

Samuel Kapusta

562-276-5016 | skapusta56@gmail.com | linkedin.com/in/samkapusta | github.com/sam-kapusta

EDUCATION

Bachelor of Science in Computer Science

Sep. 2018 – June 2022

University of California, San Diego

San Diego, CA

- GPA: 3.6/4.0

EXPERIENCE

ML Software Engineer II

Sept 2022 – Present

Amazon

San Diego, CA

- Trained, labeled, and benchmarked sparse autoencoder architectures that decompose customer behavior embeddings into interpretable audience segments (e.g. energy drink buyers); drove cross-org adoption of the winning model by Amazon Ads as the backbone of their enterprise advertiser product.
- Led a team of 4 to build the production pipeline: Step Functions orchestrating daily SageMaker inference, monitoring and alarming on delivery to Ads, and a Redshift store holding activations across 4,000 features for all 330M customers.
- Built the data layer behind Amazon's customer foundation model: a real-time stream unifying 50M+ daily behavioral events across orders, views, searches, and cart activity into per-customer sequences, vending embeddings to downstream payment risk, ads, and package consolidation systems.
- Rearchitected the Fire tablet home screen recommendation system from a pull-based poller to direct APIs to resolve latency, and launched 14 personalized and cold-start strategies with Audible, Books, Music, and Podcasts for the redesigned experience.

Data Science Intern

June 2021 – Sept 2021

Verizon

Basking Ridge, NJ

- Created Tableau dashboards using gateway metrics and geomapped data showing session data between various senders and receivers used for comparing and contrasting nationwide traffic.
- Automated Python scripts to pull gateway information hourly into PostgreSQL server for data analysis.

Software Engineering Intern

June 2020 – Jan 2021

UC San Diego, IT Services

San Diego, CA

- Built desktop application used by UC San Diego administrators to allow for grouping and querying of SQL database on secure university data.

PROJECTS

Chess Deck | *chess-deck.com* | *Python, PyTorch, React, Lambda, Fargate, SQS, DynamoDB, CDK* 2025 – Present

- LLM chess coach that explains why a player loses, not the best engine move: chess transformer Maia provides a human move prediction baseline at the player's rating, Stockfish the objective evaluation, and a 183-tag rule engine names the mistake mechanism (Missed Zwischenzug, Hung Piece) before the model writes.
- Builds a training curriculum targeting the gap to the next rating level: Maia's per-rating move distributions identify the mistakes players one band up have stopped making, so drills prioritize the skills that actually move a player forward rather than engine-only moves no human plays.
- Trained BatchTopK sparse autoencoders on the policy model's layer-7 activations and labeled 2,048 features via a two-pass Claude pipeline; at 44% crisp concepts, shipped a rule engine instead and kept the SAE as an audit tool.

SKILLS

Languages: Java, Python, TypeScript, Kotlin, SQL

ML: SageMaker, Bedrock, PyTorch, NumPy

Data: Spark, Redshift, DynamoDB, Glue, S3

AWS: CDK, CloudFormation, Lambda, ECS Fargate, Step Functions, SQS, SNS, API Gateway, CloudWatch

Testing & DI: JUnit, Mockito, Dagger, Guice, Lombok