

Tianshu (Terry) Yin

Atlanta, GA 30318 | +1(470)5211660 | Portfolio(Github) | [LinkedIn](#) | yintry314@gmail.com

AI Developer & Software Engineer

EDUCATION

| | |
|----------------------------------------------------------------------------------------------------|---------------------|
| Georgia Institute of Technology: (Bachelor of Science in Computer Engineering) GPA: 4.0/4.0 | Aug 2024 – Now |
| Shanghai World Foreign Language Academy (High School) GPA: 41/42 | Sep 2021 – Jun 2024 |
| Carnegie Mellon University Pre-College Program: National High School Game Academy | Jul 2023 – Aug 2023 |
| Main Academic Awards: USAMO Qualified, Bpho Top Gold (global top 100), USACO Gold | Oct 2022 – Feb 2023 |

SKILLS

- Language: C++, Python, Java, HTML, Javascript, mongoDB, CSS, Go, Csharp, Matlab
- Technical Skills: LINUX/UNIX, Relation DB, React, LLaMA agent, Docker, Node.js, Nginx, sentence transformer

INDUSTRIAL EXPERIENCE

| | |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------|
| OORTech LLC. AI Developer Intern | (Local, US) May 2025 – Aug 2025 |
| • Built a consulting chatbot using distributed LLaMA-3 agents across multiple servers, combining tensor parallelism and pipeline parallelism , after benchmarking Exo, prima.cpp, and Petals on GPU-less edge CPU nodes; improved inference speed by ~35% in internal load tests. | |
| • Built an AI crawling + local LLM pipeline for web scraping and LLaMA post-processing to strip irrelevant ads and restructure content by importance, enabling low-cost offline data processing for downstream analytics. | |
| • Enhanced the Knowledge Graph on filtering school information using Docker and Neo4j sandbox , cutting query lookup time by ~30% in AI agent benchmarks; developed Python tools to automate KG entity addition and metadata management. | |

| | |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------|
| TalkMeUp Inc. Website Engineer Intern | (Remote, US) May 2024 – Aug 2024 |
| • Improve performance of TalkMeUp Platform with Flask that helps students prepare for interviews with online AI coaches. Led a team of 4 interns improving the platform performance across browsers/devices, fixing 15+ critical UI/UX bugs and optimizing video delivery with port tuning and Asia-wide mirrored servers , cutting APAC load time by ~50% . | |
| • Built a React based UI for the AI Interview Chatbot with a Python (Flask) backend , and developed MongoDB aggregation dashboards for real-time user/team metrics, boosting training efficiency by over 20%. | |
| • Iterated on the TalkMeUp Analytics module with PMs during weekly sprints, improving AI service accuracy and dashboard responsiveness; performed regression testing to maintain 99.9% uptime and smooth API integration. | |

SELECTED RESEARCH/PROJECTS

| | |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------|
| SIPLab - AI Personality Simulation Gatech, US | Feb 2025 – Now |
| • Developed a desktop Multi-agent LLaMA agent that simulates human's behavior using aggregated personalities, with a GUI (CustomTkinter + llama-cpp) that loads GGUF models , applies up to 6 persona agents with normalized weights, runs parallel generation, and produces an aggregated consensus, meeting real human predictions over 80% of the time . | |
| • Ran BART tests across persona mixtures and quantified how weight proportions shift pump counts/explosion rates. | |
| • Implemented a fatigue (“tiredness”) middleware that tracks task count and cumulative effort; after high-effort turns, dynamically increased refusal probability by up to 40% based on dullness score and total trials. | |
| • Developed a matplotlib visualization pipeline to correlate LLM-generated answers with weighted personality compositions using custom prompt engineering to control bias and refine aggregation algorithms. | |

| | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------|
| Core Web Developer (Web-Dev Club) Gatech, US (https://gtmarketplace.onrender.com) | Sep 2024 – Apr 2025 |
| • Led the development of a React + MongoDB online marketplace with email sign-up/login, search/navigation for listings, and a profile dashboard. Designed MongoDB schemas for users, listings, and live message threads ; implemented CRUD logic for profile/item management and enabled direct buyer–seller messaging, with AI-based search and filtering . | |
| • Integrated Firebase Authentication for secure email login and session handling, and connected the React/JavaScript/CSS front end with MongoDB + Firebase back end, maintaining 300+ active users over the semester. | |

| | |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------|
| RAG Document QA Chatbot (Lab Research) Gatech, US | Jul 2025 – Now |
| • Built a document-QA system using Retrieval-Augmented Generation : PDF ingestion and sentence-aware overlap chunking → FAISS (cosine/IP) vector search and top-k retrieval → generations with “ Cannot be determined ” fallback. | |
| • Implemented index persistence with FAISS + JSON metadata, and 3+ multi-model embeddings selection. | |
| • Engineered a Python stack using NumPy, sentence- transformers , Gradio, OpenAI, with retries and error handling. | |

PAST ACTIVITIES

| | |
|-----------------------------------------------------------------------------------------------------------------------------|----------------|
| National Highschool Astronomy Aerospace Organization Founder and Former Leader | Oct 2022 – Now |
| • Addressed the lack of astronomy resources in high schools, by uniting 21 schools, 100+ active members. | |
| • Facilitated Saturday night seminars and speeches by university professors offering educational opportunities for members. | |
| • Led multiple long-term academic projects, including a successful nation-wide photographing to analyze the moon's surface. | |