

Post Title: PhD Studentship in Document-level Human Evaluation of Machine Translation

Location: Dublin City University

Anticipated Start Date: 1 Jan 2023

Closing Date: 14th of October 2022

Apply: <https://forms.gle/zVzRVQJ64stX13gG6>

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 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 Fully funded, 4 year PhD postgraduate studentship which includes a stipend of (€18,500 per annum - non taxed), along with equipment, annual travel funding
4. Funding for annual student fees

Context

One of the biggest challenges for Machine Translation (MT), both for developing MT systems as well as for evaluation, is the ability to handle discourse dependencies and the wider context of a document. Work on evaluation has shown that methodologies that rely on single isolated sentences often result in miscalculation because the evaluators cannot access all necessary information for proper decision about context related issues such as gender ambiguity, lexical ambiguity, gender and number agreement, etc. On the other hand, providing evaluators with context avoids such miscalculations. Furthermore, it has shown that many context-related issues vary across different

domains. It is also shown that the context span necessary to solve these types of context-related issues varies, and is often longer than the 2-sentence context commonly used in the machine translation field.

On that account, developing new-generation human evaluation metrics to improve MTEval considering discourse-level features, context span and appropriate evaluation methodology is necessary.

Under the supervision of Dr Sheila Castilho and Dr Maja Popovic, the PhD student will aim to contribute to define which human-evaluation metrics will better handle translations performed at a document-level by analysing the suitability of state-of-the-art metrics in document-level evaluation set-ups, performed by different evaluators, for different goals.

This project is multidisciplinary in nature, and there will be potential for collaboration with experts in other areas of academia. The application of results from this project will be of interest to industry partners in any area of translation.

The PhD will take place in Dublin City University, School of Applied Language and Intercultural Studies.

Minimum qualifications:

Bachelors in Translation, Natural Language Processing or related fields with knowledge of Machine Translation evaluation.

English language requirements for non-native speakers of English is available here, under Post Graduate>>Faculty of Humanities and Social Science:

<https://www.dcu.ie/registry/english-language-requirements-non-native-speakers-english-registry#tab-114966-2>

Preferred qualifications:

- Masters in Translation Technologies, Machine Translation, Natural Language Processing or related fields.
- Experience with machine translation evaluation.

Application Process

As part of your application you will be required to submit

- A letter of introduction (max 1500 words). In the letter, applicants should include the following details:
 - An explanation of your interest in this research, highlighting why you think you are a suitable candidate.

- Highlight any experience you have with machine translation evaluation, or any other relevant experience e.g., working as a translator/linguist, MT developer, or any relevant projects in the area.
- Details of your final year undergraduate project (if applicable), and or of your MSC project (if applicable).
- Details of any relevant modules previously taken, at undergraduate and/or Master level.
- Details of any relevant work experience (if applicable)
- Detailed CV, including – if applicable – relevant publications;
- Transcripts of degrees

Informal inquiries can be sent to sheila.castilho@dcu.ie

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.

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 Dublin City University, 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.

Digitally Enhanced Engagement Strand

From the individual perspective, research within this Strand will deliver proactive agency techniques that sense, understand and proactively serve the needs of individual users to deliver relevant, contextualised and immersive multimodal experiences which also offer them meaningful control over the machine agency delivering those experiences.

Digital Content Transformation Strand

From the algorithmic perspective, new machine learning techniques will both enable more users to engage meaningfully with the increasing volumes of content globally in a more measurably effective manner, while ensuring the widest linguistic and cultural inclusion. It will enhance effective, robust integrated machine learning algorithms needed to provide multimodal content experiences with new levels of accuracy, multilingualism and explainability.

Transparent Digital Governance Strand

From the enterprise and societal perspective, new structured knowledge frameworks and associated practices for AI data governance will be required to balance the needs and values of individuals, organisations and society when it comes to rich digital experiences. This requires the advancement of research in the areas of data ethics, data quality, data protection, data value, data integration, and multi-stakeholder governance models.