Northeastern University London

University of Kent

# Computer Science:<br>Northeastern University London Fully Funded PhD Scholarship

## Scalable verification of 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, capture the zeitgeist of the “AI spring” we have been experiencing for the past decade or so. At the same time, it is also widely recognised that they are vulnerable to *adversarial attacks*. Consider, for example, an image classification model. Pick a correctly classified image and add some noise to it. Even if the perturbation is small, small enough for a human to still perceive the same image, it is still possible for the model to misclassify it, in which case the model is said to be not robust to such perturbations. The existence of such adversarial examples hinder the wider adoption of neural networks in safety-critical systems (e.g., vision models in self-driving cars).

Formal verification of neural networks is looking to provide *formal guarantees* that a model behaves correctly. Correctness is specified in terms of whether a formal property is satisfied or not. For instance, the *local robustness* property states that a neural network is robust to all bounded perturbations of a given input. Violations of a property can be proved by generating one or more counterexamples: samples from the bounded input space for which the model violates the property (e.g., perturbed images misclassified by the model).

When testing a model on a single input, the values of all intermediate and output neurons are uniquely determined, so it is straightforward to check whether a property holds or not. With formal verification, however, we consider continuous input sets (subsets of real vector space). So, every neuron can take values from some interval and can generally be either activated (e.g., value greater than zero for ReLU) or not (e.g., zero value for ReLU) in that interval. This creates a combinatorial search space that needs to be explored in order to prove or disprove a property.

In order to efficiently navigate this search space, state-of-the-art verification techniques follow the branch-and-bound approach, combining the strengths of fast but incomplete and those of complete but slow approaches. Bounding typically involves incomplete methods to quickly get an answer based on some relaxation, e.g., overestimating the bounds of output neurons by the symbolic propagation of the input bounds through the network. If the approximation-based answer is not satisfactory, branching can recursively split the current problem into easier subproblems, e.g., by picking an “unstable” neuron (namely, a neuron that can be both active and inactive within the bounds considered) and creating two branches, one where the neuron is active and one where it is inactive. Good branching heuristics are crucial for the performance of branch-and-bound algorithms, as they can drastically reduce the size of the explored search space. Existing verification tools mostly employ heuristics engineered by experts for choosing which neurons to branch on. Despite recent progress, scalability remains one of the main challenges in verification of neural networks. We ask: how to derive efficient branch-and-bound algorithms that can verify models with (tens of) millions of “neurons” in reasonable time?

Northeastern University London

University of Kent

The overarching goal of the proposed PhD project is to design and implement a verification tool that is able to verify deep learning models with millions of neurons. The main ideas for conquering scalability are (i) to learn, rather than hand-craft, efficient branching heuristics to guide branch-and-bound algorithms; and (ii) to make best use of computational resources at every step of the learning and verification processes.

*Learning branching heuristics.* Crafting branching heuristics manually is time-consuming and requires a lot of insight into the problem at hand that is often hindered by the high dimensionality of intermediate neural network representations. Learning for combinatorial optimisation aims to derive good branching heuristics automatically. One aspect of this PhD project is to design and train an auxiliary machine learning model that computes efficient branching heuristics tailored to the problem of verification of neural networks.

*Hardware-efficient implementation.* Good theoretical properties of branching heuristics do not necessarily guarantee good overall speed-up of the verification tool. So, another aspect of this PhD project is to implement the overall verification process efficiently by harnessing multiple multi-core machines with heterogeneous processors such as GPU or TPU accelerators. The project will seek ways to accelerate all stages of the verification process, from feature extraction for training the auxiliary model, to training of the auxiliary model itself, to integrating the learned branching heuristic into the branch-and-bound algorithm, to processing the separate branches.

**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 '**PS1CS0223'.**