ZHENGHAO CHEN

Machine Learning Researcher

zhenghao@z-chen.net | www.z-chen.net | SF Bay Area

ABOUT ME

I am a machine learning researcher who is passionate about applying AI / ML to develop novel therapeutics that extend human healthspan.

As director of ML at Calico, I helped to grow our team to over a dozen scientists, leading projects that accelerate the drug development pipeline spanning basic research, target identification / validation through to lead gen / opt.

Our team focuses on developing state-of-the-art representation learning and generative models for applications including medical imaging, human genetics, spatial transcriptomics & proteomics, and molecular generation.

SKILLS

Languages

Native proficiency in English and Mandarin Chinese (speaking, reading, writing)

Technical Skills

Proficient in Python, Javascript + some experience working in Scala.

Skilled in ML research - developing generative models for structured output and representation learning for biological data / multiomics.

Expertise in developing data-infrastructure to support ML research and data science / analytics, and building data products.

Proficient with modern ML frameworks (currently primarily Jax-based) and ML for Science tools (e.g., LLM agents for genomics analysis / experiment design)

EDUCATION

M.S. Computer Science

(AI and Biocomputation) GPA: 4.13
Stanford University 2013

B.S. Computer Science

(with distinction, honors & minor in Math)
GPA: 4.04
Stanford University 2013

Awards

Frederick E Terman Engineering Award
(top 5% of Stanford Engineering seniors)

President's Award for Academic Excellence in the Freshman Year (top 3% in class)

Relevant Coursework

Machine Learning, Statistical Learning Theory, Probabilistic Graphical Models, Theory of Statistics, Convex Optimization, Information Theory

EXPERIENCE

Calico Life Sciences

Director of ML 2023 - Now
Associate Director of ML - 2021 - 2023
Principal ML Engineer - 2017 - 2021

Helped grow and manage team of over a dozen ML scientists and engineers focused on building generative and representation learning models for complex biological data to accelerate basic research and therapeutic development.

Developed novel AI / ML approaches in the following areas:

- Predicting how cells will respond to genetic or chemical perturbations using high-content imaging / cellular morphology using self-supervised representation learning
- Understanding how the immune microenvironment is perturbed with aging and in cancer using spatial transcriptomics / proteomics and single-cell sequencing data
- Enabling rapid development of therapeutic molecules and optimization for drug-like properties in various therapeutic modalities using generative models
- Quantifying dimensions of age-related decline using multi-omics / high-dimensional time series data from longitudinal studies (e.g., medical imaging, metabolic markers, blood biomarkers, and wearables)

Coursera

Senior Data Scientist - 2012 - 2017

Founding member of the analytics team and helped grow team to 20 data scientists.

Built key data infrastructure such as the in-house A/B testing platform which supported hundreds of experiments annually and the content recommendation system powering all aspects of content discovery on Coursera.

Led the development of data products such as the onboarding content discovery experience, weekly recommendation email, recommendations for course sequences.

Initiated and led research efforts to use machine learning to improve learning experience. Developed ML models for debiasing peer grading, automated feedback for programming assignments, recommendation models for courses / course sequences etc.

Designed, carried out and published large scale studies on the effectiveness of online instruction and instructional design as well as the long-term benefits of MOOCs on learners.

Counsyl (now Myriad Women's Health)

Intern - June - Sept 2012

Improved error correction and accuracy of in-house Fragile X screening model.

Stanford AI Lab

Research Assistant - 2009 - 2012

Applied probabilistic modelling methods to computational cancer biology and biomedical imaging with Prof. Daphne Koller.

Developed a method for automatic feature extraction of image features from whole slide images of brain tumors. Identified new subtypes of glioblastoma multiforme with different survival and treatment characteristics.

Developed models for unsupervised feature learning using deep neural networks for image recognition and speech recognition with Prof. Andrew Ng.

PUBLICATIONS AND TALKS

Chronological order

Gut metagenomes reveal interactions between dietary restriction, ageing and the microbiome in genetically diverse mice. L Litichevskiy, M Considine, J Gill, V Shandar, T Cox, H Descamps, K Wright, K Amses, L Dohnalová, M Liou, M Tetlak, M Galindo-Fiallos, A Wong, P Lundgren, J Kim, G Uhr, R Rahman, S Mason, C Merenstein, F Bushman, A Raj, F Harding, Z Chen, G Prateek, M Mullis, A Deighan, L Robinson, C Tanes, K Bittinger, M Chakraborty, A Bhatt, H Li, I Barnett, E Davenport, K Broman, M Levy, R Cohen, D Botstein, A Freund, A Di Francesco, G Churchill, M Li, C Thaiss. **Nature Microbiology 2025.**

BiDAC-dependent degradation of plasma membrane proteins by the endolysosomal system. S Villa, Q Jafri, J Lazzari-Dean, M Sangha, N Olsson, A Lefebvre, M Fitzgerald, K Jackson, Z Chen, B Feng, A Nile, D Stokoe, K Bersuker. **Nature Communications 2025.**

Baikal: Unpaired Denoising of Fluorescence Microscopy Images using Diffusion Models. S Chaudhary, S Sankarapandian, M Sooknah, J Pai, C McCue, Z Chen, J Xu. **MICCAI 2024.**

Dietary restriction impacts health and lifespan of genetically diverse mice. A Di Francesco, A Deighan, L Litichevskiy, Z Chen, A Luciano, L Robinson, G Garland, H Donato, M Vincent, W Schott, K Wright, A Raj, G Prateek, M Mullis, W Hill, M Zeidel, L Peters, F Harding, D Botstein, R Korstanje, C Thaiss, A Freund, G Churchill. **Nature 2024.**

Analysis of somatic mutations in whole blood from 200,618 individuals identifies pervasive positive selection and novel drivers of clonal hematopoiesis. N Bernstein, M Chapman, K Nyamondo, Z Chen, N Williams, E Mitchell, P Campbell, R Cohen, J Nangalia. **Nature Genetics 2024.**

Senescence rewires microenvironment sensing to facilitate antitumor immunity. H Chen, Y Ho, R Mezzadra, J Adrover, R Smolkin, C Zhu, K Woess, N Bernstein, G Schmitt, L Fong, W Luan, A Wuest, S Tian, X Li, C Broderick, R Hendrickson, M Egeblad, Z Chen, D Alonso-Curbelo, S Lowe. **Cancer Discovery 2023.**

The immunoregulatory landscape of human tuberculosis granulomas. E McCaffrey, M Donato, L Keren, Z Chen, A Delmastro, M Fitzpatrick, S Gupta, N Greenwald, A Baranski, W Graf, R Kumar, M Bosse, C Camacho Fullaway, P Ramdial, E Forgó, V Jojic, D Van Valen, S Mehra, S Khader, S Bendall, M van de Rijn, D Kalman, D Kaushal, R Hunter, N Banaei, A Steyn, P Khatri, M Angelo. **Nature Immunology 2022.**

Automated, high-dimensional evaluation of physiological aging and resilience in outbred mice. Z Chen, A Raj, P Gundannavar Vijay, A Di Francesco, J Liu, B Keyes, G Kolumam, V Jojic, A Freund. **eLife 2022.**

Find your microenvironments faster with Neural Spatial LDA. S Sankarapandian, J Xu, Z Chen. **NeurIPS Learning Meaningful Representations of Life Workshop 2022.**

Modeling multiplexed images with Spatial-LDA reveals novel tissue microenvironment, Z Chen, I Soifer, H Hilton, L Keren, V Jojic. **Journal of Comp. Bio. 2020.**

Learning good representations of cell state from cell painting, Z Chen, C Jan, F Li, J Xu. **NeurIPS Learning Meaningful Representations of Life Workshop 2019.**

Composition and decomposition of GANs, YC Harn, Z Chen, V Jojic. **arXiv Jan 2019.**

Real-time programming exercise feedback in MOOCs, Z Chen, A Nguyen, A Schlender, J Ngiam. **EDM 2017.**

Mapping the Pairwise Choices Leading from Pluripotency to Human Bone, Heart, and Other Mesoderm Cell Types, K Loh, A Chen, PW Koh, T Deng, R Sinha, J Tsai, A Barkal, K Shen, R Jain, R Morganti, SC Ng, N Fernhoff, B George, G Wernig, R Salomon, Z Chen, H Vogel, J Epstein, A Kundaje, W Talbot, P Beachy, LT Ang, I Weissman. **Cell 2016.**

Who's Benefiting from MOOCs and Why?, Z Chen, B Alcorn, G Christensen, N Eriksson, D Koller, EJ Emanuel. **Harvard Business Review 2015.**

What Matters, What Doesn't? A Coursera-Wide Look at Course Metrics, N Eriksson, A Parisi-Amon, Z Chen. **Talk at 2015 Coursera Partners Conference, Irvine, CA.**

Dissecting an Online Intervention for Cancer Survivors, Z Chen, PW Koh, P Ritter, K Lorig, E Bantum, S Saria. **Health Educ. Behavior. 2014.**

Retention and Intention in Massive Open Online Courses, D Koller, A Ng, CB Do, Z Chen. **Educause Review June 2013.**

Self-Driven Mastery in Massive Open Online Courses, CB Do, Z Chen, R Brandman, D Koller. **MOOCs Forum September 2013.**

Automated Population-scale Screening for Fragile X: Validation and Experience on 76,421 Samples, SA Patterson, MR Theilmann, Z Chen, IS Haque. **Poster at AMP 2013.**

Tuned Models of Peer Assessments in MOOCs, C Piech, J Huang, Z Chen, CB Do, A Ng, D Koller. **EDM 2013.**

A Graph Regularization Based Approach for Gene Scoring, Z Chen. **Stanford CS Honors Thesis 2012.**

Sparse Filtering, J Ngiam, PW Koh, Z Chen, S Bhaskar, A Ng. **NIPS 2011.** Spotlight paper.

Learning Deep Energy Models, J Ngiam, Z Chen, PW Koh, A Ng. **ICML 2011.**

On Random Weights and Unsupervised Feature Learning, A Saxe, PW Koh, Z Chen, M Bhand, B Suresh, A Ng. **ICML 2011.**

Tiled Convolutional Neural Networks, QV Le, J Ngiam, Z Chen, D Chia, PW Koh, A Ng. **NIPS 2010.**

Lower Bound on the Time Complexity of Local Adiabatic Evolution, Z Chen, PW Koh, Z Yan. **Phys. Rev. A 74, 2006.**