

Taylor Kang Beck

+1 702 738 3599

t@ylor.cc

taylor.id

taylorkangbeck

taylorkangbeck

Versatile machine learning engineer with 8+ years of experience pushing the boundaries of AI systems. Skilled at transforming cutting-edge research into production-ready systems across domains including language, audio, computer vision, and robotics.

Experience

Cofounder, CTO *Of Counsel AI*

07/2023 - 09/2024

Legal AI Copilot, LLM Systems & Evaluation, RAG & Knowledge Graphs, Prompt Engineering

- Architected and deployed a production-scale RAG system combining graph databases and LLMs for legal document drafting and analysis
- Developed a general-purpose agent system with distributed workflow execution, handling complex multi-step document processing tasks, increasing batch processing speed by 80%
- Implemented comprehensive LLM evaluation framework with custom datasets, enabling data-driven model selection and 25% quality improvement
- Engineered an intelligent copilot system integrating chat, document editing, and UI interaction capabilities
- Designed and deployed distributed microservices architecture supporting real-time collaboration and multi-tenant access
- Built enterprise-grade document editor with real-time collaboration using CRDTs
- Selected for Madrona Venture Labs' Launchable startup competition

Tech Stack AWS, Neo4J, Postgres, Temporal, Docker, Python, LangGraph, Typescript, Next.js

Founder *Weaver.fm*

09/2022 - 04/2023

Social Audio Speech Transcription, Audio Processing, LLMs & Embedding Models, On-Device Inference

- Developed and deployed a custom ASR pipeline combining Whisper and forced alignment for precise audio transcription, reducing errors by 40%
- Deployed efficient self-hosted embedding infrastructure for content processing and recommendations
- Engineered intelligent conversation facilitation system using LLM-generated contextual prompts, increasing average conversation length 3x
- Designed, implemented, and deployed an iOS app with iterative product improvements driven by telemetry data and user interviews

Tech Stack AWS, Postgres, Python, FastAPI, Typescript, React Native, ggml

SDE, Robotics Simulation *Amazon Scout*

07/2020 - 08/2022

Autonomous Delivery Distributed Systems, GPU Acceleration, Dataset Generation, Multimodal ML

- Led development of GPU-accelerated cloud simulation platform, handling 5,000+ photorealistic physics-based tests daily
- Optimized multi-region autoscaling to increase GPU utilization by 25% and decrease job queueing times by 50%
- Organized and coordinated cross-functional initiatives across multiple teams
- Engineered automated pipeline for generating 1M+ synthetic sensor training samples monthly, enhancing navigation and perception models
- Implemented multimodal autoencoder and clustering system to enhance incident identification and classification, resulting in discovery of new failure modes
- Enabled on-demand, GPU-accelerated environments and tools for research scientists
- Architected software distribution system supporting 1,000+ physical robot deployments weekly, with rollback capabilities

Tech Stack AWS, ROS, Python, PyTorch, C++, Nvidia Falcor, Vulkan, CUDA, Typescript, React

Generative AI Avatars **Voice Cloning, Computer Vision, Natural Language Processing**

- Led training and deployment of machine learning models across the entire engineering organization
- Developed a novel alternative to batch normalization for computer vision models, achieving faster convergence without decreasing accuracy during training
- Trained, deployed, and maintained a high-performance voice cloning pipeline
- Conducted ablation studies and surveys of embedding models and implemented a semantic similarity service, boosting response accuracy by 30%
- Implemented real-time systems for facial landmark detection and emotion recognition
- Optimized 3D animation calculations, reducing memory and compute requirements by 50%

Tech Stack Python, PyTorch, Kafka, Scala, AWS, Google Cloud, Typescript, C#, Unity3D, Docker, Websockets, WebRTC, Swift, React

SDE, Machine Learning Services *Amazon*

08/2016 - 01/2019

Alexa Shopping **Language Model Training and Inference, MLOps**

- Led development of language model infrastructure as founding engineer on Alexa Shopping ML team, serving 40M+ monthly active users
- Developed and contributed reusable neural network modules to internal PyTorch model development library
- Facilitated effective collaboration and communication between engineering and research science teams
- Built and deployed distributed machine learning inference services achieving 99.99% uptime and 50ms average latency across 10M+ daily requests
- Developed automated model deployment pipeline reducing required developer time from 2 weeks to 2 days
- Designed key components of data annotation and human-in-the-loop model evaluation tools

Tech Stack Python, PyTorch, Pandas, Numpy, Java, RabbitMQ, CuBLAS

Education**Bachelor of Science, Computer Science** *Vanderbilt University*

05/2012

- Minor: Engineering Management
- Artificial Intelligence I & II, Machine Learning, Neuroscience, Big Data, Human-Computer Interaction

Other coursework

- Machine Learning University at Amazon (internal training program)
 - 12 six-week classes including Probabilistic Graphical Models, Deep Learning NLP, CNNs, RNNs
- Statistics, Research Design in Data Science Masters Program at UC Berkeley

Additional Projects*Independent Consulting*

2022 - 2025

- Designed, built, and delivered AI prototypes to clients including: automated product listing generation from voice descriptions, internal Slack Q&A agent

Publications

- Mean Shift Rejection: Training Deep Neural Networks Without Minibatch Statistics or Normalization [🔗](#)

Hobbies

- Electronic music production
- Basketball
- Soccer
- Skateboarding
- Snowboarding

Leadership*AECJ*

2020 - 2025

- Climate nonprofit board of directors, VP
- Team lead: Tech, Recruiting

RVU Records

2014 - 2016

- Recording studio, Director