Brandon Lee Concepcion

+1 (650)-892-4233 | brandon_concepcion@berkeley.edu | linkedin.com/in/brandonconcepcion | brandonconcepcion.com
github.com/brandonconcepcion

Education

University of California, Berkeley *May 2026*
***Majors**: B.A. Data Science, B.A. Computer Science* *GPA: **3.8**/4.0*
***Technical Skills/Coursework:** Deep Learning, Neural Networks, Computer Vision, Natural Language Processing, A/B Testing, Machine Learning Theory, Data Structures, Data Engineering, Calculus, Linear Algebra, Statistics and Probability, Data Analytics*
***Organizations:** Data Science Society at Berkeley, Data C8 (Foundations of Data Science) Undergraduate Course Staff*
***Honors/Awards:** Genentech Futurelab Scholar, 6th Annual Datathon For Social Good: 2nd Place*

Skills And Interests

Languages: Python, Java, SQL, RegEx, LaTeX, HTML, Microsoft Excel, Scheme
Tools: Pandas, NumPy, TensorFlow, PyTorch, PostgreSQL, OpenCV, SciPy, StatsModels, Sci-kit Learn, Seaborn, Tableau, Plotly
Skills: Exploratory Data Analysis, Git, Data Visualization, Classification, Clustering, Data Engineering, Data Analytics, LLMs
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
Personal Interests: Movies, Volleyball, Photography, Basketball, Gym, Swimming, Dogs, Road Trips, Music, Marvel Studios

Work Experience

Data Science Society at Berkeley Berkeley, CA
President *Dec 2024 - Dec 2025*

- Directing the operations of UC Berkeley's oldest data science student organization by fostering a supportive and welcoming community environment as well as managing a leadership team of 26 members through 6 consulting projects for industry clients
- Led initiatives to promote data science accessibility by teaching 140+ students and organizing 10+ hackathons and panels

JamBase Berkeley, CA
Lead Data Scientist *Jan 2025 - May 2025*

- Led a 5-person team to the development, training, and query routing of an intelligent chatbot, architecting a full-stack pipeline that integrated BigQuery, Pinecone, and LangChain to deliver real-time recommendations from 3,000,000+ concerts records
- Built a multi-model query engine using GPT-based LLM-classification to route prompts to SQL, BigQuery ML, ARIMAX, or RAG, boosting prediction coverage by 30% and enabling 24-month forecasts across 12K+ live-music events.

UC Berkeley Data Science Undergraduate Studies Berkeley, CA
Software Developer *Jan 2024 - May 2025*

- Partnering with four El Camino College professors to develop technical curriculum for a student population of **33,000+**
- Designing and deploying two Jupyter-based data science modules, introducing students to data science and molecular biology
- Redesigned *data8.org/su24* using HTML, CSS, and JavaScript to build a web app that organizes 300+ past exam problems by topic, improving student comprehension by 8%

Doctors Without Borders Remote
Data Scientist *Aug 2024 - Dec 2024*

- Utilized Python and the Armed Conflict Location & Event Data (ACLED) API to achieve 93% accuracy in classifying global regions likely to experience fatalities from escalating political events, aiding in the identification of high-risk zones
- Preprocessed 2,000,000+ political events across 74 features and 180+ countries to train two Scikit-learn neural networks, achieving an R^2 score of 76% in predicting the number of fatalities for regions with escalating political events
- Utilized Tableau to visualize conflict severity and fatality predictions, enhancing humanitarian safety management

University of Washington Remote
Machine Learning Researcher *Jan 2024 - Jun 2024*

- Coded a Variational Auto-Encoder (VAE) neural network in PyTorch and OpenCV to convert numerical retinal 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 from an Excel spreadsheet 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

Projects

Spam Email Classifier | *Python, Pandas, Principal Component Analysis* *Nov 2023*

- Used Pandas, NumPy, and RegEx to develop an 87% accurate classification model to predict spam emails, utilizing a dataset of over 7,500 points and achieving an Area Under the ROC Curve of 91%
- Applied Principal Component Analysis (PCA) to reduce dimensionality by 70%, and enhanced model performance by 5% using GridSearch cross validation across 4 hyperparameters