Yuan Pu

yuan.pu@yale.edu | linkedin.com/in/yuan-pu | PLACE HOLDER

Education

Brown University Sept 2019 – May 2023

Bachelor of Science in Computational Biology (Computer Science track) - Honors *Providence, RI*

- GPA: 4.0/4.0; Magna Cum Laude; Computational Biology Departmental Honors; Sigma Xi Honors Society
- **Computer Science** coursework in deep learning, machine learning, computational linguistics, computer vision, computational molecular biology, computational population genetics, UIUX
- **Life Science** coursework in cell and molecular biology*, genetics, functional genomics, DNA-based techniques*, neurobiology, neural systems, general and organic chemistry* (*with experimental laboratory experience)
- **Math** coursework in calculus, linear algebra, computational probability and statistics, causal inference

Yale University Sept 2023 – Present

Postgraduate Assistant *New Haven, CT*

- Audited courses in fundamentals, recent developments, and applications of Large Language Models (LLMs)

Publications (*equal contribution)

Trajectory Flow Matching with Applications to Clinical Time Series Modelling | Xi Zhang*, Yuan Pu*, Yuki Kawamura, Andrew Loza, Yoshua Bengio, Dennis Shung, Alexander Tong

Under-review

Human-Algorithmic Interaction Using a Large Language Model-Augmented Artificial Intelligence Clinical Decision Support System | Niroop Channa Rajashekar*, Yeo Eun Shin*, Yuan Pu*, Sunny Chung, Kisung You, Mauro Giuffrè, Colleen Chan, Theo Saarinen, Allen Hsiao, Jasjeet Sekhon, Ambrose Wong, Leigh Evans, René Kizilcec, Loren Laine, Terika McCall, and Dennis Shung

Computer Human Interaction (CHI) 2024

Assessing the Usability of GutGPT: A Simulation Study of an AI Clinical Decision Support System for Gastrointestinal Bleeding Risk | Colleen Chan, Kisung You, Sunny Chung, Mauro Giuffrè, Theo Saarinen, Niroop Channa Rajashekar, Yuan Pu, Yeo Eun Shin, Loren Laine, Ambrose Wong, René Kizilcec, Jasjeet Sekhon, Dennis Shung

Machine Learning for Health (ML4H) 2023

Machine Learning on Multiple Epigenetic Features Reveals H3K27Ac as a Driver of Gene Expression Prediction Across Patients with Glioblastoma | Yusuke Suita, Hardy Bright, Jr., Yuan Pu, Merih Deniz Toruner, Jordan Idehen, Nikos Tapinos, Ritambhara Singh

Under-review

Experience

Postgraduate Assistant August 2023 – Present

Yale School of Medicine H+AIM Lab *New Haven, CT*

Clinical Time Series Modeling | *advised by Dennis Shung and Alex Tong*

- Developed a novel technique for modeling irregularly sampled clinical time series data by extending the Conditional Flow Matching method.
- A submission to NeurIPS 2024 (co-first author), a poster at 2024 AI in Medicine Symposium at Yale University, and an oral presentation at Digestive Disease Research in Progress seminar at Yale Univeristy

AI/ML-Enhanced Clinical Decision Support System Evaluation | *advised by Dennis Shung*

- Evaluated human-computer interaction of an LLM-augmented ML clinical decision support system through medical simulation, focusing on usability and user trust.
- Papers accepted at CHI 2024 (co-first author) and ML4H 2023, an oral presentation at Digestive Disease Week (DDW) 2024, an oral presentation and a poster at 2024 AI in Medicine Symposium at Yale University

Patient Data Analysis | *advised by Dennis Shung and Darrick Li*

- Conducted data extraction and analysis on patient records. Identified factors linked to post percutaneous coronary intervention gastrointestinal bleeding, rebleeding, and major cardiovascular events.

- An abstract submission to American College of Gastroenterology (ACG) 2024 Annual Scientific Meeting (first author)

Research Intern Nov 2023 – Present
Dymaxion *Remote*

AI-powered furniture arrangement
- Explored LLMs' potential in powering AI generation of indoor furniture arrangement in 2D and 3D. Implemented enhancements to an academic research method to better address real-world scenarios for commercial application.

Undergraduate Research Assistant May 2022 – May 2023
Brown University Singh Lab *Providence, RI*

Epigenetic Regulation of Gene Expression in GSCs | *advised by Ritambhara Singh*
- Implemented attention-augmented RNNs to predict gene expression from epigenetic data in glioblastoma stem cells (GSCs). Investigated epigenetic regulation of gene transcription in different GSCs by cross-patient analysis.
- Brown University Computational Biology Honors Thesis and a submission to PLOS Computational Biology

Data Science Intern Jan 2022 – May 2023
Brown University Computational Biology Core *Providence, RI*

EGME's Impact on Sperm Small RNA Expression | *advised by Daniel Spade and August Guang*
- Processed and conducted analysis on small RNA data in sperm of rats exposed to Ethylene Glycol Monomethyl Ether (EGME). Identified sensitive biomarkers and related biological pathways for EGME's testicular toxicity.

Undergraduate Research Assistant Sept 2020 – Jan 2022
Brown University Yajima Lab *Providence, RI*

Germline Factor DDX4's Role in Cancer | *advised by Mamiko Yajima*
- Discovered that DDX4 expression level influences the survival of acute myeloid leukemia (AML) patients through data mining. Identified genes co-expressed with and biological pathways influenced by DDX4 expression.

Teaching

Teaching Assistant at Brown University

CSCI1430 Computer Vision | *instructed by James Tompkin* Spring 2022, Spring 2023

CSCI1810/2810* **Computational Molecular Biology** (*graduate level) | *instructed by Sorin Istrail* Fall 2022
- Identified issues in written and coding assignments by reviewing the materials and monitoring student feedback.
- Revised tasks, instructions, and solutions to enhance student learning.
- Graded assignments and projects.
- Provided student support for assignments and projects through weekly office hours and online discussions.

Awards and Honors

Brown University Magna Cum Laude 2023
Brown University Computational Departmental Honors 2023
Sigma Xi Scientific Honors Society 2023
Hack@Brown 2022 Wolfram Award 2022
Brown SPRINT LINK Program $2000 grant for summer undergraduate research with faculty 2021

Skills

Languages: Python, R, MATLAB, Java, JavaScript, HTML/CSS

Frameworks: PyTorch, PyTorch Lightning, TensorFlow, Pandas, Scikit-Learn, React

Tools: Git, Google Cloud Platform, high performance computing clusters

Selected Class Projects (*group effort)

VAEs, Diffusion, and Score-based Models from Scratch*
- implemented and evaluated Variational Autoencoders (VAEs), LadderVAE, and Diffusion models, along with Score Matching, for 2D density estimation and data generation.

Text-guided 3D Object Editing*
- Developed and evaluated a 3D mesh stylization method utilizing both local and global text-image similarity

Computer Vision Methods and Algorithms
- Implemented: (1) Harris corner detector plus Scale Invariant Feature Transform for feature matching between 2D images; (2) random sample consensus for 3D reconstruction from stereo 2D image pairs; (3) Bag of Words plus linear support vector machine, CNNs, and self-attention networks* for 2D image classification; (4) CNN-based style transfer model for 2D images*

Computational Linguistics Methods and Algorithms
- Explored (1) Bag of Words and BERT with fine-tuning for sentiment classification; (2) LDA with tf-idf algorithm for topic modeling; (3) RNNs and transformers for machine translation; (4) CLIP zero-shot and linear-probing classification combined with GPT3 in-context learning for image captioning; (5) shift reduce algorithm for dependency parsing; (6) seq2seq with attention for semantic parsing

Computational Molecular Biology Methods and Algorithms
- Implemented (1) sequence alignment algorithms; (2) pattern matching algorithms; (3) genome assembly algorithms; (4) population genetics summary statistics calculation