

The Alan Turing Institute

SENIOR RESEARCH ASSOCIATE, FOUNDATION MODELS

THE ALAN TURING INSTITUTE

There has never been a more significant time to work in data science and AI. There is recognition of the importance of these technologies to our economic and social future: the so-called fourth industrial revolution. The technical challenge of keeping our data secure and private has grown in its urgency and importance. At the same time, voices from academia, industry, and government are coming together to debate how these technologies should be governed and managed.

The Alan Turing Institute, as the UK's national institute for data science and artificial intelligence, plays an important part in driving forward advances in these technologies in order to change the world for the better.

The Institute is named in honour of Alan Turing, whose pioneering work in theoretical and applied mathematics, engineering and computing is considered to have laid the foundations for modern-day data science and artificial intelligence. The Institute's goals are to undertake world-class research, apply its research to real-world problems, driving economic impact and societal good, lead the training of a new generation of scientists, and shape the public conversation around data and algorithms.

After launching in 2015 with government funding from EPSRC and five founding universities, the Institute has grown an extensive network of university partners from across the UK and launched a number of major partnerships with industry, public and third sector. Today it is home to more than 500 researchers, a rapidly growing team of in house research software engineers and data scientists and a business team.

BACKGROUND

In 2022, the Alan Turing Institute signalled its intention to establish a portfolio of foundational AI research, which would complement the strengths of the institute around applications of AI and AI policy. An initial portfolio of research in foundation models, game theory, and probabilistic programming will be launched in early 2023. Each of these areas is called a 'Pillar'. It is intended that this portfolio will complement the UK's current activity, rather than duplicating existing efforts, and aiming to promote emerging new areas that show promise for the future.

FOUNDATION MODELS

Foundation models are large ML models trained on large, broad data sets. Foundation models such as GPT have been shown to have remarkable capabilities for generating realistic natural language, and, to some extent, capabilities for problem solving and common-sense reasoning. Developing a Turing Foundation Model is beyond our present capacity. Instead we therefore propose work aimed at developing Turing expertise around the problem of precisely understanding the capabilities of such models. The main issue we aim to address is that of *benchmarking* such models: although such models appear to be very capable in some respects, they fail on apparently simple tasks, in unpredictable ways. In short, we don't have a clear understanding of the capabilities and shortcomings of such systems - which raises concerns for their use.

In this activity, we will work with Foundation Models such as GPT focussing on three key questions:

- Learned values. In this work we will investigate the extent to which such models learn human values. To what extent can it be said to understand the human values raised in such scenarios? Is it consistent in the application of such values? Does it hold such values immutably, or are they malleable?
- Common sense reasoning. Existing Foundation Models appear to have some common-sense reasoning capability, but to what extent do they understand (e.g.) naive physics, spatial and temporal reasoning, the concept of agency, and so on?
- Theory of mind. Much human reasoning is social, involving the beliefs and aspirations of others. To what extent can this capability be acquired by training on textual data sets?

The Alan Turing Institute

The Turing are hiring a team of Research Associates to support and enable the delivery of the main research themes under the direction of the Turing Programme Director Professor Michael Wooldridge, and the project PIs. The team will collectively have a broad range of expertise, and we will be hiring based on varied skills required for each specific challenge.

We are looking for a **Senior Research Associate** to support and enable the delivery of the Foundation Model theme, under the direction of Anthony Cohn, and in collaboration with Michael Wooldridge and Nigel Shadbolt.

ROLE PURPOSE

The successful candidate will primarily focus on evaluating the extent to which, and the conditions under which existing Foundational Models can support common-sense reasoning (such as naive physics, spatial and temporal reasoning, the concept of agency, and causality). This work will be taken in parallel and in collaboration with researchers working on other aspects of the Foundation Models Pillar.

The candidate will join a vibrant team of researchers and will have opportunities to engage with cutting-edge projects and experts at leading universities.

DUTIES AND AREAS OF RESPONSIBILITY

- Develop and implement new methods for representing, transforming, and analysing foundation models.
- Develop new project concepts and proposals and assume overall leadership for the delivery of research projects from inception to completion.
- Run computational experiments to analyse common sense reasoning and understand the efficacy of new and existing analysis methods.
- Define the research direction in collaboration with Project Investigators
- Keep on top of the state of the art in the relevant literatures, in particular at the intersection of Foundation Models and AI in the area of commonsense reasoning.
- Take the lead on writing up findings as they emerge, producing reports, and developing publications in peer reviewed journals, in collaboration with the research team.
- Help to write academic research papers and disseminate our research output to the research community, e.g., by giving talks at international conferences in computing science and other relevant disciplines, and co-authoring articles for journals.
- Present, disseminate and explain our work at internal and external events hosted by the Alan Turing Institute.
- Be a point of contact, supporting PIs in engaging with stakeholders regarding projects and deputising in meetings where necessary
- Supervise junior research team members
- Work closely and collaboratively with other members of the project team.
- Contribute to the life of the Alan Turing Institute and support its community.

Please note that job descriptions cannot be exhaustive, and the post-holder may be required to undertake other duties, which are broadly in line with the above key responsibilities. This job description is written at a specific time and is subject to changes as the demands of the Institute and the role develop.

The Alan Turing Institute

PERSON SPECIFICATION		
Skills and Requirements Post holders will be expected to demonstrate the following	Essential (E) Desirable (D)	Tested at Application (A) Tested at Interview (I)
Education		
A PhD or equivalent qualification/experience in Mathematics, Computer Science or a closely related discipline.	E	A
Knowledge and Experience		
A solid background in two or more of the following: use of Foundational Models such as GPT; AI-based methods for common sense reasoning methods such as naive physics, spatial and temporal reasoning, and causality; evaluation of AI systems.	E	A
Track record of the ability to initiate, develop and deliver high quality research aligned with the research strategy and any industrial stakeholders and to publish in peer reviewed journals and conferences.	E	A/I
Hands-on experience with Machine Learning methods	E	A/I
Track record of outstanding research and delivering impact appropriate to career stage	E	A
Experience in publishing research papers, code libraries or technical reports and giving presentations or classes on technical subjects.	E	A/I
Experience in design, development and implementation of research software tools and libraries, such as Python, Java, GPU programming (Tensorflow, PyTorch, etc), Knowledge Representation methods and tools.	D	A/I
Ability to rapidly assimilate new computational and mathematical ideas and techniques on the job and apply them successfully.	D	I
Ability to create and promote a collegial and collaborative approach to interdisciplinary research activities.	D	I
Communication		
Ability to communicate complex, specialist or conceptual information clearly and persuasively to diverse audiences.	E	A
Ability to write research reports and papers in styles accessible to both academic and lay audiences.	E	I
Analysis and Research		
Ability to organise working time, take the initiative, and carry out research independently	E	I
Ability to use own judgement to analyse and solve problems	E	I

The Alan Turing Institute

Expert in data gathering and analysis, able to develop hypotheses to explain results and confidently present findings.	E	I
Liaison and Networking		
Participates in networks within the organisation or externally to share knowledge and information in order develop practice or help others learn	E	A/I
A proven ability to collaborate successfully in a multidisciplinary environment to meet project objectives.	E	A/I
Teamwork and Motivation		
Ability to work effectively across disciplinary boundaries, both as part of an interdisciplinary team and in close collaboration with external partners in different disciplines.	E	I
Ability to develop contacts and research collaborations within the Institute and the wider community.	E	I
Other Requirements		
Commitment to EDI principles and to the Organisation values	E	I

The Alan Turing Institute

OUR VALUES

The Alan Turing Institute is committed to equality diversity and inclusion and to eliminating discrimination. All employees are expected to embrace, follow and promote our [EDI Principles](#) and Our Values.

Our values

- Trust**
We create an environment where we have trust and can be trusted
- Inclusivity**
We expect our Turing community to contribute to a culture that is inclusive and free of barriers
- Respect**
We all have different roles, priorities and challenges but our shared purpose is the same
- Leadership**
Leadership is everyone's business; Turing leaders set the right tone and lead by example
- Transparency**
Everyone should understand the how and the why of our decisions and actions
- Integrity**
We are all ambassadors for the Turing's mission of changing the world for the better

APPLICATION PROCEDURE

If you are interested in this opportunity, please click the apply button below. You will need to register on the applicant portal and complete the application form including your CV and covering letter. If you have questions about the role or would like to apply using a different format, please contact us at recruitment@turing.ac.uk.

CLOSING DATE FOR APPLICATIONS: Sunday 12 November 2023 at 23:59 (London, UK BST)

TERMS AND CONDITIONS

This full-time post is offered on a 2-year fixed-term basis. The annual salary is £53,576 - £55,125 plus excellent benefits, including flexible working and family friendly policies, <https://www.turing.ac.uk/work-turing/why-work-turing/employee-benefits>

The post can be based either at The Alan Turing Institute site in London, or at the University of Leeds. In either case you will need to travel to the other site when required (travel expenses will be paid as appropriate).

EQUALITY, DIVERSITY AND INCLUSION

The Alan Turing Institute is committed to creating an environment where diversity is valued and everyone is treated fairly. In accordance with the Equality Act, we welcome applications from anyone who meets the specific criteria of the post regardless of age, disability, ethnicity, gender reassignment, marital or civil partnership status, pregnancy and maternity, religion or belief, sex and sexual orientation.

We are committed to building a diverse community and would like our leadership team to reflect this. We therefore welcome applications from the broadest spectrum of backgrounds.

We are committed to making sure our recruitment process is accessible and inclusive. This includes making reasonable adjustments for candidates who have a disability or long-term condition. Please contact us at adjustments@turing.ac.uk to find out how we can assist you.

Please note all offers of employment are subject to obtaining and retaining the right to work in the UK and satisfactory pre-employment security screening which includes a DBS Check.

Full details on the pre-employment screening process can be requested from HR@turing.ac.uk.