



Responsible Ai UK

Gopal Ramchurn, CEO Responsible AI UK

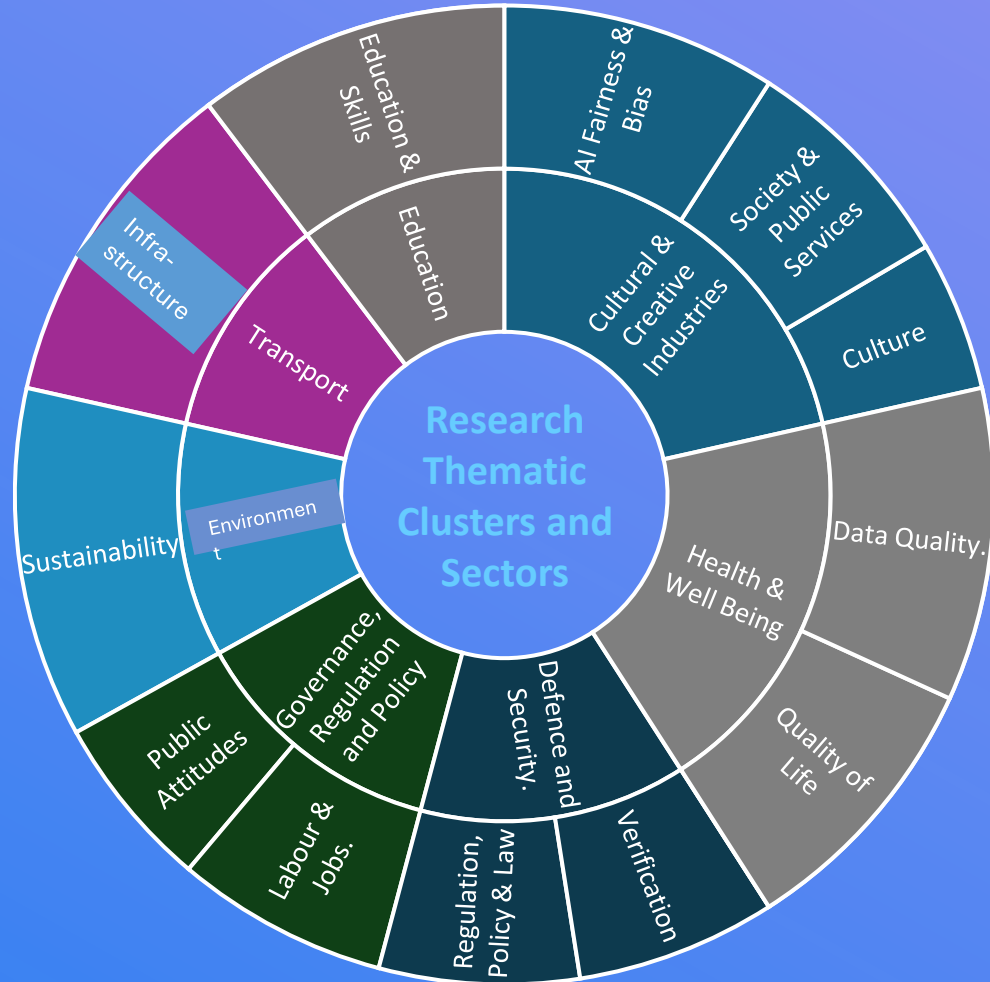
29th May 2025

[Rai.ac.uk](https://rai.ac.uk)



UK Research
and Innovation





Responsible Ai UK (RAi UK) is a **£35 million UKRI investment**, dedicated to delivering **interdisciplinary research** and has deployed over £17 million into projects to:

- **Deliver fundamental research into Responsible AI technologies, practices, and policy**
- **accelerate the adoption of responsible AI practices and technologies.**
- **bring leading research-based expertise to engage with communities, publics, industries, and governments.**

Our Vision



RAi UK will enable Responsible and Trustworthy AI to power benefits for everyday life

How we will do this



Build a national AI ecosystem, where all voices are heard, respected and debated, regardless of seniority or volume.



Deliver research that addresses societal and economic challenges.



Develop national conversations around AI – informed by research, not hype.



RAI UK Team



Our team is made up of academics from different disciplines, with strong connections to policymakers and industry



Muffy Calder



Tom Rodden



Gopal
Ramchurn



Wendy Hall



Kate Devlin



Gina Neff



Jack Stilgoe



Maire Oneill



Derek
Mcauley



Matt Jones



Joel Fischer



Elvira Perez Vallejos



Caitlin
Bentley



Prokar Dasgupta



Sana
Khareghani



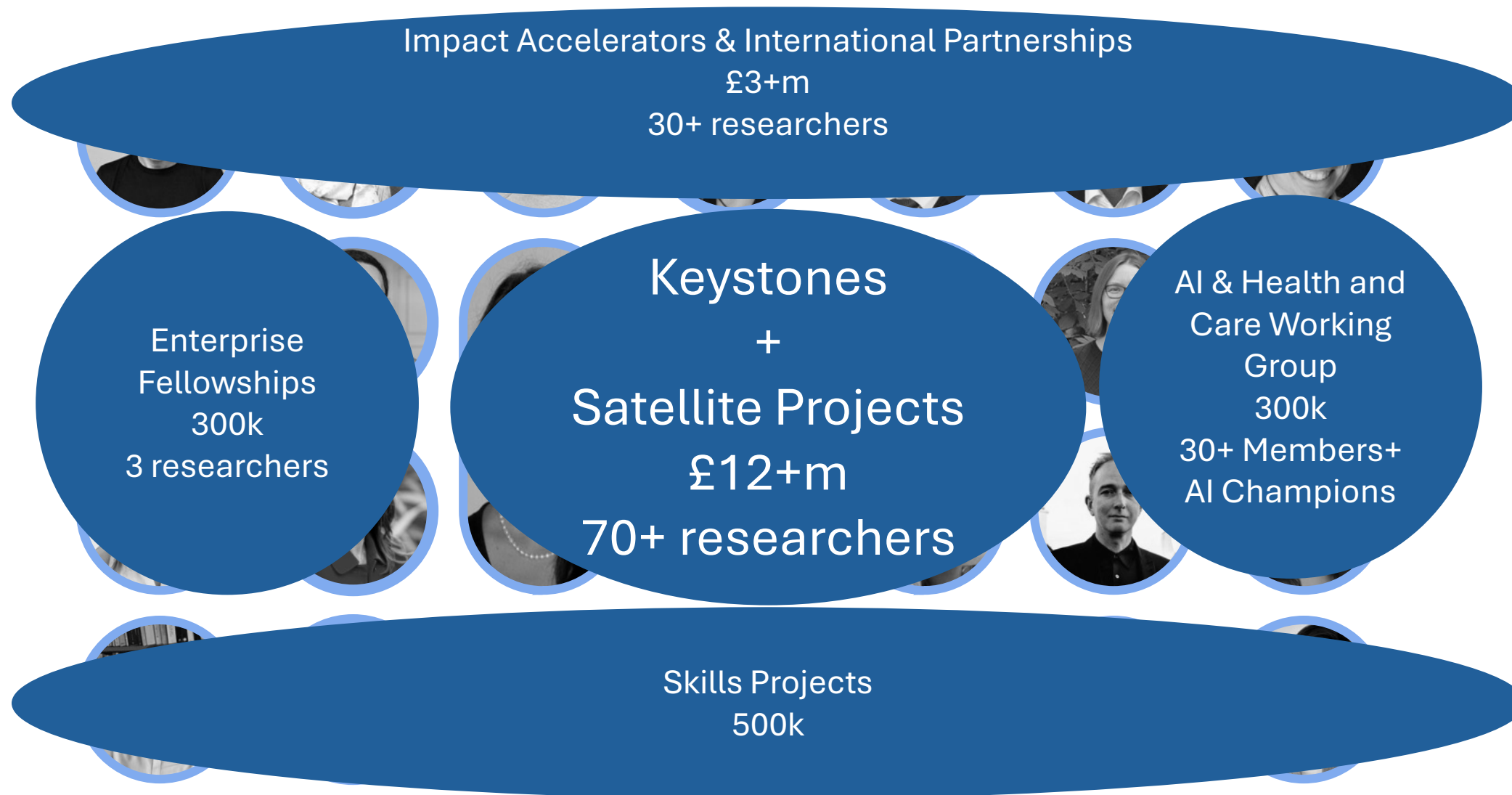
Shoaib Ehsan

RAi UK Team



Impact Accelerators & International Partnerships

The image displays a grid of 21 circular portraits of team members, arranged in three rows. The top row contains 7 portraits, the middle row contains 7 portraits, and the bottom row contains 7 portraits. A central portrait of a woman with glasses is highlighted with a larger blue border. In the bottom row, the fourth position from the left is occupied by a blue circle containing the RAi logo. The background features faint text including 'Impact Accelerators & International Partnerships', '300k', '30+ M', and 'rs+'. A dark blue circle on the right side contains the text 'AI & Ca', 'nd', '300K', and '30+ M'.



The RAI UK Health and Social Care Working Group and the AI Champions Programme



- Advancing Ethical AI: Developing robust frameworks for safe and equitable AI deployment.
- Promoting Confidence: Building workforce and public trust in AI applications.
- Fostering Collaboration: Building workforce and public trust in AI applications.
- Ensuring Equity: Safeguarding the fair distribution of AI benefits across the healthcare landscape.
- For more information please visit: <https://rai.ac.uk/working-groups/health-and-social-care/>

Responsible AI NHS Champions will act as advocates and facilitators for responsible, safe, and equitable AI use within their NHS organisations – primary, secondary and social care. Areas of involvement can include:



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

Responsible AI-powered organisations and economies



- Organisations and economies are rapidly evolving due to new ways of working in response to AI.
- Responsible AI practices need to **include systems-based approaches, beyond testing of individual AI algorithms**, to evaluate the wider implications of AI deployment into complex human/AI systems.
- This includes ensuring that innovators understand the overall consequences and measures, such as the **reskilling and upskilling** that will need to be developed. What are the responsible AI principles appropriate for sectors (e.g., healthcare and the justice system) and contexts (e.g., where there may be high consequences of failure)

Addressing harms, maximising benefits of AI



- How AI is framed and applied introduces **new opportunities but also trade-offs** for individuals, industries, and societies where the potential benefit may be outweighed by negative impacts on a wide range of issues.
- This includes issues such as **privacy, bias, accessibility, labour rights, social justice, and sustainability** (of people, organisations, and the environment). Many of these trade-offs have direct legal ramifications.
- This introduces the need for **deployment, validation, provenance, and auditing regimes** for AI, so decision makers can thoroughly understand and manage the limitations of AI systems.
- Dimensions **of AI safety** (e.g., ensuring the system functions as intended with regards to ethics, policy, legal and technical aspects), **AI security** (e.g., ensuring the system is robust to malicious interference) and the countering of **AI misuse** offer may open questions to be addressed. What is needed to be in place so that AI works for the benefit of people and societies while harms are minimised?

AI Law, governance, and regulation in an international context








- There are few established **routes to control, transparency and redress for users** in relation to automation in digital public services, as well as crucial private sector interactions. For example, the rise of generative AI has posed significant (though not unique) challenges to the creative industries.
- Currently very different approaches to **AI governance are taken by the EU, China, the US, and elsewhere**, a global debate to which the UK must make a strong contribution or risk isolation.
- While longer-term AI safety issues have recently been highlighted by the UK government, **short to medium term risks** remain outstanding.
- What are the approaches that promote **trust, provide fairness, and accountability for users**, and provide certainty for international commerce?


Perception: Human vs Machine

Human		Machine	
Rank	Survey		Comparison ML model
1	cMetastasis stage		cMetastasis stage
2	Performance status		Performance status
3	cTumour stage		Age
4	Biosy / tumour histology		Epoch (<i>not in survey</i>)
5	Comorbidities		cNodal stage
6	cNodal stage		cTumour stage
7	Patient preference (<i>not in ML model</i>)		Referring location (<i>not in survey</i>)
8	Tumour location		Comorbidities (<i>summed</i>)
9	Age		Tumour location
10	ASA grade (<i>not in ML model</i>)		Tumour histology

Themes: Barriers to Using an ML Decision Support Tool in OC

1.	Clinician Superiority	
	Intuition / 'gut feeling'. Clinicians can handle uncertainty (e.g. lung nodule or metastasis?) Should not over-ride clinician judgement.	
2.	Patient Individuality	
	Patients need individualised, holistic care. Decisions are too complex and multifactorial. ML decisions may not be reliable if patient history, tumour or circumstances are outside the norm.	
3.	Transparency and Safeguarding	
	Need open, explainable model. Informed consent from patients – right to refuse. What if clinicians disagree with model?	
4.	Need for More Evidence and Information	
	Validity, cost-effectiveness, time-saving, benefits patients and clinicians. More general information about ML tool.	
5.	Input Requirements	
	Tool should include data on new advancements e.g. molecular testing, immunotherapy. A model trained on decisions made prior to these advancements may no longer be valid.	

ML has not only provided a viable solution with reasonable accuracy but also enabled us to...

Understand decision-drivers at play 

...we can potentially uncover subconscious biases, **challenge preconceptions**

...standardize decisions – reduce variability – **health equality**

...streamline workflows, increase MDT **efficiency** – **health economy**



Responsible AI for Mental Health (RAI4MH) – Partnership

International Partnership grant between the Institute for Experiential AI at Northeastern University (US) and the University of Southampton

+

Industry stakeholders (Kooth Plc.)

- Workshops with stakeholders
- Policy engagement
- Research on bias/fairness of ML models
- Sharing resources/models/data



Rafael Mestre



Annika Schoene

Stuart Middleton



Agata Lapedriza



Challenges of AI in Mental Health Care

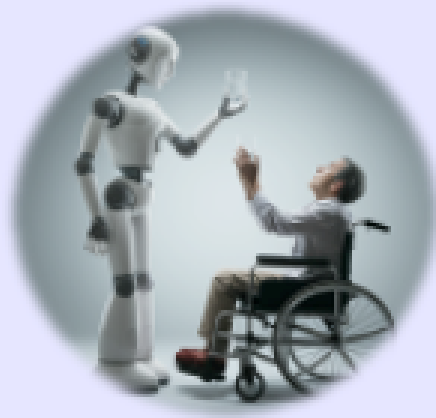
1. **Bias and Fairness in Data and Beyond**
2. **Privacy and Confidentiality**
3. **Ethical, Security and Safety**
4. **Lack of Human Oversight**
5. **Explainability and Trust**
6. **Regulatory and Legal Barriers**

Disruption Mitigation for Responsible AI (DOMINOS)

Project lead – Radu Calinescu (RAi UK AI Champion), University of York
International Partnership

Responsible AI solutions must adapt continually to comply with **Social, Legal, Ethical, Empathetic and Cultural (SLEEC)** norms in the presence of **disruptions**.

Varying user needs



User unavailability



Unexpected obstruction



DOMINOS aims to provide methodology and tools for developing and deploying such resilient, SLEEC-norm compliant AI solutions.

Images generated with Ideogram.



DOMINOS Framework for the Development and Deployment of Responsible AI Solutions

Our RAI-UK International Partnership is assembling a framework that will integrate three major Responsible AI related research outputs.

I. DOMINOS methodology

The Journal of Systems and Software 220 (2025) 112229

Contents lists available at [ScienceDirect](#)

The Journal of Systems & Software

journal homepage: www.elsevier.com/locate/jss

Specification, validation and verification of social, legal, ethical, empathetic and cultural requirements for autonomous agents^a

Sinem Getir Yaman ^{a,*}, Pedro Ribeiro ^a, Ana Cavalcanti ^a, Radu Calinescu ^a, Colin Paterson ^a, Beverley Townsend ^b

<https://doi.org/10.1016/j.jss.2024.112229>

Work in progress: systematic identification & mitigation of obstacles to achieving SLEEC goals

II. DOMINOS toolkit & guidance

Science of Computer Programming 236 (2024) 103118

Contents lists available at [ScienceDirect](#)

Science of Computer Programming

journal homepage: www.elsevier.com/locate/scico

Original software publication

Toolkit for specification, validation and verification of social, legal, ethical, empathetic and cultural requirements for autonomous agents

Sinem Getir Yaman ^a, Pedro Ribeiro ^a, Charlie Burholt, Maddie Jones, Ana Cavalcanti, Radu Calinescu

<https://doi.org/10.1016/j.scico.2024.103118>

Work in progress: toolkit extension

III. DOMINOS Use Case Repository

RESERVE REPOSITORY

ALMI (social care)	ASPEN (environment)	AutoCar (transport)
BSN (health care)	CSICobot (manufacturing)	DAISY (health care)
DPA (education)	DressAssist (social care)	SafeSCAD (transport)

<https://cutt.ly/sleec>



ICSE 2025
47th International Conference on Software Engineering

Sun 27 April - Sat 3 May 2025
Ottawa, Ontario, Canada

Tutorial 2: Social, Legal, Ethical, Empathetic and Cultural Requirements: from Elicitation to Verification

Tuesday: 14:00 to 17:30 with a half hour break:

Presenters: Lina Marusso, University of Toronto, Toronto, Canada; Sinem Getir Yaman, University of York, York, UK; Pedro Ribeiro, University of York, York, UK ; Isobel Standen, University of York, York, UK; Marsha Chechik, University of Toronto, Toronto, Canada

AI Equality by Design

Project lead – Karen Yeung, University of Birmingham
Impact Accelerator

Outputs include an AI Equality by Design (EbD) toolkit - empowering and equipping public equality defenders and other social stakeholders with the knowledge and skills to advocate for, adopt and embed EbD principles into the development and implementation of technical systems and organisational frameworks.

Outputs include an AI Equality by Design (EbD) toolkit - empowering and equipping public equality defenders and other social stakeholders with the knowledge and skills to advocate for, adopt and embed EbD principles into the development and implementation of technical systems and organisational frameworks.

Outputs include an AI Equality by Design (EbD) toolkit - empowering and equipping public equality defenders and other social stakeholders with the knowledge and skills to advocate for, adopt and embed EbD principles into the development and implementation of technical systems and organisational frameworks.

[Project website](#)

Enterprise Fellows



BrainHealthX: Responsible AI-guided solution for early dementia prediction

Dementia is stealing the lives of >55 million people worldwide, with huge societal cost (>\$1 trillion p.a.). Despite >\$56 billion R&D spend over 30 years, we lack sensitive diagnostics at early stages, when interventions may work best. We co-created—with healthcare partners, clinicians and the public—a responsible, multimodal AI tool (BrainHealthx, BHx) to improve early prediction and patient stratification to optimise interventions for each patient. This RAi Enterprise Fellowship aims to scale-up BHx into a trusted, fully deployable clinical decision support system to help identify who will benefit when from which intervention, ultimately driving precision interventions and new treatments.

Professor Zoe Kourtzi
University of Cambridge



Our Keystone Projects



Prof. Marion Oswald **Probable Futures (AI in Law Enforcement)**

“The key problem is that AI tools take inputs from one part of the law enforcement system but their outputs have real-world, possibly life changing, effects in another part – a miscarriage of justice is only a matter of time. Our project works alongside law enforcement and partners to develop a framework that understands the implications of uncertainty and builds confidence in future probabilistic AI, with the interests of justice and responsibility at its heart.”



Dr. Simone Stumpf **Participatory Harm Auditing Workbenches and Methodologies (AI Auditing)**

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



Professor Maria Liakata **Addressing the limitations of large language models for medical and social computers (Gen AI)**

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

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

Where are we heading with AI?

- Agentic AI systems that are fully autonomous and are personalised
- Agentic Meshes that bring many agents together
- A push for Sovereign AI
- AI at the Edge

What does it mean for users and practitioners?

- Personalisation requires diversity in datasets to train AI
- Multi-Agent and Edge AI environments mean bigger concerns for privacy and autonomy
- AI built by and for one country may risk being biased and non-inclusive
- If you are not getting involved with AI (or any new technology), you will lose out