

Shusei Yokoi

(415) 439-3371 | shuseiyokoi@gmail.com | Los Angeles

<https://www.linkedin.com/in/shuseiyokoi/> | <https://github.com/shuseiyokoi> | <https://shuseiyokoi.notion.site/about-me>

EDUCATION

University of Southern California

Expected Graduation May 2027

Master of Science, Computer Science Concentration in Applied Data Science

California Polytechnic State University San Luis Obispo

June 2021

Bachelor of Science, Business Administration Concentration in Information Systems & Minor in Statistics

PROFESSIONAL EXPERIENCE

USC Information Sciences Institute (ISI)

2026 - Present

Research Assistant

Trustworthy AI Research

- Conducting research on **trustworthy AI**, focusing on fairness, bias, and reliability in **LLM-based decision-making**
- Designed controlled experiments to evaluate the effects of prompt framing on responses across **LLMs**, including **ChatGPT**, **Claude**, **Gemini**, and **Qwen**
- **Identified** early **evidence** that emotional prompt framing can alter LLM behavior, causing conclusion shifts in some models and reduced confidence in others

Datify

2025 - Present

Founder / AI Product Developer

- **Founded** Datify to **build AI-powered products** that help people improve their lives through **personalization**
- Developed HealthSync, an **iOS AI health** advisor app integrating **Apple HealthKit** and **Azure OpenAI GPT-4o** to deliver personalized health insights from biometric and activity data
- **Built** data pipelines to extract **biometric data** from Apple HealthKit and convert them into structured summaries for **AI** analysis to generate dynamic, contextual health advice, helping users track trends and stay on target with personal goals

SoftBank

2022 - 2024

Data Scientist

Gym Chain Health Data Analysis

- Conducted comprehensive data analysis in **Python**, **R**, and **SQL** on **200,000 gym member** records across **150+ branches** in Japan
- Identified a 3-month weight regain trend in younger members and developed tailored retention strategies
- Implemented a notification service to re-engage members showing signs of losing interest in attending the gym, increasing re-engagement rate by **30%**

Large Software Development Team Productivity Analysis

- Performed **Difference-in-Differences** analysis on ticketing system data to diagnose productivity bottlenecks
- Determined that low productivity was driven by resource allocation inefficiencies and recommended more frequent ticket creation and resource optimization, improving team productivity by **35%**

Technical Project Manager / Solutions Engineer

2022 - 2024

Led Software/Application Development

- **Delivered end-to-end technology solutions** for clients across multiple industries, owning solution design, stakeholder communication, and product outcomes
- Built office management solutions during COVID-19, including contactless employee health management systems and a metaverse office platform, increasing employee engagement by **40%**
- Translated business needs into scalable **AWS** and **Azure** solutions, led multinational **Agile** teams across **GitLab CI/CD** and multi-platform deployments, and served as a key liaison between business stakeholders and engineers

Data Scientist Contractor

Trade Area Analysis Population Flow Analysis

- Led trade area analysis for Izumi Co. (190+ malls) under SoftBank's **Smart City** project, using **GPS**, demographic, and **Yahoo!** search data to uncover customer trends
- Built **Tableau** dashboards with **demographic attributes** and **GPS** data from diverse sources for actionable insights
- Identified retention strengths and regional growth potential, recommending targeted ads, in-store enhancements, and loyalty programs that boosted retention and informed campaigns, leading to a new Smart City **trade area analysis** department

PROJECTS

RAG ChatBot [Ask Me Blog](#) | AWS, Python, TypeScript, LLM, RAG 2025

- Developed an interactive chatbot leveraging **RAG** with **AWS Bedrock** to provide detailed, career-specific responses based on user queries
- Integrated **Amazon Kendra** for efficient data retrieval from structured documents stored in **Amazon S3**, improving answer relevance
- Designed and deployed a scalable **front-end application** using **TypeScript** on **AWS Amplify**, offering an intuitive user experience

HealthSync – AI Health Advisor iOS App [TestFlight Blog](#) | Swift, Azure OpenAI, GPT-4o 2025

- Developed an **iOS app** that analyzes personal health data and generates **GPT**-based health advice, helping users track weight trends and stay on target with personalized goals
- Integrated **Apple HealthKit** to extract real-time **biometric** and **activity data** (weight, steps, calories, heart rate) over the past 6 months and convert it into structured summaries
- Connected to **Azure OpenAI**'s Chat Completion **API** to generate dynamic, contextual advice using **GPT-4o**

LLM Evaluation with Promptfoo and AWS Bedrock [GitHub Blog](#) | Python, AWS, Promptfoo 2025

- Created automated benchmarking of **AWS Bedrock Nova Pro** vs. **Nova Micro** for **RAG**
- Developed reproducible test suites for accuracy, completeness, and consistency

HONORS AND AWARDS

SoftBank

Outstanding New Employee Recognition

Award-Winning New Hire & Data Flow Integration Lead | Vantique, Python 2022

- Collaborated with **Vantique** to lead end-to-end testing of newly developed data flow services and oversaw their successful deployment to the production environment
- Evaluated the system and data to evaluate their effectiveness in solving business problems and achieving target **ROI**

Top Candidate at SoftBank AI Competition

Ranked 10th at Data Competition | R 2020

- Made population flow prediction AI with **MAPE 5.78%** and placed **10th place** out of 110+ participants
- Applied an **ARIMA** model with seasonality, incorporating COVID-19 effects and weather attributes to capture real-world dynamics

TECHNICAL SKILLS

Programming Languages: Python (scikit-learn, TensorFlow, PyTorch, etc), R (ggplot2, dplyr, tidyr, etc), SQL, JavaScript, Rust

Tools: AWS (Bedrock, Cloud Watch Knowledge Base, Kendra, etc), Azure (Azure AI Foundry), GCP, Railway, Netlify, Tableau, Power BI, Shiny

Machine Learning: Logistic Regression, XGBoost, ARIMA, SVM, Random Forest, K-means, KNN, Hierarchical Linear Model, Log Linear Model, LLM, RAG, Promptfoo

CERTIFICATIONS

[IBM Data Science Specialization](#)

[Deep Learning Specialization](#)

[MLOps Specialization](#)

[CompTIA Data+](#)

[AWS AI Practitioner](#)

[Perform Foundational Data, ML, and AI Tasks in](#)

[Google Cloud](#)

[Perform Predictive Data Analysis in BigQuery](#)