Caleb Tucker Dame

Data Scientist | Fraud, Financial Crimes & Risk Modeling
Saint George, UT | (707) 305-8453 | calebdame@gmail.com | linkedin.com/in/caleb-dame | github.com/calebdame | calebdame.com

SUMMARY

Data scientist with 5+ years of experience building fraud, financial crimes, and risk models for a Tier 1 U.S. bank, credit unions, utilities, and TransUnion's identity fraud products. Currently the primary technical contact for a bank's sanctions and adverse media screening models, where entity resolution work cut sanctions false positives by over 95%. Started in data engineering (Databricks, Airflow, Snowflake) and am completing an M.S. in Artificial Intelligence at Johns Hopkins, with project work in transformer architectures, graph ML, tabular foundation models, and retrieval-augmented generation (RAG).

EXPERIENCE

SymphonyAI | Remote
Data Scientist II | Aug 2026 – Present

- Serve as primary technical contact for a Tier 1 U.S. bank's AI-driven sanctions, PEP, and adverse media screening models within its financial crimes compliance (AML/KYC) program.
- Developed entity resolution for sanctions screening in transaction monitoring, reducing false-positive alerts by over 95%.
- Redesigned alert reason codes built on SHAP attributions so investigators and compliance reviewers receive clear, specific explanations of why each customer or transaction was flagged.
- Write model documentation and present and defend modeling methodology to the bank's Model Risk Management team during independent model validation (SR 11-7).

BlastPoint | Remote
Senior Data Scientist | Dec 2025 – Present
Data Scientist | Oct 2023 – Nov 2025

- Built customer segmentation and predictive models for credit unions and utilities covering churn, collections prioritization, product propensity, program adoption, and energy usage forecasting, used by client teams to target outreach and retention campaigns.
- Developed federated, privacy-preserving models that pool learnings across clients into industry-wide models without exposing any client's underlying data.
- Built internal AutoML and Python tooling that standardized the team's feature engineering, modeling workflow, model performance monitoring, and SHAP-based explanations for end users, increasing project throughput at a growing startup.

Modulus Partners | Remote
Machine Learning Consultant | June 2023 – Jan 2026

- Built ML lead-scoring models for sales and marketing clients and ran ongoing A/B tests on lead qualification, removing 20–40% of low-quality leads and raising conversion rates for top salespeople 2–3x.
- Deployed and maintained cloud-hosted classifiers and connected them to client CRMs (HubSpot, Close, Zapier) through custom API integrations.

TransUnion | Remote
Data Scientist | Feb 2022 – Oct 2023

- Built and shipped identity fraud and IP address risk models (LightGBM, XGBoost) for the TruValidate fraud prevention product suite, working with highly imbalanced fraud labels.
- Maintained internal Python packages that tracked drift in fraud typologies with long label-maturity windows and estimated changes in fraud prevalence in unlabeled data.
- Built identity graphs from identity databases for feature engineering and entity linking, and benchmarked graph neural networks, autoencoders, and stacked models for detecting synthetic identities and fraud rings.
- Led cross-team work to integrate data from newly acquired subsidiaries, adding features that improved model performance.

M Science | Murray, UT
Data Engineer | Jun 2021 – Mar 2022

- Built and maintained 15+ ETL pipelines in Databricks (PySpark), Airflow, and Snowflake, processing 3.5 TB of data daily into datasets for financial analysts whose research was sold to investment clients.
- Automated anomaly detection using clustering on NLP-derived text features, improving data purity by 10–15% and feeding real-time Tableau dashboards.

Brigham Young University, Department of Economics | Provo, UT
Machine Learning Researcher | Apr 2020 – Jul 2020

- Reduced census record-linkage (entity resolution) error by 80% by engineering features that treated census records as a graph of family and location relationships, modeled with XGBoost, neural networks, and Naive Bayes on MS SQL Server data.

PROJECTS

Music Theory Priors in Transformer Attention | Johns Hopkins, Deep Learning Using Transformers, 2026
calebdame.com/static/pdf/music_transformer_2026.pdf

- Trained a 23M-parameter decoder-only transformer in PyTorch on 306K MIDI pieces (450M+ notes) from GigaMIDI, adding learned self-attention biases for harmonic (circle of fifths) and temporal (beat distance) relationships at under 0.01% added parameters.
- Across 16 runs (4 model variants × 4 data sizes) tracked in Weights & Biases, the temporal bias cut test perplexity 3.0% and reached the baseline's best validation loss in 46% fewer training steps (80% fewer on a 20% data subset). The learned bias was interpretable, favoring notes within about 2 beats of each other.

HOA Covenant Document Q&A (RAG)

- Built a containerized (Docker) pipeline that runs OCR on large PDF covenant documents, splits them into semantic chunks, indexes embeddings in FAISS, and answers questions with an LLM grounded in the retrieved passages.

Tabular Foundation Models vs. Boosted Trees for Intrusion Detection

- Benchmarked the TabICLv2 tabular foundation model against boosted-tree baselines on a hardened version of the CIC-IDS2017 intrusion detection dataset; the attention-based model outperformed the trees, a result relevant to fraud and security problems long dominated by tree

models.

EDUCATION

Johns Hopkins University | M.S., Artificial Intelligence | Expected May 2027
Coursework: Deep Learning, Deep Learning Using Transformers, Reinforcement Learning

Brigham Young University | B.S., Applied and Computational Mathematics & Economics | Apr 2021
Dean's List (2018, 2020)

SKILLS

Languages: Python, SQL, PySpark, JavaScript, Bash, HTML, Stata

Machine Learning: scikit-learn, XGBoost, LightGBM, PyTorch, TensorFlow, Keras, statsmodels, SHAP, supervised and unsupervised learning, imbalanced classification, feature engineering, graph neural networks, autoencoders, federated learning, anomaly detection, entity resolution, forecasting, drift monitoring

LLMs, NLP & Deep Learning: transformers (decoder-only, self-attention), Hugging Face Transformers, large language models (LLMs), retrieval-augmented generation (RAG), tool calling, Model Context Protocol (MCP), FAISS, embeddings, OCR, NLTK

AI Coding Tools: Claude Code, Codex, OpenCode, Cursor

Data & Cloud: Databricks, Apache Spark, Snowflake, Airflow, PostgreSQL, Oracle, MS SQL Server, MongoDB, SQLite, AWS (S3, EC2, Lambda, SageMaker, DynamoDB, Redshift), GCP

MLOps & Tooling: MLflow, Weights & Biases, Docker, FastAPI, Flask, Git, Conda, cron, model monitoring

Analytics: A/B testing, experiment design, customer segmentation, statistical modeling, Tableau, Grafana, Matplotlib, Seaborn

Domain: anti-money laundering (AML), know your customer (KYC), transaction monitoring, sanctions and politically exposed person (PEP) screening, adverse media screening, false-positive reduction, identity and synthetic identity fraud, model risk management (SR 11-7), model documentation and validation, explainable AI (XAI), privacy-preserving ML