Patrick Song

530-601-3363 | pcsong33@gmail.com | github.com/pcsong33

Education

Harvard University, GPA: 3.99 (Dual Undergrad/Masters Program)

A.B. in Computer Science: Magna Cum Laude, Highest Honors in Department — Sep 2019 – Dec 2023

S.M. in Computer Science — Dec 2022 – Dec 2023

Experience

Patreon

SOFTWARE ENGINEER, DATA — Feb 2024 – Present

• Designed and implemented a self-service DynamoDB ETL pipeline to replace Fivetran, reducing infrastructure dependency and improving scalability. Achieved ~$1.2M in yearly cost savings by eliminating Fivetran usage and downsizing its SQL Warehouse.
• Led end-to-end migration of a 50K+ line analytics codebase into the core data engineering repository. Standardized 40+ production pipelines on Airflow, integrating alerting, testing, and CI/CD frameworks. Improved reliability, observability, and development velocity across data teams.
• Designed and deployed a scalable Pyspark data pipeline powering creator post and membership insights with email and push engagement metrics. Delivered performant Elasticsearch outputs enabling real-time post analytics for 1M+ creators.

DATA ENGINEER INTERN — Jun 2023 – Sep 2023

• Developed async scheduled data ingestion tasks in Databricks, DynamoDB, and Terraform to ingest social data from creators to quantify the impact of Patreon promotion on new memberships for over 1000 of Patreon's biggest creators.
• Visualized positive sharing moments for creators on the Insights Page using React and Nivo front-end frameworks.

Harvard University: Privacy Tools Research Group

RESEARCH INTERN — Jun 2022 – Dec 2023

• Performed user studies with over 20 library developers and data scientists, examining mental models of privacy concepts. Presented research at TPDP and CSCW, two premier Differential Privacy and HCI conferences, respectively.
• Implemented a differentially private synthetic data generator in the OpenDP core Rust library, with applications for clustering, regression, and machine learning pipelines.

Microsoft

SOFTWARE ENGINEER INTERN — Jun 2022 – Sep 2022

• Developed an anomaly detection framework using Kusto queries that iterates over 3.4 trillion Windows telemetry events.
• Created visualization dashboard for developers to analyze their feature's Avg/50th/90th percentile performance measures.
• Designed a Node app that allows PMs and managers to edit configurations for team self-host performance data.

Intheon

SOFTWARE DEVELOPER INTERN — Jan 2021 – Sep 2021

• Maintained an open-source visual pipeline designer for analyzing data from research-grade EEG hardware.
• Developed 3d-brain mapping, filtering, artifact removal, and machine learning algorithms using Numpy and Scipy libraries for biosignal processing and real-time analysis.

Publications

• **Song P**, Sarathy J, Shoemate M, and Vadhan S, "I inherently just trust that it works": Investigating Mental Models of Open-Source Libraries for Differential Privacy". *Proceedings of the ACM on Human-Computer Interaction CSCW*, Nov 2024
• Chen W, Kirkby L, Kotzev M, **Song P**, Gilron R and Pepin B, "The Role of Large-Scale Data Infrastructure in Developing Next-Generation Deep Brain Stimulation Therapies". *Front. Hum. Neurosci*, Sep 2021
• Prasad S, Rankine A, Prasad T, **Song P**, Dokukin M.E., Makarova N, Backman V and Sokolov I, "Atomic Force Microscopy Detects Difference in Cancer Cells of Different Aggressiveness via Machine Learning". *Adv. NanoBiomed Res,* Aug 2021

Skills

Languages and Technologies: Python, Pyspark, Airflow, SQL, Databricks, ElasticSearch, DynamoDB, Kinesis, Amazon S3