

Sai Ram Chetla

Cleveland, USA | (330)-906-9960 | chetlasairamjp@gmail.com | linkedin.com/in/sai-ram-chetla/
<https://github.com/chetlasairam> | <https://chetlasairam.netlify.app/>

SUMMARY

AI Engineer

3+ years of professional work experience in the areas of Python Programming, Artificial Intelligence, Machine Learning.
Proven ability to seamlessly **adapt new tools and technologies** while actively pursuing ongoing skill acquisition.

EDUCATION

Master of Science, Computer Science

August 2022 – December 2023

Kent state university, Kent, Ohio, US - **GPA: 3.85**

Bachelor of Technology, Electronics and Communication

August 2017 - May 2021

VNR Vignana Jyothi Institute of Engineering and Technology, Telangana, India. - **GPA: 3.78**

EXPERIENCE

MRI Software – Cleveland, Ohio, USA

May 2024 – Present

Software Engineer

- Partnered in the creation of **AI chatbot (RAG)** tailored for **real estate software**, specializing in lease agreement comprehension and analysis, Ensuring optimal functionality with a focus on **cost efficiency** and **token count**.
- Upgraded **ES 1.7 term search index** to **ES 8.8** and implemented **Reciprocal Rank Fusion (RRF)** for **hybrid search**.
- Developed an optimized relevant segment finding solution using **SpaCy** and **FastEmbed**, leveraging **Azure Storage Class PVC** to reduce service pod initialization time by **40%**.
- Crafted a **React and Node.js-based search UI for Elasticsearch**, customizing the backend for advanced **vector search and hybrid search** capabilities, enhancing file search precision beyond standard term querying.
- Created Python script to index millions of documents into Elasticsearch, extracting text from PDFs using **OCR** and **vectorizing** the content with **OpenAI's text-embedding-3** model. Included relevant metadata to optimize search and retrieval performance.
- Partnered in an **AI system with GPT-4o** to automate the extraction of lease details from PDFs, utilizing **Pydantic functions** for accurate data validation and innovative prompt crafting for field-specific retrieval.

AMBIQUEST – Hyderabad, India

Software Engineer

Aug 2021 – Aug 2022

Intern

Aug 2020 – Aug 2021

- Developed a **cross-platform** mobile application using **Flutter** for seamless experiences on **iOS** and **Android** devices. Integrated **machine learning** models using **TensorFlow Lite** for on-device **activity recognition** and personalized **workout recommendations**.
- Implemented **computer vision** techniques with **OpenCV** to analyze and track user movements in real-time, and utilized **SQL databases** for efficient data storage and management, ensuring scalability and performance.
- Designed **RESTful APIs** for secure communication between the mobile app and **cloud-based** services, and automated deployment and **continuous integration** processes using **Bash scripts** and **CI/CD pipelines** on **GitLab**.
- Deployed the backend infrastructure on **AWS** using **EC2** instances and **RDS** for the database, leveraging **S3** for data storage. Applied advanced **data analytics** techniques to interpret user data and provide actionable insights.
- Ensured data security and user privacy by implementing **encryption standards** and robust **authentication protocols**.

PROJECTS

Kent Connects: Developed a **React.js** social site with **HTML, CSS, JavaScript, Firebase SDK (Authentication, Realtime Database, Firestore, Storage)**, version control with **Git**, and deployed via **Netlify/Firebase CLI**.

Husk: Built a **Flutter** Android chat app with **Firebase Cloud Messaging, Postman** for push notifications, **ZegoCloud/Agora** for video/voice calls, and **Firestore DB** for real-time message handling.

Face Mask Detection: Created a real-time detection system using **OpenCV, TensorFlow, PyTorch** for **CNN** training, **Flask/Django** web interface, optimized with **NVIDIA CUDA** for **GPU** acceleration.

Get Dish App: Designed a **Figma** mobile app for college canteen orders, focusing on **UI/UX** to enhance the dining experience.

SKILLS

Languages:	Python, C, C++, Java, HTML, JavaScript.
AI tools:	GPT 4o, Llama, Open AI, Azure ML, Lang Chains, Hugging Face, AWS Sage maker.
Database:	MySQL, Snowflake database, Vector databases (Elastic Search, FAISS), AWS (cloud)
Others:	Machine Learning, Deep Learning, FAST API, Kubernetes, Azure AI, Docker, REST API

CERTIFICATIONS

- **Generative AI For Beginners - Google Gemini & Google Cloud**
- Programming, **Data Structures** and Algorithms using **Python**
- Fundamentals of **Generative AI, Azure Open AI** service