YINCHENG ZHOU

(267) 788-2488 | yzhou29@seas.upenn.edu | github.com/ArtysicistZ | artysicistz.github.io

EDUCATION

University of Pennsylvania September 2025 - Present
Vagelos Integrated Program in Energy Research (VIPER) *Philadelphia, PA*

- Dual Degree: B.S.E. in Computer & Information Science (AI Concentration) | B.A. in Physics & Astronomy
- Relevant Coursework: Big Data Analysis, Computer Systems, Data Structures & Algorithms
- GPA: 4.0 / 4.0

EXPERIENCE

Franklink, Inc. [Demo] October 2025 - Feburary 2026
Founding Software Engineer *Philadelphia, PA*

- Built an iMessage-based networking assistant, automating matchmaking via multi-agent orchestration.
- Implemented persisted agent memory and selective dispatch across 5 sub-agents; cut user-facing latency by 30%.
- Designed agentic RAG with permissioned access across 8 tools (DB, email, calendar) for context-aware retrieval and recommendation.
- Engineered a Kafka-backed event pipeline for LLM-heavy messaging with at-least-once delivery, idempotency keys, and retry topics (backoff) + DLQ; supported 50+ concurrent sessions while keeping p99 end-to-end latency within 40% of single-session baseline.
- Containerize services and design test/deploy workflows; ship on AWS ECS.

University of Pennsylvania, GRASP Laboratory January 2026 - Present
Machine Learning Researcher | Advisor: Prof. Pratik Chaudhari *Philadelphia, PA*

- Designed SPD, a sparse depth prediction algorithm that decodes only at queried pixels, achieving 2.9× inference speedup over dense baselines for real-time robotics.
- Implemented a lightweight per-query decoder (11M params) with multi-scale cross-attention, cross-query communication, and deformable sampling, enabling 72 Hz inference at 256 query points with 544*416 resolution on RTX 4060.
- Architected dense canvas supervision strategy to address sparse query information loss during training, enabling convergence on NYU Depth V2 without ground truth densification.

PROJECTS

AlphaOne [GitHub] [Demo] [Play with DeBERTa-ABSA-v2] October 2025 - Present
Tech Stack: PyTorch, HuggingFace, React, Spring Boot, FastAPI, Celery, Redis, PostgreSQL

- Fine-tuned DeBERTa to classify sentiment per ticker in multi-entity financial text; achieving 82.5% accuracy (0.823 Macro F1), increasing performance by 65 percentage points from FinBERT baseline.
- Curated 6,287 training triples: scraped 7 subreddits, LLM-labeled via Ollama, hand-audited 4,501 pairs (615 corrections).
- Architected a 6-layered sentiment pipeline with idempotent content hashing for concurrent-safe batch ingestion.
- Engineered concurrent-safe, idempotent batch ingestion with content versioning; deployed via Docker Compose.

NL2SQL Bot [GitHub] [Demo] December 2025 - January 2026
Tech Stack: Google ADK, FastAPI, MySQL, Plotly.js

- Built a multi-agent NL2SQL system on Google ADK converting natural language to validated SQL with interpretive visualization.
- Orchestrated 5 agents with session state passing, loop-guarded retry, and per-agent LLM overrides for cost-performance tuning.
- Enforced read-only execution via 18 blocked SQL patterns and table allowlisting; abstracted dialect rules across MySQL, PostgreSQL, and SQLite.

SKILLS & ACHIEVEMENTS

Languages	C++, Python, Java, JavaScript/TypeScript, SQL, Assembly, OCaml
Backend & DevTools	FastAPI, Spring Boot, Kafka, Redis, Celery, Docker, AWS, Git
ML & Data	PyTorch, CUDA, spaCy, Pandas, LangGraph, Google ADK
Achievements	Team China, International Science and Engineering Fair (ISEF)
	Silver Medal, Team China, International Young Physicists' Tournament (IYPT)
	Second Prize, China Mathematics Olympiad
	Second Prize, China Physics Olympiad, Second Prize
	Champion, Start-up in a Weekend Hackathon, Photon Hybrid Intelligence Track