UKRI Trustworthy Autonomous Systems Hub

UK Research and Innovation

UKRI Trustworthy Autonomous Systems Hub
ANNUAL REPORT 2023

TAS AT A GLANCE

The TAS Programme includes one hub, 6 nodes, 4 responsibility projects, 41 Pump Priming Projects.
The TAS Hub also initiated 23 Agile/Integrator research projects with more to be announced.

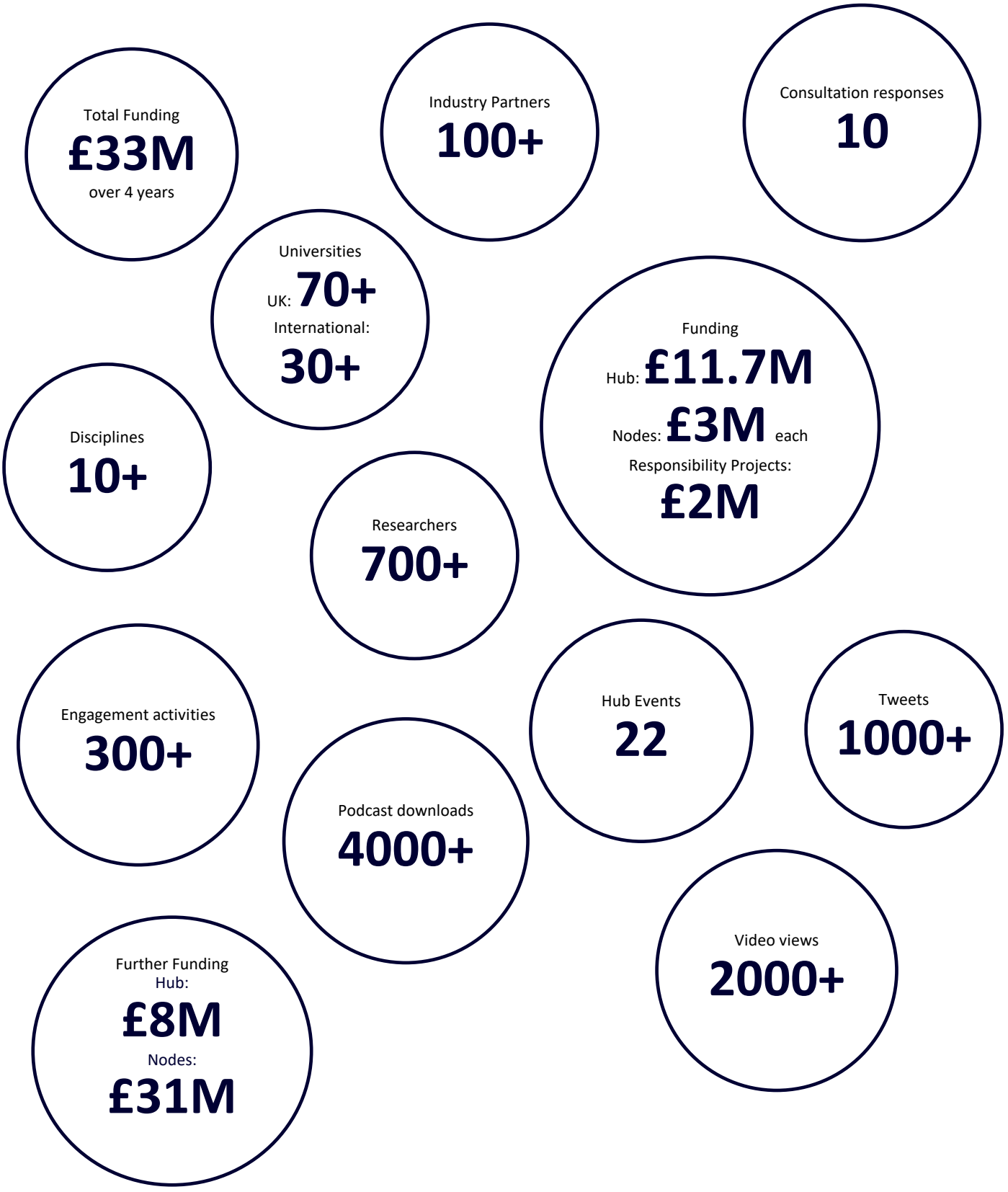


Table of contents

DIRECTOR'S NOTE

Over the last year, AI has become the focal point of governments, industries, and publics. ChatGPT is now the subject of conversations at dinners and used by children at schools. The excitement around the benefits AI technologies will bring to societies around the world are dampened by the issues they raise around copyright, devaluation of human ability, and AI-powered mis and disinformation at scale. We, as a research community, play an important role in the debate around the trade-offs AI is pushing us to make, the regulation and standards we should adopt, and the sustainability of such technologies among other things.

Generative AI technologies have been around for a long time but very few saw their potential in being used to retrieve information, summarise text, generate pictures based on text prompts, and write whole essays in a particular style. The rapid adoption of GPT by the public and industry shows that we are trusting the technology. What is it that led us to trust it so quickly? Is it due to its capabilities? The correctness or style of its answers? Its score at the Bar exam? Or is it because saved us time and money? Some would argue it is because the big technology companies are behind it and these are trusted brands. Clearly such questions can only be answered through a multi-disciplinary lens. Is such trust misplaced however? Do we really think generative AI will be beneficial in the long term?

The Trustworthy Autonomous Systems Programme has been running for three years now and has brought together a multi-disciplinary community that is well versed in Responsible Research and Innovation (RRI). As you will see in this annual report, we have been studying such questions of trust and trustworthiness across a range of projects in different sectors and where different forms of AI and autonomous systems are currently or will be deployed in the future. It is now the time for our community to respond to the growing range of questions around AI and autonomous systems.

As you will see in this report, the impacts of the TAS programme are now beginning to show in terms of our influence on policy and industry practices. With the initiation of the TAS symposium series, we are aiming to shape the legacy of the programme through the community that has been created. As we enter the final year of the programme, it is now time to translate our research results into pragmatic outcomes for society and the economy. I look forward to working with the TAS community and beyond to ensure we achieve the vision we set for the programme: to deliver best practices for the design, regulation, and operation of socially beneficial autonomous systems which are both trustworthy in principle, and trusted in practice by individuals, society, and government

Gopal Ramchurm FIET
Director, TAS Hub

RESEARCH AND POLICY IMPACT

The TAS Research Programme has three core programmes that target different audiences: the Agile Programme for TAS Hub partner universities to co-create new research projects that are responsive to current societal challenges around TAS; the Integrator Programme for the wider TAS Network (including the Nodes, Pump priming and Responsibility projects) to co-create projects that synthesise and integrate cost-cutting TAS research; and the Pump Priming Programme, that is targeted at expanding the TAS Network via a competitive research call that is open to all eligible UK research organisations.

All TAS projects have demonstrated a commitment to Responsible Research and Innovation (RRI) and Equality, Diversity, and Inclusion (EDI) as part of their research plan. We have built in the need to reflect on such underpinning values within the proposal writing and assessment process.

The TAS Hub also aims to ensure a diversity of disciplines is involved in all funded projects, as well as a diversity of project areas and representations from across the UK. Over 60 disciplines are represented in the reviewer pool for our Pump Priming Programme and projects. We also ensure that priority is given to funding early career researchers (ECRs) as project leads.

In 2023/24 we will be funding an additional 8 pump priming projects and further agile/integrator projects, to add to the 56 research projects, involving over 20 UK research institutions and more than 100 partners, that have already been supported. These projects research TAS across several core sectors (verticals – see below) of societal and economic importance, including Transportation, Health and Wellbeing, Defence and Security, Workplaces, Commerce and Finance. Analysis of the projects has given rise to five cross-cutting research areas (horizontals) across three sectors: TAS Techniques and Mechanisms, Public Engagement with Autonomous Systems, Human-Machine Interaction, Responsible Innovation Processes, and TAS Governance.

Research highlights

TAS for Autonomous Vehicles (AV)

This topic has been a popular choice for TAS research projects. There are six projects addressing challenges including what constitutes ethical and responsible use of AV data, how can you better support a driver's awareness when shifting between different levels of autonomy, how do software issues affect the driver's trust, and what should and anthropomorphic interface supporting the driver in an AV look and sound like to be trustworthy and inclusive? Thus, the work on this topic crosses all five horizontal research areas (TAS Techniques & Mechanisms; Public Engagement with AS; Human-Machine Interaction; Responsible Innovation Processes, and TAS Governance).

For example, led by Lars Kunze and colleagues at the University of Oxford, the RAILS project addresses questions of "long-term" autonomy in open-ended environments and lifelong learning systems. RAILS builds on the RoAD Pump Priming project on responsibility in the context of Autonomous Vehicle (AV) data. This strand of work helps understanding of public engagement and perception of how future Autonomous Vehicles should be designed, regulated, operated concerning important details such as what kind of data the vehicles should capture, and who should be permitted to have access to that data, and at what level of detail?

Associated projects

- RoAD—Responsible AV Data Ethical, legal, and societal Ethical, legal, and societal challenges of using data from autonomous vehicles. Main contact: Lars Kunze, University of Oxford
- Situational Awareness and trust during Shift between Autonomy levels in automated Vehicles, Main contact: UCL
- Understanding user trust after software malfunctions and cyber intrusions of digital displays: A use case of automated automotive systems. Main contact: William Payre, Assistant Professor in Transport Design & Human Factors, University of Coventry.
- The Chatty Car: Designing an Anthropomorphic, Natural Language User-Interface for Facilitating Appropriate Trust in Automated Vehicles. Main contact: Gary Burnett, University of Nottingham.
- Inclusive autonomous vehicles: the role of human risks perception and trust narratives. Main contact: Paurav Shukla, University of Southampton.
- Communicating liability in Autonomous Vehicles. Main contact: Elena Nichele, University of Nottingham.

Related Publications

- Ten Holter, C., Kunze, L., Pattinson, J. A., Salvini, P., & Jirotka, M. (2022). Responsible innovation; responsible data. A case study in autonomous driving. Journal of Responsible Technology, 11, 100038.
- Omeiza, D., Webb, H., Jirotka, M., & Kunze, L. (2021). Explanations in autonomous driving: A survey. IEEE Transactions on Intelligent Transportation Systems.
- Omeiza, D., Anjomshoae, S., Webb, H., Jirotka, M., & Kunze, L. (2022). From Spoken Thoughts to Automated Driving Commentary: Predicting and Explaining Intelligent Vehicles' Actions. arXiv preprint arXiv:2204.09109.
- Jestin, I., Fischer, J., Galvez Trigo, M. J., Large, D., & Burnett, G. (2022, July). Effects of Wording and Gendered Voices on Acceptability of Voice Assistants in Future Autonomous Vehicles. In Proceedings of the 4th Conference on Conversational User Interfaces (pp. 1-11).
- Yazdanpanah, V., Gerding, E. H., & Stein, S. (2021, September). Formal Methods to Verify and Ensure Self-coordination Abilities in the Internet of Vehicles. In International Conference on Computational Logistics (pp. 410-425). Springer, Cham.
- Rigas, E. S., Gerding, E. H., Stein, S., Ramchurn, S. D., & Bassiliades, N. (2022). Mechanism Design for Efficient Offline and Online Allocation of Electric Vehicles to Charging Stations. Energies, 15(5), 1660.
- Worrawichaipat, P., Gerding, E., Kaparias, I., & Ramchurn, S. (2021). Resilient Intersection Management With Multi-Vehicle Collision Avoidance. Frontiers in Sustainable Cities, 3, 670454.
- Zavvos, E., Gerding, E. H., Yazdanpanah, V., Maple, C., & Stein, S. (2021). Privacy and Trust in the Internet of Vehicles. IEEE Transactions on Intelligent Transportation Systems.

TAS for Health and Wellbeing

One of the key domains in which autonomous systems can benefit human beings is unsurprisingly a popular topic among TAS researchers. Overall, 11 projects are associated with this domain. Again, the research addresses challenging topics from across the five core TAS research areas (TAS Techniques & Mechanisms; Public Engagement with AS; Human-Machine Interaction; Responsible Innovation Processes, and TAS Governance). Researchers are addressing difficult questions such as, how can we make use of robotics to support health and care tasks (such as dressing, triage, and visitations) in a safe, dignified, resilient, and trusted manner? How can autonomous systems support mental health work, such as by facilitating the detection of harmful speech online for prevention purposes?

Researchers also asked what can be learnt from the large-scale deployment of digital contact tracing during the pandemic to safeguard public health in future pandemics.

In the summer of 2020, a team at the University of Nottingham started examining the public's perception and adoption of digital contact tracing during the Covid-19 pandemic in the UK. The research highlights a range of issues that have impacted trust and participation in digital contact tracing. The insights from the research, which may be drawn upon to improve the uptake and participation in future digital contact tracing have been published in Plos One and the Journal of Medical Internet Research and garnered nearly 20 citations in the 12 months since publication.

Associated projects

- DAISY: Diagnostic AI System for Robot-Assisted A&E Triage
- Imagining Robotic Care: Identifying conflict and confluence in stakeholder imaginaries of autonomous care systems
- A participatory approach to the ethical assurance of digital mental healthcare.
- COdesigning Trustworthy Autonomous Diabetes Systems.
- Trustworthy light-based robotic devices for autonomous wound healing.
- SafeSpacesNLP: Behaviour classification NLP in a socio-technical AI setting for online harmful behaviours for children and young people.
- Trustworthy Autonomous Systems to support Healthcare Experiences (TAS for Health).
- Understanding Internet and Technology Delusions of Suspicion, impact on engagement with cognitive training.
- Digital Twins for Human - Assistive Robot Teams.
- Verifiably Human-Centric Robot Assisted Dressing.
- Intersectional Approaches to Design and Deployment of Trustworthy Autonomous Systems.
- REFORMIST: Mirrored decision support fRamEwork FOR Multidiciplinary Teams in Oesophageal cancer
- Mapping trustworthy systems for RAS in social care (MAP-RAS)

Related Publications

- Dowthwaite, L., Fischer, J., Vallejos, E. P., Portillo, V., Nichele, E., Goulden, M., & McAuley, D. (2021). Public adoption of and trust in the NHS COVID-19 contact tracing app in the United Kingdom: quantitative online survey study. Journal of medical Internet research, 23(9).
- Pepper, C., Reyes-Cruz, G., Pena, A. R., Dowthwaite, L., Babbage, C. M., Wagner, H., ... & Fischer, J.E. (2022). Understanding Trust and Changes in Use After a Year With the NHS COVID-19 Contact Tracing App in the United Kingdom: Longitudinal Mixed Methods Study. Journal of Medical Internet Research, 24(10), e40558.
- Dowthwaite, L., Wagner, H. G., Babbage, C. M., Fischer, J. E., Barnard, P., Nichele, E., ... & McAuley,D. (2022). The relationship between trust and attitudes towards the COVID-19 digital contact- tracing app in the UK. Plos one, 17(10), e0276661.
- Nichele, E., Lavorgna, A., & Middleton, S. E. (2022). Identifying key challenges and needs in digital mental health moderation practices supporting users exhibiting risk behaviours to develop responsible AI tools: the case study of Kooth. SN Social Sciences, 2(10), 1-19.
- Burr, C., & Leslie, D. (2022). Ethical Assurance: A practical approach to the responsible design, development, and deployment of data-driven technologies. AI and Ethics, 1-26.
- Duckworth, C. J., Guy, M. J., Kumaran, A., O'Kane, A., Ayobi, A., Chapman, A., & Boniface, M. (2022). Explainable machine learning for real-time hypoglycaemia and hyperglycaemia prediction and personalised control recommendations. medRxiv.
- Rubio Denniss, A., Gorochowski, T. E., & Hauert, S. (2022). An Open Platform for High-Resolution Light-Based Control of Microscopic Collectives. Advanced Intelligent Systems, 2200009.
- De Saille, S. (2022). Improving Inclusivity in Robotics Design: An experiment in upstream co- creation. Frontiers in Robotics and AI, 119.

Defence & Security

A further key research domain for TAS, this topic has attracted 11 research projects. While again addressing many topics that cut across the five core TAS research areas (TAS Techniques & Mechanisms; Public Engagement with AS; Human-Machine Interaction; Responsible Innovation Processes, and TAS Governance), the research in this domain frequently raises challenges of effective and efficient human- machine interaction, such as for agriculture, search & rescue, or environments that are hazardous for humans.

For example, the research projects Trustworthy Human- Swarm Partnerships in Extreme Environments, and the follow up project XHS: eXplainable Human-swarm Systems, led by Mohammad Soorati at the University of Southampton address questions of situation awareness and control, and technical approaches to coordination of swarms for disaster response and search and rescue (Parnell et al., 2022). The latter project is a collaboration with John Hopkins University in the US, a strategic partner on the latest programme through which TAS Hub is seeking to develop international collaborations .

Associated projects

- Virtually There: Exploring Presence, Ethics and Aesthetics in Immersive Semi-Autonomous Teleoperation for Hazardous Environments
- AgriTrust: Trust Assurance in Autonomous Cyber-Physical Agriculture Farms of Future. Lincoln.
- Trust Me? (I'm an Autonomous Machine). Lancaster.
- Consent Verification in Autonomous Systems
- Trustworthy Human-Swarm Partnerships in Extreme Environments
- Verifiably Safe and Trusted Human-AI Systems (VESTAS)
- Privacy-Preserving Detection of Online Misinformation. Nottingham.
- Leap of Faith. KCL.
- Safety and desirability criteria for AI-controlled aerial drones on construction sites. Southampton.
- Preserving Marine Life in a Shipping World: AI to the Rescue (PREVAIL)
- XHS: eXplainable Human-swarm Systems.
- CHAPTER: Cognitive HumAn in the looP Teleoperations.
- Critically Exploring Biometric AI Futures.

Related publications

- Parnell, K. J., Fischer, J. E., Clark, J. R., Bodenmann, A., Galvez Trigo, M. J., Brito, M. P., ... & Ramchurn, S. D. (2022). Trustworthy UAV relationships: Applying the Schema Action World taxonomy to UAVs and UAV swarm operations. International Journal of Human–Computer Interaction, 1-17.
- Divband Soorati, M., Clark, J., Ghofrani, J., Tarapore, D., & Ramchurn, S. D. (2021). Designing a User-Centered Interaction Interface for Human–Swarm Teaming. Drones, 5(4), 131.
- Divband Soorati, M., Gerding, E. H., Marchioni, E., Naumov, P., Norman, T. J., Ramchurn, S. D., ... & Zhang, J. (2022). From intelligent agents to trustworthy human-centred multiagent systems. AI Communications, (Preprint), 1-15.
- Singh, Lokesh, Ramchurn, Sarvapali, Malik, Obaid and Clark, Jediah R (2022) Understanding the Impact of Induced Stress on Team Coordination Strategy in Multi-User Environments. In, Goonetilleke, Ravindra S. and Xiong, Shuping (eds.) Physical Ergonomics and Human Factors. 13th AHFE International Conference on Physical Ergonomics and Human Factors (24/07/22 - 28/07/22) p. 1. (doi:10.54941/ahfe1002608).
- Bossens, D. M., Ramchurn, S., & Tarapore, D. (2022). Resilient robot teams: a review integrating decentralised control, change-detection, and learning. arXiv preprint arXiv:2204.10063.
- Viganò, L. (2022). Formal Methods for Socio-technical Security:(Formal and Automated Analysis of Security Ceremonies). In Coordination Models and Languages (COORDINATION 2022).

- Snooks, K., Whitham, R., Richards, D., and Lindley, J. (2022) Beyond the body: Moving past the metricised bodily goal in self-tracking, in Lockton, D., Lenzi, S., Hekkert, P., Oak, A., Sádaba, J., Lloyd, P. (eds.), DRS2022: Bilbao, 25 June - 3 July, Bilbao, Spain. https://doi.org/10.21606/drs.2022.501
- Ramchurn, S. D., Stein, S., & Jennings, N. R. (2021). Trustworthy human-AI partnerships. Iscience, 24(8), 102891.
- Yazdanpanah, Vahid, Gerding, Enrico, Stein, Sebastian, Dastani, Mehdi, Jonker, Catholijn M. and Norman, Timothy (2021) Responsibility Research for Trustworthy Autonomous Systems. 20th International Conference on Autonomous Agents and Multiagent Systems, London (Virtual), Virtual, United Kingdom. 03 - 07 May 2021. pp. 57-62 . (doi:10.48448/w5y9-qk13).

Culture & Education

There are 9 projects pertaining to the domain of Culture & Education. Again, many of the projects cut across all five of the TAS Horizontals (Research areas Techniques & Mechanisms; Public Engagement with AS; Human-Machine Interaction; Responsible Innovation Processes, and TAS Governance). Particularly, the research area of public engagement is prominent in this domain, as are the associated enabling techniques & methods.

For instance, the TARICS project on Trustworthy Accessible Robots for Inclusive Cultural experiences, led by Marisé Galvez Trigo and colleagues at Lincoln University is working with Museum stakeholders to increase the accessibility of the space for disabled and neurodiverse people. The Kaspar Explains project led by Farshid Amirabdollahian and colleagues at the University of Hertfordshire is developing a robot that benefits the learning of children with autism.

The Open Laboratories Programme for Trustworthy Autonomous Systems (OPENTAS) led by Tony Prescott and colleagues at Sheffield has developed a robotic telepresence platform to allow the public to take part in open science event. Bridging these projects, the TAS ART: Augmented Robotic Telepresence Integrator project led by Ayse Kucukyilmaz and colleagues from Nottingham and King's College London seeks to integrate robotic telepresence across domains including offices, museums, healthcare, and education settings to increase accessibility and create novel experiences.

Associated Projects

- Reimagining TAS with Disabled Young People
- TARICS: Trustworthy Accessible Robots for Inclusive Cultural experiences
- Kaspar explains: the impact of explanation on human-robot trust using an educational platform
- An Open Laboratories Programme for Trustworthy Autonomous Systems (OPEN-TAS)
- Cat Royale
- Co-Design of Context-Aware Trustworthy Audio Capture
- Privacy Preserving Detection of Online Misinformation
- Trustworthy Autonomous Systems and Socio-Technical Innovation in Data Protection by Design and Default (DPbDD)
- TAS ART: Augmented Robotic Telepresence Integrator
- Embodied trust in TAS: robots, dance, different bodies

Related publications

- Boudouraki, A., Reeves, S., Fischer, J. E., & Rintel, S. (2022, April). Mediated Visits: Longitudinal Domestic Dwelling with Mobile Robotic Telepresence. In CHI Conference on Human Factors in Computing Systems (pp. 1-16).
- Naiseh, M., Bentley, C., & Ramchurn, S. D. (2022, March). Trustworthy Autonomous Systems (TAS): Engaging TAS experts in curriculum design. In 2022 IEEE Global Engineering Education Conference (EDUCON) (pp. 901-905). IEEE.
- Gou, M. S., Lakatos, G., Holthaus, P., Wood, L., Mousavi, M. R., Robins, B., & Amirabdollahian, F. (2022, August). Towards understanding causality–a retrospective study of using explanations ininteractions between a humanoid robot and autistic children. In 2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) (pp. 323-328). IEEE.
- Borsci, S., Malizia, A., Schmettow, M., Van Der Velde, F., Tariverdiyeva, G., Balaji, D., & Chamberlain, A. (2022). The Chatbot usability scale: The design and pilot of a usability scale for interaction with AI-based conversational agents. Personal and ubiquitous computing, 26(1), 95- 119.
- Del Duchetto, F., & Hanheide, M. (2022). Learning on the Job: Long-Term Behavioural Adaptation in Human-Robot Interactions. IEEE Robotics and Automation Letters.
- Huynh, T. D., Tsakalakis, N., Helal, A., Stalla-Bourdillon, S., & Moreau, L. (2022). Explainability-by- Design: A Methodology to Support Explanations in Decision-Making Systems. arXiv preprint arXiv:2206.06251.
- Brandao, M., Canal, G., Krivić, S., Luff, P., & Coles, A. (2021, August). How experts explain motion planner output: a preliminary user-study to inform the design of explainable planners. In 2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN) (pp. 299-306). IEEE.
- Emsley, I., & Chamberlain, A. (2021, June). Sounding out the System: Multidisciplinary Web Science Platforms for Creative Sonification. In 13th ACM Web Science Conference 2021 (pp. 50- 52).


METRA

Programme-wide collaboration

The TAS Hub is committed to building a coherent ecosystem across the UK, working alongside the TAS Nodes, Responsibility projects and other key research centres. TAS Nodes are invited to leverage their existing resources to join collaborative projects with the Hub.

Over the last year, 9 integrator projects have been funded, undertaking foundational research involving AS Hub and Node researchers. These projects are aligned with the following Nodes: Functionality, Governance and regulation, Resilience, Trust, and Verifiability. They cover a range of domains including Culture and Education, Commerce and Service, Health and Wellbeing, and Defence and Security.

Likewise, the Pump Priming Projects funded this year have involved a range of organisations, well beyond the three TAS Hub partners, including University College London, Newcastle University, University of Lincoln, University of York, and the University of the West of England.

International collaboration

TAS Hub began with several existing relationships from around the world, including Harvard University and the Max Planck Institute, and has since grown its network to span a growing number of countries and universities.

This was further boosted by the GRIP-funded activities that were reported in our Annual Report 2022, through which TAS reached out to specific groups in the US.

Strong collaborations have thus emerged with the Good Systems Group at the University of Texas in Austin and the Institute for Assured Autonomy (IAA) at John Hopkins University. This is being taken forward via two research collaborative platforms: the Agile Defence Programme and TAS Global.

Other collaboration are also emerging with related TAS centres such as the EU's TAILOR Network and Australia's TAS Defence Research Centre.

Agile Defence

The Agile Defence call, which is funded by our partners Dstl and DoD, was run as a pilot inviting submissions from TAS Hub researchers., John Hopkins University, and the Good Systems Group. The call utilised the Hub's research funding platform and processes. The intention was to identify and fund creative, inter-, and multi-disciplinary research which focuses on the challenging question of how to ensure that the design, engineering, and operation of autonomous systems generate positive outcomes and mitigate potentially harmful outcomes for people, societies, economies, and the environment.

Partners

University of Southampton

University of Nottingham UK | CHINA | MALAYSIA

King's College London

Johns Hopkins Institute for Assured Autonomy

Good Systems

U.S. Department of Defense

dstl The Science Inside

The three successful proposals seek to address factors that impact the trustworthiness of autonomous systems, such as methods of assuring effective human-machine teaming, and of understanding how effective such teaming will be under challenging conditions.

In addition to collaborating with the wider programme, these teams will also be engaging with stakeholders and users of the research, who are essential to the design, conduct and impact of application-orientated research.

Global Sandpit

In the UK, the sandpit process has been used over the last 15 years to bring creative thinking into the definition of new interdisciplinary research agendas, develop specific new research projects and build new networks between researchers. The Global Sandpit provided this opportunity to early career researchers, on a world stage and in a format we have never tried before, working with like-minded research groups from University of Texas at Austin and Johns Hopkins University's Institute for Assured Autonomy. We are very grateful to our US colleagues for investing time and effort in trialling out this mode of co-creation. Out of this exercise, we hope to establish a new benchmark for international collaboration and look forward to running similar, if not better, activities in the future with US and UK universities.

The Global Sandpit was an intensive process that kept over 25 early career researchers from across five universities, 5 mentors, 2 professional facilitators and the Sandpit Director busy over four days, at Airlie House in Northern Virginia. Besides the primary goal of defining some interesting joint projects participants were also inspired with new ideas, new perspectives, and made new friends.

The Global Sandpit deliberately brought together researchers with different views, disciplinary and cultural perspectives, who successfully worked as a group as participants not only respected, but celebrated and embraced, this diversity.

All participants received feedback on the proposals developed at the Global Sandpit, and two successful projects have commenced.

Policy Impact

The TAS Hub has directly engaged with policymakers to address emerging policy challenges via the Chief Scientific Advisors Network and using the methodology of Policy Labs, that focus on specific contentious policy questions, as well as responses to public consultations. Many of these responses were co-authored with colleagues across the TAS Network and other large research programmes. We ensure that responses expose opposing viewpoints on the policy questions posed.

The TAS Programme responded to the Governance of AI: A call for evidence from UK Parliament: Science and Technology Committee in November 2022. This, and other, consultation responses can be found on the policy impact pages of the TAS website.

The TAS Hub has also delivered a further policy landscape review on the Trusted Internet of Things (IoT), which explored the main policy issues the government should focus on and provided focus for further research, since our last annual report. This, like its predecessors, has been widely disseminated in academia and in government to engage with stakeholders.

The report from the Policy Lab on inclusive autonomous vehicles is feeding into policy at the Department for Transport (DfT). Dr George Economides, Head of AI and Autonomy at DfT is also actively engaged in the TAS Hub's Strategic Advisory Network.

Policy Lab on AI Regulation

This policy lab (a focused, collaborative workshop bringing together a wide variety of different stakeholders around the challenges of AI Regulation) funded by the Department for Media, Culture and Sport, was organised by the King's Policy Institute in March 2023.

The overarching question was 'How can the government's regulatory framework, governance tools and other measures support businesses across the AI lifecycle to optimise adherence to the proposed AI regulatory principles?'

The discussions are summarised in a concise briefing node for participants (following Chatham House rules) and inform the ongoing work of the UK's Office for AI policy teams.

TAS Regulators Workshop

In May 2023, we held a workshop in London that bridged a known divide between academic research and the priorities of UK Regulators. This workshop was attended by a variety of UK Regulators, including CMA, Ofcom, National Highways, DCMS, and HSE. In addition, we invited early career researchers and senior academics from multiple UK universities who are part of the TAS network as well as TAS-external guests from the UK and Europe. Each of the regulators presented a short talk that captured the challenges for AI regulation from their perspectives. From the challenges that the regulators highlighted, we entered into world-cafe discussions in three sessions: problematising, solutioneering at domain-specific levels, and solutioneering across domains. Our final session of the day brought together attendees to identify the cross-cutting issues where TAS could be helpful and we identified actionable next steps, including follow-on workshops that include industry stakeholders.

SKILLS AND CAPACITY BUILDING

Early Careers Researcher (ECR) Network

The TAS ECR Network continues to develop and now encompasses the Doctoral Training Network as well as those in the early stages of their academic careers. The ECR network self-organises events for the wider community such as the ECR Colloquium at the TAS '23 in July 2023, as well as connecting PhD students and post-doctoral researchers through informal group mentoring and discussion sessions, which provide the opportunity to:

- For experienced researchers: take a leadership role by chairing a session, mentoring and giving feedback.
- For more junior researchers: present work, interact and receive feedback from a wider audience.
- Share career experiences and offer/receive guidance
- Broaden networks of interdisciplinary researchers
- Find potential collaborators for future projects

Syllabus Lab

The Syllabus Lab aims to contribute to the development of future academic and professional training programmes and is led by a subgroup of the TAS Skills Committee, comprising researchers and industry partners. Members have analysed existing AI Skills Frameworks as a starting point for developing a TAS-specific Skills Framework, which will subsequently define interdisciplinary core topics for training courses and curricula. A series of workshops with a wide variety of stakeholders from Early Career Researchers through members of the Strategic Advisory Network of Industry Leaders, has influenced the development of TAS use cases and personas. These are also influencing work currently being undertaken with DCMS.

Trusted Autonomous Systems (online learning)

The course, which has been designed for professionals who want to gain an overview of TAS, will be run by the Australian National University's School of Engineering, in collaboration with the TAS Hub's Syllabus Lab. Participants will learn how to analyse, identify and communicate the needs and requirements of TAS, understand the gaps in existing standards and regulations around AS, examine the challenges of developing and implementing TAS, and develop strategies to identify and communicate with various stakeholder of TAS.

Equality, Diversity and Inclusion (EDI)

The TAS Hub is passionate about establishing and curating a programme that conveys social benefit to citizens from *all* walks of life and reaches out to *all* communities. To deliver our ambition, we have established a set of guiding principles that underpin our ways of working and decision making; we focus on openness, accessibility, equality of voice, and inclusivity..

Core to our inclusive vision for the TAS is the active integration of EDI principles in our research teams, in the research culture we promote, and in how we design research responsibly. We firmly believe that instilling EDI at the heart of what we do enables us to draw more widely on the breadth of UK talent, enhance research quality, and broaden avenues for impact.

Our EDI Framework aims to provide clear guidance that enables our leaders and teams to build EDI into all TAS activities, consistently and cohesively. It focuses on four key objectives :

- Promoting a happy, healthy research culture
- Creating diverse research teams
- Building EDI into research design
- Communicating our vision to the TAS Programme.

This year, supported by funds from Northrup Grumman, we have developed an online EDI self-assessment tool, based upon UKRI resources, and contributed to the development of an app to improve the accessibility of the EDI cards developed by the Connected Everything Programme at the University of Nottingham.

TAS Skills Day

This was held at the Connected Places Catapult in London on 17 March. The day offered an interactive Responsible Research and Innovation (RRI) session, led by Professor Perez, that explored anticipation and reflection and action/responsiveness practices using the 'RRI Prompt and Practice Cards' and relevant case studies as illustrations. There was also an industry (Thales) led session on safety assurance and certification of autonomous systems as well as a talk “From Mystical to Practical: The Path to Real World AI” by Sridhar Sundarsan, CTO of SparkCognition and a keynote by Professor Burkhard Schafer (University of Edinburgh) entitled “Law 3.0: code as law, law as code and the new modalities of regulation “.

> The TAS skills day was a brilliant experience for the Thales attendees, not only did we get to present our view on the certification of autonomous systems, we got to participate in the other sessions which gave us perspective on issues that we may face in industry. There was great engagement from the ECR’s with the interactive workshop, it generated some interesting discussions and highlighted just how complex the challenge of certification of AS is. The day also allowed us to make connections with other industry partners and in academia
>
> **Matthew Hodgkinson, UK Safety & Environmental Engineering Delivery Manager, Thales**

ADVOCACY AND ENGAGEMENT

Creative programme

One of the guiding aims of the TAS Hub is to support dialogues with a diverse set of stakeholders, including government, industry, creative, and the public. We do this via a number of routes: through academic conferences and industry summits, with public-facing events for all, by producing online resources and materials, and by means of our podcast.

Cat Royale

Working with our Cultural Ambassadors, Blast Theory, the TAS Hub is developing Cat Royale. Blast Theory make interactive art to explore social and political questions. Their work places the public at the centre of unusual and sometimes unsettling experiences, to create new perspectives and open up the possibility of change.

The artists Blast Theory and TAS researchers have designed, deployed, and are studying a high-profile interactive artwork called Cat Royale that engages the public with issues of trust in autonomous systems – especially in relation to the wider societal challenge of health and wellbeing – in a compelling, provocative, and yet safe way. The work involves the public in creating and running a paradise for cats in which AS technologies, including computer vision, robotics and IoT, monitor and tend to the wellbeing of a small community of cats that inhabit it.

Cat Royale is available as a single-screen video work, with conversation around the work being enabled via #catroyale. The work is being presented at the Science Gallery, London, Wales Millennium Centre and other UK venues in 2023. The parallel social media campaign engages the public in the design of the work, in supporting the AI in its tasks, and in debating the wider significance of the work, specifically its relationship to future human welling.

Partners and networking

All TAS projects involve at least one industrial, NGO or government agency partner, who are involved in advising on the direction of the research. Highlights of our engagements over the last year, are outlined below.

In terms of technology transfer, we have ensured that the outputs of all projects are open-source and therefore accessible to a wider community of users/industry than those just involved with the TAS Hub.

Many of the projects in train or completed have fed into technology deployed at partner organisations or have led to people being employed by these companies. Outcomes include Thales hiring two TAS researchers, Kooth Ltd internalising machine learning algorithms from the SafeSpaces project and University Hospitals reviewing its Trusted Data Environments for patient data sharing.

Further funding includes more than US$ 1m from Northrup Grumman, Thales and US DoD/Dstl for joint projects, creation of thought pieces and EDI work, and a further £1M following the co-development of successful joint bids to EPSRC (Engineering and Physical Sciences Research Council) and other funding bodies, with several partners from the TAS Network.

Thought Pieces

The TAS Hub and Nodes ran a series of workshops, sponsored by Thales, and facilitated by Dr Mark Pickering, around six themes – Functionality, Governance and Regulation, Resilience, Security, Trust, and Verifiability. Following the workshops, a series of thought pieces have been developed and published online in January. They are also available in hard copy, as a coherent set, with a foreword by the TAS Director.

Industry Day

The TAS Industry Day invited researchers from across the TAS programme (Hub, Nodes, Responsibility, Agile , Pump Priming and Global projects) to present their work and to develop new project ideas with industry ahead of the next Agile/Integrator Programme call. The keynote was given by Sridhar Sudarsan, CEO of SparkCognition

The Industry Day provided an opportunity for the TAS Programme to:

- update industry on what has been achieved and how partners may work with us
- reflect on the new challenges emerging with generative AI, AI regulation and the AI skills gap
- network and initiate new collaborations and routes to impact and
- plan for the legacy of TAS.

Other outreach

World Summit AI 2022, Amsterdam

Billed as the word's largest AI summit, the event gathered together global representatives from big tech, start ups, investors and enterprise as well as academia. Dr Kate Devlin, TAS Advocacy and Engagement Director, was in conversation with Professor Dame Wendy Hall, the TAS Skills Director, discussing governance and regulating the metaverse.

Masterclass in Swarm, Soft & Aerial Robotics

Organised by the Functionality Node, on behalf of the TAS Programme, the masterclass provided developers, operators, end users and researchers with an opportunity for hands-on experience with a range of autonomous and robotics technologies that can be developed in the logistics, manufacturing, and infrastructure industries.

Participants were able to work with researchers to try out and explore the capabilities of the latest technologies in swarm robotics, soft robotics, and aerial robotics (drones) via an interactive 'teaching and doing' format. There was also an opportunity to explore future avenues for the use of these technologies in the context of participants' own work, with the TAS Node sponsoring several follow-on summer student research projects.

AI Fest 23

A workshop, AI for Healthcare and Defence: Sharing AI Experience around Mental Health Abstract was delivered at this Dstl organised event. This workshop brought together researchers and practitioners working on and with AI for Mental Health in the Healthcare and Defence sectors to share experience gained and lessons learnt between these disciplines. It discussed the key challenges faced in these areas and attempted to signpost the current directions of travel within AI research to overcome these challenges in the short and medium term.

International Conference on Robotics and Automation (ICRA) 23 Workshop

The theme of the ICRA conference, held for the first time in the UK this year, was "Embracing the future: Making robots for humans". The ICRA workshop "Multidisciplinary approaches to co-creating Trustworthy Autonomous Systems (MACTAS)" was organised and delivered by the researchers from across the TAS Programme.

The MACTAS workshop brought together academics and industry practitioners from a wide range of disciplines and backgrounds (including robotics, engineering, AI. Computer science, social science. Humanities, design, and law), whilst being open and welcoming to researchers from the autonomous agents and multi-agent systems community.

Defining autonomous systems as systems involving software applications, machines, and people, which are able to take actions with little or not human supervision, the workshop explored different definitions of TAS and individual aspects of trust from a multidisciplinary perspective.

Good Systems Symposium "Shaping the Future of Ethical AI"

The Good Systems Symposium is the signature annual event of the UT Austin interdisciplinary research challenge to design ethical AI technologies for the benefit of society. It is free and open to the public.

Dr Kate Devlin, TAS Advocacy and Engagement Director, joined the International Perspectives Panel to discuss how global collaboration may fuel ethical collaboration.

IET Faraday Challenge Days

The Hub is supporting 7 IET Faraday Challenge Days in Bristol, Edinburgh, Lancaster, London, Nottingham, Portsmouth, and York, which in 2023 are based on a challenge provided by the UKRI's Future Flight Challenge. The TAS volunteers are supporting the delivery of these Faraday Challenge Days, a STEM competition for 12-13 years olds in the UK, and providing career mentoring to the students.


Education
Faraday
Challenge Days
The Institution of
Engineering and Technology
Find out more at

The first international symposium on Trustworthy Autonomous Systems - TAS ‘23

The first International Symposium on Trustworthy Autonomous Systems was held in Edinburgh on 11-12 July. With an early career researcher colloquium on 10 July.

TAS ‘23 showcased creative, multi- and interdisciplinary responsible research and innovation which focusses on the challenging question of how to ensure that the design, engineering and operation of autonomous systems generate positive outcomes and mitigate potentially harmful outcomes for people, societies, economies and the environment.

The proceedings, including full papers and abstracts, have been published in the ACM Digital Library.

The winners of the **Best Paper Award at TAS ‘23** were:

RE-centric Recommendations for the Development of Trustworthy(er) Autonomous Systems
Krishna Ronanki, Beatriz Cabrero Daniel, Jennifer Horkoff and Christian Berger

Honorable mentions:

An Exploration on how Trust Online Relates to Psychological and Subjective Wellbeing
Liz Dowthwaite, Elvira Perez Vallejos, Virginia Portillo, Menisha Patel, Jun Zhao, and Helen Creswick

The well-being of the Autonomous Vehicles (Avs) users under unceetain conditions
Mohammad Naiseh and Paurav Shukla

Resilient Strategies for Socially Compliant Autonnomous Assistive Dressing Robots
Katie Parnell, Siobhan Merriman, Simem Getir Yaman, Katherine Plant and Radu Calinescu

The winners of the **Best Poster Award at TAS ‘23** were

TAS Benchmarks Library and Critical Review
Peta Masters, Alan Chamberlain, Liz Dowthwaite, Yang Lu, Sachini Weerawardhana, Vitoria Young, Paul Luff, Peter McKenna and Luc Moreau.

TAS Early Career Researcher Awards

The TAS Early Career Researcher (ECR) Award scheme was launched to recognise significant contributions to TAS-related activities by PhD students and postdoctoral researchers across five impact categories. The winners were announced by Dr Horia Maior and Professor Elvira Perez Vallejos at the TAS'23 Symposium in Edinburgh.

The awardees are:
Academic Research:
Andriana Boudouraki, PhD student, Mixed Reality Lab, University of Nottingham
Knowledge Transfer:
Balint Gyevnar, PhD student, CDT in Natural Language Processing, University of Edinburgh (not pictured)
Policy:
Anastasia Kordoni, Postdoctoral Researcher, Dept of Psychology, University of Lancaster
Public Engagement:
Eike Schneiders, Postdoctoral Researcher, Mixed Reality Lab, University of Nottingham
Responsible Research and Innovation:
Helen Smith, Postdoctoral Researcher, TAS Functionality Node, University of Bristol

Each awardee will receive £4k to use towards their research and professional development.

All Hands Meeting (TAS AHM)

The TAS AHM was held in Edinburgh on 13 July, immediately following the TAS Symposium. The event brought the TAS Community together to showcase our research, forge new multi-disciplinary networks and see the scope of trustworthy autonomous systems on a global scale.

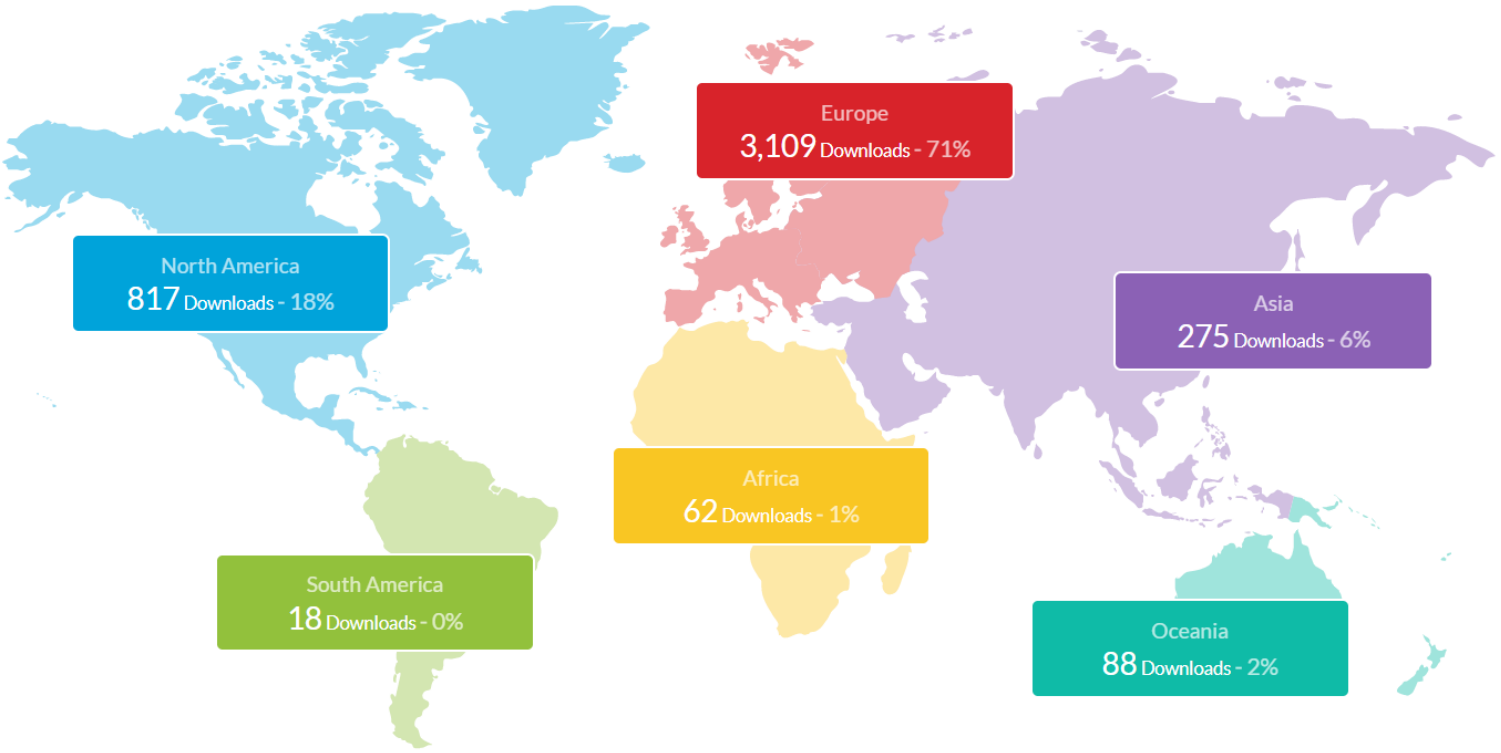
Attendees had the opportunity to interact with creative provocations, at The National Robotarium,, as well as engage with presentations and panels. The AHM also introduced Responsible AI UK, a new national initiative that aims to provide an international ecosystem for responsible AI research and innovation.

Social Media

In support of the Hub's ambition to create a 'one stop shop for all things TAS' our social media audience has increased significantly, with growing engagement across all channels – Twitter, LinkedIn, and the website. On average the Twitter account gains 50 new followers each month and traffic to the website has doubled in the last year. Referrals have shifted from predominantly direct to organic, indicating greater awareness of the TAS Hub beyond the reach of our communication channels.

Podcasts

The Living with AI podcasts, which feature a speaker and panel drawn from the TAS Community have covered such diverse topics as Track and Trace, virtual assistants, drones, driverless cars, music, robotic surgery, and legal and ethical issues of autonomous systems. All episodes are available on our website. By March 2023, the podcasts have had over 4,000 downloads from around the world. A third series showcasing the 22/23 research projects was released in May 2023. A further series of podcasts is currently being planned.



TAS Conversations

The TAS Conversations series brings together experts from a range of fields and from around the world to discuss current issues with autonomous systems, which are then made freely available via the TAS website. The most recent conversation was entitled "AI in the skies".

From commercial airlines to unmanned aerial vehicles - what are the trust challenges, what research is being done, and what are the regulations for the design and operation of AI and autonomous systems within the aviation industry? These and other questions were addressed in an hour-long, live panel discussion hosted by Sean Riley. The panelists were: Sharon Kindleysides, Chief Executive, The Chartered Institute for Logistics and Transport, John Jost, Autonomous Systems – Lead Safety Engineer, Boeing, Professor Jim Scanlon, University of Southampton, and Professor Henry Tse, Director of New Mobility Technologies, Connected Places Catapult.

LOOKING AHEAD

Translational Acceleration

The TAS Hub is currently conducting a consultation with individual TAS Hub Board members to establish a plan of action to deliver impact within the next 12 months. This includes the development of a skills framework, kickstarted through a short project funded by DCMS, and an RRI Training Framework.

Supporting the Trustworthy AI Ecosystem

The TAS Hub will look to be responsive to emerging challenges and the changing landscape of the UK economy and society. The upcoming AI Hubs and the new Responsible and Trustworthy AI Programme (RTAI) to be launched in 2023 will create more opportunities for TAS research and for TAS impact. Specifically, TAS Hub will look to give a head-start to the RTAI Programme by:

- Providing access to its network of academics, researchers, and industry through events, joint calls, and research projects.
- Providing its EDI framework and action plan to help structure the governance and operations of the RTAI Programme.
- Providing RRI training and RRI processes to support the foundational work of the RTAI Programme.
- Supporting the co-creation of specific research projects.
- Initiating an Innovation Tournament as part of the TAS Hub's Grand Challenge Programme.
- Growing the geographical reach of the TAS and the RTAI Programme through research collaborations and events across Wales and Northern Ireland in particular.

International Partnerships

Following the Global Sandpit in Washington DC and the Agile Defence Call, run in collaboration with Dstl and US DoD, the TAS Hub will further expand its international activities:

- Support for bids under the Lead Agency agreement or other US-UK funding calls building on established networks resulting from the Global Sandpit and Agile Defence projects.
- Explore collaborations with India/Australia/Japan TAS-related programmes. The TAS Hub Director has been selected to give the GREAT Talks by the British Council in India and will leverage this opportunity to broaden the impact of TAS.

UKRI Trustworthy Autonomous Systems Hub

UK Research and Innovation

UKRI Trustworthy Autonomous Systems (TAS) Hub

University of Southampton
University Road
Southampton
SO17 1BJ
www.tas.ac.uk
@tas_hub

Email: contact@tas.ac.uk