DEPARTMENT OF
COMPUTER
SCIENCE

UNIVERSITY OF
OXFORD

Professor Marta Kwiatkowska
Fellow of Trinity College
Direct Line Tel: +44 (0)1865 283509
Email: marta.kwiatkowska@cs.ox.ac.uk

Personal Assistant: Karla-Maria Perez Blanco
Direct Line Tel: +44 (0)1865 283581
Email: karla.perez@cs.ox.ac.uk

27 May 2025

Professor Yulia Timofeeva
Head of Department
Department of Computer Science
University of Warwick, UK

Dr Min Wu

I am delighted to write a recommendation for Dr Min Wu, who is applying for the position of an Assistant Professor in your department. Min completed her doctorate (DPhil in Oxford's terminology) in early 2020 with the dissertation entitled "Robustness Evaluation of Deep Neural Networks with Provable Guarantees" working under my supervision. Her thesis examiners were Prof. Patrick Frossard (EPFL), an expert in computer vision, and Prof. Niki Trigoni (Department of Computer Science), who is working with machine learning applications in sensor networks. For the period from October 2019 to March 2021, Min has been employed as a Research Assistant on one of my projects (funded by Innovate UK), and since approx. 2022 has been a postdoctoral researcher at Stanford in the group of Prof. Clark Barrett.

Before coming to Oxford, Min studied at Tongji University, Shanghai, which is one of the top universities, where she was an outstanding student academically, graduating with a rank 1/70, with a GPA 91.37/100. Min's achievements were recognized through the Ji Yang scholarship (at Tongji); a 2014 Google Anita Borg scholarship, a highly competitive scholarship for women; and a full scholarship to study for a doctorate at Oxford from the China-Oxford Foundation.

During her DPhil at Oxford, Min has focused on robustness methodologies for deep learning, a new direction initiated in my group and inspired by the need to ensure safety of autonomous driving. I began working on this topic as part of the EPSRC Programme Grant on Mobile Robotics and more recently my second ERC Advanced Grant FUN2MODEL (www.fun2model.org). Since autonomous driving relies on perception, typically implemented as a (deep) neural network, concerns have been raised as to whether such software is safe in view of neural networks being susceptible to adversarial examples, i.e., small perturbations of the input that are imperceptible to a human but change the classification of the network. I gave a keynote at CAV 2017 (the flagship verification conference) on this topic, rallying the verification community to tackle this problem. My invited paper, which was one of the first two papers in this topic in the CAV conference (the other paper being from Clark Barrett at Stanford), defined the concept of safety/robustness of neural network classification decisions with respect to adversarial 'manipulations', and a layer-by-layer analysis method based on Satisfiability Modulo Theory (SMT) to verify or falsify (under some assumptions) the safety of image classification problems. This proof-of-concept implementation showed promise but was not scalable. Working with other members of my group, Min contributed to a journal extension of the CAV 2017 invited paper, published in Theoretical Computer Science, which proposed the first practical verification framework for neural networks. Min worked on a game-based approach for computing lower and upper bounds for the "maximal safe radius" (MSR) around an input (essentially the same concept as 'safety/robustness', also studied under the name of the minimal distance to the decision boundary), within which adversarial examples are guaranteed to not exist, developing the software and contributing to the theoretical formulation. The approach employs two players, where one player selects features to manipulate, and the other player selects pixels within the feature, and players act cooperatively or competitively. This was applied both in pointwise as well as feature-based setting on state-of-the-art neural networks in computer vision. This work was also extended to global robustness for the Hamming distance (challenging because of non-differentiability), in the sense of computing the expected maximal safe radius over the test dataset (published at IJCAI 2019, leading conference in Artificial Intelligence, which Min presented at the conference), and "concolic" (concrete and symbolic)

DEPARTMENT OF
COMPUTER
SCIENCE

UNIVERSITY OF
OXFORD

Professor Marta Kwiatkowska
Fellow of Trinity College
Direct Line Tel: +44 (0)1865 283509
Email: marta.kwiatkowska@cs.ox.ac.uk

Personal Assistant: Karla-Maria Perez Blanco
Direct Line Tel: +44 (0)1865 283581
Email: karla.perez@cs.ox.ac.uk

testing for neural networks (published at ASE 2018), where Min made a strong contribution to the development of the theory and software tools.

Beyond collaborating with other members of my group, Min also led on a paper accepted to CVPR 2020 (with oral presentation) co-authored with me as supervisor. In this paper, she extended the game-based approach to videos, demonstrating not only that video inputs are susceptible to adversarial manipulations, but also how to prove their robustness against bounded perturbations. She proposed a novel and clever extension, where the game proceeds by manipulating optical flows, rather than pixels, and an additional challenge was the need to handle recurrent neural networks (LSTMs in this case) since the video processing networks combined CNNs with LSTMs. Separately, in a collaboration with a group at Leeds University, Min also developed a predictive method for gaze-based intention anticipation in the context of driving, working with data from the Leeds driving simulator, which she also presented at IROS 2019, a leading robotics conference.

Min's DPhil thesis integrated the various contributions into a comprehensive treatment of adversarial and global robustness for a range of neural network architectures and applications, focused on methodologies for lower/upper bounding of MSR. This was one of the first practically relevant approaches, with promising case studies that brought new insight into understanding of the instabilities of neural networks and their effect on performance, not only at pixel level but also a feature level.

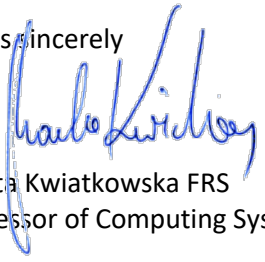
After defending her thesis in February 2020, Min contributed to my Innovate UK project (initially part-time due to the limitations of her Tier 4 visa and then full-time from June 2020 until the funding finished in March 2021) by extending the game-based method to NLP (Natural Language Processing) tasks. Along with many junior researchers at the time, she was hampered by COVID-19 and her move to Stanford was somewhat delayed. I have not followed her Stanford career but am very pleased to see the new contributions she has made and her publications success.

While at Oxford, Min was very collegial and open to collaborations. She frequently engaged in research discussions and seminars, often asking questions and making various suggestions. She is thorough, clever and has an excellent understanding of the emerging field of robustness for machine learning, both at the theoretical and practical level. Min has also engaged in teaching, and taught classes for Probabilistic Model Checking and labs on the AIMS Centre for Doctoral Training, with very positive comments from students.

Summarising, Min has been able to shape the field of AI safety more or less from the beginning of her DPhil and did this with relish. In my opinion, she is self-motivated, enthusiastic about her research topic, and was able to make a very strong contribution to this emerging and important field. Her track record of publications is outstanding for someone at her stage of career, with some well cited papers. She is a pleasure to have around. I recommend her to you strongly and without reservations.

Should you need any further information do not hesitate to get in touch with me.

Yours sincerely



Marta Kwiatkowska FRS
Professor of Computing Systems