Vision to Reality: Responsible AI UK

ANNUAL REPORT 2023-24

www.rai.ac.uk
info@rai.ac.uk

Contents

With a £35 million UKRI investment, RAi UK is a program dedicated to delivering interdisciplinary research and fostering an international ecosystem for Responsible AI research and innovation. Through extensive consultations across the UK, RAi UK has identified emerging challenges in responsible AI and has deployed over £17 million into projects aimed at accelerating the adoption of responsible AI practices and technologies. RAi UK brings research-based expertise that is connective, adaptive, and world-leading through field-building, and engagement with communities, publics, industries, and governments. For more information or to get involved, please visit rai.ac.uk.

Foreword from our CEO, Prof Sarvapali (Gopal) Ramchurn

I am pleased to present the first annual report for Responsible Ai UK (RAi UK). This report captures the key achievements of the programme and demonstrates the tireless efforts of a whole community, the collaborative spirit of the UK AI ecosystem, and the vision of both early-career researchers and research leaders from across the UK.

The AI landscape is changing rapidly, with new challenges emerging in a range of sectors, within government, industry and for society. RAi UK aims to be the trusted voice that they turn to for advice on how to develop and integrate AI into their processes, products, and daily lives to ensure AI is responsibly deployed and governed.

In the last year, we have focused on building a community. We have deployed over £15m to accelerate the translation of established responsible AI research into real-world toolkits, standards and products, to develop partnerships with other major responsible AI initiatives across the world, and, most recently, to establish the foundational projects of the programme, through our Keystone Projects. At the same time, we have contributed to the public debate on AI, through contributions at national and international events and policy consultations. Some of our key achievements include:

- Organisation of over 20 roundtables across Wales, Scotland, England, and Northern Ireland with academia, industry, NGOs, and government, to shape the core research challenges to be addressed by RAi UK. Three major research themes emerged from this consultation, namely: Responsible AI-powered Organisations and Economies, Addressing Harms, Maximising Benefits of AI, and AI Law, Governance and Regulation in an International Context.
- Launch of 23 research projects to accelerate impact, develop international partnerships, and to undertake fundamental research. These projects address AI challenges in key sectors such as health and social care, creative industries, social robotics, and international AI regulation.
- Over 15 panels and presentations, and co-organised events including the AI Fringe, Royal Society workshops, and Keystone Projects launch at CogX in Los Angeles. Under the leadership of Prof Dame Wendy Hall, the Royal Society convened a workshop and response to the UN AI Advisory Body report Governing AI for Humanity.
- Launch of the AI Champions for the Health and Care sector in collaboration with Guy's and St Thomas' hospital and a number of NHS Trusts, through the leadership of Prof Prokar Dasgupta, lead of RAi UK's Health and Social Care Working Group.
- 15 Early Career researchers joining RAi UK and developing projects for our Coordinating Keystone to be led by Prof Joel Fischer.
- Six Skills-related projects funded to develop courses and toolkits for the upskilling of the workforce in responsible AI.

Across all these activities, we ensured we adhered to our guiding principles which we worked hard to define as a collective. This ensured our processes meet the high standards required to deliver public funds, while ensuring equality, diversity, and inclusivity. In the coming years, RAi UK aims to deliver advice that is research-driven, building on empirical evidence, both qualitative and quantitative while integrating inter-disciplinary perspectives. Our advice will collate insights not just from our own team, but from other researchers and organisations from across the research ecosystem. I recognise that we are only at the beginning of a journey to ensure that all AI is Responsible AI. I would like to thank all those who helped us get here: UKRI, our partners, the research community that delivered hundreds of peer reviews, and other key parts of the UK AI ecosystem.

Introducing our team

RAi UK brings together a multidisciplinary team with a wealth of national and international research and advocacy experience. We are proud of our diversity, in gender, cultural background, and career level.

We will lead RAi UK's mission, leveraging strong relationships with senior stakeholders within government and associated agencies.

Professor Sarvapali (Gopal) Ramchurn FIET
Chief Executive Officer (CEO)

Leading expert in Responsible AI, with a focus on socio-technical applications, including energy systems and disaster management. Professor of AI at the University of Southampton, Director of the UKRI Trustworthy Autonomous Systems Hub and Co-CEO Empati Ltd.

Professor Gina Neff
Deputy CEO

Leading expert on the social impact of technology. Executive Director of the Minderoo Centre for Technology and Democracy at the University of Cambridge and Associate Director of the ESRC Digital Good Network.

Professor | Yr Athro Matt Jones
Chief Operating Officer (COO)

Leading expert in Human-Computer Interaction. Professor of Human-Centred Computing at Swansea University and the Founding Director of the Morgan Advanced Studies Institute.

Professor Tom Rodden FRS FREng FBCS FACM
Executive Management Team (EMT) Chair

Leading expert in developing novel interactive devices and interactive cooperative systems as part of multidisciplinary teams. Professor of Interactive Computing at the University of Nottingham and Pro-Vice Chancellor for Research and Knowledge Exchange.

Professor Dame Wendy Hall DBE FRS FREng MAE FIET
RAi Ambassadors Council Chair

Leading expert in Web Science, Artificial Intelligence, Data Science, and the Semantic Web. Regius Professor of Computer Science and Director of the Web Science Institute at the University of Southampton.

Professor Sana Khareghani
Public Engagement, Outreach and Policy (PEOP) Pillar Chair

Leading expert in policy and strategy at the intersection of technology and business. Professor of Practice in AI at King's College London and formerly Head of the UK Government Office for AI and co-chair of the OECD AI Compute Taskforce, focusing on Responsible AI.

Professor Dame Muffy Calder DBE FREng FRSE FBCS FIEE
Skills Pillar Chair

Leading expert in sensor-based systems, privacy intrusion, and national security. Professor of Formal Methods and serves as Vice-Principal and Head of the College of Science and Engineering at the University of Glasgow.

Introducing our team

Professor Elvira Perez Vallejos

Equities Pillar Chair

Specialist in Responsible Research and Innovation (RRI) and Equality, Diversity, and Inclusion (EDI), Patient and Public Involvement, Data and Research Ethics, and Human Factors. Professor of Digital Technology for Mental Health at the University of Nottingham.

Professor Prokar Dasgupta MSc (Urol) MD DLS FRCS (Urol) FEBU FLS FKC

Health and Social Care Working Group Chair

World-leader in Urology with shaping best practices for AI and Robotic Surgery in healthcare.Foundation Professor of Surgery at King's Health Partners and Chair in Robotic Surgery and Urological Innovation at King's College London.

Professor Joel Fischer

Research Pillar Chair

Leading expert in Human AI Interaction. Professor of Human-Computer Interaction at the University of Nottingham. Coordinating Keystone Lead for RAi UK and Research Director for the UKRI Trustworthy Autonomous Systems Hub.

Professor Kate Devlin

PEOP Pillar Deputy Chair

Leading expert in Human-centred AI and Technology and Intimacy. Professor in AI and Society at King's College London. Advocacy and Engagement Director for the UKRI Trustworthy Autonomous Systems Hub.

Dr Caitlin Bentley

Skills Pillar Deputy Chair

Specialist in AI Education, RRI, Systems Thinking, Feminist AI, and Intersectionality, Critical Theory, ICT4D and Open Development. Senior Lecturer in AI Education at King's College London and Co-Director of the UKRI Safe and Trusted AI Centre for Doctoral Training.

Professor Jack Stilgoe

Ecosystem Coordination Pillar Deputy Chair

Leading expert in Responsible Innovation, Governance of New Technologies, and Public Engagement with Science. Professor of Science and Technology Policy at University College London and a Fellow of the Alan Turing Institute.

Professor Máire O'Neill FREng, FIAE, MRIA

Leading expert in applied cryptography, hardware security and the application of AI in hardware security. Regius Professor in Electronics and Computer Engineering at Queen's University Belfast. Director of the UK Research Institute in Secure Hardware and Embedded Systems (RISE) and the Centre for Secure Information Technologies (CSIT).

Dr Shoaib Ehsan

Equities Pillar Deputy Chair

Specializes in robotics and AI, focusing on vision-based localization and mapping algorithms for mobile robotics and other computer vision systems. Associate Professor of Computer Science at the University of Southampton.

Professor Derek (Mac) McAuley

Special Projects Chair

Leading expert in ubiquitous computing, computer architecture, networking, distributed systems, and operating systems. Emeritus Professor of Digital Economy at the University of Nottingham.

Meet our researchers

Dr Ayodeji Abioye, University of Southampton

Ayodeji Abioye is a Research Fellow in human-swarm interaction at the School of Electronics and Computer Science, developing trustworthy machine-learning algorithms for practical UAV swarm applications such as search and rescue, drone delivery and logistics, and precision agriculture.

Dr Pepita Barnard, University of Nottingham

Pepita Barnard is a Human Factors Research Fellow at Horizon Digital Economy Research Institute, Pepita has both deep expertise and broad knowledge of health and social care domains and emerging technologies. Pepita's inclusive approach addresses RRI, behavioural and technology acceptance theories, trust, trustworthiness, privacy-preservation, recovery from errors and failures.

Dr Aislinn Bergin, University of Nottingham

Aislinn Gomez Bergin is a Transitional Assistant Professor. Her research is at the intersection of emerging technologies like XR and AI, and how these can be responsibly developed, evaluated, and implemented in the context of mental healthcare. She is a co-investigator on the RAi Keystone addressing the socio-technical limitations of Large Language Models (LLMs).

Dr Gisela Reyes Cruz, University of Nottingham

Gisela Reyes-Cruz is a Transitional Assistant Professor in the School of Computer Science, and is a member of the Mixed Reality Lab and Horizon Digital Economy Research. She is researching Human-Computer and Human-Robot Interaction to understand everyday life interactions with, trust in, and public acceptance of, socio-technical autonomous systems.

Dr Jayati Deshmukh, University of Southampton

Jayati Deshmukh is a Postdoctoral Researcher at the School of Electronics and Computer Science. She is part of the UKRI Trustworthy Autonomous Systems (TAS) Hub, RAi UK, and Citizen-Centric AI Systems (CCAIS) teams. Her research interests are in designing autonomous multi-agent systems which act responsibly and are trusted by the people who interface with diverse applications of these systems.

Dr Liz Dowthwaite, University of Nottingham

Liz Dowthwaite is a Senior Research Fellow in the Horizon Digital Economy Research Institute. Her research interests revolve around how motivation, attitudes, and human values relate to behaviour and wellbeing, particularly surrounding online participation, interaction with technology, and how human autonomy can be enhanced or undermined by autonomous systems.

Dr Damian Okaibedi Eke, University of Nottingham

Damian Okaibedi Eke is a Transitional Assistant Professor at the School of Computer Science. His research interests cover critical philosophical issues at the intersection of Technology, Data and Society including; Ethics and Governance of data and AI, Philosophy of Technology, Responsible Innovation and ICT4D.

Meet our researchers

Dr Imke van Heerden, King's College London

Imke van Heerden is a Research Associate in the Department of Digital Humanities, covering AI in the Creative Sector. Exploring the intersection of literary studies, computational creativity and natural language generation, her work emphasises the value of literary perspectives to the responsible design and evaluation of generative systems.

Dr Sarah Kiden, University of Southampton

Sarah Kiden is a Public Interest Technologist and Design Researcher investigating the relationship between technology, design, policy, and society. Her research explores the design of algorithms and interaction mechanisms with non-expert citizen end users to understand and trust future AI systems.

Dr Maira Klyshbekova, King's College London

Maira Klyshbekova is a Postdoctoral Research Associate working as part of the RAi UK Skills Programme. Her research interests revolve around education, language ideology, AI and education, language learning and AI. Her current work is focused on UK AI skills landscape mapping.

Dr Tuyen Nguyen, University of Southampton

Nguyen Tan Viet Tuyen is a New Frontiers Fellow in AI and a Visiting Research Fellow at King's College London. His research centres on human-centred AI, machine learning (ML), computer vision, robotics, and their applications to robot learning from and for human interactions.

Dr Dominic Price, University of Nottingham

Dominic Price is a Research Fellow and Lab Manager for the Cobot Maker Space, a facility for Human-Robot Interaction. He is a software engineer with a background in middleware platform development for human-computer interaction (HCI) research projects and expertise in software and hardware development across a spectrum of collaborative robots.

Dr Lokesh Singh, University of Southampton

Lokesh Singh is a Research Fellow. His Human-Computer Interaction research focuses on understanding induced stress's impact on performance, team coordination, and human-AI collaboration. His expertise spans diverse domains, including Trustworthy Robotics and AI, Signal Processing, Machine Learning, and Neuroergonomics.

Dr Eike Schneiders, University of Nottingham

Eike Schneiders is a Transitional Assistant Professor. His research focuses on Human-Computer and Robot Interaction. His work includes projects at the intersection between Medicine and ML, Law and AI, and Interactive art, AI, and animal welfare. Among his notable achievements is the Cat Royale project that recently earned the prestigious Webby Award.

Dr Aleksandra Landowska, University of Nottingham

Aleksandra Landowska is a Research Fellow. Her research leverages neurotechnology and brain-computer interfaces to enhance human-AI teaming, human-robot interaction, and human-swarm interaction. She is also interested in exploring brain data to understand and improve shared control with AI.

Ops Team

Angela Westley, Head of Operations

Angela Westley, Head of Operations. Angela Westley leads Operations at RAi UK, based at the University of Southampton. She has expertise in research and operations management and has previously managed operations for high-impact initiatives, including the UKRI Trustworthy Autonomous Systems (TAS) Hub.

Dr Ben Coomber, Head of Transformation

Dr Ben Coomber is the Head of Transformation at RAi UK, based at the University of Nottingham. He works closely with RAi UK leadership to drive key activities across RAi's Pillars. He also holds the position of Research Manager at the University of Nottingham, contributing his expertise to strategic initiatives.

Dr Lucy Veale, Head of Projects

Dr Lucy Veale is the Head of Projects, based at the University of Nottingham. She oversees RAi UK's portfolio of funded projects and integrator initiatives led by RAi's researchers. She plays a key role in project administration and is a member of RAi UK's Research Pillar. Lucy provides crucial support across various projects.

Ariba Fatima, Senior Administrator

Ariba Fatima is the Senior Administrator for RAi UK, based at the University of Southampton. She brings extensive administrative expertise across various sectors including education and finance, having worked with multiple institutions including University of Sussex. Efficiently managing administrative initiatives, she is ensuring seamless operations, and supporting project management function.

Danielle Higgins, Finance Coordinator

Danielle Higgins is the Finance Coordinator for RAi UK, based at the University of Southampton. Bringing extensive financial expertise to her current role, with a background in accountancy, she has held finance roles at the University of Sussex and within the NHS at Royal Sussex County Hospital in Brighton.

Kavita Mittal, Communications Manager

Kavita Mittal leads communications for RAi UK , based at the University of Southampton. A trained business manager with a keen understanding of brand building, she has previously managed the InfoCentre at the High Commission of Canada in New Delhi and corporate communications for a leading multinational group in India.

Khalid Parkar, Information Systems Manager

Khalid Parkar is the Information Systems Manager for RAI UK, based at the University of Southampton. With expertise in cybersecurity and system administration, he manages the development and maintenance essential systems such as ERP platforms, cloud services, and communication infrastructure. His competence lies in enhancing system efficiency, managing data effectively, and ensuring compliance with industry standards.

Stefania Fedele, Events Manager

Stefania Fedele is a dynamic events manager, based at King's College London. Bringing a wealth of experience, working across both industry and academia, she excels at managing high-profile B2B and B2C events. She is passionate about connecting industry leaders and academics with top experts.

Tubah Sarwar, Skills and Innovation Manager

Tubah Sarwar is RAi UK's Skills and Innovation Manager, based at King's College London. She supports the Skills Pillar and manages various skills projects, including skills programme funding calls, while also overseeing general project management for RAi UK.

Dr Virginia Portillo, RRI and EDI Coordinator

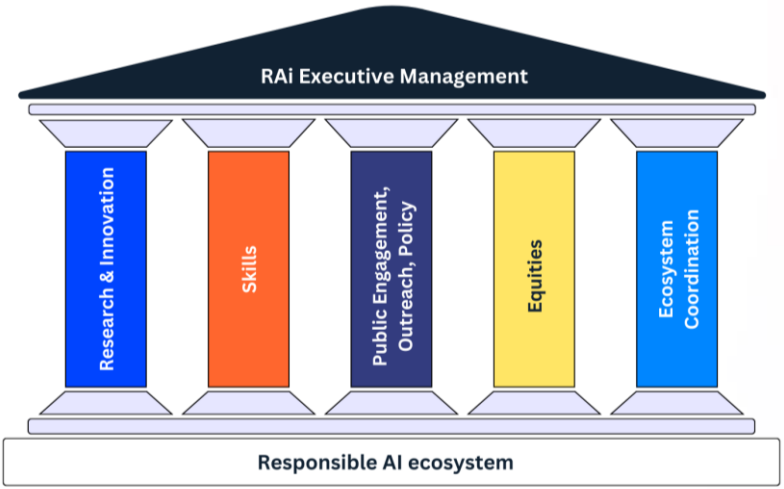
Dr Virginia Portillo is the RRI (Responsible Research and Innovation) and EDI (Equality, Diversity, and Inclusion) Coordinator at RAi UK, based at the University of Nottingham (UoN). She is a member of RAi UK Equities Pillar and supports implementation of RRI and EDI across the programme. She also holds a position of Research Fellow at Horizon DER, UoN. She is motivated in creating opportunities to foster responsible and inclusive AI practice.

Connecting the responsible AI ecosystem

At the heart of our mission for RAi UK lies an ambition to convene and connect a rich and diverse – yet noticeably fragmented – ecosystem of researchers and innovators across the UK. In essence, we believe RAi UK can act as a platform to draw together this community, championing an ethos of working together to develop national conversations about AI that prioritise responsibility and trustworthiness, and elevates end users and societal benefits as the principal concern.

The foundations of the RAi UK programme lie in our functional Pillars, grouping our teams around strategic objectives that deliver activities for:

- Research and Innovation
- Skills
- Public Engagement, Outreach and Policy (PEOP)
- Equities
- Ecosystem Coordination



Our Executive Management Team has defined, and will continue to iterate, a strategy and vision for RAi UK, ensuring activities undertaken across the Pillars are balanced and maximise the return from the resources we have available in the programme.

Within the Pillars, detailed plans are developed for the programme, driving activities, capturing outputs and reflecting on successes and delivery challenges that will enable us to streamline and fine-tune.

To support the Executive Management Team in defining and refining strategic direction, RAi UK has established two influential advisory groups: an Industry Advisory Board (IAB) and our Ambassadors Council. These groups provide platforms to share knowledge, ideas and intelligence that reciprocally benefit RAi UK and member organisations represented in these groups.

The IAB focuses on the relevance and potential impact of RAi activities for a broad set of beneficiaries, across a range of sectors including industry, government, and the public. It will help identify gaps in the responsible AI landscape and provide pathways for RAi UK to address such gaps. Input from the IAB will help to shape research activities, skills programmes, and ecosystem coordination; it will help catalyse collaborations with key parts of the ecosystem through the networks its members are part of.

Our Ambassadors Council, led by Professor Dame Wendy Hall, will connect RAi UK into influential leading organisations around the world concerned with how AI can be developed and deployed responsibly. Gathering global perspectives will enable RAi UK to develop a programme that delivers impactful research beyond UK borders, and contribute to cementing the UK as a global leader in responsible AI research and innovation.

01/05/2023
Milestone: RAi UK begins

07/07/2023
Funding call: Impact Accelerator call launched

11/09/2023
Event: Belfast Townhall

25/09/2023
Event: Glasgow Townhall

30/10/2023
Event: RAi UK at the AI Fringe

01/12/2023
Milestone: Impact Accelerator projects start

01/02/2024
Milestone: International Partnerships start

22/03/2024
Brand: Launch of new RAi UK identity

09/04/2024
Event: Responsible AI Community Building Event

01/05/2024
Milestone: Keystone Projects start

02/05/2024
Event: Integrating EDI/RRI into research

04/06/2024
Event: AI and Net Zero Showcase

14/06/2024
Research: Roundtable with Keystone Projects

27/06/2024
Milestone: AI Champions announced

10/09/2024
Event: RAi UK All Hands Meeting

26/06/2023
Event: London Townhall

31/07/2023
Funding call: International Partnerships call launched

19/09/2023
Event: Creative Industries: Writing & Publishing Roundtable

11/10/2023
Event: Cardiff Townhall

30/11/2023
Funding call: Keystone Projects call launched

04/12/2023
First RAi UK webinar/seminar

22/03/2024
Event: Partner Network Townhall

01/04/2024
Milestone: Satellite Projects start

15/04/2024
Funding call: Skills call launched

01/05/2024
Event: Responsible Tests for the Public Procurement of AI

07/05/2024
Event: CogXFestival

10/06/2024
Research: Integrator Sandpit begins

27/06/2024
Event: AI in Health and Social Care

01/09/2024
Milestone: Skills projects begin

The Ecosystem Coordination Pillar plays a critical role in drawing together the constituent parts of a fragmented ecosystem. Alongside RAi UK, we know there is an extensive community engaged in research and innovation that drives forward a responsible AI agenda. We recognise the importance – and huge potential for added value – of connecting these national investments and initiatives.

Key collaborators in drawing together the responsible AI ecosystem:

BRAID
Bridging Responsible AI Divides

The Alan Turing Institute

Ada Lovelace Institute

Funding responsible AI research and innovation

Supported by UKRI, and led by the Research & Innovation Pillar, our first year has seen RAi UK distribute £14.6M in funding through multiple highly competitive funding calls, allowing us to fuel the ecosystem and encourage its disparate parts to come together and collaborate.

In our capacity as a funder, RAi UK has sought to bring together academic researchers with industry and civil society organisations. We have engaged openly with communities across the UK to crowd source pivotal themes, and to identify gaps and critical research areas to prioritise. These have ultimately informed the design of funding initiatives that we believed could best support the responsible AI ecosystem.

Extensive consultation with the ecosystem began with a series of open RAi UK Town Hall events and roundtable discussions held in London, Belfast, Cardiff, and Glasgow between June and October 2023.

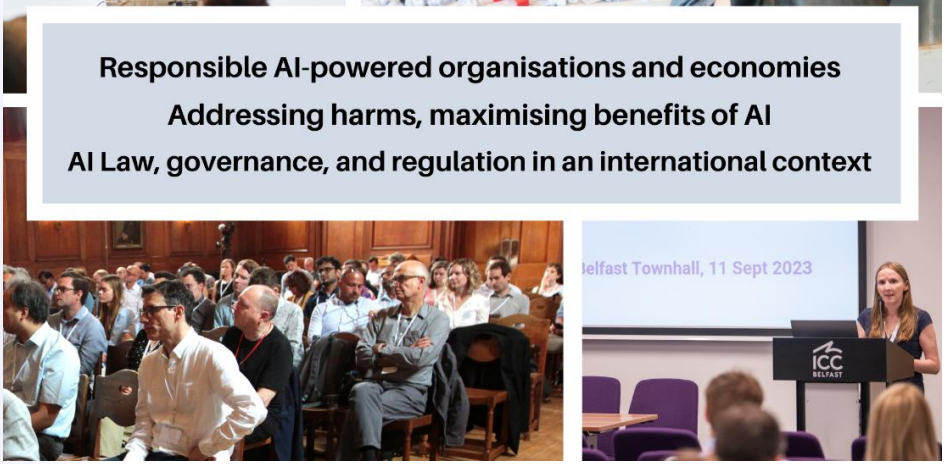
The challenges and community consensus articulated at four townhalls, 22 roundtables, and multiple conversations via various other channels with expert stakeholders in the field, including from our Ecosystem Coordination Pillar, industry, and parts of government, directly influenced our funding calls, and underpinned the thematic areas we identified, particularly for our flagship Keystone Projects funding call. In this regard, the identity of RAi UK has truly been co-created with and represents a wider community.

Community input was pivotal in defining three thematic areas to be addressed by our Keystone Projects.

01 Responsible AI-powered organisations and economies:
What sector- and context-appropriate responsible AI principles for safe and trustworthy AI provide confidence to organisations, customers, stakeholders, economies, and the wider society?

02 Addressing harms, maximising benefits of AI:
What is needed to ensure AI works for the benefit of people and societies, while harms are minimised?

03 AI law, governance, and regulation in an international context:
How do we promote trust, fairness, and accountability for users, and provide certainty for international commerce?



To complement our portfolio of research, we have also invested in a cohort of Skills projects that will commence in the next year of the programme, coordinated by the Skills Pillar. These projects, and indeed the wider work of the Skills Pillar, focus on cultivating a robust, responsible AI skills base in the UK.

We are committed to being inclusive and transparent, adhering to the principles of open science. Fundamental to the way we operate in RAi UK is our belief that we are here to convene, not to compete. Our aim is to broaden and strengthen the wider ecosystem of people who develop, deploy and use AI technologies in ways that are trustworthy, secure, and benefit people in their everyday lives – in the UK and beyond.

Multidisciplinary by design

Successfully realising, and indeed maximising, the potential societal benefits of AI in different contexts demands a range of perspectives, of specialisms, and of experiences. We view inclusivity and multidisciplinarity as being directly linked to our commitment to excellence; many of the risks and opportunities of the technologies we study arise from the diverse contexts in which they are developed and deployed.

Our Equities Pillar sets the tone for how we approach activities inclusively and responsibly. This pillar oversees our Equality, Diversity, and Inclusivity (EDI) metrics and strategy, RRI strategy, and Trusted Research. Inclusivity covers people, places, and research disciplines.

We want to encourage computer scientists and engineers to work with researchers in the arts and humanities, social scientists, clinicians and beyond.

We want academics to listen to and build productive working relationships with colleagues from industry, government, and civil society. We intend to cultivate an ecosystem where people feel seen and heard. This is especially true when those voices come from less established researchers, or when they challenge our assumptions about what makes AI trustworthy and responsible.

Developing strong and sustainable discipline-spanning collaborations between researchers takes considerable time and effort, and above all else an openness to new ideas and perspectives. We expect to see strong multidisciplinary collaborations in proposals submitted to RAi funding calls. We make a virtue of inclusivity and providing fair opportunities for all.

RAi Community Footprint

		Connect Newsletter	RAi website	RAi events
3,500+ Followers	2,000+ Followers	1,350+ Subscribers	3,600+ Monthly visitors	6,250+ Attendees

Connecting the ecosystem

During the first year of RAi UK, we have strived to facilitate networking for the extended responsible AI ecosystem led by our Public Engagement, Outreach & Policy Pillar, be that in person at our Town Halls or at bespoke events targeting specific subgroups within the ecosystem, or in other ways such as through our regular online seminar series.

In April 2024 at the Royal Academy of Engineering, we hosted over 80 colleagues from the newly announced UKRI AI Hubs and UKRI Centres for Doctoral Training in AI to encourage conversations and foster cohesion and community engagement.

> *We had very constructive conversations about how we can take what we have learned today back to our institutions. We have agreed that we should all share resources and develop sharable examples of good practice.*
>
> Prof Selin Ahipasaoglu, University of Southampton, representing the UKRI AI CDT in AI for Sustainability.

> *The RAi network can provide a lot of ways to validate whether what we develop is explainable, trustworthy, and responsible. There is already fantastic expertise in the network, and we can work all together to enhance the productivity of all the different hubs.*
>
> Prof Themis Prodromakis, University of Edinburgh, representing the AI for Productive Research & Innovation in eLectronics (APRIL) Hub.

Community building with the new UKRI AI Hubs and UKRI Centres for Doctoral Training in AI, and connecting with researchers working on AI & Net Zero, and AI in Health & Social Care.

We have also convened domain specific events and workshops. In June 2024, we brought together subject experts at the UKRI AI & Net Zero Collaborative Showcase, and later in the month at our UKRI AI in Health & Social Care Symposium, attracting 77 and 130 delegates, respectively.

The latter of these events was organised by our Health and Social Care Working Group, one of four Working Groups we have established to coalesce experts and address challenges within their domains.

Defence & Security, and the Judicial System

Brings together experts from domains where AI is increasingly used, posing technical and socio-technical challenges we are not yet equipped to address.

Led by Prof Dame Muffy Calder

> *A common research challenge for our three distinct, but overlapping, communities is presenting and evaluating probabilistic outputs in sensitive environments. We are not aware of readily identified useful taxonomies for this sector. While approaches from other jurisdictions may not be relevant to the UK, they may inform research. We are considering how to convey our challenges to the key departments HO, MoD, and DSIT.*

Health and Social Care

Brings together experts to address the design and use of AI in health and social care – particularly Large Language Models and robotics - aligning with the WHO's Global report of AI.

Led by Prof Prokar Dasgupta

> *The potential benefits of artificial intelligence in health and social care are enormous so long as we engage with it responsibly, keeping our patients at the heart of it. We are looking for deployable AI tools to augment healthcare and standardise their procurement and implementation. We have also recently launched an ambitious project of AI Champions in every hospital in the UK at the Royal Society."*

Public Dialogue

Brings together stakeholders at key organisations to maximise the value of public dialogue on AI. Connected to the Satellite project "Public Voices in AI".

Led by Prof Jack Stilgoe

> *Members of the public are currently being shut out of conversations about AI that could be hugely consequential for their futures. We're trying to change that by looking for new approaches to public participation in AI.*

AI, Law and Governance

Brings together stakeholders who conduct legal research, support regulators, and build the participation of the public into evolving AI governance.

Led by Prof Sana Khareghani

> *Governments globally are trying to understand how best to tackle AI regulation. Work is still required to bring the views of researchers, regulators, lawyers, policy makers, standards agencies, industry as well as civil society together to this right. This group is asking some of the hard questions to represent the collective voice.*

RAi UK AI Champions

Our flagship RAi UK AI Champions initiative, led by the Health and Social Care Working Group, aims to drive responsible development and adoption of AI technologies across the health and social care sector. RAi UK AI Champions combine frontline healthcare experience with AI expertise, with these individuals ideally placed to play a vital role in helping ensure AI solutions are effective, relevant, equitable, and address real world needs.

> *In healthcare, current models are unsustainable, and we must redefine pathways and delivery systems, integrating AI and other technologies to ensure high-quality care. Our collective effort involves engaging in purposeful conversations across academia, healthcare, policy, and the public, advocating for technological solutions and governance models that secure a sustainable future. Together, with our AI Champions, we must shape and advance this agenda, ensuring government and policy makers understand the imperative of beginning this transformative journey.*
>
> Professor Ian Abbs
> *Chief Executive Officer, Guy's and St Thomas' NHS Foundation Trust.*

RAi UK AI Champions, left to right: Prof Nitin Shrotri, Mr Nicholas Raison, Prof Radu Calinescu, Prof Mohamed Ibrahim collected on behalf of Prof Nikhil Vasdev, Dr Ol'tunde Asháolu, Dr Anthony Wilson, Mr Grigorios Athanasiadis, Prof Naeem Soomro, Dr Clarissa Carvalho, Prof Hutan Ashrafian, Dr Ganesh Vigneswaran collected on behalf of Prof Tim Underwood, Prof Elvira Perez Vallejos.
Pictured with: Prof Prokar Dasgupta, Prof Sarvapali (Gopal) Ramchurn, Prof Ian Abbs.

Influencing the national conversation on AI

Representing RAi UK, our team has made broad contributions to the rapidly evolving conversation around AI, notably as co-founders of the AI Fringe. Conceived to complement the AI Safety Summit held in London in the autumn of 2023, the AI Fringe brought together industry, civil society and academia, providing a platform for all communities to engage in discussion around both the potential risks and benefits of AI. This resulted in a white paper written by a consortium drawn from attendees of the AI Fringe event.

The AI Fringe white paper can be read in full on aifringe.org/#perspectives.

Subsequently, RAi UK joined the AI Fringe Seoul Summit in June 2024 as a partner. This event brought together a diverse range of voices from the UK, the Republic of Korea and France to reflect on the priorities and outcomes emerging from the AI Seoul Summit 2024.

RAi UK has also provided responses to both national and international policy consultations.

Prof Dame Muffy Calder and Prof Dame Wendy Hall provided evidence to the House of Lords Communication and Digital Select Committee in relation to an inquiry into Foundation models and LLMs. Furthermore, Prof Gopal Ramchurn and Prof Dame Muffy Calder gave evidence to the House of Lords Committee on AI in Weapons Systems, following which Prof Dame Muffy Calder was appointed Special Adviser, House of Lords Committee on AI in Weapons Systems.

We responded to a consultation launched by the UN Secretary-General's AI Advisory Body on their Interim Report: Governing AI for Humanity, which calls for a closer alignment between international norms and how AI is developed and rolled out.

Prof Gina Neff contributed significantly to the inception of the first Trade Union Congress AI Bill that aims to regulate the use of AI systems by employers in relation to workers, employees, and job seekers, to protect their rights and interests in the workplace.

Evolving the responsible AI ecosystem

Our long-term goal is to enable the wider community to leverage the convening power of RAi UK, giving voice to all perspectives in the ongoing debate around how AI is developed, when and where it should be used, and what measures are needed to ensure widespread benefits and minimise the risk of unintended harms.

Priorities for the Public Engagement, Outreach and Policy Pillar in 2024/25:

- Raise awareness of RAi UK and responsible AI more broadly, translating and communicating key messages and findings from our projects to policy makers in the UK and further afield.
- Deliver coordinated public engagement across the portfolio, utilising expert knowledge within our project teams, informing national surveys on attitudes to AI, and hosting a range of creative events to aid communication.

Research & innovation

Cultivating a portfolio of responsible AI research

In our first year, the RAi UK team has invested considerable effort to pull together an exciting portfolio of research that now includes 23 live projects, generated from four separate funding streams: Impact Accelerator, International Partnerships, Keystone Projects, and Satellite Projects.

Informed by extensive consultation undertaken at Townhall roundtables and other events, and coordinated by the Research & Innovation Pillar, all RAi UK funding calls have proved extremely popular, attracting more than 370 applications across all schemes. Our 23 funded projects involve researchers from 28 different UK universities, many working in partnership with Higher Education Institutions internationally, and alongside >150 non-academic partner organisations.

3

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, and financial services.

11

International Partnerships

Partnerships that develop strategic collaborations with world-leading research organisations to ensure society deploys and uses AI in a responsible way, beyond national boundaries

7

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.

2

Satellite Projects

Funded through additional investment by UKRI. The Satellite Projects are looking at how responsible AI can help drive productivity and how public voices can be amplified in the design and deployment of AI technologies.

View from our Research & Innovation Pillar Leads

Across all the funding schemes RAi UK has delivered in the past year, the quality of proposals has been extremely high and, consequently, competition fierce. As Research & Innovation Pillar Leads for RAi UK, we would like to take this opportunity to thank all who invested significant effort applying for RAi UK funding, and to acknowledge the valuable contributions of the wider community in participating in our review processes.

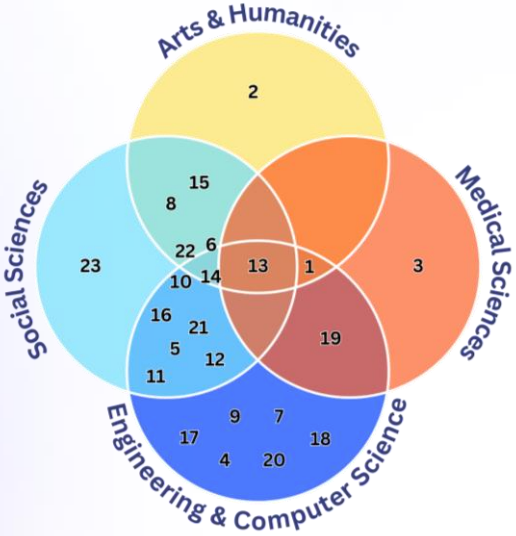
More information about the funding calls and the processes we follow is available on our website.

As the project portfolio has taken shape, it has been hugely rewarding to see widespread appreciation of the advantages that multidisciplinarity brings, a deep commitment to RRI and increasing diversity of voice as the number of research groups and universities involved continues to grow.

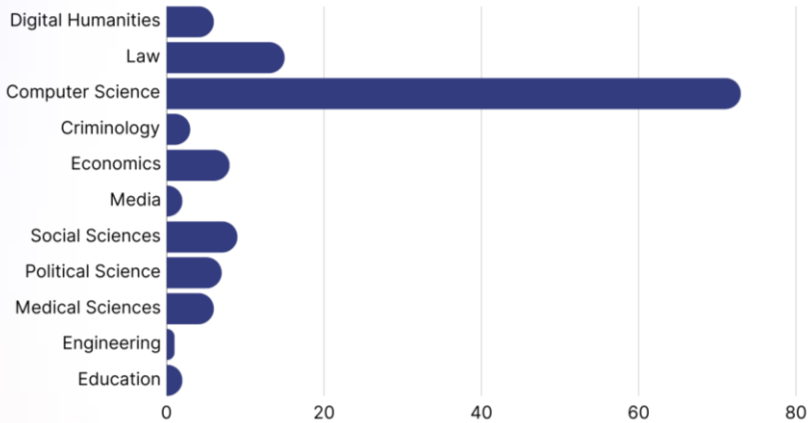
Project teams have demonstrated fantastic commitment to building the RAi UK community thus far, attending and participating in online and in-person RAi UK events, representing RAi UK at events organised by others in the ecosystem, and by engaging with our Working Groups. As they host their workshops and continue developing a range of ambitious outputs – from open-source databases and practical toolkits to guidelines and policy documents – we look forward to learning about, promoting and celebrating their successes over the coming year.

Joel Fischer and Matt Jones, Research & Innovation Pillar Leads

UK Universities formally involved in RAi UK, including Impact Accelerator Projects, International Partnerships, Keystone Projects, and Satellite Projects, and the core RAi UK programme.



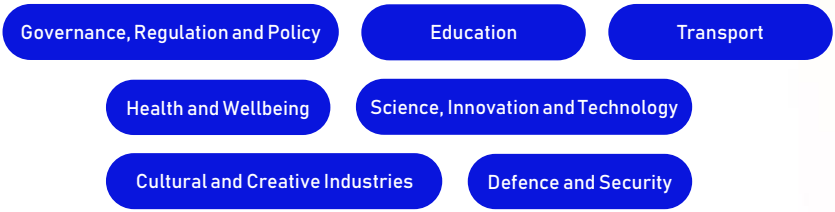
1. AdSoLve
2. AI Equality by Design
3. Ethical AI for healthcare
4. AI Regulation in Safety-Critical Systems
5. AI for electoral oversight
6. AIPAS
7. Amplify
8. AEGIS
9. MusicRAI
10. DOMINOS
11. Economic Landscape Analysis
12. For the FATES of Africa
13. PHAWM
14. PROBabLE Futures
15. Public Voices 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



Multidisciplinarity in the RAi UK project portfolio. (Top) 23 RAi projects plotted according to disciplines of researchers in the respective teams, showing the breadth of collaborations between Engineering and Computer Science, and Social Sciences, Arts and Humanities, and Medical Sciences. (Bottom) Prominent disciplines represented across a community of >130 researchers supported by RAi UK.

The RAi UK project portfolio

RAi UK projects vary considerably in their aims, size, scope and duration; they are also at different stages in their lifecycle, dependent on the funding stream. In many cases, these projects are delivering research, knowledge exchange, and impact that transcends sectors, with extensive possibilities for cross-sector benefits. Here, we present the project portfolio categorised according to primary sector for impact.



We also group our projects around a series of thematic clusters. These clusters – which were conceived and shaped with extensive input from the community at our Townhall events and other roundtables – capture the cross-sectoral nature of the RAi UK project portfolio, and provide a flexible means to explore important connections, synergies and divergent views across the portfolio.

RAi UK THEMATIC CLUSTERS

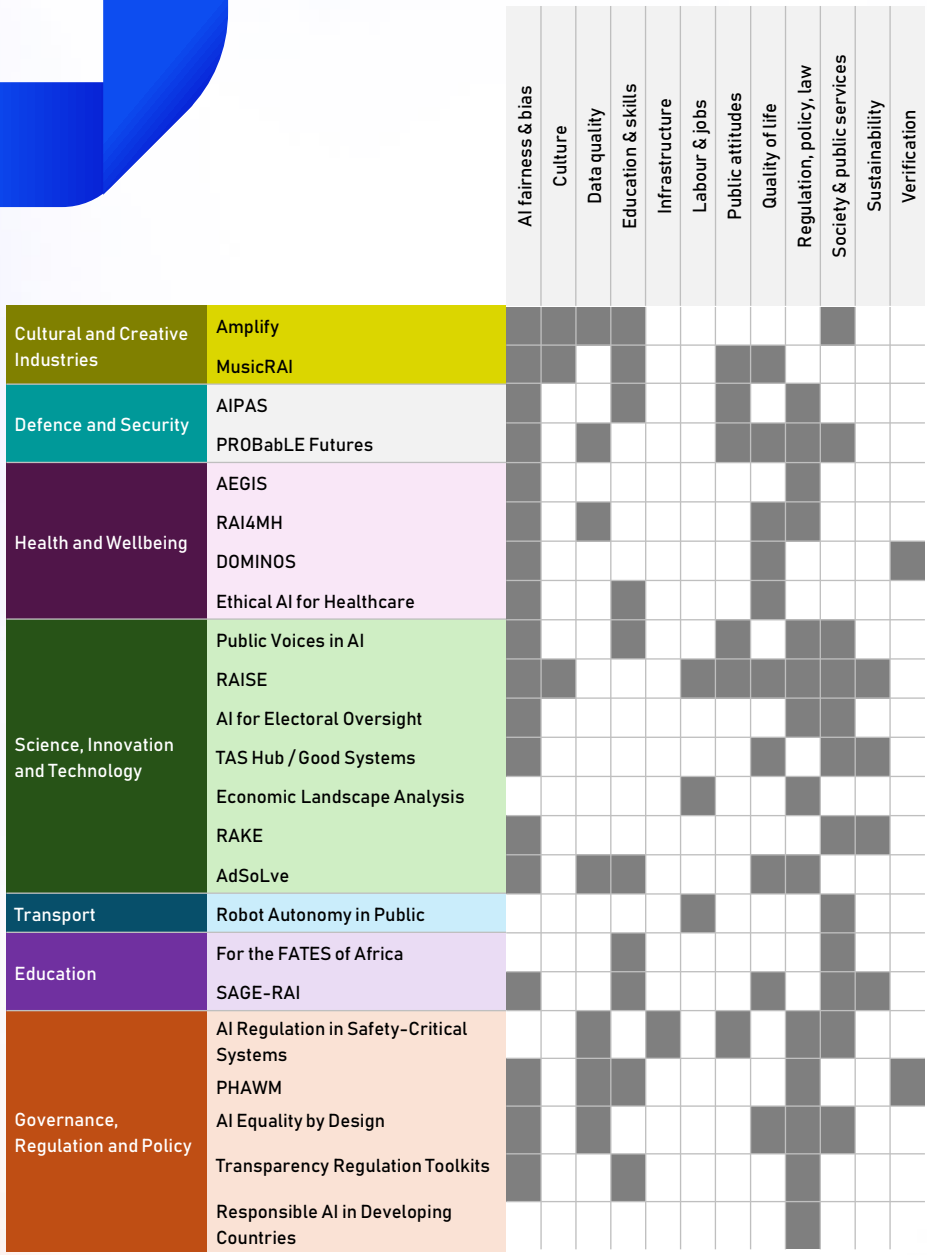
AI fairness & bias Interpretability, explainability, transparency, trustworthiness of AI	**Culture** Cultures and languages within and across borders	**Data quality** Data provenance, reliability, diversity, representativeness etc.	**Education and skills** Broadly increasing understanding of AI
Infrastructure The infrastructures that enable AI	**Labour & jobs** Labour rights, job displacement, automation, productivity	**Public attitudes** Understanding public views of AI	**Quality of life** Wellbeing, human rights, health
Regulation, policy, and law Principles, recommendations, standards, and laws	**Society & public services** Social contract, civic engagement, societal norms, social justice	**Sustainability** Sustainability of resources required to build and deploy AI	**Verification** Formal mathematical validation and verification

We will use this sectoral taxonomy and the thematic clusters to encourage conversation and support collaboration between our projects and stakeholders from the wider ecosystem, maximising impact and the return on UKRI investment in RAi UK.

Detailed project descriptions can be found in the RAi UK Project Index at the end of this report.

RAi UK THEMATIC CLUSTERS

RAi UK SECTORS



Cultural and Creative Industries: The UK has a world leading Creative Economy. RAi UK projects in this space are working to reduce bias in AI music generation [MusicRAI], and in Spoken Language Interaction systems [Amplify]. In addition to raising awareness of bias in both arenas, this research will dramatically diversify available music and language datasets. In turn, this will improve access to and increase the audience for marginalised music genres and spoken language technologies.

Defence and Security: RAi UK research in this sector will support law enforcement actors to fully realise the opportunities of AI in ways that uphold the values of the criminal justice system. Our project teams are working on the production of practical tools for policing, security and the court room [AIPAS], and a comprehensive framework for the use of responsible AI in law enforcement and criminal justice [PROBabLE Futures].

Health and Wellbeing: AI is emerging as a tool with the potential to transform mental and physical healthcare services. RAi UK projects will support the delivery of responsible AI solutions in healthcare capable of mitigating disruptions [DOMINOS], and explore 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 being deployed transparently? [Ethical AI for Healthcare] RAi research will develop new international technical standards for AI systems that purport to emulate human empathy [AEGIS].

Education: RAi UK projects are exploring the responsible use of AI tools in educational settings, demonstrating how educational providers can best use Generative AI to enhance student outcomes (with an emphasis on integrating AI into existing educational courses) [SAGE-RAI], and focusing attention on online learners in Sub-Saharan Africa through publication of an open access toolkit for protecting and championing the use of responsible AI for a population that is often left behind [For the FATES of Africa].

Governance, Regulation and Policy: RAi research will support 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 will identify gaps in AI safety, assurance and certification, specifically in the context of safety-critical systems [AI Regulation in Safety Critical Systems] and has the potential to transform AI auditing [PHAWM].

Science, Innovation and Technology: 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 will generate evidence for responsible AI as a driver for productivity and economic prosperity in the UK, and internationally. Teams are analysing the use of AI across economic sectors in the UK [Economic Landscape Analysis], developing resources to support SMEs implement generative AI tools responsibly in the UK and Africa [RAISE], creating an ambitious evaluation benchmark for LLMs that will facilitate robust self-assessment by industry and third sector partners [AdSoLve], and identifying opportunities for AI tools to uphold, rather than undermine, democratic practice in electoral processes [AI for electoral oversight]. Projects aim to ensure full consideration of public voice and will generate new data on public attitudes to AI [Public Voices in AI].

Transport: Considering both existing and emerging technologies in transport and communications, in this sector RAi UK research foregrounds everyday interactions with technologies and issues of public trust and safety. Findings from RAi UK research will inform the future design and regulation of publicly-interfacing robotics [Robot Autonomy in Public].

Coordinating Keystone

The fulcrum around which the portfolio of projects will flexibly coalesce is our Coordinating Keystone (CK).

Drawing together a talented cohort of early career researchers, and guided and supported by our large, experienced team of academics, the CK will develop a programme of research that complements and enhances the wider portfolio.

Joel Fischer, CK Lead

UKRI Trustworthy Autonomous Systems Hub

A series of underpinning research projects – specifically designed to transition into RAi UK research and minimise lead times on our CK research programme – were commissioned in April 2023 by the UKRI TAS Hub.

Go Ahead I'm Listening (TAS-GAIL)

Would you trust a robot more if it exhibits more human-like behaviour?

TAS-GAIL explored this question by building a human-like, active listening demonstrator based on the MiRo-E robot, with positive non-verbal feedback, such as nodding and blinking.

The team tested whether exhibiting these behaviours is likely to lead people to over trust or overshare and have produced an open-source system for the wider TAS community to build robots with adequate human-like behaviours and published guidelines to ensure robotics technologies are ethical by design.

REGALS: Responsible Employment of Generative AI for Legal Services

As LLMs become increasingly sophisticated, AI has the potential to impact on many professions such as psychotherapy, social services, and legal counsel.

REGALS focused on examining public willingness to use LLMs for generating legal advice and, crucially, willingness to act on advice in low-stakes legal contexts, working with legal professionals to assess the robustness of advice provided. We sought to test trustworthiness of LLM-generated legal advice, aiming to improve understanding of how to make generative AI more responsible.

Telepresence Robot Playground

Telepresence robots – semi-autonomous mobile devices that provide remote access – are being used in museums and workplaces; however, integration is challenging due to a variety of technical, infrastructural, and social factors.

This project engaged with the public to co-explore the use of telepresence robots, considering issues around trust, accessibility, and autonomy, culminating in a robot deployment in a local art gallery.

TAS RRI Legacy & Impact

RRI is at the heart of TAS. This project consolidated and disseminated experiences and resources for responsible innovation from the TAS Hub and other affiliated networks.

Workshops with the TAS community and partners fed into new guidance on embedding RRI in research activity, and the launch of the R3 network, which brings together researchers, artists and professionals from the Arts and Humanities, Design, and Technology disciplines.

A crucial aspect of the work done by the CK team will be to responsively and creatively identify gaps and potential synergies within the portfolio, and to respond to emerging opportunities in ways that integrate new collaborators and partners through co-creation activities. The CK will also play a key role in avoiding duplication of effort, encompassing both the RAi UK programme and the wider ecosystem, ensuring RAi UK delivers excellent value from this UKRI investment.

In practical terms, we will carefully coordinate activities across our three Keystone Projects – PROBabLE Futures, PHAWM, and AdSoLve – and the CK team.

We will work with our Impact Accelerator projects and International Partnerships to gain a deep understanding of the work these teams carry out: the research methodologies they use, the engagement they undertake with stakeholders to identify needs, and the value they add to the current state-of-the-art through their project outcomes and outputs.

We will undertake landscape mapping to understand and generate oversight of responsible AI research and innovation as it rapidly evolves, sharing this openly with the community to improve coherence across the ecosystem.

We will convene workshops and sandpits that generate new projects, crafted to meet emergent challenges and address pivotal questions for responsible development and deployment of AI.

Integrator Sandpits

Building on our experience from the TAS Hub and other aligned programmes, we use the EPSRC Sandpit mechanism to construct new multidisciplinary teams across universities that initiate and deliver research projects. In June 2024, we held our first RAi UK Integrator Sandpit, welcoming 27 researchers from across the programme to network, establish new collaborations, and design innovative new projects.




Integrator projects will be agile and responsive to arising societal needs and emerging research, as well as involving stakeholders in their co-creation and iterative development. Led by early career researchers, these projects will be integrative or additive, either bridging existing investments to address gaps or develop novel areas of research complementary to existing projects.

At the Integrator Sandpit, we challenged the group to focus on a pivotal question:

How can we bring delight, flourishing and fulfilment to individuals, communities, organisations and societies through the responsible design, deployment, and use of AI?

Emerging projects build on RAi UK priorities, and – in keeping with the ethos of the programme – teams will be multidisciplinary and cross-institutional, capitalising on our core research strengths such as Human-machine teaming, Science and Technology Studies, Digital Humanities, RRI, Human Robot Interaction, Multi-agent Systems, Robotics, Policy, Health, Law & Governance, and Adoption of AI.

Embodied AI in Social Spaces: Responsible and Adaptive Robots in Complex Settings

An ageing population is putting a strain on healthcare systems globally, creating a need for innovative solutions. This project will develop novel approaches for integrating adaptive multi-human multi-robot (MHMR) systems into complex social environments, specifically in healthcare settings.

ARTificial Intelligence: Show Me a Story about AI

Using AI stories authored by a variety of creative communities, this project will examine the values impacted by, and shaping responses to, (generative) AI. The research will culminate in the creation of artistic pieces to engage the general public in diverse reflections on AI and its impact.

CulturAI: Co-development of Culturally Sensitive Approaches for Prompt Engineering

CulturAI will explore culturally inclusive frameworks for prompt engineering. By incorporating cultural expertise and community input, the research will explore methods to ensure that AI outputs are not only accurate and relevant but also culturally respectful and inclusive.

Priorities for the Research and Innovation Pillar in 2024/25:

- Support and guide our diverse portfolio of projects, delivering world leading research outputs and impact that benefits society.
- Cohere the research portfolio and identify gaps and new opportunities via the Coordinating Keystone, including launching a series of new projects led by RAi researchers.
- Raise the profile of responsible AI research. Look out for the return of the webinar series in the autumn, the launch of new project pages on the RAi UK website, and an Impact Showcase to celebrate the work of our Impact Accelerator projects as they draw to a close.

Skills

Building a robust responsible AI skills base

Dame Muffy Calder, Skills Lead

The RAi UK Skills Pillar focuses on delivering impactful initiatives that cultivate a robust, responsible AI skills base in the UK. We aim to foster a thriving ecosystem of responsible AI education innovators. We will do this through Convening responsible AI skills resources and leadership across the RAi ecosystem, Accelerating innovation in responsible AI skills policy and practice, and by Supporting the ecosystem to provide practical experience and opportunities.

Core to our aspirations for the RAi UK Skills Pillar are the principles of equity and fairness in training and development opportunities. The initiatives we support are designed to empower a broader community of stakeholders, creating new opportunities for them to harness the full potential of AI in a manner that is safe, inclusive, and beneficial for all.

RAi Skills Pillar Areas of Engagement:

1. National responsible AI skills frameworks
2. Upskilling and reskilling, within and across sectors, with a focus on responsible AI
3. Raising public awareness of responsible AI
4. Addressing equity and disadvantage in AI through education

Co-creating an evaluation framework for Responsible AI skills resources

As one of our first initiatives, we brought together a diverse and multidisciplinary group of RAi UK community interest holders, including academics and partners, to explore what it means to deliver effective skills resources. An explosion of new paid and free online and face-to-face training opportunities for responsible AI skills have been launched in the past year, offering a wide range of options for learners to access responsible AI skills training. Yet, what makes RAi training effective? And, for whom? Who is being targeted, and who is excluded? How can learners, businesses or organisations trust they are making good choices when it comes to selecting what to learn and how to learn it? In our first year, we sought to bring the RAi ecosystem together to provide recommendations on what effective RAi training looks like, and to better understand how we can do better. This exercise ultimately benefits our community by providing guidance on effective practices, highlighting gaps that need to be addressed, and revealing opportunities for improved collaboration across the RAi ecosystem.

For the multidisciplinary review, 21 participants from 13 UK universities, more than 18 disciplines, policy, civil society, and industry perspectives were represented. From a cursory collection of RAi mapping, we selected four different types of publicly available resources for review by our participants.

We held a series of five workshops, first to examine the resources in depth, and then to analyse the results of our discussions and findings. We examined various facets to each resource, including the way it was described to users, intended audience, clarity of learning objectives, and overall design. A series of central themes emerged from across our discussions at the workshops, which provided key areas for inclusion in our evaluation framework. Our framework was further refined and tested through our landscape mapping exercise presented below.

Central Emerging Themes: *What are the main issues with existing resources?*

Lack of engagement with Responsible AI: Responsible AI and AI Ethics are not always synonymous - yet are often treated as so. Clear scope or definitions are rarely given, but are often implicitly assumed.

Starting with the technology rather than the context: Our reviewers often found it unhelpful to describe, in a disjointed manner, technical concerns and their societal implications or impacts in following or subsequent sections. They suggested that RAi training should start first with the context to clearly articulate RAi within that context.

Learning design is important: Learning Objectives were often unclear, with limited guidance as to how learners are supposed to engage with the materials to achieve meaningful outcomes. Occasionally, the content of resources were found to be inaccurate or misleading by reviewers.

Too general or self-interested to be effective: Resources reviewed often provided high-level descriptions of RAi principles but did not provide adequate practical insights or guidance. Frequently, the interests or qualifications of the providers were called into question by reviewers.

Mapping responsible AI skills resources in the UK

To better understand existing resources, our researchers carried out a landscape mapping exercise on a range of materials to test our evaluation framework to create benchmarks on content and quality, to identify gaps and good practice.

Through a systematic keyword search methodology, we combined 17 variations of 'Responsible', 21 variations of 'Artificial intelligence' and 13 variations of 'Training,' identifying 250 resources by searching Google. We then documented basic details about these resources for further screening to identify resources that could be benchmarked using our evaluation framework. For example, we excluded resources that were non-UK based, returned errors when browsing, or did not explicitly mention any combination of our keywords relating to 'Responsible' and 'Artificial Intelligence' together. For example, it was common for many resources to have within their keywords 'Responsible AI' but in the content of these courses to refer only to technical topics of AI, without explaining how or why these topics relate to or reinforce RAI.

Although our landscape analysis is ongoing, some of our early findings are pointing to gaps that we will aim to address through our RAi UK investments, and in collaboration with others. For instance, school educators and health professionals are underrepresented in the identified resources. Through our Skills Projects described below, we will test and research better learning and teaching methods for specific audiences and sectors.

If you have a programme you would like reviewed or would like to work with us to improve your materials or delivery strategies, do let us know. Likewise, if you have resources you would like to share to the wider ecosystem, we will also be sharing effective resources that meet our benchmarks on our website in the coming year.

Accelerating RAi Skills Innovation

We issued a funding call in April 2024 inviting the community to submit proposals that develop open educational resources for the training of AI developers, users, managers, the public, and other stakeholders. These projects will respond to critical challenges relating to upskilling, reskilling, pedagogy, or educational delivery across sectors or diverse populations.

The outcomes and resources created by our Skills projects will be made available openly to the RAi UK ecosystem, delivering impact by expanding AI skills beyond technical practitioners and promoting the upskilling and reskilling of diverse citizens and businesses.

RAi UK Skills Projects

We are delighted to introduce six new projects in the RAi UK portfolio that will have impact inside and outside of the classroom.

Further details about our cohort of Skills Projects can be found in the RAi UK Project Index at the end of this report.

Exploring AI: From Picture Books to Epic Games

Enhancing AI literacy and public awareness of responsible AI through the creation of new educational resources including an A-to-Z picture book and scenario based megagame.

Generative Artificial Intelligence Skills in Schools (GenAISis)

Amplifying student voice in discussion of the responsible use of Generative AI and co-creating an open educational toolkit with school librarians.

AI in Secondary Schools: Resource Development and Teacher Professional Learning

Supporting teachers in developing the skills and knowledge to incorporate Generative AI in their classroom teaching and learning responsibly.

Ways of (Machine) Seeing: Towards a new visual literacy in AI

Providing best practice for the use of Generative AI to deliver an enhanced visual learning experience for students.

Misinformation/disinformation and Generative AI: Building Young Refugees' Skills and Capacities

Equipping young refugees and asylum-seekers with the knowledge, skills and resources to be champions of Responsible AI within their communities.

AI, Law and Legal Training (AILLT)

Co-produced research informed resources that will enhance knowledge and awareness of, and confidence in the use of GenAI for understanding legal processes and accessing legal information.

Skills Pillar priorities for 2024/25:

- Focussed push to rapidly kickstart our newly funded Skills projects, connecting teams to other RAi UK activities and projects.
- Build on successful community building activities, collaborating with the Alan Turing Institute and AI CDTs to disseminate and share best practice in training, e.g., for schools, UK Higher Education Institutions, and the NHS via our AI Champions initiative.

Equities

Embedding responsibility and equity

When we talk about acting responsibly with AI, we envision an ambitious end goal: widespread acceptance of and adherence to the principle that the development and deployment of AI must benefit everyone and ensure it does not cause harm or unfairly disadvantage people.

In RAi UK, we believe that considering Equality, Diversity and Inclusion (EDI) and Responsible Research and Innovation (RRI) principles will be fundamental to developing and deploying responsibly sourced AI.

The Equities Pillar is on a mission to establish a culture that values diverse needs and promotes the inclusion of diverse voices as normal practice in the rapidly evolving conversation surrounding AI.

We insist that RRI and EDI are not treated as 'tick boxes' but seen for the opportunity they represent, to creatively anticipate and reflect on the issues that emerge when AI systems are trained on narrow datasets or developed and deployed by groups of people with insufficient diversity.

Our mission to promote EDI and RRI values and develop best practices across the RAi ecosystem and beyond begins on our own doorstep. Within our consortium, we assess representativeness and diversity of committees and working groups to ensure we include members from different backgrounds, disciplines, gender and seniority levels. Funding review panels are constituted in a similar way, further promoting fairness and equity in resource allocation.

It is a joy to feel we are leading by example. The recently awarded RAi UK Keystone Projects are all led by women. While gender wasn't directly a factor in selecting these projects, RAi UK's deep commitment to EDI manifested in a diverse applicant pool. This serves to emphasise how supportive, considered processes can bridge gender gaps. Women leaders in AI bring unique perspectives that influence the design, development, and implementation of responsible AI. They champion gender-sensitive approaches and promote fairer data science and ML practices. This, in turn, leads to more inclusive, representative, and impactful AI solutions. From an equity standpoint, this is a significant step towards a responsible and ethical future powered by AI.

Elvira Perez,
Equities Lead

From a personal perspective, an upcoming highlight will be an event in September, where RAi UK will collaborate with Technovation Girls in New Delhi, India. The goal: introduce young women and girls to AI and RRI concepts. Increasing female participation in STEM fields is of paramount importance. By doing our part to ignite their interest in AI, computer science, and engineering, we will contribute to empowering future generations of young women to become innovators and leaders in responsible development and deployment of AI.

During our first year, we have engaged extensively with stakeholders and the public, and this will continue throughout the life of RAi UK. From roundtables at the Royal Society (London) to grassroots community talks at the Heart Centre (Leeds), RAi UK aims to extend its reach and access audiences that traditionally have been excluded from the debate on AI. We are therefore bringing the RAi UK programme to non-conventional audiences, at art festivals and museums for example, in addition to academic and policy contexts. RAi UK is here to engage meaningfully on the national and international stage.

As we plan events and other outreach activities for the next phase of the programme, we are carefully considering their audience and potential accessibility barriers in choosing locations and venues. Our goal is to ensure we continue to build in geographical diversity to RAi UK's programme of events to promote inclusivity, and to provide experienced facilitators to make sure all voices are heard. Support is available for those with caring responsibilities, and our EDI Allies will be on hand at events to provide responsive assistance.

Promoting gender equity in AI

At an open event in London in November 2023, run in parallel with AI Fringe events hosted across the UK, we convened a group of EDI experts, artists, industry innovators and gender advocates to provoke discussion and debate on how to address gender disparity and exclusion that persists within the field of AI.

A key focus of this event – and indeed for the wider programme of activities RAi UK will deliver – is the distinction between equality and equity, for gender but also for a range of other characteristics.

For RAi UK, surfacing the most pernicious and prevalent issues around equity in AI head-on from the outset is critical to ensure we design our research and develop a wider portfolio of activities that tackles these issues. Roundtable discussions at our Gender Equity in AI event identified common themes which can inform our research:

- The need for robust data ethics and more representative data sets for large language models
- Designing AI systems to represent plurality of voice, tackling imbalances and power dynamics
- Developing AI to support cultural diversity and multiple languages, and avoid associated bias
- Broadening and enhancing AI literacy among the general population

Integrating RRI and EDI into our research

In May 2024, we welcomed representatives from our funded projects to Nottingham for a workshop to discuss the benefits and challenges of applying RRI and EDI principles to responsible AI research.

This was a fantastic opportunity for our teams to network and identify synergies between their projects, in addition to discussing the main order of the day: RRI and EDI best practice. It was inspiring and thought-provoking to hear from our funded projects about their own experiences with EDI and RRI, and sharing lessons learnt so far.

To evoke a participatory feel to this event, we spent the afternoon in small groups reflecting on our own RRI and EDI practices within our RAi UK projects. Facilitated 'hands-on' activities with two decks of structured ideation cards: the RI Prompts and Practice Cards (RI Cards)[1] and Equality, Diversity and Inclusion Cards[2] provided a platform for groups to compare notes on the RRI and EDI strengths and weaknesses within their projects, and to explore opportunities to enhance these components of their projects and troubleshoot potential risks and mitigation.





[1] V Portillo et al. 2023. Responsible Research and Innovation (RRI) Prompts and Practice Cards: a Tool to Support Responsible Practice. In Proceedings of the First International Symposium on Trustworthy Autonomous Systems (TAS '23). Association for Computing Machinery, New York, NY, USA, Article 57, 1–4. https://doi.org/10.1145/3597512.3599721

[2] P Craigon et al. 2023. The Equality, Diversity and Inclusion Cards– Introduction and Work in Progress. In Proceedings of the First International Symposium on Trustworthy Autonomous Systems (pp. 1-7).

The RI Prompts and Practice Cards are a deck of ideation cards based on the AREA-Plus Framework for RRI. Each card emphasises one element of the AREA framework (Anticipate, Reflect, Engage, Act) and one of the "4 Ps" (Purpose, Product, People, Process). The cards can be used to explore RRI, to reflect on a project and plan RRI activities, or to facilitate a discussion about priorities for RRI in a research or application area. The RI cards were developed by the Horizon Digital Economy Research Institute in collaboration with the TAS Hub, supported by UKRI.

The RI Prompts and Practice card deck and EDI card deck are available online and are free to access.

RI Cards

EDI Cards

We highly encourage all researchers and actors involved in the development and deployment of AI to consider how their project addresses the challenges and embraces opportunities for embedding best practice concerning EDI and RRI.

Please do get in touch via info@rai.ac.uk for an informal conversation or for support in using the tools.

Priorities for the Equities Pillar in 2024/25:

- Launching in Autumn 2024, our Research Culture Survey will engage the wider community in setting expectations and encouraging change to ensure we have an open and inclusive research culture.
- Establish benchmarks and resources that support best practice in RRI and EDI, working in collaboration with the wider RAi ecosystem, including our Keystone Projects.
- Monitor and reflect on EDI and RRI across the programme and listening to and acting on feedback from the community; in the membership of our groups, participation in our events, and meaningful consideration in research approaches.

Global engagement with RAi UK Research

RAi UK aims to connect the broad ecosystem of responsible AI research across the UK into leading research centres and institutions around the world. Our ambition is to build a globally connected programme, incorporating and embracing diversity of cultures, economies and societal challenges, e.g., surrounding attitudes to the use of personal data, or public acceptance of new technologies.

Building an internationally recognised and globally leading programme over the course of the next four years will be essential to deliver best practices on the design, evaluation, regulation, and operation of AI-systems that benefit all people, all societies and all nations, irrespective of national borders and socioeconomic status.

Our team has worked hard over the past year to establish RAi UK as a recognisable brand within the global arena, and to ensure our position on what it means to design and deploy AI responsibly is part of the ongoing conversation.

Outside of the UK, the team have represented RAi UK at conferences, including the UK Malaysia AI Conference in Kuala Lumpur in October 2023, sent delegations for research visits to Singapore, Australia, and the US amongst other places, and presented at events including CogX LA in April 2024, where we were joined by representatives from our Keystone Projects to announce these investments.

Extending the reach of RAi UK

The first year in RAi UK saw a flurry of activity across our team to leverage existing contacts and forge new strategic relationships with international partners. We quickly recognised that realising our ambition to exert influence beyond the UK would require empowering an extended team from the responsible AI ecosystem. Thus, one of our earliest activities was to issue the funding call for International Partnership projects.

Through this call, we invested in a suite of projects that focused on developing deeper collaborations with experts based at organisations around the world.

The International Partnerships funding call ran in three sequential cycles, enabling colleagues with 'oven-ready' projects to bid and win funding immediately, or for colleagues with highly promising connections to spend time developing ideas prior to applying.

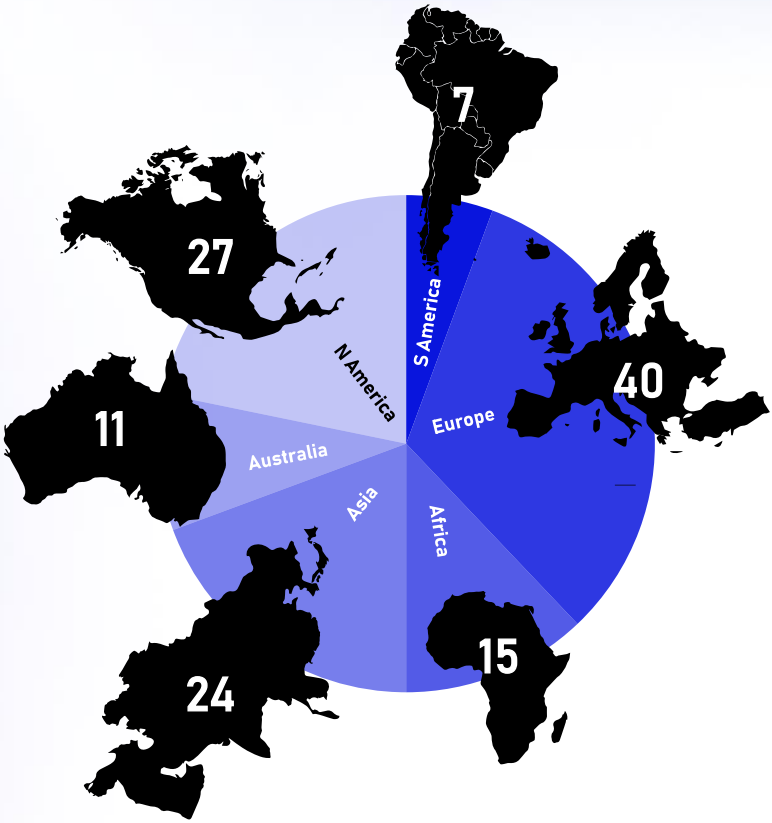
Colleagues across the responsible AI ecosystem were asked to propose collaborations with international research organisations, including universities, public sector research organisations and publicly funded research institutes. We sought proposals that fostered strategic connections, with a 'centre-to-centre' feel, and capable of priming collaborations for sustained activities beyond the life of these projects. Partnerships were not restricted to single organisations or single countries, and we were equally keen to see either high-quality partnerships that developed new collaborations or those that cultivated existing relationships.

The opportunity to bid for funding for these Partnerships proved highly popular. In total, 78 proposals were submitted, led by colleagues from more than 40 different universities and other eligible organisations across all four nations of the UK.

We received an impressive and exciting array of bids, proposing intriguing collaborations with responsible AI institutes, regional, or national initiatives around the world, in 45 different countries, and spread across six continents. The scale of proposed projects ranged from building partnerships with single organisations, through to bids involving collaborators in up to six other countries.

Collaborations with countries in Europe were the most prevalent, followed by North America and Asia, but we were pleased to see colleagues also identify prospective opportunities to connect with organisations in Africa, Australia, and South America.

While the focus of this funding call was not specifically to support projects that connect with organisations in the global south, it was also highly encouraging to see diversity and originality in proposals that aimed to explore challenges faced by regions that differ both culturally and socioeconomically.



Continental distribution of proposed collaborations observed in our International Partnerships funding call.

Investing in an international portfolio

With limited funding available, we could only support a relatively small cohort from these high-quality proposals; 11 projects were selected and commenced in February 2024. Successful projects are now developing collaborations with organisations in 11 countries (US, Canada, Chile, Sweden, Belgium, Estonia, Spain, South Africa, China, the Philippines, and Australia), with project activities including representatives from a much larger number of countries. Further details about all our International Partnership projects can be found in the RAi UK Project Index.

Our International Partnership projects are complemented by a number of other RAi investments, giving an open, outward-looking feel to the research portfolio at this early stage. Outside of the UK, our research projects reference working with 24 individual countries, and the European Union. For example, several Impact Accelerator projects are working with partners around the world: RAISE will work with SMEs in Nigeria, Kenya, and South Africa; AEGIS collaborates with colleagues at the National Institute of Informatics in Tokyo, Japan; and Amplify is leveraging pre-existing relationships with organisations in other countries, including India and South Africa. Our Keystone investments have also endeavored to bring global perspectives into their project design, by engaging with a range of academic and industrial partners from Europe and beyond.

Initiating new international projects

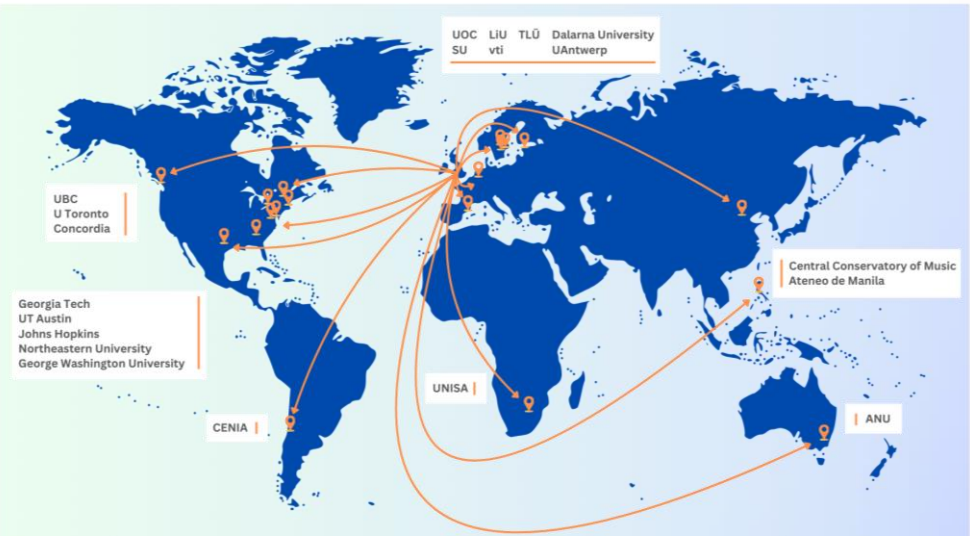
Transnational collaborations ensure the work we undertake in the UK on responsible AI is informed by, and complements, research carried out in other parts of the world.

In September 2024, early career researchers from RAi UK will join forces with researchers from the Centre for Responsible AI (CeRAI), a virtual interdisciplinary research centre at the Indian Institute of Technology Madras (IIT Madras), to network, share ideas and co-create new projects that will address the challenges of building, operating, and regulating artificial intelligence across different countries and cultures.

RAi UK is organising a residential four-day sandpit in collaboration with CeRAI for a cohort of 30+ researchers. The sandpit will bring together diverse and multi-disciplinary research teams, primarily comprising early career researchers, with different perspectives on national and international challenges.

Sandpits are intensive, and require significant commitment from those involved, but are well established as a mechanism to seed novel, ambitious projects and break down boundaries between disciplines.

Together, RAi UK and CeRAI coordinate over £40m of investment in multi-disciplinary research across India and the UK. We are excited to see the innovative projects that emerge and anticipate announcing these new projects later in the year. Further to the national agreement established between UKRI and the Government of India in 2023 to co-invest in AI research and innovation, this sandpit will also serve to uncover risks associated with UK-India partnerships, and to derisk similar investments in the future.



Collaborations funded by RAi UK through our International Partnerships funding call. UBC – University of British Columbia, U Toronto – University of Toronto, UT Austin – University of Texas Austin, CENIA - National Center of Artificial Intelligence in Chile, UOC – Open University of Catalonia, SU – Stockholm University, LiU - Linköping University, vti - Swedish National Road and Transport Research Institute, TLÜ - Tallinn University, UAntwerp – University of Antwerp, ANU – Australian National University.

Establishing partnerships

RAi UK's partner network is a key component in our mission to connect the responsible AI ecosystem. It will act as the catalyst for our research programmes, helping to propagate research and innovation projects with new challenges, and to translate the outputs of our research into real world impact. We aim to create a range of entry points and modes of engagement for colleagues in industry, government officials, policy makers, global thought leaders, and beyond, to work with the RAi UK Pillars, through which focused activities are delivered for RAi UK.

Collective responsibility for ensuring AI is deployed responsibly is at the heart of what we want RAi UK to represent. This will only happen by facilitating meaningful engagement from all stakeholders, allowing room for diversity of opinion, perspective and priorities.

We are committed to engaging and working with a network of partners, exploring how they can best contribute to the programme and develop working relationships that bring benefits to all parties.

Our partners receive regular briefings from our CEO and other team members, helping to foster links between partners and the extended RAi UK programme, and offering opportunities to influence our agenda and shape our strategy. Proposing use-cases and challenges for our research projects is a hugely valuable input, ensuring we address research questions that matter and providing clear pathways to impact.

Our partners also benefit from early access to research publications, white papers, and thought leadership pieces, all with the potential to impact on their own R&D. Looking forwards we will welcome RAi UK partners to our events and provide a platform to address the wider community. There will be opportunities to come together with other colleagues from other organisations in the sector at our domain-specific Working Groups, and accompanying invitations to workshops and roundtables with government, industry and research leaders.

Reflecting on a first successful year for RAi UK, we committed from the outset to engage as broadly as possible. Our Townhalls in London, Belfast, Glasgow and Cardiff were open to all. Although popular with the academic community eager to hear how the programme would support researchers, we were also really pleased to welcome non-academic stakeholders to these events to hear their perspectives, and grateful to them for their input at the roundtables, which ultimately helped to shape the thematic focus for the RAi UK research programme.

There is a clear and pressing need to maintain and strengthen ongoing dialogue between the academic research community and non-academic stakeholders, to cope with the pace of change that surrounds AI development and deployment. We intend to utilise the resources and networks available through RAi UK to provide this for the extended community.

Earlier this year, we held our first RAi UK Partner Network Townhall, designed to complement other events RAi UK has hosted – such as our earlier regional Townhalls – but with a notable twist. Here, we put the spotlight firmly on non-academic partners, focusing on their agendas, talking about how they can best feed into the programme, and gathering their input via a series of facilitated sector-orientated roundtables. These covered the creative industries, security and defence, financial services, transport, and health & wellbeing, with an extra catch-all, cross-cutting roundtable to talk about Equality, Diversity and Inclusion in AI.

We would like to take this opportunity to once again thank all our partners and research colleagues who joined us for the event and made it such a success; the turnout was fantastic, and we were especially pleased with the diversity we saw in terms of sectors, areas of interest, and types of organisations. This led to some interesting discussions and hugely valuable input that we used to shape a range of in-train activities, including our recent Integrator Sandpit held in June as part of the Coordinating Keystone.

Lara Carmona

Director of Policy & Engagement, Creative UK

Responsible Ai UK seems really open to developing work that is responsive to the needs of the people that they are seeking to influence. I was really struck by the fact that they were open to changing direction, to allowing people to shape and influence the way they think and design the work.

Prof Henry Tse

Global Head of Products and Services, BAE Systems Digital Intelligence

Responsible AI is a topic that we all talk about at the moment. At BAE systems as Global Head of Products and Services, we keep trying to think about how we enhance the productivity for our clients and customers and here is an opportunity to engage with other industry sectors and also academia to work together.

Sundeep Bhandari

Chief Digital Innovation Officer, National Physical Laboratory

The most valuable thing for me is that it's not just one set of people from one community thinking about "how we bring an ecosystem together or how we maximise an opportunity to catalyse the safe, responsible and positive outcomes that AI can deliver". It's a grouping of inspiring, open and motivated minds from different backgrounds, disciplines, and focus areas - just brilliant to see and more importantly to be part of. NPL are proud to play its part in supporting the responsible adoption of AI, ensuring that public can have confidence in the technology and in drawing business and investment into the UK.

To hear more about what some of our attendees had to say about their experience and aspirations for RAi UK, click here to view our exclusive video.

We would like to restate our open invitation for colleagues from industry and beyond to join us as we cultivate a robust and internationally leading ecosystem for responsible AI research and innovation. As our portfolio of research expands and evolves, our growing research community will need your input to shape their research, and your support to amplify impact across sectors.

RAi Project Index

www.rai.ac.uk

RAi UK Project Index

RESEARCH & INNOVATION

The RAi UK research project portfolio includes 23 funded projects, awarded through a number of distinct funding schemes (Impact Accelerator, International Partnership, Keystone, Satellite), and distributed across seven sectors of impact.

Governance, Regulation and Policy

PHAWM - Participatory Harm Auditing Workbenches and Methodologies

Team: Simone Stumpf, Monika Harvey, Yunhyong Kim, Craig Macdonald, Graham McDonald, Iadh Ounis, Alesandro Vinciarelli, Mark Wong, Evangelia Fringi, Zoe Bartliff, Jane Blanchard, Paul McMillan, Adam Turner, Vyron Christodoulou, Elizabeth Black, Elena Simperl, Daniele Quercia, Dan Hunter, Kathryn Simpson, Leonardo Bezerra, Alexander Brownlee, Leif Azzopardi, Ian Ruthven, Yashar Moshfeghi, Siamak Shahandashti, Laura Whyte, Patrizia Di Campli San Vito

AI fairness & bias | Data quality | Education & skills | Verification | Regulation, policy, law

AI auditing has, until now, been left solely in the hands of AI experts. The PHAWM project team is driving change in AI testing and evaluation through the novel concept of participatory AI auditing, where a diverse set of stakeholders without a background in AI, such as domain experts, regulators, decision subjects and end-users, are empowered to undertake audits of predictive and generative AI, either individually or collectively.

Research is grounded in four use cases that cover predictive and generative AI. For predictive AI, the use cases cover health (hospital readmission and identification of children with insecure attachment) and media content (fairness in a search engine and detection of hate speech in social media). For generative AI, the use cases will explore cultural heritage (historical materials in the collections of partner organisations), and collaborative media creation with AI (Wikipedia articles on health exploring under-resourced languages). For each use case, regulatory frameworks in the UK, USA, Japan and the EU will be reviewed, and policy and decision makers interviewed.

Outputs will include workbenches that support stakeholders in auditing AI, with audits embedded in methodologies that define how, when and who carries out the audit. Following rigorous evaluation, the focus will move to training stakeholders to carry out the audits, to embed project outputs in practice. The team will work towards a certification framework for AI solutions, democratising AI development and ensuring that AI is safe and trustworthy.

https://phawm.org/

> *Our project will put auditing power back in the hands of people who best understand the potential impact [...] 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.*
>
> Simone Stumpf, Project Lead, PHAWM

Transparency Regulation Toolkits for Responsible Artificial Intelligence

Team: John Downer, Peter Winter, Joshua Krook, Jan Blockx

AI fairness & bias | Education & skills | Regulation, policy, law

This partnership between the University of Bristol and the University of Antwerp aims to harness the opportunities of a comparative analysis of AI transparency and to understand how AI transparency law is viewed on-the ground, by holding workshops with industry professionals, stakeholders, and end users in the EU and UK, and starting a collaborative dialogue between policymakers and researchers.

The research team hosted a successful SME AI Transparency Workshop at the University of Bristol in June 2024, attended by leading experts from industry and academia. Outputs will include ethical toolkits, giving practical and straightforward advice for Small to Medium sized Enterprises (SMEs) on how to abide by AI transparency laws in the UK and EU, and responses to calls for evidence by the UK parliament and EU government bodies, helping to shape the evolving regulatory landscape.

> *We hope to assist in the upskilling of SMEs in the UK and EU to harness the full potential of AI while making sure that AI systems are transparent towards consumers. This is an exciting opportunity to build collaborative economic opportunities across the two jurisdictions for transparent technological design and development.*
>
> Joshua Krook, University of Antwerp

Responsible AI Networks for Industries and Governments in Developing Countries (Latin America)

Team: Nestor Castaneda, Madeline Carr, Maria Paz Sandoval

Regulation, policy, law

A collaboration between UCL and Chile's National Centre for Artificial Intelligence (CENIA), the main aim is to provide a model for international partnerships to address the challenges of adopting responsible AI in the Global South. The expert team are exploring the economic and social consequences of implementing AI in countries that are primarily technology recipients and regulation-takers. Understanding the challenges that developing countries face will significantly enhance understanding of the issue in the UK and Europe.

An international workshop in Santiago, Chile, will see a gathering of experts from Argentina, Brazil, Chile, Colombia, Mexico, and Uruguay and provide a unique opportunity to exchange knowledge on current practices. A follow up workshop in London will focus on knowledge exchange initiatives with potential European partners. The team will publish and disseminate a policy paper offering actionable insights for policymakers, practitioners, and researchers in Europe and Latin America.

AI Equality by Design, Deliberation, and Oversight

Team: Karen Yeung, Patricia Shaw, James MacLaren, Aaron Ceross, Milla Vidina, Sandra Fredman, Fabian Lutz

AI fairness & bias | Data quality | Quality of life | Regulation, policy, law | Society & public services

The EU AI Act delegates technical standard-setting to the European Committee for Standardization (CEN) and the European Committee for Electrotechnical Standardization (CENELEC). Unless these new standards include effective, demonstrable equality protection measures (reflecting equality by design, deliberation and oversight (EbD) thinking), there are real dangers that adherence to these standards will fail to uphold the fundamental right to protection from unfair discrimination.

The overarching aim of this project is to develop and embed EbD principles into European AI standards (and, in turn, British AI standards), helping to ensure that they are properly interpreted, understood and applied, to provide effective protection against AI-generated unfair discrimination across Europe and the UK

It seeks to do so in three ways:

- By representing Equinet, actively participating in standards-drafting work
- By creating and making available practical toolkits
- To gauge the extent to which UK tech firms understand how equality law affects digital services (through survey), and based on that knowledge (with input from the RAKE project team – see Science, Innovation and Technology) to develop and provide training and resources in equality law to UK tech firms to help address gaps and misunderstandings in their current knowledge.

AI Regulation Assurance in Safety-Critical Systems

Team: Jennifer Williams, Peng Wei, Zena Assad

Infrastructure | Public attitudes | Regulation, policy & law | Data quality | Society & public services

Safety-critical systems that use AI can pose a variety of challenges and opportunities and come with the risk of real consequential harms. Public trust in a technology can quickly evaporate if there is a unexplained safety incident. Working across borders in the UK, US, and Australia, this partnership between the University of Southampton, George Washington University and the Australian National University is exploring AI safety risks for technologies from aerospace, maritime and communications.

The project's main objectives are to:

- Identify emerging and existing technologies that are safety-critical systems involving AI in each sector
- Develop a case study framework (to serve as an RRI toolkit) for specific technologies across sectors by defining scenarios, AI components, and safety metrics
- Elicit, identify, and quantify the potential harms to safety and the likelihoods of these harms
- Identify the technical and regulatory gaps in AI safety, assurance and certification

For each case study, technical and regulatory gaps have now been categorised and solutions for the mitigation of safety risks proposed. The project is bridging the gap between the wider scientific community and international government stakeholders, influencing how AI in safety-critical systems is developed and regulated.

AdSoLve - Addressing Socio-technical Limitations of LLMs for Medical and Social Computing

Team: Maria Liakata, Matt Purver, Michael Schlichtkrull, Greg Slabaugh, Claude Chelala, Domenico Giacco, Rob Procter, Tom Sorell, Nikos Aletras, Jiahong Chen, Aislinn Bergin, Dimitris Gkoumas, Jenny Chim, Joshua Kelsall, Emily Thelwell

AI fairness & bias | Data quality | Quality of life | Regulation, policy, law | Education & skills

> *LLMs are being rapidly adopted without forethought for repercussion. For instance [...] 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.*
>
> Maria Liakata, Project Lead, AdSoLve

AdSolve will assess the socio-technical limitations of Large Language Models (LLMs) in the sensitive and safety-critical settings of healthcare and law. In these domains, implementing poorly understood, or untested and black box LLM systems, is likely to result in life changing implications for patients, litigants, and other stakeholders.

The interdisciplinary project team, working closely with stakeholders, will co-create a strong evaluation benchmark to enable standards and policy partners to implement responsible regulations for the use of LLMs, and industry and third sector partners to robustly assess their own systems.

Novel mitigating technical solutions (augmenting LLMs with temporal reasoning and situational awareness) will then be devised, again informed by expertise in law, ethics and healthcare and co-created with domain experts.

Using three use cases (Multi-modal medical diagnostics and monitoring, AI support for mental health, and AI legal support), the work will enhance the capabilities of LLMs for Responsible Research and Innovation with sensitive and private data, especially in dialogue systems and medical applications, where they offer significant benefits such as the potential for faster, more accurate diagnostics and personalised interventions for healthcare, and improved access to justice and faster decision-making in the law.

The ambition is to provide an evaluation landmark and novel LLM augmentation methods that are transferable to other domains, offering practical solutions and setting the foundation for the safe and responsible development of LLM-based societal applications globally.

https://adsolve.github.io/

Public Voices in AI

Team: Helen Kennedy, Ros Williams, Reema Patel, Susan Oman, Helen Margetts, Octavia Reeve, Jack Stilgoe, David Leslie, Florence Enock, Roshni Modhvadia, Mhairi Aitken, Ellie O'Keeffe, Cian O'Donovan, Tvesha Sippy, Sara Cannizzaro, Ruth Lauener, Sarah Givans

AI fairness & bias | Public attitudes | Education & skills | Regulation, policy, law | Society & public services

As concern about the societal impacts of AI grows and pressure for its effective regulation mounts, understanding and anticipating societal needs and values can inform responsible AI developments and deployments.

The Public Voices in AI project aims to ensure that public voices are front and centre in AI research, development, use and policy, and to centre those most impacted and underrepresented.

Following early-stage scoping work, the project is distributing funds through the Public Voices in AI Fund. The Fund will support participatory projects led by Voluntary Community Enterprise Sector organisations, either negatively impacted by AI directly or working with negatively affected or underrepresented groups.

Outputs will include an open access evidence database; a participation framework; new data on public attitudes to AI (gathered using a large-scale survey); and resources for the meaningful and appropriate inclusion of public voice. All outputs of the project are designed to support a range of beneficiary groups, including AI makers, AI policy makers, and those most negatively affected by AI. Outputs will support equitable and inclusive engagement of diverse members of the public across the AI innovation lifecycle.

https://digitalgood.net/dg-research/public-voices-in-ai/

Public voices need to inform AI research, development and policy much more than they currently do."

Helen Kennedy, Project Lead, Public Voices in AI

RAISE - Responsible generative AI for SMEs in UK and Africa

Team: Bernd Stahl, Damian Eke, Paschal Ochang, Harriet Cameron, Rowena Rodrigues, David Barnard-Wills, Matt Buckley, Ben Howkins, Christoher Nathan

AI fairness & bias	Culture	Labour & jobs
Quality of life	Public attitudes	Society & public services
Regulation, policy, law	Sustainability	

Building on insights developed across several ethical AI projects, notably the EU-funded projects SHERPA and SIENNA, this work will provide actionable guidance to SMEs on how generative AI systems can be developed and used responsibly. Working with Trilateral Research, (using the 'with SMEs, for SME's' approach) the project has access to a company leading on socially responsible AI to co-design and test practical and actionable guidance on how generative AI can be integrated into innovative products.

Launched at AI UK 2024, the team now welcome feedback on the RAISE Guidelines Version 1.0 https://raise-project.uk/guidelines as part of their agile co-creation process. In providing targeted guidance on responsible AI adoption for SMEs, particularly those working with generative AI, the project seeks to foster economic growth and innovation for SMEs in the UK and Africa, as well as ensuring that these technologies are deployed in a manner that benefits society (mitigating risks, reducing potential for harm, and facilitating benefits).

By engaging with SMEs in diverse cultural settings, the project will inform the development of culturally sensitive and context-specific guidance on responsible generative AI adoption, thereby fostering cultural diversity and inclusivity in the AI landscape.

The team have authored a Policy Brief with British Computer Society (BCS), Living with AI and emerging technologies: Meeting ethical challenges through professional standards, and engaged with a Kenyan Government Heads of Delegation meeting organised by the Foreign, Commonwealth and Development Office (FCDO). This has already helped to forge further collaborations with SMEs in Kenya and facilitated ongoing discussions with Innovations Hubs in Kenya.

https://raise-project.uk

Project team members with Kenyan Government Heads of Delegation, FCDO, London, June 2024.

TAS HUB and Good Systems Strategic Partnership

Team: Helena Webb, Justin Hart, Liz Dowthwaite, Sam Baker, Kate Devlin, John Robert Bautista, pepita Barnard, Arya Farahi, Jennifer Williams, Elliott Hauser, Anna-Maria Piskopani, Joel Fischer

AI fairness & bias | Quality of life | Society & public services | Sustainability

This international partnership supports and strengthens the existing relationship between two world-leading centres in Responsible and Trustworthy AI and Autonomous Systems: TAS Hub, UK and Good Systems at the University of Texas, at Austin.

The project team are leading innovative and ambitious activities for research development, knowledge exchange and sharing of best practices. These activities support existing members of both centres and bring in new members from other institutions into the community. This includes non-academic members and colleagues from the Global South, to nurture an international community around Responsible AI.

In March 2024, 'Encountering Robot Autonomy in Public' was held at the Peter Flawn Academic Center at UT Austin. The workshop focused on the role of autonomous robotic systems being deployed into public settings for a range of deployed uses, including delivery robots and autonomous transport shuttles, as well as more unusual or proposed applications of 'public' robotics. In July, in Matlock Bath, the HEAD (Health Equity in AI Decision) Residency was attended by 15 participants. HEAD is an initiative that aims to revolutionise approaches to public health and healthcare technology access disparities, with focus on the US and UK contexts. By the end of the week, participants had formed groups and presented proposed projects that the HEAD community can champion over the coming year.

RAKE - Responsible Innovation Advantage in Knowledge Exchange

Team: Marina Jirotka, Carolyn Ten Holter, Elvira Perez Vallejos, Virginia Portillo, Rob Procter, Alan Chamberlain

AI fairness & bias | Society & public services | Sustainability

Building on existing Responsible Innovation (RI) work developed through previous projects such as ORBIT and TAS Hub, the overall aim in RAKE is to collectively support a new generation of Responsible Innovation, through the development of new methods for the responsible design, deployment, use and assessment of AI technologies in society, thereby strengthening the responsible AI research agenda. RAKE takes a longitudinal view from shaping funding to commercialisation (working with funders, businesses, projects, CDTs, and university spinouts), to investigate how RI can be better embedded within these pipelines to improve the development and deployment of AI.

Objectives are to:

- Map and assess current RI provision, tools, and practice within the UKRI ecosystem
- Identify training gaps in assessment and RI provision for responsible AI, and develop best-practice recommendations and documentation for training and delivery
- Connect with current providers of responsible practice standards and tools, and centres of RI training provision, to create a networked knowledge system
- Develop a programme of support for novel-technology companies
- Build a community of interdisciplinary researchers
- Create longevity for project outputs, including card-based kits and other tools

RAKE is working with the ESPRC AI CDTs, as well as industrial partners, standards bodies and arts organisations.

Responsible & Trustworthy AI: Economic Landscape Analysis

Team: Diane Coyle, Adrian Weller, Ramit Debnath, Sara Abdelaziz, John Lourenze Poquiz, Lucy Hampton, Bart van Art, Joel Hoskins, Shobna Gowda, Rebecca Riley, Arjun Shah

Labour & jobs | Regulation, policy, law

How can Responsible AI drive productivity and economic prosperity, and conversely how can productivity and economic considerations be taken into account when developing Responsible AI?

These are the questions being asked by the Economic Landscape Analysis project team, with an underlying hypothesis being that the best opportunities for societal benefit, improved economic performance and commercial success for the UK lie in a focus on trustworthy and responsible development and deployment of AI.

> *This is an opportunity for the UK to drive forward research globally at the intersection of technical and social science disciplines, particularly where there has been relatively little interdisciplinary research to date. We are keen to enhance connections between the research communities and businesses and policymakers.*
>
> Diane Coyle, Project Lead,
> Economic Landscape Analysis

Building on the Generative AI Framework for HM Government (2024), the team are engaging with policy and regulatory stakeholders to scope the knowledge and policy gaps in competition policy and other sectoral regulatory domains (such as telecoms, electricity, rail, water), and developing a framework for tracking the definitions, risks and innovations that need to be on the regulatory roadmap.

Using a combination of broad landscape mapping and focused examples from two public service sectors, policing and social care, the team are exploring incentives and barriers to the use of AI tools and the associated impact of use of AI on public perceptions and trustworthiness. Outputs will include the construction of an open access database of AI usage across UK's economic sectors.

https://www.productivity.ac.uk/research/projects/responsible-trustworthy-ai-economic-landscape-analysis/

Harnessing AI to Enhance Electoral Oversight

Team: Sam Power, Katharine Dommett, Judita Preiss, Adam Sheingate

AI fairness & bias | Society & public services | Regulation, policy, law

Democracies across the world establish regulators and management bodies to oversee electoral activity. AI has been so far underutilised by electoral regulators and Electoral Management Bodies (EMBs). Empowering EMBs with the use of responsible AI techniques and methods can enhance their oversight mechanisms.

This benefits the everyday citizen in democracies across the world, who expect elections – at a bare minimum – to be free and fair.

The main aim of this partnership between the University of Sussex, and University of Sheffield, with Johns Hopkins University, is to develop an international Community of Practice (CoP) to identify opportunities for the application of AI tools to electoral processes, and to develop standards such that the application of these tools upholds, as opposed to undermines, democratic practice.

The partnership builds on the TAS Hub project, 'Building Trustworthy Autonomous Oversight'. The interdisciplinary team will develop and share software for the automated classification of electoral spending in the UK, with coding schemes, and directions for how it can be adapted for different countries.

Objectives:

- The hosting of two workshops/sandpit events in the USA and the UK to a) identify practical applications of AI in this space, b) outline barriers to these applications and ethical concerns to c) develop a set of agreed upon standards and protocols.
- Agreement around the development of a comparative open access database which will collect information on political spending and donations in countries across the world.
- A practitioner-facing report to be published with International IDEA outlining how regulators and EMBs are using automation and prospects for future use.
- To further develop a Natural Language Processing (NLP) tool being designed in the UK for the Electoral Commission (to aid with their compliance activities) such that it can be adapted and utilised in other cases and contexts, by other EMBs.

Understanding Robot Autonomy in Public

Team: Stuart Reeves, Hannah Pelikan

Labour & jobs | Society & public services

Autonomous robots are being deployed in more and more public urban spaces – whether it is for delivery, transport, or cleaning – yet there is little systematic understanding of how everyday interactions with and around such technologies impact citizens living and working amongst them.

Autonomous shuttle at Linköping Science Park, Sweden (photo Stuart Reeves).

By drawing together networks and partners spanning transport, Human Computer and Human Robot Interaction, robotics, sociology and linguistics, the Robots in Public international partnership with the Swedish National Road and Transport Research Institute (VTI) is building partnerships between existing and new collaborators to share empirical data (ethnographic video data) on human robot interactions in public, and jointly developing interdisciplinary insights to inform the future design of responsible public robotics.

Objectives:

- Share and translate study findings with an interdisciplinary cohort of partners in research and in government, in the UK and Sweden
- Develop empirically grounded frameworks for responsible, inclusive and sustainable design of autonomous robots for public settings
- Contribute to the development of interdisciplinary methods, that will help to bridge the still large divide between technical development of AI and societal consequences.
- Develop and share rich, annotated data 'fragments' (e.g., exemplars, instructive episodes or moments) of 'natural' human-robot interaction in public environments that designers, roboticists, educators and governmental/legislative bodies can take inspiration from.

https://robotsinpublic.org/

Starship robots in Tallin, Estonia (photo Hanah Pelikan)

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

Team: John Domingue, Alexander Mikroyannidis, Aisling Third, Joseph Kwarteng, David Tarrant, Kwaku Kuffour-Duah, Siobhan Donegan, Gráinne O'Neill

AI fairness & bias | Society & public services | Education & skills | Sustainability | Quality of life

The SAGE-RAI team are using partner-applied education-oriented Generative AI tools to unlock the potential of AI to lead to improved student outcomes, while widening access for more, and more diverse, learners. The goal is to create a platform that supports assessment and student guidance while responsibly applying Generative AI, addressing the challenges of misinformation, copyright, and bias.

Based on preliminary work at the Open University, the development of plugins allows for the integration of AI into existing educational courses.

The plugins are deployed in two settings:

- ODI learning environment which provides free and paid courses for learners across the globe.
- OpenLearn the Open University's high profile free learning resource.

Key outputs from the project will include a business case, structural and process models that present why and how educational providers can use Generative AI to enhance student outcomes without undue resource implications, and ethical guidance (using the classification which underpins the EU AI Act) on how to apply Generative AI in educational settings.

Technical work on the project is well underway (https://github.com/SAGE-RAI) for pilot evaluations to begin in September with ODI learners.

https://sage-rai.kmi.open.ac.uk/

For the FATES of Africa: A Co-Developed Pipeline for Responsible AI in Online Learning Across Africa

Team: Nora McIntyre, Sebastian Stein, Judy van Biljon, Adele Botha, Teresa Cerrato-Pargman, Richard Marrett, Raissa Malu, Zhaoxing Li, Indu Bodala, Etienne van der Poel, Mac van der Merwe, Bester Chimbo, Ronell van der Merwe, Junaid Mubeen, Wangeci Thuo, Ruth Mghoi, Obadiah Biwot, Sarah Hawkes.

Education & skills | Society & public services

Online learning is widely celebrated as the gateway to universal educational access. Yet, the opposite is more often true, with online learning found to exacerbate rather than alleviate societal divides.

This international partnership between the University of Southampton, University of South Africa and Stockholm University, focuses attention on a population that is vast on a global scale, yet is often left behind: namely, online learners in Sub-Saharan Africa (SSA).

This partnership will develop a 'how-to' pipeline for protecting and championing responsible AI among SSA online learners: from the conception of an online learning idea, through iterative design and development, until the moment of scale.

Following a comprehensive programme of consultation across the edtech ecosystem in SSA, an openly available Toolkit will be published (available online, but downloadable and operable offline), alongside policy and academic papers. Principles of Fairness, Accountability, Transparency, Ethics, and Safety (FATES) will be cross-cutting in all areas of the pipeline.

In the first few months of the project, the team has established a comprehensive team of Research Fellows who are now carrying out asynchronous and synchronous within-partnership consultations, with in-country visits planned for September and October 2024 to engage in student-led, community-led, and policy-maker informed insights into responsible AI in online learning.

https://raifatesafrica.wordpress.com/

RAI4MH - Exploring Fairness and Bias of Multimodal Natural Language Processing for Mental Health

Team: Rafael Mestre, Annika Marie Schoene, Stuart Middleton, Cansu Canca, Agata Lapedriza

AI fairness & bias | Data quality | Regulation, policy, law | Quality of life

The world is grappling with an escalating mental health crisis. Emergent technologies like AI, and in particular NLP are often presented as promising tools to solve these challenges, by tackling online abuse and bullying, suicide ideation detection and other behavioural online harms.

The overarching aim of this partnership between the University of Southampton and the Institute of Experiential AI at Northeastern University is to co-design with stakeholders a new evaluation framework suitable for cutting-edge interdisciplinary AI research and mental health practitioners. The team is committed to contributing to the evidence base that will help shape policy and ensure AI's integration into mental health is both ethical and beneficial to society.

Objectives:

- To expand the reach of resources (like datasets or fine-tuned models)
- To ensure responsible use of AI in mental health monitoring and detection by evaluating existing models and datasets for bias and fairness and assessing their impact and ethical implications
- To achieve policy impact by offering actionable guidelines and policy recommendations

https://generic.wordpress.soton.ac.uk/rai4mh/

The Human Role to Guarantee an Ethical AI in Healthcare

Team: Raquel Iniesta, Albert Royo, Manuel Armayones

AI fairness & bias | Education & skills | Quality of life

The integration of AI-based tools into routine clinical care opens the door to a completely new paradigm where doctors and machines collaborate to decide diagnosis or treatment for a patient, based on an individual's biomedical information. In this project (a partnership between KCL, Open University of Catalonia and Government of Catalonia), the team will seek to give answer to emerging ethical concerns, as how to build trust between practitioners and AI; how to ensure an AI for routine clinical care that is transparent, fair and trusted; who is accountable for the automated decisions; what is the role and responsibility of patients in the AI medical era; or how we can avoid dehumanisation and disempowerment in an AI medicine.

In previous work, the project team characterised five-facts that contributed to translate ethical principles into human action that can align AI and ethical principles in healthcare.

This international partnership aims to:

- Generate discussion and reflection at national and international levels by a diversity of agents including academia, industry, governments, medical practitioners, and patients on (a) the ethical concerns arising with the integration of AI in healthcare, (b) which ethical principles are considered as prominent, (c) efficient human actions that can serve to align AI with the ethical prospects.

- Work on the refining and validation of the ethical five-facts model with national and international agents.
- Generate knowledge by summarising and analysing the conclusions reached on the different panel discussions to identify the most prominent factors discussed.

DOMINOS - Disruption Mitigation for Responsible AI

Team: Radu Calinescu, Marsha Chechik, Reem Ayad, Ana Cavalcanti, Nick Feng, Sinem Getir Yaman, Lina Marsso, Victoria Oldemburgo de Mello, Jason Plaks, Pedro Ribeiro, Isobel Standen, Beverley Townsend

AI fairness & bias | Quality of life | Verification

AI solutions used in healthcare, transportation, and other critical domains are affected by regular disruptions ranging from environmental change to user error. To mitigate these disruptions, AI solutions must adapt and evolve, while complying with the social, legal, ethical, empathetic, and cultural (SLEEC) norms of their stakeholders. This is extremely challenging and unsupported by existing AI development approaches.

DOMINOS will address this significant gap by delivering a methodology and toolkit for development, deployment, and use of responsible-AI solutions capable of mitigating a wide spectrum of disruptions in ways that are compliant with SLEEC norms across a range of sectors.

The partnership with the University of Toronto is strengthening the UK-Canada collaboration on responsible AI.

Objectives:

- To assemble an open access repository of responsible AI use cases illustrating the disruptions that AI solutions from different application domains need to mitigate, and the SLEEC norms they must comply with.
- To devise the DOMINOS methodology (made available through a technical report) for developing AI solutions capable of mitigating disruptions without violating stakeholder-specified SLEEC norms.
- To implement the DOMINOS toolkit (comprising software tools and reusable software components).

Automated Empathy – Globalising International Standards (AEGIS): Japan and Ethically Aligned Regions

Team: Andrew McStay, Ben Bland, Vian Bakir, Phoebe Li, Alex Laffer

AI fairness & bias | Regulation, policy, law

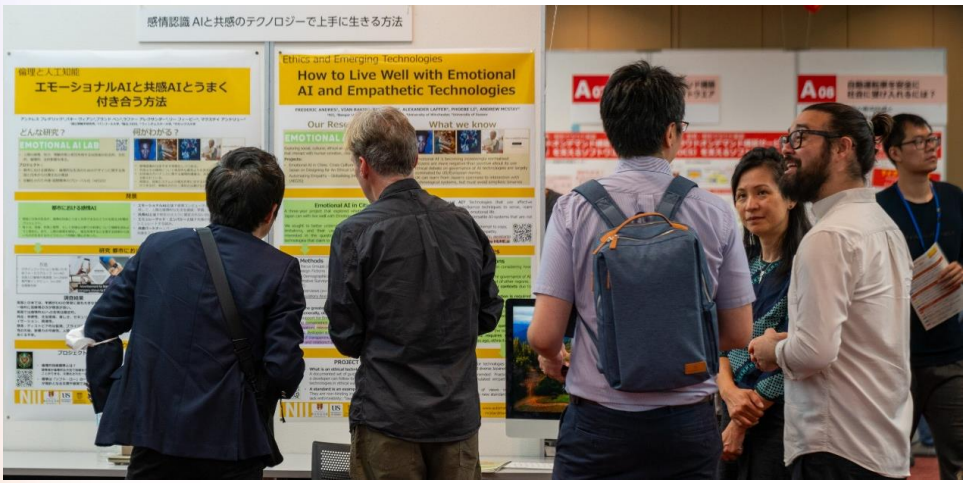
Partnering with standards developer the Institute of Electrical and Electronics Engineers (IEEE) and Japan's National Institute of Informatics, and engaging the Information Commissioner's Office (UK), this project is developing a methodology for responsible AI for systems that purport to emulate human empathy.

Building on academic work led through the Emotional AI Lab and, separately, the IEEE P7104 Standard for Ethical Considerations in Emulated Empathy in Autonomous and Intelligent Systems, the team are developing IEEE P7014.1 (receiving approval for the Project Authorisation Request in March, with the Working Group launched in June 2024). This technical standard and soft law will address the overlap between emulated empathy and human-AI partnering. Use cases encompass general-purpose artificial intelligence products marketed as 'empathic partners', 'personal AI', 'co-pilots', 'assistants', and related phrasing for 'human-AI partnering'.

Current and nascent domains of use include work, therapy, education, life coaching, legal problems, fitness, entertainment, and more.

Sensitive to extreme time-zone and language differences, by curating in-person workshops of regional experts in Japan and East Asia and creating an online public framework of resources, the team are deriving learning for the UK and EU from a region immersed in social robotics and steeped in ethical questions about systems that process intimate data. The first suite of workshops is complete, and the team's white paper of findings, insights and points of global learning for policymakers worldwide, is soon to be published through IEEE.

https://automatingempathy.ai/



AEGIS workshops, June 2024. Hosted at the National Institute of Informatics, Tokyo, Japan.

PROBabLE Futures – Probabilistic AI Systems in Law Enforcement Futures

Team: Marion Oswald, Gita Gill, Kyle Montague, Kyriakos Kotsoglou, Yifeng Zeng, Adrian Weller, Carole McCartney, Claire Paterson-Young, Michael Maher, Elizabeth Tiarks, Muffy Calder, Michele Sevegnani, Sabine Schindler, Peter Style

AI fairness & bias | Quality of life | Public attitudes
Data quality | Regulation, policy, law | Society & public services

> *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.*
>
> Marion Oswald, Project Lead, PROBabLE Futures

Probabilistic systems supported by AI, such as facial recognition, predictive tools, large language models (LLMs) and network analysis are being introduced at pace into law enforcement. Whilst these systems offer potential benefits, decision-making based on 'Probabilistic AI' has serious implications for individuals.

The key problem for responsible AI is that the uncertain or probable nature of outputs is often obscured or misinterpreted, and the underlying data is sensitive and of varying quality. Haphazard ad hoc standards, a lack of monitoring, independent oversight, or evaluation mechanisms, all heighten possible risks and exacerbate potential inconsistencies.

The research of the PROBabLE Futures project team endeavours to deliver an 'operational-ready' blueprint of an evidence-based, contextually informed and future-oriented framework for responsible AI in law enforcement and criminal justice.

Building on the National Police Chiefs' Council Covenant for AI in UK Policing (September 2023), and in partnership with policing, governmental and commercial bodies, the team are drawing on their combined expertise in computer science, law, criminal justice, social innovation and participatory methodology. Activities will include mock trials with AI outputs as evidence, applying lessons from prior technological developments to study the use of 'Probabilistic AI' in policing, intelligence, probation and wider criminal justice contexts.

https://probablefutures.rai.ac.uk/

AIPAS - AI Accountability in Policing and Security

Team: Babak Akhgar, Fraser Sampson, Petra Saskia Bayerl, Martin Snowden, Helen Gibson, Tony Day, Rowan Dennis, Christopher Spencer, Shalini Singh, Rebecca Chapman, Steve Leach, Martin Wilson, Hazel Biggs, Zale Johnson, Shazia Khan

AI fairness & bias | Education & skills | Public attitudes | Regulation, policy, law

A major challenge for Law Enforcement Agencies (LEAs) is balancing the significant opportunities AI presents for the safeguarding of society with the societal concerns and expectation about its responsible use.

A practitioner-focused consortium, led by CENTRIC (Centre of Excellence in Terrorism, Resilience, Intelligence and Organised Crime), AIPAS builds on work conducted by the groundbreaking AP4AI (Accountability Principles for AI) project which investigated assessment and compliance mechanisms for Accountability of AI in Justice and Home Affairs Agencies, with a global reach. As a practitioner-focused consortium, AIPAS is now adapting this foundational work to meet the exact needs of the Policing and Security domain in the UK and into mature practical solutions.

The experienced team are designing practical mechanisms and software tools that will allow LEAs in the UK to assess and implement AI Accountability for their AI applications. With relevant governance bodies and a large-scale citizen consultation included in this process by design, these solutions are not only supporting the prevention and mitigation of AI challenges in policing and security but also AI across the full lifecycle of AI.

In March, AIPAS was introduced to a cohort of industry and academic professionals at AI UK 2024, and in June the team presented to senior police figures from across the forces at the 42nd Annual Police Conference. The first AIPAS Delphi study took place in May, hosted by the Metropolitan Police Forensics Services.

Responsible AI International Community to Reduce Bias in AI Music Generation and Analysis

Team: Nick Bryan-Kinns, Rebecca Fiebrink, Phoenix Parry, Zijin Li, Nuno Correia, Alex Lerch, Sid Fels, Gabriel Viglensoni, Andrei Coronel, Rafael Alampay, Rikard Lindell

AI fairness & bias | Education & skills | Public attitudes | Culture | Quality of life

Music is a core sector of the Creative Economy generating revenue and value across Creative Industries (£17bn global revenue) and the UK is one of the few net exporters. The impact of AI is felt across the Music Industry from creation and production, protection, distribution, to consumption.

This project is building an international community to address Responsible AI challenges of bias in AI music generation and analysis. Bringing together an interdisciplinary team of researchers, musicians, and industry experts, the project's outputs will include AI tools, expertise, and datasets which improve access to marginalised music genres, and to build an international ecosystem to reduce bias in AI music generation and analysis. The work is directly benefitting musicians and audiences through engagement with a wider range of musical genres and benefits the creative industries by offering new forms of music consumption.

https://nickbknickbk.github.io/MusicRAI/

The objectives are:

- To bring together and grow the international community of researchers, creative practitioners, and AI experts interested in using musical genres marginalised by current AI systems
- To establish an open repository of marginalised musical genre datasets for use in AI
- To bring together and make available methods and tools for how artists might use low-resource deep learning models to generate music using marginalized music genres
- To commission a small number of speculative artistic projects resulting in international showcase of generative AI music using marginalised musical genres to raise awareness of the bias of deep-learning AI models
- To explore the translational potential of the AI techniques identified in this project to other creative practices

Ethical and Responsible AI Music Making workshop, July 2024. Live performance by Gabriel Viglensoni.

Amplify: Empowering Underserved Voices in Spoken Language Interaction

Team: Simon Robinson, Thomas Reitmaier, Jennifer Pearson

AI fairness & bias | Education & skills | Society & public services | Culture | Data quality

AI-driven systems for spoken language interaction (SLI) have enjoyed an accelerating rise in availability and popularity in recent years. However, SLI systems are currently only available for a small fraction of the world's languages, a consequence of the challenges in developing the required capabilities for both automatic speech recognition and downstream processing tasks.

This project adopts a radical new approach to the provision of SLIs for language communities where the availability of transcribed speech training data is limited or non-existent; and, most significantly, for languages with no written form at all.

Building on research from the EPSRC-funded project Unmute: Opening Spoken Language Interaction to the Currently Unheard, the project team are facilitating the uptake of their speech toolkit and working with a network of community partners and NGOs to prove and refine the tools and build a network around spoken language technologies.

The UnMute Toolkit was formally launched at the Indian Institute of Technology Guwahati in January 2024, and the project team are now working on refinements, following additional events to test-and-challenge with languages in India and South Africa. The team also plan to work with additional groups in India, Kenya and further afield to further refine the toolkit.

https://unmute.tech

Launching the UnMute toolkit, Indian Institute of Technology, Guwahati, January 2024.

SKILLS

The RAi UK project portfolio also includes six Skills projects which all start in September 2024.

Exploring AI: From Picture Books to Epic Games

Team: Oli Buckley, Magda Zajaczkowska, Sarah Morris, Jake Montanarini

Raising public awareness

The team aim to enhance AI literacy and public awareness of responsible AI by developing two educational resources. The first is an A-to-Z picture book aimed at KS1 and KS2 children, introducing core AI concepts through engaging illustrations and age-appropriate text to foster early engagement.

The second resource is a scenario-based megagame for older children and adults, blending tabletop, role playing, and applied games to immerse players in responsible AI scenarios. These tools will foster critical thinking, problem-solving, and a deeper understanding of AI.

Generative Artificial Intelligence Skills in Schools (GenAISis project)

Team: Konstantina Martzoukou, Pascal Ezenkwu, Sean McNamara, Kirsten MacQuarrie, Emma Grey, Ioannis Panayiotakis, Diane Scott

Raising public awareness | Addressing equity

Partnering with secondary school students, the team will explore the responsible use of Generative AI (GenAI) in their co-creation of open educational resources, articulating student voice and enacting student experience via fictitious characters and cartoons. The project outcomes will also recognise and empower school librarians as key agents in fostering the responsible use of GenAI.

A co-produced open educational toolkit with resources on GenAI academic integrity, information literacy and critical thinking skills in schools, will be widely disseminated via open workshops. GenAISiS offers a transformative pedagogical approach, raising public awareness, promoting equity and amplifying the student voice.

Misinformation/disinformation and Generative AI: Building Young Refugees' Skills and Capacities

Team: Reem Talhouk, Kyle Montague, Bridget Stratford, Jamie Mahoney, Lauren Scott, Ana Pires, Sarah Rüller

Upskilling and reskilling | Addressing equity

Refugees and asylum-seekers are disproportionately impacted by AI propagated misinformation/disinformation.

Young refugees and asylum-seekers (YRAS) are ideal actors to upskill in this area as they already take on the role of technology educators within their communities.

The project will leverage YRAS as educators and equip them with knowledge, skills and resources to be champions of Responsible AI. More specifically, working with YRAS, educators and third-sector organisations, the team will develop 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.

AI, Law and Legal Training (AILLT)

Team: Francine Ryan, Liz Hardie, Kim Barker, Kevin Waugh, Mike Richards and Daniel Gooch

Raising public awareness | Upskilling and reskilling

This project brings together a multidisciplinary team to co-produce research informed resources that will enhance knowledge and awareness of, and confidence in, the use of GenAI for understanding legal processes and accessing legal information. By explaining GenAI, its application, implications, and ethical use in legal contexts, the project activities will educate and empower the public, organisations that provide legal advice, small and medium-sized law firms, students, and academics.

The team will create open access and engaging online courses (including interactive and audio-visual materials, checklists, and case studies) that provide ethical and responsible knowledge of, and skills to use GenAI.

AI in Secondary Schools: Resource Development and Teacher Professional Learning

Team: Judy Robinson, Serdar Abac, Kate Farrell, Jasmeen Kanwal

Upskilling and reskilling

Recent advances in AI may catalyse a transformation in education. Current GenAI tools are powerful, yet brittle. They can act as useful learning companions, but are known to be unreliable, often biased and potentially misleading. There is an urgent need to develop critical AI literacy in learners, and to support teachers in developing the skills and knowledge to incorporate GenAI in their classroom teaching and learning responsibly.

This project will develop resources to support the acquisition of key critical AI literacy skills that will equip young people for living and working in the age of AI, and offer professional learning for teachers.

Ways of (Machine) Seeing: Towards a new visual literacy in AI

Team: Geoff Cox, Annie Davey, Tim Fransen, Yasmine Boudiaf, Janice McLaren

National skills frameworks

How does image-based AI challenge existing approaches to the teaching of visual creativity? New digital tools such as GenAI are easily adopted by young people through everyday media, but there is little or no formal guidance on how to use these technologies safely or responsibly. Our research is designed to generate educational materials about AI, working directly with teachers to offer an enhanced visual learning experience for students.

The team aim to produce responsible AI skills resources that offer best practice for teacher training and curriculum development, introducing the necessary skills to address AI literacy in classroom settings.


Skills Programme Pillars & Activities


RAi
Empowering and Connecting the Responsible AI Ecosystem
rai.ac.uk

UK Research and Innovation

www.rai.ac.uk
info@rai.ac.uk