYSIA

Enabling the development of socially beneficial autonomous systems that are both trustworthy in principle and trusted in practice by the public, government, and industry.