
Enrichment scheme focus research areas appendix

**The Alan Turing Institute
Enrichment scheme 2024-2025**

The key focus research areas appendix within Health, Defence and Security, Environment and Sustainability, as well the big society issues being tackled within Digital Society and Public Policy, and the Core Capabilities such as Fundamental AI. These will enable us to deliver challenge-led Science and Innovation and achieve our mission of changing the world for better with data science and AI.

| Key focus research area | Project | Project description |
|-------------------------|--|--|
| Health | Machine Learning for cell and molecular biology | Developing and using Machine Learning tools to decipher the “rules of life”. |
| | Public health | Using machine learning to design robust public health interventions that mitigate health inequalities. |
| | Digital twins and enabling technologies | This includes research on digital twin credibility and trustworthiness, and model creation, calibration, updating, forecasting and uncertainty quantification. |
| | Further information | Transformation of healthcare The Alan Turing Institute and Health and medical sciences The Alan Turing Institute |

| Key focus research area | Project description |
|-----------------------------|--|
| Defence and Security | <p>General themes under Defence and Security are detection of rare events/learning from limited data, autonomous decision-making fundamentals, human-machine teaming fundamentals, machine learning for applied cyber defence, autonomous vehicles, and implementation of AI algorithms at the edge.</p> <p>Work on cutting edge research and engineering projects that help keep people safe. Our projects include building state of the art autonomous cyber defence agents, establishing the technical risks and capabilities of AI models, and shaping national security policy with interdisciplinary research on emerging technology issues.</p> <p>We encourage prospective PhD Enrichment students to consider Defence and Security as their alignment area. Due to the nature of the research the call text outlines high level areas of activity within the Defence and Security programme. If you choose alignment with the Defence and Security programme and are selected you will be matched to a suitable programme by the Defence and Security staff taking your research, experience, and preferences into account.</p> |

| | |
|--|---|
| | The Defence and Security projects are only open to applicants who are UK citizens. Successful applicants will be asked to go through Security Clearance ahead of aligning with a project. |
|--|---|

| Key focus research area | Project | Project description |
|---------------------------------------|---|---|
| Environment and Sustainability | Ocean monitoring/Scivision | Scivision is an open-source software tool allowing users to explore a range of computer vision models and datasets hosted in the Scivision catalogue in order to find new ways to apply computer vision in their research. Scivision The Alan Turing Institute |
| | AI to predict Arctic sea ice loss (DEFIANT/IceNet) | Drivers and Effects of Fluctuations in sea Ice in the ANT arctic (https://defiant.ac.uk/) & IceNet (Artificial intelligence to help predict Arctic sea ice loss The Alan Turing Institute) |
| | Scivision/Pixelflow | This project will focus on the Pixelflow subpackage of Scivision GitHub - alan-turing-institute/pixelflow Scivision The Alan Turing Institute |
| | Environmental Data Science book (EDS book) | EDS book (www.edsbook.org) is a community-driven initiative aligned to the Turing Way project values. EDS book invites PhD students interested in environmental informatics. It is desirable for collaborators to be familiar with some of the following: GitHub, Jupyter Notebooks, Reproducibility, community-building, and Pangeo (a python software stack for geoscience) or other open-source software for processing and analysing environmental data. Previous enrichment students provided their testimony in the EDS book project in the 21/22 Turing Enrichment Student yearbook (see pages 45 and 68). |
| | Agriculture | AI and Machine Learning in agriculture aims to help farmers and policy makers to make decisions regarding resource application, such as determining the appropriate amount of water, nutrition or other resources needed. More complex on-farm systems for precision agriculture include digital twins and robotic systems. For large-scale applications, we use remote sensing data to work on the problems like land cover segmentation, boundary delineation, yield estimation, and pests and disease detection. |
| | Data Centric Engineering for Critical | Data Centric Engineering for Critical Infrastructure: Research related to leveraging data to support the continued availability of assets, facilities, networks and workers for |

| | | |
|--|---|--|
| | Infrastructure | critical infrastructure sectors (including energy, communications, transportation systems and water). See DCE webpage for more information. |
| | Understanding Daily Urban Population Flows for Digital Twins | Many 'smart' cities publish data that describe movements of people around urban areas. These data could be transformational in the creation of (near) real-time digital twins e.g. to promote resilience and environmental sustainability, but deriving broad insights into urban flows from the discrete data can be extremely challenging. This project will apply state-of-the-art statistical and machine learning methods to pedestrian count data to create a more holistic model of pedestrian dynamics across whole cities. Aligned with work on City Digital Twins. |
| | Creating Synthetic Human Trajectories | Trajectory data are essential for understanding mobility patterns which underpin cities which are equitable, sustainable and resilient. They are also extremely sensitive. This project will develop machine learning methods to create synthetic trajectory data, either from real trajectories or by imputing trajectories from aggregate data. |
| | Land Use | Research on deploying AI to understand the evolution of land uses in urban environments. This is aligned, in part, with the DemoLand project. |
| | Digital Manufacturing | AI and data science research related to supply chains, logistics and factory operations. Methodologically, complex system models, network science, multi-agent systems and predictive analytics (using machine learning) within these areas, for improving resilience and sustainability are particularly welcome. Research relating to trustworthy AI in these areas also very welcome. |

| Key focus research area | Project | Project description |
|--------------------------------|---|--|
| | Multi-modal data integration to facilitate research ready data | Multi-modal data in a Health setting (link, link) will often include images (MRIs, X-rays) and outputs from various omics approaches, alongside other data (from EHRs, or clinical study records). We are interested in implementing, or developing, the best approaches in integrating this variety of data into research-ready formats and building tools, standards and processes around those approaches, in order to extract maximum value of the data and keep the data journey efficient and reproducible. Due to the complexity and cost of collecting multiple data types, we are also interested in exploring questions surrounding the quantity and/or quality of data needed for Machine Learning model development. |

| | | |
|--------------------------|--|--|
| Core Capabilities | Synthetic data for data engineering workflows | Synthetic data is a potentially untapped resource that can speed up research and innovation. What accessible synthetic datasets exist? Are they being utilised fully within data engineering workflows? Existing work on synthetic data has stemmed from the NIHR funded AIM-RSF project; the team are developing data engineering and management workflows with CPRD's synthetic datasets, alongside a growing collaboration with the data provider. Other prominent organisations provide synthetic datasets (ONS, NHS Digital, UK Biobank) and there is scope to better understand how these are generated, categorised, catalogued, and ultimately utilised within the research community. |
| Core Capabilities | The Turing Way | The Turing Way is an internationally-recognised, open source, open collaboration, and community-driven project capturing best practices in open, ethical, reproducible and collaborative data science and AI. Whether you're interested in the way that research is done on a global scale, have ideas about reproducible practices you'd like to share, or just want to get stuck into some open source experience, everyone is welcome! The engagement possibilities are varied, but previous Enrichment students have written chapters on ethics and self-reflection, contributed their experiences with other research projects, run workshops for other Enrichment students, and joined Book Dash events to collaborate with other community members. You can see our open issues here: Issues · the-turing-way/the-turing-way (github.com) and find out more here: Welcome to The Turing Way! (start.page) |
| | Data Safe Haven | For those interested in getting under the hood of how sensitive research data is managed and kept safe in the cloud, the Data Safe Haven offers an opportunity to contribute to a core piece of infrastructure at the Turing. Dive into the national conversation on what the "trust" means in a trustworthy research environment, help run user testing for valuable feedback, or learn how to technically contribute to an open-source project codebase with the team. Find out more here: Data safe havens in the cloud The Alan Turing Institute |
| | AutSPACES citizen science | AutSPACES is a citizen science platform co-created by autistic people and their supporters alongside researchers and the open source community. Initiated in collaboration with the Autistica Foundation, the project is developing an online platform to collect qualitative experiences on sensory processing in autistic people, alongside recommendations and coping strategies to support autistic people, their friends/relatives, and policy makers. If you want to explore contributions to a unique participatory research project, find out more here: alan-turing-institute/AutisticaCitizenScience: Project management and resource repository for the Autistica/Turing Citizen Science project (github.com) |

| | | |
|--------------------------|---|--|
| Core Capabilities | <i>Fundamental AI: Participatory budgeting for chores</i> | Participatory budgeting is an approach to allocating budgets to different projects based on users' preferences. For instance, a local council can propose a list of 10 projects with different costs (such as, a new park or bike path), announce a budget, elicit the residents' preferences and then select a set of projects to implement to satisfy as many participants as possible while staying within the budget. Participatory budgeting for chores |
| | <i>Fundamental AI: Automated analysis of strategic interactions</i> | Current AI increasingly considers that automated agents interact strategically with other agents, both automated and human. Game theory is a formal framework for reasoning about strategic interactions. This project aims to develop a robust and standard framework for working with and analysing game models at scale, in an automated way. Automated analysis of strategic interactions |
| | <i>Fundamental AI: Understanding the capabilities of foundation models</i> | Current methods used to evaluate Large Language models have several limitations that restrict our understanding of the models' capabilities. This project aims to develop new methods to evaluate the strengths and weaknesses of foundation models so informed decisions can be made about their safety and utility. Understanding the capabilities of foundation models |
| | <i>Fundamental AI: Adjoint-accelerated Programmable Inference for large PDEs</i> | The aim of this project is to extend Programmable Inference to large non-self-adjoint partial differential equations (PDE) problems, such as the Navier–Stokes equations. |
| | <i>Fundamental AI: Probabilistic Programming for Ocean Digital Twins</i> | A new infrastructure for tunable aquatic ecosystem models Project objectives: 1) Develop the next generation of accessible and extensible aquatic ecosystem models that can account for the uncertainties of key ecological processes; 2) Apply the new models to provide novel ecological insight into the underlying functioning of aquatic ecosystems; and 3) Establish an international probabilistic aquatic modelling community. |
| | <i>Fundamental AI: Development of composable, parallelisable and user-friendly inference and growing the community of the Turing probabilistic</i> | Turing.jl, a general-purpose probabilistic programming language (PPL), written in Julia is rapidly growing. This project will extend Turing.jl to facilitate collaborative, multi-dataset analyses through compositionality and modularity; to enable Bayesian PPL-enabled analysis of huge data through embedding massive parallelisation; and to improve its usability through community consultation to inform improved application programming interface (API) design. |

| | | |
|--------------------------|--|--|
| Core Capabilities | programming language | |
| | Fundamental AI: A UK Spatial, Climate, & Health Probabilistic Programming Language (PPL) Community | While spatial modelling is a well-developed area within computational statistics, PPLs are not yet mature, robust, and scalable enough to meet the challenges of modern data analysis. Climate change models and the voluminous and ever-changing data they produce pose another huge computational burden. Integrating them within PPLs is an important and open area of future research. This project aims to make conceptual research advances in areas of strategic importance to the UK (climate, health, and spatial modelling), run forecasting competitions and deliver training sessions. |
| | Fundamental AI: Robust inference with probabilistic answer set programs scaffolds for large language models | There are major gaps in LLMs reliability and manageability for real-world use especially in reasoning and planning related tasks. This project aims to develop a inference-time approach to enforcing syntactic, semantic and logical constraints on the outputs of LLMs using neural probabilistic answer set programming (P-ASP) where P-ASP allows for the integration of sound probabilistic inference with logical reasoning. |

| Key focus research area | Project | Project description |
|--------------------------------|----------------------|--|
| Digital Society | Online Safety | Ongoing research agenda surrounding the mapping of the scope and prevalence of online harms through live data collection on the Online Harms Observatory , surveys on public attitudes, experiences, and behaviours online, and the development of novel methods to detect online harms. |

| | | |
|--|---|---|
| and Public Policy | AI for Public Services | Ongoing research relating to the use of AI to improve the provision of public services by mapping use cases across government and collaborating with government partners to design and implement AI applications for prediction, detection, or simulation problems. |
| | Ethics & Responsible Innovation | Ongoing research involving advancing world-leading approaches to AI ethics and governance through research, policy innovation, public sector collaborations, and advisory initiatives. |
| | AI Governance, Regulation, and Standards | Continued research surrounding effective AI governance approaches, the development of regulatory practices, and the coordination of international standards relating to AI for the AI Standards Hub platform. |
| Digital Society and Public Policy | Women in Data Science and AI | Research projects relating to mapping the participation of women in data science and AI in the UK and globally, investigating diversity and inclusion in online and physical workplace cultures, and exploring how the gender gap shapes scientific knowledge and tech innovation. See most recent report on Women, AI, and Venture Capital . |
| | Modelling for Policy | Ongoing research agenda dedicated to developing novel computational social science tools (i.e. for labour flow mapping and skill sorting) combining economic theory, network analysis, machine learning and other methods to uncover large-scale patterns and inform policy, while working closely with governments and international. |