Wenda Zhou

Research interests

I am currently a researcher at OpenAI.

Experience

2023- **Member of Technical Staff**, *OpenAI*

2020-2023 **Assistant Professor / Moore-Sloan Faculty Fellow**, *Center for Data Science, New York University*

2020-2023 **Flatiron Research Fellow**, *Flatiron Institute, Simons Foundation*

Preprints and Publications

- *A code superoptimizer through neural Monte-Carlo tree search* (W. Zhou, O. Solodova, R. P. Adams), NeurIPS 2022 MLSys Workshop (Spotlight).
- *Compressive Sensing of Synthetic Aperture Radar* (W. Zhou, S. Jalali, A. Maleki), IEEE Trans. I.T., 2022.
- *Vitruvion: A Generative Model of Parametric CAD sketches* (A. Seff, W. Zhou, N. Richardson, R. P. Adams), ICLR 2022.
- *Autobahn: Automorphism-based Graph Neural Nets* (E. H. Thiede, W. Zhou, R. Kondor), NeurIPS 2021.
- *Asymptotics of Cross Validation* (M. Austern, W. Zhou), in revision.
- *Error Bounds in Estimating the Out-of-sample Prediction Error Using Leave-one-out Cross Validation in High Dimensions* (K. Rad, W. Zhou, A. Maleki), AISTATS 2020.
- *Discrete Object Generation with Reversible Inductive Construction* (A. Seff, W. Zhou, F. Damani, A. Doyle, R. Adams), NeurIPS 2019.
- *Denoising Structured Random Processes* (W. Zhou, S. Jalali), arXiv pre-print 2019. A short version was presented at ISIT 2019.
- *Non-Vacuous Generalization Bounds at the ImageNet Scale: a PAC-Bayesian Approach* (W. Zhou, V. Veitch, M. Austern, R. P. Adams, P. Orbanz), ICLR 2019.
- *Empirical Risk Minimization and Stochastic Gradient Descent for Relational Data* (V. Veitch, M. Austern, W. Zhou, P. Orbanz, D. Blei), AISTATS 2019.
- *Approximate Leave-One-Out for Fast Parameter Tuning in High Dimensions* (S. Wang*, W. Zhou*, H. Lu, A. Maleki, V. Mirrokni), ICML 2018.
  A version of this paper received the runner up for best paper award at the DMDA Workshop at INFORMS 2019.
- *Analysis of Genotype by Methylation Interactions through Sparsity-Inducing Regularized Regression* (W. Zhou, S. Lo), BMC Proceedings 2018 (GAW 20).

Education

2015-2020 **Ph.D. student**, *Columbia University*, Department of Statistics
GPA: 4.0

2014-2015 **MMaths (Part III)**, *University of Cambridge*

2011–2014 **B.A. Mathematics**, *University of Cambridge*

2009–2011 **Baccalauréat Scientifique**, *Lycée International*, St Germain en Laye, Mention *Très Bien*

Academic Service

I have served as a reviewer for ICML, NeurIPS, JMLR, ICLR, AAAI, Annals of Statistics, JASA, IEEE Trans. I.T.. I was the Ph.D. student representative to the faculty for my department in 2016.

Prizes and Awards

2016 **Howard Levene Outstanding Teaching Award**, *Deparment of Statistics, Columbia University*

2011 **1st prize**, *Concours Général de Physique-Chimie*

2011 **3rd prize**, *Concours Général de Mathématiques*

2010 **2nd prize**, *Olympiades Académiques de Mathématiques (French Mathematical Olympiads)*

Work Experience

2019 **Intern**, *Google AI*, Zürich
I worked in the compression team to develop the next generation of audio compression codec based on human hearing studies.

2018 **Intern**, *Nokia Bell Labs*, Murray Hill, New Jersey
I worked with Shirin Jalali on information theroetic aspects of signal processing. We designed information-theoretic approaches to signal denoising with provable guarantees under signal structure assumption.

2017 **Summer Instructor**, *Columbia University*, New York
I taught the calculus-based introduciton to statistics course at Columbia in the 2017 summer session. The material is available at `https://github.com/wendazhou/STATS1201`.

2014 **Intern**, *AHL*, London
I had the opportunity to work at AHL for four weeks during the winter and ten weeks during the summer. I implemented a dashboard aggregating and visualizing the risk exposure of various strategies traded by the fixed income team. I worked on collecting and analyzing several data sets, and exploring the use of machine learning techniques to develop trading strategies.

2013 **Software developer**, *Futurmaster*, Paris
I investigated the behaviour of sales for suppliers in the case of promotional events using statistical techniques from linear mixed effect models.

2012 **Junior Program Manager**, *Microsoft*, Paris
PM on the "Music Intelligence" team, responsible for algorithms concerning classification, extrapolation and verification of music metadata.

Volunteer experience

2016-2017 **Ph.D. student representative**, *Columbia University*
I was responsible for coordinating student activities and communication between the Ph.D. students and the faculty.

2014 **Mathematician in residence**, *Comberton village college*, Cambridge
As part of the Cambridge "Mathematicians in residence" program, I spent two weeks teaching mathematics every day in a local sixth form. The responsibilities varied from simply assisting in the classroom to leading the classes.

2012 **Volunteer**, *STIMULUS*, Cambridge
Stimulus is a "community service programme which gives Cambridge University students the opportunity to work with pupils in local schools". I have helped in a Year 11 class, teaching mathematics.

2011-2014 **Volunteer**, *St John's college access team*, Cambridge
I volunteer on a regular basis with my college's access team, helping with tours given to visiting school and answering students' questions during open days.

References

Arian Maleki, Columbia University — *Ph.D. Advisor*

Peter Orbanz, Gatsby Unit at UCL — *Ph.D. co-advisor*

Ryan P. Adams, Princeton University — *Senior Collaborator*

Languages and Skills

I am a bilingual French and English speaker. I also have some knowledge of Chinese, Spanish and German.
In addition to my core skills in mathematics, statistics and machine learning, I have taken graduate level classes in the following topics: convex optimization, statistical physics, formal verification.

Programming and Software

I am proficient in **C++** and **Python**. I have particular experience in scientific computing and deep learning (Tensorflow and Pytorch), including writing custom C++ code (and CUDA) for deep learning frameworks. I also have significant experience in low-level code optimization and SIMD vectorization. I am familiar with `R`, `Matlab` and `C#`.

- Neural network training and compression (Python, Tensorflow).
  `https://github.com/wendazhou/nnet-compression-generalization`
- Approximate leave-one-out (R, C++, BLAS / LAPACK).
  `https://github.com/wendazhou/alocv-package`
- Graph Samplers (Python, Tensorflow, C++).
  `https://github.com/wooden-spoon/relational-ERM`.
- Discrete object generation (Python, Pytorch, C++, CUDA).
  `https://github.com/PrincetonLIPS/reversible-inductive-construction`