**JACK LUCAS CHANG**

564-229-5879 | jacklucaschang@gmail.com | linkedin.com/in/jacklucaschang | Seattle, WA

**PROJECTS**

##### ML: MULTI-LABEL CONVOLUTIONAL NEURAL NETWORK *Final project for Applied Machine Learning at University of California, Berkeley* 05/2024-08/2024

##### Developed and implemented advanced CNN classifiers for multi-label image classification, resulting in models with 82%, 84% and 98% accuracy across characteristics including type, evolutionary stage, and legendary status.

##### Fine-tuned hyperparameters to optimize performance and improve prediction accuracy by systematically adjusting learning rates, batch sizes, and network architecture parameters.

##### NLP: OPTICAL CHARACTER RECOGNITION, NAMED ENTITY RECOGNITION

##### *Converting medical records to Fast Healthcare Interoperability Resources (FHIR)* 07/2024 - present

##### Led the first-ever integration of FHIR data with a biobank and digital tissue slides library in the canine-cancer research community by developing the conversion of over 17,000 unstructured veterinary medical records into FHIR using Amazon Textract and Healthlake, democratizing access to NLP tools for the veterinary research community.

##### XGBOOST & CLOUD DATA ARCHITECTURE

##### *Capstone project at University of California, Berkeley* 04/2025 – 08/2025

* Engineered end-to-end cloud data architecture processing 150+ million bird observation records using AWS S3 data lake, Glue ETL jobs with automated daily ingestion pipelines.
* Developed real-time prediction capabilities with XGBoost model analyzing noise impact on bird populations, implementing interactive visualization dashboard with REST API integration.

**EXPERIENCE**

##### MORRIS ANIMAL FOUNDATION

*Data Engineer*09/2021 – present

* Led the design and maintenance of business-critical ETL pipelines, ensuring uninterrupted flow and integrity across systems pivotal for decision-making processes, driving impactful insights in a $40 million project and facilitated the publication of over 10 scientific papers.
* Drove architectural planning and implementation for future-proof data storage, reporting, and analytic solutions, aligning technological advancements in data pipeline tools with organizational objectives, resulting in 75% reduced time to build millions of rows of transformed data.
* Successfully migrated a legacy PHP and SQL data transformation framework to Python, enhancing scalability and efficiency to accommodate an additional 4 million rows.

##### MICROSOFT STREAM

*Database Engineer Intern*01/2020 – 06/2020

* Designed and implemented scalable database architecture for Microsoft Stream using SQL Server 2016 with advanced SQL operations, managing enterprise-level video content and user engagement data platforms.
* Built ETL pipeline infrastructure using Python to evaluate UI and content recommendation impact on user engagement metrics, driving 25% increase in user satisfaction and 20% boost in feature adoption.

**SKILLS**

##### Programming Languages: Python, SQL, R

##### Cloud Platforms: AWS (S3, Glue, Redshift, Lambda), Google Cloud Platform (BigQuery)

##### Machine Learning: XGBoost, PyTorch, TensorFlow, Scikit-learn

##### Data & Visualization: BigQuery, SQL Server, PostgreSQL, LookerStudio, Tableau

**Tools:** Git, Docker, Jenkins, Jupyter, Apache Spark (learning)

**EDUCATION**

##### UNIVERSITY OF CALIFORNIA, BERKELEY Berkeley, CA

*Master of Information and Data Science* *(Machine Learning)*08/2025

##### UNIVERSITY OF WASHINGTON, SEATTLE Seattle, WA

*Bachelor of Arts, Philosophy (Ethics); Bachelor of Science, Informatics (Database Engineering)* 06/2021