

# Shifan Chen

(646) 724 -3792 | lovekano233@gmail.com | Brooklyn, NY | <https://github.com/csf233csf>

## EDUCATION

### CUNY Baruch College

BBA in Computer Information System

Manhattan, NY

August 2020 – August 2023

### CUNY Hunter College

BA in Computer Science

Manhattan, NY

Starting August 2024, Expected Graduation: June 2026

## SKILLS

**Programming:** Python, C#, JavaScript, Java, HTML, Rust, C++, TypeScript

**Technologies:** Git/GitHub, Linux, Pytorch, Firebase, MongoDB, Paddle, Vue.js, Tailwind, MySQL, Node.js, Flask, Burp

**Languages:** Mandarin (Native)

## PROJECTS

### The UFO Project

A full-stack website for alien enthusiasts, serving as a small community with a co-creation platform.

- Built with Vue3 in TypeScript and fully composition API. Hosted with Cloudflare CDN and a Linux Server with Nginx for reverse proxy. Firebase is used for user co-creation Realtime Database storage.
- For the artistic part of this website, GSAP and Three.js are used for its model animations and user interactions.
- The users can upload/draw their own picture/photo to become a piece of this website.
- <https://www.thelighthouseproject.life/> (Site) <https://github.com/csf233csf/UFO-project> (Code)

### RAG is all you need

A simple and research-oriented RAG System with WebUI to run locally. Supporting SOTA models from Transformers.

- Built with LangChain (RAG), HuggingFace (Models), Streamlit (Web), and FAISS (Vector Database).
- Documents and PDFs uploaded to the system will be stored in FAISS as vectors. The user can query or ask logical questions about information in the documents uploaded. The AI will be able to query and answer them based on the information provided.
- Implemented Semantic Document Chunking by combining dense and sparse methods. Improved the document chunking.
- Added Kmeans Clustering Algorithm to cluster documents and vectors. Allowing the user to query the clusters.
- Detailed and highly customizable settings such as temperature, topK, repetition penalty, max token, and system prompt.
- <https://github.com/csf233csf/rag-is-all-you-need>

### Nike Account Generator

A fully automatic script to generate Nike Accounts using Selenium and C#.

- A Selenium-based web scraper that generates nike accounts automatically.
- Bypasses the akamai antibot detection using random delays and clicks.
- Bogus is used for random User-Agent Generation.
- SMSCODE API was implemented but not in public code.
- <https://github.com/csf233csf/Nike-Account-Generator-written-in-Csharp>

## RELEVANT EXPERIENCE

### Sigma AI (J.D. Technology Building Office)

Beijing, China

Machine Learning Engineer Intern

Feb 2024 – Jun 2024 (4 months)

- Collaborated with China's largest online educational platform to process millions of academic documents on their website every week, integrating them into Machine Learning tasks.
- Trained BERT models for various NLP tasks, including named entity recognition, sequence tagging, and text classification. Resolved issues related to missing content and special paddings in the BERT tokenizer.
- Developed complex Regex patterns for semantic chunking verifications, and significantly improved HTML document processing speed by rewriting the HTML parsing pipeline.
- Integrated APIs from closed-source LLM models into the task pipeline and created a JSON format checker to ensure the correct return format from the LLMs.
- Gained a deep understanding of how to scientifically build software, including stages like system design, coding, unit tests, and refactor. Learned how to handle large volumes of irregular and real-world data, including how to clean, preprocess, and visualize them.

### The Oil Technology

Remote

Group Administrator and developer

Aug 2018 - Dec 2020

- Worked as a group admin for an online community group with 500+ paid members.
- Took part in various projects that involved website development, web scraping, discord api integrations, git version controls and server hosting.
- Gained skills such as Teamwork, Self-Learning, and Leadership.