Caleb Kha-Uong

(303) 618-2650 | ck2867@princeton.edu | LinkedIn | GitHub | Website

Education

Princeton University Aug. 2024 – May 2028
Bachelor of Engineering in Computer Science; Minors in Statistics & ML and Quant Econ *Princeton, NJ*

Relevant Coursework: Data Structures & Algorithms, Vector Calculus, Linear Algebra with Applications, Machine Learning (Fall '25), Programming Systems (Fall '25)

Experience

Princeton Computational Imaging Lab Sep. 2025
Computer Vision Researcher *Princeton, NJ*

- Supervised by Prof. Felix Heide

Remora May 2025 – Present
Quantitative Software Engineer *Lexington, VA*

- Building predictive models for **S&P 400 MidCap** stocks to forecast 2-week returns and rank top gainers/losers from market and news data.
- Designing sector-aware ensembles that, in backtests, beat the benchmark by **1.5–2%** in 2-week returns for top-quintile picks.

MyChance.ai Apr. 2025 – Present
Founding Engineer *Princeton, NJ*

- Built and launched a centralized college admissions platform (match-making, chance-me, essay review, application tracking), attracting **3,000+** sign-ups.
- Designed a scalable **Supabase/PostgreSQL** backend for profiles, college data, and recommendations; integrated **Stripe** for subscriptions.

Princeton University IPA Lab Sep. 2024 – Present
AI Researcher *Princeton, NJ*

- Led large-scale LLM hallucination experiments with **1,000+ trials** on **20,000 WikiData facts** across six spacing parameters (G values).
- Found significant schedule effects ($p < 0.001$): blocked training reduced hallucinations by an average of **21%** in Geography and **29%** in Literature vs. interleaved.
- Ran DRM false-memory experiments (**4,000+ prompt–response pairs**); semantic clustering raised false recall by **7%** on average and up to **35%** in some models.

Projects

Distill | *Python, PyMuPDF, Transformers, OpenRouter, BM25, Rank Fusion/MMR* Aug. 2025 – Present

- Achieved up to **98% context reduction**, saving **10,000+** tokens per query under tight context budgets.
- Increased retrieval efficiency by filtering **500+** candidate chunks to ~ 250 high-value segments (≈50%), cutting processing costs nearly in half while preserving citation accuracy.

Volley | *Python, YOLO-v8, PyTorch* Jun. 2025 – Present

- Developed a real-time tennis tracking system (60 FPS) that detects players and the ball, drawing a 2D overlay of positions on the court.
- Estimated ball speed/velocity from keypoints with **95% accuracy**, showing live response in output video.

Skills and Interests

Languages: Java, Python, C (Fall '25), JavaScript, HTML, CSS
Frameworks/Libraries: React, Next.js, pandas, NumPy, Matplotlib, PyTorch, GSAP, Tailwind
Databases/Platforms: Supabase, PostgreSQL/SQL, Firebase, Git, Vercel, Docker