

# Enhanced PhD Scholarship in Machine Learning

## National Robotarium, Heriot Watt University, Edinburgh

### Funding

Industrially sponsored enhanced EPSRC Scholarship: Fees paid, plus annual Stipend of £18,622 plus £3,000/yr enhancement.

Stipend normally tax free & normally incremented annually with inflation. Stipend expected to rise for Sep. 2023 intake.

### Technical Area

Application of Large Language models and Hierarchical Topic Modelling for Interactive Interpretation of Massive Knowledge Repositories

### Eligibility

This scholarship is available only to 'home' students.

To be treated as a home student, candidates must meet one of these criteria:

- be a UK national (meeting residency requirements)
- have settled status
- have pre-settled status (meeting residency requirements)
- have indefinite leave to remain or enter.)

Further information from <https://www.ukri.org/what-we-offer/developing-people-and-skills/esrc/funding-for-postgraduate-training-and-development/eligibility-for-studentship-funding/>

### Industrial Sponsor

STMicroelectronics Edinburgh

ST is a large industrial organization which is currently implementing machine learning based tools for search, discovery and interrogation of its large document repositories. It will provide strong support for this PhD in terms of industrial context, and access to large data repositories and computational resource.

### Industrial Context

Advances in machine learning applied to natural language processing (NLP) are revolutionising our ability to search, find, understand and generate textual information. This is an extraordinarily rapidly developing and exciting field of artificial intelligence (AI) research.

There are two key areas where significant advancement has been made. Firstly, training on massive, open, datasets for development of large language models (e.g., ChatGPT) and secondly, training in

specialist technical fields where taxonomies, lexicons and semantics are a key part of the subject knowledge. In these cases the natural control of specialist terminology provides key features to assist the development of machine learning models and the use of AI assistance in processing and understanding large unstructured document sets.

Many technical fields exhibit a less controlled, ambiguous terminology set where context may change the semantics of the terms; the subject field being defined via the topics addressed. The development of topic models for these specialist fields is important when seeking to search, discover and contextualise analysis of document corpora. Such topic models provide critical context which enhances our ability to understand and efficiently obtain value from these document sets.

Being able to build and define such rich topic models is expected to aid additional natural language processing (NLP) tools in searching, separating and understanding document content and provide thematic overviews of these repositories.

This project seeks to research the ability to form these topic models in narrow technical fields using large language models and classical machine learning techniques, and then use and embed them within additional NLP tools to digest, understand, search and sort documentation sets.

## Location

National Robotarium, Heriot Watt Riccarton and STMicroelectronics Edinburgh

## Requirements

2:1 or First-class degree in Computer Science or related discipline, plus bags of curiosity and drive.

This is an ideal project for someone who wishes to deepen their existing knowledge of natural language processing or language models while researching new methodologies which improve and enhance the human / machine interface in the area of document search, sort and understanding.

## Duration & Start Date

3.5 years. September 2023 – December 2023 start

## Further info

Email Prof. Mike Chantler ([m.j.chantler@hw.ac.uk](mailto:m.j.chantler@hw.ac.uk)) with brief CV in first instance.