

# Du Xiang

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

EDUCATION	<b>University of California, Berkeley</b> , Electrical Engineering and Computer Science Master of Engineering in EECS, Data Science & Systems Track Honors: Fung Institute Excellence Scholarship ( <b>20,000 dollars</b> ) <b>2022 - 2023</b>
	<b>University of California, San Diego</b> , Halicioğlu Data Science Institute Bachelor of Science in Data Science   GPA: <b>3.99/4.0</b> Honors: <b>Summa Cum Laude</b> , <b>Data Science Scholar's Award</b> (Faculty Nomination) <b>2018 - 2022</b>
TECHNICAL STRENGTH	<b>Programming Languages:</b> Python, Java, SQL, Bash, Javascript, HTML/CSS, R, MATLAB <b>Software &amp; Tools:</b> Kubernetes, Docker, Pandas, React Native, Scikit-learn, Jupyter, PostgreSQL, Neo4j, TigerGraph, Apache Spark, Dask, AWS, Matplotlib, Seaborn, PyTorch, Keras, Git, Tableau, Matlab, Anaconda, Vim <b>Text mining &amp; NLP:</b> Web crawling with Scrapy/Beautifulsoup, word cleaning, Tfidf, Topic Modeling, Relation Extraction, Named Entity Recognition, XML Parsing, Bert, NLTK, Language Models, Mallet
COURSES	Theories and Applications of Data Science; Algorithms and Programming/Data Structures; Recommender Systems and Web Mining; Convex Optimization for Data Science; Data/Web Visualization; Probability and Statistics. Numerical Analysis; Systems for Scalable Analytics; Modeling and ML; Advanced Algorithms; Deep Learning; Reinforcement Learning; Parallel Computing (in 2023); User Interface Design and Development (in 2023);
EXPERIENCE	<b>Synthetic Biology Knowledge System, San Diego Supercomputer Center</b> <u>Machine Learning and Natural Language Intern</u> <b>02/2021 – 07/2022</b> <ul style="list-style-type: none"><li>Applied <b>state-of-the-art language models</b> (BioBERT) to <b>extract relations</b> between chemicals and genes</li><li><b>Topic Modeling</b> trained on 1000+ academic articles: achieved high topic interpretation score rated by domain experts. Results were used to find similar articles and build <b>author-article network</b>.</li><li>Organized Text Parsing code (<b>XML parsing</b>) and <b>containerized</b> different tasks using <b>Docker</b>.</li></ul>
	<b>Halicioğlu Data Science Institute, UC San Diego</b> <u>Instructional Assistant</u> <b>03/2020 – 06/2022</b> <ul style="list-style-type: none"><li>IA for <i>Theories for Intro Machine Learning; Data Science Principles; Practice of Data Science &amp; ML Pipeline</i></li><li>Help with structuring classes, holding <b>office hours</b>, and <b>debugging</b> in programming assignments on Gradescope</li></ul>
	<b>CNS Laboratory of Memory and Brain</b> <u>Undergraduate Research Assistant</u> <b>01/2020 – 06/2021</b> <ul style="list-style-type: none"><li><b>Data cleaning/image processing:</b> worked on fill in missing values using statistical method, automating image edits</li><li>Brain <b>data visualization</b> and <b>statistical analysis:</b> visualized using <b>matplotlib</b>, experimented <b>correlations</b> with <b>PCA</b></li><li>Individual Research Project on White Matter Hyperintensity with <b>prediction to Alzheimer's Disease</b></li><li>Wrote a <b>data pipeline</b> for quantifying WMH volume analysis for 50+ subjects using <b>Bash and Python</b></li></ul>
	<b>China Telecom, Suzhou, China</b> <u>Intern, Assistant to the Network Optimizer</u> <b>06/2019 – 08/2019</b> <ul style="list-style-type: none"><li>Worked with wireless network optimization including concepts like <b>coverage</b>, <b>band frequencies</b>, etc.</li><li>Analyzed and cleaned user data using <b>Pandas and Python</b> to assist creating network solutions</li></ul>
PUBLICATIONS	McInnes, Bridget, " <b>Discovering Content through Text Mining for a Synthetic Biology Knowledge System</b> " ACS Synthetic Biology 2021 (ACS Synth. Biol. 2022, 11, 6, 2043–2054) <ul style="list-style-type: none"><li>Contributor of the <b>Topic Modeling &amp; Relation Extraction</b> parts</li><li>Previous Paper: Monte, Jeanet, "<b>Synthetic Biology Knowledge System</b>", ACS Synthetic Biology 2021, 10, 9, 2276-2285</li></ul>
PROJECTS	<b>Faculty Information Retrieval System</b> (Topic Modeling   Python, Flask, Docker, Kubernetes   <a href="#">website</a> ) <ul style="list-style-type: none"><li>Topic Modeling and NLP to group faculty members based on their publications to serve industry needs</li><li>Implemented the backend structure including Data ETL, Preprocess, Modeling, and Web Deployment</li></ul>
	<b>Image Captioning Tool</b> (Deep Learning   PyTorch, Python, Flask) <ul style="list-style-type: none"><li>Encoder-Decoder Framework using CNN + LSTM to generate image captions.</li></ul>
	<b>Pocket Health App</b> (App Development   React Native, AWS Amplify, DynamoDB, JavaScript) <ul style="list-style-type: none"><li>Designed and implemented the major components for the front end</li><li>Designed the backend database using DynamoDB and optimized the App logic to allow fast edit and retrieval</li></ul>
	<b>Interactive Author-Articles-Institutions Graph Network</b> (Research   Neo4j, Aura Cloud, SQL) <ul style="list-style-type: none"><li>Graph relational database constructed to show a graph network for publications in Synthetic Biology</li></ul>