

Jun Yu (Johnny) Chen

LinkedIn ▪ (310) 962-5942 ▪ jochen030327@ucla.edu

EDUCATION

- ❖ **University of California, Los Angeles** **GPA: 3.922/4.0** **(Fall 2021 – Expected June 2025)**
Major: Bachelor of Science, Cognitive Science with Specialization in Computation & **Minor:** Data Science Engineering

HONORS

- ❖ **Dean’s Honor List & UCLA Statistics Club & UCLA DataRes Research & UCLA Bruin Sports Analytics**

COURSEWORK

- ❖ **Data Science:** Regression and Data Analysis, Machine Learning Specialization, Data Mining, Convolutional Neural Network, Generative Data Science,
- ❖ **Computer Science:** Advanced C++ Programming, Algorithms, Computer Vision and Deep Learning, Sequence Models, Fundamentals of Reinforcement Learning, Artificial Intelligence and Symbolic Reasoning
- ❖ **Cognitive Science:** Cognitive Psychology and Psychobiology, Human Memory and Decision Making, Principles of Behavioral Neuroscience

RESEARCH INTEREST

- ❖ Multi agent reinforcement learning and collaboration
- ❖ Open-world LLM and multimodal model driven agents
- ❖ Computer vision and Object Detection

PUBLICATIONS

- ❖ Andrea Kang, **Jun Yu Chen**, Youngzie (Zoe) Lee, Shuhao Fu, “Synthetic Data Generation with LLM for Improved Depression Prediction”, submitted to **AAAI 2025 W42: Large Language Model and Generative AI for Health**.
 - ❖ **Arxiv:** <https://arxiv.org/abs/2411.17672>
 - ❖ **OpenReview:** <https://openreview.net/forum?id=94h6UVEroD>
- ❖ **Jun Yu Chen**, Tao Gao, “APT: Architectural Planning and Text-to-Blueprint Construction Using Large Language Models for Open-World Agents”, submitted to **AAAI 2025 W17: Planning in The Era of Large Language Models**.
 - ❖ **Arxiv:** <https://arxiv.org/abs/2411.17255>
 - ❖ **OpenReview:** [https://openreview.net/forum?id=PCE0vO1sZx&referrer=%5Bthe%20profile%20of%20Jun%20Yu%20Chen%5D\(%2Fprofile%3Fid%3D~Jun_Yu_Chen1\)](https://openreview.net/forum?id=PCE0vO1sZx&referrer=%5Bthe%20profile%20of%20Jun%20Yu%20Chen%5D(%2Fprofile%3Fid%3D~Jun_Yu_Chen1))
- ❖ **Jun Yu Chen**, David Lipkin, ... “A graphical user interface that integrates real-time marker-less pose detection and experimental interventions”, presented at **Neuroscience 2024**, in preparation for submission to **Frontiers in Systems Neuroscience**.
- ❖ Sean Wu*, **Jun Yu Chen***, “Self-supervised denoising of visual field data improves glaucoma progression detection” in preparation for submission to **IEEE Journal of Biomedical and Health Informatics**. Available on arXiv: <https://arxiv.org/abs/2411.12146>
- ❖ Vahid Mohammadzadeh, Sean Wu*, **Jun Yu Chen***, ... “Prediction of Final Visual Field and Glaucoma Progression with Deep Survival Analysis”, in preparation for submission to **American Academy of Ophthalmology**.

RESEARCH & WORK EXPERIENCE

Visual Intelligence Lab (Statistics & Data Science Department)

Machine Learning Researcher

March 2024 – Present

- ❖ Engineered LLM-based agent frameworks with RAG memory pools, primitive actions, and interactive environment using Mineflayer and LangChain for strategic planning and task executions in simulated environment, under the guidance of Dr. Tao Gao.
- ❖ Developed and evaluated agent performance benchmarks, testing the agent's spatial reasoning, building accuracy, and time-to-completion with various architecture settings, achieving over 90% success rate for task type-1 with GPT-4o backend.
- ❖ Developed multimodal LLM-based workflow for single-agent and multi-agent collaboration, featuring self-learning and self-reflection mechanisms. The workflow enables agents to translate image and textual construction instructions into coordinate-based 3D blueprints, leading to the accurate execution of complex structures in Minecraft, such as wooden houses and redstone-driven architecture.
- ❖ Implemented multi-agent reinforcement learning algorithms, including Multi-agent Deep Deterministic Policy Gradient (MADDPG) and QMIX, from scratch using PyTorch to develop codebase for benchmark testing in multiplayer war simulations.

Ahmet Arac Lab (Neurology Department)

Data Science & Machine Learning Researcher

May 2023 – Present

- ❖ Conducted interdisciplinary research focusing on the neural and behavioral mechanisms of skilled motor movements in mice, utilizing advanced machine learning algorithms and statistical models in collaboration with Dr. Ahmet Arac.
- ❖ Built DeepBehavior, a realtime, all-in-one Graphical User Interface(GUI) pipeline featuring real-time object detection with OpenCV, customized labeling, model training modules, and interface to control parameters for optogenetic manipulations.
- ❖ Utilized Pytorch torchvision and adapted MobileNet, along with NIDAQ boards and FLIR cameras, proving effective in benchmark tests for detecting specific mouse movements, enabling precise multi-step optogenetic interventions based on predefined probability thresholds.
- ❖ Implemented Adversarial Inverse Reinforcement Learning(AIRL) from Imitation Library and OpenAI Gym for recovering reward networks, integrating these into a Variational Gaussian Hidden Markov Chain model to identify hidden states in mouse trajectories.

Job Junior

Founder & Machine Learning Engineer

Sep 2023 – Present

- ❖ Founded Job Junior, a start-up that utilizes real-time scraping and large language models (LLMs) to automate the job search process across multiple industries, creating a structured, user-friendly tracking database for job seekers and career planning organizations.
- ❖ Developed and customized LLM agent workflows using LangChain and OpenAI API, enabling automated job analysis and real-time data autocorrection, outperforming traditional keyword and rule-based heuristic method used by the current market by 30%.
- ❖ Fine-tuned Named Entity Recognition (NER) models, classifier, and LLaMA3 pipeline leveraging spaCy transformer models and Unsloth QLoRA, achieving 93% accuracy in job classification, application timelines, and experience level identification.
- ❖ Led strategic planning and contract negotiations with companies like DBC, OSG, and Jumpstart. Directed software engineering and data annotation teams, deploying over 480 company websites and annotating 50,000 training data using Prodigy within 3 months.

Stein Eyes Institute

Data Science Research Intern

June 2023 – March 2024

- ❖ Collaborated on research initiative under the guidance of Dr. Fei Zhe, focusing on deep learning models and the evaluation of Masked and Variational Autoencoders (VAEs) for their effectiveness in Visual Field reconstruction and glaucoma diagnosis
- ❖ Implemented multiple evaluation metrics such as Percent Loss Ratio (PLR), Mean Deviation (MD), and Glaucoma Rate Index (GRI) to streamline pipeline for automated preprocessing, diagnosis, and progression index of over 25,000 rows of patient VF data
- ❖ Developed and finetuned Convolutional Neural Networks, achieving MAE of 3.1 in predicting final visual field (VF) data and TTP.
- ❖ Discovered critical insights through the Masked Autoencoders reconstructed data, resulting in a quantifiable 10% improvement in the diagnostic accuracy for detecting progressive stages of glaucoma; the results are in preparation of peer-reviewed publication.

PROJECT EXPERIENCE

Synthetic Data Generation with LLM for Improved Depression Prediction

Generative Data Science Grad Course

April 2024 – Nov 2024

- ❖ Generated synthetic patient transcripts using fine-tuned Llama and chain-of-thought (CoT) prompting, enhancing downstream BERT-based predictions for PHQ-8 depression scores.
- ❖ Designed a CoT LLM framework that reconstructs patient transcripts into synopsis and sentiment analyses, achieving a root mean square error (RMSE) of 4.39 and a mean absolute error (MAE) of 3.53—outperforming baseline results from dual encoder models and original data
- ❖ Evaluated the fidelity of synthetic data using KL divergence with word-level frequency distribution and assessed privacy with TF-IDF similarity metrics, achieving favorable results for both

Bruin Sports Analytics (BSA)

Research Journalist

Feb 2022 – Spring 2023

- ❖ Developed neural network model with cost-sensitive training method, accurately predicted the 2022-2023 NBA champion and top 3 contenders, achieving 90% test accuracy by leveraging pre-trade, post-trade, playoff data, Elo scores, and other engineered features.
 - Performed model evaluation and tuning utilizing grid search, key metrics including PCA, confusion matrix, and ROC curve.
 - Executed feature importance analysis using Gradient Boosting Classifier, unveiling critical factors like Elo score and playoff eff that significantly contributed to the model's predictions; Authored another BSA headline research article on this project.
- ❖ Deployed a content-based deep neural network recommendation system for League of Legends (LoL) players using Streamlit, offering personalized champion recommendations based on player preferences, historical data, and champion dynamics.
 - Utilized RiotWatcher, web scraping, and custom functions to extract and preprocess 50,000 rows of data from Riot APIs.
 - Performed model evaluation and tuning utilizing grid search, key metrics including PCA, confusion matrix, and ROC curve.
 - Achieved 70% accuracy in predicting match-up win rates based on selected champions of teammates, opponents, and team positions; Authored headline BSA research article and developed an interactive dashboard on this work for users.

SKILLS

- ❖ **Machine Learning:** TensorFlow, Pytorch, Imitation
- ❖ **Programming languages:** Python, R Studio, C++, Git, SQL, Selenium, Request
- ❖ **Document Preparation & Design Tools:** LaTeX, Procreate, Final Cut Pro