

Post Title: PhD Studentship (48 months) - Ethics of Large Language Models

Location: Dublin City University (DCU)

Anticipated Start Date: October/November 2023

Closing Date: 1st September 2023

Apply: <https://forms.gle/xMojc9pxZJgSsEK47>

The [ADAPT Centre](#) is inviting applications for a four year fully funded (stipends & registration fees) PhD position focused on the Ethics of Large Language Models.

The PhD

Large language models (LLMs) are having a major impact on industries such as journalism, health care, legal practice, education and publishing. Researchers are finding use cases for LLMs retrospectively (Bubeck et al. 2023). Online news outlets have already published AI-produced content containing major errors (Leffer 2023). Schools and universities are scrambling to create workable policies for the use of AI for learning and assessment (Farnell 2023) as well as course preparation (Kasneci et al. 2023). LLMs can help navigate electronic health records (Yang et al. 2022) and assist healthcare workers and patients in a variety of ways (Meskó & Topol, 2023). LLMs may also function as digital legal assistants or even automate some of lawyers' work (Ziffer 2023). Consequently, LLMs would drastically impact the role of professionals working in various fields, as well as the nature and quality of the services available to the general public.

Surprisingly, there has so far been relatively little research produced on the ethics of LLMs or the suggested use cases. The most elaborate study is Weidinger et al. (2021), which is a solid review of the risks of LLMs. Yet it does not address any other ethical aspects. The launch of ChatGPT launched on November 30th, 2022, mainstreamed the debate about LLMs, triggering many journalistic articles and opinions (Gordijn, 2023). In the next few years it will be pivotal to systematically assess the ethical aspects of the development and use of LLMs, and develop a set of ethical recommendations in order to accommodate responsible research and innovation as well as ethical deployment.

The proposed PhD project will contribute to this debate by developing a thorough analysis of some of the most urgent ethical issues in development and/or deployment of LLMs. The PhD will be supervised by Prof Bert Gordijn ([Institute of Ethics](#)) and Dr Joss Moorkens ([School of Applied Language and Intercultural Studies](#)) who are both Funded Investigators at the [ADAPT centre](#). On completion of this project, the graduate should have published work that adds to the field. Experience in this topic should be highly beneficial, leading to opportunities for further research at postdoctoral level in industry or academia. There is an expectation that the student would contribute to ethics teaching, which should also increase employability on graduation. The project is co-funded by the [ADAPT Centre](#) and [the Institute of Ethics](#) at [Dublin City University](#).

Informal enquiries may be directed to Professor Bert Gordijn at bert.gordijn@adaptcentre.ie or bert.gordijn@dcu.ie

References

- Bubeck, S, Chandrasekaran, V, Eldan, R, Gehrke, J, Horvitz, E, Kamar, E, Lee, P, Lee, Y T, Li, Y, Lundberg, S, Nori, H, Palangi, H, Ribeiro, M T, & Zhang, Y. (2023). Sparks of Artificial General Intelligence: Early experiments with GPT-4. <https://arxiv.org/abs/2303.12712>
- Farnell A 2023 AI will replace academics unless our teaching challenges students again. Times Higher Education
- Gordijn, B., Have, H.t. ChatGPT: evolution or revolution?. *Med Health Care and Philos* **26**, 1–2 (2023). <https://doi.org/10.1007/s11019-023-10136-0>
- Kasneci, Enkelejda, Kathrin Sessler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, et al. 'ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education'. *Learning and Individual Differences* 103 (April 2023): 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Leffer L 2023 CNET Is Reviewing the Accuracy of All Its AI-Written Articles After Multiple Major Corrections. Gizmodo.
- Meskó, B., Topol, E.J. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *npj Digit. Med.* **6**, 120 (2023). <https://doi.org/10.1038/s41746-023-00873-0>
- Weidinger, L., M. Mellor, and M. Rauh, et al. 2021. Ethical and social risks of harm from language models. preprint arXiv. <https://doi.org/10.48550/arXiv.2112.04359>
- Yang, Xi, Aokun Chen, Nima PourNejatian, Hoo Chang Shin, Kaleb E. Smith, Christopher Parisien, Colin Compas, et al. 'A Large Language Model for Electronic Health Records'. *Npj Digital Medicine* 5, no. 1 (26 December 2022): 194. <https://doi.org/10.1038/s41746-022-00742-2>

Minimum qualifications:

Applicants should:

1. *Hold a relevant undergraduate degree at 2.1 or first-class honours level (or equivalent) in in Philosophy, Ethics, STS or an aligned field. Ideally, they should also hold a relevant master's qualification. Candidates who are currently completing a master's qualification are welcome to apply. (DCU's entry requirements: Link [here](#))*
2. *Meet, where relevant, the English language requirements for non-native speakers of English. Link [here](#).*

Preferred qualifications:

- *Demonstrable research competence from previous projects at undergraduate or postgraduate level*
- Excellent academic track record as well as research and analytical skills
- Good academic writing and presentation skills
- Ability to combine rigorous theoretical work with practical applications
- Active interest in interdisciplinary work which connects philosophical analysis to other relevant disciplines and societal challenges.

Application Process

Applications should be submitted as a single pdf file. Each application should only consist of

1. Detailed curriculum vitae, including – if applicable – relevant publications;
2. Transcripts and results of final years of degrees,
3. A writing sample (for example, a copy of your Master thesis)
4. The name and email contacts of two academic referees,
5. A cover letter/letter of introduction (max 1000 words). In the letter, applicants should include the following details:

- a. An explanation of your interest in the research to be conducted and why you believe they are suitable for the position.
- b. Details of your final year undergraduate project (if applicable)
- c. Details of your MA/MSc project (if applicable)
- d. Details of any relevant modules previously taken, at undergraduate and/or master's level.
- e. Details of any relevant work experience (if applicable).

About the ADAPT Centre

ADAPT is the world-leading SFI research centre for AI Driven Digital Content Technology hosted by Trinity College Dublin. ADAPT's partner institutions include DCU, University College Dublin, Technological University Dublin, Maynooth University, Munster Technological University, Athlone Institute of Technology, and the National University of Ireland Galway. ADAPT's research vision is to pioneer new forms of proactive, scalable, and integrated AI-driven Digital Content Technology that empower individuals and society to engage in digital experiences with control, inclusion, and accountability with the long term goal of a balanced digital society by 2030. ADAPT is pioneering new Human Centric AI techniques and technologies including personalisation, natural language processing, data analytics, intelligent machine translation human-computer interaction, as well as setting the standards for data governance, privacy and ethics for digital content.

Our Research Vision

Governments and civil society are starting to recognise the need for urgent and concerted action to address the societal impact of the accelerating pace of digital content technologies and the AI techniques that underpin them. ADAPT provides an ambitious, ground-breaking, integrated research programme that assembles three interlocking Strands that together are capable of addressing this challenge. Each of these complementary and reinforcing research Strands takes one of the different perspectives on the provision of personalised, immersive, multimodal digital engagement, i.e. the individual's experience and control of the engagement, the algorithms underlying digital content processing, and the balanced governance by enterprise and societal stakeholders.

Why ADAPT?

- **Contribute** to the ADAPT research agenda that pioneers and combines research in AI driven technologies: Natural Language Processing, Video/Text/Image/Speech processing, digital engagement & HCI, semantic modeling, personalisation, privacy & data governance.
- **Work** with our interdisciplinary team of leading experts from the complementary fields of Social Sciences, Communications, Commerce/Fintech, Ethics, Law, Health, Environment and Sustainability.
- **Leverage our success.** ADAPT's researchers have signed 43 collaborative research projects, 52 licence agreements and oversee 16 active commercialisation funds and 52 commercialisation awards. ADAPT has won 40 competitive EU research projects and obtained

€18.5 million in non-exchequer non-commercial funding. Additionally, six spinout companies have been formed. ADAPT's researchers have produced over 1,500 journal and conference publications and nearly 100 PhD students have been trained.

As an ADAPT/DCU co-funded PhD researcher you will have access to a network of 85 global experts and over 250 staff as well as a wide multi-disciplinary ecosystem across 8 leading Irish universities. We can influence and inform your work, share our networks and collaborate with you to increase your impact, and accelerate your career opportunities. Specifically we offer:

1. Opportunity to build your profile at international conferences and global events.
2. A solid career pathway through formalised training & development, expert one-on-one supervision and exposure to top specialists.
3. A 4 year PhD postgraduate studentship with research student fees covered in full and including a non-taxable stipend (currently €19,000 per annum), along with equipment and travel funding
4. Open to EU and non-EU applicants

Diversity

ADAPT is committed to achieving better diversity and gender representation at all levels of the organisation, across leadership, academic, operations, research staff and studentship levels. ADAPT is committed to the continued development of employment policies, procedures and practices that promote gender equality. On that basis we encourage and welcome talented people from all backgrounds to join ADAPT.