

Brandon Lee Concepcion

(650)-892-4233 | brandon_concepcion@berkeley.edu | linkedin.com/in/brandonconcepcion | brandonconcepcion.com
github.com/brandonconcepcion

EDUCATION

University of California, Berkeley

Majors: B.A. Data Science, B.A. Computer Science

Technical Skills/Coursework: Deep Learning, Neural Networks, Computer Vision, Natural Language Processing, AB Testing, Machine Learning Theory, Data Structures, Advanced Calculus, Linear Algebra, Discrete Mathematics, Statistics and Probability
Organizations: College of Computing, Data Science, and Society, Data Science Society at Berkeley, Berkeley Unboxing Data Science, Data C8 (Foundations of Data Science) Undergraduate Course Staff, Data Scholars, Data Science Undergraduate Association
Honors/Awards: Genentech Futurelab Scholar, 6th Annual Datathon For Social Good: 2nd Place

May 2026

GPA: 3.8/4.0

SKILLS AND INTERESTS

Languages: Python, Java, SQL, RegEx, L^AT_EX, HTML | Learning Ruby, Javascript

Tools: Pandas, NumPy, TensorFlow, PyTorch, CV2, SciPy, StatsModels, SkLearn, Seaborn, Plotly, Matplotlib, Tableau, Streamlit

Skills: Exploratory Data Analysis, Git, Data Visualization, Classification, Clustering, Linear and Logistic Regression, Data Analytics

Soft Skills: Efficient Communication, Adaptability & Flexibility, Teamwork and Collaboration, Organizational Ability, Simple and Creative Problem Solving, Leadership, Critical-Thinking, Taking Initiative, Self-Starter, Attention to Detail, Cross-Functional

Academic Interests: Computer Vision, Deep NNs, Teaching, Probability, Mathematics, Transformers, NLP, Pedagogy, Astronomy

Personal Interests: Movies, Volleyball, Photography, Basketball, Gym, Swimming, Web Development, Music, Marvel Studios

EXPERIENCE

UC Berkeley College of Computing, Data Science, and Society

Undergraduate Student Instructor - DATA c8: Foundations of Data Science

Berkeley, CA

Jun 2024 - Present

- Developed and delivered educational content to a student population of over 3,000, providing comprehensive and effective instruction through office hours, review sessions, and management of 6 tutors to answer student questions and address concerns
- Led the restructuring of data8.org/su24 using HTML, CSS, and Javascript for front-end web development; adding 19 dynamic tabs to filter 250+ past exam problems by topic, leading my students to achieve the highest median final exam scores (90th percentile), as well as the highest overall median cumulative course grade among all student instructors (93rd percentile)

Data Science Society at Berkeley

Instructor, Course Director

Berkeley, CA

Aug 2023 - Present

- Managing 8 Teaching Assistants and 6 tutors to operate the “An Introduction to Real World Data Science” course, promoting accessibility by hosting educational workshops from industry and academic leaders for the 70+ diverse students in the course
- Created the course website dssdecal.org by utilizing tool such as Jekyll, Ruby, and Git, and currently developing a 12-chapter introductory data science textbook at dssdecal.org/textbook for the course’s 14 student groups

University of Washington

Data Scientist

Remote

Jan 2024 - Jun 2024

- Utilized natural language processing (NLP) techniques, including Latent Dirichlet Allocation and Bidirectional Encoder Representations from Transformers (BERT), to uncover hidden themes in 107 student health and wellness survey responses
- Applied supervised learning models to classify topics in student feedback, enhancing the analysis of student engagement patterns and the effectiveness of self-care initiatives, providing insights for professors at the School of Pharmacy

PROJECTS

SafeZone 📍 | Streamlit, Python, Pandas, Scikit-Learn, Application Programming Interfaces

Nov 2024

- Utilized the Armed Conflict Location & Event Data (ACLED) API to achieve 93% accuracy in classifying regions likely to experience any fatalities from escalating political events, aiding Doctors Without Borders to identify high-risk zones.
- Preprocessed over 2 million political events across 74 features and 180+ countries to train 2 Multilayer Perceptrons, achieving an R² score of 0.76 in predicting the number of fatalities for regions with escalating political events
- Developed a Streamlit app to visualize conflict severity and fatality predictions, enhancing humanitarian safety management.

To Spam, or Ham? 📧 | Python, Scikit-Learn, Principal Component Analysis

Apr 2024

- Engineered a logistic regression model for classifying emails as Spam or Ham, using a dataset of over 7,500 points and resulting in a classifier that achieved 87% testing accuracy and an Area Under the ROC Curve of 0.91
- Applied Principal Component Analysis (PCA) to reduce dimensionality from 18 to 5 features, mitigating overfitting and ensuring 92% of legitimate (ham) emails in the test set of 1065 were correctly identified
- Enhanced model performance through GridSearch Cross Validation optimization across 4 hyperparameters, resulting in a 5% increase in accuracy when classifying a validation set of 1250 emails

Gen AI Retinal Video Sequences 📺 | PyTorch, CV2, Deep Neural Networks, Long Short Term Memories

Nov 2023

- Coded a Variational Auto-Encoder (VAE) neural network to convert retinal movement numerical data into generative AI video simulations of retinal movement afflicted by one of three different diseases
- Implemented a data preprocessing pipeline that converts .avi files into sets of 300 individual frames
- Ran training data through a Long-Short Term Memory (LSTM) network to encode data into latent space, then decoded data using a Gated Recurrent Unit (GRU), producing video simulations in 512x512 resolution