

# Shifan Chen

(646) 724 –3792 | SHIFAN.CHEN25@myhunter.cuny.edu | Brooklyn, NY | <https://github.com/csf233csf>

## EDUCATION

**CUNY Baruch College**

BBA in Accounting

**CUNY Hunter College**

BA in Computer Science

Manhattan, NY

August 2020 – August 2023

Manhattan, NY

August 2024 – June 2026

## SKILLS

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

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

**Languages:** English, Mandarin

## RELEVANT EXPERIENCE

### Sigma AI

Machine Learning Engineer Intern

**Beijing, China**

Feb 2024 – Jun 2024

- Processed millions of academic documents collaborating China's largest online education resource platform. Converting these raw data into Json formats and integrating them into Machine Learning tasks.
- Trained BERT models for various NLP tasks, including named entity recognition, sequence tagging, and text classification. Solved issues related to missing content and special paddings with the BERT tokenizer in the task of Sequence Tagging.
- Constructed complex Regex patterns for semantic chunking verifications, and significantly improved HTML document processing speed by over 90% while rewriting the HTML parsing pipeline.
- Integrated APIs from closed-source LLM models into the task pipeline and integrated a JSON format checker to ensure the correct return format from the LLM outputs.
- Gained a deep understanding of how to scientifically build and test software, including development stages like system design, debugging, unittests, and code-refactor. Learned how to handle large volumes of irregular and real-world data, including how to preprocess, convert, and visualize raw shape data.

## PROJECTS

### RAG is all you need | Python

A succinct 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>

### The UFO Project | Typescript

A Full-Stack project for a multi-media platform, serving as a creative community for alien enthusiasts.

- 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)

### Nike Account Generator | C#

A fully automatic script to generate Nike Accounts using Selenium.

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