# UKRI AI CDT in Trustworthy Responsible Artificial Intelligence (TOLERANCE)

Fit to Call

Centre vision. We will train a new generation of AI thought leaders and professionals, equipping them to realise the potential of AI in strategic UK Priority Sectors, with the critical skills to explore, advocate for and advance the role of social, ethical and regulatory issues in AI development. Leveraging established links with external partners and gaining research and subject specific knowledge through a tailored programme of cross-disciplinary courses will enable our graduates to help realise societal and economic benefits for industry, academia and public policy, leading the way in the development and ethical

Fig. 1.

adoption of safe, reliable and accessible AI technologies. TOLERANCE CDT will be a national beacon of excellence for doctoral training in the **priority area of responsible and trustworthy AI**.
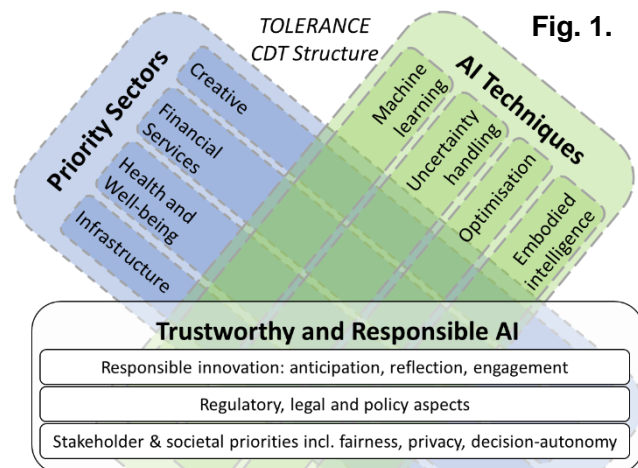
Plans. We will deliver *multidisciplinary training* for a cohort of 60 students leveraging UKRI's initial investment of 40 studentships with at least 20 more funded by the University and external partners, strategically balancing contributions from financially strong sectors with those from charities and the creative sector, providing in-kind support, and maintaining sector diversity of projects/partners in each cohort to maximise opportunities for cross-sectorial synergies and impact. The TOLERANCE CDT is structured around three interacting strands as illustrated in Fig. 1:

Trustworthy and Responsible AI (TRAI)*:* TOLERANCE will integrate *responsible innovation* into training for AI research and development, equipping students to conceptualise the future (*anticipation*), embed reflexivity into their research (*reflection*), and fruitful ways of working with the *stakeholders* and *society* at large (*engagement*) – while developing technical AI expertise. Students will participate in *regulatory*, *legal* and *policy* discourses, leveraging their insights in and becoming an advocate for the design of next-generation responsible AI techniques and processes.

AI Techniques comprise the plurality of technical approaches to AI, the fundamental advancement of which underpins the realisation of trustworthy and responsible AI. Such AI techniques for safe, reliable and equitable real-world deployment include fair, explainable and privacy preserving *machine learning*; *uncertainty handling*, communication for informed and transparent decision making; and advanced approaches to *optimisation* and *embodied intelligence* supporting the complex set of trustworthy and responsible AI requirements, going substantially beyond statistical *performance* as the traditionally primary driver of AI. Aligned with UK priorities we will provide graduates with in-depth knowledge of AI techniques, their design and application, and experience of and sensitivity to the principles of trustworthiness and responsibility.

Priority Sectors: TRAI will be attained across four priority sectors, where public trust is crucial and responsible AI innovation will have direct impact on people's lived-experience, jobs, well-being and more, providing social benefits across the UK, and improve business productivity. We have already recruited >40 committed partners with research and innovation needs aligned with our vision from *creative industries*, *financial services*, *health and well-being*, and *infrastructure* development towards net zero.

National need. The TOLERANCE CDT speaks to the UK National AI Strategy-AI Action Plan, addressing the knowledge and skills gaps, to support the transition to an AI-enabled economy across the *Priority Sectors*. The biggest barriers to adoption and deployment of AI include lack of expertise to understand and deal with complexity, bias and mistrust. Our co-creation conversations with >40 private, public and third sector partners have emphasised a clear need for doctoral level recruits aligned with our training programme. As our initial network of partners evolves and grows as the CDT develops, so will our ability to support cross-sectorial fertilisation and development of TRAI innovation and growth in these sectors. A key aspect of Trustworthy and Responsible AI is democratising AI, which in turn empowers more people to benefit. (see *Priority Sectors* above).

Research environment. TOLERANCE will provide a *diverse and inclusive research environment*

across for students *and* supervisors, building on a 15-year Horizon CDT that has trained 150 PhDs over 14 cohorts; established a flourishing, multidisciplinary research culture and community across the University; and a framework for meaningful engagement with a network of >50 partners. We will exploit the rich experience in AI of the delivery and supervisory team, who are actively engaged in the UK AI research landscape, including via the, the UKRI TAS Hub and the Horizon Hub. TOLERANCE will be housed in a dedicated space on Nottingham's new £100M Castle Meadow postgraduate campus, strategically co-located with key financial sector partners, and close to the city's creative industries and community partners.

Centre Management and Pastoral Care

Applicant team. TOLERANCE will employ a robust management and governance structure. Director, **Ozcan** is an internationally recognised scientist in optimisation for intelligent decision support (180> AI publications). He has contributed to externally funded projects worth more than £11m. He was a PI on the OPTIMACS - H2020 ITN and leads the Faculty of Science DTC in AI at UoN. Deputy Director, **Stahl**, Professor of Critical Research in Technology, was PI of the EPSRC Observatory for Responsible Research and Innovation in ICT. He has contributed to projects >£30m and is Ethics Director of the €450m EU FET Flagship Human Brain Project. Deputy Director, **Wagner**, Professor of Computer Science is leading the Lab for Uncertainty in Data and Decision Making (LUCID). He has published extensively (150>) in AI and led/co-led a £2.84m / £25.75m research portfolio and was Training Programme Manager of the Horizon CDT 2012-2015.

They will be supported by a Management Board (MB) which will shape the overall strategic direction of the CDT, monitoring progress, ensuring balanced and EDI-driven allocation of studentships and other resources across themes and partners, regularly reviewing and revising the programme. The MB will include the following group of CIs, *diverse* in terms of protected characteristics and career stage, each representing a different supporting School from across the University, contributing their network and external partners. **Smith** (Business School) is the director of N/LAB, a centre of excellence in behavioural analytics promoting 'machine learning for social good' across sectors including health, retail, mobility, and energy. **Perez** (Medicine) is a member of MindTech, a national centre focusing on the development, adoption and evaluation of new technologies for mental healthcare and dementia. **Rodrigues** (Architecture and Built Environment) is Head of Department, Deputy Director of the UoN Energy Institute, and founder and Director of Transport, Mobility & Cities initiative at UoN. **Hyde** (Law) is Professor of Law, Regulation and Governance and Deputy Head of School. **Spence** (Psychology) has been involved in research within social, economic, and environmental psychology with experience in both the academic and public sectors. **Watson** (Chemical and Environmental Engineering) is a member of the Food, Water, Waste research group and associate member of the Advanced Manufacturing Technologies research group. **Martindale** (Cultural, Media and Visual Studies) is a Nottingham Research Fellow working at the interface of creative industries and digital technologies. **Caleb-Solly** (Computer Science) researches safe and ethical robots for health and social care, currently leading an EPSRC Healthcare Technologies Network+ and Turing AccessibleAI@Nottingham network.

In support of the MB, we will employ full time CDT and Training Programme Managers and an Impact Officer. Finally, the governance structure will include a *Learning Community Forum*, an *Independent Advisory Board* to shape our programme, and an *Admissions and Progression Board* for academic quality assurance processes.

Student supervisory capacity of UoN. At EoI stage, we already have 60 committed supervisors from across disciplines, with an established track record in delivering PhD training and supervision. The team comprises of a range of experts which is diverse in terms of protected characteristics and career stage, for example, with comparable numbers of established vs. early/mid-career researchers. About half of the team consists of computer scientists, complemented by colleagues from across the humanities, social sciences, engineering and arts faculties, enabling interdisciplinary supervisory teams and cross-sectorial micro-cohorts addressing novel AI. Key investments in leading researchers over the last two years include Caleb-Solly (embodied intelligence) and Stahl (responsible research innovation) who will join our MB, as well as contributors to our supervisory team, e.g. Furnell (cybersecurity), and Vear (AI and music); in turn, these colleagues will support the recruitment of up to 20 further academics into Computer Science, ensuring the growth of our supervisory capacity with new talent and expertise.

Student Training Experience

Recruitment. TOLERANCE will run EDI-driven recruitment campaigns, seeking applicants from across scientific, technical as well as social science and humanities disciplines amongst others. Applicants will have strengths in at least one of the three *strands* of the CDT structure, but few, if any, will be confident in all. The CDT programme is therefore structured with a flexible and modular ethos allowing all students to develop research skills across the core strands. Each student will be supported by a multi-disciplinary team of supervisors and industry partner to ensure that their PhD addresses a combination of the three core strands as described in the following examples:

Student A enters with a background in law and engages in a PhD to work on AI-driven credit scoring with a focus on 'logical flaws' that might lead to 'irresponsible' behaviour, having academic supervisors with expertise in *economics* and *machine learning*. *Student B* enters with formal training in Computer Science. They establish AI music technologies and assistive robotics to enable profoundly disabled musicians to participate in orchestral performances, having academic supervisors with expertise in *music* and *embodied intelligence*.
Both PhDs involve *AI systems* which need to be technically vigorous as well as compliant with data protection and other regulations, but also be publicly acceptable. *Both students* go through the multidisciplinary training and engage with the relevant stakeholders, exploring possible impacts and unintended uses of the technology, embedding these "new" requirements into their solutions and developing innovative responsible and trustworthy AI solutions to the challenges.

Micro-Cohort Training. Building on our experience of developing rounded and technically skilled subject experts, we are committed to the added value of a cohort-based and practice-led approach to deliver a *multidisciplinary training programme* supporting peer-to-peer learning and intra-cohort cohesion, and engagement opportunities with our external partners, leading to *diverse and flexible career tracks* for our graduates. We will: **(i)** provide focused technical training in cutting-edge AI techniques and technologies, and cover core and advanced topics in trustworthy and responsible AI, **(ii)** equip students with the skills to identify and articulate human, societal, ethical and legal challenges in AI, and transfer this insight into the design of AI techniques and technologies, **(iii)** empower individual students with broad *and* tailored skills based on their experience and challenge from a sector that their research is addressing, reflecting their future career plans, **(iv)** facilitate *collaboration and teamwork* at every stage of our programme, **(v)** leverage purposely-designed sector-led micro-cohort-based training, grouping students with heterogeneous skills but a shared sector, aiding students to value and benefit from diverse views across sectorial contexts.

Our programme (Fig. 2) will deliver core research training in year 1, followed by advanced skills training in year 2. During the next two years, forward facing training will be provided to ensure that students complete their writing-up and PhD within the funded period and deliver impact.

Fig. 2. *Overview of the TOLERANCE training programme*

| ● Research-led activity | ● Impact-led activity | ● Partner input |

● Research skills ● Core courses ●● Placement ● Proposal	● Advanced courses ● Peer reviewed paper ●● Sprint project ● Summer school	● Optional advanced courses ● Thesis planning ● Impact plans	● Thesis writing ● Peer viva ● Thesis submission
Annual Events	● Week One Conference (Autumn) – ● All-hands Retreat (Spring) – ● Programme Review (Summer) ● Industry day (Autumn) – ● Networking activity (Spring) – ● Advisory board (Summer)		
Year 1: Core research training	*Year 2: Advanced skills*	*Year 3: Forward training*	*Year 4: Towards viva*

Partner Contributions to Training. Our external partners will support PhD co-design, provide resources to our students, host *placements* (internships/visits), set and evaluate industry-led challenges via *sprint projects,* and more. Our programme will include three annual events, each paired with activities supported by partners: *week one conference* (*industry day*), *all-hands retreat* (*networking event*) and *programme review* (*advisory board*) for fine-tuning the training programme for subsequent years. All students will attend the week-one conference to showcase their work (e.g., placement reports, publications, impact plans) to alumni and *external partners*. An all-hands retreat will focus on well-being and targeted professional skills training delivered by (guest) speakers from *industry* and academia (on e.g., entrepreneurship, mindfulness, career advice). We will seek synergies and engage with the other CDTs across the UK, partners, (inter)national research groups, and organisations offering wider training opportunities and support, including Nottingham's Institute for Policy and Engagement and Haydn Green Institute.