Graham Soto

Jersey City, NJ · graham.soto@example.com · (201) 555-0139 · linkedin.com/in/example-graham-soto

My Key Qualifications

- **7 years deploying production ML systems**, unbroken from Alloy in 2019 through Marqeta today, all in fintech and payments.
- **Real-time fraud scoring ownership** at Marqeta: redesigned the scoring service and cut p95 latency from 180 ms to 92 ms.
- Improved fraud model **PR-AUC from 0.41 to 0.53** with XGBoost and expanded behavioral features for card-present and card-not-present screening.
- Full **MLOps stack hands-on**: MLflow, Airflow, Terraform, and CI/CD cut model release cycles from 14 days to 6.
- AWS production depth across **S3, EKS, ECS, Lambda, Kinesis, SageMaker, and CloudWatch** for training and serving infrastructure.
- Monitoring and governance in regulated environments: drift alerts and SLA escalation raised scoring uptime **from 99.3% to 99.9%**.
- **Mentors a team of 4 engineers** through design reviews and production-readiness checks, with 3 promotions over 2 review cycles.

Experience

Senior Machine Learning Engineer — **Marqeta**, card issuing and payments technology company

Aug 2022 – Present · 4 yrs 3 mos

- Led a redesign of the real-time fraud scoring service, cutting p95 latency from 180 ms to 92 ms by moving inference to a streamlined Python service on Kubernetes, tightening feature retrieval paths, and profiling request bottlenecks.
- Built a gradient-boosted fraud model for card-present and card-not-present transaction screening that improved PR-AUC from 0.41 to 0.53 in offline validation, expanding behavioral features, retraining XGBoost models, and formalizing calibration and threshold review with risk teams.
- Established model monitoring and incident-response standards for production scoring services, raising uptime from 99.3% to 99.9% with CloudWatch dashboards, drift alerts, on-call runbooks, and SLA-based escalation paths.
- Implemented reproducible training and deployment workflows across fraud models, cutting release cycle time from 14 days to 6 days by standardizing MLflow tracking, Airflow retraining jobs, Terraform-managed infrastructure, and CI/CD checks.
- Mentored a team of 4 engineers through code reviews and system design guidance, supporting 3 internal promotions over 2 review cycles.
- Partner with product, risk, platform engineering, and data science stakeholders on roadmaps and success metrics, and lead design reviews across feature engineering, model training, real-time inference, and monitoring.

Machine Learning Engineer — **Alloy**, an identity risk and fraud decisioning platform for financial institutions

Feb 2019 – Jul 2022 · 3 yrs 5 mos

- Built production ML systems for fraud detection, anomaly detection, and application risk scoring, including the model and feature pipelines behind fraud and risk decisioning.
- Developed an application fraud risk model for account onboarding workflows that reduced the false positive rate from 7.8% to 5.9%, combining device, velocity, and identity-consistency features in XGBoost with threshold tuning against analyst labels.
- Introduced unsupervised anomaly detection for emerging identity abuse patterns, surfacing 23% more high-risk cases for analyst review in the first quarter after launch.
- Established governed model documentation and validation artifacts for production releases, standardizing experiment records, dataset lineage notes, and release checklists in MLflow, and cut audit-prep turnaround from 5 business days to 2.
- Partnered with risk, product, and platform teams on model evaluation, governance, and deployment practices in a regulated environment, using Python, SQL, Docker, Kubernetes, Airflow, and AWS services including S3, Lambda, and CloudWatch.

Data Scientist — **AxiomSL**, regulatory reporting and risk management software provider for financial services

Mar 2016 – Jan 2019 · 2 yrs 10 mos

- Built a transaction anomaly scoring prototype for compliance investigations, engineering SQL-based features and training tree-based models on historical alert outcomes, reducing investigator review backlog by 12 cases per week in a pilot team.
- Automated recurring model input validation by writing Python checks embedded in scheduled batch workflows, cutting monthly data-quality defects from 9 to 3.
- Supported product and engineering teams on model evaluation and data-quality analysis across risk and compliance-focused financial software products.

Software Engineer — **TrueAccord**, a digital debt collection and recovery technology company

Apr 2015 – Feb 2016 · 10 mos

- Improved a batch account-prioritization service used by collections operations teams, cutting nightly processing time by 37 minutes by refactoring Python jobs and optimizing SQL queries against account event tables.
- Maintained backend APIs and batch processing jobs supporting digital collections operations, including data modeling and application reliability improvements.

Education

- **Rutgers University** — Bachelor of Science in Computer Science · 2010 – 2014

Skills

- **ML & Modeling:** Python, SQL, XGBoost, PyTorch, Feature Engineering, Anomaly Detection
- **MLOps & Infrastructure:** Kubernetes, Docker, Airflow, MLflow, Terraform, CI/CD
- **Cloud:** AWS (S3, EKS, ECS, Lambda, Kinesis, SageMaker, CloudWatch)
- **Domain:** Fraud Detection, Model Governance, Real-Time Inference