Ethan Lu

ethanlu0929@gmail.com | 973-997-3240 | Richmond Hill, GA | ethanlu.netlify.app | elu139 | ethan-lu

EDUCATION

Georgia Institute of Technology **Atlanta, GA**

B.S. in Computer Science *Expected May 2027*

- **GPA:** 4.00 | **Concentrations in** Information Internetworks and Intelligence
- **Coursework:** Data Structures and Advanced Algorithms, Systems and Networks, Computer Architecture, Databases, Ethics in Computing, HCI Research Methods, Discrete Math, Linear Algebra, OOP
- **Organizations**: Data Science at Georgia Tech, Investment Management Club, SynapseX, CHEFS at Tech

EXPERIENCE

Georgia Tech Century for 21st Century Universities Lab **Atlanta, GA**

Lead Machine Learning+Data Engineer for Data-Driven Education *Jan 2025 – Present*

- Increased scalability of model training by generating a synthetic dataset with ~90% statistical similarity to the institution's course forum data, using OpenAI's o1 model and validating similarity through Latent–Dirichlet Allocation (LDA) topic modeling and personalized natural language processing tokenization
- Improved instructor insight by building a React dashboard with pyLDAvis visualizations for topic modeling, with planned integration of sentiment analysis
- Developed a BERT-encoded classification model to categorize forum posts into cognitive presence categories, increasing accuracy and interpretability of decision-making and discussion analytics

PROJECTS

factr | *PyTorch, FastAPI, OpenCLIP, LLaVA, AWS Lambda, Hugging Face* ***May 2025 – Present***

- Reduced misinformation detection false negatives by ~20% over unimodal baselines by developing an AI-backed pipeline that cross-validates image-text-vision models
- Designed cloud-native infrastructure enabling horizontal scaling, rapid experimentation, and production-ready API services for multimodal inference
- Trained models on LIAR, DFDC, and other benchmark datasets to ensure cross-modal robustness for real-world misinformation scenarios

EEG-Based Cursor Platform | *Python, PyTorch, WebSocket, pylsl, Node.js, Flask* ***Aug 2025 – Present***

- Building brain-computer interface (BCI) prediction pipeline to transform livestreamed EEG data into real-time cursor control while creating data replay functionality and optimizing latency and stability

Company Moat Impact Score Model | *Python, PyTorch, SQL, PostgreSQL* ***Aug 2024 – May 2025***

- Delivered interpretable 0–1 scaled growth potential and market impact scores for companies in a simulated $100K investment portfolio, by designing a scoring framework using Gaussian Mixture Models and log-probability transformations
- Built data pipelines to collect and normalize real-time and historical financial metrics from the Financial–ModelingPrep API, enabling unsupervised sector-based performance analysis

Premier League Transfer Analysis | *Python, Pandas, Numpy, Matplotlib* ***Aug 2024 – Dec 2024***

- Determined that investing in fewer, high-impact rather than many, low-impact players correlated with higher league survival rates for newly-promoted clubs, based on analysis of ~500 Premier League transfer records (2008–2024) filtered from a dataset of 75K+ entries using BeautifulSoup and visualizing results

MathWorks Math Modeling Challenge | *Python, Pandas, MATLAB* ***Mar 2024 – Mar 2024***

- Achieved a predictive accuracy of 0.86 AUC in forecasting homelessness trends in U.S. cities by developing a binomial logistic growth model to analyze complex socioeconomic datasets using Pandas and MATLAB

The Climate Reality Project | *JavaScript, React.js, Node.js, HTML, CSS* ***Mar 2022 – Oct 2022***

- Expanded access to 100+ climate education resources integrated into the curricula of 15 local schools and libraries upon live, in-person demonstrations, by developing a user-friendly web platform
- Enabled real-time content publishing by integrating the TinyMCE rich text editor, allowing multiple teammates to upload and update educational material directly to website

Senior Home Sign-In Application | *Python, Tkinter, React.js, Node.js* ***Dec 2021 – Dec 2022***

- Improved accuracy and efficiency of daily resident and visitor log management at a local senior community, by developing a Python-based web application with an accessible Tkinter interface tailored for elderly users
- Implemented JSON-based data storage to enable efficient read/write operations and offline admin access

PROFESSIONAL SKILLS

- **Programming Languages:** Python, Java, JavaScript, TypeScript, SQL, C, MATLAB, HTML, CSS
- **Frameworks, Libraries & Tools:** FastAPI, Django, React.js, Node.js, Pandas, NumPy, Bootstrap, Tkinter
- **Databases & Platforms:** MySQL, PostgreSQL, SQLite, AWS, Docker, CI/CD, Jira
- **AI & Machine Learning:** BERT, OpenCLIP, LDA, NLTK, LLaVA, Hugging Face, PyTorch
- **Leadership:** Events Exec of CHEFS, Sports Analysis Lead of DS @GT, ML Pipeline Team Lead of SynapseX