Xinyu (Ivy) Li

ivy@studioxolo.com | (718) 406-6863 | Jersey City, NJ | linkedin.com/in/ivyxli | ivylee.github.io

WORK EXPERIENCE

Independent Consultant | Studio Xolo

AI software development & advisory

- Building GrocerBird.com, an app written in Django and Javascript that extracts structured data from grocery receipts snapshots using Vision Language Models, and delivers analytics using Pandas and Plotly for optimizing grocery expenditure.
- Built HoodKit, an aggregated database and API based on 47 NYC real estate data feeds updated daily to accelerate investment property research, using Apache Airflow and SQL for data ingestion, AWS RDS (PostgreSQL) as database, FastAPI for REST API backend service and JSON Web Token for secure integration with frontend service.
- Built RoughCut, an AI-assisted documentary editing app using Google Cloud Vision to annotate footage content for search and retrieval.
- Prototyped real-time object detection system on NVIDIA Jetson equipped drones for UAV-based inspection and surveillance services..
- Collaborated with Professor David Parke's EconCS research group at Harvard University to refactor, modularize and modernize a Multi-Agent Reinforcement Learning codebase for researching Mechanism Design; taught git branching, code review and software engineering best practices to academic researchers.
- Mentoring machine learning engineers and leaders.

Lead Machine Learning Engineer | Turing Labs

AI-powered B2B SaaS for Fortune 500 Consumer Packaged Goods companies

- As the first ML Engineer, designed and built a Probabilistic AI system from scratch to help CPG R&D leaders to reduce testing, decrease formulation cost, optimize product performance and find new ingredients, delivered \$65M+ in business impact.
- As a tech lead manager, recruited and led a 5-person ML Engineering team including research scientists, machine learning engineers, and software engineers in a remote-first culture across four time zones with a focus on diversity, inclusion, psychological safety, and best practices.
- Built a REST API backend service using Python, FastAPI and PostgreSQL to enable automated surrogate model training, model deployment and inference, as well as Bayesian Optimization (Bayesian Experimental Design), reducing time to experiment by 90%.
- Designed and implemented AI system architecture and infrastructure with Docker and AWS (ECS, RDS, S3) to work with an existing Web application in AWS. Prepared cloud and application architecture diagrams for due diligence.
- Leveraged Large Language Models via OpenAI API to build a quantitative Knowledge Graph of formulation science to solve the cold start problem.
- Reimplemented algorithm research prototypes in R into production Python codebase with performance optimizations.

Co-founder & CTO | Phiar Technologies

Auto Edge AI & Augmented Reality navigation

- Built and led a 6-person agile R&D team of software engineers and 3D perception experts. Led technical strategy and development, and technical due diligence.
- Researched, implemented and deployed efficient Deep Neural Networks achieving real-time semantic segmentation on mobile devices using PyTorch, Onnx, Swift and Apple Metal Framework.
- Created Deep Neural Network training pipeline in PyTorch optimized for research iteration speed, experimented with cutting edge algorithms, leveraged Git branches for reproducible experiments.

Sep 2019 – Present Remote

Dec 2020 - Jan 2024

Remote

Jun 2018 - Aug 2019

Remote

- Co-authored 2 patents for augmented reality navigation systems integrating computer vision and advanced driver assistance functionalities.
- Accepted into Y Combinator S18; acquired by Google.

Senior Machine Learning Engineer | Shutterstock

Leading global marketplace licensing 500M+ digital media content to businesses and creatives

- Implemented a large-scale (100m+) next-generation visual search engine consisting of a custom Vector Database for image embeddings and retrieval algorithm written in C++ and gRPC, and a REST API written in Python. Containerized using Docker, set up continuous integration on GitHub and deployed to AWS ECS.
- Built an experimental search ranker using a machine learning model, including a data pipeline written in Apache Spark (Scala) and SQL (Hive on Hadoop) for gathering large amounts of historical data for training, and a REST API service written in Python.
- Prototyped and productionized image auto-tagging (keyword recommendation) AI feature, created back-end API design doc, collaborated with Mobile and Web product managers and developers for integration. This feature gained press coverage, increased mobile content submission by 60%.
- Created a visual clustering algorithm to improve diversity of image search results and make it easier to find the right content, won a company hackathon.
- Created the back-end of a visual search chrome plugin for migrating customer's saved assets from competing marketplaces, won a company hackathon, and patented the algorithm.
- Trained and evaluated Deep Neural Networks for image understanding using PyTorch and Tensorflow.

Data Scientist | Socure

Fintech B2B SaaS unicorn providing AI-powered digital identity verification and fraud prevention solution

- As an early employee, researched and developed a novel real-time face similarity scoring algorithm written in Java and Scala, achieved 0.83 AUC, and improved identity fraud detection rate by 10%.
- Contributed an algorithm bug fix to the open source OpenCV project on GitHub.

PATENTS

- Computer-vision based positioning for augmented reality navigation. U.S. Patent #10,871,377.
- Mobile real-time driving safety systems and methods. U.S. Patent #10,573,183.
- Augmented reality navigation systems and methods. U.S. Patent #10,495,476.
- Migrating image histories between different databases. U.S. Patent #10,585,953.
- Analyzing facial recognition data and social network data for user authentication. U.S. Patent #9,147,117.

EDUCATION

Columbia University M.A., Statistics Guangdong University of Foreign Studies B.S., Mathematics & Applied Mathematics

SKILLS & INTERESTS

- Skills: Python, PyTorch, AWS, Docker, CI/CD, Git, Apache Spark, Apache Airflow, SQL, Java, Scala, R
- Interests: Rowing, Reading, Cooking, Art Museums

Oct 2013 – Jul 2014

New York, NY

Feb 2014 New York, NY May 2012 Guangzhou, China

Jan 2015 – May 2018

New York, NY