



School of Computing, Engineering, and the Built Environment Edinburgh Napier University

PHD STUDENT PROJECT

Funding and application details

Funding status: Self-funded students only

Application instructions:

Detailed instructions are available at <https://www.napier.ac.uk/research-and-innovation/research-degrees/how-to-apply>

Prospective candidates are encouraged to contact the Director of Studies (see details below) to discuss the project and their suitability for it.

Project details

Supervisory Team:

- DIRECTOR OF STUDY: David Haynes (Email: D.Haynes@napier.ac.uk)
- 2ND SUPERVISOR: Sarah Thomson

Subject Group: Applied informatics

Research Areas: Artificial Intelligence

Machine Learning
Digital Media

Project Title: Using ontologies to explain AI based risk assessment

Project description:

The proposed new EU legislation on artificial intelligence will increase demand for systems that can explain decisions made by AI systems.

The research will investigate the ways in which AI systems are used to assess risk. The objective will be to develop a method of explaining decisions made using AI. This will build on previous work that has distilled the knowledge embedded in an AI system into taxonomies. Ontologies offer a richer way of representing knowledge. You will investigate whether ontologies are a more effective way of

distilling the knowledge from an AI system. This research will focus on discovering the implied relationships between concepts and the axioms or assumptions on which AI decisions are based. Criteria for assessing the effectiveness of the techniques will need to be established at an early stage.

The successful candidate will be a part of both the Applied Informatics and Evolutionary Computing groups at Edinburgh Napier University.

References:

- [1] Haynes, D. (2020). Understanding Personal Online Risk To Individuals Via Ontology Development. In Knowledge Organization at the Interface: Proceedings of the Sixteenth International ISKO Conference, 2020, Aalborg, Denmark (171-180). <https://doi.org/10.5771/9783956507762-171>
- [2] Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531.
- [3] Tang, Y., da Costa, A. A. B., Zhang, J., Patrick, I., Khastgir, S., & Jennings, P. (2023). Domain Knowledge Distillation from Large Language Model: An Empirical Study in the Autonomous Driving Domain. arXiv preprint arXiv:2307.11769.
- [4] Thomson, S. L., van Stein, N., van den Berg, D., & van Leeuwen, C. (2023). The Opaque Nature of Intelligence and the Pursuit of Explainable AI.

Candidate characteristics

Education:

A second class honour degree or equivalent qualification in Information Science, Computing, Machine Learning

Subject knowledge:

- Machine Learning
- Computational Linguistics
- Ontology development
- Knowledge Organizing Systems (KOS)

Essential attributes:

- Curiosity and persistence
- Creativity
- Attention to detail

Desirable attributes:

- Do you have a distinction at MSc level in Computing or Information Science? If you have a good understanding of AI systems and ontologies, come and join our lively and supportive research community. A background in knowledge organization with experience