

Dear Applicant,

The School of computer Science & Informatics at Cardiff University will soon advertise several PhD projects, which will be fully funded by the School. I am involved in two such projects, one as the lead supervisor and the other as a co-supervisor.

Project #1: Natural language processing

Title: Integrating numbers into document representation models for improved text classification

Clinical text often includes numbers of various types and formats. However, most current text classification approaches do not take advantage of these numbers. A recently published article "The Value of Numbers in Clinical Text Classification" showed that even a handful of numerical features can significantly improve text classification performance. It concluded that commonly used document representations (vector space model and word embedding model) do not represent numbers in a way that machine learning algorithms can effectively utilize them as features. Although this study demonstrated that traditional information extraction can be effective in converting numbers into features, further research is required to systematically incorporate numbers into different document representation models. This project aims to do exactly that by exploring an array of different supervision approaches, including but not limited to self-supervised representation learning, distant supervision, reinforcement learning, etc. The success of these approaches will be measured by comparing the performance of classification algorithms on such representations relative to existing document representations.

Miok, Kristian, Pdraig Corcoran, and Irena Spasić. 2023. "The Value of Numbers in Clinical Text Classification" Machine Learning and Knowledge Extraction 5, no. 3: 746-762.

<https://doi.org/10.3390/make5030040>

Keywords: natural language processing, artificial intelligence, large language models.

Contact for more information on the project: **Professor Irena Spasić**; spasici@cardiff.ac.uk

Project #2: Cryptocurrencies

Title: Efficient and scalable consensus algorithms for decentralised cryptocurrencies

A decentralised cryptocurrency, such as Bitcoin invented by Satoshi Nakamoto, is a digital currency that can be transferred between users in a peer-to-peer manner without a centralised authority such as a Bank. Cryptocurrencies offer many potential benefits relative to traditional centralised fiat currencies such as the US dollar. This includes cheaper transaction costs relative to traditional payment channels, improved privacy and providing financial services to the 1.7 billion unbanked individuals in the world. To achieve decentralisation, cryptocurrencies use a consensus algorithm to achieve consensus with respect to the ownership of the cryptocurrency in question. Many cryptocurrencies, including Bitcoin, use a consensus algorithm that is based on a proof-of-work (POW) algorithm. POW is computationally expensive, and this leads to more expensive transaction fees and environmental concerns. Consequently, developing more efficient and scalable solutions to consensus in decentralised and permissionless systems represents an important and impactful research problem. This project aims to develop novel solutions to this problem through the application of artificial intelligence (AI) and network science. The deliverables from this project will be a system that is evaluated in simulation and on historical transaction data. This evaluation will aim to demonstrate that the proposed system achieves greater efficiency and scalability.

Keywords: cryptocurrency, consensus algorithm, scalability, blockchain.

Contact for more information on the project: **Dr Padraig Corcoran**; corcoranp@cardiff.ac.uk

Please submit your application before the application deadline of the **6th May** via [Computer Science and Informatics - Study - Cardiff University](#)

Academic criteria: A 2:1 Honours undergraduate degree or a master's degree, in computing or a related subject. Applicants with appropriate professional experience are also considered. Degree-level mathematics (or equivalent) is required for research in some project areas. Applicants for whom English is not their first language must demonstrate proficiency by obtaining an IELTS score of at least 6.5 overall, with a minimum of 6.0 in each skills component.

How to apply:

Please develop an individual research proposal that builds on the information provided in this advert. Once you have developed the proposal, please share it with your supervisor to check if she would provide an initial agreement of support. If so, please submit your application following the instructions provided below.

In order to be considered candidates must submit the following information:

- Supporting statement
- CV
- In the 'Research Proposal' section of the application enter the name of the project you are applying to and upload your Individual research proposal, as mentioned above in BOLD
- In the funding field of your application, insert "I am applying for the 2024 PhD Scholarship in Computer Science and Informatics", and specify the project title and supervisors of this project in the text box provided.
- Qualification certificates and Transcripts
- References x 2
- Proof of English language (if applicable)

If you have any questions on the application process, please contact COMSC-PGR@cardiff.ac.uk

The projects will soon be advertised on <https://www.findaphd.com/>. However, you now have all relevant information to help you start preparing an application if you are interested in doing a PhD. The shortlisted candidates will be interviewed by a panel of academic staff who will make the final decision. One of the criteria for shortlisting will be your ability to research and solve problems. For example, you should be able to research online how to write research proposals.

Best wishes,

Prof. Irena Spasić

<https://users.cs.cf.ac.uk/I.Spasic/>