Yiwei Yang

University of Washington

819 Northeast 67th st
Seattle, WA 98115

yanyiwei.github.io
yanyiwei@uw.edu

Education

10/20 - present **University of Washington**
Seattle, WA PhD Student, Information School
Research Interests: Interpretability and Fairness in Machine Learning
Advisor: Bill Howe

09/15 – 05/19 **University of Michigan**
Ann Arbor, MI B.S. in Computer Science and Engineering

Professional Experience

10/20 - present **University of Washington**
Seattle, WA Graduate Student and Researcher

05/18 – 08/18 **IBM Research, Almaden**
San Jose, CA Research Intern (Mentor: Yunyao Li, Eser Kandogan, Prithviraj Sen)

Research Projects

02/22 - Present **Group-level Feature Visualization for Fair Image Classification**
UW Working on a novel feature visualization method that surfaces what an image classifier has learned about a target class belonging to a group of interest (e.g. female doctor vs male doctor)

05/21 - 01/22 **XAI meets Fairness: Co-Regularization of Procedural and Distributive Fairness in Neural Networks**
UW Introduced a novel fairness metric for procedural fairness by computing the divergence of feature attributions between different groups (*In Submission to FAccT 2022*)

07/19 – 01/20 **Generating User-efficient Clarification Questions**
U of M Worked on a human-machine hybrid approach that leverages human intelligence and ranking by information gain to select questions that balance the trade-off between information gain and answerability

05/18 – 05/19
IBM Research

Learning Linguistic Expressions with Deep Learning and Human-in-the-Loop
Worked on a novel human-in-the-loop interaction paradigm which learns first-order-logic rules with deep learning and recruits people to select rules that generalize beyond the training data; this work is featured on IBM Research Blog https://www.ibm.com/blogs/research/2019/07/heidl-acl2019/

05/17 – 09/17
U of M

Bolt: Instantaneous Crowdsourcing Via Just-in-Time Training
Worked on a "look-ahead" approach which breaks the cognitive barrier of humans by predicting possible future states and fetching actions to each state beforehand through crowdsourcing

Publications

Conference Full Papers

C.04 A. Lundgard, **Y. Yang**, M.L. Foster, W.S. Lasecki. Bolt: Instantaneous Crowdsourcing via Just-in-Time Training. *In Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2018). Montreal, Canada.*

C.03 S.W. Lee, Y. Zhang, I. Wong, **Y. Yang**, S. D. O'Keefe, W.S. Lasecki. SketchExpress: Remixing Animations for More Effective Crowd-Powered Prototyping Of Interactive Interfaces. *In Proceedings of the ACM Symposium on User Interface Software and Technology (UIST 2017). Quebec City, Canada.*

C.02 H. Kaur, M. Gordon, **Y. Yang**, J. Teevan, E. Kamar, J. Bigham, W.S. Lasecki. CrowdMask: Using Crowds to Preserve Privacy in Crowd-Powered Systems via Progressive Filtering. *In AAAI Conference on Human Computation Demos (HCOMP 2017), Quebec City, Canada.*

C.01 Y. Chen, S.W. Lee, Y. Xie, **Y. Yang**, W.S. Lasecki, S. Oney. Codeon: On Demand Software Development Assistance. *In Proceedings of the International ACM Conference on Human Factors in Computing Systems (CHI 2017), Denver, USA.*

Workshop/Demo/Posters

P.03 **Y. Yang.**, E. Kandogan, Y. Li, W.S.Lasecki, P.Sen. HEIDL: Learning Linguistic Expressions with Deep Learning and Human-in-the-Loop. *In Proceedings of the Association for Computational Linguistics (ACL 2019). Florence, Italy.*
(**Best Poster** at Michigan AI Symposium, 1/55)

P.02 **Y. Yang.**, E. Kandogan, Y. Li, W.S.Lasecki, P.Sen. A study on Interaction in Human-in-the-Loop Machine Learning for Text Analytics. *Joint Proceedings of the ACM IUI 2019 Workshops co-located with the 24th ACM Conference on Intelligent User Interfaces (ACM IUI 2019), Los Angeles, USA,* March 20, 2019.

P.01 S. W. Lee, **Y. Yang**, S. Yan, Y. Zhang, I. Wong, Z. Yan, M. McGruder, C. M. Homan, W. S. Lasecki. Creating Interactive Behaviors in Early Sketch by Recording and Remixing Crowd Demonstrations. *In AAAI Conference on Human Computation Demos (HCOMP 2016), Austin, TX*.