YANG YAN

BA. COMPUTER SCIENCE & MATHEMATICS

MIT, 2017-2021, 2025 · 4.5/5

Matrix Algorithms · Database Systems · Advanced Algorithms · Randomized Algorithms · Probability & Random Variables · Group Theory

Traders@MIT · AI@MIT · DFA

USACO FINALIST '14

USAMO QUALIFIER '13 '15 '16

ISEF FINALIST '16

C++: 17 · (n)make · WinAPI

Javascript: ES6 · React · Vue · TS

Python: 3.8 · torch · scikit · flask

Postgres · Markdown · TeX · Docker

EMILIA HTTP & SMTP

gilgamesh.cc, 2015+

Reverse-engineered TCP, HTTP, SMTP servers written in native, cross-platform C++17. Original front-end design in native ES6.

MIT CONFESSIONS SIM.

fb.com/mitconfessionssim, 2018

Char-RNN and Bayesian text models for MIT Confessions. Data scraping & posting integration in Selenium.

XENA, 2024

E-ink, handwriting-optimized, theme-aware SVG notes.

ML ALIGNMENT & THEORY SCHOLAR

with Ethan Perez @ Anthropic, 2024-2025

"Towards Safe Language Model Fine-tuning APIs"

- Developed datasets & defenses against CMFT-like and other attacks against LLM finetuning.

PRODUCT MANAGER, PAYMENTS

Nuvo Technologies, 2023-2024

- Led customer, banking, service provider, and internal stakeholder conversations to scope and build out Nuvo's first payments product: features, engineering, regulatory, banking, pricing.

SOFTWARE ENGINEER, RISK

Ramp, 2021-2022

- Modeled credit, fraud, and operations risks, for an expected ~80% ≈ 5mm annual fraud averted.
- Optimized production model evaluation latency, speeding up every single transaction and transfer that ever occurs at Ramp by ~20ms ≈ 10%.

INTERN, QUANTITATIVE RESEARCH

D. E. Shaw & Co., 2020

- Designed and built a market simulator, each agent having personalized strategy and private forecasts, optimizing P&L under different training scenarios.
- Explained effects of varying number of agents, forecast horizons/accuracies correlations, on PnL.

PREVIOUS: Intern @ ***Scale AI***, Intern @ ***Microsoft***

UNDERGRADUATE RESEARCHER

with Greg Wornell @ MIT, 2020-2021

on "Adversarial Examples in Simpler Settings"

- Derived robustness measure for classifier features (pen-ultimate NN layer), discarding the lowest of which will naively improve model robustness.
- Verified hypothesized robustness inheritance effects in select transfer learning scenarios.

+1 (801) 558-2948 · yang@gilgamesh.cc · github.com/GilgameshxZero