Shusei Yokoi

(415) 439-3371 | shuseiyokoi@gmail.com | Los Angeles, CA | GitHub | Portfolio | LinkedIn

Graduation: M.S., May 2027 | Available full time from June 2027

SUMMARY

Data scientist finishing an M.S. in Computer Science (Applied Data Science) at USC, with **3+ years of industry data science** at SoftBank. Defines **success metrics**, runs **experiments**, does **root cause analysis**, and uses **causal inference** (difference-in-differences) to **quantify the impact** of changes, then turns the result into **actionable insights** that product and engineering teams act on. **Python**, **SQL**, **R**; **forecasting**, classification and segmentation.

EDUCATION

University of Southern California, Los Angeles, CA May 2027 (expected)
M.S. in Computer Science (Applied Data Science) | *GPA 3.92* | *Expected May 2027*

California Polytechnic State University, San Luis Obispo, CA Jun 2021
B.S. in Business Administration (Information Systems), Minor in Statistics | *GPA 3.63*

EXPERIENCE

USC Information Sciences Institute (ISI) — Student Researcher, Trustworthy AI May 2026 - Present

- **Research experience in data science:** designed **controlled experiments** on 2024 California HMDA mortgage data testing whether prompt framing shifts LLM judgments (n=300 per condition, **chi-square/Fisher** tests); built the **Python** pipeline (**19,200 calls, 14 models**).
- **Root cause analysis:** found the framing effect depends on input format (emotional framing pushed one model from 0% to 100% "biased" verdicts on summary tables but removed every one on raw loan files), then ran a thinking-off ablation showing most of the drop in bias verdicts from Claude Haiku 4.5 to 5.5 came from model generation, not reasoning.
- **Defined the evaluation metrics:** designed ground-truth labels and scored verdicts with hit rate, false-alarm rate and accuracy vs an always-NO baseline; in a 935-call pilot, a stated Black vs White identity raised "discrimination" conclusions from 4.9% to 15.1% ($p < 0.001$).

Datify — Founder / AI Product Developer Apr 2025 - Present

- Shipped **HealthSync**, an iOS LLM health-advisor app (Swift, Azure OpenAI): **30+ TestFlight users**, **160+ sessions**.

SoftBank — Data Scientist & Technical Project Manager Apr 2022 - May 2024

- Set **retention** as the **success metric** with the client, then analyzed **200K+ gym-member records** (**SQL, Python, R**) and found a 3-month weight-regain pattern that signaled churn; built a **churn-risk notification service**, and member **re-engagement** rose **30%** after launch.
- Ran a **difference-in-differences** study on a 50-person engineering org's ticketing data; traced the bottleneck to resource allocation and drove changes that raised throughput by **35%**.
- Built an **XGBoost** classifier with **SHAP** on a hotel chain's account data and found two user types (business vs leisure) with different favorite hotels, so marketing could target each.
- Led a **30-engineer** multinational team as technical PM, the **cross-functional** link between client stakeholders and engineers; shipped a health-check and virtual-office platform that raised **daily active users 40%**.

SoftBank — Data Scientist (Contract, part-time) Apr 2021 - Mar 2022

- Analyzed **GPS** data from **~4M devices** for trade-area analysis of **190+ shopping malls**; built **Tableau** dashboards for **15 stakeholders**, and the recommendations led SoftBank to form a trade-area team.
- **Segmented** 3,000+ members (**K-means**) for an executive report that informed the launch of the Station AI hub.

ABC Cooking Studio — Data Scientist Intern Jul 2020 - Dec 2020

- Modeled e-commerce sales with **hierarchical mixed models** capturing category price elasticity; in **Google Analytics**, found that add-to-cart pushed shoppers to the cart page and many went back to keep browsing; the CEO backed it and the **user flow** was changed.

PROJECTS

News-to-Podcast Platform — led a 2-person team; ranked episodes per listener from listening history | Django Jul 2026

TECHNICAL SKILLS

Languages: Python (pandas, NumPy, scikit-learn, PyTorch), SQL, R (tidyverse)
Experimentation & Inference: A/B testing, experiment design, causal inference (difference-in-differences), statistical inference and hypothesis testing (chi-square, Fisher exact), regression, hierarchical models
Modeling: forecasting (ARIMA), classification (XGBoost, random forest, logistic regression), K-means segmentation
Tools: Tableau, Power BI, Spark (coursework), GCP (BigQuery), AWS, Azure, Google Analytics, Git/GitLab CI/CD; Claude Code

HONORS

Participants' Choice Award — USC AI4SE Hackathon Jul 2026
Outstanding New Employee Recognition — SoftBank Oct 2022
Top 10 of 110+ — SoftBank AI Competition (seasonal ARIMA forecaster, MAPE 5.78%) Aug 2020