

# JADELYNN DAO

Curriculum Vitae (Last Updated November 2024)

[jdao@caltech.edu](mailto:jdao@caltech.edu) | +1 (408) 310-7529

## RESEARCH INTERESTS

Multimodal machine learning and representation learning. I am particularly interested in autonomous learning, Bayesian inference, and state estimation techniques that integrate diverse data for holistic health monitoring. Exploring the intersection of time-series signal processing, wearable devices, and multimodal generative AI to create personalized, real-time feedback systems.

## EDUCATION

California Institute of Technology  
*Bachelor of Science, Computer Science*

Expected: June 2025  
*GPA: 4.1*

## RESEARCH EXPERIENCE

**California Institute of Technology**

**September 2022- Present**

*Undergraduate Researcher, Gao Research Group*

Modeling acute stress and emotional states in real-time for a wearable biosensor system (EEG, EOG, EDA, skin temperature).

Journal First-Authorship, Spring 2025 [\[ArXiv Preprint Here\]](#) | Journal Co-Authorship, Spring 2025

- Conducted human trial experiments and developed a GUI for synchronized multi-sensor data collection.
- Classified natural eye blinks in EOG sensor data from aggressors like head movements, electrical noise, and signal drift, reaching an on-device accuracy score of 98.1% and F1-score of 87.3% in patient trials. Conducted post-inference feature importance and noise analysis on EOG and EDA signals for state anxiety biomarker discovery.
- Contributed to a meta-learning pipeline that pulls open-source datasets to improve the detection of emotion and mood from real-time physiological signals. Time-series sensor data embedding using a CNN autoencoder.
- Built a feedback therapy system that generates images, text, and audio based on the user's emotion profile and adapts in real time using OpenAI API.

**California Institute of Technology**

**September 2023- Present**

*Undergraduate Researcher, Valen Valen Lab*

Predicting antigen (CDR region) and antibody interactions using 3D representation learning techniques.

Conference Publication Co-Authorship, January 2025

- Created a foundation model for zero-shot antibody generation using sequence and structural information of proteins.
- Designed an equivariant GNN encoder and contrastive learning (CLIP) to predict binding of 3D molecular objects
  - 50-300% binding prediction improvement from state-of-the-art (SOTA) method via Palepu et. al 2022
    - Top-1 accuracy from 3.7% to 14.8%, and top-10 accuracy from 18.9% to 53.9%
  - Data from Protein Data Bank (PDB) and The Structural Antibody Database (SAbDab)
- Applied [Evolutionary Scale Modeling \(ESM-2\)](#) to improve tokenized representations of protein sequences

**NASA Jet Propulsion Laboratory**

**Summer 2022**

*Research Internship, Mars Sciences*

Mapping geomorphic surface features on Mars' Southern Hemisphere.

Poster Presented at American Geological Union Fall Mtg, 2022 [\[Poster Here\]](#) | Oral Presentation for JPL Science Division, 2022

- Mapped Martian araneiforms and their correlation to cryptic terrain on Mars' Southern Hemisphere, organizing observations into a Java database. Image data from the Context Camera (CTX) and High Resolution Imaging Science Experiment (HiRISE) on the Mars Reconnaissance Orbiter.
- Identified strong correlations and developed data-driven hypotheses on past and present Martian seasonal dynamics.

Independent Research

Touchless, gesture-based computer interface device

Human Computer Interaction

Patent No. 12,105,884 Issued [\[Patent Here\]](#) , October 2024

- Designed and prototyped a unique gesture-based, touchless communication between user and electronic device.
- System identifies hand or finger gestures using depth-mapping-based (time-of-flight) sensor technology that is complemented by an inertial measurement unit (IMU) or similar gyroscopic sensors.

September 2019- August 2021

TEACHING EXPERIENCE

CMS 9: Introduction to Computer Science Research

Teaching Assistant, 54 students

Organized professor guest lecturers and oversaw course logistics.

Fall 2024

CS/CNS/EE 156b: Learning Systems (Projects)

Teaching Assistant, 108 students

Gave introductory lecture, graded student presentations, organized project groups

Spring 2024

CS/CNS/EE 156a: Learning Systems (Theory)

Teaching Assistant, 187 students

Hosted office hours, graded problem sets

Fall 2023

INDUSTRY EXPERIENCE

Google

Devices, Sensors Algorithms | Software Engineering (ML) Intern

Designing on-device machine learning and heuristics algorithms for motion-based sensing for wearables.

Whitepaper, 2024 | Oral Presentation for Google Sensor Algorithms, 2024

- Designed a CNN model to predict uncertainty of temporal magnetometer measurements in real-time, combining approaches from control systems, deep learning, signal processing, and bayesian inference.
- Incorporated model into an extended Kalman filter for 6 DOF device orientation tracking.
- Built model into a demo app, improving performance of orientation in real-time validation tests.

Summer 2024

Apple

Health Research, Multimodal + Longitudinal ML | Machine Learning Intern

Investigated correlations in nightly aggregated user data and medical conditions.

Oral Presentation for Apple Health Sensing, 2023

- Identified biomarker correlations and variations across age, gender, race, and medication used to inform and support future Apple Health features (women’s health and mental health).
- Applied interpretable models to quantify correlations and classify aggressors in multimodal, longitudinal sensor data.
- Created new sensing data pipelines for HID feature incubation projects using Apache Spark and Hadoop.

Summer 2023

Pivotal Systems

Manufacturing Systems | Engineering Intern

Optimized product calibration and testing procedures, reducing QA cycles by 18%.

Summer 2021

## RESEARCH ARTIFACTS

### “State Anxiety Biomarker Discovery: Electrooculography and Electrodermal Activity in Real-Time Stress Monitoring”

Target Journal Publication: *IEEE Journal of Biomedical and Health Informatics*, 2025 | *ArXiv Human-Computer Interaction*, 2024 [\[Preprint Here\]](#)

Jadelynn K. Dao, Ruixiao Liu, Sarah Solomon, Samuel Solomon\*, Wei Gao\*

### Manuscript on Wearable Electronic for Real-Time Emotion Monitoring and Therapy Feedback

Target Journal Publication: *Nature Mental Health*, 2025

Samuel Solomon, Yadong Xu, Ruixiao Liu, **Jadelynn K. Dao**, José Lasalde Ramírez, Jihong Min, Dickson Yao, Sarah Solomon, Canran Wang, Sijie Ji, Wei Gao\*

### “CDR Sequence-Structure Codesign with Classifier-Free Diffusion”

Target Conference Publication: *International Conference on Machine Learning (ICML)*, 2025

Rohit Dillip, **Jadelynn K. Dao**, Ayush Varshney, Evan J. Zhang, Katherine L. Bouman\*, David Van Valen\*

### “A Comprehensive Map of Martian South Polar Region Araneiform Locations and Morphological Analysis”

Conference Poster Presentation: *American Geological Society Fall Meeting*, 2022 [\[Poster Here\]](#)

Lauren Mc Keown, Serina Diniega, **Jadelynn K. Dao**, Candice J. Hansen, Ganna Portyankina, Sylvain Piqueux, Klaus-Michael Aye\*

### “Machine Learning to Quantify Device Orientation Uncertainty in Real-Time”

Company White Paper: *Google*, 2024

### “Touchless, Gesture-Based Human Interface Device”

Patent No. 12,105,884: *US Patent & Trademark Office*, 2024 [\[Patent Here\]](#)

**Jadelynn K. Dao**, Advit Deepak

---

Oral Presentation for JPL Science Division, 2022

Oral Presentation for Apple Health Sensing, 2023

Oral Presentation for Google Sensor Algorithms, 2024

## PROJECTS

### CS 159: Advanced Topics in ML, Large Language Models for Reasoning

Spring 2024

*Embodied Lifelong Agents* [\[Report Here\]](#)

WAYFARER, an LLM-powered embodied lifelong learning agent in Minecraft that excels at zero-shot generalization, automatic curriculum generation, and prompt efficiency. WAYFARER leverages an automatic curriculum driven by GPT-4o to maximize exploration and task completion using a flexible controller framework, MINEWIRE. WAYFARER employs high-level task planning and tree-like decision-making, enabling it to achieve complex crafting goals with minimal computational expense.

### CS 101: Special Topics in Computer Science

Spring 2024

*Contrastive Language-Image Pretraining + Drug Discovery* [\[Repository Here\]](#)

Understanding and identifying protein-protein interactions is essential for developing targeted treatments and eradicating diseases at their root. Our study investigates the application of CLIP (Contrastive Language–Image Pretraining) and FILIP to predict protein-peptide and protein-protein interactions. We have extended the utility of these models beyond their conventional domains to address the complex challenge in bioinformatics of protein binding.

## CS/CNS/EE 156b: Learning Systems (Projects)

Spring 2024

### *Image Transformers + Medical Imaging*

Participated in a term-long competition focused on minimizing MSE across 9 binary-classified pathologies. Our model, a multi-stage vision transformer using ensemble learning and pre training techniques, placed 2nd on the class leaderboard.

## CS 3: Introduction to Software Design

Spring 2022

### *Physics Modeling and Game Design*

Designed and programmed a multiplayer, multi-level physics simulation game in C++ where characters collaborate and compete to navigate obstacles and collect gems. The game incorporates simulated gravity, acceleration, elasticity (bounce), and friction for realistic interactions.

## Synopsys Science

Spring 2019

### *Using Reinforcement Learning (RL) to Modernize Traffic Control Algorithms* [\[Poster Here\]](#)

Developed a Deep Q-Network Reinforcement Learning model to control traffic lights intelligently. After interviewing the director of San Jose's Department of Transportation, we received access to real time data from several traffic intersections and used urban mobility simulations to train a model that demonstrated a 26% reduction in average commute time. Using an embedded processor and LED strips to represent roads, we built a prototype to visualize our findings. 1st prize Synopsys Science.

## ACADEMIC HONORS + SOCIETIES

### Interests & Leadership

Undergrad Research Journal Editing Team  
US Dept of State Science Exchange Advisory Board  
Society of Women Engineers, Science Colloquium Speaker  
NCAA DIII Athlete - Swim & Dive  
Red-Cross Certified First Responder/Campus Leader

### Fellowships

Kleiner Perkins Engineering Fellow 2024  
NCAA Swim & Dive: Conference All-Academic 2024  
Caltech AI Alignment Group Intro to ML Safety Fellow 2024  
Harvard Public Health Mathematical Modeling and Public Health Workshop Scholar 2024  
Neo Scholar Finalist 2023, 2024  
Meta Above & Beyond CS Fellow 2022  
U.S. Department of State Science Exchange Programs 2019, 2020

### Awards & Scholarships

National Center for Women in Information Technology Aspirations in Computing National Winner 2021  
Hispanic Scholarship Fund National Scholar 2020, 2021  
FIRST Robotics Excellence in Robotics Award 2021  
U.S. Department of State: Certificate of Recognition 2020  
Cisco MVP of Externship Program 2019  
Straubel Engineer of the Future 2018

## Personal Statement for Computer Science PhD at Columbia University

*Jadelynn Dao, Fall 2025 Matriculation*

I find the greatest fulfillment in using science and technology to address complex challenges in healthcare and medicine. I was born with a cleft palate, a medical condition that required me to spend my early years in a children's hospital, undergoing craniofacial surgeries. Through seven years of extensive speech therapy and countless appointments, my journey to recovery was heavily influenced by the dedication of my doctors, nurses, and specialists. These experiences taught me that humanity and health are universal endeavors that transform lives. They inspired my excitement for scientific innovations that provide life-changing solutions in critical moments of vulnerability. To me, science is as human as it is technical—a field that thrives on compassion, curiosity, and the drive to improve the quality of life for others.

### **Research Motive: Multimodal Machine Learning for Context-Aware Healthcare**

I envision a system of non-invasive personal devices that enables people to live healthier, happier lives by integrating biosignals, genetic profiles, health histories, and culturally informed practices for tailored treatments that contextualize individual health profiles. Embedding Artificial Intelligence (AI) and Machine Learning (ML) in real-time health monitoring could improve healthcare efficiency and outcomes, promoting health equity across diverse populations.

Over the past six years, I have built numerous sensor-packed, embedded devices and designed the algorithms that power them. I hold a patent for touchless computing and I have designed deep learning models for research projects in academia and industry-scale devices currently in production. My ML research has increasingly focused on AI-health, where I have recognized that healthcare's complexities require more than existing statistical and ML methods. Current models often lack a basis in human physiology and struggle to handle diverse, noisy, and unseen health data—limiting their generalizability and transparency. By pursuing a PhD, I will

## Personal Statement for Computer Science PhD at Columbia University

*Jadelynn Dao, Fall 2025 Matriculation*

create multimodal, intelligent systems that blend ML with core scientific principles to capture and analyze critical context, improving the robustness needed for medical applications.

Columbia's Computer Science doctoral program, with its focus on developing computational approaches to address complex scientific and engineering challenges, offers an ideal environment for my graduate studies. My AI-health research experience at Caltech, along with ML internships at Google and Apple, have equipped me with a strong foundation to contribute to Columbia's interdisciplinary research and innovations in data-driven healthcare.

### **Medical Engineering, Wearable Devices, Representation Learning**

I have always been curious about how unique data sources can be integrated into intelligent systems. In 2022, I met Professor Wei Gao, and his pioneering work in non-invasive wearable bioelectronics inspired my focus on bringing his biosensors to life, using my computational background to develop real-time, data-driven health monitoring agents.

My project group is prototyping a forehead patch for monitoring affectivity and emotional state inference in real-time using electrooculography (EOG), electroencephalogram (EEG), and electrodermal activity (EDA) sensors. I conducted a game theoretic analysis to decompose and analyze signal artifacts in EOG and EDA data and their correlation to state anxiety, discovering new biomarkers that connect physiological responses to affectivity states. I published a first-author manuscript preprint, now publicly available, and we plan to submit it to the IEEE Journal of Biomedical and Health Informatics.

Furthering this work, I incorporated public datasets into a meta-training pipeline and developed convolutional neural network (CNN) autoencoders, proposing generalizable ways to represent high-frequency biosensor signals for emotional profiling. My interest in delivering

## **Personal Statement for Computer Science PhD at Columbia University**

*Jadelynn Dao, Fall 2025 Matriculation*

actionable outcomes led me to develop a feedback therapy system that generates personalized images and audio in real-time to mitigate the psychological effects of emotional arousal.

Taking ownership over project components—and the opportunity to formulate and publish my own research—has solidified my decision to pursue graduate education, where I can propose larger projects that span across science and engineering disciplines.

### **Bioengineering, Drug Discovery, Equivariant Graph Neural Networks**

My growing interest in bridging ML with rigorous mathematical principles influenced my work in drug discovery, where I found conventional ML methods often struggle with the complexities of 3D molecular structures. Since 2023, under the guidance of Professors Katie Bouman and David Van Valen, I have been developing a pipeline integrating Contrastive Language-Image Pretraining (CLIP), Variational Autoencoders (VAEs), and Equivariant Graph Neural Networks (EGNNs) to predict antibody binding to target proteins. By embedding molecular principles like symmetry, chirality, and complementarity into the model, we improved antibody-antigen interaction prediction top-k accuracy by over 25%. In my graduate research, I want to continue to incorporate scientific relationships to guide statistical priors underlying the problem space for representation learning and foundation models.

### **Machine Learning Internships at Apple Health and Google Devices**

Spending my summers in industry research groups alongside teams of accomplished research scientists, all with doctorate degrees, has deeply shaped my decision to pursue a PhD. Working on high-impact projects with a billion-person reach has honed my research aptitude and approach to open-ended, product-driven questions. For my 2023 internship with Apple's Health Research team, I collaborated with ML research scientists and practicing physicians to identify

## Personal Statement for Computer Science PhD at Columbia University

*Jadelynn Dao, Fall 2025 Matriculation*

correlations in multimodal and longitudinal physiological data collected from the Apple Watch and iPhone. I used Bayesian models to identify dependencies for women's menstrual cycles and mental health outcomes across demographic groups. My time at Apple introduced me to model design that follows principles of health decision-making and medical rigor.

Continuing to integrate scientific and mathematical insights in my models, I spent the Summer of 2024 at Google, where I enhanced an Extended Kalman Filter with an on-device deep-learning model for Inertial Measurement Unit sensor fusion. I reviewed textbooks and relevant papers, and collaborated closely with physicists and roboticists to pair these theoretical foundations with empirical testing. Effective real-time models require a thoughtful, problem-driven approach. Tailoring solutions to the challenges of sensor fusion taught me techniques to implement and deploy ML with domain-specific principles.

I am excited to deepen my expertise and advance my quantitative research skills with the goal of one day leading a research group. A PhD would provide the resources and intellectual environment to pursue these ambitions, and I hope to find ways to bridge cutting-edge academic research with the feasibility required for user-level impact.

### **Leadership, Teaching, and Mentorship**

As an NCAA Division III swimmer and Red Cross-certified Emergency Medical Responder, I am passionate about health and empowering others to lead healthier lives. On campus, I serve as an on-call first responder for physical and mental health crises and lead initiatives to improve women's health resource access. This human-centered perspective in medical practice complements my technical work in healthcare.



## Personal Statement for Computer Science PhD at Columbia University

*Jadelynn Dao, Fall 2025 Matriculation*

Teaching and mentorship has deepened my passion for creating collaborative, inclusive learning environments in science. I was a TA for two upper-division computer science and ML courses at Caltech, where I lectured on medical imaging to over 100 students. Currently, I facilitate a freshman CS/Math research seminar, introducing students to Caltech research. I also serve on the advisory board of a U.S. Department of State global STEM exchange program. In 2024, we reached 125 girls from 37 countries, empowering them to lead K-12 STEM education initiatives in their communities, advancing a global network of young scientists and engineers.

### **Why Computer Science and Machine Learning at Columbia**

Columbia SEAS is a perfect environment for me, where I can translate computer science theories into practical applications with social impact. The university's emphasis on developing robust, scalable technologies aligns with my commitment to creating ML systems that enhance healthcare outcomes. Professor Xia Zhou's work on sensor algorithms and wearables, particularly her focus on on-device ML and health sensing, closely aligns with my research interests and experience. Similarly, Professor Yunzhu Li's research on multimodal representation learning and interacting agents directly connects to my work with time-series sensor data. Professor Steven Feiner's work on human-computer interaction through novel data sources and interface design aligns with my interest in conducting innovative, technically rigorous research that is centered around human perception and impact.

Columbia is the ideal setting for me to thrive as a researcher, foster cross-domain collaboration, and amplify the impact of my work. I am committed to leveraging and expanding my expertise in machine learning to bridge disciplines and drive innovative solutions for real-time, context-aware applications that have potential for broader societal impact.