David (Yunlang) Dai

Haverford, PA
ddai@haverford.edu • yunlangdai.com
david-dai-2b099b267

Education

Haverford College **Haverford, PA**
B.S. Mathematics & Computer Science *2026 (Expected)*
GPA: 3.9 **Minor:** Data Science
Honors: Class of 1896 Prize (Mathematics)
Selected Coursework: Deep Learning; Data Mining; Statistical Methods (Categorical, Nonparametric, Multivariate); Algorithms and Data Structures; Operating Systems; Real Analysis

Carnegie Mellon University **Pittsburgh, PA**
LearnLab Summer School *July 2025*
Selected participant in intensive program on Intelligent Tutoring Systems.

The Hudson School **Hoboken, NJ**
High School Diploma *2022*
Graduated summa cum laude (GPA: 4.0).

Research Experience

Haverford College **Haverford, PA**
Large Language Models (LLMs) Moderation Auditing and Steering *Feb 2025 – Present*

Advisor: Prof. Sorelle Friedler

Summary: Investigating how large language models (LLMs) handle socially sensitive content, and developing methods to steer and evaluate moderation behavior over time.

- **Longitudinal Moderation Auditing** (*Feb 2025 – Sep 2025*): Studied temporal drift and alignment stability in commercial LLM moderation. Built auditing pipeline and interactive dashboard (*genAIaudits.github.io*). **First co-author** on *"Longitudinal Monitoring of LLM Content Moderation of Social Issues"* (*arXiv:2510.01255*). Preparing submission to **CHI 2026**.
- **Senior Thesis** (*Aug 2025 – Present*): Extending this work to design interpretable steering mechanisms using concept directions. Developing quantitative metrics for steerability, consistency, and semantic fidelity across model checkpoints.

University of Pennsylvania **Philadelphia, PA**
Human-Centered AI for Learning and Instruction *Dec 2024 – Present*

Advisors: Prof. Ryan Baker & Dr. Anthony Botelho

Summary: Investigating how learners engage with online courses and AI tutoring systems, and how student mastery evolves over time. Combining large-scale behavior analytics with interpretable modeling to improve educational feedback and support.

- **MOOC Dropout Modeling** (*Dec 2024 – Mar 2025*): Analyzed 9M+ learners across 200+ massive open online courses (MOOCs); identified course design features influencing dropout. **First author** on *"Understanding MOOC Stopout Patterns: Course and Assessment-Level Insights"*, published in **Learning@Scale 2025** (*doi:10.1145/3698205.3733944*).
- **LLM Teaching Assistant Evaluation** (*May 2025 – Oct 2025*): Examined student–LLM interactions (JeepyTA) in live course deployments. Built engagement analytics pipelines to understand how LLM feedback affects learning process and performance.
- **Knowledge Tracing Model Evaluation** (*Aug 2025 – Present*): Comparing Bayesian and deep

knowledge tracing approaches on ASSISTments. Focusing on interpretability and modeling of evolving student mastery. Contributing to evaluation framework design.

Haverford College **Haverford, PA**
Senior Thesis in Mathematics *Aug 2025 – Present*

Advisor: Prof. Lynne Butler

Summary: Developing theoretical results on the Poisson binomial distribution and its extensions to sequentially dependent trials. Deriving moment and convergence properties with proofs emphasizing heterogeneity among events in stochastic sequences.

Bryn Mawr College **Bryn Mawr, PA**
Auto-grading Pipeline Reproduction and Extension *Jun 2024 – Sep 2024*

Advisor: Dr. Ratnik Gandhi

Summary: Reproduced Stanford's generative grading approach for richly structured problems (*Malik et al., 2019*). Extended the original VGG and KNN models to small pre-trained transformer architectures, and implemented a probabilistic-programming data simulator to support both text-based and vision-based grading workflows.
Source code: `github.com/davidhhhhhhh/research2024.git`.

Selected Projects & Competitions

Massachusetts Institute of Technology **Cambridge, MA**
Solo Developer, HackMIT 2025 *September 2025*

Summary: Built a web-based intelligent study assistant that identifies and resurfaces students' confusion points through AI-driven **syllabus mapping** and **historical conversation data**. Developed the full-stack system in Flask with SQLite and the Anthropic Claude API.
Source code: `github.com/davidhhhhhhh/confusion-bank`.

Tencent **Online**
Team Lead, AI Arena Global Open *August 2025*

Summary: Led a team of five undergraduates in *Intelligent Agent Gaming Algorithm* (*Intermediate*) *Track*. Implemented a **PPO-driven GRU agent** for a partially observable Markov decision process environment. Awarded the **Gold Prize (top 1.6% out of 1000+ teams)** in the Undergraduate division.

Carnegie Mellon University **Pittsburgh, PA**
Participant, LearnLab Summer School *July 2025*

Summary: Participated in a competitive summer program directed by Prof. Vincent Aleven and Jonathan Sewall. Designed and implemented an **Intelligent Tutoring System** for JavaScript using the CTAT framework, including skill modeling and feedback logic. Presented final system and results in a poster session.

Bryn Mawr College **Bryn Mawr, PA**
Team Lead, PraxIS Course Project *Jan 2025 – May 2025*

Summary: Led a data engineering project for the Philadelphia Solar Energy Association (PSEA). Automated data ingestion, cleaning, and visualization workflows, **reducing manual reporting time by 90%**. Deployed interactive dashboards to the public-facing `PA Solar Energy Snapshot` website, improving accessibility of solar installation data across the state.

Villanova University **Villanova, PA**
Team Lead, DataFest Philly 2025 *March 2025*

Responsibilities: Led a team in analyzing commercial office leasing trends pre- and post-pandemic across U.S. cities and leasing types, using data from Savills and statistical modeling. The competition was co-hosted by Villanova University and the American Statistical Association. Awarded **Top External Data Source Prize** (one of three prizes awarded).

Roundhouse One **San Francisco, CA**
Data Analyst Intern *Jun 2024 – Aug 2024*

Summary: Conducted multivariate environmental data analysis from **over 200 K–12 schools** in Hawaii. Modeled the effects of HVAC systems on indoor environmental quality using statistical and causal inference methods. Drafted two research papers linking environmental factors to adaptive building design strategies.

Honors & Awards

2025: **Gold Prize**, Tencent AI Arena Global Open (Top 1.6% of 1000+ teams). Recognized for implementing adaptive PPO–GRU agent in a POMDP environment.

2025: **Student Travel Scholarship**, Educational Data Mining Society (EDM 2025, Palermo, Italy). Awarded for research on large-scale modeling of learner behavior.

2025: **Koshland Summer Scholar**, Marian E. Koshland Integrated Natural Sciences Center, Haverford College. Supported summer research on AI auditing pipeline.

2025: **Top External Data Source Prize**, DataFest Philly 2025 (Villanova University, ASA). Recognized for integrating external datasets to uncover post-pandemic leasing pattern shifts.

2024: **Class of 1896 Prize in Mathematics**, Haverford College. Awarded to the top sophomore in the Department of Mathematics and Statistics.

2022: **Distinguished Academic Award**, The Hudson School. Graduated *summa cum laude* as top of class.

Skills

Adaptive & Machine Learning Systems: Deep Learning (PyTorch, Hugging Face, CUDA); Reinforcement Learning (PPO, DQN, SARSA, GRU-based agents); Representation Steering, LangChain / LangGraph, Retrieval-Augmented Generation (RAG), LoRA / QLoRA fine-tuning.

Probabilistic & Statistical Modeling: Bayesian inference, Sequential modeling, Hidden Markov Model, Knowledge tracing models (BKT, DKT), Uncertainty quantification, Model selections, Statistical hypothesis testing.

Programming Languages: Python (primary), R, C++, Java, SQL, C.

Data Infrastructure & Tooling: Jupyter, Git / GitHub, PostgreSQL, SQLite, Shell scripting (Bash / PowerShell), Data visualization (Matplotlib, ggplot2).

Systems & Deployment: Linux (Debian), MacOS, Windows; Google Cloud Platform (GCP), Microsoft Azure; Distributed and server management (SSH, cron).

Full-Stack Development: Flask, HTML, CSS, JavaScript, AJAX, Marked.js.

Languages: Mandarin (fluent), Latin (intermediate).