Computer Science:
Northeastern University London Fully Funded PhD Scholarship

A comprehensive test framework for deep learning models

Northeastern University London > Computer Science

Deadline: 13 March 2023

Funded PhD Project (UK or International Students)

Funding provider: Northeastern University London (NU London)

Subject areas: Computer Science, Artificial Intelligence

Project start date: 24 April 2023 or September 2023

Supervisors (*lead):

- Dr Alexandors Koliousis* (Northeastern University London)
- Dr Elena Botoeva and Prof Alex Freitas (University of Kent)

Aligned programme of study: PhD in Computer Science

Mode of study: Full-time

Northeastern University London

As part of a major investment, Northeastern University London (NU London) has multiple, fully-funded PhD studentships available to accelerate its interdisciplinary research in the humanities, social sciences and digital sciences. Each scholarship is fully-funded for three and a half years (UKRI rates) and includes fees, an annual stipend, an additional London allowance and associated costs, such as training.

NU London is the European campus of Northeastern University, a large, top-tier research intensive, Boston-based institution. With campuses across the United States, Canada and London, students will have the opportunity to engage with, and visit the Northeastern University network overseas, as part of their London-based doctoral studies, providing a truly unique and highly sought-after dimension to their research training.

The Project

Deep learning models, usually based on deep neural networks, are highly prominent at the moment and being extensively used both in academia and industry for prediction tasks such as image classification, image captioning, and question answering. But we have also begun to observe errors, typically detected when encountering one or more counterexamples: inputs that, when slightly modified, are no longer correctly predicted by the model (e.g., a rotated image, or an image with an imperceptible perturbation of its pixel values). A good software development practice is to introduce a test case whenever an error is found and fixed. Put together, these test cases create a “safety net” for users. However, developers still lack *a comprehensive test framework to systematically detect and characterise errors in deep learning models*.

Deep learning models use known input-output pairs to learn (millions of) parameters so that, layer by layer, they can represent inputs in such a way that each input class is separable from one another in the output space. They err when the learned representation of a perturbed input falls outside the decision boundaries learned for its input class. The problem of error testing is to define a set of test cases based on the model’s deployment requirements and architecture design to assert that the system behaves as specified. This set should be considered sufficient when it includes tests on each decision boundary and inputs on each side of each boundary for each class of inputs.

The overall goal of this PhD project is to streamline the specification and evaluation of well-formed test cases for deep learning models that can be easily fine-tuned to a particular deployment environment and model architecture with just a few parameters. There are three main challenges in this project.

Global probabilistic guarantees. Consider that we are given a trained model and the data used to train and evaluate it; our task is to check for errors. We could use formal neural network verification techniques to test that, for a given input, the model is robust to all perturbations within bounds. Suppose that the model passes the test. In other words, we have proven that all bounded perturbations of that particular input do not affect the model’s prediction. How can we ensure that this local property holds for all inputs in that class or, globally, for each input class? The idea is to provide probabilistic guarantees about the overall correctness of the model by efficiently sampling our input data distribution. Samples must be part of the input distribution, they must be diverse enough to capture all modes of the data distribution, and they must generalise beyond the training data. One way forward, for example, is to learn a generative model, such as variational auto-encoders, for generating inputs. As with traditional software testing, the challenge remains to ensure high test coverage and do so in a scalable way.

Error summarisation. Suppose that our model fails to pass the aforementioned local robustness test. In other words, we have proven that there is at least one input perturbation that causes the model to mispredict it. Which counterexample(s) to report? Within the bounded input region considered, there can be more than one counterexample; and different counterexamples may cause different misclassifications depending on which decision boundary they cross, thus representing different errors. We want to devise a way to systematically navigate the bounded input region, characterise errors found therein, and return a summary.

Property-based test generation. Different model architectures use different learning algorithms (e.g., convolution, or attention) to accomplish a prediction task. Robustness to adversarial inputs is a desirable property for all of them, but robustness criteria depend on the task's deployment requirements (e.g., robustness to semantic perturbations such as rotation, scale, and brightness for image perception models), as well as the inductive bias(es) of learning algorithms (e.g., translation equivariance for convolution, or permutation equivariance for attention). We want to devise a way for developers to specify properties to verify and then automatically generate tests from them so as to cover all relevant test cases.

The successful candidates will:

- Have a proven, strong educational background in computer science, mathematics, or related subject (see eligibility criteria)
- Have a good background in software development
- Be highly motivated and excited to engage in research in the proposed project area
- Be an independent learner, willing to challenge themselves
- Have strong communication skills

The successful candidates will benefit from a brand new campus on the banks of the River Thames next to Tower Bridge; an interdisciplinary, vibrant research environment; international collaboration and networking opportunities, dedicated research space and a highly experienced, multi-institution supervisory team from NU London and the University of Kent.

Shortlisted candidates will be interviewed in March 2023. Candidates are welcome to contact the NU London supervisor with informal enquiries before the application deadline: Dr Alexandros Koliousis (alexandros.koliousis@nchlondon.ac.uk).

Eligibility

- Bachelor's degree (essential) or Master’s degree (optional) in a relevant subject: upper-second class (2:1) or first-class honours (1st).

English Language requirements

If applicable – IELTS 6.5 overall (with a score of at least 6.5 in each individual component) or equivalent.

Nationality

Applications are open to UK and international students. Please indicate if you are likely to require a visa on your application form.

Funding

This scholarship covers the full cost of tuition fees, an annual stipend and an additional London allowance (set at UKRI rates) for 3.5 years. For the 2022/2023 academic year the annual stipend is £19,668 (£17,668 UKRI stipend plus £2,000 London allowance). Annual increments will be in line with UKRI rates.

International travel

Students will have the opportunity to optionally travel to Northeastern University in North America to further their research training and experience.

How to Apply

Please submit a **CV** and a **1-page covering letter** stating how you meet the requirements and why you are interested in the proposed research project. Please also include transcripts of your qualifications (incl. English, if applicable) with your submission. Apply by clicking on this link. Please reference your application **‘PS2CS0223’.**