Zhezhi (Elliot) He
Ph.D. candidate at Arizona State University

Title:
"Secure and Efficient AI Computing System: A Software and Hardware Co-design Perspective"
Monday, March 16, 1:00pm - 2:00pm HEC 438

Abstract
Recent advances in Deep Neural Network (DNN) have demonstrated its great potential for achieving human-surpassing performance in multiple domains. On the one hand, researchers still have insufficient knowledge about the principles of DNN. Thus, DNN is like "the sword of Damocles" where its security or reliability is an essential concern that cannot be circumvented. On the other hand, the impressive performance achieved by DNNs normally accompanies with the drawbacks of intensive memory and power usage, due to enormous model size and high computation workload. It significantly hampers their deployment on the resource-limited cyber-physical systems or internet-of-things on the edge. In this talk, I will first discuss a new paradigm of the adversarial attack on the DNN weights, where a few malicious bit-flips can cause fully malfunction of target DNN, and its corresponding defense method. After that, I will discuss the DNN compression technique using the low bit-width quantization together with its hardware accelerations. I will summarize my talk with a vision of the future software-hardware collaborative design of trustworthy and efficient systems for artificial intelligence and big-data processing.

Biography
Zhezhi (Elliot) He is currently a Ph.D. candidate at Arizona State University advised by Dr. Deliang Fan. Before that, he received the M.Eng. degree in electrical and computer engineering from Oregon State University, Corvallis, OR, USA, in 2016, and the B.S. degree in Information Engineering from Southeast University, Nanjing, China, in 2012. His research interests are in the area of secure and efficient deep learning, brain-inspired in-memory computing, and post-CMOS technologies. During his Ph.D. study, he has been the primary author of 47 publications on IEEE/ACM top-tier journal and conference (e.g., CVPR, ICCV, AAAI, DAC, DATE, ICCAD, ISLPED, TCAD, TETC, etc) in the aforementioned areas. His works also result in the receipt of 2 Best Paper Awards from IEEE ISVLSI 2017 and 2018, respectively. He is also serving as the member of technical program committee and reviewer for various conferences and Journals (TNNLS, AI, CVPR, HPCA, DAC, DATE, GLSVLSI, TCAS-I, TED, etc.).