UNIVERSITÀ
DI PAVIA

Bachelor Thesis

HealthVision AI: Leveraging Multimodal Large Language Models for Streamlining Health Insurance Claims and Expense Reimbursements

Andrea Lolli
Artificial Intelligence

Academic Year 2023/2024

Bachelor in Artificial Intelligence
Pavia 2024

Bachelor in Artificial Intelligence

Andrea Lolli

HealthVision AI: Leveraging Multimodal Large Language Models for Streamlining Health Insurance Claims and Expense Reimbursements

Supervisor: **Matteo Zignani, University of Milan**

Acknowledgments

First and foremost, a heartfelt thanks to my family and the relatives who gave me this opportunity, who had blind faith in me and opened up possibilities I could never have imagined.

I extend my gratitude to my friends who have been with me during these three years, through various trains and trips, between the humidity of Pavia and the fog of Milan. To those who followed me from afar and encouraged me, with whom I shared successes and defeats.

A special thanks to the Accenture team, who allowed me to have this wonderful first work experience in the industry in a healthy and respectful environment. To the colleagues who guided me throughout the journey and lightened the burden of the office between coffee breaks.

I am thankful for your support and the love we have shared. I couldn't have asked for better companions than you. Life is good.

Everbody wanted to know what I would do if I didn't win.
I guess we'll never know.

-

Abstract

This thesis explores the development and implementation of HealthVision AI, an innovative system designed to enhance the accuracy and efficiency of health insurance claims and expense reimbursements through the integration of multimodal Large Language Models (LLMs). The primary objectives of this project include improving text recognition accuracy, streamlining document processing workflows, reducing operational costs, and enhancing user experience.

HealthVision AI leverages LangChain for creating LLM-based applications, Streamlit for rapid web app development, and Azure AI Services to augment GPT-4 Vision's Optical Character Recognition (OCR) capabilities. Initial performance tests have shown an 18% increase in OCR accuracy compared to existing solutions, particularly for handwritten documents and complex medical forms. This thesis provides a comprehensive overview of the project, including the internship context, a detailed literature review of OCR technologies and LLMs, system design and architecture, and the development and deployment processes.

The results section presents a thorough performance and cost analysis, demonstrating HealthVision AI's superior accuracy and efficiency despite higher costs compared to traditional OCR systems. This project significantly contributes to the field of artificial intelligence by demonstrating the practical application of LLMs in addressing critical OCR challenges. The thesis concludes with a discussion of the industry implications of HealthVision AI and potential areas for future research and development.

Keywords

HealthVision AI, Optical Character Recognition (OCR), Large Language Models (LLMs), GPT-4 Vision, Streamlit, LangChain, Azure AI Services, Health Insurance Claims, Expense Reimbursements.

Contents

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Overview

During my internship at Accenture, I developed HealthVision AI, a cutting-edge solution designed to revolutionize Optical Character Recognition (OCR) by leveraging advanced Large Language Models (LLMs). Traditional OCR systems often struggle with accurately interpreting handwritten text and documents in non-standard formats. HealthVision AI addresses these challenges by utilizing LLMs adept at processing and understanding a wide range of visual and textual data. This innovative approach transforms extracted text into structured formats, significantly enhancing data processing and validation. Furthermore, HealthVision AI's multilingual capabilities enable seamless recognition and processing of documents in multiple languages, making it a versatile tool for global applications[17][57].

HealthVision AI, while specifically optimized for medical and expense documents, is a versatile system that can adapt to handle a wide range of document categories. It has been rigorously tested with real data from a prominent Italian client, proving its effectiveness in processing health insurance claims and expense reimbursements.

The project takes a comprehensive approach, integrating cutting-edge technologies such as LangChain for creating LLM-based applications, Streamlit for developing user-friendly web interfaces, and Azure AI Services to augment GPT-4 Vision's OCR capabilities[57]. This approach ensures the robust and reliable performance of HealthVision AI. Initial performance testing has demonstrated a remarkable 18% average increase in OCR accuracy compared to existing solutions, highlighting the project's potential to revolutionize document processing workflows.

This thesis provides a comprehensive overview of HealthVision AI, including the internship context, a detailed literature review of OCR technologies and LLMs, system design and architecture, and the development and deployment processes. The results section presents a thorough performance and cost analysis, underscoring HealthVision AI's efficiency and costs of implementation. By demonstrating the practical applications of LLMs in addressing critical OCR challenges, this project makes significant contributions to artificial intelligence. The thesis concludes with a discussion of the industry implications of HealthVision AI and potential areas for future research and development.

1.2 Objectives

One of the primary goals of HealthVision AI is to significantly improve text recognition accuracy, particularly for handwritten documents and documents in non-standard

or complex formats. Traditional OCR systems often struggle with the variability in document formats and handwriting styles. By integrating advanced Large Language Models (LLMs) such as GPT-4, HealthVision AI aims to leverage deep learning capabilities that understand context and semantics, resulting in higher accuracy[44][63]. As noted by Vincent Perot et al. in their 2023 paper, "LMDX: Language Model-based Document Information Extraction and Localization," LLMs have revolutionized Natural Language Processing (NLP), improving state-of-the-art performance across various tasks[58].

Another key objective is to reduce the processing time required for health insurance claims and expense reimbursements. Efficient data extraction processes are essential for minimizing manual effort and speeding up workflows. HealthVision AI seeks to automate and streamline these processes, allowing for quicker and more accurate document processing. According to UiPath, AI-powered automation can significantly increase processing capacity and efficiency in healthcare document processing[74].

Developing a cost-effective solution that minimizes operational expenses compared to traditional OCR methods is also critical. Reducing the dependency on manual corrections and extensive training can significantly cut costs. The cost analysis will primarily focus on the expenses associated with calling LLMs and utilizing Azure resources for enhancing OCR capabilities. Traditional OCR methods often incur high costs due to the need for extensive manual corrections and specialized training. HealthVision AI aims to provide a scalable LLM-centric solution that is both efficient and economical. This cost reduction will be achieved through improved extraction accuracy, streamlined administrative processes, efficient resource allocation, and reduced manual data processing. Integrating AI and machine learning techniques in document processing has been shown to reduce operational costs significantly, as highlighted by Fintelite[30].

Another essential objective is enhancing the user experience by providing a user-friendly interface that facilitates easy interaction with the system. HealthVision AI employs Streamlit to create an intuitive web application, enabling users to easily upload, process, and extract information from documents. It is generally recognized in the industry that user-friendly interfaces are crucial for effective document processing, especially when dealing with multiple languages and formats.

Ensuring scalability and versatility is crucial for the success of HealthVision AI. The ability to process different types of documents effectively makes HealthVision AI a versatile tool. Initially optimized for medical documents, the system is designed to be robust and scalable to meet diverse document processing needs. Integrating advanced LLM prompting techniques ensures the system can adapt to new document types without significant performance degradation, instilling confidence in its adaptability. This aligns with the findings of HyperVerge, which emphasize the importance of adaptability in AI-powered OCR systems[36].

Leveraging multilingual capabilities is another significant goal. In a globalized world, the ability to handle multilingual text is essential. HealthVision AI integrates multilingual support within the LLM framework, ensuring accurate processing of documents in various languages. This capability is crucial for global clients who deal with documents in multiple languages. We are building on the advancements made by Google Cloud in OCR technology to ensure our system is at the forefront of multilingual document processing[32].

Finally, the project aims to create a system that does not require extensive training or document-specific knowledge for users. By enabling interactions with the system using natural language commands, HealthVision AI reduces the learning curve and makes the platform highly accessible. This feature allows users to seamlessly upload, process, and

extract information from documents without specialized skills. Integrating LLMs ensures that the system can effectively understand and respond to user queries.

1.3 Context of the internship

The internship at Accenture's data and AI department provided a unique opportunity to contribute to the advancement of optical character recognition (OCR) technology in the health insurance and expense reimbursement sector. The project, which was at the forefront of innovation, focused on an alternative solution to the existing "iOCR" software, previously developed by the team, for processing health insurance claims and expense reimbursements by integrating an LLM-centric approach to improve accuracy and efficiency leveraging GPT-4 Vision model.

Accenture, a global leader in professional services, has been at the forefront of implementing AI solutions in various industries, including insurance. The company's commitment to innovation and partnerships with leading technology providers created an ideal environment for exploring cutting-edge AI applications[4]. The internship involved close collaboration with a diverse and talented team of AI engineers, data scientists, and project managers. This inclusive setting facilitated knowledge exchange and guided the project's development phases.

To prepare for the HealthVision AI project, I conducted an in-depth study of various resources, encompassing a wide range of cutting-edge technologies and methodologies. This comprehensive research laid the foundation for developing a robust, scalable, and efficient AI-powered application. The key areas of study included:

- **Generative AI:** I examined the principles and applications of generative AI, focusing on:
  - Design principles for generative AI applications.
  - Unique challenges in user experience design for generative AI systems.
  - Strategies for implementing effective and safe generative AI applications.
  - Ethical considerations and potential societal impacts of generative AI technologies.
- **Multimodal Large Language Models (GPT-4):** I explored GPT-4 and its potential applications in OCR, including:
  - GPT-4 Turbo with Vision capabilities for advanced image analysis.
  - Integration techniques with Azure AI Vision for enhanced object detection and OCR.
  - Performance optimization and fine-tuning strategies for GPT-4 models.
  - Limitations and pricing considerations for GPT-4 Turbo with Vision in production environments[47].
- **Prompt Engineering:** I studied various techniques for optimizing prompts to improve LLM performance, such as:
  - Chain-of-Thought (CoT) prompting for complex reasoning tasks.
  - Few-shot and zero-shot learning approaches in prompt engineering.
- **LangChain Framework:** I investigated this framework designed to simplify the creation of applications using LLMs, focusing on:
  - Core components for building LLM-powered applications.

  - Integration strategies with various LLM providers and tools.
  - Best practices for developing scalable AI applications using LangChain.
  - Performance optimization techniques within the LangChain ecosystem.

- **Streamlit Framework:** I explored this tool for building and sharing data apps quickly and efficiently, emphasizing:
  - Rapid prototyping techniques for AI-powered web applications.
  - Integration methods with machine learning models and data visualization tools.
  - User-friendly interface design principles for AI applications.
  - Deployment strategies and performance optimization for Streamlit apps.

- **Azure AI Services for GPT-4 Vision Enhancements:** I studied how to leverage Azure AI services to enhance GPT-4 Vision's OCR capabilities, including:
  - Integration of Azure AI Vision with GPT-4 Turbo.
  - Advanced object detection and OCR capabilities within the Azure ecosystem.

- **Docker:** I explored containerization and deployment of applications using Docker, covering:
  - Best practices for containerizing AI applications for consistent deployment.
  - Efficient management of dependencies and environment variables in Docker containers.
  - Scalability and portability considerations for containerized AI solutions.

One of the biggest difficulties encountered during the internship was the rapidly evolving nature of technologies, frameworks, and modules for Generative AI. Given that these tools are relatively new and constantly being updated, finding resources that provided clear explanations on how to code specific parts was challenging. Often, the available documentation was either incomplete or outdated. This lack of comprehensive resources sometimes made it difficult to understand and implement certain features. To overcome these challenges, I relied heavily on collaboration and knowledge-sharing within the team. We engaged in regular brainstorming sessions and sought advice from more experienced colleagues. Additionally, I invested extra time in experimenting with the tools and frameworks, learning through trial and error. This hands-on approach, coupled with team support, allowed us to navigate the complexities and successfully implement the required functionalities.

The development followed the Model-View-Controller (MVC) framework, incorporating APIs through Azure Services and the LangChain framework for improved OCR capabilities. The user interface was developed using Streamlit, ensuring a user-friendly experience for end-users. Performance testing was a crucial aspect of the project, comparing the GPT-4 Vision model's accuracy with the previously developed model, iOCR, across various document types, including digital and handwritten texts. This evaluation focused on extracting information from medical and expense documents such as invoices and prescriptions.

The internship provided hands-on experience with state-of-the-art AI technologies and methodologies, offering valuable insights into AI's practical applications in the industry. This experience enhanced technical skills and provided a deeper understanding of the challenges and intricacies involved in developing and deploying advanced AI systems in a real-world context.

1.4 Outline

The rest of this thesis is organised as follows:

Chapter 2 presents the company and project context, detailing Accenture's role and the project's inception.

Chapter 3 provides a literature review covering the historical development and current state-of-the-art in Optical Character Recognition (OCR), Large Language Models (LLMs), generative AI, and multimodal AI systems.

Chapter 4 discusses the design and architecture of HealthVision AI, including system components, user interface design, technology stack, and the development and deployment process. It highlights the iterative testing, challenges faced, and solutions implemented.

Chapter 5 presents the cost analysis and performance evaluation of HealthVision AI, comparing it with the existing "iOCR" solution and discussing its cost-effectiveness and efficiency.

Chapter 6 concludes the thesis by summarizing the key achievements, discussing industry implications, and suggesting potential areas for future research and development.

Chapter 2
Presentation of the company and project

2.1 Company Overview

Accenture is a leading global professional services company that provides a wide range of services in strategy, consulting, digital, technology, and operations. Founded in 1989 as a technology consulting division of Arthur Andersen, Accenture has become one of the world's largest consulting firms, with approximately 750,000 employees serving clients in more than 120 countries[4]

The company's history is marked by significant milestones, including its rebranding from Andersen Consulting to Accenture in 2001[78]. This change signaled a new era for the company, emphasizing its unwavering commitment to innovation and "accent on the future".

Accenture's services span various industries, including healthcare, finance, manufacturing, and technology. The company is known for its digital transformation, cloud computing, and artificial intelligence (AI) expertise. In recent years, Accenture has made substantial investments in AI, including a $3 billion investment over three years, announced in 2023 to accelerate clients' AI-driven reinvention[3].

Accenture has positioned itself as a leader in AI and technology innovation. The company offers a range of AI services and solutions, helping clients leverage generative AI and other advanced technologies to transform their businesses[6]. Accenture has strong partnerships with leading technology providers, including Microsoft, with whom it announced a partnership in 2017 to create business software on iOS[78].

Accenture's company culture is built on six core values: Client Value Creation, One Global Network, Respect for the Individual, Best People, Integrity, and Stewardship[2]. The company's strong commitment to sustainability and corporate social responsibility, aiming to achieve net-zero emissions by 2025[8][7], is a testament to its responsible business practices.

Accenture has received numerous awards and recognitions for its work environment and diversity initiatives. It has been ranked among the top companies on various lists, including Fortune's 100 Best Companies to Work For and the FTSE Diversity & Inclusion Index[5].

As a global leader in professional services, Accenture continues to drive innovation and digital transformation across industries, helping clients navigate the rapidly evolving technological landscape and achieve sustainable growth.

2.2 Project Context

The primary motivation behind HealthVision AI was to integrate cutting-edge AI technologies into document processing solutions, staying at the forefront of industry advancements. The project aimed to build a system relying on multimodal LLMs to improve accuracy compared to current solutions, streamline the user experience by reducing the need for extensive training and specialized knowledge, and achieve faster processing times with more accurate results. These goals align with the industry trend towards AI-powered solutions that can work alongside humans, as noted in recent studies on the impact of AI in the insurance sector[34][50].

Before HealthVision AI, the existing OCR solution (iOCR) utilized machine learning techniques for document classification and information extraction. However, it faced several limitations, including difficulty handling diverse document formats such as medical records, insurance claims, and handwritten prescriptions. The system comprised multiple models, complicating analysis and reducing overall efficiency. It also needed an interactive question-answering functionality about document content and struggled significantly with handwritten content, a common challenge in OCR technology[49][81].

By integrating advanced Large Language Models (LLMs), and utilizing Azure AI Services for OCR, HealthVision AI processes sensitive and personal data with high precision. This centralized, multimodal approach integrates text and visual data processing, performing document classification, Q&A, and fields extraction relying on the latest Generative AI technologies, significantly outperforming traditional OCR methods in both accuracy and efficiency[44][63].

By leveraging the power of LLMs and addressing the limitations of traditional OCR systems, HealthVision AI represents a significant step forward in applying LLMs capabilities to document processing. This innovation has the potential to significantly improve how insurers and healthcare providers manage and extract value from their vast document repositories, leading to a future of improved operational efficiency and better customer experiences.

Chapter 3
Literature Review

3.1 Overview of OCR Technology

3.1.1 Definition and Explanation of OCR Technology

Optical Character Recognition (OCR) is a transformative technology that converts various documents, such as scanned paper documents, PDF files, or images captured by a digital camera, into editable and searchable data. The primary goal of OCR is to recognize and extract text from these visual inputs, empowering computers to process and analyze the information within them[80].

3.1.2 Key Components of OCR

1. **Image Acquisition:** OCR begins with capturing an image of the document containing text. This can be done through various means:

   - Scanning physical documents using flatbed scanners or specialized OCR scanners.
   - Capturing images with digital cameras or smartphone cameras.
   - Processing existing digital files such as PDFs or image formats (JPEG, PNG, etc.)[9].

   During this stage, the OCR software analyzes the scanned image and classifies the light areas as background and the dark areas as potential text[39].

2. **Preprocessing:** Before the actual text recognition can occur, the image undergoes several preprocessing steps to improve recognition accuracy:

   - Deskewing: Correcting any tilt or rotation in the scanned document to ensure horizontal text lines.
   - Despeckling: Removing digital noise or spots from the image.
   - Binarization: Converting the image to black and white to simplify text detection.
   - Line Removal: Cleaning up the image's boxes, lines, or other non-text elements.
   - Layout Analysis: Identifying and separating text blocks from images, tables, or other document elements.
   - Character Isolation: Separating individual characters for recognition[9].

3. **Text Recognition:** This is the core process of OCR, where the system attempts to identify individual characters and words. Two main approaches are used:

- Pattern Matching (Matrix Matching): Compares each character image to a stored library of character templates. This method works well for typed text in known fonts but needs help with new fonts or handwriting.
- Feature Extraction: Decomposes characters into features like lines, closed loops, line directions, and intersections. This method is more flexible and better at handling variations in text appearance, often using artificial intelligence techniques like neural networks for improved accuracy[80].

4. **Post-processing:** After initial recognition, OCR systems employ various techniques to improve accuracy:
   - Spell-checking and Error Correction: Using dictionaries and language models to correct recognized text.
   - Context Analysis: Disambiguating similar-looking characters based on context.
   - Applying Language-specific Rules: Enhancing recognition accuracy by applying grammatical and syntactical rules.
   - Machine Learning Algorithms: Continuously improving recognition capabilities through learning from corrections and feedback[9].
5. **Output:** The final step is to produce the recognized text in a usable format:
   - Plain text files.
   - Word processing documents.
   - Searchable PDFs.
   - Structured data for form processing[39].

3.1.3 Applications of OCR

OCR technology has found applications across numerous industries and use cases:

- **Document Digitization:** Conversion of physical documents into digital formats for more accessible storage, retrieval, and analysis[11].
- **Data Entry Automation:** Extracting information from forms, invoices, and other structured documents, reducing manual data entry and potential errors[39].
- **Accessibility:** Enabling visually impaired individuals to access printed text through text-to-speech conversion[9].
- **Historical Document Preservation:** Digitizing and making searchable historical manuscripts and archives, preserving valuable information for future generations[80].
- **License Plate Recognition:** Automating vehicle identification for parking, toll collection, and law enforcement[11].
- **Mobile Banking Applications:** Facilitating check deposits and document verification through smartphone cameras, making banking more convenient for users [9].
- **Language Translation:** Combining OCR with translation software to provide real-time translation of printed text[39].

3.1.4 Recent Advancements

OCR technology continues to evolve, with recent advancements focusing on:

- Improved handling of complex layouts and multiple languages[27].
- Better handwritten text recognition[39].
- Integration with AI for context understanding and error correction[11].
- Real-time OCR processing for mobile and augmented reality applications[9].

OCR plays a crucial role in digital transformation across various industries. It bridges the gap between physical documents and digital information systems by converting printed and handwritten text into machine-encoded text, thereby making it an indispensable tool in the modern professional's toolkit[39].

3.1.5 Current State-of-the-Art OCR Techniques

Optical Character Recognition (OCR) has witnessed significant advancements in recent years, primarily driven by the integration of deep learning and artificial intelligence. These advancements have led to developing more accurate, robust, and versatile OCR systems capable of handling many document types and text recognition scenarios. Here, we explore the current state-of-the-art OCR techniques and how they have improved the technology.

1. **Deep Learning-Based OCR:** Modern OCR systems heavily rely on deep learning algorithms, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). These approaches have significantly improved accuracy and robustness compared to traditional methods.

    - Convolutional Neural Networks (CNNs): CNNs are highly effective in image recognition tasks because they can learn and generalize features from input images. In OCR, CNNs are used for feature extraction, directly identifying relevant visual characteristics such as edges, shapes, and patterns from the training data. This allows CNN-based OCR systems to adapt to various fonts, sizes, and layouts with higher accuracy[76]. CNNs also provide robustness to variations in text appearance, making them more versatile than rule-based OCR solutions[62].
    - Recurrent Neural Networks (RNNs): RNNs, particularly Long Short-Term Memory (LSTM) networks, are used for sequence modeling in OCR. They process sequences of characters and words, considering the context and dependencies between them. RNNs can handle connected text and cursive writing more effectively, improving overall recognition accuracy[62].

      By combining CNNs for feature extraction and RNNs for sequence processing, modern OCR systems can handle complex document layouts and varying text styles more effectively[76].

2. **End-to-End OCR Systems:** End-to-end OCR systems integrate text detection, recognition, and post-processing into a single, trainable model. This holistic approach often leads to better overall performance compared to systems with separate components for each stage of the OCR process. Examples of end-to-end OCR systems include:

    - EAST (Efficient and Accurate Scene Text Detector): An end-to-end text detection system that is both accurate and efficient.
    - FOTS (Fast Oriented Text Spotting): Combines text detection and recognition in a single model, improving speed and accuracy[60].

These systems often outperform traditional pipeline methods in accuracy and speed, making them suitable for real-time applications.

3. **Attention Mechanisms:** Inspired by transformer models' success in natural language processing, attention mechanisms have further improved OCR performance. Models like ASTER (Attentional Scene Text Recognizer) use attention to focus on relevant parts of the input image, enhancing recognition accuracy[60].

4. **Multi-lingual and Multi-script OCR:** State-of-the-art OCR systems now support multiple languages and scripts within a single model. This is achieved through large, diverse datasets and advanced training techniques. For instance, Google's Cloud Vision OCR supports over 200 languages and can automatically detect and process multiple languages within the same document[18].

5. **Handwritten Text Recognition (HTR):** Modern OCR systems have made significant strides in handwritten text recognition, a traditionally challenging area. Techniques like Multidimensional Long-Short-Term Memory (MDLSTM) networks have shown promising results in this domain[26].

6. **Scene Text Recognition:** Recognizing text in natural scenes (e.g., street signs and product labels) has significantly improved. Models like CRNN (Convolutional Recurrent Neural Network) and ASTER have pushed the boundaries of accuracy in this challenging domain[60].

7. **Few-Shot and Zero-Shot Learning:** Recent research has focused on developing OCR models that recognize new character types or languages with minimal or no additional training data. This approach is beneficial for handling rare or historical scripts[10].

8. **Post-OCR Processing:** Advanced language modeling techniques are now being used to improve OCR output quality. These methods leverage context and linguistic knowledge to correct errors and improve accuracy[26].

3.1.6 Leading OCR Systems

Some of the leading OCR systems that incorporate these state-of-the-art techniques include:

- Google Cloud Vision OCR: Supports a wide range of languages and provides high accuracy for various document types[18].
- Microsoft Azure Computer Vision OCR: Offers robust text recognition capabilities and integrates well with other Azure services [47].
- Amazon Textract: Provides accurate text extraction for structured and unstructured documents [11].
- ABBYY FineReader: Known for its high accuracy and support for complex document layouts[1].
- Tesseract (open-source): A widely used OCR engine that supports multiple languages and scripts[76].

These systems can achieve character-level accuracy above 99% for high-quality printed documents and above 90% for challenging handwritten texts[60]. Despite these advancements, challenges remain in heavily degraded historical documents, extremely stylized fonts, and highly cursive handwriting. Ongoing research continues to push the boundaries of OCR technology, focusing on improving robustness, efficiency, and adaptability to new domains[10].

3.2 Overview of Large Language Models (LLMs)

Large Language Models (LLMs) stand out in the field of artificial intelligence and natural language processing due to their unique features. These computational models, with their remarkable proficiency, are designed to understand, generate, and manipulate human language. LLMs are probabilistic in nature, learning the statistical relationships and patterns in language from the training data. Their massive scale, characterized by the amount of data they are trained on and the number of parameters they contain, sets them apart [79].

3.2.1 Definition and Overview

An LLM is a type of artificial intelligence algorithm that uses deep learning techniques and massively large datasets to understand, summarize, generate, and predict new content [70]. These models are notable for their ability to achieve general-purpose language generation and other natural language processing tasks such as classification, translation, and question-answering [79].

The term "large" in LLMs refers to several aspects:

1. Data scale: LLMs are trained on vast corpora of text, often comprising hundreds of gigabytes or even terabytes of data.
2. Model size: Modern LLMs can contain billions or even trillions of parameters.
3. Computational resources: Training LLMs requires enormous computational power, often utilizing distributed computing systems and specialized hardware like GPUs and TPUs [79].

3.2.2 Historical Development

The concept of language models in AI traces back to the early days of artificial intelligence. The Eliza language model, developed in 1966 at MIT, is one of the earliest examples [70]. However, the modern era of LLMs began in 2017 with the introduction of the transformer architecture [79].

Critical milestones in LLM development include:

- 1966: Eliza, one of the earliest AI language models
- 2017: Introduction of the transformer architecture
- 2018: BERT (Bidirectional Encoder Representations from Transformers) by Google
- 2020: GPT-3 by OpenAI, marking a significant leap in scale and capabilities
- 2022-2023: Emergence of ChatGPT and GPT-4, bringing LLMs into mainstream use[16]

3.2.3 How LLMs Work

LLMs are based on the transformer architecture and consist of multiple neural network layers, including embedding, feedforward, recurrent, and attention layers. The model receives a text input, encodes it to capture semantic and syntactic meaning, and then decodes it to generate an output prediction. This process involves unsupervised pre-training on large text corpora, followed by fine-tuning or prompt-tuning for specific tasks.[79] During training, LLMs learn the probabilities of word sequences, allowing them to predict the most likely next word given a prompt. By capturing these statistical relationships, LLMs can generate coherent and contextually relevant text. The attention mechanism enables the model to focus on relevant parts of the input, enhancing its ability to understand context and generate accurate outputs.[29]

3.2.4 Technical Foundations

The architecture of modern LLMs is primarily based on the transformer model introduced by Vaswani et al. in 2017 [75]. The transformer architecture revolutionized NLP by introducing the self-attention mechanism, allowing the model to weigh the importance of different input parts when processing each element.

Key components of the transformer architecture include:

1. Embedding Layer: This layer converts input tokens (words or subwords) into dense vector representations.
2. Positional Encoding: Since transformers process input in parallel rather than sequentially, positional encodings are added to provide information about the relative or absolute position of tokens in the sequence.
3. Multi-Head Attention: This mechanism allows the model to attend to different parts of the input simultaneously, capturing various types of relationships between tokens.
4. Feed-forward Neural Networks: These networks process the output of the attention layers, allowing for non-linear transformations of the data.
5. Layer Normalization and Residual Connections: These components help in training very deep networks by mitigating the vanishing gradient problem [38].

The transformer architecture consists of two main parts:

- Encoder: Reads and processes the input text
- Decoder: Generates the output text based on the encoded information [72]

3.2.5 Training Process

The training process of LLMs typically involves two main phases:

1. Pre-training: Using self-supervised learning objectives, the model is trained on a large corpus of unlabeled text data. Common pre-training tasks include:
    - Masked Language Modeling (MLM): Predicting masked tokens in a sentence.
    - Next Sentence Prediction (NSP): Determining if two sentences follow each other in the original text.
    - Causal Language Modeling: Predicting the next token given the previous tokens [38].
2. Fine-tuning: The pre-trained model is further trained on specific tasks or domains, often with much smaller datasets. This process adapts the general knowledge acquired during pre-training to specific applications [38].

3.2.6 Word Representation and Embeddings

LLMs use advanced techniques to represent words and understand their relationships:

1. Word Embeddings: LLMs employ multi-dimensional vectors to represent words instead of using simple numerical tables. This allows words with similar contextual meanings or relationships to be close to each other in the vector space.
2. Contextual Understanding: Through the training process, the model learns to adjust these embeddings based on the context in which words appear, allowing for a nuanced understanding of language [16].

3.2.7 Scaling Laws and Emergent Abilities

Research has shown that the performance of LLMs often improves predictably with increases in model size, dataset size, and computational resources. This observation has led to the development of scaling laws, which help researchers estimate model performance in advance [40].

As LLMs scale up, they sometimes exhibit unexpected capabilities that are not present in smaller models. These "emergent abilities" can include improved reasoning, task generalization, and even basic arithmetic skills [40].

3.2.8 Applications and Use Cases

LLMs have found applications across numerous fields:

1. Natural Language Processing: Text generation, summarization, translation, and question-answering.
2. Content Creation: Assisting in writing, generating marketing copy, and creating educational content.
3. Code Generation: Helping programmers by generating code snippets and providing explanations.
4. Healthcare: Medical literature summarization, clinical decision support, and generating synthetic medical data for research.
5. Customer Service: Powering chatbots and virtual assistants for more natural interactions.
6. Education: Serving as tutoring aids and personalized learning assistants.

3.2.9 Challenges and Ethical Considerations

Despite their impressive capabilities, LLMs face several challenges:

1. Ethical Concerns and Bias Mitigation: Researchers are actively working on methods to reduce biases in LLMs and ensure their outputs are fair and ethical.
2. Hallucination and Factual Accuracy: Improving the factual consistency of LLM outputs remains an important area of focus.
3. Interpretability and Explainability: As LLMs become more complex, understanding their decision-making processes becomes increasingly important, especially for high-stakes applications.
4. Efficiency and Environmental Impact: A growing emphasis is on developing more efficient training and inference methods to reduce the computational and environmental costs associated with LLMs [40].

In conclusion, Large Language Models represent a significant leap forward in artificial intelligence and natural language processing. Their ability to understand and generate human-like text has opened up numerous applications across various industries. As research in this field continues to advance, we expect to see even more sophisticated and capable language models in the future, potentially revolutionizing how we interact with technology and process information.

3.3 Overview of Generative AI

3.3.1 Definition and Overview

Generative AI is a subset of artificial intelligence that focuses on creating new content, such as text, images, audio, and video, by learning patterns from existing data. Unlike traditional AI models designed to recognize patterns and make predictions, generative AI models can produce novel outputs that mimic the characteristics of the training data. This capability has opened up various applications across various industries, from content creation and design to healthcare and finance [37].

Generative AI models are typically built using deep learning techniques, particularly Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs). These models learn to generate new data by understanding the underlying distribution of the training data, allowing them to create realistic and diverse outputs [37].

3.3.2 Technical Foundations

1. **Generative Adversarial Networks (GANs):**
    - GANs consist of two neural networks: a generator and a discriminator. The generator creates new data samples while the discriminator evaluates their authenticity. The two networks are trained simultaneously in a process where the generator aims to produce increasingly realistic samples, and the discriminator strives to better distinguish real from fake samples[37].
    - This adversarial training process continues until the generator produces data that the discriminator can no longer reliably distinguish from accurate data.
2. **Variational Autoencoders (VAEs):**
    - VAEs are another generative model that learns to encode input data into a latent space and then decode it back into the original data. This encoding-decoding process allows VAEs to generate new data samples by sampling from the latent space [37].
    - VAEs are particularly useful for generating data that follows a specific distribution, making them suitable for applications like image and audio synthesis.
3. **Transformer Models:**
    - Recent advancements in generative AI have leveraged transformer architectures, particularly for text generation. Models like GPT-3 and GPT-4 use transformers to generate coherent and contextually relevant text based on input prompts [51].
    - Transformers use self-attention mechanisms to weigh the importance of different parts of the input sequence, enabling them to capture long-range dependencies and generate high-quality text [75].

3.3.3 Generative AI in Document Processing

Generative AI has significant potential in document processing, particularly in automating and enhancing the extraction of information from unstructured and semi-structured documents. This is highly relevant to the HealthVisionAI project, which aims to streamline health insurance claims and expense reimbursements.

1. **Document Extraction and Classification:**
    - Generative AI can create document processors that automate data extraction from various document formats, such as PDFs, scanned images, and text files[33].

- These processors can classify documents, extract relevant information, and convert it into structured data, improving the efficiency and accuracy of document processing workflows[15].

2. **Summarization and Data Insights:**

   - Generative AI models can summarize large documents, extract key information, and present it in a concise format. This is particularly useful for processing lengthy insurance claims and medical records[33].
   - By structuring and digitizing information from documents, generative AI helps businesses gain deeper insights and make better decisions[33].

3. **Automated Data Entry:**

   - Generative AI can automate the tedious and error-prone task of data entry by extracting and validating information from documents and transforming it into machine-readable formats[15]

4. **Handling Complex Document Structures:**

   - Advanced generative AI models can handle complex document structures, such as tables, forms, and mixed-content documents, ensuring accurate data extraction and classification[33].

3.3.4 Ethical Considerations and Challenges

Generative AI raises several ethical concerns and challenges that need to be addressed to ensure its responsible use:

1. **Misinformation and Deepfakes:**

   - Generative AI can create highly realistic fake content, such as deep fake videos and synthetic news articles, which can be used to spread misinformation and manipulate public opinion[69].
   - The potential for harm is significant, as deep fakes can damage reputations, influence elections, and incite violence.

2. **Bias and Discrimination:**

   - Generative AI models can perpetuate and amplify existing biases present in the training data, leading to discriminatory outcomes in applications such as hiring, lending, and law enforcement[43].
   - Addressing bias requires careful selection of training data and ongoing monitoring of model outputs.

3. **Copyright and Intellectual Property:**

   - Generative AI can produce content that resembles existing copyrighted works, raising legal issues related to intellectual property infringement[69].
   - Companies must navigate the legal landscape and ensure that their use of generative AI complies with copyright laws.

4. **Data Privacy and Security:**

   - Generative AI models require large amounts of training data, including sensitive and personal information. Data privacy and security is crucial to prevent unauthorized access and misuse[73].

- Compliance with data protection regulations, such as GDPR, is essential.

5. **Explainability and Interpretability:**
    - Many generative AI models operate as "black boxes," making it difficult to understand how they produce specific outputs. This lack of transparency can hinder trust and accountability[25].
    - Developing methods for explaining and interpreting model decisions is an ongoing area of research.
6. **Workforce Impact:**
    - Generative AI has the potential to automate tasks traditionally performed by humans, leading to concerns about job displacement and workforce morale[25].
    - Companies need to invest in reskilling and upskilling employees to prepare them for new roles created by AI technologies.
7. **Ethical Governance and Regulation:**
    - Establishing ethical guidelines and regulatory frameworks is crucial to ensuring the responsible development and deployment of generative AI[69].
    - Organizations should implement ethical oversight committees and conduct regular ethical impact assessments.

3.3.5 Future Directions

The future of generative AI holds exciting possibilities, but it also requires careful consideration of ethical implications and challenges:

1. **Advancements in Model Architecture:**
    - Continued research into improving the efficiency and capabilities of generative models, including developing new architectures and training techniques[82].
2. **Ethical AI Development:**
    - Integrating ethical considerations into the AI development lifecycle, from data collection to model deployment[25].
    - Developing technologies to detect and mitigate bias, ensure data privacy, and enhance model explainability.
3. **Regulatory Frameworks:**
    - Governments and regulatory bodies will play a crucial role in establishing guidelines and standards for the ethical use of generative AI[69].
    - International collaboration will be essential to address the global impact of generative AI.
4. **Interdisciplinary Collaboration:**
    - Collaboration between AI researchers, ethicists, legal experts, and policymakers will be necessary to navigate the complex ethical landscape of generative AI[25].

In conclusion, generative AI represents a powerful and transformative technology with the potential to revolutionize various industries. However, its development and deployment must be guided by ethical principles and robust governance frameworks to realize its benefits while minimizing potential harm.

3.4 Multimodal AI Systems Overview

3.4.1 Definition and Overview

Multimodal AI systems are advanced artificial intelligence models capable of processing and integrating multiple types of data inputs, such as text, images, audio, and video. These systems aim to mimic human-like perception and understanding by combining information from various sensory modalities. Unlike traditional AI models focusing on a single data type, multimodal AI can analyze and interpret complex, real-world scenarios involving multiple data streams simultaneously[31].

Integrating different modalities in AI systems enables a more comprehensive and contextual understanding of information. For example, in document processing applications, a multimodal AI system can analyze a document's textual content and visual layout, allowing for more accurate information extraction and classification. By leveraging the strengths of each modality, these systems can overcome limitations associated with single-modal approaches, leading to improved performance in tasks such as document analysis, content retrieval, and automated decision-making[31].

3.4.2 Applications in Document Processing

Multimodal AI has several applications in document processing, making it a valuable tool for automating and enhancing various tasks:

1. **Document Classification and Segmentation:** Multimodal AI can classify documents based on their content and layout, making organizing and retrieving information easier. It can also segment documents into coherent sections, improving the accuracy of data extraction[61].
2. **Data Extraction and Normalization:** By combining text and visual analysis, multimodal AI can extract relevant data from documents and normalize it into structured formats. This is particularly useful for processing forms, invoices, and other structured documents[31].
3. **Content Summarization and Analysis** Multimodal AI can summarize large documents by extracting essential information and presenting it concisely. This capability is valuable for processing lengthy reports, legal documents, and medical records[31].
4. **Enhanced Search and Retrieval:** By understanding both the textual content and visual layout of documents, multimodal AI can improve search and retrieval capabilities, making it easier to find relevant information within large document repositories[61].

3.5 Prompt Engineering Overview

Prompt engineering is a rapidly emerging field that focuses on designing and optimizing prompts in order to incerease the performance and accuracy of LLMS without the need for extensive fine-tuning or retraining. By carefully designing prompts, developers and users can leverage the full potential of LLMs and generative AI tools for various applications, including document content extraction[56].

3.5.1 Techniques and Best Practices

Fundamental techniques and best practices in prompt engineering include:

1. **Clear and Specific Instructions:** Providing clear and specific instructions helps the model understand the task and produce accurate outputs. For example, using delimiters like triple quotation marks or XML tags can help demarcate sections of text that should be treated differently[56].

2. **Context and Examples:** Including relevant context and examples in the prompt can improve the model's understanding and performance. For instance, providing a few examples of the desired output format can guide the model in generating similar responses[56].

3. **Iterative Refinement:** Iteratively refining prompts based on model responses can help achieve more accurate and relevant results. This involves testing different prompts, analyzing the outputs, and adjusting as needed[56].

4. **Chain-of-Thought Prompting:** Encouraging the model to break down complex tasks into smaller, logical steps can improve reasoning and problem-solving capabilities. This technique, known as chain-of-thought prompting, helps the model reason its way toward correct answers more reliably[56].

5. **Few-Shot and Zero-Shot Learning:** Few-shot learning involves providing a few examples along with the instruction, while zero-shot learning relies on the model's pre-trained knowledge without any examples. Both techniques can effectively guide the model to perform specific tasks[56].

3.5.2 Application in Document Content Extraction

Practical prompt engineering can guide LLMs in document content extraction to accurately identify, extract, and categorize relevant information from diverse document types. For instance, a well-crafted prompt might instruct the model to focus on specific document sections, recognize particular data formats, or extract information based on specific criteria. This approach allows for more flexible and adaptable document processing solutions, capable of handling many document structures and content types without extensive pre-programming or rule-based systems[59].

By leveraging prompt engineering techniques, organizations can enhance the capabilities of LLMs and generative AI models, improving the efficiency and accuracy of document processing workflows. This is particularly valuable in industries like healthcare and insurance, where accurate and timely information extraction is critical for decision-making and operational efficiency[59].

3.6 Microsoft Azure and Azure AI Services

Microsoft Azure is a comprehensive cloud computing platform that offers a wide range of services for building, deploying, and managing applications through Microsoft's global network of data centers. Azure provides software as a service (SaaS), platform as a service (PaaS), and infrastructure as a service (IaaS) offerings, supporting various programming languages, tools, and frameworks[48].

One of the key components of Azure is its AI services, which enable developers to incorporate intelligent features into their applications without needing deep expertise in machine learning. Azure AI Vision, in particular, offers computer vision capabilities such as optical character recognition (OCR), image analysis, and facial recognition[48]. These services can be easily integrated into applications to extract insights from visual data.

Azure AI Vision's OCR service, powered by advanced machine learning models, can extract printed and handwritten text from images and documents in multiple languages. It supports various input formats, including scanned documents, photos, and digital PDFs[45]. By leveraging Azure AI Vision OCR, developers can automate data extraction processes and reduce manual data entry efforts.

Furthermore, Azure OpenAI Service allows developers to access powerful language models like GPT-4 and integrate them with Azure AI services for enhanced functionality. GPT-4 Turbo with Vision, a variant of GPT-4 available through Azure OpenAI Service, can be combined with Azure AI Vision to improve OCR performance and enable more accurate text extraction[46].

The integration of GPT-4 with Azure AI Vision OCR brings several benefits:

- **Improved accuracy:** GPT-4's deep learning capabilities enable it to understand context and semantics, resulting in higher OCR accuracy, especially for handwritten text and complex layouts.
- **Language support:** Azure AI Vision OCR covers a wide range of languages, making it suitable for multilingual document processing scenarios.
- **Object grounding:** GPT-4 Turbo with Vision can visually distinguish and highlight important elements in images, enhancing data analysis and user interaction.

By leveraging the power of Azure AI services and integrating them with advanced language models like GPT-4, developers can build intelligent applications that accurately extract and process text from visual data. This combination of technologies opens up new possibilities for automating document processing workflows and deriving valuable insights from unstructured data sources.

In the context of HealthVision AI, utilizing Azure AI Vision OCR and GPT-4 Turbo with Vision enables the system to handle a wide variety of documents, including handwritten prescriptions and complex forms, with high accuracy. This integration forms a critical component of the project's architecture, contributing to its enhanced performance compared to traditional OCR solutions.

3.7 Advantages of LLM-Based Solutions Over Traditional OCR Systems

Traditional Optical Character Recognition (OCR) systems have long been used to digitize printed text, converting it into machine-readable data. However, these systems have several limitations, particularly when dealing with complex, unstructured, or varied document types. LLM-based solutions offer significant improvements over traditional OCR systems.

- **Enhanced Comprehension and Context Understanding:** Traditional OCR systems primarily rely on pattern recognition and keyword spotting, which can be effective for structured documents but struggle with unstructured data. In contrast, LLMs interpret language nuances and context, enabling them to handle complex and varied document types more effectively[14].
- **Flexibility with Unstructured Data:** Traditional OCR systems often fail with unstructured documents, such as emails or reports, leading to high error rates and the need for manual intervention. LLMs solutions excel in handling unstructured data, extracting meaningful information from diverse formats and contexts[14].
- **Adaptive Learning:** Updating traditional OCR systems for new formats or languages is time-consuming and resource-intensive. LLM-based solutions, however, can easily handle new data, adapting to changes in language usage or document formats without extensive manual reprogramming[14].
- **Improved Accuracy and Reduced Errors:** LLM-based solutions significantly boost text recognition accuracy, even in challenging conditions such as poor-quality images

or varied fonts. This leads to fewer errors and more reliable data extraction, which is crucial in high-stakes industries like healthcare and finance[35].

- **Comprehensive Data Analysis:** LLM-based solutions not only extract text but also analyze the data for deeper insights, identifying patterns and trends that are invaluable for business intelligence. This capability extends the usefulness of document processing beyond mere digitization[28].

By integrating LLM-centric approach, HealthVision AI aims to overcome the limitations of traditional OCR systems, providing a more robust, accurate, and efficient approach to document processing.

Chapter 4
Design and Architecture of HealthVision AI

4.1 System Architecture

4.1.1 Overview of the System Architecture

HealthVision AI employs a modular architecture inspired by the Model-View-Controller (MVC) design pattern. The system uses Python and leverages various Azure AI services and the LangChain framework to enable powerful document processing and conversational capabilities. The architecture, built for flexibility, ensures that the system can be tailored to meet your specific needs by separating the application logic into three main components:

- **Model:** This encapsulates the core AI functionalities and business logic related to document classification, document content investigation, and information extraction using large language models (LLMs) provided by Azure OpenAI. The GPT4 class represents the central Model component.
- **View:** Represents the application's user interface, built using the Streamlit framework. The UserInterface class handles the rendering of the UI and user interactions and manages the application flow.
- **Controller:** This class acts as an intermediary between the Model and the View, processing user inputs and orchestrating the interactions between the front end and the LLMs. It facilitates communication and abstracts the complexities of the underlying AI services.

4.1.2 Components and Their Interactions

The critical components of HealthVision AI and their interactions are as follows:

- **User Interface (UI):** Built using Streamlit, the UI allows users to upload documents, select document types, and interact with the system through an intuitive interface. It allows the user to chat with the LLM about the document context and display the extraction result in a structured format that can eventually be downloaded as a CSV file. The UserInterface class communicates with the Controller to send user inputs and receive updated data for display.
- **Azure AppServices:** Hosts the application and manages the flow of documents and questions between the user interface and the AI modules.

- **Controller:** The Controller class is the central orchestrator, facilitating interactions between the UI and the LLMs. It provides simplified commands for fetching responses, classifying documents, and retrieving prompts. The Controller interacts with the GPT4 and can interact with other services (Model components) to leverage their capabilities.
- **GPT4 Service:** The GPT4 class represents the core Model component, encapsulating the business logic and data processing tasks related to document classification, content investigation, and information extraction. It integrates with Azure OpenAI and Azure Computer Vision services to enable powerful OCR and LLM capabilities.
- **LangChain Integration:** LangChain, a framework for developing LLM-powered applications, simplifies the integration of LLMs into HealthVision AI. It provides tools for prompt orchestration and manages interactions with the LLMs, enhancing the system's capabilities.
- **Azure AI Services:** HealthVision AI leverages various Azure AI services, such as Azure OpenAI for LLM integration, Azure Computer Vision for OCR, and Azure Form Recognizer for structured data extraction. These services seamlessly integrate into the Model components, enabling advanced document processing functionalities and handling documents containing confidential and sensitive information.

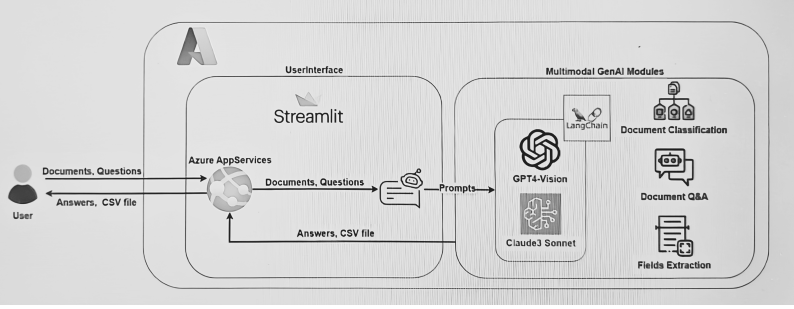

Figure 4.1: This diagram illustrates the overall architecture of the HealthVision AI system, showing the flow of data and interactions between the user interface, Azure AppServices, and the Multimodal GenAI Modules. It provides a visual representation of how documents and questions from users are processed through the system, utilizing various AI services for document classification, Q&A, and field extraction.

4.1.3 Integration and Data Flow

The modular architecture of HealthVision AI allows for seamless integration and data flow between the components. When a user interacts with the UI, the UserInterface class captures the user inputs and sends them to the Controller. The Controller processes the inputs and communicates with the appropriate Model component (GPT4 or other models) to perform the required tasks, such as document classification, information extraction, and conversation about the document content.

The Model components leverage the integrated Azure AI services and LangChain framework to process the documents, extract relevant information, and generate responses. The results are then returned to the Controller, which prepares the data for display and updates the View (UserInterface) accordingly.

This modular design allows for easy extension and customization of the system. New LLMs services or models can be integrated into the Model components without impacting the overall architecture. The separation of concerns between the Model, View, and Controller ensures the maintainability and scalability of the HealthVision AI system.

4.1.4 Image Preprocessing for Enhanced LLM Performance

HealthVision AI incorporates several image preprocessing techniques to optimize LLM performance in document analysis tasks:

1. **Image Resizing:** Ensuring that input images are of a consistent size improves processing speed and efficiency. While LLMs can handle various image sizes, providing a standardized input size reduces computational overhead and potentially improves response times [71].
2. **DPI Adjustment:** Modifying input images' dots per inch (DPI) enhances image quality and readability. A minimum of 300 DPI is often recommended for document processing to ensure clear text recognition [64].
3. **Grayscale Conversion:** Converting color images to grayscale improves the readability of written characters, especially in documents with complex backgrounds or low contrast. This step enhances the LLM's ability to accurately extract text from images [71].

Implementing these preprocessing techniques in HealthVision AI leads to several benefits:

- Improved Processing Speed: Optimizing images before feeding them to the LLM reduces the overall processing time, allowing for faster document analysis and information extraction.
- Enhanced Accuracy: Cleaner, more standardized inputs lead to more accurate text recognition and information extraction by the LLM.
- Reduced Computational Load: Preprocessing reduces the complexity of the task for the LLM, potentially allowing for more efficient use of computational resources.
- Consistency Across Various Document Types: These techniques help standardize inputs from various sources, ensuring consistent performance across different document types and qualities.

4.1.5 Overcoming Privacy Challenges with Azure Vision Enhancements

A critical challenge encountered during the development of HealthVision AI was handling sensitive personal information in medical and insurance documents. This limitation became particularly evident when processing documents containing names, surnames, and fiscal codes, as OpenAI's GPT-4 model tended to block such processes due to privacy concerns[53].

To address this challenge, we implemented Azure AI Services for Enhanced OCR, which proved to be a game-changer for our project. Azure AI Vision offers innovative computer vision capabilities:

- **Azure Cognitive Services:** Azure Cognitive Services provides powerful OCR capabilities to accurately recognize and extract text from various document formats, including PDFs and images. This integration leverages Azure's advanced algorithms to handle complex document layouts and ensures high text recognition accuracy [13].
- **Azure Form Recognizer:** Azure Form Recognizer extracts key-value pairs, tables, and other structured data from documents. This service enhances HealthVision AI by providing detailed and accurate data extraction, crucial for processing health insurance claims and expense reimbursements [12].

This technology allowed us to:

1. **Enhance Privacy Compliance:** Azure's OCR capabilities could process sensitive information without exposing it directly to the GPT-4 model, ensuring compliance with data privacy regulations.
2. **Improve Text Extraction Accuracy:** Azure's OCR technology extracts high-quality text from dense medical documents and financial records, including handwritten notes and complex layouts[46].
3. **Maintain Performance:** Despite the additional processing step, the integration of Azure OCR maintained the system's responsiveness, as evidenced by the quick responses in the chat interface. This impressive performance was a testament to the system's robustness.

To implement this solution, we modified LangChain's source code to integrate Azure AI Services seamlessly. This adaptation required a deep understanding of LangChain's architecture and Azure's OCR capabilities. The result was a robust system that could process sensitive documents while maintaining strict privacy standards[45].

This solution demonstrates the project's commitment to innovation, privacy, and regulatory compliance. By leveraging Azure's advanced OCR capabilities, we not only overcame a significant technical hurdle but also enhanced HealthVision AI's overall capabilities, setting it apart in the field of LLMs-powered document processing for healthcare and insurance. Our innovative approach to privacy and compliance, coupled with the advanced OCR technology, ensures that HealthVision AI is at the forefront of document processing solutions.

4.1.6 LLMs, Prompts and Their Roles

HealthVision AI efficiently leverages Large Language Models (LLMs) at three distinct stages to significantly improve the document processing experience:

- **Document Upload and Verification:** Users can choose between different document options (General Document, Expense Document, Medical Document) when uploading a document. The LLM verifies if the chosen category matches the uploaded file, using specific critical points associated with each document type and extracting contextual information. This verification process ensures that the subsequent processing is aligned with the LLM's instructions, enhancing the accuracy and relevance of data extraction.

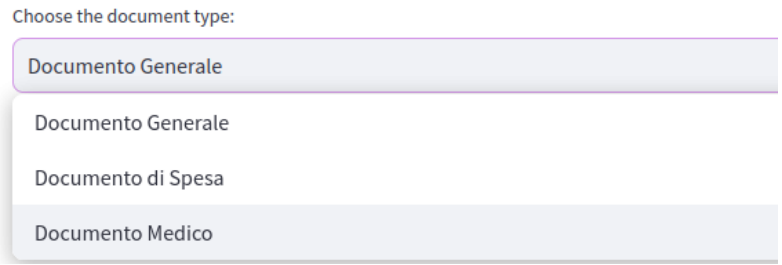

Figure 4.2: Options for selecting the type of document (General Document, Expense Document, Medical Document).

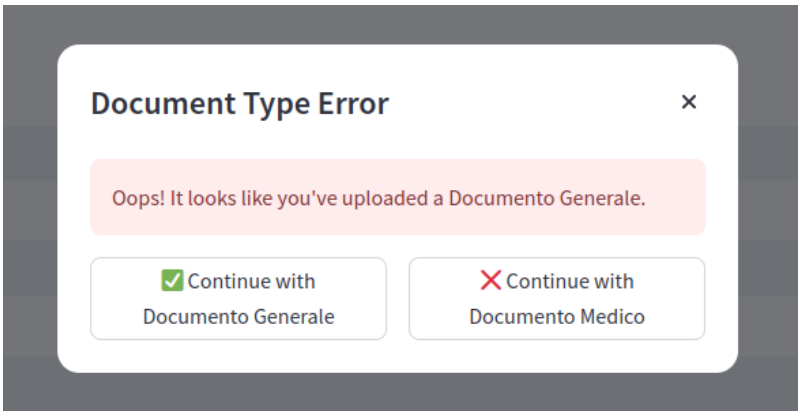

Figure 4.3: Document category classification error message.

- **Interactive Q&A and Document Content Analysis:** The system provides a user-friendly interactive Q&A interface between the user and the LLM in the second stage. The chosen document type determines the system prompt, which contains information about possible content, relevant information, and expected formats. This stage ensures that the LLM accurately extracts the relevant information upon request, making the user experience seamless.

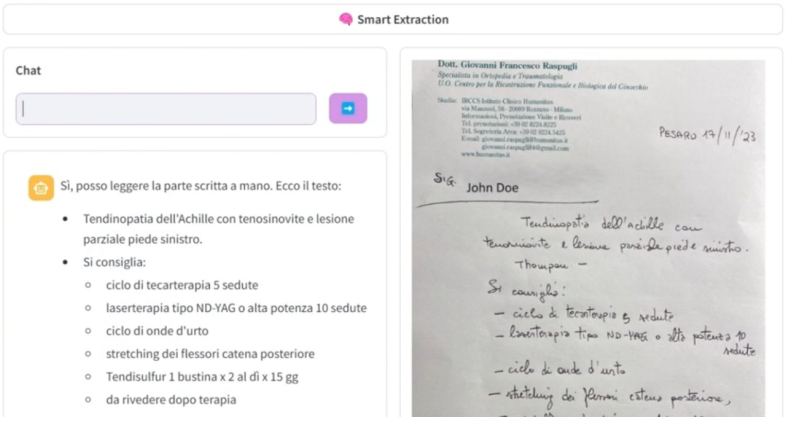

Figure 4.4: Example of conversation with LLM about document content.

- **Information Extraction and Formatting:** In the final stage, the LLM extracts and formats the requested information in a structured JSON format. A special message triggers the extraction process to separate the chat interaction from the extracted data. This approach ensures that the LLM can follow specific instructions and maintain context, resulting in accurate and reliable data extraction.

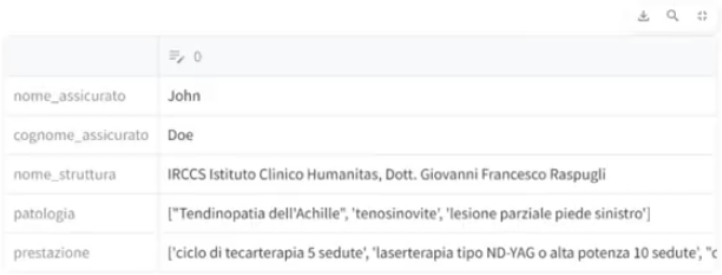

Figure 4.5: Extracted information is presented in a structured format upon user request.

HealthVision AI uses prompt engineering techniques to optimize the performance of LLMs in document content extraction:

- **Clear and Specific Instructions:** Clear and specific instructions play a key role in helping the LLM understand the task at hand and produce accurate outputs. This approach, as outlined in the Prompt Engineering Guide, significantly improves the overall efficiency of the document processing workflow [59].
- **Chain-of-Thought Prompting:** Encouraging the model to break down complex tasks into smaller steps enhances its reasoning and problem-solving capabilities, leading to more reliable results in data extraction [59].

Carefully crafted prompts guide the LLMs in HealthVision AI at each stage of the document processing workflow. These prompts are crucial in directing the AI models to perform specific tasks and generate relevant outputs. Here is an overview of the critical prompts and their roles:

1. **Document Classification Prompt:**
    - This prompt is used to determine the class or type of the uploaded document based on visual features and content description.
    - It instructs the LLM to output only the label of the document analyzed, ensuring concise and accurate classification results.
    - The prompt provides detailed descriptions of three document categories: Expense Document, Medical Document, and General Document, allowing for more informed and precise classification.
    - The LLM analyzes the uploaded document and classifies it into one of these categories based on the provided descriptions and rules.
2. **General Document Processing Prompt:**
    - This prompt guides the LLM in processing general documents and extracting relevant information.
    - It emphasizes the importance of accurate OCR, tracking user-specified fields, and responding to user queries related to the document contents.
    - The prompt instructs the LLM to provide extracted data in a structured JSON format when the user requests it, ensuring consistent and machine-readable output.
3. **Expense Document Processing Prompt:**
    - This prompt is designed to process expense documents and extract essential information. It focuses on tracking fields such as the transferee's name and surname, tax code, date, document number, amount, VAT number, and tax code of the transferor.

- The prompt guides the LLM in extracting and providing the data in a structured JSON format when the user requests, facilitating easy integration with other systems.

4. **Medical Document Processing Prompt:**

   - This prompt is tailored to process medical documents and extract relevant information about medical treatments and pathologies. It directs the LLM to track fields such as the insured's name and surname, the name of the facility/doctor, pathologies, and services provided.
   - The prompt instructs the LLM to extract and present the data in a structured JSON format when the user requests it, enabling efficient analysis and reporting.

These prompts are carefully designed to guide the LLMs in performing specific tasks at each stage of the document processing workflow. They provide clear instructions, specify the fields of interest, and define the expected output format. By leveraging these prompts, HealthVision AI ensures accurate document classification, information extraction, and interactive Q&A capabilities.

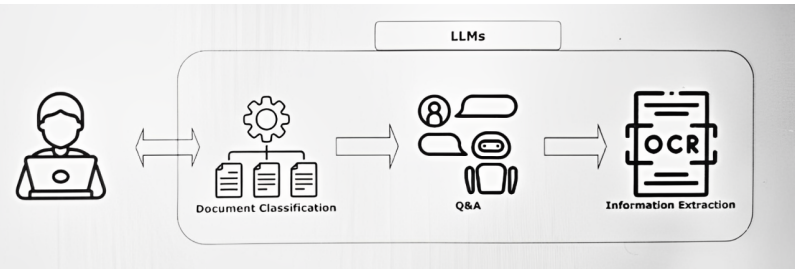

Figure 4.6: All the various tasks the LLMs perform, showing the centricity of this service in handling the various tasks.

The prompts are integrated into the Model components (GPT4 and other services) and are invoked by the Controller based on user interactions and the selected document type. This modular architecture allows for easy extension and customization of the prompts to accommodate new document types or additional processing requirements.

HealthVision AI's prompts follow best practices, such as providing detailed queries, using clear separators, and outlining the steps involved in each task. This approach ensures that the LLMs can understand and execute the desired actions accurately, leading to more reliable and relevant outputs.

Moreover, the prompts in HealthVision AI are tailored to specific document types and processing requirements, leveraging domain-specific knowledge to enhance the accuracy and effectiveness of the LLMs. As mentioned in the search results, this task-specific knowledge enrichment is crucial for achieving optimal performance in intelligent document processing.

4.2 User Interface Design

The HealthVision AI user interface is designed to focus on simplicity, clarity, and user-friendliness. It leverages the Streamlit framework to create an intuitive and responsive

web application. The design follows best practices for document processing applications, ensuring a smooth user experience throughout the document analysis workflow.

4.2.1 Design Principles

- **Simplicity and Clarity:** The interface uses a clean, uncluttered layout to reduce cognitive load on users.
- **Visual Hierarchy:** Essential elements are emphasized through size, color, and positioning.
- **Consistency:** A uniform design language is maintained across all pages.
- **Responsiveness:** The UI is designed to adapt seamlessly to different screen sizes and devices, ensuring that all users can access and use the application comfortably.
- **Feedback:** Clear visual cues and messages are strategically placed to inform users about the system's status and actions, empowering them with a sense of control over the process.

4.2.2 Page 1: Login and Document Upload

This page serves as the entry point for users, combining authentication and document upload functionalities. It is designed to guide users through the initial steps of the document processing workflow.

Key Elements:

- **Sidebar:**
  - Application logo and title: Provides branding and immediate recognition.
  - Purpose statement: Clearly states the purpose of the Company.
  - Login form: Ensures that only authorized users can access the system.
  - Document analysis instructions: Guides users on how to use the system effectively.
  - Document Analysis Instructions: provides information on the correct use of the application for new users.
  - Disclaimer: Provides legal and usage information informing about the AI's limitations and unaccountability.
- **Main Content Area:**
  - Document type selection dropdown: Allows users to specify the type of document they are uploading.
  - Model selection dropdown: Let users choose the model for document processing.
  - File upload area: Supports drag-and-drop functionality, enhancing user experience by simplifying the upload process.
  - "Browse files" button: Offers an alternative to drag-and-drop for file uploads.
  - "UPLOAD" button: Initiates the document upload process.
- **Design Considerations:**
  - Context and Authentication: The sidebar provides essential context and ensures secure access through authentication.

- User Guidance: Clear instructions and a disclaimer guide users through the document upload process, reducing potential errors.
- Flexibility: The upload area supports drag-and-drop and traditional file browsing, catering to user preferences.
- Accessibility: Ensure all interactive elements are accessible via keyboard navigation and screen readers.

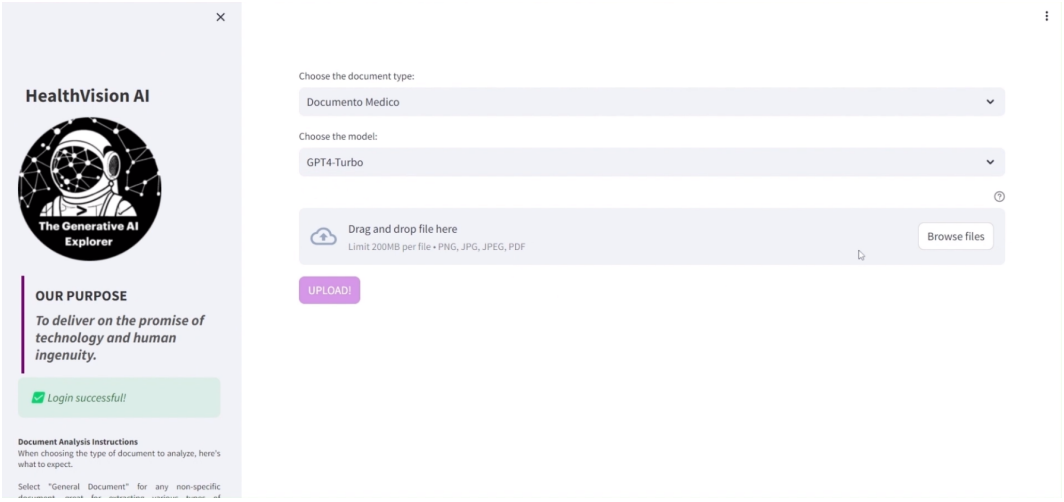

Figure 4.7: The HealthVision AI login and document upload page combines authentication and document upload functionalities in a clean and intuitive layout. The sidebar provides branding, purpose, and document analysis instructions. At the same time, the main content area allows users to select the document type and model and upload files either by dragging and dropping or using the browse button. The "UPLOAD" button initiates the upload process.

4.2.3 Page 2: Chat with Document

This page facilitates user and LLM interaction regarding the uploaded document.
Key Elements:

- **Two-Column Layout:**
  - Left column: Chat interface allows users to interact with the LLM.
  - Right column: Document viewer references the uploaded document.
- **Chat Interface:**
  - Text input field for user questions: Allows users to type queries and instructions for the LLM.
  - Send button: Submits the user's query to the LLM.
  - Chat history display: The conversation history provides context for ongoing interactions.
- **Document Viewer:**
  - Displays the uploaded document: Ensures users can reference the document while interacting with the LLM.
- **Smart Extraction Button:**

  - Prominently placed at the top of the page: Encourages users to proceed with the extraction process.

- **Navigation**
  - Button to return to the main page: Allows users to navigate back to the document upload page easily.

- **Design Considerations:**
  - Two-Column Layout: This design allows users to seamlessly reference the document while conversing with the LLM, enhancing usability.
  - Intuitive Chat Interface: The chat interface mimics familiar messaging apps, making it intuitive for users to interact with the LLM.
  - Highlighting Key Actions: The Smart Extraction button is highlighted to guide users toward the next step in the workflow.

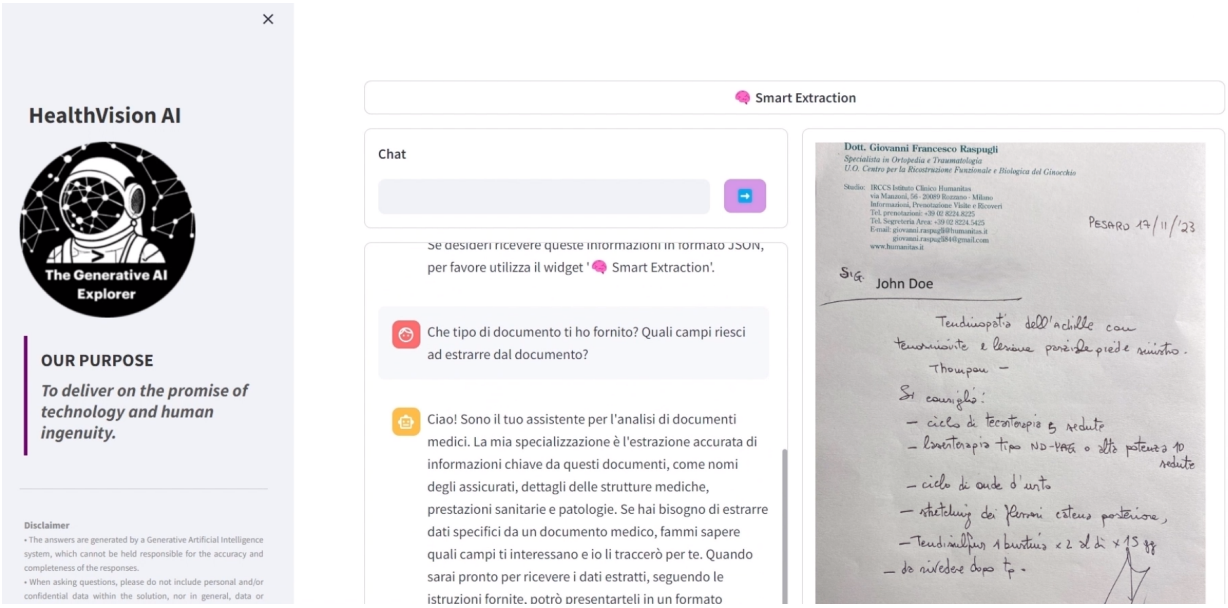

Figure 4.8: The chat with document page in HealthVision AI enables users to interact with the LLM regarding the uploaded document. The two-column layout features a chat interface on the left and a document viewer on the right. Users can ask questions about the document, receive responses, and use the "Smart Extraction" button to initiate data extraction. This layout allows users to quickly reference the document while engaging in a conversation with the LLM.

4.2.4 Page 3: Extraction Results

This page presents the extracted information in a structured format.
Key Elements:

- **Two-Column Layout:**
  - Left column: Extracted data table displays the document analysis results.
  - Right column: Document viewer references the original document.

- **Extracted Data Table:**
  - Editable fields: Allow users to make corrections directly within the interface.

  - Download button: Enables users to export the extracted data as a CSV file.

- **Document Viewer:**
  - Displays the original document: Facilitates easy verification and comparison with the extracted data.

- **Navigation Buttons:**
  - Return to document upload: Allows users to start a new upload process.
  - Return to chat interface: Let users return to the previous step for further interactions.

- **Design Considerations:**
  - Side-by-Side Layout: Facilitates easy verification by displaying the extracted data and the original document.
  - Editable Fields: Enhance user control by allowing quick corrections to the extracted data.
  - Prominent Download Button: Ensures easy access to the extracted data for further use.
  - Responsive Design: Ensure the table and document viewer are usable on various screen sizes.

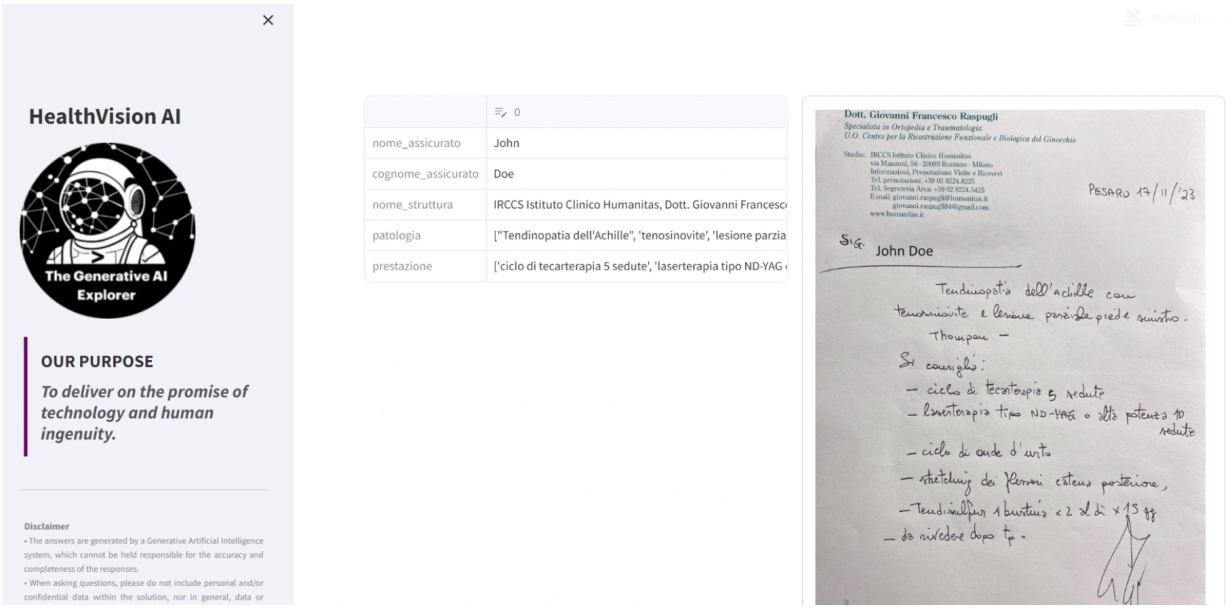

Figure 4.9: The extraction results page in HealthVision AI presents the extracted information in a structured format. The two-column layout displays the extracted data table on the left and the original document on the right. Editable fields in the data table allow users to make corrections, and a download button enables exporting the data as a CSV file. This layout facilitates easy verification and comparison of the extracted data with the original document.

4.2.5 General UI Enhancements

- **Progress Indicators:** Long-running processes like document parsing, document classification, answer generation, information extraction, progress bars, and loading wheels with explanatory comments help keep users informed of the process's status.

- **Error Handling:** Clear error messages and suggestions for resolution when issues occur are shown clearly whenever an error occurs.
- **Tooltips and Help Text:** Additional information and guidance through tooltips and help text for complex features is added to clarify the use of each button and widget.
- **Responsive Design:** The application is usable on various devices and screen sizes thanks to the Streamlit nature.

4.3 Technology Stack

HealthVision AI leverages a cutting-edge technology stack to deliver powerful document processing and analysis capabilities. The core components of this stack are LangChain and Streamlit, which work together to create an efficient and user-friendly application.

4.3.1 LangChain

LangChain is an open-source framework designed to simplify the creation of applications using large language models (LLMs). It is the backbone for HealthVision AI's natural language processing capabilities [41][42].

Critical components of LangChain utilized in HealthVision AI:

1. **LLMs Integration:** LangChain seamlessly integrates with GPT4-Turbo and other potential models. [41].
2. **Prompts:** Custom prompt templates efficiently interact with LLMs for document classification and information extraction [42].
3. **Chains:** LangChain's chains create a workflow for document processing, from upload to classification and information extraction [41][42].
4. **Memory:** This module plays a crucial role in maintaining context during user interactions, ensuring that the LLM system can remember previous interactions and provide more personalized responses [42].
5. **Agents:** LangChain's agents can be used for decision-making components that determine the best course of action based on input, context, and available resources [42].

Benefits of using LangChain in HealthVision AI:

- Simplified development of complex LLM-powered document processing workflows.
- Enhanced flexibility in integrating different LLMs (GPT4-Turbo, Claude3 Sonnet)
- Improved consistency in LLM interactions across different document types (General, Medical, Expense)

4.3.2 Streamlit

Streamlit forms the foundation of HealthVision AI's user interface. This open-source Python library enables the creation of interactive web applications with minimal front-end development expertise [67][68].

Critical features of Streamlit utilized in HealthVision AI:

1. **Intuitive UI Components:**
    - File uploader for document submission.

- Dropdown menus for document type and model selection.
- Chat interface for user-LLM interaction
- Sidebar for login and application information [68]
- Buttons for page navigation.

2. **Real-time Interactivity:** The chat interface and document viewer update in real time based on user inputs and AI model outputs [67].
3. **Data Visualization:** Streamlit's support for data presentation is evident in the structured display of extracted information. This feature enhances the user experience by presenting complex data in a clear and understandable format, making it easier for users to interpret the AI's outputs [68].
4. **Session State Management:** This feature is crucial for maintaining login status, conversation history, and document context across user interactions [68].

Benefits of using Streamlit in HealthVision AI:

- Rapid development of a professional-looking user interface.
- Seamless integration with Python backend, where LangChain orchestrates the AI models.
- Built-in support for file uploading, data visualization, anduser input widgets [67][68]

4.3.3 Integration in HealthVision AI

The combination of LangChain and Streamlit in HealthVision AI creates a powerful synergy:

1. **Document Upload and Classification:** Streamlit provides the interface for document upload and type selection, while LangChain manages the backend logic for document classification.
2. **Interactive Chat:** The chat interface is built using Streamlit's components, with LangChain handling the conversation flow, context management, and interaction with the selected LLM.
3. **Information Extraction:** LangChain's chains and agents process the document to extract relevant information, which is then displayed in a structured format using Streamlit's data presentation capabilities.
4. **User Authentication:** While LangChain or Streamlit do not directly handle login functionality, it is seamlessly integrated into the Streamlit interface.
5. **Modular Design:** The use of these technologies allows for a modular design, evident in the separate pages for document upload, chat interface, and extraction results.

4.4 Development Process and Deployment

The development of HealthVision AI followed a structured approach, beginning with extensive research and culminating in deploying a sophisticated application. This section details each process step, highlighting key activities, challenges encountered, and solutions implemented.

4.4.1 Initial Research and Preparation

Before the company-provided hardware was even provided, my first task was to study deep into the foundational knowledge and resources related to generative AI and relevant technologies. This preparatory phase was not just a formality, but a crucial step that set the stage for our entire development journey. The resources I studied were not just random materials, but carefully selected to provide a solid understanding of the tools and concepts we would later apply. These resources included:

- **Theoretical Resources:**
  - OpenAI's blog post on ChatGPT's capabilities, including visual and auditory enhancements[52][53].
  - The GPT-4 System Card and related research articles from OpenAI.[55]
  - Azure AI Services tailored enhancements, when combined with Azure AI Vision, it enhances your chat experience by providing the chat model with more detailed information about visible text in the image and the locations of objects.
- **Practical Courses:**
  - Generative AI: A comprehensive course on generative AI fundamentals[20].
  - ChatGPT Overview and API Integration: Courses on using ChatGPT, focusing on prompt engineering and API usage[19][54].
  - Prompt Engineering: A course on advanced prompt engineering with models like Llama-2[56][24].
  - LangChain: Basic and advanced courses on using LangChain for LLM applications, including Q&A functionalities[22][21][23].
  - Streamlit: Tutorials on building conversational apps and integrating LangChain with Streamlit[65][66].
  - Git: A tutorial on using Git for version control and collaboration[77].

4.4.2 Project Conceptualization and Initial Testing

Following the research phase, several meetings were conducted to discuss the project's vision and goals. The primary objective was to enhance the company's existing OCR solution by leveraging the capabilities of multimodal LLMs. The focus was on processing expense and medical documents for insurance claims and reimbursements.

The first primary task was to test the model's ability to extract relevant information from these documents, comparing its performance to the existing "iOCR" solution. We evaluated the model using computer-typed and handwritten information, performing separate evaluations for each. The tests showed promising results, leading us to progress with full-scale development.

4.4.3 Development Phase

The development phase centered on integrating the LLM's information extraction capabilities into a user-friendly application. Streamlit was chosen as the framework for building the application due to its simplicity and seamless integration with Python. This choice allowed us to maintain consistency in our technology stack, using Python for both backend and frontend development.

LangChain was selected to handle LLM interactions. Its popularity and standardized syntax for integrating generative AI models simplified the development process. The project followed an MVC pattern:

- Model: Encapsulated core AI functionalities for document classification and information extraction.
- View: Handled the user interface and built it using Streamlit.
- Controller: Managed the interactions between the model and the View, processing user inputs and orchestrating the application's workflow.

4.4.4 Challenges and Solutions

1. **Multimodal Processing Limitations:** One of the primary challenges faced was the integration of multimodal processing capabilities with LangChain and Streamlit. Since multimodality was a new concept, LangChain did not initially support settings for processing documents containing personal data. This limitation was particularly evident when dealing with documents/images with names, surnames, and fiscal codes, as OpenAI's GPT-4 tended to block such processes. However, our team's adaptability and quick thinking led us to a solution. We utilized Azure AI Services for Enhanced OCR and modified LangChain's source code to integrate these resources, ensuring compliance with data privacy regulations and achieving the desired functionality.
2. **Performance Optimization:** We implemented Streamlit's *st.cache_resource* decorator to improve the application's performance for LLM instance creation. This optimization ensured that LLM instances were cached and reused across user sessions, significantly reducing the initialization time when users navigated between pages or when multiple users accessed the system simultaneously.
3. **User Interface Design:** Based on stakeholder feedback, we refined the user interface to enhance usability. One notable improvement was implementing a two-column layout that displayed the original document alongside the extracted information or chat interface. This layout allowed users to verify extraction results against the original document quickly. Additionally, we added a chat interface later in the development process to facilitate interaction with the LLM for document analysis based on feedback that indicated the need for a more interactive and informative tool.

4.4.5 Iterative Development and Testing

Daily meetings with supervisors were integral to the development process. These sessions allowed us to review progress, brainstorm new features, and ensure alignment with team goals. Multiple demos were created and presented to gather feedback from team members and other stakeholders in the AI and data departments. This iterative feedback loop was critical for refining and improving the application's functionality.

Testing methodologies included manual verification against ground truth data. Ground truth files containing all fields of interest were prepared for each document type. The extracted information from the application was then compared with these ground truth files to assess accuracy. Based on these evaluations, prompts were continuously refined to enhance extraction performance. Although automated testing frameworks were not employed, the manual testing approach ensured thorough validation of the application's functionality.

4.4.6 Deployment

We used Docker for containerization for deployment, which provided a consistent deployment environment. The application was deployed on Azure App Services, leveraging Azure's scalability and reliability to ensure high performance and availability. The deployment process involved creating a Docker file following Streamlit's tutorial on deploying

applications with Docker. This file facilitated the application deployment on Azure App Services, ensuring a seamless transition from development to production.

Performance analysis measured response times, throughput, and resource utilization. Tools like Azure Monitor were used for real-time performance monitoring. The accuracy of document processing and information extraction was evaluated using accuracy metrics.

Cost analysis compared the expenses associated with developing, deploying, and operating HealthVision AI to traditional OCR solutions.

4.4.7 Conclusion

The development of HealthVision AI demonstrates the power of combining cutting-edge AI technologies with a user-centric design approach. By leveraging Streamlit, LangChain, and Azure services, we created a scalable, user-friendly application that significantly advances document processing capabilities in the health insurance domain. The iterative development process and continuous stakeholder feedback were crucial in refining the solution, ensuring that the end product met the needs of our users and stakeholders. This approach resulted in a powerful tool for streamlining insurance claims and reimbursements.

Chapter 5
Cost Analysis and Performance Evaluation

5.1 Cost Structure of HealthVision AI

The cost model for HealthVision AI's document processing system is primarily built on two crucial components: the utilization of GPT-4 Vision API and Azure AI services. This section outlines the cost structure and provides a comprehensive analysis of the financial implications of system usage, with these two components at its core.

5.1.1 GPT-4 Vision API Costs

The GPT-4 Vision API employs a token-based pricing model with distinct rates for input and output tokens:

- Input tokens: $0.01 per 1,000 tokens
- Output tokens: $0.03 per 1,000 tokens

The token consumption is bifurcated into two principal stages:

1. **Document Classification Stage:**
    - System Prompt: 638 tokens
    - Image Processing: 765 tokens (estiamted)
    - Classification Output: 4 tokens
2. **Document Q&A and Extraction Stage:**
    - System Prompt: 730 tokens (average)
    - Image Processing: 765 tokens (estimated)
    - Extraction Command: 3 tokens
    - Q&A Interaction: 50 tokens per query (estimated)
    - Extraction Output: 750 tokens (estimated)

5.1.2 Azure AI Services Costs

Supplementary to the GPT-4 Vision API costs, Azure AI services incur an additional expense of $1.5 per 1,000 transactions. Each document processing cycle necessitates N+2 transactions, where N represents the number of queries during the Q&A phase.

5.1.3 Consolidated Cost Formula

The following equation expresses the amalgamated cost per document processed:

$$C = [1406 + 1495(N+1) + 50(N^2 + 2N)](10^{-5}) + (754 + 50N)(3 \times 10^{-5}) + (N+2)(1.5 \times 10^{-5})$$

Where:

- C is the total cost per document in dollars.
- N is the number of queries in the Q&A phase.

5.1.4 Cost Analysis and Optimization Strategies

The cost structure of HealthVision AI exhibits several noteworthy characteristics:

- **Query Dependency:** The cost escalates non-linearly with the number of queries due to increased token usage and additional Azure AI service transactions.
- **Image Processing Overhead:** The high-resolution image processing (765 tokens) constitutes a significant portion of each document's base cost.
- **System Prompt Impact** The substantial token count for system prompts (638 and 730 tokens for classification and Q&A, respectively) contributes significantly to the base cost.
- **Extraction Output Dominance:** The extraction result (750 tokens) significantly contributes to the output token count, which is charged at a premium rate.
- **Transaction-based Azure AI Costs:** The Azure AI services cost adds a fixed amount per transaction, scaling linearly with query count.

To optimize costs while maintaining system efficacy, the following strategies are proposed:

1. Implementation of a query limit in the Q&A phase to limit variable costs.
2. Refinement of system prompts to reduce token count without compromising effectiveness
3. Exploration of lower resolution image processing for suitable document types

5.1.5 Scalability and Cost Projections

The cost structure of HealthVision AI demonstrates a high degree of scalability, with costs directly proportional to usage. However, as the volume of processed documents increases, vigilant monitoring becomes imperative to ensure cost-effectiveness and identify optimization opportunities.

Future work should include developing a comprehensive cost projection model that accounts for varying document complexities, user interaction patterns, and potential volume discounts from service providers. Such a model would facilitate more accurate budgeting and inform strategic decisions regarding system scaling and feature development.

5.1.6 Comparative Analysis with Existing OCR Solution

While HealthVision AI demonstrates superior accuracy and versatility compared to the company's existing OCR solution, the current cost structure renders it more expensive to operate. This cost differential is primarily attributed to utilizing advanced LLM technologies and Azure AI services, which, while providing enhanced capabilities, come at a premium.

Several factors justify the increased expenses associated with HealthVision AI:

1. **Enhanced Accuracy:** The LLM-based approach significantly reduces error rates in document processing, particularly for complex medical terminology and handwritten notes.
2. **Versatility:** Unlike traditional OCR, HealthVision AI can handle many document types without requiring specific templates or extensive rule-based programming.
3. **Natural Language Understanding:** The system's ability to comprehend context and respond to user queries provides value beyond mere text extraction.
4. **Continuous Improvement:** LLMs have the potential to improve over time with minimal manual intervention, potentially reducing long-term maintenance costs.

Despite these advantages, the current cost structure may limit widespread adoption, particularly for high-volume, routine document processing tasks where the existing OCR solution remains more cost-effective.

However, it is crucial to consider the rapidly evolving landscape of AI technologies. There is a strong possibility that the costs associated with LLMs and cloud-based AI services will decrease in the future, driven by:

1. Advancements in model efficiency, reducing computational requirements.
2. Increased competition among AI service providers.
3. Economies of scale as adoption of these technologies becomes more widespread

HealthVision AI is poised to become an increasingly attractive solution as these cost reductions materialize, potentially surpassing traditional OCR in performance and cost-effectiveness. This prospect underscores the importance of continued investment and development in this technology, positioning the company at the forefront of document processing innovation.

In the interim, a hybrid approach may be considered, leveraging HealthVision AI for complex documents or high-value processes where its advanced capabilities justify the additional cost while retaining the existing OCR solution for more routine, high-volume tasks. This strategy would allow the company to balance cost considerations with the need for advanced document processing capabilities while preparing for a future where LLM-based solutions become the standard in the industry.

5.2 Performance Evaluation of HealthVision AI

HealthVision AI, our LLM-centric approach to document processing, which leverages large language models for accurate information extraction, demonstrates significant performance improvements over the traditional Azure-based OCR solution (iOCR). It is important to note that iOCR is the system developed by the team before I joined the project, which has been referred throughout this thesis. This evaluation process was designed to provide a comprehensive and fair comparison between these two systems, focusing on their ability to extract information from various document types accurately.

5.2.1 Dataset Preparation

We utilized an existing dataset comprising diverse documents obtained from a company's primary client. This ensured that our evaluation was based on real-world data encountered in healthcare and insurance contexts, a factor that directly relates to the practical application of our work in your industries. The dataset was divided into two main categories:

- Computer-typed information documents such as tickets, invoices, receipts and policy documents.
- Handwritten information documents such as red and white prescriptions and reports.

This division allowed us to assess HealthVision AI's performance across different document types, reflecting the variety of formats typically encountered in medical and expense documentation.

5.2.2 Extraction and Comparison

For the evaluation process, we focused on testing HealthVision AI against the ground-truth data in a comprehensive manner. Comparisons with iOCR were based on historical performance data, as the client had been using iOCR for an extended period and had tracked its performance over time.

The process involved:

- Processing the entire dataset through HealthVision AI.
- Comparing the extracted information from HealthVision AI against the ground truth data.
- Using existing performance metrics of iOCR as a baseline for comparison.

5.2.3 Performance Metric

The primary measure used to evaluate performance was accuracy, a metric chosen for its direct relevance to information extraction in document processing. This choice assures the audience of the precision and validity of our evaluation.

This approach allowed us to assess the real-world performance improvements offered by HealthVision AI in the context of actual client data, providing practical insights into its effectiveness in handling both computer-typed and handwritten documents in healthcare and insurance settings. The comparison with iOCR, the pre-existing system, offers valuable context for understanding the advancements made by implementing our LLM-centric approach.

5.2.4 Key Findings

HealthVision AI achieved an 18% overall increase in accuracy for extracting 11 distinct fields from documents compared to iOCR. Notably, it demonstrated perfect accuracy for critical fields like names and dates in invoices and showed an average improvement of 15% for handwritten documents.

Field	iOCR	HealthVision AI
Name	80%	100%
Surname	77%	96%
Date	80%	100%
Doc. number	71%	84%
Amount	68%	96%
Structure Vat	73%	94%
Structure Fiscal Code	91%	94%
Structure Name	25%	42%

Table 5.1: Computer-typed Documents Extraction Performance

Field	iOCR Accuracy	HealthVision AI Accuracy
Name	68%	83%
Surname	70%	78%
Date	77%	82%
Doc. number	71%	76%
Amount	64%	62%

Table 5.2: Handwritten Documents Extraction Performance

5.2.5 Analysis

1. **Digital Document Processing:** HealthVision AI excels in processing digital documents with perfect or near-perfect accuracy for critical fields like names, dates, and amounts.
2. **Handwritten Document Analysis:** Significant improvements are observed in handwritten document processing, with increases ranging from 5% to 15% across most fields.
3. **Consistent Performance:** HealthVision AI demonstrates more consistent performance across different fields than iOCR, which showed high variability (25% to 91% accuracy).
4. **Areas for Improvement:** While HealthVision AI shows improvements in most areas, there is a slight decrease in accuracy for the 'Amount' field in handwritten documents (64% to 62%). This could be an area for future optimization.

5.2.6 Conclusion

HealthVision AI represents a substantial advancement in document processing technology, particularly in the healthcare sector. Its ability to accurately extract information from digital and handwritten documents with high precision is a response to a critical need in the industry. The system's improved performance in handwritten document analysis is particularly noteworthy, as this has traditionally been a challenging area for OCR technologies. These performance improvements translate to significant practical benefits, including reduced need for manual data entry, decreased error rates, and increased overall efficiency in document processing workflows. As HealthVision AI continues to evolve, it has the potential to revolutionize document handling in healthcare and other industries that rely heavily on accurate information extraction from diverse document types.

Chapter 6
Conclusion and Future Directions

The development and evaluation of HealthVision AI represent a significant advancement in the field of document processing for the healthcare and insurance industries. This project has demonstrated the potential of integrating Large Language Models (LLMs) with traditional OCR technologies to create a more powerful, versatile, and accurate document analysis system.

6.1 Summary of Key Achievements.

- **Performance Improvements:** HealthVision AI achieved an overall 18% increase in accuracy compared to the traditional iOCR system, with perfect accuracy for critical fields like names and dates in medical documents. This improvement is particularly notable in the processing of handwritten documents, an area that has traditionally been challenging for OCR systems.
- **LLM-Centric Technology:** The core innovation of HealthVision AI lies in its LLM-centric approach, which represents a paradigm shift in document processing technology. By leveraging advanced Large Language Models and integrating them with Azure AI Services for enhanced OCR capabilities, HealthVision AI has created a powerful, multimodal system capable of processing both textual and visual data with unprecedented accuracy.
- **User-Centric Design:** The development of an intuitive user interface, as evidenced by the application screenshots, demonstrates a focus on user experience. Features such as document type selection, model choice, and a clear login process contribute to a system that is not only powerful but also accessible to end-users.
- **Versatility:** HealthVision AI's ability to handle both computer-typed and handwritten documents across various medical and expense document types illustrates its versatility and potential for wide-ranging applications in the health insurance claims and expense reimbursement sector.

6.2 Implications for the Industry

The success of HealthVision AI has several important implications for the healthcare and insurance industries:

- **Efficiency Gains:** The improved accuracy and speed of document processing can lead to significant time and cost savings in healthcare administration and insurance claim processing.

- **Error Reduction:** By minimizing manual data entry and improving accuracy, HealthVision AI can help reduce errors in expense reimbursement and insurance claims, potentially improving end-user satisfaction and reducing financial discrepancies.
- **Scalability:** The system's ability to handle diverse document types suggests it could be scaled to address document processing needs across various departments and even different industries.

6.3 Challenges and Limitations

Despite its successes, HealthVision AI faces some challenges:

- **Cost Considerations:** As discussed in the cost analysis section, the current implementation of HealthVision AI is more expensive than traditional OCR solutions. This cost differential may impact widespread adoption, particularly for high-volume, routine tasks.
- **Performance Variability:** While overall performance improved, some areas, such as the extraction of 'Amount' fields in handwritten documents, showed a slight decrease in accuracy. This highlights the need for continuous refinement and possibly specialized approaches for certain data types.
- **Data Privacy and Security:** Given the sensitive nature of medical and insurance documents, ensuring robust data protection measures remains a critical ongoing concern.

6.4 Future Research Directions

Based on the findings of this project, several avenues for future research and development emerge:

- **Expanded Language Support:** Enhancing the system's capabilities to process documents in multiple languages, broadening its applicability in diverse systems globally.
- **Cost Optimization:** Investigating ways to reduce operational costs without compromising performance, possibly through more efficient use of API calls or by developing hybrid models that combine LLM capabilities with more traditional, cost-effective methods for simpler tasks.
- **Advanced Natural Language Processing:** Further developing the system's ability to understand context and nuance in expertise terminology, potentially improving its performance in extracting complex information.
- **Continuous Learning Mechanisms:** Implementing features that allow the system to learn and improve from user feedback and corrections, enhancing its accuracy over time.

6.5 Closing Thoughts

The creation of HealthVision AI marks a significant milestone in the application of AI and LLMs to healthcare and insurance document processing. While challenges remain, the achievements thus far indicate the potential for AI to revolutionize these critical industries and many other fields. We acknowledge these challenges and are committed to overcoming them, as the system significantly boosts operational efficiency and precision, setting the stage for groundbreaking innovations that will revolutionize public administration and office processes.

As we look to the future, the possibilities for HealthVision AI are boundless. This project has demonstrated the profound impact of interdisciplinary collaboration, innovative technology, and most importantly, user-centered design on solving real-world problems. By continuing to push the boundaries of what is possible, we can create solutions that meet today's needs and pave the way for a brighter, more efficient, and patient-focused future in healthcare and insurance.

In essence, HealthVision AI is more than just a technological advancement, it is a step towards a future where technology and humanity work hand in hand to improve lives and outcomes. The journey does not end here, it is only the beginning of a new era of possibilities.

References

[1] ABBYY. Abbyy finereader pdf. https://www.abbyy.com/en-us/finereader/, 2024. Accessed: 2024-07-08. (Cited on page 18)

[2] Accenture. Accenture code of business ethics. https://www.accenture.com/content/dam/accenture/final/a-com-migration/pdf/pdf-63/accenture-cobe-brochure-english.pdf, 2022. Accessed: 2024-07-08. (Cited on page 13)

[3] Accenture. Accenture to invest $3 billion in ai to accelerate clientsâ reinvention. https://newsroom.accenture.com/news/2023/accenture-to-invest-3-billion-in-ai-to-accelerate-clients-reinvention, 2023. Accessed: 2024-07-08. (Cited on page 13)

[4] Accenture. Accenture fact sheet q3 fiscal 2024. https://newsroom.accenture.com/fact-sheet, 2024. Accessed: 2024-07-08. (Cited on pages 10 and 13)

[5] Accenture. Awards & recognition. https://www.accenture.com/cr-en/about/awards-recognition, 2024. Accessed: 2024-07-08. (Cited on page 13)

[6] Accenture. Data and ai services solutions. https://www.accenture.com/il-en/services/data-ai, 2024. Accessed: 2024-07-08. (Cited on page 13)

[7] Accenture. Environmental sustainability. https://www.accenture.com/it-it/about/responsible-business/environment, 2024. Accessed: 2024-07-08. (Cited on page 13)

[8] Accenture. Responsible company & citizen. https://www.accenture.com/fi-en/about/responsible-business/responsible-company-citizen, 2024. Accessed: 2024-07-08. (Cited on page 13)

[9] Adobe. Ocr meaning: What is ocr and why it's important. https://www.adobe.com/acrobat/guides/what-is-ocr.html, 2024. Accessed: 2024-07-08. (Cited on pages 15, 16, and 17)

[10] AIMultiple. State of ocr in 2024: Is it dead or a solved problem? https://research.aimultiple.com/ocr-technology/, January 2024. Accessed: 2024-07-08. (Cited on page 18)

[11] Amazon Web Services. What is ocr? - optical character recognition explained. https://aws.amazon.com/what-is/ocr/, 2024. Accessed: 2024-07-08. (Cited on pages 16, 17, and 18)

[12] Azure AI Services. Document intelligence di azure ai. https://azure.microsoft.com/it-it/products/ai-services/ai-document-intelligence, 2024. Accessed: 2024-07-08. (Cited on page 31)

[13] Azure AI Services. Ocr - optical character recognition. https://learn.microsoft.com/en-us/azure/ai-services/computer-vision/overview-ocr, 2024. Accessed: 2024-07-08. (Cited on page 31)

[14] Blanc Labs. Ai vs. ocr: Understanding the differences and applications. https://blanclabs.com/blog/ai-vs-ocr, 2024. Accessed: 2024-07-08. (Cited on page 27)

[15] Blue Prism. How to use generative ai for document extraction and processing. https://www.blueprism.com/resources/blog/generative-ai-document-extraction-processing/, 2023. Accessed: 2024-07-08. (Cited on page 23)

[16] Sahil Chugani, Reagan Bourne, and Carlos Bravo. A brief overview of large language models. https://fsulib.com/a-brief-overview-of-large-language-models/, 2024. Accessed: 2024-07-08. (Cited on pages 19 and 20)

[17] Jill Daley and Esther Adediran. What is optical character recognition? ocr explained by google. https://cloud.google.com/blog/products/ai-machine-learning/what-is-ocr. (Cited on page 8)

[18] Jill Daley and Esther Adediran. What is ocr? | google cloud blog. https://cloud.google.com/blog/products/ai-machine-learning/what-is-ocr/, 2024. Accessed: 2024-07-08. (Cited on page 18)

[19] DeepLearning.AI. Building systems with the chatgpt api. https://learn.deeplearning.ai/courses/chatgpt-building-system, 2024. Accessed: 2024-07-08. (Cited on page 42)

[20] DeepLearning.AI. Generative ai for everyone. https://www.deeplearning.ai/courses/generative-ai-for-everyone/, 2024. Accessed: 2024-07-08. (Cited on page 42)

[21] DeepLearning.AI. Langchain chat with your data. https://learn.deeplearning.ai/courses/langchain-chat-with-your-data, 2024. Accessed: 2024-07-08. (Cited on page 42)

[22] DeepLearning.AI. Langchain for llm application development. https://learn.deeplearning.ai/courses/langchain, 2024. Accessed: 2024-07-08. (Cited on page 42)

[23] DeepLearning.AI. Langchain functions, tools, and agents. https://learn.deeplearning.ai/courses/functions-tools-agents-langchain, 2024. Accessed: 2024-07-08. (Cited on page 42)

[24] DeepLearning.AI. Prompt engineering with llama-2. https://learn.deeplearning.ai/courses/prompt-engineering-with-llama-2, 2024. Accessed: 2024-07-08. (Cited on page 42)

[25] Somdip Dey. Which ethical implications of generative ai should companies focus on? https://www.forbes.com/sites/forbestechcouncil/2023/10/17/which-ethical-implications-of-generative-ai-should-companies-focus-on/, 2023. Accessed: 2024-07-08. (Cited on page 24)

[26] Docsumo. How has advanced image pre-processing improved ocr accuracy? https://www.docsumo.com/blog/how-has-advanced-image-pre-processing-improved-ocr, 2024. Accessed: 2024-07-08. (Cited on page 18)

[27] Docsumo. What is optical character recognition (ocr)? https://www.docsumo.com/blog/optical-character-recognition-definition, 2024. Accessed: 2024-07-08. (Cited on page 17)

[28] DocuClipper. Ai in document processing. https://docuclipper.com/blog/ai-in-document-processing, 2024. Accessed: 2024-07-08. (Cited on page 28)

[29] Elastic. What are large language models? https://www.elastic.co/what-is/large-language-models, 2024. Accessed: 2024-07-08. (Cited on page 19)

[30] Fintelite AI. Cost benefits of ocr document extraction for businesses. https://fintelite.ai/cost-benefits-of-ocr-document-extraction-for-businesses/, 2024. Accessed: 2024-07-08. (Cited on page 9)

[31] Google Cloud. Multimodal ai. https://cloud.google.com/use-cases/multimodal-ai, 2023. Accessed: 2024-07-08. (Cited on page 25)

[32] Google Cloud. Detect text in images | cloud vision api. https://cloud.google.com/vision/docs/ocr, 2024. Accessed: 2024-07-08. (Cited on page 9)

[33] Google Cloud. Document ai | google cloud. https://cloud.google.com/document-ai, 2024. Accessed: 2024-07-08. (Cited on pages 22 and 23)

[34] HETT Insights. How does ai reduce costs in healthcare? https://blog.hettshow.co.uk/how-does-ai-reduce-costs-in-healthcare, 2024. Accessed: 2024-07-08. (Cited on page 14)

[35] HyperVerge. Enhancing ocr with ai. https://hyperverge.co/blog/ai-enhanced-ocr, 2024. Accessed: 2024-07-08. (Cited on page 28)

[36] HyperVerge. Ocr in the insurance industry. https://hyperverge.co/blog/ocr-insurance/, 2024. Accessed: 2024-07-08. (Cited on page 9)

[37] IBM. What is generative ai? https://research.ibm.com/blog/what-is-generative-AI, 2023. Accessed: 2024-07-08. (Cited on page 22)

[38] IBM. What are large language models (llms)? https://www.ibm.com/topics/large-language-models, 2024. Accessed: 2024-07-08. (Cited on page 20)

[39] IBM. What is ocr (optical character recognition)? https://www.ibm.com/topics/optical-character-recognition, 2024. Accessed: 2024-07-08. (Cited on pages 15, 16, and 17)

[40] Kili Technology. Large language models (llms). https://kili-technology.com/large-language-models-llms, 2024. Accessed: 2024-07-08. (Cited on page 21)

[41] LangChain. Introduction to langchain. https://python.langchain.com/docs/get_started/introduction, 2024. Accessed: 2024-07-08. (Cited on page 40)

[42] LangChain. Langchain github repository. https://github.com/langchain-ai/langchain, 2024. Accessed: 2024-07-08. (Cited on page 40)

[43] Lark. Ethical issues in generative ai. https://www.larksuite.com/en_us/topics/ai-glossary/ethical-issues-in-generative-ai, 2023. Accessed: 2024-07-08. (Cited on page 23)

[44] Yuliang Liu, Zhang Li, Hongliang Li, Wenwen Yu, Yang Liu, Mingxin Huang, Dezhi Peng, Mingyu Liu, Mingrui Chen, Chunhua Shen, and Lei Zhang. On the hidden mystery of ocr in large multimodal models, 2023. (Cited on pages 9 and 14)

[45] Microsoft. Azure ai document intelligence. https://azure.microsoft.com/en-us/products/ai-services/ai-document-intelligence, 2024. Accessed: 2024-07-08. (Cited on pages 26 and 32)

[46] Microsoft. Gpt-4 turbo with vision - concepts. https://learn.microsoft.com/en-us/azure/ai-services/openai/concepts/gpt-with-vision, 2024. Accessed: 2024-07-08. (Cited on pages 27 and 32)

[47] Microsoft Azure. Azure ai vision with ocr and ai. https://azure.microsoft.com/en-us/services/cognitive-services/computer-vision/, 2024. Accessed: 2024-07-08. (Cited on pages 10 and 18)

[48] Microsoft Azure. What is azure? https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-azure/, 2024. Accessed: 2024-07-08. (Cited on page 26)

[49] Motics AI. Streamlining administrative tasks in healthcare with ai. https://www.motics.ai/post/streamlining-administrative-tasks-in-healthcare-with-ai, June 2024. Accessed: 2024-07-08. (Cited on page 14)

[50] News Medical. Ai in healthcare: A double-edged sword? study reveals impact on diagnostic accuracy. https://www.news-medical.net/news/20231219/AI-in-healthcare-A-double-edged-sword-Study-reveals-impact-on-diagnostic-accuracy.aspx, December 2023. Accessed: 2024-07-08. (Cited on page 14)

[51] OpenAI. Gpt-3 powers the next generation of apps. https://openai.com/blog/gpt-3-apps/, 2020. Accessed: 2024-07-08. (Cited on page 22)

[52] OpenAI. Chatgpt can now see, hear, and speak. https://openai.com/blog/chatgpt-can-now-see-hear-and-speak, 2023. Accessed: 2024-07-08. (Cited on page 42)

[53] OpenAI. Gpt-4v(ision) system card. https://cdn.openai.com/papers/GPTV_System_Card.pdf, 2023. Accessed: 2024-07-08. (Cited on pages 31 and 42)

[54] OpenAI. Api reference. https://platform.openai.com/docs/api-reference, 2024. Accessed: 2024-07-08. (Cited on page 42)

[55] OpenAI. Gpt-4. https://openai.com/research/gpt-4, 2024. Accessed: 2024-07-08. (Cited on page 42)

[56] OpenAI. Prompt engineering techniques. https://platform.openai.com/docs/guides/prompt-engineering, 2024. Accessed: 2024-07-08. (Cited on pages 25, 26, and 42)

[57] Joe Filcik Patrick Farley and Kathryne Browne. Gpt-4 turbo with vision concepts. https://learn.microsoft.com/ro-ro/azure/ai-services/openai/concepts/gpt-with-vision. (Cited on page 8)

[58] Vincent Perot, Aditya Kusupati, Hao Wu, Romain Sauvestre, Ashwin Swaminathan, Stefano Soatto, and Aniruddha Kembhavi. Lmdx: Language model-based document information extraction and localization, 2023. (Cited on page 9)

[59] Prompt Engineering Guide. Information extraction with llms. https://www.promptingguide.ai/prompts/information-extraction, 2024. Accessed: 2024-07-08. (Cited on pages 26 and 34)

[60] Roboflow. Best ocr models for text recognition in images. https://blog.roboflow.com/best-ocr-models-text-recognition/, 2024. Accessed: 2024-07-08. (Cited on pages 17 and 18)

[61] Roots Automation. How different document ai models process automation of documents. https://www.rootsautomation.com/blogs-and-news/segmenting-documents-with-llms-and-multimodal-document-ai-part-1, 2024. Accessed: 2024-07-08. (Cited on page 25)

[62] Parikshit Sharma, Priyanka Sharma, and Priyanka Sharma. Advancements in ocr: A deep learning algorithm for enhanced text recognition. *International Journal of Engineering and Advanced Technology*, 12(6):131–137, 2023. (Cited on page 17)

[63] Yongxin Shi, Jingye Chen, Lianwen Jin, Zhaoyang Liu, Matthias Rottmann, and Rui Zhang. Exploring ocr capabilities of gpt-4v(ision): A quantitative and in-depth evaluation, 2023. (Cited on pages 9 and 14)

[64] Stack Overflow. Image processing to improve tesseract ocr accuracy. `https://stackoverflow.com/questions/9480013/image-processing-to-improve-tesseract-ocr-accuracy`, 2012. Accessed: 2024-07-08. (Cited on page 31)

[65] Streamlit. Building conversational apps with streamlit. `https://docs.streamlit.io/knowledge-base/tutorials/build-conversational-apps`, 2024. Accessed: 2024-07-08. (Cited on page 42)

[66] Streamlit. Llm quickstart with streamlit. `https://docs.streamlit.io/knowledge-base/tutorials/llm-quickstart`, 2024. Accessed: 2024-07-08. (Cited on page 42)

[67] Streamlit. Streamlit documentation. `https://docs.streamlit.io/`, 2024. Accessed: 2024-07-08. (Cited on pages 40 and 41)

[68] Streamlit. Streamlit github repository. `https://github.com/streamlit/streamlit`, 2024. Accessed: 2024-07-08. (Cited on pages 40 and 41)

[69] TechTarget. Generative ai ethics: 8 biggest concerns and risks. `https://www.techtarget.com/searchenterpriseai/tip/Generative-AI-ethics-8-biggest-concerns`, 2023. Accessed: 2024-07-08. (Cited on pages 23 and 24)

[70] TechTarget. 19 of the best large language models in 2024. `https://www.techtarget.com/whatis/definition/large-language-model-LLM`, 2024. Accessed: 2024-07-08. (Cited on page 19)

[71] Towards Data Science. Pre-processing in ocr. `https://towardsdatascience.com/pre-processing-in-ocr-fc231c6035a7`, 2023. Accessed: 2024-07-08. (Cited on page 31)

[72] TrueFoundry. Transformer architecture in large language models. `https://www.truefoundry.com/blog/transformer-architecture`, 2024. Accessed: 2024-07-08. (Cited on page 20)

[73] UCSD LibGuides. Challenges and possibilities of generative ai. `https://ucsd.libguides.com/c.php?g=1322935&p=9734831`, 2024. Accessed: 2024-07-08. (Cited on page 23)

[74] UiPath. Transforming healthcare document processing with ai. `https://www.uipath.com/blog/ai/healthcare-document-processing-with-ai`, 2023. Accessed: 2024-07-08. (Cited on page 9)

[75] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Åukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, pages 5998–6008, 2017. (Cited on pages 20 and 22)

[76] Veryfi. Tesseract ocr vs. cnn-based ocr: Which is right for you? `https://www.veryfi.com/technology/tesseract-ocr-vs-cnn-based-ocr/`, 2024. Accessed: 2024-07-08. (Cited on pages 17 and 18)

[77] W3Schools. Git tutorial. `https://www.w3schools.com/git/default.asp?remote=github`, 2024. Accessed: 2024-07-08. (Cited on page 42)

[78] Wikipedia contributors. Accenture. `https://it.wikipedia.org/wiki/Accenture`, 2024. Accessed: 2024-07-08. (Cited on page 13)

[79] Wikipedia contributors. Large language model. `https://en.wikipedia.org/wiki/Large_language_model`, 2024. Accessed: 2024-07-08. (Cited on page 19)

[80] Wikipedia contributors. Optical character recognition. `https://en.wikipedia.org/wiki/Optical_character_recognition`, 2024. Accessed: 2024-07-08. (Cited on pages 15 and 16)

[81] Hao Wu, Xiaoyu Lu, and Hanyu Wang. The application of artificial intelligence in health care resource allocation before and during the covid-19 pandemic: Scoping review. *JMIR AI*, 2(1):e38397, 2023. Accessed: 2024-07-08. (Cited on page 14)

[82] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, and Quoc V. Le. Xlnet: Generalized autoregressive pretraining for language understanding, 2024. (Cited on page 24)