Xingfan Xia

8502 134th Ct NE, Redmond, WA 98052
xingfanxia@gmail.com • +1 (507) 403-1689 • https://xiax.xyz

WORK EXPERIENCE

Compute Labs, Remote, Redmond, WA

- Co-Founder and CTO — Aug 2024 – Present
  - **Overseeing the entire technological vision and strategy of the company, ensuring alignment with business goals and market demands.**
  - Managing and leading the engineering team, including hiring, mentoring, and developing top-tier talent.
  - Spearheading product launches, from conceptualization to market release, ensuring technical excellence and market fit.

Duckie (YC W24), Remote, Redmond, WA

- Founding Engineer — Jul 2024 – Aug 2024
  - **Engineering Lead working on the revolution of T2 technical support.**

Amazon Web Service – Athena, Redmond, WA

- Software Engineer — Apr 2022 – Present
  - **Building the next-gen storage infrastructure for large-scale distributed systems with billions of rows and terabytes of data, improving the performance of AIML and other big data analytics applications.**
  - **Tech Stack: Java, Python, Spark, Iceberg, AWS Services(S3, CodePipeline, CodeBuild, Cloudwatch, etc.)**
  - Implemented CTAS(Create Table As Select) feature for Iceberg Tables in Athena query engine. Released Iceberg Table GA Preview RE:INVENT 2022 and millions of CTAS queries being executed since the release.
  - Overhauled the existing Hive JSON Serde library, improved the query execution performance by up to 30%, and fixed non deterministic query results, reduced related ops tickets by 60%.
  - Implemented a virtualized table connector on top of the Iceberg table connector so users can leverage the performance of the Iceberg table engine to access other table formats with limited scalability. Improved the query execution speed of these table formats by approximately 40%, avoided congestion, and is projected to save 4 million in hardware costs per year.
  - Worked on a cross-team endeavor to overhaul the query routing logic from the Athena engine to multiple external services, reduced excessive retry behaviors, and reduced service congestion by up to 32%

Apple, Cupertino, CA

- Machine Learning Engineer — Aug 2020 – Apr 2022
  - **Improving users' experience with proactive AI assistant on iOS, macOS, watchOS, tvOS.**
  - **Tech Stack: Python, Objective-C, Sklearn Kit, Pytorch, Airflow, Hive, Presto, Spark**
  - Implemented an on-device user profiling framework that consists of several models to predict users' overall interest and preferences as an upstream service that provides signals to downstream consumers like Apple Map, and Apple News.
  - Improved the on-device location model with 4 times of training data and better engineered aggregated features, increased user engagement rate by 9%. Also redesigned the model inference logic on device to reduce latency by up to 22%.
  - Overhauled the topic prediction model leveraging transformer-based model BERT, increased topic prediction accuracy by 14% and user engagement rate by 11%. Also improved the inference pipeline to reduce latency by up to 12%.
  - Developed an automatic machine learning pipeline that retrains models with data source updates, launches A/B testing experiments, and deploys new models if a specified improvement is observed, saving engineering hours across the team.
  - Worked on action prediction service which predicts user intention at a particular time and location by aggregating various signals from multiple models input which is used in Siri Suggestions and Siri Shortcuts.

Airbnb, San Francisco, CA

- Software Engineer — Jan 2019 – Aug 2020
  - **Fighting financial fraud with machine learning models and rule-based engine.**
  - **Tech Stack: Python, Java, Scala, Sklearn Kit, Airflow, Hive, Presto, Spark**
  - Developed various models(XGBoost and logistic regression) to tackle the fake review problem at Airbnb, cutting the volume of fake reviews to 1% of the peak volume and preventing 2 million losses. Also deployed it with a new inference mechanism, improve inference performance by 30%.
  - Trained a logistic regression model to produce a comprehensive risk score with Sklearn, evaluating the overall risk of users on Airbnb, especially on Airbnb hosts, similar to Alipay's Zhima/Seasame score.
  - Rolled out a new policy that delays the payout of risky Airbnb hosts, prevented 60% of fraudulent payouts being made and hence in an estimated 10 million in annual savings.
  - Built a Java microservice that provides downstream services with consolidated risk signals using Redis, Hive, and Kafka.
  - Designed and Implemented a scoring Java service that produces risk scores of certain events and scenarios supporting multiple downstream consumers using Redis, Hive, and Kafka.

EDUCATION

Carleton College, Northfield, Minnesota

- B.A. in Computer Science — Graduated Jun 2018