

Zhanlin(Kevin) Liu

APPLIED SCIENTIST/ENGINEER MANAGER · GREEN CARD HOLDER

14823 NE 64th St, Redmond, WA, 98052

☎ (206) 422-3451 | ✉ liuzhanliniowa@gmail.com | 🌐 <https://zhanlin-site.onrender.com/>

AI Scientist and Engineering Leader with 7+ years of experience building and scaling AI products across large language models, knowledge graphs, multimodal AI, retrieval systems, and deep learning applications. Proven track record leading cross-functional AI teams and driving end-to-end development from research and model training to production deployment and evaluation. Published author of 20+ research papers spanning NLP, audio intelligence, multimodal learning, and machine learning systems.

Education

University of Washington

PHD IN INDUSTRIAL & SYSTEM ENGINEERING
M.SC IN STATISTICS

Seattle, WA

September 2014 - August 2020

September 2014 - June 2016

University of Iowa

BS IN STATISTICS WITH DISTINCTION

Iowa City, IA

August 2012 - May 2014

Industrial Experience

ProRata.ai

Bellevue, WA

PRINCIPAL AI ENGINEERING

October 2024 – Present

- Serve as a technical lead across AI research and engineering, driving the architecture, development, and productionization of large-scale agentic AI, retrieval, evaluation, Knowledge Graph, and multimodal AI systems.
- **PACE – Publisher-Adaptive Content Extractor:** Proposed and developed **PACE**, an agentic human-in-the-loop framework that dynamically adapts extraction strategies to publisher-specific content structures, improving extraction quality while reducing crawling and processing costs across **1,500+ publishers**.
- **Knowledge Graph:** Architected a large-scale Knowledge Graph platform and designed entity and relation taxonomies spanning **people, organizations, places, events, works, products, science, and knowledge**. Built a Wikidata-derived entity index by cleaning and normalizing **100M+ raw entities into approximately 34M canonical entities**, and constructed intrinsic and topical entity relationships to support retrieval, reasoning, and entity discovery.
- **Entity Extraction, Linking, and Expansion:** Developed scalable entity extraction pipelines over publisher content, linking extracted entities against the canonical Knowledge Graph and expanding entity coverage to **100M+ entities and entity mentions**. Productionized entity services for **Gist Answers, search, and Generative Engine Optimization (GEO)**, enabling entity-aware retrieval, contextual reasoning, and content discovery.
- **Discovery Questions / Gist Widget:** Built LLM-based systems to generate contextual discovery questions from publisher content and deployed the Gist discovery experience across **30+ publishers**, reaching **20M+ monthly widget views**, approximately **0.5% engagement**, and generating **65K+ monthly Gist queries**.
- **Discovery Question Evaluation:** Developed a continuous quality-improvement framework combining **LLM-based evaluation, human-in-the-loop validation, and online CTR feedback** to measure and improve generated-question relevance, usefulness, and engagement.
- **EPOCH – Agentic System Optimization:** Architected and developed **EPOCH**, an agentic evaluation-driven optimization framework that establishes task-specific baselines, automatically modifies prompts, rules, configurations, and code, executes evaluations, and iteratively retains improvements. Applied EPOCH to production tasks including **query understanding and discovery-question generation** and contributed to the associated research publication.
- **Gist Answer and LLM Evaluation:** Designed **LLM-as-a-Judge and human-in-the-loop evaluation frameworks** for complex generative AI systems. Developed binary-preference and multiple-choice annotation tasks to validate automated evaluation against human judgments, and used human agreement, disagreement, and confidence signals to improve evaluator robustness, calibration, and reliability across text denoising, chunking, question generation, and answer generation.
- **Music Attribution:** Developed a retrieval-based music attribution system for identifying reference music associated with generative-AI outputs using music-industry reference catalogs. Improved retrieval quality while optimizing the system to reduce inference latency and computational cost.
- **Multimodal Retrieval and Attribution:** Led multimodal retrieval and attribution initiatives spanning **text, images, audio, and video**, developing systems for multimodal content understanding, retrieval, matching, provenance, and source attribution.

AstrumU

Bellevue, WA

LEAD APPLIED DATA SCIENTIST, SKILL-BASED EDUCATION & HIRING

March 2023 – October 2024

- Led and scaled a multidisciplinary AI team from **2 to 7+ members**, owning the end-to-end development of large-scale AI/ML systems spanning data annotation, model training, evaluation, deployment, and production inference.
- **SkillLLM**: Led the development of **SkillLLM**, fine-tuning Large Language Models for contextual skill extraction from job descriptions and professional profiles and achieving state-of-the-art performance on skill identification and extraction.
- **Large-Scale Skill Extraction**: Architected and productionized large-scale extraction pipelines processing **millions of job descriptions and professional profiles**, transforming unstructured workforce data into normalized and structured skill representations.
- **Skill Taxonomy**: Architected a comprehensive skill taxonomy spanning technical and soft skills. Built automated pipelines for skill normalization, deduplication, linking, and taxonomy expansion using extracted workforce data and semantic relationships.
- **Skill Matching and Recommendation**: Built retrieval- and embedding-based skill matching systems for skill normalization, disambiguation, deduplication, and semantic matching, powering **personalized job recommendations, skill-gap analysis, career intelligence, and workforce analytics**.
- **AI Career Intelligence Platform**: Integrated SkillLLM, large-scale skill extraction, taxonomy expansion, and semantic skill matching into end-to-end production systems for AI-powered career intelligence and recommendation.

Microsoft Research

Redmond, WA

APPLIED SCIENTIST II, NLP & LARGE LANGUAGE MODELS

August 2022 – March 2023

- **Science Engine / Antigen Language Model**: Trained large-scale BERT-based antigen/protein language models on **2B+ multi-species antigen sequences** using distributed multi-GPU training, improving perplexity from **1.7 to 1.4** and outperforming the state-of-the-art **AntiBERTy** model.
- **Antigen Naturalness Scoring**: Introduced naturalness-scoring methods to identify unnatural antigen sequences and support the identification of higher-quality biological candidates.
- **Scientific Entity Extraction**: Developed transformer-based entity extraction systems for scientific papers and patents and benchmarked transformer fine-tuning against **GPT-2 and GPT-3 few-shot learning** across multiple NLP tasks.
- **GPT Document Intelligence**: Built GPT-based document intelligence pipelines for summarization, entity-relation extraction, question answering, and document generation, including fine-tuning GPT-3 on **European Medicines Agency** datasets for domain-specific applications.

Warner Music Group

Remote / Redmond, WA

DATA SCIENTIST, ENTITY MATCHING & MUSIC AI

September 2021 – August 2022

- **Music Copyright Identification and Entity Matching**: Enhanced large-scale music copyright identification and metadata reconciliation systems by combining traditional string-matching approaches with **deep learning-based entity embeddings** to resolve noisy artist, recording, and catalog metadata.
- **Open-Set Music Recognition**: Led research on music genre classification and trend analysis, developing **open-set recognition models** capable of identifying emerging and previously unseen music genres outside a predefined taxonomy.
- **Lyric Recognition and Restoration**: Collaborated with Warner Records to develop deep learning-based lyric recognition systems for restoring and generating lyrics from historical music recordings.

Wyze Labs, Inc.

Kirkland, WA

RESEARCH SCIENTIST II, AUDIO AI

December 2019 – September 2021

- **Audio Event Dataset and Annotation**: Led the organization and management of annotation workflows for large-scale sound-event datasets sourced from public audio and customer feedback. Designed annotation guidelines, tutorials, quality metrics, and evaluation processes to ensure consistent, high-quality training data.
- **Cloud Audio Event Recognition**: Researched and developed deep learning-based audio classification models and collaborated with engineering teams to deploy them as scalable cloud services for Wyze cameras, enabling detection of events including **baby cries, dog barks, and gunshots**.
- **Edge Audio Event Detection**: Partnered with edge-device engineers to develop an efficient **C++ on-device audio event detection algorithm**, filtering more than **95% of irrelevant/noise audio events during testing** and reducing unnecessary downstream cloud processing.
- **Multi-Object Tracking**: Mentored a University of Washington capstone project on multi-object tracking using Wyze cameras, guiding data collection, model training, experimentation, and technical research.

Patent and Award

Patent: Cross-Document Intelligent Authoring And Processing, Including Format For Semantically-Annotated Documents

US20210081601A1

Zhanlin Liu, Andrew Begun, Steven DeRose, Taqi Jaffri, etc.

July 19th, 2022

Award: 3rd place Data+AI Summit Hackathon (>200 participants) by Databricks.

San Francisco

Zhanlin Liu, Mohamed Elhoseiny, Julien Witty, and Shreshth Sonkiya.

June 13th, 2024

Award: 12th place Deep Research Tool Competition by OpenAtom Foundation.

China

Liang Chen, Zhanlin Liu, Ziyou Li

April 15th, 2025