“Trustworthy and Explainable Artificial Intelligence”

Wednesday, May 17, 2023 • 11:00AM • HEC 101A • Zoom

Abstract:
Artificial Intelligence (AI) is becoming ubiquitous and a key that enables humankind to reach unprecedented horizons (e.g. in protein folding). However, in high-stake applications (e.g. cancer diagnosis, automated target detection, or face identification), a single incorrect decision of AIs could change lives entirely. My three burning research questions are: (1) How to build AIs that are accurate in new edge cases in the evolving world? (2) How to build AIs that maximize the human-AI team accuracy? (3) How to build AIs whose decision-making can be understood and debugged by humans? In this talk, I will share our recent research in both Computer Vision and Natural Language Processing in these three thrusts. I am also excited to discuss Bottleneck XAI, my proposed architecture for future Trustworthy and Explainable Artificial Intelligence systems. Bottleneck XAI would enable users to (a) understand and learn from AIs; (b) audit; and (c) edit AIs.

Bio:
Anh completed his Ph.D. from the University of Wyoming working with Jeff Clune and Jason Yosinski in 2017 and since then has been an Assistant Professor at Auburn University. He also worked at Apple and Geometric Intelligence (acquired by Uber). In a previous life, he enjoyed building web interfaces at Bosch and invented a 3D input device for virtual reality (covered on MIT Tech Review). Anh is interested in making AIs more trustworthy as well as explainable, and in understanding their inner workings. His research has won 3 Best Paper Awards (CVPR 2015, GECCO 2016, ICML 2016 Visualization workshop), a Best Application Paper Honorable Mention (ACCV 2020), and 2 Best Research Video Awards (IJCAI 2015 & AAAI 2016). His work has been covered by many media outlets e.g. MIT Technology Review, Nature, Scientific American, and lectures at various institutions. Anh was awarded an NSF CAREER award (2022).