# Yunfeng Zhang Machine Learning Engineer / Al 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 - Nov 2022

- Architected and developed Fairness Evaluator, Twitter's first distributed model
  performance and fairness evaluation system based on TFX and Apache Beam. Fairness
  Evaluator has been used throughout the company to evaluate and improve
  recommendation, ranking, and content moderation models and is a cornerstone of
  responsible machine learning at Twitter.
- Researched and developed new AI fairness and performance metrics, including the metrics published at ACM FAccT 2022 (citation below).

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

May 2016 - June 2021

- Helped develop IBM's AI Fairness 360 and AI Explainability 360 toolkits, the first open source ethical AI toolkits in the industry. Helped integrate them into IBM OpenScale.
   Received IBM outstanding research accomplishment award.
- 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 Al-driven multimodal interaction techniques for smart meeting rooms by incorporating gesture, speech, and face recognition techniques.

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

May 2014 - September 2014

• Researched methods to remediate cognitive biases in 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

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**

Published over 50 papers with more than 2000 citations, 3 patents, and 9 pending patent applications. See my <u>Google scholar</u> page for more information . Selected publications:

- <u>De-biasing "bias" measurement</u>
- <u>AI Fairness 360: An extensible toolkit for detecting, understanding, and mitigating</u> <u>unwanted algorithmic bias</u>
- 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, R, Java, and Go. Familiar with C++, Scala, and Julia.
- Proficient in Tensorflow, Keras, PyTorch, scikit-learn, pandas, numpy.
- Proficient in TFX, Kubeflow, Apache Beam, BigQuery, Apache Spark, GCP, AWS.