STANFORD
COMPUTER SCIENCE

David L. Dill
Donald E. Knuth Professor, Emeritus, in the School of Engineering
Department of Computer Science

630 Park Rd
Emerald Hills, CA 94062

May 27, 2025

Dear Search Committee,

I am very pleased to write this letter of reference for Min Wu for a faculty position in computer science or a related field. I have been working with her informally at Stanford for several months. I have found her to be a talented researcher who is very knowledgeable about both formal methods and artificial intelligence, and she is creative and hard-working. She has a strong research and publication record as a result.

To introduce myself, I was a professor of computer science at Stanford University for thirty years, doing research in formal verification and computational biology. I became emeritus at Stanford in 2017, took a job at Facebook/Meta in 2018 and finally retired 2023. I am a member of the National Academy of Engineering and the American Academy of Arts and Sciences. I am a co-author of one of the more highly cited papers in Min's area on formal verification of deep neural networks ("Reluplex").

Min's research is in AI safety, and it combines formal methods and deep learning. A basic conflict between the two fields is that formal methods rely on precise logical specifications of system properties, but neural networks are trained by example to solve poorly specified problems. Min was a co-author of a one of the founding papers combining these fields (in 2017, published earlier as a tech report), which tackled the specification problem by focusing on proving the robustness of a network to adversarial perturbations of its inputs. Specifically, the authors found a method to show that perturbations within a specified radius of a particular input vector did not change the classification of the input. This paper now has over a thousand citations on google scholar.

In later work, Min expanded on the idea of robustness in various ways. I won't try to be comprehensive but instead focus on a few examples. One advance was to quantify robustness by computing bounds on the maximal safe radius around an input point where the classification of the point remains constant. The maximal safe radius idea was applied to the very different domains of image classification and natural language. The notion of a maximal safe radius has turned out to be so fundamental that it has been incorporated into an ISO standard for AI safety.

More recently, Min has focused on provable versions of other existing neural network analysis methods. For example, Min has developed a method to "explain" results of neural nets by finding a minimal sufficient set of inputs to determine the outputs, which she calls a *minimal explanation set.* The minimal explanation set is useful for evaluating the trustworthiness of a neural net. For example, the classification of a road sign should depend on meaningful pixels with the shape and content of the sign, not bolts or raindrops. Impressively, she was able to apply minimal explainable sets to a real-world system for autonomous aircraft taxiing that used images of runway markings under the aircraft.

Phone: (650) 743-7677 • E-mail: dill@cs.stanford.edu

Several of the analysis methods above make use of Marabou, a sophisticated optimization program developed specifically for neural nets. Min was a member of the team that developed this widely used tool.

In my recent conversations with Min, it became clear that she has many more problems to work on. She has convinced me that other problems on neural networks, such as interpretability, and benefit from being made provable. Also, while small neural networks are important in critical applications, many networks are gigantic, so scaling her methods to larger systems will be a continuing challenge, for which Min has the appropriate talents.

Min has been successful in a very difficult research area for several reasons. She is creative about finding new problems and formulating them to be both solvable and useful. She is clever at coming up with new algorithms and heuristics for solving hard computational problems, and she puts in the effort to evaluate and improve them on real networks. She also seeks out real-world applications of her methods.

In summary, she has the potential to be a great researcher, and I believe that she'll be very successful in a faculty position. In many departments, she would also be valuable as a bridge between researchers in formal methods and artificial intelligence (I expect she could teach in both areas). I recommend her with great enthusiasm.

Sincerely,

David L. Dill

Phone: (650) 743-7677 • E-mail: dill@cs.stanford.edu