**John Doe**

Greater NYC Area | JohnDoe@gmail.com | L[inkedIn.com/in/yasinehsa](http://linkedin.com/in/yasinehsan/)n | [GitHub.com/YasinEhsa](https://github.com/YasinEhsan)n | [YasinEhsan.com](http://www.yasinehsan.com)

**EDUCATION** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Queens College, City University of New York (CUNY)**  ***Expected Graduation: Aug 2025***

Bachelors of Arts in Computer Science | 3.9 GPA

**SKILLS** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Languages** *Proficient* JavaScript(5yrs) · Python(3yrs) *Intermediate* Swift(3yrs) · Java(1yrs) *Beginner* SQL(1yrs) · PHP(2yrs)

**Software**  AWS · GCP · Langchain · Docker · Kubernetes · TensorFlow · JupyterNotebook · iOS · Bootstrap · Git · Bash

**EXPERIENCE** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Software Engineering Intern *Aug 2024 - Present***

FUZZY BUZY AI *New York, NY*

• Creating 6+ web services for AI chatbot cutting cloud costs by 10% for 1B+ monthly replies

• Building framework bridging customer APIs with Snowflake streams run on AWS Lambda

• Owning 12+ APIs in servicing components with Typescript backend, React frontend

• Leading component to production, deploying using container orchestration, Jenkins CI/CD

• Implementing visibility of AWS resources placing triggers when ML models in dev overspend

**Software Engineering Resident  *Sep 2024 - Present***

HEADSTARTER *New York, NY*

• Built 14+ machine learning, ai-engineering and full-stack projects in fast-paced software team environments

• Developed 5+ neural networks in Python, 11 apps in Typescript on AWS/Vercel with dev and production environments

• Implemented llm-chaining, hyperparameter tuning, fine tuning on 10+ LLM models controlling for latency & accuracy,

• Coached by Google Machine Learning, Google Kubernetes, Two Sigma, Tesla, Figma and Citadel Engineers

• Created 321+ commits on github with 7-day deadlines getting a Career Capital increase of 40% from start date

**Software Engineering Intern**  ***Feb 2023 - Aug 2024***

RILLA VOICE *Dallas, TX*

• Built 5+ iOS features pushed to production resulting in $1Mn+ in auto- loans within 8 weeks of launch

• Developed geo-fences on map around bookmarked cars which trigger local notifications when breached

• Created an algorithm in Swift to optimize user movement; a solution around Apple’s location limitations

**Software Engineering Tutor *Apr 2023 - Jun 2023***

CS CLUB / CS DEPT / ONLINE TEACHING *New York*, *NY*

• Spearheaded cloud workshops on compute/storage/databases as part of 200 ambassadors nationwide

• Distributed $15,000+ computing credits to 300+ students/professionals on open-source and agile apps

**AI PROJECTS** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**US-Bank Churn Prediction** | Team Project *(~30 hours)*  - [Github da fudge](https://bit.ly/US-Bank-Churn-Prediction) ***Jan 2025 - Feb 2025***

• Used 30k+ data set, Llama 3.1b, Groq and Vercel to evaluate accuracy of predicting when banking customer quits

• Created an end-to-end solution complete with sending automated personalized email to banking customer based

on feature engineering, normalization, model training, evaluating and hyperparameter tuning across 5 LLM models

**Brain Tumor Classification**  | Open-Source 166 stars *(~10 hours)*  - [Github hehe](https://bit.ly/Brain-Tumor-Classification) ***Nov 2024 - Dec 2024***

• Used neural networks in Python to classify 1000 MRI scans into 3 types of possible brain diseases with custom model

• Generated multimodal MRI reports for neurosurgeons in under 200MS after image classification, construction & training

**Vercel v0 Clone**  | Team Project & Open-source *(~30 hours)*  - [Github MuaHaHa](https://bit.ly/Vercel-v0-Clone) ***Oct 2024 - Nov 2024***

• Used open-source libraries in Typescript and LLM-chaining in Python to dynamically craft react assets for a brand

• Defined constraints to **automate front-end** engineering needs like landing pages, contact forms, login pages to dynamically create and match the theme of popular brands and icons like Duolingo, Talor Swift, Apple, etc

**Discord Live Video** | Published Web-App 1.6K Users *(~40 hours)*  - [Landing Page NO click](https://bit.ly/Discord-Live-Video) ***Sep 2024 - Oct 2024***

• Used AWS, WebRTC, React and Vercel to deploy real-time chat and video functionality of popular gaming service

• Built 2+ production backend services that handle upto 10 real-time conversations processing 100GB of video every second

**Model Training Infra** | Published AI-App *(~80 hours)*  - [AI demo NOT for you](https://bit.ly/Model-Training-Infra) ***Sep 2024 - Sep 2024***

• Created 5 iterations of 2-3 recurrent neural network (RNN) layers to identify the ideal model variant before overfitting

• Developed an algorithm to reduce loss function with the number of layers and units saving RNN model training time by 200%

**Activities** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

AI Blogs - Published 5 blogs on AI experiments and trade offs with inference / tuning ***Oct 2024 - Present***AI Hiring Hackathon - Top-3 in vetted hackathon after building a multimodal agent for Bland, Retell and Rilla  ***Dec 2024***

Technical Keynote - Presented to software engineers on inference and neo4j databases for embeddings ***Nov 2024***