Pinlin [Calvin] Xu

pinlinxu@stanford.edu | linkedin.com/in/pinlin-xu | pinlinxu.com | github.com/Calvin-Xu | KN6YCY (Radio)

EDUCATION

Stanford University — Stanford, CA

B.S. Computer Science (AI Track) | *CS Honors Program* | GPA 3.95 — *September 2022 – June 2026*

EXPERIENCE

Undergraduate Researcher & Engineer — Dec 2024 –

Stanford Department of Emergency Medicine & Mussallem Center for Biodesign — *Stanford, CA*

- Working on releasing My Heart Counts AI dataset and creating baselines at the Ashley Lab for precision medicine
- Led building Stanford Feedbridge, an iOS + Firebase baby tracker to prevent physiologic poor feeding in newborns
- Support & DevOps for a React Native physician well-being study platform on AWS, Supabase, Apple App Store.

Software Engineer (part-time) — Sep 2024 – Mar 2025

Ideaflow Inc. — *Palo Alto, CA*

- Developed core product (collaborative semantic knowledge graph and editor) in TypeScript & Next.js. Created custom NLP pipelines. Spec-ed & shipped graph query system. Supporting and pitching to researchers.

Undergraduate Researcher — Jan 2024 –

Stanford Causal AI Lab (SCAIL) — *Stanford, CA*

- Researched generative genetic model for perturbation prediction using PubMed-derived causal graphs
- Developed semantic entity resolution for knowledge-graph powered causal discovery; collaborated w/ Ideaflow Inc

Software Engineer Intern — Jun 2023 – Sep 2023

JuniorKids Group / Le Groupe JuniorKids — *Montreal, QC*

- Led intern team of 7 automating marketing and admin workflows with OpenAI, Shopify GraphQL APIs
- Contributed to traffic growth through August for the first time and 37% YoY growth

SKILLS & SELECTED COURSEWORK

C, C++, Python, JavaScript, TypeScript: web, Swift: iOS, Java, SQL, Spark, Git, CI/CD

Machine Learning, Deep Learning: NLP, Graph (CS 224N/W), Operating Systems, Compilers, Networking, Algorithm & Data Structures, Mining Massive Data, HPC, Theoretical Statistics

LLM from Scratch (CS 336): every aspect of language model creation, including distributed training infrastructure, tokenizer, model architecture, optimizer, Triton & CUDA kernels, FlashAttention, DPO, GRPO

SOME PROJECTS

Magi: High-Throughput Ontology Creation and Causal Discovery System (Honors Thesis) — Jan 2024

- Building a scalable, high-performance LLM-powered knowledge processing system using Apache Spark, Docker
- Designed fully containerized application leveraging PySpark, Postgres (pgvector) and Memgraph for storage and a novel embedding-based coreference resolution method, and Valkey (Redis) for rate limiting and message queues; ingest your data from S3 and build high-quality ontology at the speed of light

FLFL: Grounded Japanese Furigana Generation using Aligned Whisper Transcription — 2024

- Evaluated and released model, datasets, & codebase for processing 20+ GB of public-domain audiobook data released by the Japanese National Diet Library | *HuggingFace Trainer, axolotl, Modal.com*

Predicting Hospital Length of Stay from Imbalanced Data | *Python, scikit-learn, XGBoost* — 2024

- Led and presented a strong classification-regression pipeline using synthetic oversampling, ensemble learning

Allegorical Lisp Machine | *C, Lisp, ARMv6 Assembly, Calculus & Discrete Math* — 2023

- Built a freestanding graphical Lisp environment on Raspberry Pi A+ with low-level routines in bare-metal C
- Implemented Lisp interpreter, system calls, exception handling, REPL, etc. from relevant papers

Hikari Ray Tracer | *Typed Racket, RackUnit* — 2022

- 3D ray tracing renderer from scratch in functional Typed Racket (PLT Scheme, Lisp dialect)

Flow Browser | *SwiftUI, UIKit, WKWebView* — 2019

- Designed and built a tree-style tab browser for iOS & iPadOS; enabled iCloud integration, etc. with native APIs

The Shades of Meaning: LLMs' Cross-lingual Representation of Grounded Structures | *NLP, LLMs*
- Led an outstanding CS 224N custom project poster and report | *Python, PyTorch, transformers*
- Collected and built a dataset of cross-lingual cultural color words; performed a series controlled experiments on how the quality of representations varies with language, model, context, and fine-tuning
- Introduced a novel visualization for cross-lingual perceptions of color spaces using orthogonal Procrustes analysis

GNN-Powered Predictive Drug Repurposing via Augmenting Biomedical Ontologies | *GNNs, Bioinformatics*
- Developed a novel approach leveraging Graph Neural Networks (GNNs) to enhance predictive drug repurposing using the Drug Repurposing Knowledge Graph (DRKG) | *PyTorch, PyGeometric, GraphML*
- Engineered an augmented DRKG by deriving protein interactions from STRING and AlphaFold 3
- Implemented and compared various GNN designs, achieving a 55% accuracy boost over KG embedding baselines in identifying COVID-19 treatments with a custom Heterogeneous Graph Transformer (HGT) architecture

Roots of Probability: A New Commentary on *De Ratiociniis in Ludo Aleae* (1657) | *Latin, Science*
- Composed a new partial translation of & commentary on *De Ratiociniis in Ludo Aleae* by Christiaan Huygens, the first printed text on probability
- Compared against 1714 translation and other commentaries on that translated version; pointed out errors in the original Latin version, the translation, and subsequent commentaries.

Delere imperium* and *animi imperio*: the semantics of *imperium* in *Cic.Cat.* and *Sall.Cat. | *Latin Philology*
- Conducted an in-depth linguistic and historical analysis of *imperium* in Cicero's *In Catilinam* and Sallust's *Bellum Catilinae*, leveraging the *Thesaurus Linguae Latinae* (TLL) to track its varied meanings and rhetorical significance; identified and corrected an error in the *TLL* for future editions
- Engaged with primary Latin texts and lexicographical sources to reconstruct the ideological and rhetorical evolution of *imperium* for fresh insights into Cicero's and Sallust's political narratives
- Contrasted Cicero's rhetorical monopolization of *imperium* with Sallust's teleological metaphor of *imperium* as a symbol of Rome's moral and political decline within his historiography