Post Title: PhD researcher in AI audits and the science of evaluation
Location: AI Accountability Lab, School of Computer Science & Statistics, Trinity College Dublin
Under the supervision of: Dr. Abeba Birhane
Duration: 4 years. In-person only, full-time, fully-funded
PhD Start Date: 01 / 03 / 2026

About AIAL

The Artificial Intelligence Accountability Lab (AIAL) is a research lab housed in the ADAPT Research Centre and the School of Computer Science and Statistics in Trinity College Dublin. We are driven by the urgent need to demystify, critically assess, and publicly communicate the operations and functionality of AI systems by looking under the hood as well as at their downstream impact on the public with the aim of shaping public knowledge, shifting power, and reserving fundamental rights, freedoms, and autonomy. The AIAL is an innovative lab at the cutting-edge of AI dedicated to greater transparency on the use of AI systems in the public domain and concrete accountability for the downstream societal impact of these technologies.

The AIAL operates under the core principle that academic research, particularly research in AI, should be of a high-relevance and utility to the public. Thus, we actively engage and partner with local communities, civil society, and rights groups to ensure our work is grounded in real challenges and remains relevant and useful to these stakeholders. We are deeply committed to ensuring our work contributes to fostering transparency and to driving a culture of AI accountability. Academic freedom and autonomy are of utmost importance to us and we take great care to ensure our lab is free from direct conflict of interest or implicit influence or pressure from the industry that we are studying. We are funded by philanthropic organisations such as the MacArthur foundation and government grants.

About the PhD

Audits and evaluation of AI systems — and the broader context that AI systems operate in — have become central to conceptualising, quantifying, measuring and understanding the operations, failures, limitations, underlying assumptions, and downstream societal implications of AI systems. Existing AI audit and evaluation efforts are fractured, done in a siloed and ad-hoc manner, and with little deliberation and reflection around conceptual rigour and methodological validity.

This PhD is for a candidate that is passionate about exploring what a conceptually cogent, methodologically sound, and well-founded AI evaluation and safety research might look like. This requires grappling with questions such as:

- What does it mean to represent “ground truth” in proxies, synthetic data, or computational simulation?
- How do we reliably measure abstract and complex phenomena?
- What are the epistemological or methodological implications of quantification and measurement approaches we choose to employ? Particularly, what underlying presuppositions, values, or perspectives do they entail?
- How do we ensure the lived experiences of impacted communities play a critical role in the development and justification of measurement metrics and proxies?

Through exploration of these questions, the candidate is expected to engage with core concepts in the philosophy of science, history of science, Black feminist epistemologies, and similar schools of thought to develop an in-depth understanding of existing practices with the aim of applying it to advance shared standards and best practice in AI evaluation.

The candidate is expected to integrate empirical (for example, through analysis or evaluation of existing benchmarks) or practical (for example, by executing evaluation of AI systems) components into the overall work. Subsequently, this PhD is an undertaking that marries critical assessment of existing state-of-the-art AI evaluation practices with the aim of driving audits and evaluation towards rigorous science through applied audits.

Requirements:

- Hold a bachelor's degree in the following disciplines: philosophy of science, science and technology studies, computer science, machine learning, artificial intelligence, information systems studies, or cognate disciplines
- In-depth experience and/or publication track-record in the areas of AI evaluation or measurement science
- Familiarity with Black feminist epistemologies, decolonial studies and STS (Good to have)
- Self-driven, ambitious, passionate, and committed to driving AI evaluation towards rigorous science
- Good communicator who is able to articulate the gap between existing evaluation practices and rigorous science and its implications for accountability and AI regulations

Closing date: 10/*02/2026*

Application Process

Interested candidates can submit their application to the email aial@tcd.ie with the subject line “AIAL measurement phd – application” with CV and cover letter to be included. Late applications will not be accepted. Informal enquiries about the role can also be sent to the email aial@tcd.ie with the subject line “AIAL measurement phd – enquiry”.

Please ensure the subject line matches exactly as above. Emails without this subject line may not be considered.