

Jai Dhiman

San Francisco, CA | +1 (707) 583-6129 | jai.dhiman@outlook.com

www.linkedin.com/in/jai-d/ | www.github.com/Jai-Dhiman | www.jaidhiman.com

SKILLS

Core ML Stack: JAX/Flax, PyTorch, Hugging Face (Transformers, PEFT, TRL), NumPy, Pandas, scikit-learn, XGBoost

Audio ML Specialized: MERT, Wav2Vec2, AST, CQT/mel-spectrograms, librosa, nnAudio, music info retrieval

Languages: Python (expert), Rust (production deployment), TypeScript, Golang, SQL

Infrastructure & MLOps: Docker, Modal, GCP, AWS, Redis, PostgreSQL, Git/CI/CD, Weights & Biases

Techniques: Multi-modal fusion, hierarchical modeling, active learning, ensemble methods, DPO/LoRA

EXPERIENCE & PROJECTS

Piano Performance Analysis System (github.com/Jai-Dhiman/crescendai) | September 2025 - Present

- Designed hierarchical multi-modal architecture using MERT-330M as audio backbone, fused with MusicBERT symbolic encoding via cross-attention, achieving 21% performance gain over audio-only baseline.
- Built a comprehensive evaluation framework spanning 16 performance dimensions with automated benchmarking pipeline: technical metrics (pitch, rhythm, dynamics, articulation) achieved $r=0.76$ correlation with expert consensus, representing 8-19% improvement over prior SOTA and exceeding single expert ICC reliability.
- Implemented hierarchical temporal aggregation (note→phrase→piece) using bidirectional LSTMs + Music Transformer relative attention, enabling processing of 2-10 minute performances with $O(LD)$ rather than $O(L^2D)$ complexity while maintaining temporal coherence across thousands of time steps.

Constitutional AI Implementation (github.com/Jai-Dhiman/ml-learning) | August 2025 - September 2025

- Implemented four-stage Constitutional AI pipeline in JAX/Flax: safety classifier training, supervised fine-tuning, critique-revision generation, and Direct Preference Optimization (DPO) as alternative to PPO alignment.
- Trained transformer safety classifier from scratch achieving 85%+ accuracy across four safety categories (0.89 AUROC); applied LoRA-based parameter-efficient fine-tuning to Gemma 2B-IT, reducing harmful outputs 23%.

Founding Engineer - Capture (Privacy Focused Social Platform Startup) | December 2024 - August 2025

Built ML-powered social platform from 0→50 beta users as sole engineer, later scaled to 4-person team.

- Architected and deployed recommendation engine combining collaborative filtering + content-based ranking, achieving 0.68 Recall@5 on cold-start users through systematic A/B testing and offline evaluation framework; drove 40% engagement increase over chronological baseline via data-driven feature prioritization.
- Designed distributed systems infrastructure: Redis-based rate limiting + circuit breaker patterns handling 10K+ requests/min with <50ms p99 latency; implemented comprehensive monitoring with custom metrics tracking.

Software Engineering Intern - Southern Glazer's Wine & Spirits | May 2025 - July 2025

- Automated deployments with GitHub Actions + Docker, cutting deployment time 2 hrs → 15 min and eliminating recurring config errors; collaborated with DevOps team on internal tools at a \$26B enterprise.

EDUCATION & COMMUNITY

Actualize Coding Bootcamp | June 2024 - December 2024

- Full-stack development + CS fundamentals (algorithms, data structures, distributed systems design).
- Self-directed ML/AI: Karpathy Neural Networks, Fast.ai Deep Learning, AI Safety Fundamentals (BlueDot).

Bachelor of Music in Performance (2x Dean's List) - Berklee College of Music | May 2020 - December 2023

- Developed discipline, creativity, and collaborative problem-solving through rigorous performance training.
- Performed in 2 international orchestra tours, honing high-pressure execution and teamwork skills.

Hackathons & Community

- 3-time Hackathon winner (Meta Executorch, Man vs Machine, Inception Buildathon).
- Active in AI/ML open-source (LangChain, JAX/Flax, Hugging Face); regular attendee at SF AI/ML dev meetups.

Jai Dhiman

San Francisco, CA | +1 (707) 583-6129 | jai.dhiman@outlook.com
www.linkedin.com/in/jai-d | www.github.com/Jai-Dhiman | www.jaidhiman.com

Founding engineer who builds production ML systems rapidly and explains them clearly. Proven ability to go from concept → deployed demo in days through 3 hackathon wins and 0→50 user product launch. Combines deep ML expertise (Constitutional AI, multi-modal systems, recommendation engines) with customer-facing communication skills honed through music performance background.

EXPERIENCES & PROJECTS

Founding Engineer - Capture (<https://www.captureapp.org>) | December 2024 - August 2025

- Built and launched platform from 0→50 beta users as sole engineer, later scaling to 4-person team.
- Delivered recommendation engine in 6 weeks from conception → production deployment, enabling product launch and driving 40% engagement increase through iterative A/B testing.
- Built and demoed privacy-preserving architecture to potential investors and users, explaining complex technical trade-offs (decentralized data + centralized ML) in accessible terms.

Piano Performance Analysis System (<https://crescend.ai>) | September 2025 - Present

- Built AI-powered feedback system for piano students in underserved markets lacking access to private teachers, validated need through 12 educators and 20+ student interviews.
- Designed interface prioritizing actionable guidance ("crescendo more at phrase ending") over technical metrics, demonstrating product sense and user empathy.

Constitutional AI Implementation (github.com/Jai-Dhiman/ml-learning) | August 2025 - September 2025

- Implemented complete Constitutional AI pipeline from Anthropic papers in 4 weeks, including safety classifier, fine-tuning, critique generation, and DPO alignment, demonstrating 85%+ safety accuracy after fine-tuning.

Intern - Southern Glazer's Wine & Spirits (<https://southernglazers.com>) | May 2025 - July 2025

- Automated deployments at \$26B enterprise, reducing time from 2 hours → 15 minutes through GitHub Actions and Docker pipeline, eliminating recurring config errors; collaborating with DevSecOps teams on internal tooling.

SKILLS

Rapid Prototyping: POC development, demo environments, customer integrations, technical presentations

ML/AI Expertise: PyTorch, JAX/Flax, Hugging Face (Transformers, PEFT, TRL), LLMs, Constitutional AI/RLHF/DPO, RAG systems, recommendation engines, multi-modal models, music AI (MERT, Wav2Vec2, AST)

Infrastructure & Integration: Docker, Modal, GCP, AWS, Redis, PostgreSQL, APIs, Git/CI-CD, Weights & Biases

Languages: Python (expert), SQL (expert), TypeScript, Rust (production deployments), Golang

Customer-Facing: Technical demos, solution scoping, user research, stakeholder communication

EDUCATION & COMMUNITY

Actualize Coding Bootcamp | June 2024 - December 2024

- Full-stack development + CS fundamentals, with self-directed ML/AI specialization (Fast.ai, AI Safety).

Bachelor of Music in Performance (5x Dean's List) - Berklee College of Music | May 2020 - December 2023

- Developed communication and presentation skills through 100+ performances and 2 international tours requiring high-pressure execution and cross-cultural collaboration.
- Created multilingual practice materials for 100+ students (Simply Strings Program volunteer), demonstrating resourcefulness and user-centered design under constraints.

Hackathons & Technical Communication | 2024 - Present

- Won 3 hackathons (Meta Executorch, Man vs Machine, Inception Buildathon) building and presenting working demos under time pressure to technical and non-technical judges, average build time 24-48 hours.
- Regular attendee at AI community events, comfortable discussing technical topics with varied audiences.

Jai Dhiman

San Francisco, CA | +1 (707) 583-6129 | jai.dhiman@outlook.com
www.linkedin.com/in/jai-d | www.github.com/Jai-Dhiman | www.jaidhiman.com

Data science professional with 2+ years of experience building ML-powered systems, ETL pipelines, and predictive models. Proven track record architecting data infrastructure from scratch, implementing automated workflows, and translating complex analytics into actionable insights for non-technical stakeholders. Strong background in cloud environments (GCP, AWS), data modeling, and cross-functional collaboration.

EXPERIENCES & PROJECTS

Founding Engineer - Capture (<https://www.captureapp.org>) | December 2024 - August 2025

- Developed and deployed recommendation engine combining collaborative filtering with content-based ranking, achieving 0.68 Recall@5 on cold-start users; drove 40% engagement increase through systematic A/B testing and data-driven feature prioritization.
- Designed comprehensive metrics framework tracking 50+ KPIs across user engagement, content quality, and system performance; built automated reporting dashboards for stakeholder insights.
- Built automated ETL processes for extracting user interaction data, transforming it into feature representations, and loading into recommendation models; reduced data processing time by 65%.
- Designed distributed systems infrastructure on cloud platforms (GCP) with Redis-based caching, rate limiting, and circuit breaker patterns handling 10K+ requests/min with <50ms p99 latency.

Piano Performance Analysis System (<https://crescend.ai>) | September 2025 - Present

- Designed hierarchical multi-modal ML architecture processing audio and symbolic music data, achieving 21% performance gain over baseline through systematic experimentation and evaluation.
- Built comprehensive evaluation framework spanning 16 performance dimensions with automated benchmarking pipeline; achieved $r=0.76$ correlation with expert consensus through rigorous statistical validation.
- Processed and analyzed large-scale audio datasets, implementing efficient data pipelines for feature extraction and model training on 2-10 minute audio files.

Constitutional AI Implementation (github.com/Jai-Dhiman/ml-learning) | August 2025 - September 2025

- Implemented four-stage ML pipeline: safety classifier training, supervised fine-tuning, critique generation, and preference optimization; achieved 85%+ classification accuracy across safety categories (0.89 AUROC).

Intern - Southern Glazer's Wine & Spirits (<https://southernglazers.com>) | May 2024 - July 2024

- Automated deployment pipelines using GitHub Actions and Docker, reducing deployment time from 2 hours to 15 minutes and eliminating recurring configuration errors at \$26B enterprise.

SKILLS

ML & Analytics: PyTorch, JAX/Flax, scikit-learn, XGBoost, NumPy, Pandas, statistical modeling, regression analysis

Data Engineering: ETL pipeline design, PostgreSQL, data extraction/transformation, data modeling, Redis

Cloud & Infrastructure: GCP, AWS, Docker, Modal, automated deployment, CI/CD pipelines, Git

Languages: Python (expert), SQL, TypeScript, GoLang, Rust

Techniques: Predictive modeling, A/B testing, collaborative filtering, ensemble methods, feature engineering

EDUCATION & COMMUNITY

- 3-time hackathon winner (Meta Executorch, Man vs Machine, Inception Buildathon).
- Regular attendee at San Francisco AI/ML developer meetups and technical community events.

Actualize Coding Bootcamp | June 2024 - December 2024

- Intensive program covering algorithms, data structures, distributed systems, and software engineering.

Bachelor of Music in Performance (5x Dean's List) - Berklee College of Music | May 2020 - December 2023

- Developed analytical rigor, discipline, and collaborative problem-solving through performance training.