



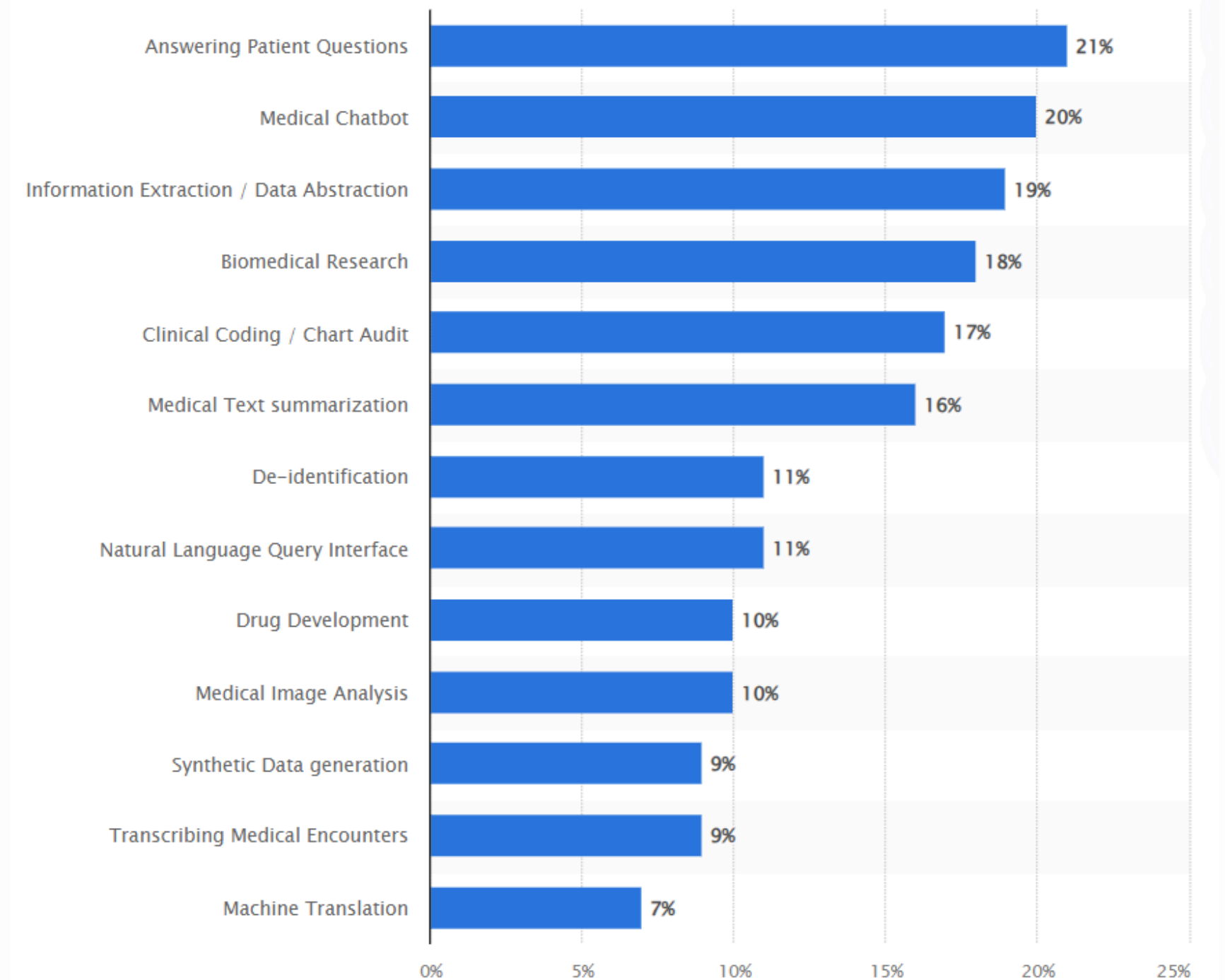
PRIVATE LLM FOR HEALTHCARE FILES

A custom Large Language Model (LLM) for securely processing, analyzing, and summarizing healthcare files, ensuring privacy.

Aarthy R

LLMs FOR HEALTHCARE

As of 2024, at least one fifth of respondents working in healthcare organizations reported that they used large language models for answering patient questions and medical chatbots. Furthermore, 18 percent of healthcare organizations used LLMs for biomedical research.



<https://www.statista.com/statistics/1469378/uses-for-llm-use-in-healthcare-in-the-us/>



Problem

Data Security

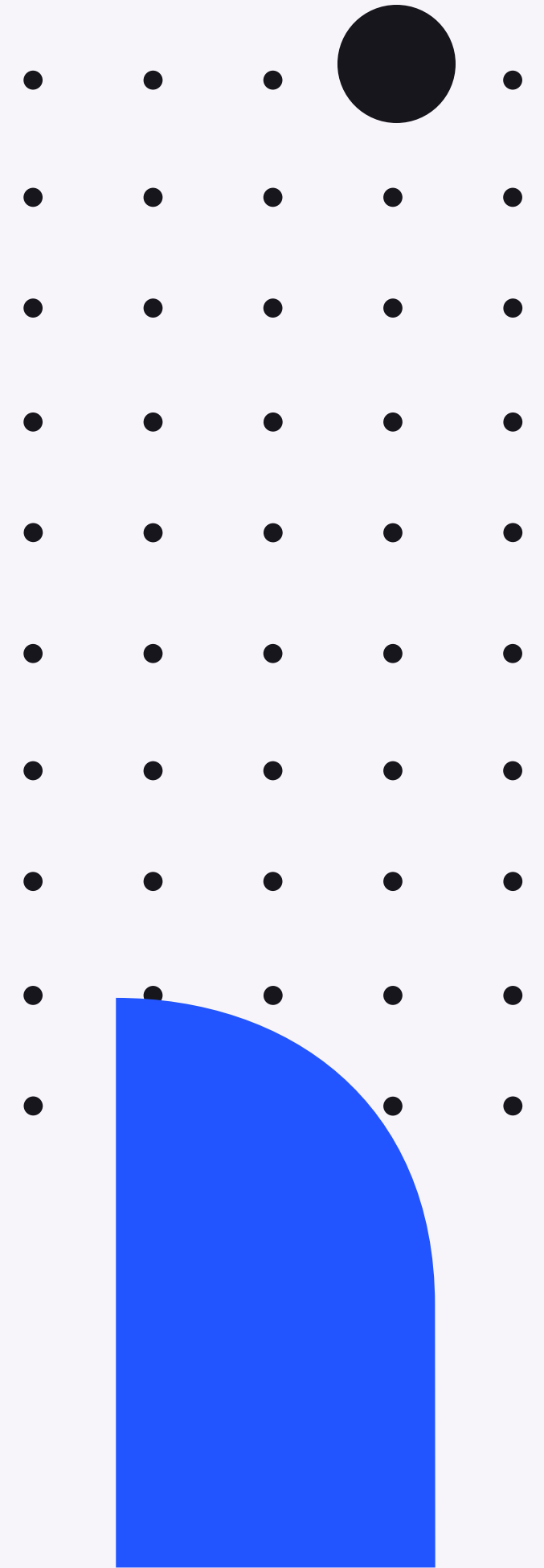
Healthcare organizations face challenges in securely processing sensitive patient data due to privacy regulations like HIPAA, which demand strict compliance and confidentiality.

Limited General Models

General-purpose language models struggle with domain-specific medical knowledge, leading to inaccurate data processing, analysis, and insights for healthcare professionals.

Need for Customization

Healthcare providers need a private, fine-tuned LLM that can accurately process, analyze, and summarize healthcare files while ensuring data privacy and regulatory compliance.



Domain-Specific Contextual LLM

Using a fine-tuned Large Language Model that can query healthcare data with context, improving the accuracy of insights drawn from medical files and reports.

Enhanced Data Security

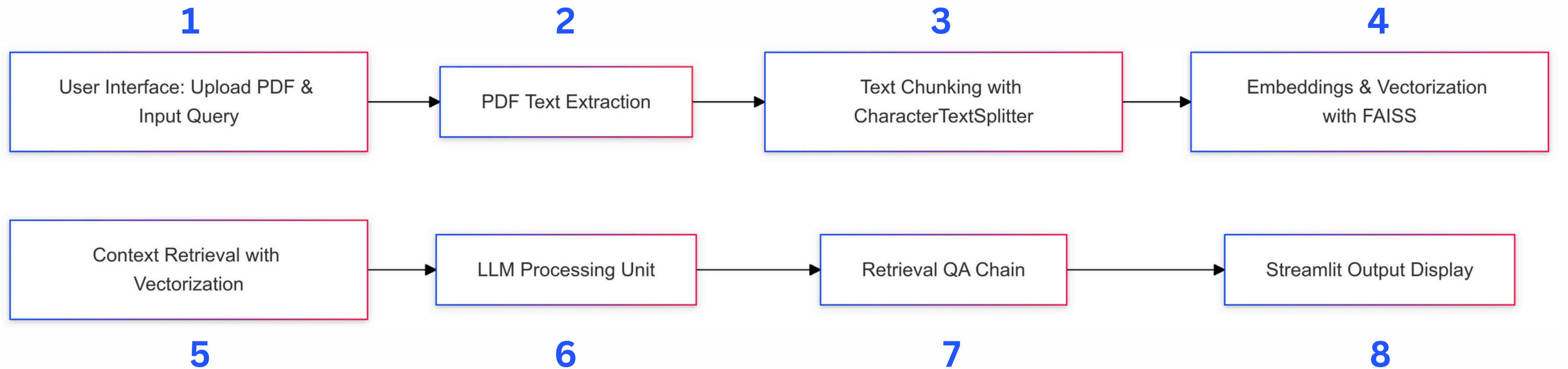
The solution operates on a private LLM setup, ensuring secure and compliant handling of sensitive healthcare data, adhering to regulations like HIPAA.

Efficient Data Querying

The model supports efficient querying of healthcare data, enabling healthcare professionals to extract relevant information quickly while ensuring data privacy and integrity.

Proposed Solution

Architecture

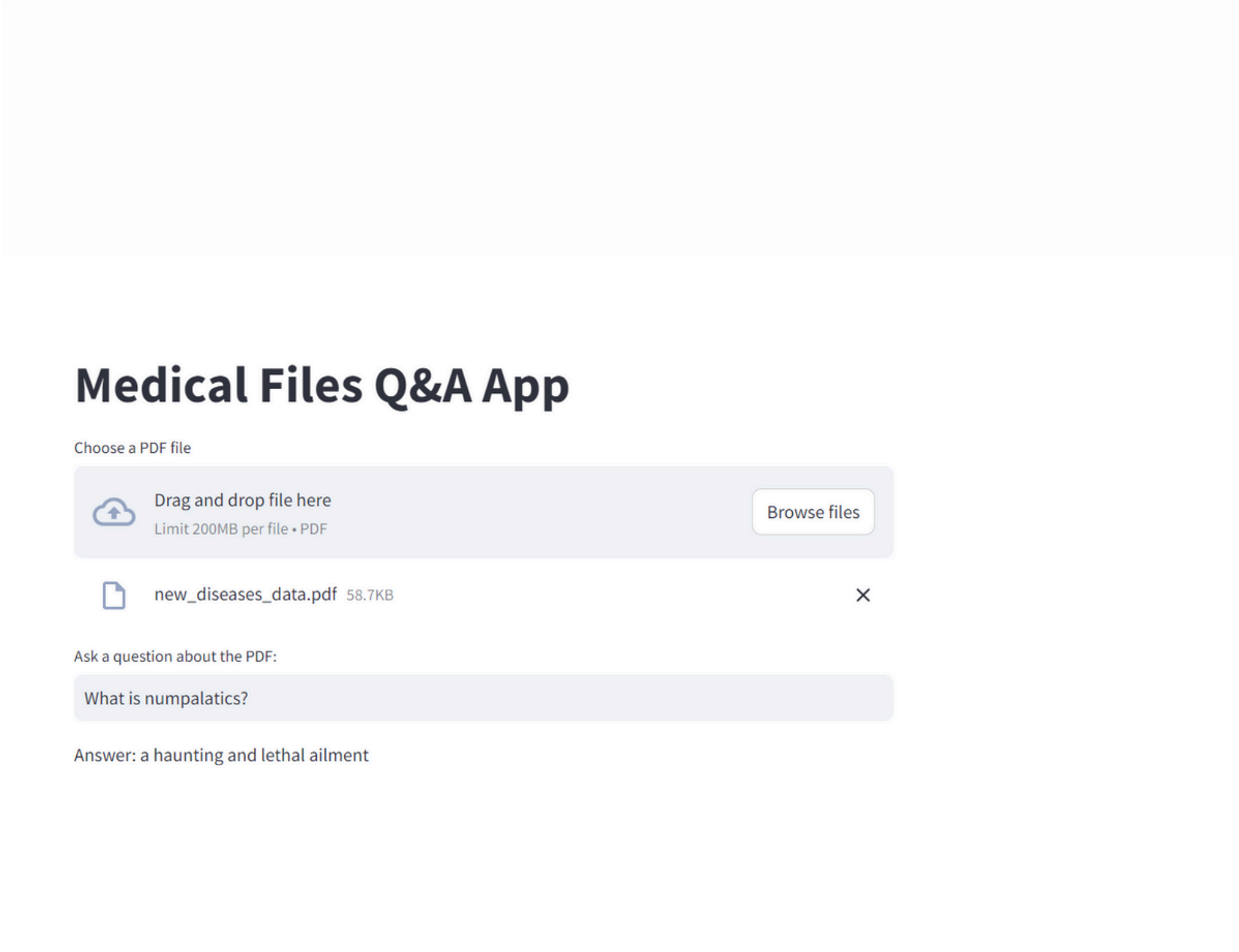
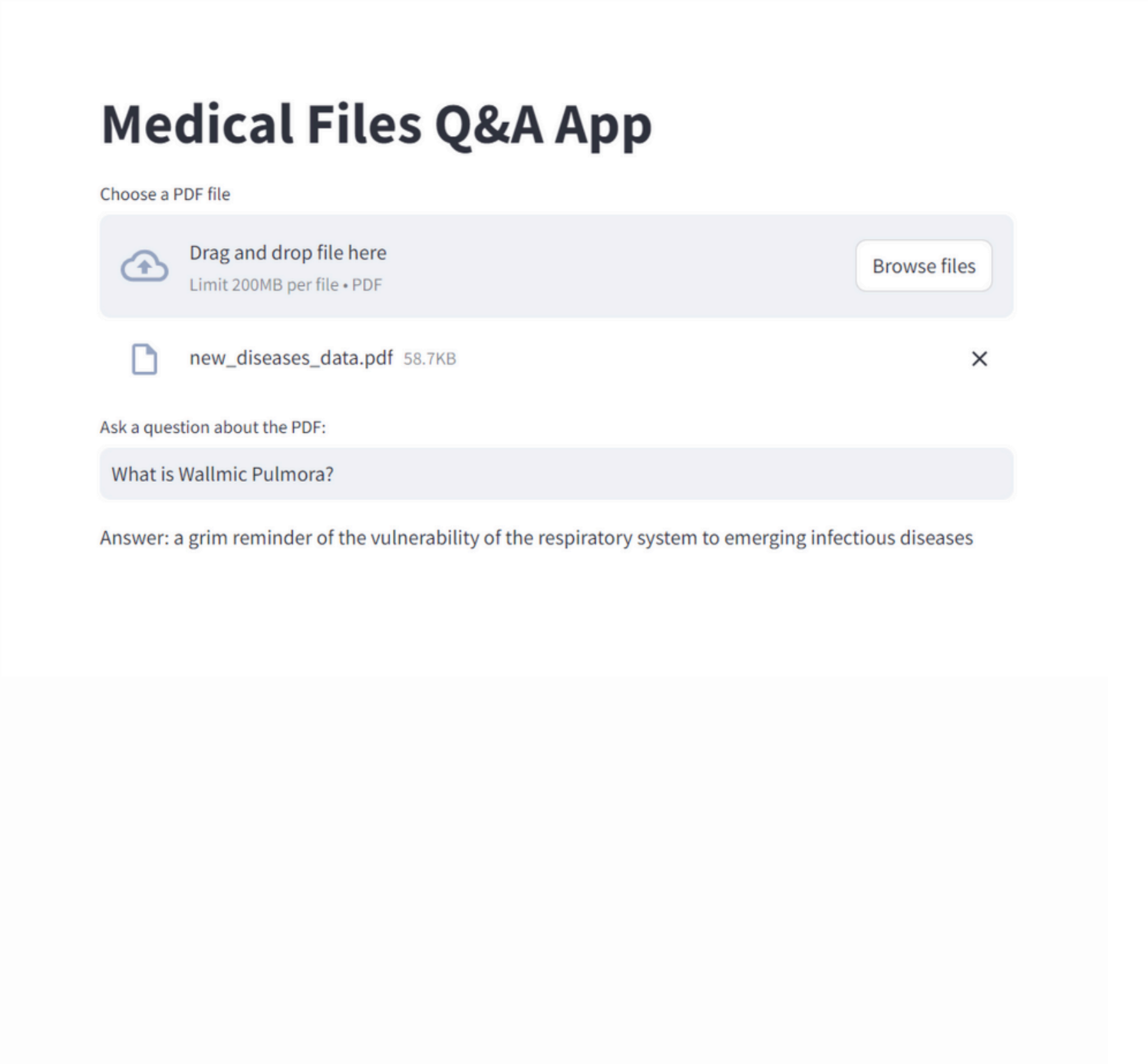


Future Enhancements



- **Security & Data Privacy:** Implement end-to-end encryption for data storage and transmission, ensuring secure handling of sensitive healthcare data.
- **Login & Authentication:** Add user authentication with multi-factor authentication (MFA) to restrict access to authorized users only.
- **Role-Based Access Control:** Implement role-based permissions to manage different levels of access, ensuring only authorized healthcare professionals can view specific data.
- **Regulatory Compliance:** Ensure compliance with regulations such as HIPAA or GDPR by incorporating data anonymization and proper audit logging.
- **Improved Accuracy:** Fine-tune the LLM with domain-specific medical datasets to improve the precision of answers related to healthcare.
- **Performance Optimization:** Implement caching mechanisms and optimize the use of embeddings to reduce query response time and system latency.
- **Scalability:** Deploy the system on cloud infrastructure with auto-scaling capabilities to handle larger volumes of queries and PDF files efficiently.

Screenshots



Screenshots

```
from tqdm.autonotebook import trange
Downloading (...)7f436/.gitattributes: 100% 1.48k/1.48k [00:00<00:00, 97.4kB/s]
Downloading (...)_Pooling/config.json: 100% 270/270 [00:00<00:00, 17.9kB/s]
Downloading (...)2_Dense/config.json: 100% 116/116 [00:00<00:00, 8.16kB/s]
Downloading pytorch_model.bin: 100% 3.15M/3.15M [00:00<00:00, 13.5MB/s]
Downloading (...)0daf57f436/README.md: 100% 66.3k/66.3k [00:00<00:00, 5.58MB/s]
Downloading (...)af57f436/config.json: 100% 1.52k/1.52k [00:00<00:00, 144kB/s]
Downloading (...)ce_transformers.json: 100% 122/122 [00:00<00:00, 7.41kB/s]
Downloading pytorch_model.bin: 100% 4.96G/4.96G [00:45<00:00, 86.4MB/s]
Downloading (...)nce_bert_config.json: 100% 53.0/53.0 [00:00<00:00, 1.89kB/s]
Downloading (...)cial_tokens_map.json: 100% 2.20k/2.20k [00:00<00:00, 102kB/s]
Downloading spiece.model: 100% 792k/792k [00:00<00:00, 54.3MB/s]
Downloading (...)7f436/tokenizer.json: 100% 2.42M/2.42M [00:00<00:00, 23.2MB/s]
Downloading (...)okenizer_config.json: 100% 2.40k/2.40k [00:00<00:00, 180kB/s]
Downloading (...)f57f436/modules.json: 100% 461/461 [00:00<00:00, 35.4kB/s]
ad INSTRUCTOR Transformer
```

```
[ ] path_to_pdf = ['./content/MyDrive/new_diseases_data.pdf']

from PyPDF2 import PdfReader

def extract_text_from_pdf(pdf_path):
    """
    Extracts all text from a PDF file.


    Parameters:
    pdf_path (str): Path to the PDF file.

    Returns:
    str: Extracted text from the PDF.
    """
    text = ""
    with open(pdf_path, 'rb') as pdf_file:
        pdf_reader = PdfReader(pdf_file)
        for page in pdf_reader.pages:
            text += page.extract_text()
    return text

# Example usage:
# Replace 'your_pdf_file.pdf' with the actual filename of the PDF you want to use.
pdf_path = './content/MyDrive/new_diseases_data.pdf'
pdf_text = extract_text_from_pdf(pdf_path)
```



Screenshots

```
[ ] sol=bot(ques[1])
    print(ques[1])
    print(sol['result'])
```

 what are the stages of diseases progression in Ramtronephiach Oculosis

The disease's progression in Ramtronephiach Oculosis can be classified into four stages:
1. Early symptoms: This stage involves mild symptoms such as blurry vision, light sensitivity, and occasional eye irritation.
2. Acute symptoms: This stage involves severe symptoms such as eye pain, headaches, and retinal detachment.
3. Chronic symptoms: This stage involves severe symptoms such as complete loss of vision and permanent damage to the optic nerve.
4. Severe symptoms: This stage involves severe symptoms such as complete blindness and loss of vision.

```
▶ sol=bot(ques[2])
   print(ques[2])
   print(sol['result'])
```

 what is mortality rate in Wallmic Pulmora

The mortality rate of Wallmic Pulmora is not known, as the disease is relatively rare.

```
▶ ques=['what are the origins od Numpalofich Legatrosis',
        'what are the stages of diseases progression in Ramtronephiach Oculosis',
        'what is mortality rate in Wallmic Pulmora',
        'is Numpalactics incubation period short?',
        ' what is Numpalactic',
        ' What are the symptoms of a disease that causes blindness?',
        'what are the origins of Ramtronephiach Oculosis']
```

```
[ ] sol=bot(ques[0])
    print(ques[0])
    print(sol['result'])
```

 what are the origins od Numpalofich Legatrosis



THANK YOU

raarthy1553@gmail.com

<https://github.com/Aarthy1553/>

