WANCONG ZHANG

Github / LinkedIn / Wancongzhang@gmail.com / ☏ (917) 484–0079

Education

MS in Computer Science – GPA: 3.9/4.0 *09/2018-09/2020*
New York University – Courant Institute, New York, NY

BA in Chemistry– GPA: 3.7/4.0 *09/2009-05/2013*
Bard College, Annandale-on-Hudson, NY

Research / Employment

Deep Learning Research Lead, AssemblyAI *02/2021-Present*
- Built and shipped transformer automatic speech recognition (ASR) models for both offline and realtime applications, improving accuracy by 28% over predecessors.
- Implemented custom C++ multithreading beam search kernels with word boost and external language model fusion capabilities, increasing inference speed by 200%.
- Applied self supervised learning to train ASR models for low resource languages.
- Leading a team of 5 to continually improve the ASR model by scaling parameters and data through multi-node training using Google Cloud.

Affiliate Researcher, Yann LeCun Group, NYU *03/2021-Present*
- Designed an efficient probing benchmark to evaluate the fitness of unsupervised visual representations for reinforcement learning. Applied it to systematically improve upon the state-of-art pretraining setup for Atari. Preprint

Graduate Researcher, Sam Bowman & Rajesh Ranganath Groups, NYU *09/2019-06/2020*
- Proposed MixUp algorithms for sentence classification using transformers, regularizing model training and reducing calibration error by 50%. Preprint
- Built customized, fused CUDA kernels for a novel gateless RNN architecture (GATO), increasing its training speed by 4x over its Python version.
- Derived the optimal chain rule sequence for computing the Jacobian of autoencoders, increasing backpropagation speed by >10x. Report Git

Research Intern, Marzyeh Ghassemi Group, University of Toronto *06/2019-09/2019*
- Implemented a multi-modal representation learner for healthcare data that utilizes lab vitals and medical notes to solve clinical tasks. Paper

Software Engineer, Guardant Health *07/2017-07/2018*
- Parallelized background jobs across multiple workers through a queue-based system.
- Created a configurable algorithm that matches patient genomic data to clinical trials.

Research Staff, Harvard & Stanford Universities *10/2013-10/2016*
- Applied genome editing to build cellular glucose sensors for diabetes monitoring.
- Studied the molecular mechanisms of stem cell differentiation.

Publications

Light-weight probing of unsupervised representations for Reinforcement Learning
W Zhang, A Chen, V Sobal, Y LeCun, N Carion
arXiv preprint 2022[Preprint]

A Comprehensive EHR time series pre-training benchmark
M McDermott, B Nestor, W Zhang, P Szolovitz, A Goldenberg, M Ghassemi
The ACM Conference on Health, Inference, and Learning 2021 [Paper]

Mixup training leads to reduced overfitting and improved calibration for the transformer architecture
W Zhang, I Vaidya
arXiv preprint 2021 [preprint]

Skills (years of practice)

C++(4), Python(6), PyTorch(4), Tensorflow(2), NumPy(4), Pandas(4), CUDA(1), Scipy(4)
Distributed Systems(1), Docker(6), Git(6), AWS(2), Slurm(3), SQL(3), VertexAI(1)