RAi UK ANNUAL REPORT

2025

UK Research and Innovation

We estimate that RAi UK has secured at least £3 million in 'follow on funding' from other research grants and activities.

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 across the ecosystem.

Professor Sarvapali (Gopal) Ramchurn

Contents

Responsible AI UK (RAi UK): A welcome

By Professor Sarvapali (Gopal) Ramchurn

CEO of RAi UK from the University of Southampton

As we reflect on 2024-25, we are immensely proud of the collective strides made in advancing responsible AI. This year has been defined by a concerted effort to translate our research vision into tangible, real-world impact, directly addressing the priorities and objectives we set out for each of our core pillars in the previous report.

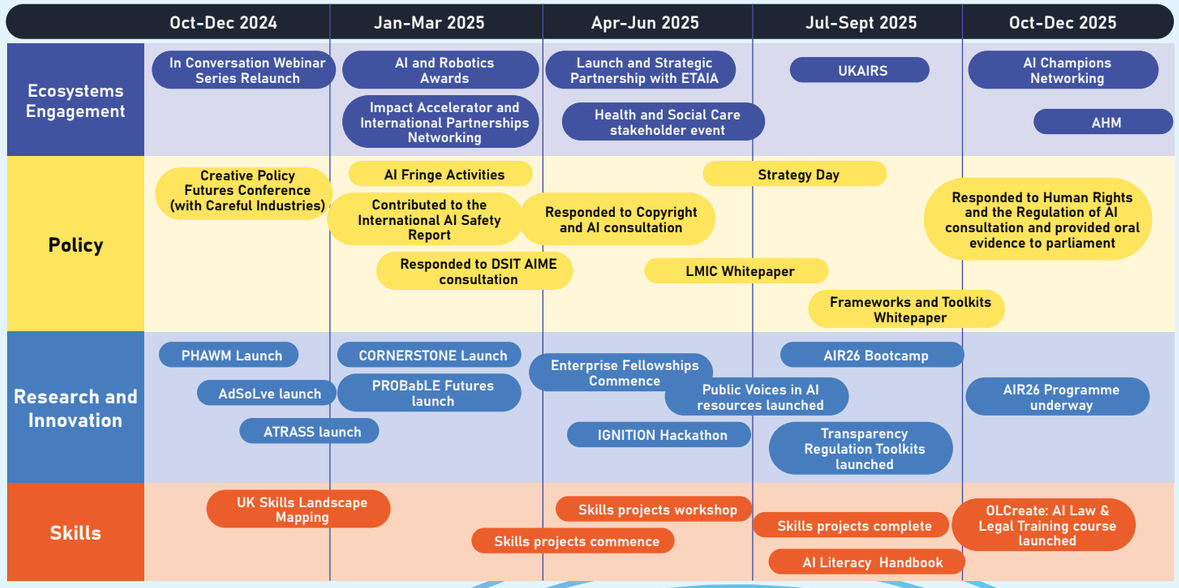
The landscape of Artificial Intelligence (AI) continues its rapid evolution, presenting both transformative potential, and complex challenges for the UK in terms of ethics, governance, and societal impact. RAi UK has resolutely positioned itself at the forefront of this dynamic environment, nurturing a vibrant ecosystem of interdisciplinary research, fostering crucial collaborations, and championing the development of AI that has genuine positive impacts for people and society. We are delighted to report significant progress in translating research into impact, delivering in our mission to support the adoption of responsible AI in the UK. Our deepened engagement with policymakers and industry has already resulted in more impactful and practical solutions to complex challenges, from developing pilots for trialling responsible AI adoption in Integrate Care Trusts to pilot in 2026, to establishing new avenues for inter-disciplinary engagement across the UK AI ecosystem through the first UK AI Research Symposium. Our funded "Research & Innovation projects have delivered groundbreaking insights and practical tools, particularly concerning advanced generative AI and critical Ai assurance methodologies. Our "Skills" initiatives have empowered and broadened access to essential AI literacy across age-groups and sectors, preparing society for an AI-integrated future. Our "Equities" work has placed inclusion at the heart of all we do addressing and mitigating bias in systems, in data, and in debate about the future of this transformative technology. RAI is committed to AI development for all. Finally, our global engagement efforts have solidified the UK's position as a world class thought leader and innovator, and a trusted, collaborative partner in the formation of a global responsible AI ecosystem, whether creating collaborative projects through sandpits in India; developing lasting research partners with US colleagues, or cementing the UK's leadership in Europe through membership of the European Trustworthy AI Framework.

This report showcases our significant achievements and the collaborative spirit that underpins our work. It is a testament to the dedication of our researchers, the strategic guidance of our leadership, and the invaluable contributions of our partners, and the strong foundations we are building for the future of AI in the UK and beyond.

As we look to 2026 and beyond, we will continue to build upon these foundations to accelerate our impact. We have reassessed our strategic approach to focus on greater present impacts while anticipating future threats and challenges. Our annual All Hands Meeting will focus on the responsible deployment of AI within critical national infrastructure. We will also expand our policy and public engagement programmes to foster greater understanding and trust in AI. These initiatives will further strengthen our mission: to ensure that the transformative power of AI is harnessed ethically and equitably for the benefit of all.

Our Story

This section visually charts Responsible AI UK's activities during 2025-26, illustrating the scope and momentum of our initiatives through key events and milestones.



Welcoming our new Chair of the Research Pillar

In July this year, RAi had the privilege of welcoming our new Chair of the Research Pillar, Professor Bashar Nuseibeh, who is Professor of Computing and Head of Software Engineering and Design at the Open University. Bashar has previously served as Chief Scientist of Lero – The Irish Software Research Centre – and Head of Software Engineering Laboratory at Imperial College London. He is an Honorary Professor at University College London (UCL), and a Visiting Professor at University College Dublin (UCD), Ireland, and the National Institute of Informatics (NII), Japan. Bashar's interdisciplinary research has focused on responsible software engineering and on the engineering of resilient socio-technical and cyber-physical systems.

Professor Bashar Nuseibeh

"I am excited to join the RAi UK leadership team as Chair of the Research Pillar. Research lies at the heart of the RAi programme, and in the dynamic world of AI, championing responsibility is not just an ethical duty—it's a powerful interdisciplinary challenge that spans the entire ecosystem.

My core vision is to establish RAi UK as the undisputed, trusted source of expertise in responsible AI. Drawing on my experience, I am committed to igniting a trans-disciplinary research approach that drives forward the highest possible quality of outputs and practice.

Looking ahead, we are set for strategic growth! We will capitalize on RAi UK's strong foundations in skills, policy, and public engagement, while boldly expanding our portfolio. This includes making critical contributions to dependable and trusted AI engineering and proactively tackling emerging frontiers like AI deployment and sustainability.

I am incredibly eager to connect with our community and partners, building on successful initiatives like the Collaboration Grants. This focused energy ensures every step we take is geared towards our ultimate goal: unlocking the immense potential of AI for decisive societal benefit!"

Summary of our impact and publications

The results below, captured via Researchfish in August 2025, quantify the significant real-world impact delivered by the RAi UK research community:

Shaping Policy – Nationally and Internationally

RAi UK actively shapes robust evidence-based AI policy both domestically and internationally, with responsibility and trust at its heart. We are dedicated to translating research into actionable guidance for governments and international bodies, ensuring that the latest insights inform real-world policy and regulation.

In 2024/25, our efforts were heavily focussed on UK national policymaking, accounting for 75% of our activities and solidifying our influence within the national discourse, reinforcing the foundations for the UK's global leadership.

Our comprehensive approach to embedding responsible AI principles was delivered by extensive engagement across key government departments (Department for Science, Innovation & Technology (DSIT), Foreign, Commonwealth & Development Office (FCDO) and Home Office); Parliament (House of Lords, House of Commons, select committees and APPGs); international organisations (UN, IEEE, UNICEF, WHO), and professional associations. We have responded to a number of UK government consultations on significant topics including copyright, human rights and regulation, and AI toolkits. RAi also contributed to the AI Action Summit International AI Safety Report (published in January 2025) alongside 100 international AI experts who collaborated to establish an internationally shared scientific understanding of risks from, and mitigations for, advanced AI.

This broad and dynamic approach has built the broad coalitions of support necessary to ensure that responsible AI principles are able to evolve from academic theory into concrete policy and professional practice. Policy impact workshop whitepapers

Throughout the year, RAi UK convened a series of collaborative policy workshops to tackle complex regulatory challenges through interdisciplinary thematic studies. We produced the first three White Papers in a series designed to translate ideas into impact.

Title	Key Research Impact
RESPONSIBLE AI for Low- middle-Income Countries (LMICs) (July 2025)	**Challenging Global Inequality:** This report, co-authored with partners in India, Indonesia, Africa, and South America, provided an evidence-based critique of the direct transfer of AI governance from high-income countries. It highlights how concentrated AI development exacerbates existing inequalities and fails to address the unique needs of LMICs. The paper issues a call to action for internationally inclusive policy.
Frameworks and Toolkits for Assuring Responsible AI (Oct 2025)	**Operationalising Accountability:** This report dives into the growing ecosystem of tools designed to make AI more transparent, accountable, and trustworthy. The report builds on the collaborative workshop co-hosted with Confiance.ai, earlier in April 2025, bringing together experts from across government, public services, and industry.
Advancing Trustworthy Artificial Intelligence: Lessons Learned and Emerging Challenges (joint publication with ETAIA)	**Establishing International Alignment:** Synthesising insights from major UK and European research programmes (including Confiance.AI and the UKRI Trustworthy Autonomous Systems (TAS) Programme), this paper established an aligned, cross-national perspective on operationalising trustworthy AI, building inclusive AI through co-design, and enhancing governance across the UK and EU

RAi UK Submission to policy consultations

RAi UK submitted a series of responses to policy consultations, to find out more and to access the following reports, please visit https://rai.ac.uk/media/reports/.

Report	Summary	Authors
Responsible AI Governance	Response to UN's consultation on Interim Report: Governing AI for Humanity. The report calls for a closer alignment between international norms and how AI is developed and rolled out.	Dr S Kiden, Prof. S Stein and Prof. S D Ramchurn
Perspectives on the AI Fringe	With the EU AI Act coming into force and the US Executive Order on AI in the US, industry is being challenged to develop measures to navigate an emerging regulatory environment for AI-based systems. The UK believes myriad benefits can exist by creating a regulatory framework for safe and responsible AI that remains pro-innovation at its core.	Written by a consortium drawn from the attendees of the AI Fringe event in November 2023, with Responsible Ai UK ensuring a diversity of views were collected from the academic community.
The International Scientific Report on the Safety of Advanced AI	Building a shared scientific understanding in a fast-moving field. It is the work of 100 international AI experts who collaborated in an unprecedented effort to establish an internationally shared scientific understanding of risks from advanced AI and methods for managing them.	CHAIR Prof. Yoshua Bengio RAi UK researchers were among the international experts who collaborated to establish an internationally shared scientific understanding of the risks from advanced AI and methods for managing them. The RAi UK's CEO, Professor Gopal Ramchurn, served as a Senior Adviser to the report, with the core focus on AI Sustainability.
RAi UK response to the DSIT Public Consultation on the AI Management Essentials Tool	Response to the Department for Science, Innovation and Technology (DSIT) consultation on the new AI Management Essentials (AIME) tool on behalf of Responsible AI UK (RAi UK), an open and multidisciplinary network that brings together researchers from across the four nations of the UK to understand how we should shape the development of AI to benefit people, communities and society.	Dr Isabela Parisio; Dr Sarah Kiden; Professor Gina Neff; Professor Sarvapali (Gopal) Ramchurn
RAi UK response to the Public Consultation on Copyright and Artificial Intelligence	Response to the Intellectual Property Office (IPO), the Department for Science, Innovation and Technology (DSIT) and the Department for Culture, Media and Sport (DCMS) joint consultation on Copyright and Artificial Intelligence on behalf of Responsible AI UK (RAi UK), an open and multidisciplinary network that brings together researchers from across the four nations of the UK to understand how we should shape the development of AI to benefit people, communities and society.	Prof. SD Ramchurn, Prof. G. Neff, Prof. K. Devlin, Dr. I. Parisio, Prof. M. Oswald, Dr. T. Lawal, Dr. F. Ryan, Prof. B. Stahl, Dr. D. Eke, Dr. P. Ochang, Dr. H. Cameron, Dr. M. Voigts and Dr. C.Brinck
PROBabLE Futures Submission in response to the Call for Evidence on 'Human Rights and the Regulation of AI' (2025)	Response to the Joint Committee on Human Rights' inquiry on 'Human Rights and the Regulation of Artificial Intelligence' published in July 2025. The submission has been prepared by a team of academic researchers with extensive expertise in the ethical, legal, and technical implications of AI use, directly relating to human rights. Our work is anchored in the UKRI Responsible AI (RAi) UK Keystone Project, "PROBabLE Futures"	Dr M Jorgensen, Dr A Paul and Professor M Oswald
RAi UK Submission in response to the Call for Evidence on 'Human Rights and the Regulation of AI' (2025)	This is a submission, on behalf of RAi UK and the Citizen-Centric AI Systems (CCAIS) project, in response to the Joint Committee on Human Rights' inquiry on 'Human Rights and the Regulation of Artificial Intelligence' published in July 2025. The submission addresses human rights issues, existing and possible changes to legal and regulatory frameworks. Oral evidence, based on this response, was presented by Dr Sarah Kiden, to the Human Rights (Joint Committee) on 29 October 2025.	Dr S Kiden, Dr V Yazdanpanah, Dr R Mestre, Dr Ingi Iusmen, Dr J Atkinson, Dr I Parisio, Prof S D Ramchurn and Prof S Stein

RAI International Ambassadors Council

RAi launched its "International Ambassadors Council" (IAC) in October 2025. The RAi UK Ambassador Council (IAC) is to be established as a strategic, internationally focused collective comprising prominent individuals whose professional missions and objectives align intrinsically with those of Responsible AI UK.

The primary function of the IAC is to cultivate a global network of world-leading stakeholders, drawing membership from a broad, multi-sectorial community that includes Academic and Research Institutions, Governmental bodies, Civil Society organisations, Policymakers, Regulators, Standards Bodies, and Industry. For the full list of Council members, please visit https://rai.ac.uk/team_category/ambassadors-council/

This structure is designed to leverage international expertise and provide a crucial platform for collaborative insight generation and the influencing of global challenges pertinent to responsible AI. The anticipated outcomes of the Council's activity include:

- Amplification of the RAi UK mandate on the international stage.
- The promotion and substantiation of international collaboration and knowledge exchange.
- Systematic sharing of best practice and lessons learned.
- The active fostering of mentorship within the responsible AI domain.

The inaugural meeting was chaired by Dame Wendy Hall and Professor Sana Khareghani. The sessions brought together global leaders from China, Australia, the USA, and India.

"The launch of the RAi UK International Ambassadors Council marks an important step in our mission to build a globally connected, inclusive, and accountable AI future. By bringing together leading voices from around the world, we aim to accelerate the exchange of knowledge and promote practices that make AI systems more transparent, fair, and beneficial to all."

Dame Wendy Hall

"I am so excited to be working alongside Dame Wendy Hall to bring together leading voices from across the globe on Responsible AI. There are many people working in this area, and we want to shine a light on the work, research and momentum of this community. Responsible AI is not about one thing, it is all encompassing, end-to-end and requires the thoughts and participation of every single one of us!"

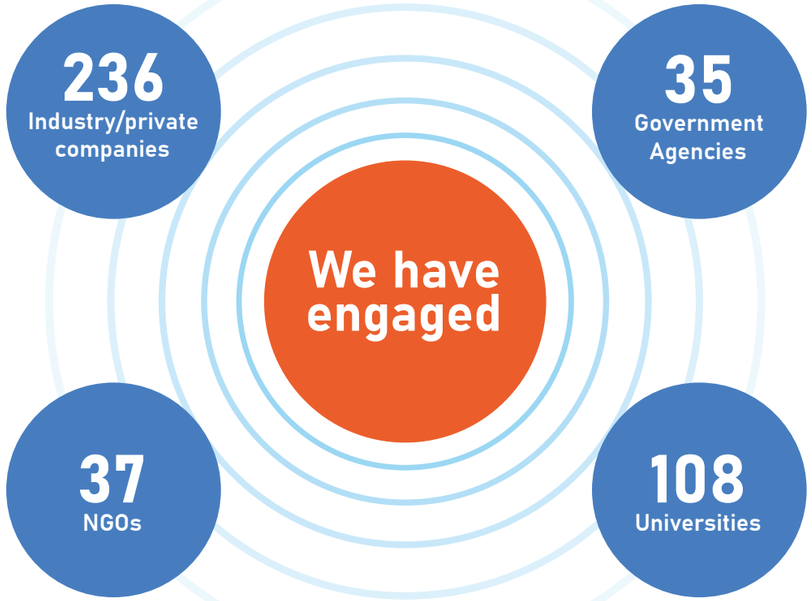
Professor Sana Khareghani

Key themes which emerged from the inaugural meetings were:

- Build an international RAI network, led by RAi UK, connecting global institutions and AI Safety Institutes.
- Focus on policy-to-practice: practical playbooks, standards, and toolkits that show how Responsible AI delivers real value.
- Identify distinct, apolitical niches where the IAC can make a visible impact.
- Promote interoperability of frameworks rather than uniformity, and learn to work together across diverse contexts.
- Prioritise joint testing, open data, and shared learning, underpinned by scientific collaboration.

Members will meet in-person in Delhi, India in February 2026 to continue to co-create the IAC's roadmap, establish a working group/task force for future whitepapers, and formalising its global agenda on Responsible AI.

Building Collaborative Communities



RAi UK – Notable Recognitions

- Prof Prokar Dasgupta was elected as an Honorary Fellow of the Royal Academy of Engineering (HonFREng), recognising his outstanding contributions to engineering and his work exploring public perceptions concerning robotic surgery (2025).
- Prof Tom Rodden was awarded a CBE in the New Year Honours for services to Science, Technology and Academia (2025).
- Prof Marion Oswald, MBE was awarded an Honorary Fellowship by the Royal Statistical Society (2025).
- Prof Gina Neff was featured in this incredible round-up of remarkable women in STEM, by Department of Science, Innovation and Technology to celebrate International Day of Women and Girls in Science (2025).
- Prof Prokar Dasgupta was awarded the inaugural Medal of Excellence in Robotic Surgery at the 9th Portsmouth Colorectal Robotic Congress PCC 2025.
- Prof Elvira Perez Vallejos was named in the list of "One Hundred Brilliant Women in AI Ethics" by Women in AI Ethics™ (2025).
- Prof Sana Khareghani was featured on Computer Weekly's Most Influential Women in UK Tech 2025 longlist (2025).
- The paper "Aggregating Multiple Bio-Inspired Image Region Classifiers for Effective and Lightweight Visual Place Recognition" co-authored by Bruno Arcanjo, Bruno Ferrarini, Maria Fasli, Michael Milford, , Klaus D. McDonald-Maier, and Shoaib Ehsan, was awarded the Best Paper Honorable Mention Prize for IEEE Robotics and Automation Letters at ICRA 2025 in Atlanta (USA) (2025).
- The Demo titled "Serious Games for Ethical Preference Elicitation" by Jayati Deshmukh, Zijie Liang, Vahid Yazdanpanah, Sebastian Stein, Sarvapali Ramchurn won the 'Best Demonstration Award at the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025) (2025).
- AIRR Early Access Project Award - Dr Shuang Ao's project was awarded 50,000 GPU hours on the Isambard-AI AIRR supercomputing cluster (2025).
- 'Generative Artificial Intelligence Skills in Schools (GenAISiS)' RAi UK Skills project received a University-wide Recognition Award for 'Engaging the Public with Research' presented as part of the Curious Minds Research Festival (2025).
- Prof Dame Wendy Hall was named 18th in the list of Top Women in AI by Daily AI (2024).
- Prof Gina Neff received an Honorable Mention for the Career Achievement Award from CITAMS (Communication, Information Technologies, and Media Sociology) (2024).
- Prof Muffy Calder was selected as one of the first four Honorary Fellows of Health Data Research UK (HDR UK), in recognition of her outstanding contributions to the Institute (2024).
- Prof Prokar Dasgupta received European Association of Urology (EAU) Innovators Award, presented at the EAU Congress 2024, recognizing him as a pioneer in robotic surgery who completed his 10,000th robotic procedure (2024).
- Prof Marion Oswald, MBE was named as one of six new members of the Biometrics and Forensics Ethics Group (BFEG), to provide independent expert advice to the Home Office on ethical considerations in the use of biometrics, forensics, machine learning and complex data sets (2024).
- Prof Helen Kennedy was elected as Fellow of the British Academy (2024).
- Dr Aislinn Gómez Bergin – University of Nottingham Health Policy Impact Leadership Awardee (2024).

Right: Prof Prokar Dasgupta receiving the inaugural Medal of Excellence in Robotic Surgery at the 9th Portsmouth Colorectal Robotic Congress PCC 2025.

Below: Our researchers celebrating the 'Best Demonstration Award received at the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025).

UK Highlights in 2024-2025

Key events to note include the workshop with the European Trustworthy AI Association, aimed at aligning and influencing responsible AI activities with large investments across Europe, particularly in France and Germany; and UKAIRS which brings the UKRI funded AI Research community together for the first time.

Strategic Partnership with the European Trustworthy AI Association (ETAIA)

RAi UK joined the newly established foundation to collaborate with the foundation, established by industrial leaders including Air Liquide, Naval Group, Safran, Sopra Steria, and Thales, to develop open-source tools, methodologies, and frameworks for trustworthy, secure, and regulation-compliant AI systems. The partnership will focus on nine key areas including research dissemination, open-source maintenance, standards development, and policy shaping, with immediate collaborations included the Trustworthy AI Summit in France this September and the ATRASS seminar series, aiming to create a robust international ecosystem that connects UK research with European institutions and translates responsible AI principles into real-world applications that benefit society.

UKAIRS

The inaugural UK AI Research Symposium (UKAIRS), a landmark event organised by RAi UK, and the UKRI National Research Hubs for Collective Intelligence (AI4CI), and Edge AI for Real Data, successfully convened the UKRI-funded AI research community for the first time.

Hosted at Northumbria University in September 2025, the symposium achieved a critical mass, bringing together 340 AI researchers and academics from 59 institutions: along with industry and government partners. The event was a major success, forging essential links between theoretical research and real-world applications designed to benefit society.

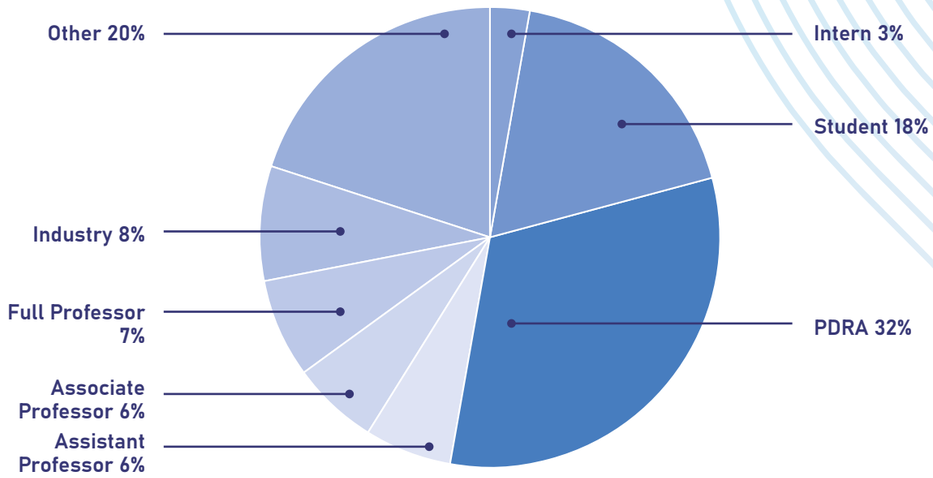
With over 50 papers and more than 100 posters presented, along with demos and engaging panels, the symposium highlighted the strength and diversity of the UK community, focussing not only on technological advancement but also on the critical examination of AI's impact. Key thematic areas of research presented included:

- Innovative technological subject matter and citizen-centred AI
- AI in law enforcement and public sector governance
- Emerging research into the environmental impacts and educational use of AI

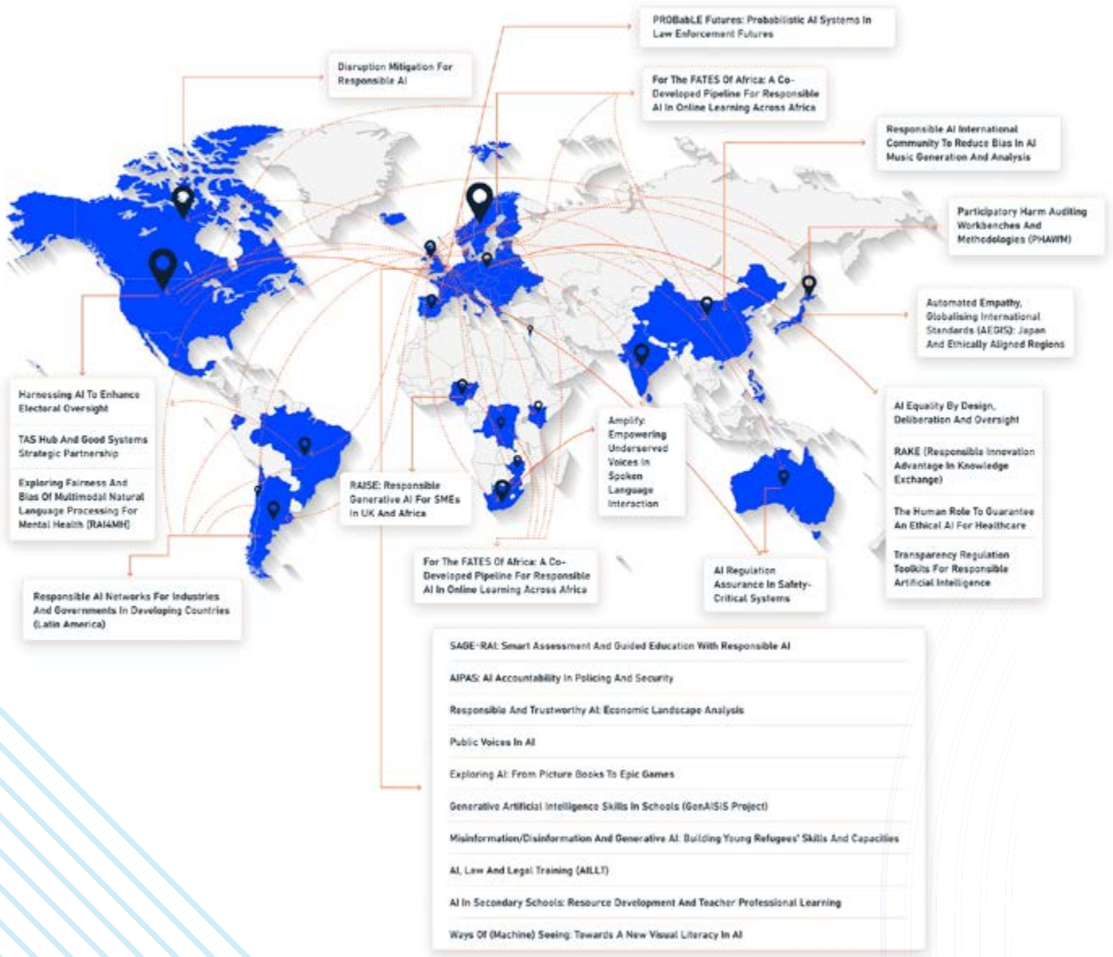
The symposium was lauded for showing "the value of connecting across disciplines, and the impact that research in responsible AI is having on driving AI adoption to improve people's lives", according to Dr James Dracott, Deputy Director for AI at UKRI's Engineering and Physical Sciences Research Council (EPSRC).

For more information and details on UKAIRS 2026, please visit the UKAIRS (ukairs.ac.uk).

UKAIRS Attendees



Key Global Connections in 2024-2025



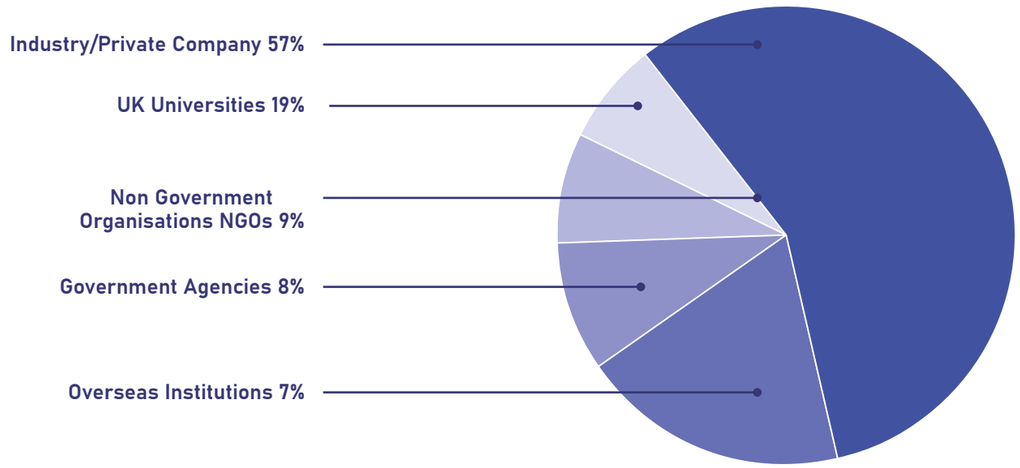
Global reach of RAi projects:

- **North America (USA, Canada):** formalisation of relationship with Partnership on AI, hosting academic visitors including Professor Missy Cummings in support of developing new ecosystem led working groups in Transport and Defence and Security.
- **Europe (EU, Germany, France, Netherlands):** developing relationship with the European Trustworthy AI Association ETAIA, which was cemented at a workshop held at the Royal Academy of Engineering in April.
- **Asia-Pacific (Japan, Singapore, Malaysia, India, Indonesia, Australia):** working with FCDO to deliver a workshop in Malaysia with ASEAN countries, developing relationship with CeRAi in India, including 2 joint research projects, and NUS in Singapore.
- **Latin America (Brazil, Colombia):** working with Chile's National Center for Artificial Intelligence (CENIA) through an International Partnership project creating a Responsible AI Networks for industries and Governments in Developing Countries (Latin America); and through supporting international AI entrepreneurship through Opticut.
- **Africa (Rwanda, Kenya, South Africa and Sub-Saharan Africa):** working with local organisations to support and enhance opportunities for local communities through the FATES of Africa, and the RAISE: Responsible Generative AI for SMEs in UK and Africa projects
- **International Organisations:** Continued active participation and advisory roles with the OECD, UNESCO, and the Council of Europe on developing international norms and standards for responsible AI.

This expanded global footprint underscores RAi UK's commitment to a collaborative, international approach to ensuring AI benefits humanity worldwide.

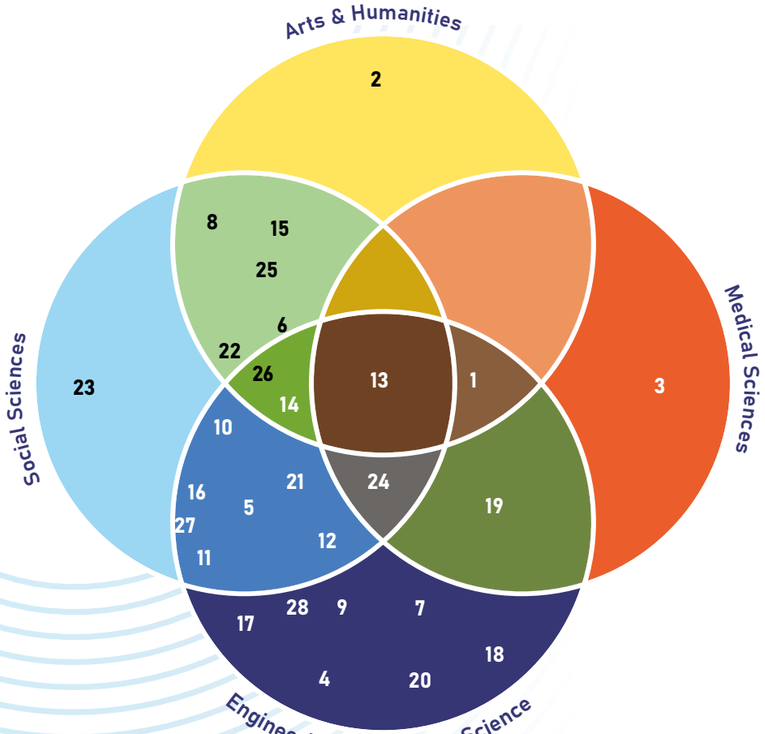
Global Collaborations and Partnerships

An analysis of RAi UK collaborations and partnerships reveals that private industry, including major tech companies such as Amazon, makes up over half of all organisations engaged with. There's also a strong academic presence with 108 universities combined (both UK and international), and significant representation from government bodies and NGOs, such as the World Health Organisation.



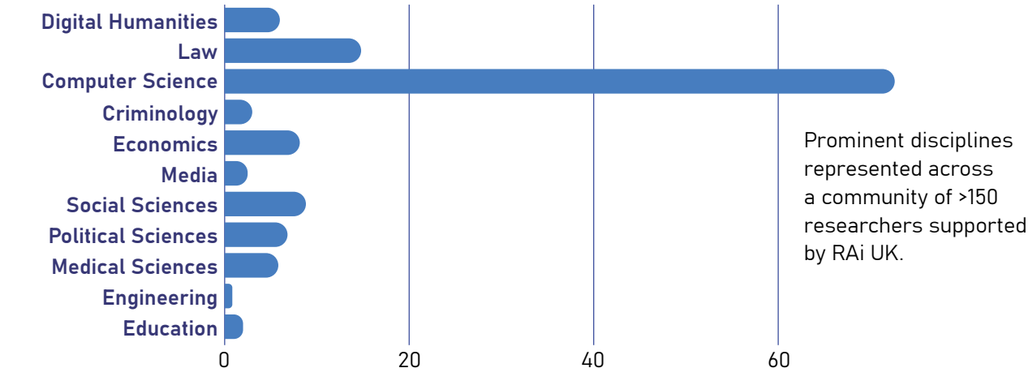
The RAi UK project portfolio

The portfolio of the RAi UK initiative comprises a diverse set of projects, exhibiting substantial variation in their aims, magnitude, scope, and temporal parameters. The duration of these projects is intrinsically linked to the corresponding funding stream. Much of this undertaking yields cross-sectoral research, knowledge exchange, and tangible impact.



1. AdSoLve
2. AI Equality by Design
3. Ethical AI for healthcare
4. AI Regulation of safety-critical systems
5. AI for electoral oversight
6. AIPAS
7. Amplify
8. AEGIS
9. MusicRAi
10. DOMINOS
11. Economic Landscape Analysis
12. For the Fates of Africa
13. PHAWM
14. PROBabLE Futures
15. Public Voices in AI
16. Responsible AI in developing countries
17. RAI4MH
18. RAISE
19. RAKE
20. Robot Autonomy in Public
21. SAGE-RAI
22. TAS Hub/Good Systems
23. Transparency Regulation Toolkits
24. REACT
25. ART-I
26. CulturAI
27. TRAIS
28. Off-grid AI deployments

Multidisciplinarity in the RAi UK project portfolio, according to the disciplines of all the researchers in the respective teams, showing the breadth of collaborations between engineering and computer science, and social sciences, arts and humanities and medical sciences.



These projects are systematically catalogued and presented based on their primary sector of anticipated impact, which encompasses the following domains:

- Cultural and Creative Industries
- Defence and Security
- Education
- Governance, Regulation, and Policy
- Health and Wellbeing
- Science, Innovation, and Technology
- Transport

Research Programme Architecture

The overarching research programme within the RAi UK portfolio is delineated by the following structural component:

- **Keystone Projects:** Large, multi-disciplinary collaborations that will tackle emerging concerns linked to generative and other forms of AI currently being built and deployed across society. Our Keystone Projects cover health and social care sectors, law enforcement, cultural heritage, media content generation, legal profession and education.
- **International Partnerships Projects:** Partnerships that develop strategic collaborations with world-leading research organisations to ensure society deploys and uses AI in a responsible way, beyond national boundaries.
- **Impact Accelerator Projects:** Projects that maximise impact from existing responsible AI research to rapidly realise benefits for the economy, society, culture, policy, health, the environment, and quality of life.
- **Cornerstone Projects:** A programme of research that complements and enhances the wider portfolio, led by RAi UK's postdoctoral researchers. The purpose of the programme is to facilitate agile, multidisciplinary and collaborative research projects on emergent topics in responsible AI.
- **Satellite Projects:** Funded through additional investment by UKRI. The Satellite Projects looked at how responsible AI can help drive productivity and how public voices can be amplified in the design and deployment of AI technologies.

This report proceeds to feature a selection of case studies drawn from across this structured research programme, providing illustrative evidence of the impact generated by RAi UK's projects within their designated sectoral contexts.

Cultural and Creative Industries

The UK government recognises the Cultural and Creative industries as a major economic driver, contributing over £124 billion annually and growing faster than the rest of the UK economy. RAI UK projects directly align with the UK government's strategy to position the UK as a global leader in ethical and responsible Ai development, whilst supporting the creative economy.

Our research is actively working to mitigate systemic bias in AI music generation (MusicRAI), and in spoken language interaction systems (AMPLIFY). The outcomes of these projects serve both a cultural and economic purpose by diversifying available music and language datasets, thus improving access for, and increasing the audience of, marginalised music genres and spoken language techniques.

Case Study: MusicRAI

MusicRAI, led by Professor Nick Bryan-Kinns at the University of the Arts, London, employed a community-driven approach to address systemic bias in AI music generation by championing low-resource AI models and small datasets as alternatives to mainstream systems that favour Western musical traditions. The project convened an interdisciplinary team of researchers, musicians, and industry experts through workshops, funded three artistic mini-projects exploring marginalised genres, and developed both academic research on low-resource models and policy frameworks for government guidance.

Achievements and Value Added: The project successfully mobilised over 100 participants in its major workshop, produced two academic papers and a policy brief that directly informed regulatory strategy, and created an openly accessible AI Music Generation Model Explorer repository. It democratised AI music creation by proving that effective generative systems don't require massive datasets, MusicRAI empowered creators from underrepresented musical traditions with practical tools and frameworks, while simultaneously influencing policy to ensure future AI development in creative sectors prioritises diversity and human agency over computational scale.

Case Study: Amplify: Empowering Underserved Voices in Spoken Language Interaction

AMPLIFY, led by Professor Simon Robinson at Swansea University, refined and scaled the UnMute Toolkit through a participatory design process centred on community sovereignty and ethical data governance. The research involved international workshops and hackathons in collaboration with the University of Edinburgh, Swansea University, and partners in India, focusing on integrating user feedback regarding usability and critical ethical concerns—particularly data sovereignty and privacy—directly into the toolkit's architecture for low-resource speech recognition systems.

Achievements and Value Added: The project successfully demonstrated the toolkit's real-world viability at its launch event in Guwahati, India, and established training programs that built local capacity for autonomous deployment and data governance in minority language communities. The transformative value added is empowerment at scale—by enabling community members without technical expertise to actively participate in developing speech technology for their own languages, AMPLIFY shifted power dynamics in AI development, ensuring that hundreds of millions of speakers of "low-resource" languages can finally access speech-based interactive systems designed by and for their communities rather than being perpetually excluded from voice technology.

Case Study: ARTificial Intelligence (ART-I): show me a story about AI

ART-I, a cornerstone project led by Dr Liz Dowthwaite at the University of Nottingham, employed a value-sensitive design approach to diversify AI narratives beyond dystopian/utopian extremes by engaging creative communities through surveys and participatory workshops. The project collected 36 responses from artists and creatives, identifying critical tensions in how the creative sector perceives AI (empowerment vs. disempowerment, tool vs. partner, democratisation vs. disparity, augmentation vs. replacement), and used these insights to inform the co-creation of artworks serving as mediums for public engagement.

Achievements and Value Added: The project will culminate in a group exhibition featuring new artwork and documentary videos across multiple Nottingham locations in autumn 2025, translating abstract AI ethics into tangible cultural experiences. The unique value-added lies in legitimising artistic perspectives as essential evidence in responsible AI discourse by centring creative practitioners nuanced, embodied experiences with AI rather than purely technical or economic considerations, ART-I creates accessible entry points for public understanding while demonstrating that cultural impact assessment must be integral to, not peripheral to, AI governance frameworks.

Case Study: Ways of (Machine) seeing: towards a new visual literacy in AI

This project, led by Professor Geoff Cox at London South Bank University, employed a co-creation methodology with UK secondary art and design teachers to develop accessible AI literacy activities requiring no specialised technical tools or expertise. Drawing inspiration from John Berger's Ways of Seeing and emphasising "in visual literacy," researchers worked directly with teachers to understand practical classroom needs and constraints, then iteratively designed activities exploring computer vision, machine learning, and generative art that simultaneously produce creative work and prompt critical discussions on authorship, bias, automation, and human imagination.

Achievements and Value Added: The project produced a complete suite of ready-to-implement activities and best-practice resources for teacher training and curriculum development, specifically addressing the systemic challenge that art education faces potential marginalisation in AI literacy discourse despite its unique capacity for exploring ethical dimensions of technology. The critical value added is pedagogical legitimacy - by demonstrating that AI literacy belongs in arts classrooms as much as STEM subjects, the project ensures students develop not just technical competency but critical consciousness about AI's influence on perception, culture, and creativity, while simultaneously supporting teachers who have been under-resourced in engaging with emerging technologies despite their essential role in fostering holistic digital citizenship.

Defence and Security

RAi UK sees the Defence and Security sector as a cornerstone of both national security and economic growth, a sector valued in the tens of billions of pounds annually. The report on "Developing AI capability and expertise in UK defence", published by the Defence Committee in January 2025, underscores the significant implications for responsible AI research in the UK, while the 2025 Strategic Defence Review cites 'greater use of autonomy and Artificial Intelligence within the UK's conventional forces' as an 'immediate priority.' The report signalled a strong need for more than just technological development. To truly embed responsible AI, the UK requires a cultural and structural shift, alongside easy access to advanced tools and the right governance that can keep pace with AI capabilities. A key emphasis must be placed on improving data management and governance to mitigate bias and ensure transparency, which is a core tenet of responsible AI research.

Our showcased projects in this sector support law enforcement actors to fully realise the opportunities of AI in ways that uphold the values of the criminal justice system. Project teams have worked on the production of practical tools for policing, security and the court room (Case study: AIPAS) and a comprehensive framework for the use of Responsible AI in law enforcement and criminal justice (Case Study: PROBabLE Futures).

Case Study: AI Accountability for Policing and Security (AIPAS)

AIPAS, led by Professor Babak Akhgar at Sheffield Hallam University, employed a grounded, participatory methodology combining expert sessions with policing specialists across forensics, counterterrorism, serious organised crime, and child sexual exploitation, alongside large-scale UK-wide public consultations to gather citizen perspectives. This approach developed 12 evidence-based AI Accountability principles and translated them into two practical outputs: technology-agnostic guidance providing step-by-step procedures for implementing AI accountability in policing, and an assessment tool enabling evaluation across the full AI lifecycle while ensuring alignment with public values, legal requirements, and operational needs.

Achievements and Value Added: The toolkit was presented directly to UK Cabinet Office Parliamentary Secretary Abena Oppong-Asare MP in May 2025, who initiated ministerial discussions about adoption, while five police forces are currently reviewing the guidance and the project maintains visibility through bimonthly columns reaching over three million readers internationally. The transformative value added is bridging the implementation gap—AIPAS doesn't merely articulate accountability principles but provides operational mechanisms that law enforcement can immediately deploy, ensuring AI systems enhance rather than undermine public trust while the citizen consultation findings revealed nuanced public support for police AI use coupled with clear expectations for responsible deployment, directly informing both technical development and democratic legitimacy.

PROBabLE Futures (Probabilistic AI Systems in Law Enforcement Futures)

"AI can help police and the courts to tackle digital data overload, unknown risks, and increase operational efficiencies. But the key problem is that AI tools take inputs from one part of the law enforcement system, but their outputs have real-world, possibly life changing, effects in another part – a miscarriage of justice is only a matter of time. Our project works alongside law enforcement and partners to develop a framework that understands the implications of uncertainty and builds confidence in future probabilistic AI, with the interests of justice and responsibility at its heart."

Professor Marion Oswald,
Principal Investigator, PROBabLE Futures

This four-year Keystone Project investigates probabilistic AI systems (facial recognition, predictive tools, LLMs, network analysis) across the full law enforcement pipeline from policing through prosecution, trials, sentencing, and probation. The research employs real-world investigation of AI applications in actual law enforcement contexts, evaluates governance effectiveness including independent scrutiny panels, and produces systematic deliverables including a Responsible AI Checklist for Policing, policy briefs analysing lessons from past technology deployments, and the first international collation of mock trial studies examining how uncertain AI outputs affect legal proceedings.

Achievements and Value Added: The project has achieved exceptional policy influence, with the team providing research-informed evidence to multiple government bodies including the Joint Human Rights Committee, Ministry of Justice, DSIT, College of Policing, and HMICFRS, plus the European Commission's consultation on high-risk AI systems, while Professor Marion Oswald delivered a private briefing to the House of Lords Justice and Home Affairs Committee. The critical value added is systematic risk mitigation—by identifying how the "probable" nature of AI outputs becomes obscured or misinterpreted as they move through the criminal justice system, and by demonstrating through real-world analysis how AI tools taking inputs from one part of the system can produce life-changing effects in another, PROBabLE Futures provides the evidence base necessary to prevent miscarriages of justice before they occur rather than addressing failures retrospectively.

Education

RAi UK is actively engaged in researching the ethical and practical integration of AI tools within educational settings. Through its Skills Programme, RAi UK is dedicated to fostering and scaling the impact of multidisciplinary and cross-sectoral exchange across the UK's AI ecosystem.

The Programme's mission is to cultivate robust and responsible AI skills nationally by supporting diverse, multidisciplinary teams in co-creating resources with businesses, government, educational institutions, and the public, while concurrently sharing these lessons learnt international.

The RAi-funded projects within the Skills Programme reflect this mission and are structured to deliver impact across the following key areas:

- **National Responsible AI Skills Framework:** Establishing a standardised approach to defining and assessing responsible AI competencies.
- **Sectoral Upskilling and Reskilling:** Developing targeted educational interventions focused on responsible AI within and across industries, including formal schooling.
- **Public Awareness and Literacy:** Conducting initiatives to raise public understanding and engagement regarding responsible AI principles.
- **Equity and Disadvantage Mitigation:** Addressing issues of fairness and accessibility in AI through focused educational and outreach efforts.

The following case studies illustrate the broad scope and impact of projects within the Skills Programme.

Education

Case Study: SAGE-RAI: Smart Assessment and Guided Education with Responsible AI

SAGE-RAI, led by John Domingue at The Open University, measured pedagogical benefits through real learner testing comparing AI-guided and traditional methods, then used findings to develop a comprehensive framework for ethically integrating Generative AI into higher education. The research emphasised responsible and human-centred approaches focused on pedagogical soundness, established clear principles of accountability, inclusivity, and human oversight including AI guidelines and output auditing, and promoted transparency with students and educators about AI-driven learning material origins while ensuring high-quality content through technologies like Retrieval-Augmented Generation (RAG) and knowledge graphs (GraphRAG) for accuracy.

Achievements and Value Added: The project's open-source SAGE-RAI technical platform has been integrated into RAG libraries like "embed-js," ensuring broader systemic impact on AI development globally, while achieving institutional adoption by the Open University (affecting massive-scale distance learning) and the Open Data Institute (embedding responsible AI into data professional training). The exceptional value added is dual-track impact: Technical contributions to open-source communities ensure ethical innovations become standard practice in AI development worldwide, while **the ODI's successful external funding collaboration with Mexican authorities to apply the SAGE-RAI framework to a 2026 FIFA World Cup visitor AI tool demonstrates how UK-led responsible AI governance can be operationalised internationally.**

Case Study: For the FATES of Africa: A co-developed pipeline for responsible AI in online learning across Africa

The project, led by Dr Nora McIntyre at the University of Southampton, employed a co-development approach beginning with in-person student consultations in Kenya, followed by remote consultations with educators, community members, policymakers, and academics across East, West, Central, and Southern Africa. The research applied the "AREA-Plus Framework" (Anticipate, Reflect, Engage, Act) for Responsible Research and Innovation with dedicated RRI champions monitoring every consultation session, expanded traditional AIED focus beyond classroom competencies to encompass the whole AIED ecosystem, and uniquely concentrated on online learning within the sub-Saharan African context to create concrete, step-by-step guidance rather than general principles.

Achievements and Value Added: The project produced the FATES of Africa Toolkit decision-making guide available in PDF and app formats (web version launching before December 2025), delivered keynotes including at UN Headquarters for Africa in Nairobi, and achieved paper acceptances at high-profile conferences including CIES 2025 and ECER 2025. The critical value added is culturally embedded equity—by working with local partners rather than imposing frameworks, the project ensures AI educational tools are designed for and with sub-Saharan African contexts, directly addressing how existing AIED guidance excludes non-Western populations while the collaboration with Whizz Education provides a direct pathway for research findings to be integrated into real-world online learning platforms, ensuring the toolkit becomes operational rather than theoretical while fostering international research networks dedicated to responsible AI.

Case Study: AI, Law and Legal Training (AILLT)

AILLT, led by Dr Francine Ryan at The Open University, adopted a highly collaborative, stakeholder-informed methodology through three online workshops involving participants from advice, legal, and education sectors, ensuring the eight free online courses developed were directly responsive to real-world concerns and opportunities identified by end-users. The research co-produced modular training covering Ethical and Responsible use of GenAI, Risk Management, and Legal Regulation and Compliance, with particular focus on addressing lawyers' and self-representing members of the public's vulnerability to GenAI misinformation, while collaboration with Citizens Advice, Fieldfisher, Mishcon de Reya, and Browne Jacobson ensured professional validation of content.

Achievements and Value Added: The eight open-access courses hosted on Open University's OpenLearn Create platform have generated over 56,000 course views, demonstrating urgent demand, while the official launch event initiated conversations with the Bar and Law Society to promote courses across their memberships, positioning the research team as key trainers for the legal profession. The transformative value added is democratised expertise by **providing free, trusted resources co-created with Citizens Advice (whose head of Data Science confirmed the training "will ensure our advisers stay up to date and ultimately help protect the people we are here to support")**, the project directly addresses power imbalances where those least able to afford legal advice are most vulnerable to GenAI misinformation, while the 'Wise in 5' policy guide actively engages policymakers across all four UK nations to advocate for discipline-specific AI literacy as essential rather than optional for responsible AI adoption.

The courses are hosted on The Open University's Open Learn Create platform: OLCreate: AI Law and Legal Training | OLCreate

Case Study: Misinformation/ Disinformation and Generative AI: Building Young Refugees' Skills and Capacities

The project, led by Dr Reem Talhouk at Northumbria University, established a Computer Club for young refugees and asylum seekers as both research and action space to (1) explore their understandings and uses of AI and media for gaining information, (2) co-design and iterate on learning activities and resources, and (3) pilot and evaluate the developed learning resources. Working with YRAS, educators, and third-sector organisations, the team developed Self-Organised Learning Environment activities, Action Learning Sets, and resources focused on increasing YRAS' skills and capacities in misinformation/ disinformation, Generative AI, Responsible AI, and engaging with their wider communities on these issues.

Achievements and Value Added: The project successfully built AI and media literacies among participants, while establishing a partnership with The Guardian Foundation interested in using learning resources as part of their NewsWise programme for young children in schools, providing a pathway for scaling impact beyond initial cohort. The critical value added is amplifying marginalised voices by recognising that young refugees and asylum seekers are disproportionately impacted by AI-propagated misinformation/ disinformation yet already serve as technology educators within their communities. The project transformed vulnerability into agency, ensuring those most at risk from scams and misleading information about fundamental rights became advocates to protect their communities, while The Guardian Foundation partnership demonstrates how resources co-created with specific marginalised groups can have broader educational applicability, thereby ensuring equity considerations inform mainstream AI literacy education rather than being siloed as separate specialised interventions.

Dr Sarah Kiden providing oral evidence to the Human Rights (Joint Committee)

Governance, Regulation and Policy

RAi research is supporting SMEs to harness the full potential of AI in the UK and EU (Transparency regulation toolkits), and shape the development of technical standards in both UK and EU AI policy to ensure that fundamental legal rights are upheld (AI equality by design), as well as informing policy for AI adoption in Latin America (Responsible AI in developing countries). Research has identified gaps in AI safety, assurance and certification, specifically in the context of safety critical systems (AI regulation assurance in safety critical systems) and has the potential to transform AI auditing (PHAWM).

Case study: Participatory Harm Auditing Workbenches and Methodologies (PHARM Keystone)

"Our project will put auditing power back in the hands of people who best understand the potential impact in the four fields these AI systems are operating in. By the project's conclusion, we will have developed a fully featured workbench of tools to enable people without a background in artificial intelligence to participate in audits, make informed decisions, and shape the next generation of AI."

Professor Simone Stumpf, Principal Investigator, PHAWM

PHAWM is one of the three foundational Keystone Projects. The project is a major research initiative developing new, non-technical methods and tools for auditing AI systems. Led by Professor Simone Stumpf at the University of Glasgow and comprising a consortium of 31 researchers across seven leading UK universities, and 28 external partners. The project aims to rigorously maximise the benefits of predictive and GenAI while systematically minimise their potential for harm.

PHAWM developed participatory AI auditing methods and a common infrastructure workbench through deep stakeholder research across four use cases (Health and Media Content for predictive AI; Cultural Heritage and Collaborative Content Generation for generative AI). The methodology involved structured 3-hour workshops with 17 participants in Health (patients, teachers, parents) focusing on information requirements, harm measurement definition, and iterative mock-up feedback, and engagement with 96 Cultural Heritage professionals across 6 workshops and 2 evaluation sessions exploring existing workflows and developing realistic auditing scenarios, with both revealing preference for the term "Event" over "Harm" to ensure comprehensive scope covering both risks and positive outcomes.

Achievements and Value Added: The consortium established 28 external partners including Nokia, Microsoft, Ada Lovelace Institute, and ICO, built an 'AI auditing group' of 210 DataLab community members, released Version 1 of the PHAWM workbench and methodology by October 2025 with openly available code, and published findings at high-impact venues including CHI 2025, EMNLP 2025 (PRISM methodology), and ECIR 2025. The transformative value added is democratising audit power—by creating tools enabling non-technical end-users and decision-subjects to assess AI applications themselves rather than relying solely on technical experts, PHAWM fundamentally shifts who can hold AI systems accountable while the project's investigation of how certification using the PHAWM approach can provide compliance evidence for emerging AI regulation ensures this participatory methodology becomes operationally viable for organisations seeking to demonstrate responsible AI practices.

Case Study: AI Equality by Design, Deliberating and Oversight

The project, led by Professor Karen Yeung at the University of Birmingham, developed theoretical foundations of 'equality-by-design, deliberation and oversight' (EbD) approaches, transforming broad legal principles into actionable technical knowledge through three strategic interventions: directly representing European equality defenders (including Equinet) within CEN/CENELEC AI standard-setting processes, developing knowledge and skills resources to empower public equality defenders and social stakeholders to advocate for and embed EbD principles, and providing targeted training in equality law to UK technology firms to address knowledge gaps and improve the quality of UK-developed AI products.

Achievements and Value Added: The project successfully influenced technical experts' perceptions and awareness of fundamental rights within CEN/CENELEC JTC 21 working groups, securing provisions reflecting 'fundamental rights by design, deliberation, and oversight' principles across critical areas including AI risk management systems, provider quality management systems, high-risk AI system trustworthiness, and data governance. The exceptional value added is preventive governance—by intervening before the window for European AI standards drafting closes, the project ensures fundamental rights safeguards are systematically embedded throughout the entire AI lifecycle rather than retrofitted after deployment, with conditional legislative influence meaning these principles will govern the design, development, and deployment of high-risk AI systems across EU and UK markets if successfully reflected in published Harmonised European Standards (hENs).

DAISY robot system (DOMINOS project)

Health and Wellbeing

AI is emerging as a tool with the potential to transform mental and physical healthcare services. Our projects are supporting the delivery of responsible AI solutions in healthcare capable of mitigating disruption (DOMINOS) and exploring the benefits arising from the application of AI in medical diagnostics and support for mental health, whilst assessing the limitations of LLMs/NLPs and ensuring use is responsible (RAI4MH). Our projects seek to provide answers to emerging ethical questions about AI in healthcare: is it fair? Can it be trusted? Is it deployed transparently? (The Human Role to Guarantee an Ethical AI for Healthcare). RAI research has contributed to the development of new international standards for AI systems that purport to emulate human empathy (AEGIS).

Case Study: Disruption Mitigation for Responsible AI (DOMINOS)

DOMINOS, led by Professor Radu Calinescu at the University of York, developed a tool-supported framework for analysis and debugging of normative requirements, extending it with capabilities for disruption identification, analysis, and mitigation, resulting in a toolkit supporting formal specification, validation, and verification of complex requirements for autonomous agents. The research co-created a diverse, open-access repository of real-world AI use cases and associated normative requirements covering health and assistive care (DAISY diagnostic system, ALMI home assistance system), manufacturing, and environment protection (ASPEN), validating the methodology through engagement with diverse stakeholders to ensure usability for both technical and non-technical users.

Achievements and Value Added: The project received a Distinguished Paper Award at the 2024 International Conference on Software Engineering, directly informed IEEE Guide for the Verification of Autonomous Systems development (with the Project Lead serving as founding Working Group member), contributed significantly to the international research roadmap on resilience and antifragility of autonomous systems, and launched the inaugural annual RETRAI workshop expanding the research community. The critical value added is pre-emptive resilience—by providing systematic methods to detect and mitigate disruptions across environmental shifts, human errors, and adversarial attacks before AI systems fail rather than analysing failures retrospectively, DOMINOS enables responsible AI solutions that maintain compliance and safety under real-world conditions, with current application supporting an AI-enabled robotic system for inclusive breast screening demonstrating how theoretical resilience translates into tangible patient benefit.

Case study: Automated Empathy, Globalising International Standards (AEGIS): Japan and Ethically Aligned Regions

Project AEGIS, led by Professor Andrew McStay at Bangor University, engaged diverse global stakeholders through practical means including scheduling meetings across multiple time zones to ensure truly international input, directly addressing UK Parliament's call for improved global AI governance coordination. The research led to completion and publication of IEEE Standard P7014 on "Emulated Empathy" in 2024, providing the first ethical global guidelines for AI systems interpreting human emotions, while initiating IEEE P7014.1 focusing on ethical recommended practices for automated empathy in general-purpose partner-based AI (expected publication 2026), demonstrating pioneering governance by prioritising inclusive participation over convenience.

Achievements and Value Added: Research insights attracted high-level attention from the European Commission (keynote at "Better Internet for Kids" conference), European Parliament presentations, and a commissioned report for Canadian regulators on children and emotional AI, while the team actively advises key UK safeguarding organisations including the Molly Russell Foundation and Internet Watch Foundation on improvements to the UK's Online Safety Act, plus provides guidance on emulated empathy for the BBC and Microsoft. The exceptional value added is establishing 'soft' governance that augments 'hard' law—by developing technical standards with greater specificity and adaptability than legislation alone can provide, AEGIS creates internationally recognised frameworks for responsible emotional AI while ensuring child safety provisions are embedded at the design stage, with the project's expansion from UK-Japan origins to include Indonesia, Tanzania, and Zanzibar highlighting crucial variations in how different societies interpret empathy in AI, thereby preventing culturally insensitive technology deployment.

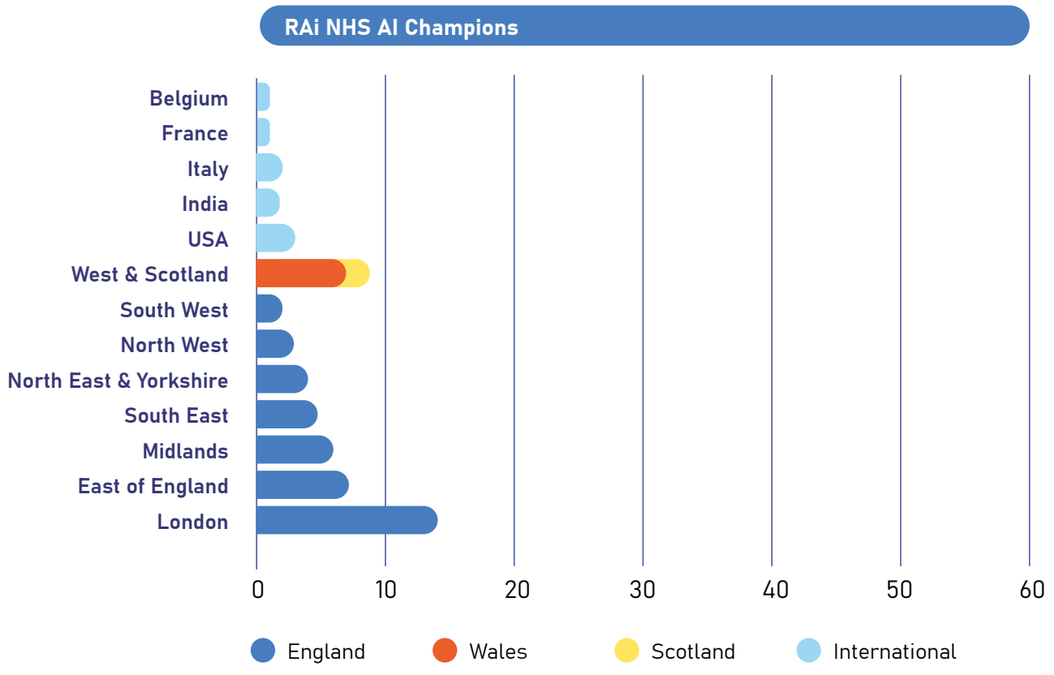
Championing Responsible AI in the NHS: Innovation with Purpose

Healthcare systems worldwide face the dual challenge of harnessing AI's transformative potential while ensuring its safe, equitable, and responsible deployment across diverse care settings.

The Responsible AI NHS Champions programme has established a growing network of over 60 advocates embedded within NHS organisations across primary, secondary, and social care settings. These champions serve as critical facilitators, promoting responsible AI practices within their local healthcare environments. Responsible AI NHS Champions will act as advocates and facilitators for responsible, safe and equitable use of AI within their NHS organisations – primary, secondary and social care. Areas of involvement can include:

- **Advocacy and awareness:** Promote understanding of the potential of AI
- **Training and capacity building:** Support colleagues in the best practice of AI
- **Ethical oversight:** Ensure compliance with national policies and ethical standards
- **Collaboration and engagement:** Work with patients, carers and community
- **Monitoring and evaluation:** Assess impact of AI on outcomes, equity and efficiency
- **Policy development and alignment:** Contribute to the development and alignment of AI-related policies at local & national levels.
- **Network engagement:** Share best practice and dialogue nationally

Network Expansion: The programme continues to grow beyond its initial 60 champions, creating a sustainable infrastructure for responsible AI adoption across the NHS



RAi UK is championing responsible research in AI (TAS Hub/Good Systems), and developing new methods for the responsible design, deployment, use and assessment of AI technologies in society (RAKE). Our projects are generating evidence for responsible AI as a driver for productivity and economic prosperity in the UK, and internationally. Teams have analysed the use of AI across economic sectors in the UK (Economic landscape analysis), developing resources to support SMEs implement generative AI tools that facilitate robust self-assessment by industry and third sector partners (AdSoLve), and identifying opportunities for AU tools to uphold, rather than undermine, democratic practice in electoral processes (AI for electoral oversight). Projects aim to ensure full consideration of public voices and have generated new data, and resources for researchers, on public attitudes to AI (Public Voices in AI).

Designed by Freepik

Science, Innovation and Technology

Case Study: RAKE – Responsible Innovation Advantage in Knowledge Exchange

Project Lead:
Marina Jirotka (University of Oxford)

RAKE, employed an action research methodology centred on a four-month engagement with Kainos (multinational digital transformation provider), beginning with a survey assessing Delivery Managers' understanding of Responsible Innovation, followed by introductory workshops and an eight-week experimentation phase where DMs integrated RI Prompt and Practice cards into standard product development cadence with clients, concluding with weekly reporting and follow-up surveys gauging perspectives and reception. The approach emphasised participation and direct impact rather than solely data-gathering, facilitating mutual knowledge exchange where researchers gained insight into Agile and public sector development processes while Kainos teams learned to embed RI principles, with subsequent internal workshops using RI cards to develop methodology for incorporation into Agile product development cycles.

Achievements and Value Added: Key findings were provided to Kainos to support prototyping of an RI Delivery Tool anticipated as an extension of their current Delivery Management Tool guiding new responsible-innovation-focused working methods, while the project produced a published white paper on "Frameworks and Toolkits for Assuring Responsible AI" (August 2025), contributed to "Advancing Trustworthy AI" white paper, and has forthcoming papers on embedding RI challenges (December 2025) and UK RI uptake in businesses (January 2026). The critical value added is practical implementation pathways—by demonstrating how RI principles can be operationalised within existing Agile workflows rather than requiring separate processes, RAKE proves that responsible innovation enhances rather than hinders commercial development, with the validated RI Prompt and Practice Cards providing immediately usable tools that other organisations can adopt, thereby scaling impact beyond the single case study to establish new standards for how technology companies integrate ethical considerations into daily practice.

Case Study: AdSoLve – Addressing Socio-technical Limitations of LLMs for Medical and Social Computing

"LLM's are being rapidly adopted without forethought for repercussion. For instance, UK judges are allowed to use LLMs to summarise court cases and, on the medical side, public medical question answering services are being rolled out.

Our vision addresses the socio-technical limitations of LLMs that challenge their responsible and trustworthy use, particularly in medical and legal use cases."

Professor Maria Liakata, Principal Investigator,

The AdSoLve keystone project, led by Professor Maria Liakata at Queen Mary University of London, addresses socio-technical limitations of LLMs particularly in medical and legal contexts through developing comprehensive evaluation benchmarks and RRI criteria for rigorous assessment of LLM systems, emphasising the need for bias auditing and explainability as standard procedures. The research employs multidisciplinary approaches and stakeholder co-creation methodology to bridge gaps between technical AI capabilities and practical ethical deployment, focusing on ensuring LLM applications in sensitive domains like healthcare align with public values, legal requirements, and clinical needs.

Achievements and Value Added: The developed toolkit has been successfully piloted across three NHS trusts, achieving a 15% reduction in identified demographic bias in diagnostic AI tools and a 20% increase in clinician confidence in AI-generated insights, translating directly to more equitable patient care and trusted diagnostic enhancement, while the toolkit is currently under consideration for integration into national healthcare guidelines for AI procurement representing potential systemic impact standardising responsible AI practices across UK healthcare. The exceptional value added is measurable harm reduction—by providing concrete assessment frameworks enabling organisations to identify and mitigate bias before deployment rather than discovering problems through adverse patient outcomes, AdSoLve establishes new baselines for what constitutes acceptable AI performance in high-stakes domains, with the project's RRI criteria and evaluation benchmarks offering essential tools for regulatory bodies to implement enforceable responsible AI policies while the emphasis on real-time bias detection mechanisms and patient feedback loop integration charts necessary directions for next-generation medical AI systems.

Case Study: Public Voices in AI

"Public voices need to inform AI research, development and policy much more than they currently do. This project represents a commitment from UKRI and RAi UK to ensuring that happens. It brings together some of the best public voice thinkers and practitioners in the UK, and we're excited to work with them to realise the project's aims."

Helen Kennedy, Professor of Digital Society, University of Sheffield

The project, led by Professors Helen Kennedy and Ros Williams at the University of Sheffield, employed a multi-faceted approach including reviewing over 300 global studies on public attitudes toward AI to map who is included/excluded, surveying over 4,000 international AI researchers to understand their perceptions and inclusion of public input, conducting qualitative community-level studies across diverse UK communities (Belfast, Southampton, Lambeth) using Community Researchers to centre minoritised and excluded groups, and establishing a competitive Public Voices in AI Fund distributing up to £195,000 to four participatory projects led by or working with underrepresented communities such as platform workers and asylum seekers, with a People's Advisory Panel comprising public members with prior AI engagement experience guiding design, dissemination, and resource development.

Achievements and Value Added: The project produced a comprehensive Framework and Self-Assessment Workbook for including public voices in AI providing clear guidance on engaging people across the AI lifecycle spectrum from low to high power participation, created an Evidence Database and Repository of Use Cases Report synthesising existing public views and effective engagement projects, and generated evidence on how structural inequities lead to disproportionate negative AI impacts (e.g., in policing and welfare) intended to inform policy-making and industry codes of practice. The transformative value added is systematic participation infrastructure—rather than one-off consultations, the project establishes replicable methodologies and practical tools enabling sustained, meaningful public engagement in AI development, with the Fund's support for community-led research (including The Workers' Observatory and Migrants' Rights Network projects) directly fostering agency and power for those most affected by AI to shape discourse rather than merely be subjects of research, while the Framework's explicit acknowledgment that different engagement levels require different resources and approaches provides honest guidance preventing tokenistic participation, thereby creating foundations for genuinely democratic AI governance where diverse public values inform rather than retrospectively critique technology design.

This video provides an overview of the Public Voices in AI project: Public Voices in AI.

Transport

Artificial Intelligence (AI) within the transport sector necessitates rigorous research to ensure responsible innovation, particularly concerning public interaction, trust, and safety.

The RAi UK's research agenda in transport and communications strategically foregrounds everyday interactions with both existing and emerging technologies. This focus is crucial, as the introduction of autonomous systems—such as delivery robots and self-driving shuttles—shifts the nature of human interaction within shared public spaces.

Transport

Case Study: Understanding Robot Autonomy in Public

Project Lead:
Stuart Reeves (University of Nottingham)

The project employed naturalistic fieldwork capturing "anyone and everyone" interacting with or around robots in public streets, examining delivery robots in the UK and autonomous transport shuttles in Sweden as comparison cases to develop understanding of how members of the public respond to autonomous robots in shared spaces and how associated personnel (operators, vendors) work with these systems. The research emphasised mundane, in situ settings deliberately overlooked by conventional practices, bringing together a robust network of academics across transport, HCI, HRI, robotics, sociology, and linguistics disciplines to share empirical data and develop interdisciplinary insights that can inform future design and responsible governance of civic robotics in public environments.

Achievements and Value Added: Findings have been cited in a Policy Document by the Swedish National Road and Transport Research Institute (VTI) and disseminated via the Urban Robotics Foundation (URF), demonstrating direct path from empirical research to regulatory influence, while the project developed sustained support for a researcher network investigating robot autonomy in public, constructed interdisciplinary linkages between sociological studies and partners in video-based interaction studies, engineering, and HRI research, and informed emerging conceptual and regulatory frameworks underpinning civic robotics by incorporating comprehensive regulatory, governance, and responsibility considerations. The critical value added is grounding policy in lived experience by prioritising everyday, mundane interactions rather than controlled laboratory studies or aspirational scenarios, the project ensures regulatory frameworks evolve in response to real-world data about how actual people (not ideal users) encounter autonomous systems in public spaces, with the emphasis on "anyone and everyone" capturing diverse circumstances including children, elderly, disabled individuals, and those with varying technological literacy, thereby preventing governance models that only account for technically sophisticated users and ensuring civic robotics development serves entire communities rather than narrow demographics.

Investing in skills at every level

Next generation skills development

The RAi UK Skills Pillar is strategically focused on delivering impactful initiatives designed to cultivate a robust and responsible AI skills base across the United Kingdom, with a vision of fostering a thriving national ecosystem of responsible AI education innovators.

The Skills projects have surpassed all expectations, bringing new people, energy, and ideas into RAI. Within a short period of time, the 6 projects have produced outstanding and impactful resources, including videos, games, online activities, a book, and an OU online course. My own understanding of skills needs has grown greatly, and I have enjoyed hearing about the range of approaches and audiences, from young people to Professionals.

Professor Dame Muffy Calder

Vision and Core Objectives

The central vision is to establish national expertise in Responsible AI (RAI) by fostering a community of education innovators and supporting a broad range of stakeholders. The core objectives are achieved through a three-pronged strategy:

- Convening responsible AI skills resources and leadership across the entire RAi ecosystem.
- Accelerating innovation in responsible AI skills policy and practice.
- Supporting the ecosystem to provide practical experience and opportunities for skills development.

The Pillar's engagement areas are broad, encompassing the development of national responsible AI skills frameworks, driving upskilling and reskilling within and across sectors, raising public awareness of responsible AI, and critically, addressing equity and disadvantage in AI through education. Core to all aspirations is the principle of equity and fairness in training and development.

Impact and Initial Initiatives

The immediate impact of the Pillar has been to address the lack of established benchmarks in the burgeoning field of RAI training. Recognising the recent explosion of diverse training opportunities, the first key initiative convened a multidisciplinary group of 21 participants from 13 UK universities, encompassing 18+ disciplines, policy, civil society, and industry.

This review sought to answer critical questions regarding the efficacy of current RAI training, including: what makes RAI training effective, for whom is it intended, and who might be excluded. The exercise ultimately benefits the community by providing recommendations and guidance on effective practices, highlighting existing knowledge gaps that require addressing, and revealing new opportunities for improved collaborative efforts across the entire RAi ecosystem. By empowering a broader community of stakeholders, the supported initiatives are designed to ensure that the full potential of AI is harnessed in a manner that is fundamentally safe, inclusive, and beneficial for all.

Four Markers of Effective Responsible AI Skills Resources

The RAi UK Skills Pillar commissioned a multidisciplinary review involving extensive online search and content analysis of existing UK RAI learning resources, with critical evaluation conducted by a specially convened working group of 21 participants from 13 UK universities representing 18+ disciplines plus perspectives from policy, civil society, and industry. This diverse composition enabled co-creation of an evaluation framework accounting for varied dimensions of Responsible AI, which was then applied to four distinct types of publicly available resources to identify weaknesses and generate concrete recommendations beyond superficial landscape mapping.

The core findings were synthesised into Four Markers of Effective Responsible AI Skills Resources, designed to serve as a practical evaluation framework for both providers and learners:

Four Markers of Effective Responsible AI Skills Resources

Marker	Strategic Focus for Skills Development
Contextually grounded and relevant	Training must explicitly address the full AI system lifecycle and be practical to UK or international contexts, covering both the technical and non-technical dimensions of RAI.
Trusted Resource Developers	Developers must possess demonstrable qualifications and credibility. Resources must maintain a balanced engagement with topics and align with UN Sustainable Development Goal 4 (Quality Education), demonstrating a commitment to inclusive and equitable education.
Addresses Clearly the Audience and their Needs	Resources must provide multiple pathways for learning engagement, effectively implementing accessibility standards and adopting inclusive education strategies that acknowledge and support learners' varied backgrounds and objectives.
Promotes and Inspires Further Dialogue	Content must explicitly encourage reflective practice, discussion, and critical evaluation of AI's societal implications, fostering deeper intellectual engagement beyond mere technical instruction.

The ultimate impact of this report lies in providing clear, actionable guidance for the development of future RAI training initiatives, thereby directly contributing to the quality assurance of responsible AI education in the UK. By addressing identified gaps and emphasising the necessity of multidisciplinary, inclusive, and contextually grounded training, the Skills Pillar aims to ensure that the nation's workforce is equipped with the critical literacy required to design, deploy, and govern trustworthy AI systems. These markers are designed to spark essential conversations toward establishing national RAI education quality benchmarks, supporting collaborative efforts, and ultimately fulfilling the overarching mission of creating a safe and equitable AI ecosystem.

AI in Secondary Schools: Resource Development and Teacher Professional Learning

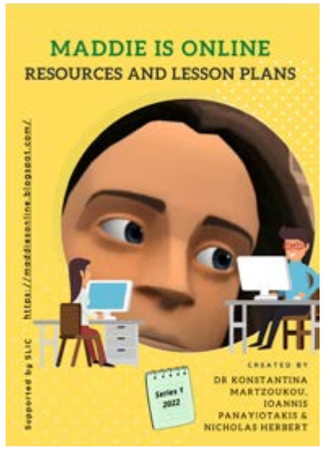
The project, led by Professor Judy Robertson at the University of Edinburgh, co-created resources with six 'AI in Residence Teachers' from various subjects to develop subject-specialist materials in five contrasting disciplines plus additional resources for supporting learners with additional support needs, while running professional learning sessions for approximately 200 teachers. The research designed and developed an evaluation website enabling teachers to provide expert/peer review on AI resources and classroom uses so future teachers can search for evidence about tools and topics they're considering, with data collection from teachers who attended professional learning and subsequently tried resources in classrooms informing iterative development.

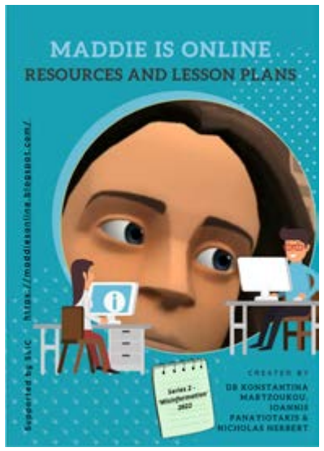
Achievements and Value Added: The project produced the Teach AI Literacy Handbook: including a 'work in progress' AI curriculum for Scotland (with support from Scottish Government, Education Scotland, and Scottish Qualifications Authority) posted to every secondary school in Scotland in August 2025, created co-designed educational materials available on the project website, and delivered well-received professional learning workshops and conference talks with another series planned for Autumn 2025 including a showcase seminar for c.200 teachers in September 2025. The critical value added is systemic educational transformation—by working directly with classroom teachers rather than imposing top-down curricula, the project ensures resources address practical pedagogical needs and constraints while the evaluation website creates ongoing quality assurance infrastructure enabling evidence-based resource selection, with the 'work in progress' curriculum acknowledgment demonstrating adaptive rather than prescriptive approach that can evolve with rapidly changing AI landscape, thereby establishing Scotland as exemplar for how nations can develop comprehensive, teacher-validated, subject-integrated AI literacy education at scale.

Generative Artificial Intelligence Skills in Schools (GenAISiS)

GenAISiS, led by Dr Konstantina Martzoukou at Robert Gordon University, partnered directly with secondary school students to explore responsible GenAI use through their co-creation of open educational resources, with students articulating voice and enacting experience via fictitious characters and cartoons. The project recognised and empowered school librarians as key agents in fostering responsible GenAI use, developing a co-produced open educational toolkit covering GenAI academic integrity, information literacy, and critical thinking skills in schools for wide dissemination via open workshops.

Achievements and Value Added: The project was awarded a university-wide Recognition Award for 'Engaging the public with research' as part of the Curious Minds Research Festival, demonstrating external validation of its impact. The transformative value added is student agency—by positioning young people as co-creators rather than passive recipients of AI education, GenAISiS ensures resources reflect authentic student concerns and experiences while the creative approach using fictitious characters and cartoons makes complex ethical concepts accessible, with the strategic focus on empowering school librarians addresses an often-overlooked professional group who are ideally positioned to support information literacy across all subjects, thereby creating sustainable infrastructure for responsible AI education that doesn't depend solely on computer science teachers or require schools to hire new specialised staff.





RAi UK: Supporting and Training Early Careers Researchers

The RAi UK Cornerstone programme plays a pivotal role in skills development by actively supporting and training its cohort of Early Career Researchers (ECRs) and Transitional Associate Professors (TAPs). The programme's mission is fundamentally centred on facilitating research that employs an agile, multidisciplinary, and collaborative approach, co-curated with stakeholders to benefit societies in the UK and internationally. ECRs are positioned within this responsive and creative research programme to complement and enhance the wider RAi UK portfolio.

Framework for Skills Development and Training

The Cornerstone structure is specifically designed to provide a rich, supportive, and practical training framework for its researchers. This framework offers broad experience in research project management, including proposal writing, budgeting, and other essential aspects of research operations. ECRs are immersed in multi- and interdisciplinary work, gaining invaluable experience in collaborating across various disciplines, with national and international institutions, and with non-academic organisations. This exposure is critical for broadening networks and advancing complex, cross-disciplinary research.

Furthermore, the programme explicitly cultivates impact development, particularly through engagement in policy and public engagement activities. ECRs are integrated into formal RAi UK mechanisms, such as coordinating responses to UK government consultations (e.g., on Copyright and AI and the Digital Inclusion Plan) and submitting written evidence to Parliamentary inquiries.

The programme also embeds Responsible Research and Innovation (RRI) principles into its methodology and governance. This commitment is demonstrated through projects like the ongoing review of the EPSRC Sandpit Mechanism, which directly responds to feedback regarding EDI challenges in non-inclusive funding formats. Additionally, ECRs lead studies focused on understanding how researchers experience multidisciplinary collaboration in AI, aiming to develop practical, research-informed resources to support effective teamwork in Responsible AI.

Positive Career Impact and Esteem Measures

RAi UK actively encourages its ECRs and TAPs to pursue external career development opportunities. The Cornerstone framework has yielded tangible, positive impacts on the careers of its participants:

- Securing Permanent Academic Roles: Several researchers have successfully secured new, permanent appointments while retaining their involvement with RAi UK, including appointments to Lecturer positions at prestigious institutions such as the University of Southampton and the University of Liverpool.
- External Funding Success: ECRs have successfully competed for new research funding. For example, Shuang Ao was awarded 50,000 GPU hours on the UK's national AI Research Resource Isambard-AI, a prestigious allocation set to power cutting-edge experiments at the intersection of AI, climate policy, and finance.
- Awards and Esteem: ECRs have received international recognition, such as Jayati Deshmukh, who was part of the winning team for the Best Demonstration Award at AAMAS 2025 (a world-leading conference in AI and multiagent systems) for an interactive game exploring ethical AI.
- Developing International Partnerships: Cornerstone projects provide an excellent platform for developing lasting international relationships, as evidenced by submitted grant applications to schemes like the UNESCO Women in Science scheme and the Royal Society ISPF (Japan) call.

By offering this blend of practical management experience, disciplinary breadth, policy engagement, and dedicated career support, the RAi UK Cornerstone projects serve as a vital incubator, successfully establishing the next generation of entrepreneurial and responsible AI research leaders.

Enterprise

RAi UK - Enterprise Fellows

Background

The RAi UK Enterprise Fellowship is a 12-to-18-month accelerator programme designed to support university-based researchers in transforming their academic expertise into commercially viable ventures centred on responsible AI. The programme takes a human-centred approach to co-create an inclusive ecosystem that addresses critical societal needs of justice, equity, sustainability, and inclusivity while embedding Equality, Diversity and Inclusivity (EDI) and Responsible Research and Innovation (RRI) principles from the outset. The programme aims to maximise the beneficial impact of AI research across multiple domains including the economy, society, culture, health, and the environment through ethical and trustworthy AI deployment.

The Enterprise Fellowship call evaluated projects across five key dimensions: Innovation Potential (originality, protectable IP, and significant impact potential), Commercial Credibility (clear problem articulation, strong value proposition, identified target markets, and evidence of market need), Ambition (entrepreneurial capability and pathway to securing long-term investment), Responsible AI Integration (alignment with RAi UK's vision for societal benefit through embedded EDI and RRI principles), and Feasibility (scalability, team commitment, institutional support, and likelihood of creating a self-sustaining business). Through this rigorous assessment process, three Enterprise Fellowships were awarded to support the responsible transfer of academic innovation into marketable AI products and services that address technical, social, legal, and ethical challenges.

BrainHealthX: Responsible AI-guided solution for early dementia prediction

Professor Zoe Kourtzi

UNIVERSITY OF CAMBRIDGE

Dementia currently affects over 55 million people worldwide, imposing a substantial societal cost exceeding $1 trillion annually. Despite over $56 billion invested in R&D over the past three decades, a critical deficiency persists in sensitive diagnostics for early-stage detection, the period when interventions are projected to be most efficacious. Addressing this gap, a responsible, multimodal AI tool known as BrainHealthx (BHx) was co-created with healthcare partners, clinicians, and the public. BHx is specifically designed to enhance early prediction and patient stratification, thereby optimising personalised interventions. Prof. Kourtzi's Fellowship aims to scale BHx into a trusted, fully deployable clinical decision support system, enabling the precise identification of which patients will benefit from which intervention and at what time, ultimately driving the development of precision interventions and novel treatments. Prodomic has recently been accepted onto Innovate UK's Kings Med Tech Accelerator.

PACER (ProActive Causal ExplaineR)

Professor Lars Kunze

UWE Bristol | University of the West of England

Our vision centres on enhancing the safety, accountability, and trustworthiness of Autonomous Systems (AS) by mandating the integration of robust data recorders, or 'black-boxes', mirroring the crucial flight-data recorders used in aviation. Following successful development and demonstration of these recorders and associated recording standards in collaboration with industry partners, the objective of the Enterprise Fellowship is to operationalise a new cloud-based Software-as-a-Service (SaaS) platform: PACER (ProActive Causal ExplaineR). This platform is designed to leverage and meticulously analyse recorded event-data logs to proactively anticipate potential safety issues, provide precise causal explanations for incidents, and subsequently enhance safety measures. This mechanism is critical for fostering improved understanding and securing the necessary trust among developers, operators, and all other relevant stakeholders.

OPTIcut – AI-Powered Decision Support System for Sustainable Banana Processing

Professor David Elizondo

DE MONTFORT UNIVERSITY LEICESTER

The OPTIcut project, directed by Professor David Elizondo, leverages computational intelligence to transform banana farming by addressing waste reduction and enhancing agricultural sustainability. This collaborative initiative, which includes partners such as CORBANA, the University of Malaga, Jaen University, and De Montfort University, is focused on the refinement of segmentation algorithms and the automation of fruit damage detection, with funding allocated for essential GPU upgrades and field equipment. The project is underpinned by a commitment to Responsible Research and Innovation (RRI) principles, ensuring that the resulting impactful and scalable solutions foster equitable growth and actively contribute to the advancement of the circular economy within banana-producing regions.

IGNITION Hackathon

The IGNITION AI Startup Hackathon was a three-week program designed to transform academic research into commercial ventures, specifically targeting early career researchers and innovators. Supported by RAI UK, UCL Centre for Digital Innovation, and Future Worlds (based at the University of Southampton), the initiative addressed a critical gap: fewer than 5% of research-active academics in the UK lead startups, with this figure being even lower for those from underrepresented backgrounds. IGNITION provided a dedicated platform where participants could translate groundbreaking research into impactful startup concepts that leveraged AI to create positive social impact across areas like healthcare, education, sustainability, and public services, while carefully considering ethical implications and potential risks.

For participants, the hackathon offered exceptional value by providing hands-on experience in startup development regardless of their technical background, with expert mentorship and structured support throughout the journey. It created a unique opportunity for early career researchers from any academic discipline to develop viable business concepts while working in diverse, multidisciplinary teams of 3-6 members. The programme not only equipped participants with practical entrepreneurial skills and access to a guiding vision for their AI-driven ventures, but also opened doors to potential commercial opportunities, fostering innovation that bridges the gap between academic research and real-world impact. This inclusive approach allowed ideas from underrepresented groups to be transformed into accessible, impactful solutions.

FUTUREWORLDS

UCL

AIR26 Bootcamp

The AIR26 Bootcamp was a two-day, in-person event serving as the mandatory initial phase for the full six-month AIR26 acceleration programme.

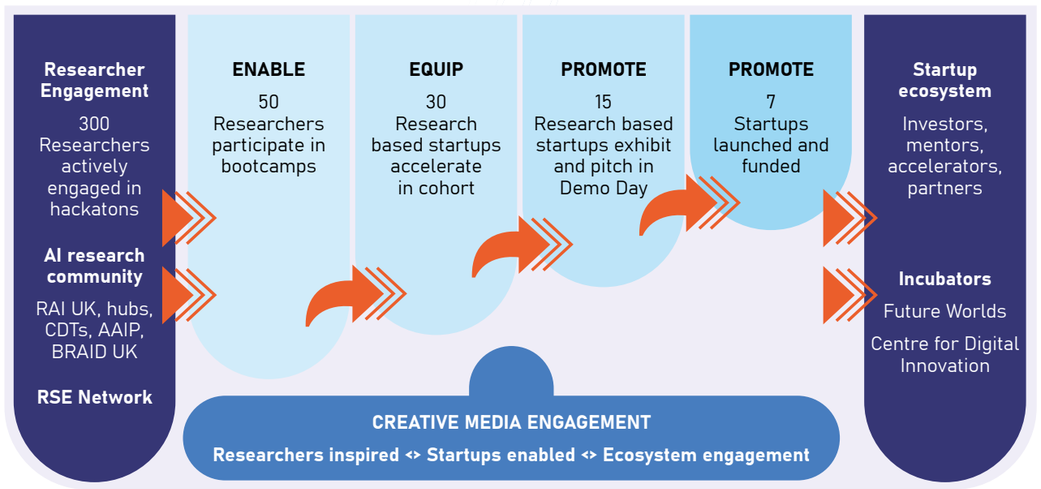
The event, hosted at the University of Southampton's Winchester campus on 2–3 September 2025, was targeted at UKRI-funded researchers who were actively developing an AI-based idea or technology with commercial potential. Crucially, the innovation must embody a responsible use of Artificial Intelligence.

Aspiring founders received intensive immersion training from startup experts, covering core entrepreneurial principles such as product-market fit, value proposition, business models, and market potential. The primary objective was to enable participants to rapidly assess the viability and capacity for growth of their idea. At the conclusion of the two days, attendees were afforded the opportunity to pitch for a fully funded space on the subsequent six-month acceleration programme designed to make their startup investor ready.

Artificial Intelligence Responsibly (AIR) Programme

Funded by RAi UK, the Artificial Intelligence Responsibly (AIR) programme is a high-level, cross-institutional initiative spanning all four nations of the UK, established to address the critical gap between world-leading academic research in Artificial Intelligence and its successful commercial implementation. Led by Future Worlds, with the UCL Centre for Digital Innovation and RAi UK, AIR is positioned as a launchpad to accelerate the translation of research into globally scalable, responsible AI startups.

Through its "triple lock" approach, making entrepreneurship Accessible, Authentic, and Aspirational, AIR seeks to create an inclusive ecosystem that transforms diverse researchers into successful AI entrepreneurs. The programme bridges the research-commercialisation gap while embedding responsible AI principles from conception, ultimately aiming to produce globally scalable, ethical AI startups that deliver significant socio-economic benefits and inspire the next generation of diverse STEM founders.



Looking Forward

As we look to 2026, RAi will continue to convene the UK AI research ecosystem, with an emphasis on translating research into real world impact for societal good. Our focuses will include:

Supporting economic growth by continuing to support the transformation of research into entrepreneurship, connecting research to industry, and developing the responsible AI landscape to allow policymakers and business leaders to accelerate adopting AI responsibly and effectively.

Supporting NHS adoption of AI through the continued growth of the RAI NHS AI Champions network, delivering our integrated care system pilot in its first board, and building the foundations for wider rollouts.

Breaking down barriers to opportunity by engaging with the public, policymakers and educators to ensure that people have the skills and knowledge they need to make the most of an AI future, whether the skillsets to harness new job opportunities, or to become intelligent customers of AI systems as these become more integrated into every day interactions.

Support responsible, trustworthy security at home and abroad, by continuing to work hand in glove with policing and the defence sector, ensuring that research, ethical principles and responsibility, and operational impact are closely linked so that government can fulfil its first obligation to protecting the public with safety and reliability at its core.

Quantifying, auditing and evaluating AI performance through engagement with metrology, policymakers and industry to create meaningful evaluation and assessment metrics for AI systems to maximise the value of adoption, and to create more intelligent customers in the marketplace of emerging technologies

B60, University of Southampton,
Southampton, SO17 1BJ

Rai.ac.uk