

Du Xiang

+1 (858) 666-5552 | xd00099@berkeley.edu | Website: <http://du-xiang.com> | LinkedIn: <https://linkedin.com/in/du-xiang/>

EDUCATION	University of California, Berkeley , Electrical Engineering and Computer Science Master of Engineering in EECS, Data Science & Systems Track Honors: Fung Institute Excellence Scholarship (20,000 dollars) 2022 - 2023
	University of California, San Diego , Halicioğlu Data Science Institute Bachelor of Science in Data Science GPA: 3.99/4.0 Honors: Summa Cum Laude, Data Science Scholar's Award (Faculty Nomination) 2018 - 2022
TECHNICAL STRENGTH	Programming Languages: Python, Java, SQL, Bash, Javascript, HTML/CSS, C, R, MATLAB Software & Tools: Kubernetes, Docker, Pandas, React Native, Scikit-learn, Jupyter, PostgreSQL, Neo4j, TigerGraph, Apache Spark, Dask, AWS, Matplotlib, PyTorch, Hugging face, Gym, Keras, Git, Tableau, Matlab, Anaconda, Vim Text mining & NLP: Scrapy/Beautifulsoup, Text Preprocessing, Text Parsing, BoW/Tfidf, Topic Modeling, Relation Extraction, Named Entity Recognition, Bert, NLTK, Language Models, Mallet
SELECTED COURSES	Data Science and Machine Learning; System for Scalable Analytics; Natural Language Processing; Deep Learning Data Structures and Algorithms; Data/Web Visualization; Convex Optimization; Recommender Systems; Deep Reinforcement Learning; Parallel Computing (2023); User Interface Design and Development (2023)
EXPERIENCE	UC Berkeley Capstone Project Partner Simplr, Asurion 09/2022 – continue <ul style="list-style-type: none">Generating customer service response template using state-of-the-art language Transformer modelsCreating a generative conversational agent that would solve certain customer inquiries coherently and accurately
	Machine Learning and Natural Language Intern San Diego Supercomputer Center 02/2021 – 08/2022 <ul style="list-style-type: none">Applied and tuned BioBERT to extract 9000+ relations (chemicals and genes) and achieved 0.9+ F1 scoreTopic Modeling trained on 1000+ academic articles: achieved high topic interpretation score rated by domain experts. Results were used to find similar articles and build author-article graph database.Worked on Text Parsing Pipeline (XML parsing) and containerized different tasks using Docker.
	Instructional Assistant Halicioğlu Data Science Institute, UC San Diego 03/2020 – 06/2022 <ul style="list-style-type: none">IA for <i>Theories for Intro Machine Learning; Data Science Principles; Practice of Data Science & ML Pipeline</i>Help with structuring classes, holding office hours, and debugging in programming assignments
	Undergraduate Researcher CNS Laboratory of Memory and Brain 01/2020 – 06/2021 <ul style="list-style-type: none">Data cleaning/image processing: filled in 20% of missing data using statistical sampling, automating image editsMRI image analysis: applied decision tree to analyze important brain features, and filter top correlations with PCAWrote an automated pipeline that boosts WMH volume segmentation efficiency 10+ times w/ Bash and Python
	Network Optimizer Intern China Telecom, Suzhou, China 06/2019 – 08/2019 <ul style="list-style-type: none">Worked with wireless network optimization including coverage, band frequencies, locations, etc.Tested 5G devices and used software to calculate viable network distances for optimized solutions.
PUBLICATIONS	McInnes, Bridget, " Discovering Content through Text Mining for a Synthetic Biology Knowledge System " ACS Synthetic Biology 2021 (ACS Synth. Biol. 2022, 11, 6, 2043–2054) <ul style="list-style-type: none">Contributor of the Topic Modeling & Relation Extraction parts
PROJECTS	Faculty Information Retrieval System (Topic Modeling / NLP Python, Flask, Docker, Kubernetes website) <ul style="list-style-type: none">Used unsupervised learning (LDA) to categorize faculty members for industry needs; presented to the boardImplemented the backend structure including Data ETL, Preprocess, Modeling, and Web DeploymentBuilt a dashboard web app that integrates the fine-tuned model, search tools, and Sankey visualization
	Image Captioning Tool (Deep Learning / Web App PyTorch, Python, Flask, Heroku) <ul style="list-style-type: none">Built an encoder-decoder deep learning framework to identify images and generate caption sentences.Trained ResNet + RNN/LTSM with 30+ sets of hyperparameter; deployed as a web-app using Flask, Heroku
	Pocket Health App (IOS / Android App Development React Native, AWS Amplify, GraphQL, JavaScript) <ul style="list-style-type: none">Initiated a project that aims to help the people in need of free healthcare (a team with Med and CS students)Designed and implemented the reusable components for the front end using React-Native and JavaScriptBuilt user authentication system using AWS and designed the backend database schema in DynamoDB
	Interactive Author-Articles-Institutions Graph Network (Research / NLP Neo4j, Aura Cloud, SQL) <ul style="list-style-type: none">Constructed a graph relational database to show graphical relationships for 2000+ publication dataUsed text mining to bring together PubMed data and ACS data to highlighted novel topic clusters