# Sarah Liaw

sliaw@caltech.edu | linkedin.com/in/sarah-liaw | github.com/SarahLiaw

## Education

**California Institute of Technology (Caltech)** Expected May 2025

*B.S., Computer Science; Mathematics. GPA 4.2/4.0*

**Coursework**: Statistics and Probability, Linear Algebra, Discrete Mathematics, Machine Learning/Data Mining, Relational Databases, Data Structures, Algorithms, Software Design, Computing Systems, Functional Programming

## Technical Skills

**Proficient**: Java, Python, C, MATLAB, HTML/CSS, SnakeMake, Git, JUnit, Pytorch, OpenCV, Tensorflow
**Familiar**: x86-64 Assembly, Swift, SQL, R, Dash, Django, Bootstrap, bash

## Experience

**Undergraduate Researcher** March 2023 - Present

*Computing and Mathematical Sciences at Caltech*

- Developing mathematical algorithms to learn the Markov graph structure in non-Gaussian settings for improved information retrieval of dense datasets (with MATLAB and Python (PyTorch)).

**Data Science Intern (Machine Learning)** June 2022 – Dec. 2022

*National Cancer Institute, Aadel Chaudhuri Lab at the Washington University in St. Louis*

- Leveraged Python, SnakeMake, and R to solve a challenging computational problem in Machine Learning involving a dataset with fewer than 200 samples and over 3000 features.
- Improved the accuracy of an existing ML model by 8% using feature engineering and classification algorithms.
- Developed a novel ML classification using logistic regression and regularization, resulting in 95% accuracy and outperforming all existing models; expected to be published as second author in an upcoming publication.

**Undergraduate Researcher (Software Engineering)** Feb. 2022 – June 2022

*Aerospace and Control Lab at Caltech*

- Expedited validation of M-SEPS space algorithm by maintaining interal APIs using Python and computing camera calibration and background removal tools (morphology, image segmentation) in ROS2 on Ubuntu and CUDA.
- Analyzed the trade-offs of different background removal techniques and contributed to the codebase with 200+ lines in implementation of extracted ROS2 bags. Contributed to open source community with ArUco calibration tools.

## Projects

**Citadel Women's Datathon 1st Place** | *Python, Data Science* Feb. 2023 – Feb. 2023

- Successfully led a team of 4 in creating an ML model and composing a comprehensive report on the impact of urban planning policies on mental health in the US, leveraging 30+ extensive datasets in 7.5 hours.
- Developed an impactful ML model employing techniques such as Random Forest Regressor, XGBoost, LightGBM and Bayes Optimization, producing an RMSE of less than 1 and significantly reduced bias.
- Won 1st place in in-person datathon of 84 selected from 1200, advancing to Citadel World Championships.

**Scalable Uncertainty-Aware Models for Imaging Analysis** | *Python, ML, Research* April 2023 – June 2023

- Proposed an deep kernel learning (DKL) algorithm leveraging stochastic variational inference to fine-tune a Gaussian process-based model (implemented in PyTorch) to enhance predictive accuracy for image analysis.
- Improved RMSE on the multiclassification 2017 RSNA pediatric bone age medical dataset by 4% using proposed algorithm which incorporates mini-batch training and localized kernels.

**2D Physics Engine** | *C, JavaScript, Python* March 2022 – June 2022

- Led team of 4 members to engineer a physics engine by implementing a tech stack and data structures library in C.
- Improved loading performance of game by troubleshooting 15 errors through code refactoring and unit testing.
- Utilized available REST APIs to handle large volumes of requests and ensure critical services uptime.
- Implemented queues to validate body collisions, resulting in a 10% improvement in the algorithm's system time.

## Selected Actvities

Head Teaching Assistant for CS 12 (Computational Neuroscience)
Selected Caltech Board of Control Representative; Academic and Research Committee At-Large Representative
Society of Women Engineers Committee Member (SWE)
11th (out of 35) in 2019 Asia Pacific Informatics Olympiad; 1st (out of 400) in Malaysia Computer Science Assessment