Yunfeng Zhang

Machine Learning Engineer / AI Researcher

Chappaqua, NY 10514
zywind@gmail.com
linkedin.com/in/zywind/

SUMMARY

- AI/HCI researcher dedicated to developing fair and interpretable AI systems.
- Proficient in machine learning operations (MLOps) and infrastructure development.
- Published over 50 papers with more than 2000 citations. Filed 13 patent applications.

EXPERIENCE

Twitter Inc., NYC - *Senior Machine Learning Engineer*

June 2021 - PRESENT

- Architected and developed Twitter's first AI fairness evaluation system based on Tensorflow and Apache Beam. The system has been applied to recommendation and content moderation models for hundreds of millions of users.
- Researched and developed new and more effective AI fairness and performance metrics.

IBM, T. J. Watson Research Center - *Research Staff Member*

May 2016 - May 2021

- Developed IBM's **AI Fairness 360** and **AI Explainability 360** toolkits. Received **IBM outstanding research accomplishment award**.
- Researched and developed novel methods to detect and remediate AI/machine learning bias, as well as methods for explaining machine learning models and AI systems.
- Researched and developed various prototype systems on active learning, drift detection, and chatbot authoring to inform designs of production systems such as Watson OpenScale, IBM AutoAI, and IBM Watson Assistant. Received **IBM outstanding accomplishment award**.

IBM, T. J. Watson Research Center - *Postdoctoral Researcher*

June 2015 - May 2016

- Designed and implemented AI-driven multimodal interaction techniques for smart meeting rooms by incorporating gesture, speech, and face recognition techniques.
- Designed and developed an application development framework for distributed, multimodal applications.

IBM, T. J. Watson Research Center - *Research Intern*

May 2014 - September 2014

- Researched methods to remediate cognitive biases in AI-assisted human decision making.

Palo Alto Research Center - *Research Intern*

May 2013 - December 2013

- Developed computational models to simulate and predict how humans detect changes in stochastic environments.

EDUCATION

University of Oregon, Eugene - *Ph.D. in Computer and Information Science,* June 2015

Thesis: *Towards a principled, data-driven approach for developing and evaluating computational cognitive models: Applying high performance computing and advanced sampling algorithms to cognitive modeling*

University of Oregon, Eugene - *M.S. in Computer and Information Science,* June 2013

Beijing Normal University, Beijing - *B.S. in Computer and Information Science,* June 2007

PUBLICATIONS and PATENTS

I have published over 50 papers with more than 2000 citations, 3 patents, and 9 pending patent applications. For more details, check out my Google scholar page. Selected publications:

- De-biasing "bias" measurement
- AI Fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias
- One explanation does not fit all: A toolkit and taxonomy of AI explainability techniques

AWARDS

- IBM Outstanding Accomplishment Award for Research Advancements to Conversational Technology, 2020.
- IBM Outstanding Research Accomplishment Award for Trustworthy AI, 2019.
- Annual Conference of the Cognitive Science Society, Computational Modeling Award for Applied Cognition, 2014.
- ACM CHI Conference on Human Factors in Computing Systems, Best Paper award, 2014.
- First place, Green Driver Programming Contest, 2011.
- First place, Fifth Annual UO Eugene Luks Programming Contest, 2011.
- ACM CHI Conference on Human Factors in Computing Systems, Honorable Mention award, 2010.
- International Conference on Cognitive Modeling, Siegel-Wolf Award for Best Applied Paper, 2010.

TECHNICAL SKILLS

- Proficient in Python, Java, and R. Familiar with C++, Scala, and Julia.
- Proficient in Tensorflow, Keras, PyTorch, scikit-learn, pandas, numpy.
- Proficient in Apache Beam, TFX, kubeflow, Apache Spark, GCP, BigQuery, Dataflow.