

CONNIE YANG

contactconnieyang@gmail.com ◇ Distinguished AI & Machine Learning Leader ◇ San Francisco, CA ◇ 484.320.0672

CORE EXPERTISE

- **Generative AI & Agentic Frameworks:** Built AI assistants using LangChain, AutoGen, and Semantic Kernel integrated with enterprise systems (Azure, Vertex AI, Agentspace).
- **LLMOps & AI Infrastructure:** Deployed LLM inference, vector search, and guardrail systems on Azure, GCP, and AWS; optimized latency, throughput, and cost for production workloads.
- **Experimentation & Evaluation:** Designed A/B testing, uplift modeling, and structured evaluation pipelines for model safety, performance, and business impact.
- **Enterprise AI Strategy:** Partnered with VP-level stakeholders to set AI roadmaps, align experimentation with growth strategy, and deploy AI systems with measurable ROI.

PROFESSIONAL EXPERIENCE

3Cloud (Microsoft Solutions Partner)

San Francisco, CA

Director of Machine Learning and AI

Mar 2024 – Present

- Lead GenAI & Insights delivery for LinkedIn's Workplace division, deploying LLM-powered assistants, AI employees and agentic workflows across 5+ stakeholder teams enterprise-wide.
- Expanded LinkedIn engagement from \$5M to \$8M through successful delivery and scale-up of AI initiatives.
- Architected LLMOps pipelines with Azure OpenAI, Azure Functions, and Databricks; deployed production-grade assistants with vector search, guardrails (Nemo, Semantic Kernel), and governance.
- Implemented multi-agent orchestration workflows for enterprise insights and automation, reducing research and reporting time by 40%.
- Scaled and led an 8-person ML/AI engineering team; established best practices in experimentation, observability, and responsible AI deployment.

DesignMind

San Francisco, CA

Vice President of Machine Learning and AI

May 2023 – Mar 2024

- Founded and scaled the AI/ML practice; established model deployment pipelines, experimentation frameworks, and responsible AI standards across client engagements.
- Delivered GenAI copilots for LinkedIn using LangChain, Azure AI Foundry, Autogen and Semantic Kernel, accelerating adoption of LLM-powered assistants and contributing to DesignMind's acquisition by 3Cloud.
- Led design and delivery of two PoC-to-production GenAI applications for KLA: (1) custom text-to-SQL RAG engine, and (2) RAG-based intelligent contract analyzer; doubled contract value through demonstrated impact.

Passes

Miami, FL

Founding 1st Employee, Head of Operations & Technology

May 2022 – May 2023

- Founding team member at venture-backed startup with seed funding led by Multicoin Capital and participation from David Sacks, Paris Hilton's 11:11 Fund, and other top VCs.
- Built and scaled core operations from seed stage to a successful go-to-market and eventual \$40M Series A, hiring and managing teams across engineering, legal, finance, HR, and go-to-market.
- Owned GTM strategy, product, and marketing execution; drove exponential early growth with \$700K+ GMV in first 3 months post-launch.
- Developed customer onboarding and retention strategy, increasing activation and lifetime value across multiple verticals.
- Designed initial success metrics, dashboards, and database schema to guide decision-making and scale operations.

Pallet

New York, NY

Head of Data Science & 1st Employee

Jul 2021 – May 2022

- Built and deployed recommendation systems and real-time analytics pipelines at scale, boosting customer conversion by 30% and average order value by 12%.
- Designed and productionized a transformer-based multi-label classification pipeline, increasing data-labeling throughput 10x and reducing outsourcing costs by 50%+.

Microsoft

Redmond, WA

Data Scientist I-II

Sep 2017 – Jul 2021

- Developed predictive models and causal inference pipelines improving OS build quality and release confidence.
- Built ML platforms serving 100+ engineering teams; led adoption of experimentation for product decisions.
- Created forecasting and usage models for Xbox/HoloLens with \$150M+ revenue impact.
- Collaborated with Microsoft Research and product teams across Office, Windows, Azure, and Xbox on multimodal learning projects for HoloLens, bridging research prototypes into production systems.
- Filed two patents on ML-driven insights platforms; presented at international conferences on AI system reliability.
- Founded and scaled Microsoft's Women in Data Science community to 1,000+ members, organizing technical workshops and research presentations.

KEY QUALIFICATIONS

- **10+ Years AI/ML Experience:** Microsoft, LinkedIn, and enterprise-scale GenAI deployments.
- **LLMOps:** Hands-on with Vertex AI, Azure ML, SageMaker, and custom pipelines for training, inference, and monitoring.
- **Agentic Frameworks:** 2+ years experience with LangChain, AutoGen, Semantic Kernel, CrewAI; deployed multi-agent systems.
- **AI Safety & Guardrails:** Implemented structured evaluation, content filters, memory, and retrieval-based grounding.
- **Cross-Cloud Infrastructure:** Deployed scalable AI systems on Azure, GCP, and AWS; optimized cost and performance.
- **Leadership:** Led multi-team AI initiatives; advised VP/C-level on strategy and technical roadmaps.

PUBLICATIONS & PATENTS

- **Publication:** "Comparing Insights From Inductive Qualitative Analysis Versus Automated NLP Algorithms For Analyzing Feedback In Digital Randomized Controlled Trials."
- **Patent (1st Author):** Feature deployment readiness prediction.
- **Patent (1st Author):** Propensity score matching for metrics estimation in unsampled computing devices.

EDUCATION

Carnegie Mellon University

Pittsburgh, PA

B.S. Computational & Applied Mathematics and Computer Science

2013 – 2017

– Undertook accelerated graduate-level coursework in linear algebra, partial differential equations, and early theoretical work in machine learning (pre-transformers) before joining Microsoft.

University of Washington

Seattle, WA

Board Member, MS in Data Science Industry Advisory Board

2020 – Present