Jerome Yutai Shen, Ph.D.

Sunnyvale, CA 94089 | Tel: 310-309-0228 | Email: yutai.shen@gmail.com
LinkedIn: https://www.linkedin.com/in/jerome-s-986123113/
10 years of experience in software development

Skills

- **Languages & Framework:** C/C++, Golang, JAVA, JMP/JSL, Julia, Octave/MATLAB, Python, Dash, R, Scala, Shell, SQL.
- **Tools:** Azure (Data Factory, Synapse), AWS (SageMaker, EC2, ECS, S3), GCP, Apache Spark, Apache Storm, Apache Flink, Hadoop, MySQL, PostgreSQL, Oracle SQL, Data Stage, Redis, Big Query, Docker, Jenkins, Argo, Bazel, MinIO, Clickhouse..

Professional Experience

❖ **A Stealth Startup.**   *07/2023 – present*
Lead Python Developer
- Architect and implement scalable solutions, ensuring clean, maintainable, and efficient code aligned with industry best practices.
- Collaborate with cross-functional teams to define software requirements and project goals.
- Lead and mentor a python developer team, provided technical guidance, and support to ensure the team's success in delivering high-quality software solutions.
- Develop and maintain back-end systems and APIs, emphasizing seamless integration with front-end applications, including AWS web interface development.
- Achieved optimal API performance through query optimization, caching strategies, and best practices, supported by thorough testing that reduced bugs by 40%.

❖ **Meta, Menlo Park, CA, via Pyramid Consulting Inc.**   *07/2022 – 01/2023*
Software Engineer III
- Computed benchmarks of GPU collectives with Meta's **PARAM** platform, developed **Python 3.11+** modules and scripts to automate computation.
- Implemented asynchronous log parsing, designed **regexp** patterns for different granularities, increased the correct rate to over 99%.
- Customized the output formats, structured the data to JSON and XML, validated the benchmark results, analyzed the throughputs, performance metrics and system metrics of GPU collectives.
- Largely improved the efficiency of the whole team by 2-5 times on average.

❖ **Kaiser Permanente, Oakland, CA, via Pyramid Consulting Inc.**   *10/2020 – 04/2022*
Software Engineer
- Architected and redesigned another system for drug education coordinators in Kaiser Permanente, by integrating data analysis, machine learning and data visualization functionalities via **Python 3.7+** and **Flask**, **Plotly**, **Dash** framework and **pandas**, **statsmodels**, **Scikit-learn**, **Keras** and **Tensorflow**.
- Implemented authentication and authorization mechanisms, enhancing the security and preventing unauthorized access.
- Documented API endpoints and the modules, request/response payloads, and authentication methods, the architecture of the project, easing the onboarding process for new developers.
- Used **Redis** as the cache layer to optimize system read performance. Implemented **Redis** distributed lock to handle race conditions among different worker nodes.
- Integrated **DataDog** and **Sentry** for better error alerting and performance monitoring.
- Introduced **Azure Front Door Content Delivery Network** for rapidly delivering pictures and videos.
- Designed schema and migrated data in relational databases (**Hive**, **PostgresQL, Oracle SQL** and **Azure Data Studio**) to serve new features. Designed and implemented database models, ensuring efficient data storage and retrieval using **SQLAlchemy ORM**.
- Created **ETL data pipelines** and **data flows** in **Azure Data Factory**, generated tables in **Azure Synapse**.
- Leveraged **Azure Pipelines** to automatically build and test code projects for **CI/CD**.
- Ported the reported features to **Power BI** with a Direct Query option in **SQL** to get better insights on forecast.
- Worked to leverage Big Data **Hadoop** and **Spark** infrastructure to empower the organization.

❖ **KLA-Tencor, Milpitas, CA, via Valiantica Inc.**   *12/2019 – 06/2020*
Software Engineer
- Led a 5 people group to port the **JMP** apps to **Python 3.7+**, leveraged **AWS SageMaker**, **pandas**, **numpy**, **scikit-learn**, **matplotlib and jupyterlab**.
- Developed APIs with **Python Flask**, **FastAPI** frameworks to provide data storage and access services with **AWS S3** storage.
- Improved legacy codebase by constant refactoring, reorganization of dependencies, introduction of unit tests, integration tests, and end to end tests, achieved bug free and averaging twice faster than before, and high scalability with the framework **Pytest**.

❖ **HP Labs, Palo Alto, CA, via Radiant Systems.**   *10/2018 – 10/2019*
Software Engineer
- Worked at Immersive Experience Lab which develops the next generation VR device, built the off-line and on-line analysis pipeline.
- Built a video transcoding system (which has six main components: preprocessor, **DAG scheduler**, resource manager, task workers, temporary storage, and encoded video as the output) to encode the videos collected from VR device into compatible bitrates and formats.
- Implemented and optimized signal preprocessing and feature engineering programs employing **zero-phase filtering**, **Hilbert transform**, **Welch method** and **multi-taper windowed FFT** based **Nonparametric Power Density Spectral estimation** and **leave-one out cross validation**.
- Applied **Continuous Wavelet Transform** and **Wigner-Ville transform** to compute Spectrogram with **PyWavelets** and **PyTFTB**.
- Introduced **CNN** model and **LSTM** model to optimize feature engineering by dint of **TensorFlow** and **Pytorch**.

❖ **Shanxi Mount Heng Real Estate Development Limited, Taiyuan, China**   *06/2017 – 10/2018*
Software Engineer
- Built an internal house recommendation and analysis system for the company, it can help our realtor to build better pricing and sale strategies.
- Created crawlers to collect descriptions and reviews of open or rentable houses from local realtor websites and structured the data to JSON.
- Calculated an inverted index table, leveraged **Apache Pig** to index on stemmed word and used **Hive** to load details table to **HBase**.
- Designed the **DAG models** which defines tasks in stages so they can be executed sequentially or parallelly.
- Implemented the crawler job scheduler to fetch houses' new reviews and posted them into **Kafka**.
- Used **Apache Storm** to consume messages from **Kafka** and integrated the new data with existing records in **HBase**.
- Developed **Golang** scripts for cron jobs.

❖ **Department of Physics and Astronomy, UCLA.**   *11/2013 - 11/2016*
Staff Research Associate 3
- Designed signal processing algorithms with **MATLAB** for **FIR filters designing**, **zero-phase anti-aliasing filtering**.
- Computed selected features, performs dimension reduction machine learning algorithms such as **PCA**, **t-SNE**, and clustering such as **k-means++, Mixture of Gaussians** and **Spectral Clustering**, with **MATLAB**.
- Adopted a **Generalized Linear Model (GLM)** to minimize the influence of behavioral biases, and obtained an **unbiased estimate**, with **MATLAB**.
- Ported a **Python 2.7.x** based standalone app to **Python 3.3.x** in accordance with updates of dependent packages.

❖ **David Geffen School of Medicine, UCLA.**   *11/2012 - 10/2013*
Postdoctoral Scholar
- Developed experiment controlling and visual stimulus creation system in **C/C++ 11** and **RTOS QNX**.
- Designed algorithms with **MATLAB**, combined **spline interpolation**, **FIR filters designing** and **zeros-phase filtering** for eye movement signal.

Education Experience

❖ **Beijing Normal University,** *Ph.D. in Cognitive Neuroscience*   **Beijing, China** *09/2009 - 06/2012*
❖ **Shanxi Medical University,** *M.Sc. in Physiology*   **Taiyuan, China** *09/2006 - 06/2009*
❖ **East China Normal University,** *B.Sc. in* Biotechnology   **Shanghai, China** *09/2001 - 06/2005*