International Journal of Research in Engineering, Science and Management
Volume 9, Issue 4, April 2026
www.ijresm.com | E-ISSN: 2581-5792 | RESAIM Publishing | www.resaim.com

CareChain AI: An AI Powered Universal Health Record System

Prithviraj Kshirsagar[1*], Sudesh Jambhulkar[1], Tanishq Dhote[1], Krish Rawat[1], Nitish Das[1]
[1]*School of Computing, MIT Art, Design and Technology, Pune, India*

***Abstract*: Due to the fact that healthcare data is often spread out across many hospitals (as well as laboratories) and in many different formats (i.e., PDFs, images, handwritten prescriptions), it is often difficult to obtain urgent/critical patient information during emergencies. This fragmentation of data may cause delays in diagnosis and inefficient clinical decision making; additionally, patients are not as engaged in their own healthcare when the data they need to manage their own health is fragmented. Current digital health systems only focus on storing data, without providing the contextual intelligence and safety mechanisms that can assist users in understanding complex medical data. In response to these issues and challenges, this paper presents CareChain AI, an AI-based universal healthcare record management system that serves both to consolidate and intelligently process lifelong patient medical records in a secure (encrypted) and patient-centered manner. CareChain AI integrates Optical Character Recognition (OCR) functions to extract text from various documents, along with a Retrieval-Augmented Generation (RAG) framework to allow for natural language queries from a user-friendly interface. As text is extracted from documents, it is converted into vector embeddings and stored using FAISS indexes for highly efficient document retrieval methods.**

***Keywords*: Healthcare AI, Universal Health Records, Retrieval-Augmented Generation (RAG), Optical Character Recognition (OCR), FAISS, Natural Language Processing (NLP), Secure Data Management.**

1. Introduction

The digitization of healthcare systems at a rapid rate has allowed for many patient records to be collected (examples: clinical reports, laboratory work, medical images, and prescriptions). Many times, the patient information is disparate amongst different healthcare providers using heterogeneous formats (examples: PDF Files, scanned images, and written notes). Because of this, when a health provider is trying to access a data history about a patient in times of great need, they often cannot get all the information in a timely manner. This can create delays in a health care provider's ability to diagnosis and/or treat a patient.

In addition, a patient doesn't often have enough medical knowledge to read and properly understand the complex reports that the doctor provides for them. Therefore, when a patient or his/her provider utilizes traditional Electronic Health Record Systems (EHR), the system is mainly storage and retrieval based; as such there is little intelligent functionality to provide useful insights, or meaningful and contextual explanations associated to the patient and/or his/her diagnosis.

Furthermore, EHR Systems lack the ability to meet interoperability, real-time access, and provide secure emergency access to medical records. With a growing need for solutions that are centered around the patient who receives care from multiple health systems, there is a growing demand from consumers for solutions that deliver centralized medical records that are also user-friendly and easily accessible over the internet.

To solve these problems, we will be implementing an AI-based UHR (Universal Health Record) system called CareChain AI which utilizes OCR (Optical Character Recognition) technology to convert documents into readable text, NLP (Natural Language Processing) to analyze the content of each document, and RAG/13703 (Retrieve Augmented Generation) technologies to assist users with extracting the key data from their medical records and creating a chronological timeline regarding their overall health. Users may also submit natural language queries of their medical records to obtain access quickly and easily.

The CareChain system has multiple built-in security features including secure access protocols for physicians vs. non-physicians, and the ability for emergency responder personnel to have controlled access to patient information while at the same time the healthcare providers' privacy will be maintained as well as providing relevant information to facilitate better decision support for treating patients.

A. Problem Statement

Hospital, laboratory, and clinic patient records are frequently fragmented throughout the healthcare industry and they are very often stored in unstructured format (PDFs, images, handwritten prescriptions, etc.). As such, getting complete information about a patient's medical history is often difficult during emergencies, resulting in delays in diagnosing or treating a patient.

Additionally, patients have difficulty understanding many of the complex medical reports they receive. Existing systems are typically focused on just storing information and are lacking in intelligent (or smart) ways to interpret data and apply security to ensure compliance, creating a need for one integrated

*Corresponding author: prithvikshirsagar007@gmail.com

solution for health record management that is intelligent, secure, and provides access to data in a meaningful way while respecting data privacy.

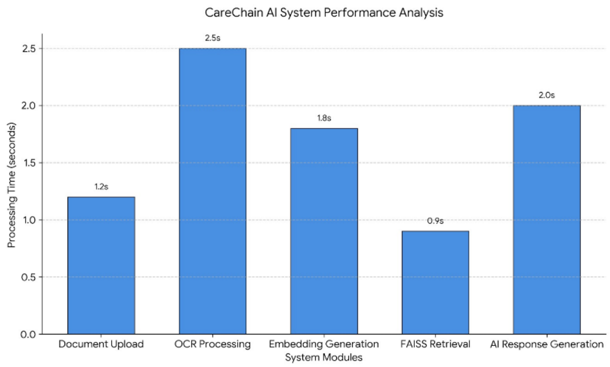

Fig. 1. Performance analysis

B. Objectives

The main goal of this project is to establish a novel, secure system for the management of medical records (known as patient records). Project objectives include the improvement of the access, organization and comprehension of medical data through the use of artificial intelligence (AI) methods.

- Create an integrated system to allow for storage and access of medical records.
- Use of Optical Character Recognition (OCR) and Natural Language Processing (NLP) to automatically capture and organize data.
- Allow for natural language search emergency retrieval capabilities through the use of Robotic Process Automation (RPA).
- Create a chronological health history for each patient.
- Provide a strong security model for storing patient records while at the same time being able to provide access as required during emergencies.

2. Literature Survey

1) Electronic Health Record (EHR) Systems

EHR systems are digital platforms designed to store patient medical history, diagnoses, treatments, and reports in electronic form. They reduce paperwork and improve data management efficiency. However, most EHR systems are limited to individual hospitals and lack interoperability.

2) Hospital Management Systems (HMS)

HMS applications manage patient records, billing, and appointments within a hospital. These systems improve operational efficiency and reduce manual work. However, they are limited to a single organization and do not provide a unified view of a patient’s complete medical history across multiple hospitals.

3) Cloud-Based Health Systems

Patients can access their medical records remotely through a cloud platform that provides both increased availability and scalability of those records. In addition to increasing convenience, the cloud adds new challenges of data privacy and security risk along with reliance on an Internet connection for access.

4) Optical Character Recognition (OCR)

OCR allows for the conversion of scanned documents/images into machine-readable text. However, OCR technology may produce inaccurate results when trying to read handwritten text, or when there is poor-quality or complex layout of the scanned document/image.

5) Natural Language Processing (NLP)

The use of natural language processing methods aids in the processing of large quantities of unstructured medical information so that you can identify key entities like disease, medication or symptom. This means that it allows the raw medical data to be turned into a structured, usable kind of data to be used for analysis.

6) AI Chatbots in Healthcare

The usage of automated AI chatbots is extremely common today as a source of basic health care goodies and with retrieving responses to users' inquiries about the service. There is an inherent risk of the AI sending an inaccurate or unsafe response if no measure has been taken to validate and monitor processes that manage the accuracy of the database from which the AI is accessing.

7) Retrieval-Augmented Generation (RAG)

The Retrieval-Augmented Generation (RAG) methodology combines text information retrieval along with text generation to enable context-based responses generated from accurately retrieved data. RAG employs a database to locate and secure data and generate the respective response for that record.

8) Vector Databases (FAISS)

Fast Approximate Nearest Neighbors (FAISS) methodology employs an analytic process and supports an environment conducive to efficiently performing high-speed similarity searches on large amounts of vector embeddings and hence provides rapid retrieval for all relevant data elements.

9) Data Security Techniques

The use of encryption, authentication, and access control are key security mechanisms to protect sensitive health care data. These techniques provide confidentiality, integrity, and controlled access. Proper security mechanisms provide a foundation for building trust in health care systems.

10) QR-Based Emergency Access Systems

QR-based systems can provide quick access to essential information about a patient should something happen in an emergency. They provide fast access but do not require full login credentials, but they must ensure that there is limited amount of time the data is accessible through the QR code and that there are secure methods to access this data to prevent misuse of the sensitive information or the data associated with the QR code.

3. Proposed System

CareChainAI is proposed to improve and provide a better structured system for handling the fragmented, unstructured data that exists in today's healthcare environment. This is accomplished by providing users with a place to upload any of the many types of documents having to do with a patient's healthcare (PDF, images, etc.) for the purpose of processing those documents with Optical Character Recognition (OCR) to extract text from those documents.

Once the text from the documents have been extracted, we will then "organize, and structure" the text so that it can be accessed easily. Natural Language Processing (NLP) in combination with a Retrieval Augmented Generation (RAG) method, will allow patients to use natural language to query their patient records. We will use a Vector database (FAISS) to store the vector representation of the data we extracted to allow for retrieval of information that matches the user's query quickly and accurately.

Additionally, we will provide patients with a timeline of their health care history so that they can track their health care history over time. We have incorporated an ability to identify the validity of the information returned from the AI by providing a means to filter out any false results, as well as other safeguards for security (like authentication and controlled access in case of an emergency through QR codes).

The CareChain AI System has been created to solve some of the difficulties created by fragmented and unintegrated health care data, as well as to develop an inclusive and intelligent data system that will be driven by artificial intelligence (AI), combining all of a patient's health care-related data throughout their life, i.e. all of a patient's health care-related data, will be available to them in one place.

With this platform, users will be able to upload their medical-related documents (e.g. PDF's, photos, prescription) into the CareChain AI application. The documents are processed through Optical Character Recognition (OCR) when they are uploaded to CareChain AI. After OCR processes the documents, text will be created and the text is cleansed through the use of natural language processing (NLP) methods.

Furthermore, data security and patient privacy will be ensured through user authentication, encryption of data at rest and with role-based access control over each user's record. QR codes will allow rapid access to patient data in the case of an emergency. Medical professionals will be able to quickly access critical patient data without compromising data security.

Overall, the proposed CareChain AI project will benefit healthcare providers by enhancing their access to healthcare data by providing both increased accessibility of healthcare data and improved usability of healthcare data.

A. System Architecture

CareChain AI's Architecture consists of three layers for managing and processing patient medical records. The user interface layer is considered the top layer in the design. This is where the patient, physician, or other end user can upload a patient medical record into the application and also perform any other function related to the patient using a web-based application**. I**ngestion layer is used to record the upload of patient medical records in many different file formats (PDF, image, etc.).

The data ingestion layer will also have an Optical Character Recognition (OCR) Module; this module will identify the text and extract text from the scanned images of patient medical records and then store that text digitally (and will eventually be stored on paper).

An important function of the CareChain AI architecture is to utilize many different security layers, such as user authentication, access control and the secure storage of patient medical records, in order to provide secure and quickly accessible patient medical records for patients.

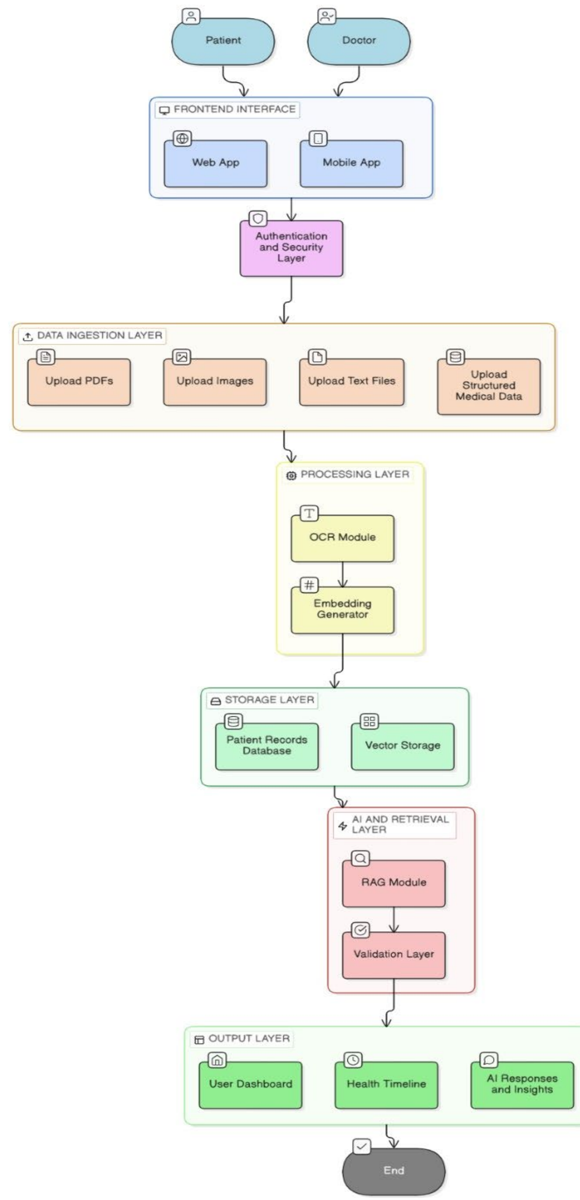

Fig. 2. System architecture

B. Methodology

CareChain AI was created as a way to process and manage healthcare-related documents, allowing customers to upload their healthcare documents in either a PDF or picture format. After uploading a document, CareChain will use optical character recognition (OCR) technology to convert the text from that document into a machine-readable format.

System will use vectorization technology to create what are known as "vector embeddings" of these documents. These vector embeddings will then be stored in one of CareChain's primary databases and/or in a third-party vector service (e.g. FAISS) for the purpose of providing fast, reliable search results.

Steps:

- Upload medical documents
- Extract text using OCR
- Convert text into embeddings
- Store data in database and FAISS
- Retrieve information using RAG
- Display response to user

C. *Implementation Details*

The CareChain AI system is built in a way that includes everything from the front to the back. This means it combines the parts that users see the parts that do the work, behind the scenes and the parts that use intelligence to get data. The front part of the CareChain AI system is made using Next.js, TypeScript and Tailwind CSS.

FastAPI to build the backend APIs for your site allows for easy integration between your mobile application and your relational database, either SQLite or PostgreSQL. These will act as the relational databases' backends to allow access to the relational databases through SQLAlchemy. When it comes to processing documents, you can use libraries like pdfplumber, pytesseract, or Pillow for text extraction from PDF files or images with optical character recognition (OCR).

Following the text's extraction, pre-processing will occur via an embedding model (e.g., a Text to Vector model) to create embeddings/representations; these will be stored using systems such as Facebook AI Similarity Search (FAISS) – both for performing and accessing similar to available medical records.

An interface with an AI will also provide access via the Gemini API to a Retrieval Augmented Generation (RAG) Model for the generation of responses based upon the context of the stored medical records.

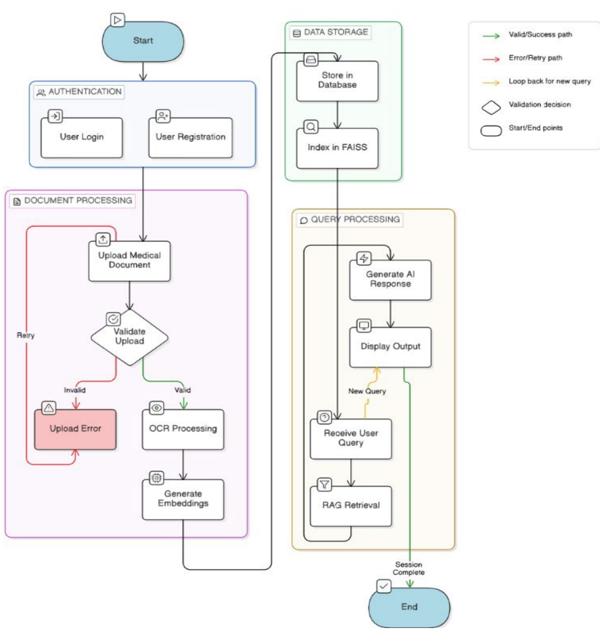

Fig. 3. Implementation flow

D. *Evaluation*

We checked if the CareChain AI was really quick and accurate and whether it was user-friendly in retrieving healthcare records by uploading a variety of types of medical documents (e.g., images, PDFs) to see if the system's OCR would be able to extract text correctly from those uploads. CareChain AI succeeded in locating and organizing most of the important medical data. We then attempted to ask questions on the data to test whether or not the CareChain AI could retrieve the requested data and provide a clear description of it. CareChain AI was very fast and efficient in retrieving and answering questions due to the use of Retrieval-Augmented Generation and FAISS vector storage.

Overall, it made finding and understanding health records a lot easier. Users get better access to their own data and can pull up what matters most without a hassle. It includes security features to protect sensitive, health info. Honestly, the system gets the job done.

4. Results

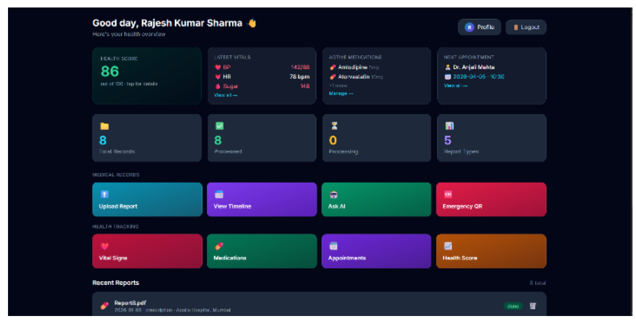

Fig. 4.

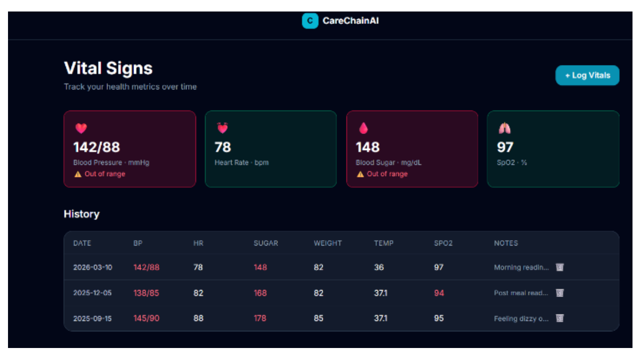

Fig. 5.

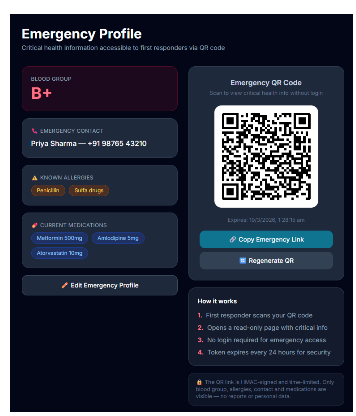

Fig. 6.

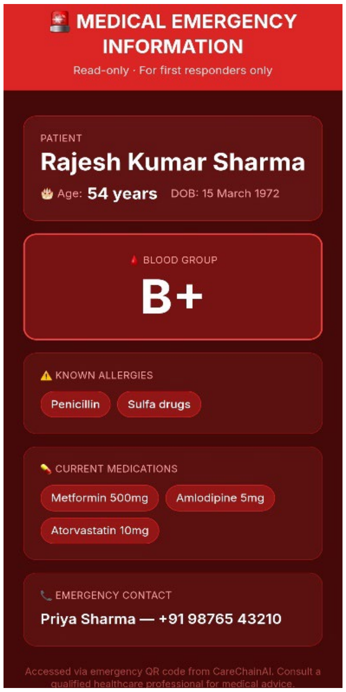

Fig. 7.

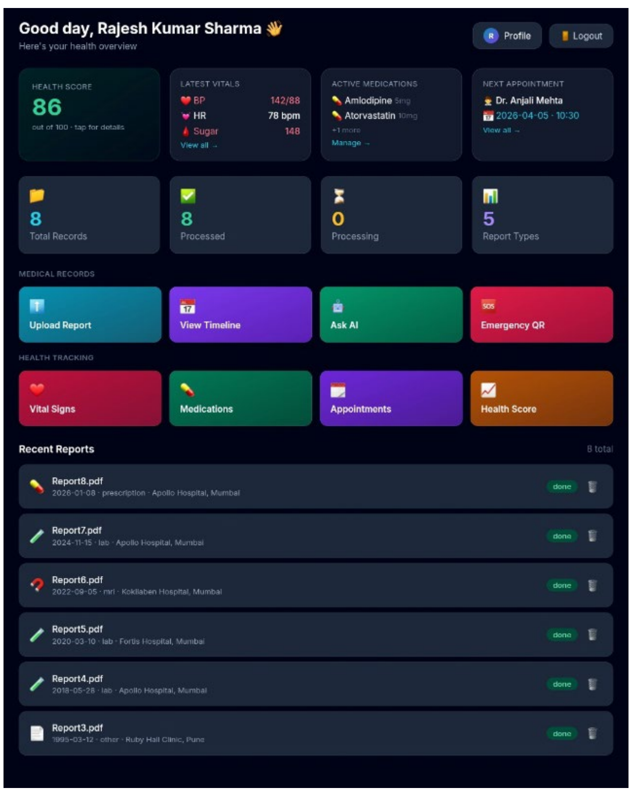

Fig. 8.

5. Discussion

CareChainAI is an example of integration of artificial intelligence with management of healthcare data to create an innovative approach to extracting and organizing healthcare information from multiple sources.

By taking advantage of a combination of tools like optical character recognition (OCR), vector databases using artificial intelligence-based retrieval methods.

CareChainAI system allows for a quick retrieval and organization of medical information from a database based on the type of documents used. This method solves the issue of having health records distributed among multiple locations and provides users with a single, universal location to manage their medical information.

Furthermore, when charters respond to a user's inquiry via Retrieval-Augmented Generation (RAG), the data retrieved for the inquiry will be much more accurate than if it were simply produced based on what the system previously learned, as occurs in older artificial intelligence systems. In addition, CareChainAI protects user privacy by employing stringent security protocols to protect the private nature of the user's medical records.

CareChainAI is an example of why Artificial Intelligence combined with the management of healthcare data is beneficial. We collect and organise patient health information from various types of documentation quickly by using tools like Optical Character Recognition (OCR), Vector Databases, and Retrieval Processes and systems.

Retrieval-Augmented Generation (RAG) is a process of searching and gathering information before providing an answer. This allows the AI system to generate more accurate responses as the information that AI is based on must be referenced against historical documents before being presented to the user. The addition of RAG to the process also increases the perceived reliability of the AI response since previous AI systems did not have the benefit of being able to cross-reference learned knowledge with accurate historical documents.

The quality of the documentation that is entered into the system will dictate how accurately the text is extracted from the uploaded documents. All AI-generated responses produced by this system must be validated before being presented to the end-user to protect from the misrepresentation of errors as factual information. With advances in OCR, direct connectivity to the hospital's database, and the development of better graphical models of the data, CareChainAI can be further leveraged to provide better analytical tools for evaluating trends in healthcare. It is clear that the implementation of an AI system will greatly improve how we can access and manage healthcare data.

The system also improves patient engagement by allowing users to understand their health data through simple questions or queries. This can help patients understand their medical history and conditions better. However, challenges such as OCR inaccuracy for low-quality document and dependence on AI models to generate responses may still exist.

A. Limitations

- API Rate Limits: Gemini free tier allows only 200 requests/day, limiting scalability.
- OCR Accuracy: Tesseract struggles with handwritten prescriptions and low-resolution scans.
- Information Extraction: Regex-based extraction fails on non-standard date formats and regional conventions.
- Scalability: AISS flat index and SQLite are not

suitable for large-scale production deployment.

- Security: Two-factor authentication and document encryption are not implemented.
- Doctor-Patient Linking: Direct record sharing between doctors and patients is absent.
- Medical Disclaimer: AI responses are informational only and not validated by medical professionals.

B. Future Scope

- Develop a mobile or web-based application for easier user access and real-time predictions.
- Use cloud platforms for scalable data storage and faster processing of large datasets.
- Expand the system to handle more input parameters and complex datasets.
- Improve security and privacy mechanisms for safe data handling. User Feedback & Personalization: Creating a feedback process to get user feedback for improving accuracy and tailor conversations with users based on their history.
- Application-Based: Develop a separate application to improve usability and extend usage range.

6. Conclusion

CareChainAI system provides an AI based record management system that focuses on the needs of patients. The system uses multiple modes of document processing, RAG-question answering, and agent orchestration. There is also a Critic Agent which ensures that each of the five agents in the pipeline is following all safety protocols and automates the extraction, classification and querying of records. To maintain the confidentiality of the patient's sensitive information, the vector retrieval system is scoped to the user only; therefore, the patient will not have to log onto the system when they would want to access vital health records using the QR mechanism. The premise of the CareChainAI solution is to demonstrate that both safety and new AI technologies can be combined to create a platform for managing health records.

Future work includes the addition of a medical NER model, cloud-scalable vector storage, sharing doctor-patient health records and HL7 FHIR integration of hospitals.

References

[1] A. Vaswani *et al.*, "Attention is all you need," in *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, 2017, pp. 5998–6008.
[2] J. Johnson, M. Douze, and H. Jégou, "Billion-scale similarity search with GPUs (FAISS)," *IEEE Trans. Big Data*, vol. 7, no. 3, pp. 535–547, Sep. 2017.
[3] T. Brown *et al*., "Language models are few-shot learners," in *Proc. Neural Inf. Process. Syst. (NeurIPS)*, 2020.
[4] R. Smith, "An overview of the Tesseract OCR engine," in *Proc. Int. Conf. Document Anal. Recognit. (ICDAR)*, 2007, pp. 629–633.
[5] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using siamese BERT-networks," in *Proc. Conf. Empirical Methods Natural Language Process. (EMNLP)*, 2019, pp. 3982–3992.
[6] FastAPI Documentation. "FastAPI: Fast code, high performance, easy to learn, production ready." [Online]. Available: https://fastapi.tiangolo.com
[7] FAISS Documentation. "FAISS: A library for efficient similarity search and clustering of dense vectors." [Online]. Available: https://faiss.ai
[8] Google Generative AI (Gemini) Documentation. "Google Generative AI developer site, including the Gemini API and Gemma." [Online]. Available: https://ai.google.dev