

ARLEN KUMAR

Berkeley, CA | (669) 292-8054 | arlenkumar2829@gmail.com | arlenkumar.com
linkedin.com/in/arlen-frederick-kumar-1198592b8 | github.com/arlenk2021 | arxiv.org/abs/2509.10762

SUMMARY

Co-Founder & CTO and published NLP/IR researcher building LLM retrieval and AI search infrastructure. Designed and shipped production RAG systems, LLM evaluation pipelines, and AI crawler analytics end-to-end; co-author of GEO-16 (arXiv:2509.10762). Targeting Founding Engineer and Member of Technical Staff roles in LLM infrastructure, retrieval, and AI search.

EXPERIENCE

Co-Founder & CTO — Wrodium | Berkeley, CA | 2024 – Present

Generative Engine Optimization (GEO) and knowledge freshness infrastructure. Berkeley SkyDeck Batch 21; winner, Most Innovative Technology, SkyDeck Pad-13 Pitch Competition.

- Own end-to-end architecture of the full production surface: **Living Articles** publishing pipeline, **Semantic Resonance Engine**, **ClawdBot/OpenClaw** AI crawler and citation-monitoring system, **WebMCP** infrastructure layer, and the Wrodium Roast audit tool.
- Designed and shipped production **Retrieval-Augmented Generation (RAG)** systems: hybrid retrieval (dense embeddings + BM25) with reciprocal rank fusion, **pgvector** indexes (HNSW, IVF-Flat), cross-encoder re-ranking, structured-output extraction, and **LLM-as-judge evaluation harnesses**; optimized for latency and cost across ChatGPT, Gemini, Perplexity, and Claude.
- Built AI-bot crawl analytics and **citation drift monitoring** on Cloudflare (GPTBot, ClaudeBot, PerplexityBot, Google-Extended): recrawl scheduling, change-detection pipelines, and staleness scoring; production stack spans **Python, TypeScript, PostgreSQL, FastAPI, Next.js, and Cloudflare Workers**.
- Managed a **5-engineer development team** and led a **13-person intern cohort** across engineering, research, and GTM tracks plus a founding marketer; authored Wrodium's **deployment guide and release runbooks**, standardizing CI/CD and production release practices; owned hiring frameworks and technical mentorship.
- Delivered GEO infrastructure for clients in legal AI, construction software, and luxury e-commerce, with integrations for six major CMS platforms — schema.org/JSON-LD rollouts, Wikidata entity disambiguation, citation tracking; client-reported **+20% AI citations within one week** and 70% content-workload reduction.
- Closing a **\$1.5M pre-seed at a \$15M post-money SAFE cap**; built investor data room and 18-month financial model; designed proprietary revenue-weighted GEO measurement framework (Conversion Potential Score, Prompt-to-Revenue Velocity, AI Traffic CAC).

Undergraduate Researcher, NLP & Information Retrieval — Hearst Lab, UC Berkeley | 2024 – Present

Advised by Prof. Marti Hearst.

- Co-authored **GEO-16 (arXiv:2509.10762)**, an empirical analysis of AI answer engine citation behavior: audited 1,100 URLs across 70 prompts and 1,702 citations across three engines; pages meeting quality thresholds achieved a **78% cross-engine citation rate** (odds ratio 4.2).
- Co-authored **CHASE** — knowledge freshness and citation behavior in retrieval-augmented LLMs; under review at **COLM 2026**.
- Produced a **Lean 4 formal verification** of the GEO-16 framework (CS 294-268), bridging empirical NLP findings with machine-checked proofs.

Founder — Air Quake Simulations (Acquired) | VR flight-simulator hardware

- Founded, scaled, and **exited** a VR flight-simulator hardware company before age 21; shipped **hundreds of F-18 cockpit units at 10x below incumbent pricing** via 3D-printed components.
- Engineered an AI-powered VR training platform in **Python and PyTorch** with reinforcement learning, real-time adaptive feedback, and GPU-accelerated pipelines integrated with Unity3D.

EDUCATION

University of California, Berkeley — B.A. Computer Science, Data Science, and Economics | May 2026

SKILLS

Languages: Python, TypeScript, JavaScript, SQL, Lean 4, Bash, Java

ML & Retrieval: PyTorch, Hugging Face Transformers, Retrieval-Augmented Generation (RAG), embedding fine-tuning (OpenAI, Cohere, BGE, E5), pgvector, HNSW, BM25, hybrid retrieval, reciprocal rank fusion, cross-encoder re-ranking, retrieval evaluation (nDCG, MRR, recall@k), LLM-as-judge, prompt engineering, Model Context Protocol (MCP), agentic tool use

Infrastructure: PostgreSQL, Supabase, FastAPI, Next.js, Cloudflare Workers, CI/CD, web crawlers, schema.org / JSON-LD, knowledge graphs, Wikidata, Unity3D

Open Source: FreeLLMsTxt — llms.txt generator with sitemap-aware crawler, FastAPI web UI, and CLI (github.com/arlenk2021/FreeLLMsTxtGenerator)

PUBLICATIONS & TALKS

- GEO-16: "AI Answer Engine Citation Behavior" (Kumar & Palkhouski), arXiv:2509.10762 (2025).
- CHASE — knowledge freshness in retrieval-augmented LLMs; under review, COLM 2026.
- "The Ad-ification of AI" — invited talk, CITRIS, UC Berkeley (Feb 2026).