### Postdoctoral Research Associate, in Natural Language Processing

### 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.

#### THE PUBLIC POLICY PROGRAMME

The Public Policy research programme works alongside policy makers to explore how data-driven public service provision and policy innovation might help to solve long running societal problems. Our work is interdisciplinary, with machine learning engineers, NLP experts, data scientist and social scientists collaborating across different projects.

The Online Safety Team is a new part of The Turing's Public Policy Programme. It provides objective, evidence-driven insight into the technical, social, empirical and ethical aspects of online safety, supporting the work of public servants (particularly policymakers and regulators), informing civic discourse and extending academic knowledge. High-priority topics include online hate, personal attacks, extremism, misinformation and disinformation. We have three main workstreams:

- 1. Data-centric Artificial Intelligence for online safety
- 2. Building an Online Harms Observatory
- 3. Policymaking for online safety

### **ROLE PURPOSE**

The successful candidate will work as a Postdoctoral researcher in Natural Language Processing (NLP), joining The Turing's Online Safety Team. They will work in the Data-centric AI for online safety workstream, which aims to create a step-change in the socially responsible use of AI for online safety. Candidates who are experienced in machine learning and NLP but who have little experience in online safety will be considered provided they can demonstrate interest and willingness to learn about the topic. Key objectives of the Data-centric AI for online safety workstream are:

- **Objective 1.** To build high performing, robust, fair and explainable tools for detecting content and activity that could inflict harm on people online.
- **Objective 2.** To understand the limitations, weaknesses, biases and failings of tools and technologies for detecting content and activity that could inflict harm on people online.
- **Objective 3.** To produce datasets and tools that can be used by the research community to understand and increase online safety.

The successful candidate will be responsible for conducting original research and implementing innovative data-driven AI, which could include active learning, adversarial learning, few shot learning, and/or domain adaptation approaches. The

successful candidate will drive forward research by co-designing innovative projects, training and evaluating models, and processing data. There is scope for the successful candidate to support with creation of new labelled datasets and other high-impact community resources. The research activities will lead to the publication of high-impact academic papers at NLP and machine learning venues (e.g. ACL, EMNLP and NAACL).

Opportunities for non-academic scientific communication (e.g. blogs and public facing talks) will be provided, if desirable to the successful candidate. The successful candidate will be enabled to explore new areas of NLP research in online safety and to develop models which can be used for real-time analysis of content in The Turing's Online Harms Observatory.

The successful candidate will report to Dr. Bertie Vidgen. They will collaborate with other NLP and data science experts and the rest of The Online Safety Team and Public Policy Programme, including Dr. Scott Hale and Prof Helen Margetts. The successful candidate will be supported with opportunities for ongoing training and support, including funding to attend courses and conferences. There are no teaching requirements and the successful candidate will be encouraged to publish in top tier venues (with appropriate guidance and support). They will also be enabled to form new collaborations and to support the wider NLP community, such as Chairing conferences and joining workshop committees. This is a research-focused role which offers an excellent opportunity for ambitious early career researchers. Opportunities to apply for further funding will be encouraged and supported.

## DUTIES AND AREAS OF RESPONSIBILITY

The successful candidate will:

- Conduct original research to achieve the objectives of the Data-centric AI for Online Safety workstream.
- Train, evaluate and improve new machine learning models using innovative NLP methods.
- Identify and evaluate labelled training datasets.
- Write and submit academic papers to top-tier conferences.
- Support and lead with the creation of new labelled datasets.
- Support with the deployment of models in live settings, such as the Online Harms Observatory.
- Support the work of team members in engaging with non-technical stakeholders.

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.

PERSON SPECIFICATION		
	Essential (E)	Tested at application(a)
Skills and Requirements	Desirable (D)	Tested at interview (i)
Post holders will be expected to demonstrate the following		
Education/Qualification		
PhD-level degree in computer science, statistics, data science, or a related discipline.	E	A
Knowledge and Experience		
Experience in training and evaluating NLP models	E	A
Experience in publishing in top-tier academic venues (e.g. ACL, NAACL, EMNLP)	E	A
Experience in creating, assessing and/or improving labelled training datasets	D	A
Experience in working on online safety related topics, such as hate, harassment, misinformation and extremism (etc.)	D	А
Communication		
Excellent writing skills and proven ability to communicate complex, specialist or conceptual information/research findings clearly and persuasively to diverse audiences.	D	A
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	А
Decision Making		
Independently makes decisions which are low risk and that mainly affect themselves or a small number of people and are guided by regulation and practice.	E	I
Initiative and Problem Solving		
Creatively solve problems, working both independently and with other team members	E	I
Highly motivated and committed to achieving the project goals	E	I
Analysis and Research		
Ability to use own judgement to analyse and solve problems.	E	I
Ability to organise working time, take the initiative, and carry out research independently, under the guidance of the PI.	E	I
Other Requirements		
Commitment to meeting deadlines	E	I
Flexible attitude towards work	E	I
Commitment to EDI principles and to the Organisation values	E	I

## 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 <u>EDI Principles</u> and <u>Rules of the Game</u>



**Respect** – We treat everyone with respect, dignity and kindness and acknowledge the experiences, skills and contributions of others.

**Trust** - We communicate openly and honestly to support an environment where we have trust in each other.

**Transparency** – We seek to ensure that everyone understands the how and the why of our decisions and actions. We take on board to feedback when those decisions are challenged.

**Inclusivity** – We are committed to continuously learning how to be more inclusive by listening to those facing barriers.

**Leadership** – We recognise creating an inclusive, diverse and equitable institute requires leadership from all. We stand up and speak out when change is needed.

**Integrity** – We recognise that how we work is as important as our outputs and seek to exemplify best practice in all our decisions.

## **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 on 0203 862 3340, or email recruitment@turing.ac.uk.

## **TERMS AND CONDITIONS**

This full time post is offered on a fixed term basis for two years. The annual salary range is £37,000-£42,000 plus excellent benefits, including flexible working and family friendly policies, <u>https://www.turing.ac.uk/work-turing/why-work-turing/employee-benefits</u>

Candidates who have not yet been officially awarded their PhD will be appointed as Research Assistant at a salary of £34,500 per annum.

### 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 teal to reflect this. We therefore welcome applications from the broadest spectrum of backgrounds.

Reasonable adjustments to the interview process will be made for any candidates with a disability.

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 <u>HR@turing.ac.uk</u>.