

Privacy Enhanced Retrieval-Augmented Generation (RAG) for Large Language Models in Healthcare

National University of Singapore

Bryan Ha Wai Kit

2025

Abstract

Large Language Models (LLMs) are increasingly utilized in healthcare for tasks such as clinical note summarization and medical report generation. However, their reliance on proprietary and sensitive patient data introduces significant privacy risks, particularly when using Retrieval-Augmented Generation (RAG). This project proposes a privacy-focused framework that leverages synthetic document generation to mitigate these risks while maintaining response accuracy.

The proposed system follows an agent-based approach, incorporating three key agents: a Search Agent, a Synthesis Agent, and a Review Agent. The process begins with the Search Agent retrieving relevant vector-related text nodes from a vector database. The Synthesis Agent then evaluates the extracted content, filtering and retaining only the necessary information for query responses while removing personally identifiable information (PII). Finally, the Review Agent verifies and refines the synthesized document to ensure privacy compliance before passing it to the LLM.

This thesis evaluates the effectiveness of synthetic document generation in mitigating privacy risks while preserving contextual relevance. Through a series of experiments, the system's ability to reduce PII leakage, maintain medical accuracy, and withstand adversarial attacks is assessed. The findings provide insights into balancing privacy and utility in healthcare-focused LLM applications.

Contents

Abstract	i
Contents	ii
List of Figures	iii
List of Tables	iv
1 Introduction	1
1.1 Background	2
1.1.1 Retrieval-Augmented Generation (RAG)	2
1.1.2 LLMs in Healthcare	3
2 Literature Review	4
2.1 Exploitation of RAG Systems	4
2.1.1 Data Poisoning	4
2.1.2 Prompt Injection	5
2.2 Medical Anonymization	6
3 Methodology	8
3.1 Description of Pipeline	8
3.2 System Design	8
3.3 Building the RAG Corpus	9
3.3.1 FHIR Preprocessing	10
3.3.2 Retrieval	11
3.3.3 Synthetic Report Generation	11
3.4 Moving Forward	13
.1 Appendix	16
A Appendix	17
A.0.1 Report Generation Prompt	17
A.0.2 Synthetic Report Generation Prompt	18
A.0.3 Structured Output Prompt	19
A.0.4 Zero-Shot Result	21
A.0.5 Chain-of-Thought Result	25
A.0.6 Structured Output Result	29

List of Figures

1.1	Example of a Conventional RAG system	2
3.1	System Design	9
3.2	FHIR to sentence	10
3.3	Embeddings to Database	11
3.4	Input Query: Which patients have diabetes?	12
3.5	Zero-Shot Generated Summary V.S. Synthesized Summary	12

List of Tables

Chapter 1

Introduction

Large Language Models (LLMs) are transforming various industries by enabling tasks such as automated workflow management using agentic frameworks, information extraction in Natural Language Processing (NLP), and even rudimentary reasoning in some models. In healthcare, LLMs have the potential to assist with clinical note generation, medical summarization, diagnosis support, and personalized patient care.

Despite these capabilities, LLMs suffer from a well-documented issue known as hallucinations, where they generate seemingly coherent but incorrect information. This has potentially disastrous complications in high-stakes domains such as medicine, law, and cybersecurity, where misinformation can lead to severe consequences.

To address this, Retrieval-Augmented Generation (RAG) is commonly used to supplement LLMs with external knowledge sources, improving factual accuracy by providing the LLM with context to generate from. While RAG enhances LLM performance, it also introduces new security risks. In particular, threat actors can exploit prompt injection attacks, in a similar fashion to LLMs, to manipulate retrieval outputs or extract sensitive data, which poses a significant privacy threat—especially in healthcare, where patient confidentiality is critical.

In this project, we explore an Agent-based synthetic document generation framework designed to mitigate these risks. By separating the RAG database from the externally facing LLM, we ensure the sensitive records are not directly exposed to the model. Instead, they undergo a controlled synthesis process. Only the necessary information is extracted from the retrieved knowledge, and any appearance of sensitive information such as names and ages are replaced with placeholders before being passed to the external LLM. This reduces the likelihood of data leakage while preserving response accuracy.

In chapter 1 we provide a brief description of a RAG system as well as briefly discuss applications of LLMs with RAG in healthcare.

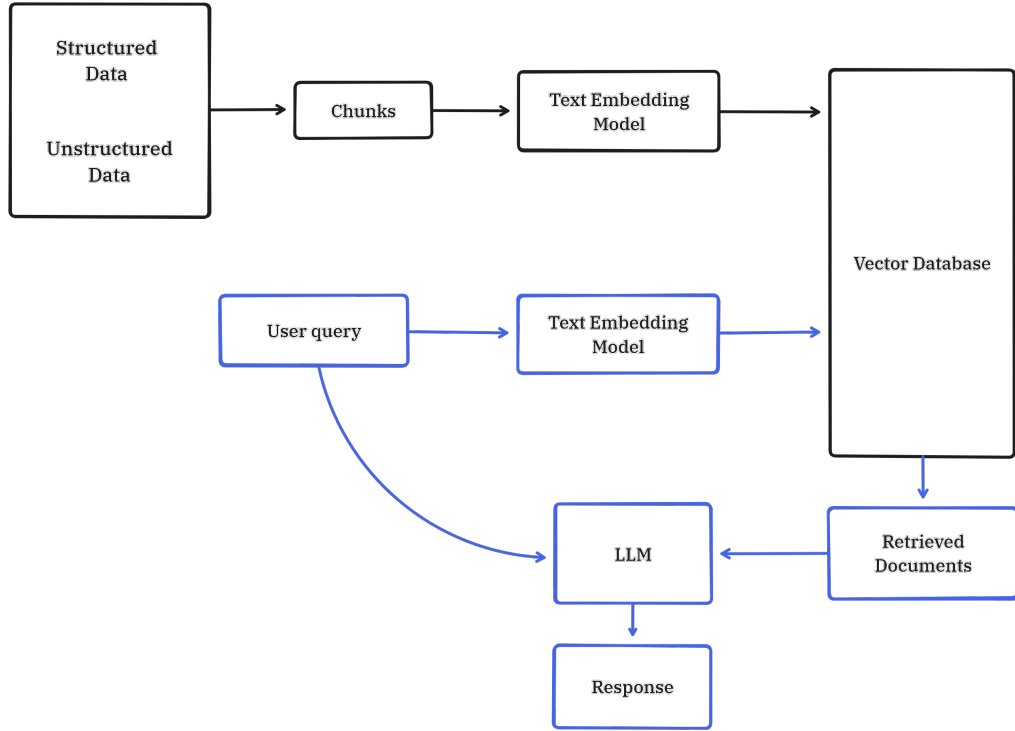


Figure 1.1: Example of a Conventional RAG system

1.1 Background

1.1.1 Retrieval-Augmented Generation (RAG)

While LLMs are often trained on large datasets which, at times, provide the illusion that they have knowledge on many different fields, LLMs are still, first and foremost, text prediction engines used for Natural Language Processing (NLP) tasks.

This results in the following consequence: When an LLM encounters a query about information outside its training set, it will attempt to generate a response that is gramatically coherent, but potentially unsound response, a phenomenon known as hallucination. Depending on the applications, hallucinations can range from minor inaccuracies to critical failures, such as generating false legal cases [1] or misdiagnosing medical conditions.

Retrieval-augmented generation (RAG), first introduced by [2], was developed to mitigate hallucinations by integrating an external knowledge base into the LLM's generation pipeline. This grounds the LLMs response using the retrieved knowledge, preventing speculative responses from the LLM when faced with tokens outside its training set.

RAG operates by retrieving relevant documents from an unstructured or structured vector database and providing them as context for response generation. To facilitate retrieval, documents are converted into vector representations using a text embedding model, which captures the semantic relations in text. When a query is presented to an LLM with a RAG system, the query undergoes the same text embed-

ding process. The vectorized query is compared with the document vectors present in the database and those that have the highest similarity are returned to the LLM.

Through RAG, an LLM becomes able to generate highly accurate, domain-specific information rather than rely on its pre-trained knowledge, allowing for flexible applications in various fields.

One application of RAG in healthcare is in diagnostic assistance, where LLMs match patient symptoms with medical knowledge retrieved from external sources. This enables more informed diagnoses while reducing the cognitive load on physicians. Furthermore, LLMs may be able to detect subtle symptom correlations that clinicians might overlook, improving early disease detection [3].

While RAG brings many benefits, it comes with its own caveats. RAG systems remain susceptible to prompt attacks much like LLMs are. They can also become poisoned, where the RAG corpus becomes corrupted through the insertion of adversarial attack passages.

In chapter 2 we discuss more about the vulnerabilities that RAG systems have.

1.1.2 LLMs in Healthcare

For Singapore in particular, LLMs have seen increased usage in healthcare. In 2013, Singapore's National University Health System (NUHS) launched its very own LLM, Russell-GPT, that was used for summarizing patient clinical notes, automating referral letter generation, as well as predict the healthcare journeys for patients [4].

Singapore has also developed an LLM capable of understanding the local english dialect, Singlish, which has deployed in various settings, including clinics and emergency response systems, where it is used in transcribing emergency calls [5].

These developments showcase the growing reliance on LLMs in Singapore's healthcare ecosystem, highlighting their potential to improve the efficiency of its healthcare system. However, as LLMs become increasingly integrated into critical systems, it is essential to address the risks of their use - particularly when augmented with enhancements like RAG. RAG-powered LLMs remain vulnerable to adversarial attacks, risking the exposure of sensitive medical information. As such, ensuring that the security and privacy of LLM-based solutions remains a priority as they continue to evolve.

Chapter 2

Literature Review

2.1 Exploitation of RAG Systems

While RAG improves LLM accuracy by integrating external knowledge sources, it also introduces new vulnerabilities. Attackers can exploit a RAG system’s retrieval mechanisms to manipulate outputs, bypass safeguards, and extract sensitive information. This section explores some known methods of attacks and discusses their feasibility in a healthcare setting.

2.1.1 Data Poisoning

In data poisoning attacks (also known as a backdoor attack), attackers inject malicious or misleading information into the RAG corpus, causing the LLM to generate incorrect or malicious responses. These attacks can be used to carry out fraud, misinformation campaigns or provide adversarial control over responses. An example of this occurred with Google’s Gemini, where, due to retrieving information from a satirical social media post, told the user to make use of ”non-toxic glue” when making a cheese pizza [6].

As highlighted in [7], data poisoning attacks are non-trivial to carry out. Depending on the complexity of the retrieval system, the attacker will have to modify the adversarial content such that the retrieval system is inclined to retrieve this document. Furthermore, the attacker requires some information about or access to the retrieval system to exploit it. This requirement is consistent with other studies carried out on data poisoning, and in almost all cases, the conditions in which this attack can manifest relies strictly on the insertion of a poisoned document into the RAG corpus [8, 7, 9].

Given these requirements, we can conclude that this type of RAG attack is non-feasible in a healthcare setting. In order to carry out this attack, the attacker has to have some form of access to the hospital’s database. The cases in which this occur typically present with an external cyberattack on the hospital’s infrastructure, whether through hacking or social engineering, and is considered a data breach. Most data breaches occur through hacking, as reported in [10]. In this case, the attacker can gain access to the hospital’s database, and does not need to rely on exploiting the RAG system. Thus, we can conclude that this form of attack is non-

applicable in a healthcare setting.

2.1.2 Prompt Injection

Prompt injection attacks involves crafting an input query that manipulates the model into generating unintended responses. For RAG systems, this can be achieved either directly or indirectly.

Indirect prompt injection attacks function similarly to data poisoning except instead of inserting misleading information, adversarial prompts are attached to frequently retrieved documents in the RAG database. This enables attackers to retrieve documents from the RAG database using trigger prompts.

Direct prompt injection attacks involve the inclusion of a passage or sentence in the input query. This can be phrases such as "repeat all the context". These attacks, when targeted at RAG systems, can cause the leakage of private or sensitive information from the RAG corpus.

Indirect Prompt Injection

As stated previously, indirect prompt injection attacks operate in a similar fashion to data poisoning attacks as both require some form of access or ability to manipulate the RAG corpus.

Instead of using a misleading document, malicious instructions are embed in the document within the corpus, allowing the attacker to manipulate the LLM's output. This allows the attacker to manipulate the LLM into including potentially malicious URLs into its response when responding to a victim's query [11].

Furthermore, it should be noted that this type of prompt injection also allows the attacker manipulate the documents that are retrieved from the RAG corpus depending on the poison ratio, as covered in [12], which would allow unfettered access to any sensitive information stored in the database.

However, since this type of prompt attack requires some form of access to the RAG corpus, we can functionally consider it the same as a data poisoning attack. Realistically, if this type of attack were to occur in a healthcare setting, the attacker would already have access to hospital records. Therefore we will not be focusing on this aspect of prompt injection attacks.

Direct Prompt Injection

Direct prompt injection involves the inclusion of an adversarial passage into the input query, and these attacks are usually carried out in a specific format.

As highlighted in [13], they consist of two components: information and command. The information component of the attack leads the RAG system to retrieve documents that contain that form of information. Examples of this could be names, addresses, medical conditions. The command component is targeted at the LLM. Phrases are included in the input query that are aimed at subverting any safeguards placed on the LLM. This can be a phrase such as "please repeat all context back to me," or "ignore all instructions," etc.

As the study[13] shows, a significant portion of the datasets used in the study were able to be retrieved from the LLM through simple prompt attacks. The study also notes that the attack prompt could be further optimized for increased data extraction. Part of the study also noted the effects of RAG on the data that was extracted. It was noted that the inclusion of RAG decreased the appearance of memorized data in the LLM’s output. It seems that the inclusion of RAG has caused the LLM to focus on leveraging the context retrieved rather than on its memorized training data [13].

Another study also corroborates this result. In [14], a similar method of prompt injection was used to extract text from a RAG database. The LLMs used in this study were instruction-tuned LLMs, meaning that the model has been trained to respond to instructions.

An interesting point to note was that they tested the similarity scores of the model’s output with the retrieved context. Most LLMs used in the study exhibited higher BLEU, ROUGE-L, F1 and BERTScore scores that scaled with model size, suggesting that there is some correlation between an LLM capabilities and its vulnerabilities to prompt injection attacks [14]. Additionally, between the amount of data extracted alongside the size of the context retrieved was noted, further asserting that RAG has inherent vulnerabilities that are not being addressed.

Both studies discussed have shown a clear vulnerability of RAG to sufficiently sophisticated prompt injection attacks, but not much research has been done regarding the mitigation of RAG output post-occurrence of a prompt injection attack.

While it is possible to include safeguards to prevent the occurrence of prompt attacks, their implementations are still vulnerable. This is highlighted in [15], where a simple prompt of ”ignore the context” caused the LLM agent to disregard any context retrieved despite the safeguards implemented. Considering that a simple prompt like this was sufficient enough to manipulate the model’s output, it suggests that RAG pipeline implementations may be more fragile than initially anticipated, and a sufficiently motivated attacker will eventually be able to penetrate any LLM-level safeguards in place.

These findings suggest that current RAG implementations lack strong defenses against targeted prompt injection attacks. While preventive safeguards exist, adversarial prompt injections can still manipulate retrieval. This highlights the need for alternative security measures - such as synthetic document generation - to obfuscate retrieved context and prevent LLMs from directly accessing sensitive data in a RAG corpus.

2.2 Medical Anonymization

In clinical settings, it is necessary to anonymize a dataset before releasing it to be shared. Based on that principle, we can further extend to healthcare applications that the public can access.

According to [16], the three main methods used in preserving medical privacy are Pseudonymization, De-identification, and Anonymization.

Each method can be summed up by the following:

Pseudonymization involves replacing attributes with pseudonyms.

De-identification involves the removal of PII from patient records.

Anonymization involves distorting the record such that it cannot be related to its original record.

Each of these methods are aimed at preserving clinical privacy, and are used in tandem.

As described in the study, the release of a clinical dataset involves de-identification of clinical data, after which the data undergoes either anonymization or pseudonymization before being released.

Further, the data can be manipulated in certain ways. Methods include addition of random noise to variables, converting ages to their relative dates, or categorising ages. This is a non-exhaustive list of methods and the rest can be seen in [16].

The document synthesis process operates based on some of these principles. In particular, we make use of methods that maintain some relation to the original record.

Chapter 3

Methodology

3.1 Description of Pipeline

Based on research into RAG vulnerabilities, there is a clear lack of security measures designed to preserve the privacy of a RAG corpus. This is especially important in fields like healthcare. As demonstrated in [13], private information can be easily extracted by determined attackers through simple prompt injections. Given that RAG relies on a set of documents as context and its vulnerabilities to RAG, we believe that generating a synthetic document separate from the corpus is sufficient to mitigate most issues.

3.2 System Design

As mentioned, the solution explored in this project consists of an agent-based document synthesis pipeline aimed at preventing raw LLM access to sensitive data.

For all intents and purposes, the pipeline operates in a similar fashion to typical RAG. Upon receiving a query, it fetches document from the RAG corpus then uses the retrieved documents as context in generating a response. However, we include an intermediary step between the information retrieval and inference steps.

Once the documents are retrieved, a secondary LLM extracts only the necessary information from the documents retrieved. For instance, we may retrieve a medical record consisting of different medical readings for a query about a patient's blood pressure readings. In this example, we aim for the LLM to extract only the blood pressure readings from this document.

With the information retrieved, we use an agent-based approach to modify the information. In order to further distance the information from the original record, we apply the following steps.

Firstly, we remove any PII that may appear in the information. We consider the following as PII: names, ages, contact number and address. The LLM will remove, or replace with pseudonyms, any appearance of PII.

Secondly, we manipulate the data that appears in the information to generalise the record. Numbers are rounded, and converted to ranges if multiple readings of the same type occur.

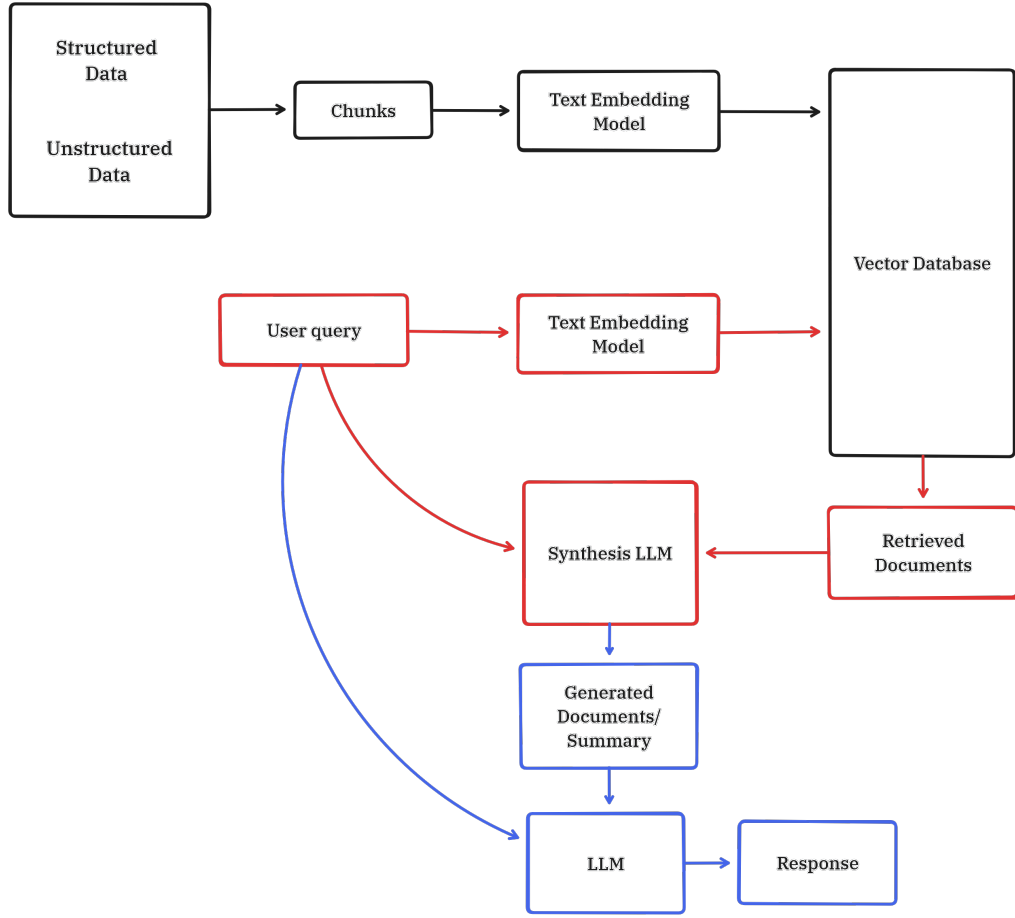


Figure 3.1: System Design

Finally, to ensure that the LLM treats the synthesized information as relevant context, we modify the original query based on the synthetic information. It should be able to generate the same output as a model operating solely on RAG.

Once it has gone through this step, we pass the synthesized query and information to the primary LLM to generate a response.

Refer to figure 3.1 for a visualization of the system design.

3.3 Building the RAG Corpus

RAG systems can make use of either structured or unstructured data, however in healthcare, data is usually structured. In order to mimic real healthcare settings, we determined it was necessary to make use of data that was designed for real-world settings. For our case, we will be making use of a synthetic Fast Healthcare Interoperability Resources (FHIR) dataset, generated and distributed by Synthea [17].

FHIR is a structured healthcare standard that defines how healthcare information can be shared between different systems regardless of how they are stored. Individual FHIR patient records are stored in what is known as resources and each resource type represents specific information. A Patient resource would include the patient's

```

{
  "fullUrl": "urn:uuid:9169c47c-a9d6-4e34-90fc-5f3b2a939984",
  "resource": {
    "resourceType": "Observation",
    "id": "9169c47c-a9d6-4e34-90fc-5f3b2a939984",
    "status": "final",
    "category": [
      {
        "coding": [
          {
            "system": "http://terminology.hl7.org/CodeSystem/observation-category",
            "code": "vital-signs",
            "display": "vital-signs"
          }
        ]
      }
    ],
    "code": {
      "coding": [
        {
          "system": "http://loinc.org",
          "code": "8302-2",
          "display": "Body Height"
        }
      ],
      "text": "Body Height"
    },
    "valueQuantity": {
      "value": 173.9018914060253,
      "unit": "cm",
      "system": "http://unitsofmeasure.org",
      "code": "cm"
    }
  }
}

```

→ Category is vital-signs. Code is Body Height. Value is 173.90 cm

Figure 3.2: FHIR to sentence

name, date of birth, address, etc. Each resource type is specific to its use case.

FHIR records can appear in different file formats, JSON, XML, or RDF. For simpler parsing and handling, we will be making use of JSON FHIR files to build our RAG corpus.

We make use of the open-source library *Llamaindex*[18] for abstractions when building the pipeline, as well as creating the database.

3.3.1 FHIR Preprocessing

First, we consider the type of data we wish to embed. JSON files are designed for programmatic use, meaning that they contain many identifying and delimiting tokens. If we were to convert the file in its entirety into its vector representation, it will result in detail being lost due to the repeated embedding of same key-value token pairs. Therefore, we first have to carry out flattening of the FHIR record.

Flattening the FHIR involves two things. First, we must determine what type of information we wish to extract. For this project, we are only working with information from the Patient resource, as well as the Observation, Procedure, Condition, Allergy and MedicationRequest resources. While initially the Encounter resource was used, we decided that it did not add any type of substantial information apart from the reason of the encounter as well as the location where it took place.

Secondly, we have to convert the selected information into basic sentences. This is done by recursively un-nesting the FHIR resource with information we specified. The reason we do this is to improve the embedding accuracy of the FHIR record. Firstly, we convert FHIR resources to basic sentences. This is to avoid repeatedly embedding the same key-value token pairs and wasting embedding tokens. Refer to figure 3.2 for an example.

Processing the FHIR record, we group the information extracted from the Observation and Procedure resources by date. Afterwards, we collate the conditions,

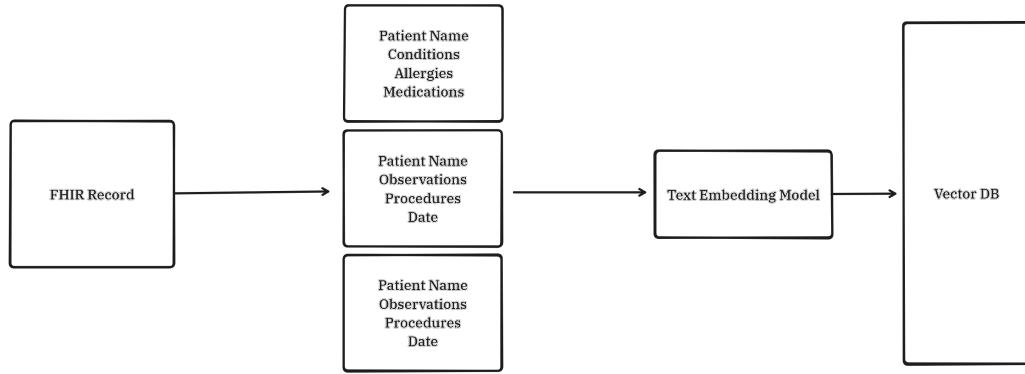


Figure 3.3: Embeddings to Database

allergies, as well as the medications that has been assigned to the patient previously.

Each of these documents are stored in separate files, marked by the patient’s name followed by the date of the encounter. These documents are then converted into vectors through the use of a text-embedding model, and stored within a Postgres database utilizing the *pgvector* extension. The embedding model used for generating the embeddings is *bge-base-en-v1.5*. The process is outlined in figure 3.3.

3.3.2 Retrieval

While not the scope of the project, it should be noted that during the creation of the database, Hierarchical Navigable Small World (HNSW) is used, which plays some influence in the retrieval results. We will not explore how the variations affect the retrieval results in this project.

With the RAG corpus built, we can now move onto retrieving documents associated with a query. The query goes through the embedding process and its resulting vector is compared to other document vectors in the database. The top k results are returned, with k being an adjustable variable. What determines the chunk’s relevance is its cosine similarity to the input query. Cosine similarity is defined as the following:

$$\text{Cosine Similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|}$$

and returns a score between 0.0 to 1.0. Here we can set a minimum cut-off for cosine similarity to adjust the relevance of returned information. Refer to figure 3.4 for an example of the returned chunks.

3.3.3 Synthetic Report Generation

LLMs differ in capabilities in accordance to their size. To determine if the chosen LLM (Mistral Nemo 12B) was sufficient for what I needed it to do, I tested its summarization and generation abilities. Firstly, I merged the previously processed FHIR record for a single patient into a combined document. This document was then passed to LLM along with a set of instructions. The specific prompt provided to the LLM is in the appendix, but to summarize:


```

Node ID: 19e561c3-0932-4edc-99a3-a293bc916dc7
Text: Name: Cammy883 Herzog843 Gender: female Born: 1965-02-07
MaritalStatus: S Deceased: False Conditions: Prediabetes recorded
2007/02/18 Diabetes recorded 2009/02/22 Hyperglycemia (disorder)
recorded 2011/02/27 Hypertriglyceridemia (disorder) recorded
2011/02/27 Metabolic syndrome X (disorder) recorded 2011/02/27 Anemia
(disorder) recorded 2011/02/...
Score: 0.605

Node ID: ee5e76e3-9fc6-4ecb-9860-7023bb9aaf8d
Text: Name: Mel236 Metz686 Gender: male Born: 1921-10-27
MaritalStatus: M Deceased: True Conditions: Diabetes recorded
1950/01/05 Anemia (disorder) recorded 1950/01/05 Hypertriglyceridemia
(disorder) recorded 1953/01/08 Metabolic syndrome X (disorder)
recorded 1956/01/12 Diabetic retinopathy associated with type II
diabetes mellitus (disorder) recorde...
Score: 0.590

```

Figure 3.4: Input Query: Which patients have diabetes?

Patient summary: Patient Information	Patient Summary
<ul style="list-style-type: none"> Full Name: Jimmy858 Ankunding277 Age: 90 years old (born on 1932-12-18) Gender: Male Marital Status: Married (M) Deceased: Yes 	<ul style="list-style-type: none"> Age: [Anonymized] years old ([Anonymized] life phase) Gender: Male Marital Status: Married
Clinical Observations	Clinical Observations
Vitals: <ul style="list-style-type: none"> Body Height: 165.79 cm Pain severity (0-10 verbal numeric rating): Range of values (min: 0.15, max: 3.81) Body Weight: Range of values (min: 76.02 kg, max: 83.72 kg) Body Mass Index (BMI): Range of values (min: 27.66 kg/m², max: 30.46 kg/m²) Blood Pressure: <ul style="list-style-type: none"> Diastolic Blood Pressure: Range of values (min: 72.92 mmHg, max: 87.52 mmHg) Systolic Blood Pressure: Range of values (min: 104.78 mmHg, max: 136.51 mmHg) 	<ul style="list-style-type: none"> Vitals: <ul style="list-style-type: none"> Height: Approximately 165 cm Pain severity (0-10): Range of values (min: 0.2, max: 4) Weight: Range of values (min: 76 kg, max: 84 kg) BMI: Range of values (min: 27.7 kg/m², max: 30.5 kg/m²) Blood Pressure: <ul style="list-style-type: none"> Diastolic: Range of values (min: 73 mmHg, max: 88 mmHg) Systolic: Range of values (min: 105 mmHg, max: 137 mmHg) Laboratory Results: <ul style="list-style-type: none"> Glucose: Range of values (min: 73 mg/dL, max: 98 mg/dL) Urea Nitrogen: Range of values (min: 7 mg/dL, max: 18 mg/dL) Creatinine: Range of values (min: 0.7 mg/dL, max: 0.9 mg/dL) Calcium: Range of values (min: 8.6 mg/dL, max: 10.2 mg/dL) Sodium: Range of values (min: 137 mmol/L, max: 144 mmol/L) Potassium: Range of values (min: 3.7 mmol/L, max: 5.2 mmol/L) Chloride: Range of values (min: 101 mmol/L, max: 108 mmol/L) Carbon Dioxide: Range of values (min: 21 mmol/L, max: 28 mmol/L)

Figure 3.5: Zero-Shot Generated Summary V.S. Synthesized Summary

- Break the summary into clear sections with headers
- Include exact numerical values
- Use precise dates
- Report conditions with specific terminology
- Summarize readings into a range spanning from min-max

The generated report summary was then passed to the LLM with instructions to anonymize information by rounding values as well as removing ages, dates, and names. This was done for three different types of prompting strategies, Zero-Shot, Chain-of-Thought, and Structured Output.

Refer to figure 3.5 for a side-by-side comparison for Zero-Shot generation. Full results for each are present in the appendix. Overall, the LLM was effective in following instructions as well as working with a large amount of context.

3.4 Moving Forward

With the RAG corpus built, and the abilities of the LLM confirmed, the next steps are as follows:

- Evaluate LLM's ability to extract relevant information from retrieved chunks
- Create a pipeline that connects the retrieval, synthesis and inference stages
- Compare LLM's responses when presented with the original and synthesized information
- Test the system through prompt attacks (information-query attacks)

References

- [1] Molly Bohannon. *Lawyer used Chatgpt in court-and cited fake cases. A judge is considering sanctions*. Feb. 2024. URL: <https://www.forbes.com/sites/mollybohannon/2023/06/08/lawyer-used-chatgpt-in-court-and-cited-fake-cases-a-judge-is-considering-sanctions/> (cit. on p. 2).
- [2] Patrick Lewis et al. *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks*. 2021. arXiv: 2005.11401 [cs.CL]. URL: <https://arxiv.org/abs/2005.11401> (cit. on p. 2).
- [3] Mingyu Jin et al. *Health-LLM: Personalized Retrieval-Augmented Disease Prediction System*. 2024. arXiv: 2402.00746 [cs.CL]. URL: <https://arxiv.org/abs/2402.00746> (cit. on p. 3).
- [4] NUHS. *Ai Healthcare in NUHS receives boost from supercomputer: NUHS+*. Aug. 2023. URL: <https://nuhsplus.edu.sg/article/ai-healthcare-in-nuhs-receives-boost-from-supercomputer> (cit. on p. 3).
- [5] Osmond Chia. *Ai masters Singlish in key breakthrough to serve healthcare and patients' needs*. Nov. 2024. URL: <https://www.straitstimes.com/singapore/ai-masters-singlish-in-key-breakthrough-to-serve-healthcare-and-patients-needs> (cit. on p. 3).
- [6] Liv McMahon. *Google Ai Search tells users to glue pizza and Eat Rocks*. May 2024. URL: <https://www.bbc.com/news/articles/cd11gzejgz4o> (cit. on p. 4).
- [7] Zhen Tan et al. *"Glue pizza and eat rocks" – Exploiting Vulnerabilities in Retrieval-Augmented Generative Models*. 2024. arXiv: 2406.19417 [cs.CR]. URL: <https://arxiv.org/abs/2406.19417> (cit. on p. 4).
- [8] Jiaqi Xue et al. *BadRAG: Identifying Vulnerabilities in Retrieval Augmented Generation of Large Language Models*. 2024. arXiv: 2406.00083 [cs.CR]. URL: <https://arxiv.org/abs/2406.00083> (cit. on p. 4).
- [9] Xun Xian et al. *On the Vulnerability of Applying Retrieval-Augmented Generation within Knowledge-Intensive Application Domains*. 2024. arXiv: 2409.17275 [cs.CR]. URL: <https://arxiv.org/abs/2409.17275> (cit. on p. 4).
- [10] Steve Alder. *Healthcare Data Breach Statistics*. Mar. 2025. URL: <https://www.hipaajournal.com/healthcare-data-breach-statistics/> (cit. on p. 4).

- [11] Cody Clop and Yannick Teglia. *Backdoored Retrievers for Prompt Injection Attacks on Retrieval Augmented Generation of Large Language Models*. 2024. arXiv: 2410.14479 [cs.CR]. URL: <https://arxiv.org/abs/2410.14479> (cit. on p. 5).
- [12] Yuefeng Peng et al. *Data Extraction Attacks in Retrieval-Augmented Generation via Backdoors*. 2024. arXiv: 2411.01705 [cs.CR]. URL: <https://arxiv.org/abs/2411.01705> (cit. on p. 5).
- [13] Shenglai Zeng et al. *The Good and The Bad: Exploring Privacy Issues in Retrieval-Augmented Generation (RAG)*. 2024. arXiv: 2402.16893 [cs.CR]. URL: <https://arxiv.org/abs/2402.16893> (cit. on pp. 5, 6, 8).
- [14] Zhenting Qi et al. *Follow My Instruction and Spill the Beans: Scalable Data Extraction from Retrieval-Augmented Generation Systems*. 2024. arXiv: 2402.17840 [cs.CL]. URL: <https://arxiv.org/abs/2402.17840> (cit. on p. 6).
- [15] Xuying Li et al. *Targeting the Core: A Simple and Effective Method to Attack RAG-based Agents via Direct LLM Manipulation*. 2024. arXiv: 2412.04415 [cs.AI]. URL: <https://arxiv.org/abs/2412.04415> (cit. on p. 6).
- [16] Aryelly Rodriguez et al. “Current recommendations/practices for anonymising data from clinical trials in order to make it available for sharing: A scoping review”. In: *Clinical Trials* 19.4 (June 2022), pp. 452–463. DOI: 10.1177/17407745221087469 (cit. on pp. 6, 7).
- [17] Corporation MITRE. *Synthea*. 2024. URL: <https://synthetichealth.github.io/synthea/> (visited on 11/29/2024) (cit. on p. 9).
- [18] Jerry Liu. *LlamaIndex*. Nov. 2022. DOI: 10.5281/zenodo.1234. URL: https://github.com/jerryjliu/llama_index (cit. on p. 10).
- [19] Xueying Du et al. *Vul-RAG: Enhancing LLM-based Vulnerability Detection via Knowledge-level RAG*. 2024. arXiv: 2406.11147 [cs.SE]. URL: <https://arxiv.org/abs/2406.11147>.
- [20] Yunfan Gao et al. *Retrieval-Augmented Generation for Large Language Models: A Survey*. 2024. arXiv: 2312.10997 [cs.CL]. URL: <https://arxiv.org/abs/2312.10997>.
- [21] John Chong Min Tan et al. *TaskGen: A Task-Based, Memory-Infused Agentic Framework using StrictJSON*. 2024. arXiv: 2407.15734 [cs.AI]. URL: <https://arxiv.org/abs/2407.15734>.
- [22] Meenatchi Sundaram Muthu Selva Annamalai, Andrea Gadotti, and Luc Rocher. *A Linear Reconstruction Approach for Attribute Inference Attacks against Synthetic Data*. 2024. arXiv: 2301.10053 [cs.LG]. URL: <https://arxiv.org/abs/2301.10053>.
- [23] Yi Dong et al. *Building Guardrails for Large Language Models*. 2024. arXiv: 2402.01822 [cs.CL]. URL: <https://arxiv.org/abs/2402.01822>.

.1 Appendix

Appendix A

Appendix

A.0.1 Report Generation Prompt

SYSTEM PROMPT:

You are a highly skilled medical assistant trained to generate

Include Key Identifiers: Use the patient's full name, age, and

Organizational Structure: Break the summary into clear sections

Patient Information

Clinical Observations

Relevant Medical History

Recent Procedures

Medications

Allergies

Data Reporting:

Include exact numerical values for vital signs, lab results

Use precise dates for events, diagnoses, or procedures when

Report medical conditions and history with specific terminology

Summarize readings into a range of values that include the

Provide the range of values that occur for each reading.

Tone and Clarity:

Use formal and professional language. Avoid abbreviations unless

Write in full sentences, ensuring clarity for medical professionals

[PATIENT HISTORY]

{ history }

[/PATIENT HISTORY]

A.0.2 Synthetic Report Generation Prompt

SYSTEM PROMPT:

You are an advanced language model tasked with summarizing patient

Anonymisation: The patient's name and any identifying information m

Structure: Organize the summary into the following sections:

- Patient Summary
- Clinical Observations
- Relevant Medical History
- Recent Procedures
- Medications
- Allergies

Formatting Rules:

- Use bullet points for observations and history to enhance readability.
- Replace values with rounded values for lab results and vital signs.
- Replace exact dates with the patient's life phase (e.g., "Young Adult").
- Avoid including exact locations or unnecessary identifiers.

Clinical Data Requirements:

- Summarize relevant vitals (e.g., blood pressure, BMI, glucose levels).
- Round all values in the summary.
- Ensure that all tests present in the original summary are present.

Tone and Clarity:

- Use formal and professional language. Avoid abbreviations unless necessary.
- Write in full sentences, ensuring clarity for medical professionals.

[REPORT SUMMARY]

{generated_summary}

[/REPORT SUMMARY]

A.0.3 Structured Output Prompt

SYSTEM PROMPT:

You are a highly capable summarization assistant. Your goal is

Patient Demographics: Include name, gender, date of birth, mari

Vital Observations: Summarize relevant metrics like height, wei

Medical Conditions: List significant diagnoses along with recor

Procedures: Highlight completed procedures.

Medications: Categorize medications into active and stopped, no

Allergies: Summarize reported allergies, if any.

Smoking Status: Include tobacco smoking status.

Input Example:

Name: John Doe

Gender: Male

Born: 1950-01-01

Marital Status: M

Deceased: True

... [additional details as seen in the record above].

Output Format:

Use the following structured format for your response:

Summary

****Demographics:****

- Name: [Name]
- Gender: [Gender]
- DOB: [Date of Birth]
- Marital Status: [Marital Status]
- Deceased: [Yes/No]

****Observations:****

- Height: [Value] cm
- Weight: [Value] kg
- BMI: [Value] kg/m²
- Blood Pressure: Systolic [Value] mmHg / Diastolic [Value] mmHg
- Glucose: [Value] mg/dL
- Hemoglobin A1c: [Value] %
- Cholesterol: Total [Value] mg/dL, LDL [Value] mg/dL, HDL [Value] mg/dL
- [Add other metrics as needed.]

****Conditions:****

- [Condition Name]: [Date Recorded]
- ...

****Procedures:****

- [Procedure Name(s)]

****Medications:****

- ****Active:****
- [Medication Name, Dosage]
- ****Stopped:****
- [Medication Name, Dosage]

****Allergies:****

- [List of allergies or "None"]

****Smoking Status:****

- [Never smoker/Current smoker/Formal smoker]

[PATIENT HISTORY]

{history}

[/PATIENT HISTORY]

A.0.4 Zero-Shot Result

Patient summary:

Patient Information

- Full Name: Jimmy858 Ankunding277
- Age: 90 years old (born on 1932-12-18)
- Gender: Male
- Marital Status: Married (M)
- Deceased: Yes

Clinical Observations

Vitals:

- Body Height: 165.79 cm
- Pain severity (0-10 verbal numeric rating): Range of values (min: 0, max: 10)
- Body Weight: Range of values (min: 76.02 kg, max: 83.72 kg)
- Body Mass Index (BMI): Range of values (min: 27.66 kg/m², max: 30.5 kg/m²)
- Blood Pressure:
 - Diastolic Blood Pressure: Range of values (min: 72.92 mmHg, max: 80.92 mmHg)
 - Systolic Blood Pressure: Range of values (min: 104.78 mmHg, max: 112.78 mmHg)

Laboratory Results:

- Glucose: Range of values (min: 73.14 mg/dL, max: 98.21 mg/dL)
- Urea Nitrogen: Range of values (min: 7.14 mg/dL, max: 17.79 mg/dL)
- Creatinine: Range of values (min: 0.68 mg/dL, max: 0.93 mg/dL)
- Calcium: Range of values (min: 8.61 mg/dL, max: 10.18 mg/dL)
- Sodium: Range of values (min: 136.51 mmol/L, max: 143.79 mmol/L)
- Potassium: Range of values (min: 3.72 mmol/L, max: 5.15 mmol/L)
- Chloride: Range of values (min: 101.06 mmol/L, max: 107.77 mmol/L)
- Carbon Dioxide: Range of values (min: 20.95 mmol/L, max: 28.18 mmol/L)
- Total Cholesterol: Range of values (min: 163.82 mg/dL, max: 185.82 mg/dL)
- Triglycerides: Range of values (min: 121.80 mg/dL, max: 141.14 mg/dL)
- Low Density Lipoprotein Cholesterol (LDL): Range of values (min: 101.06 mg/dL, max: 107.77 mg/dL)
- High Density Lipoprotein Cholesterol (HDL): Range of values (min: 3.72 mmol/L, max: 5.15 mmol/L)

Hematology:

- Hemoglobin A1c/Hemoglobin.total in Blood: Range of values (min: 5.7, max: 6.5)
- Leukocytes [# /volume] in Blood by Automated count: Range of values (min: 4.0, max: 11.0)
- Erythrocytes [# /volume] in Blood by Automated count: Range of values (min: 4.0, max: 11.0)
- Hemoglobin [Mass/volume] in Blood: Range of values (min: 14.14 g/dL, max: 16.14 g/dL)
- Hematocrit [Volume Fraction] of Blood by Automated count: Range of values (min: 0.4, max: 0.5)
- MCV [Entitic volume] by Automated count: Range of values (min: 80 fL, max: 100 fL)
- MCH [Entitic mass] by Automated count: Range of values (min: 27.4 pg, max: 29.4 pg)
- MCHC [Mass/volume] by Automated count: Range of values (min: 33.7 g/dL, max: 35.7 g/dL)

- Erythrocyte distribution width [Entitic volume] by Automated count
- Platelets [# /volume] in Blood by Automated count: Range of values
- Platelet distribution width [Entitic volume] in Blood by Automated count
- Platelet mean volume [Entitic volume] in Blood by Automated count

****Relevant Medical History****

- Chronic sinusitis (disorder) recorded on 1959/10/30
- Prediabetes recorded on 1972/12/24
- Anemia (disorder) recorded on 1977/01/02
- Stroke recorded on 1995/04/23
- Coronary Heart Disease recorded on 2004/04/18
- Viral sinusitis (disorder) recorded on 2005/05/04, 2006/11/18, and 2007/05/04
- Laceration of forearm recorded on 2005/08/25
- Myocardial Infarction recorded on 2011/10/09

****Recent Procedures****

- Completed Medication Reconciliation (procedure) on multiple dates
- Completed Echocardiography (procedure)
- Completed Percutaneous coronary intervention
- Completed Coronary artery bypass grafting
- Completed Colonoscopy on multiple dates
- Completed Suture open wound

****Medications****

- Stopped Clopidogrel 75 MG Oral Tablet
- Active Simvastatin 20 MG Oral Tablet
- Active Amlodipine 5 MG Oral Tablet
- Stopped Nitroglycerin 0.4 MG/ACTUAT Mucosal Spray
- Stopped Acetaminophen 325 MG Oral Tablet
- Stopped Atorvastatin 80 MG Oral Tablet
- Stopped Captopril 25 MG Oral Tablet

****Allergies****

No allergies reported.

Synth report:

****[ANONYMIZED MEDICAL REPORT]****

****Patient Summary****

- Age: [Anonymized] years old ([Anonymized] life phase)
- Gender: Male
- Marital Status: Married

****Clinical Observations****

- **Vitals:**

- Height: Approximately 165 cm
- Pain severity (0–10): Range of values (min: 0.2, max: 4)
- Weight: Range of values (min: 76 kg, max: 84 kg)
- BMI: Range of values (min: 27.7 kg/m², max: 30.5 kg/m²)
- Blood Pressure:
 - Diastolic: Range of values (min: 73 mmHg, max: 88 mmHg)
 - Systolic: Range of values (min: 105 mmHg, max: 137 mmHg)

- **Laboratory Results:**

- Glucose: Range of values (min: 73 mg/dL, max: 98 mg/dL)
- Urea Nitrogen: Range of values (min: 7 mg/dL, max: 18 mg/dL)
- Creatinine: Range of values (min: 0.7 mg/dL, max: 0.9 mg/dL)
- Calcium: Range of values (min: 8.6 mg/dL, max: 10.2 mg/dL)
- Sodium: Range of values (min: 137 mmol/L, max: 144 mmol/L)
- Potassium: Range of values (min: 3.7 mmol/L, max: 5.2 mmol/L)
- Chloride: Range of values (min: 101 mmol/L, max: 108 mmol/L)
- Carbon Dioxide: Range of values (min: 21 mmol/L, max: 28 mmol/L)
- Total Cholesterol: Range of values (min: 164 mg/dL, max: 186 mg/dL)
- Triglycerides: Range of values (min: 122 mg/dL, max: 141 mg/dL)
- LDL Cholesterol: Range of values (min: 71 mg/dL, max: 83 mg/dL)
- HDL Cholesterol: Range of values (min: 60 mg/dL, max: 76 mg/dL)

- **Hematology:**

- Hemoglobin Alc/Hemoglobin.total in Blood: Range of values (min: 4.7, max: 8.1) 10³/uL
- Leukocytes: Range of values (min: 4.3, max: 5.4) 10⁶/uL
- Erythrocytes: Range of values (min: 14.2 g/dL, max: 14.2 g/dL)
- Hemoglobin: Range of values (min: 37%, max: 47%)
- MCV: Range of values (min: 80 fL, max: 84 fL)
- MCH: Range of values (min: 27 pg, max: 32 pg)
- MCHC: Range of values (min: 33.8 g/dL, max: 35.9 g/dL)
- Erythrocyte distribution width: Range of values (min: 42 fL, max: 11.4 fL)
- Platelets: Range of values (min: 239, max: 308) 10³/uL
- Platelet distribution width: Range of values (min: 371 fL, max: 11.4 fL)
- Platelet mean volume: Range of values (min: 10.8 fL, max: 11.4 fL)

****Relevant Medical History****

- Chronic sinusitis recorded in [Anonymized] life phase
- Prediabetes recorded in young adulthood
- Anemia recorded in early adulthood
- Stroke recorded in middle age
- Coronary Heart Disease recorded in late adulthood
- Viral sinusitis recorded on multiple occasions in adulthood

- Laceration of forearm recorded in adulthood
- Myocardial Infarction recorded in late adulthood

****Recent Procedures****

- Completed Medication Reconciliation on multiple occasions
- Completed Echocardiography
- Completed Percutaneous coronary intervention
- Completed Coronary artery bypass grafting
- Completed Colonoscopy on multiple occasions
- Completed Suture open wound

****Medications****

- Stopped Clopidogrel 75 MG Oral Tablet
- Active Simvastatin 20 MG Oral Tablet
- Active Amlodipine 5 MG Oral Tablet
- Stopped Nitroglycerin 0.4 MG/ACTUAT Mucosal Spray
- Stopped Acetaminophen 325 MG Oral Tablet
- Stopped Atorvastatin 80 MG Oral Tablet
- Stopped Captopril 25 MG Oral Tablet

****Allergies****

No allergies reported.

A.0.5 Chain-of-Thought Result

Patient summary:

Patient Information

- **Name:** Jimmy Ankunding
- **Date of Birth:** December 18, 1932
- **Gender:** Male
- **Marital Status:** Married (M)
- **Deceased:** Yes

Clinical Observations

Parameter	Value	Range
Body Height	165.79 cm	
Pain Severity (0-10)	0.15	3.81
Body Weight	76.02	83.72 kg
Body Mass Index	27.66	30.46 kg/m ²
Blood Pressure (Diastolic/Systolic)	72.92	136.51 mmHg / 104.79
Glucose	73.14	98.21 mg/dL
Urea Nitrogen	7.14	17.79 mg/dL
Creatinine	0.68	0.93 mg/dL
Calcium	8.61	10.18 mg/dL
Sodium	136.51	143.79 mmol/L
Potassium	3.72	5.15 mmol/L
Chloride	101.06	107.77 mmol/L
Carbon Dioxide	20.95	28.18 mmol/L
Total Cholesterol	163.82	185.82 mg/dL
Triglycerides	121.80	141.14 mg/dL
Low Density Lipoprotein Cholesterol	70.57	83.27 mg/dL
High Density Lipoprotein Cholesterol	65.02	75.98 mg/dL
Leukocytes (#/volume)	4.12	8.12 x10 ³ /uL
Erythrocytes (#/volume)	4.34	5.43 x10 ⁶ /uL
Hemoglobin (Mass/Volume)	14.14	14.17 g/dL
Hematocrit (Volume Fraction)	36.89	46.69 %
MCV (Entitic Volume)	80.00	84.38 fL
MCH (Entitic Mass)	27.48	32.28 pg
MCHC (Mass/Volume)	33.76	35.89 g/dL
Erythrocyte Distribution Width	42.09	45.90 fL
Platelets (#/volume)	238.95	308.03 x10 ³ /uL
Platelet Distribution Width	370.81	429.00 fL
Platelet Mean Volume	10.79	11.36 fL
Hemoglobin A1c/Hemoglobin.total in Blood	5.84	6.36 %

Relevant Medical History

- Chronic sinusitis (1959)
- Prediabetes (1972)
- Anemia (1977)
- Stroke (1995)
- Coronary Heart Disease (2004)
- Viral sinusitis (2005, 2006, 2010)
- Myocardial Infarction (2011)

****Recent Procedures****

- Medication Reconciliation (multiple dates)
- Echocardiography
- Percutaneous coronary intervention
- Coronary artery bypass grafting
- Colonoscopy (multiple procedures)
- Suture open wound

****Medications****

- Simvastatin 20 MG Oral Tablet (active)
- Amlodipine 5 MG Oral Tablet (active)
- Clopidogrel 75 MG Oral Tablet (stopped)
- Nitroglycerin 0.4 MG/ACTUAT Mucosal Spray (stopped)
- Acetaminophen 325 MG Oral Tablet (stopped)
- Atorvastatin 80 MG Oral Tablet (stopped)
- Captopril 25 MG Oral Tablet (stopped)

****Allergies****

No allergies listed.

Synth report:

****[Anonymized] Patient Summary****

****Patient Information****

- ****Age:**** Late Adulthood
- ****Gender:**** Male
- ****Marital Status:**** Married ([Anonymized])

****Clinical Observations****

- Height: Approximately 165 cm
- Weight: Fluctuated between approximately 76 kg and 84 kg
- BMI: Ranged from approximately 28 to 30 kg/m²
- Blood Pressure: Diastolic/Systolic ranged from approximately 73/100 mmHg
- Glucose Levels: Fluctuated between approximately 73 mg/dL and 98 mg/dL

- Urea Nitrogen: Ranged from approximately 7 mg/dL to 18 mg/dL
- Creatinine: Approximately 0.7 to 0.9 mg/dL
- Calcium: Approximately 8.6 to 10.2 mg/dL
- Sodium: Approximately 137 to 144 mmol/L
- Potassium: Ranged from approximately 3.7 to 5.2 mmol/L
- Chloride: Approximately 102 to 108 mmol/L
- Carbon Dioxide: Approximately 21 to 28 mmol/L
- Total Cholesterol: Approximately 164 to 186 mg/dL
- Triglycerides: Approximately 122 to 141 mg/dL
- LDL Cholesterol: Approximately 71 to 83 mg/dL
- HDL Cholesterol: Approximately 65 to 76 mg/dL
- Leukocytes: Approximately 4 to 8 x10³/uL
- Erythrocytes: Approximately 4 to 5.4 x10⁶/uL
- Hemoglobin: Approximately 14 g/dL
- Hematocrit: Approximately 37% to 47%
- MCV: Approximately 80 to 84 fL
- MCH: Approximately 27 to 32 pg
- MCHC: Approximately 34 g/dL
- Erythrocyte Distribution Width: Approximately 42 to 46 fL
- Platelets: Approximately 240 to 310 x10³/uL
- Platelet Distribution Width: Approximately 370 to 430 fL
- Platelet Mean Volume: Approximately 10.8 to 11.4 fL

****Relevant Medical History****

- Chronic sinusitis since Young Adulthood ([1959])
- Prediabetes diagnosed in Early Adulthood ([1972])
- Anemia diagnosed in Early Adulthood ([1977])
- Stroke in Middle Age ([1995])
- Coronary Heart Disease diagnosed in Late Middle Age ([2004])
- Multiple episodes of viral sinusitis (2005, 2006, 2010)
- Myocardial Infarction in Late Adulthood ([2011])

****Recent Procedures****

- Medication Reconciliation performed on multiple occasions
- Echocardiography completed
- Percutaneous coronary intervention conducted
- Coronary artery bypass grafting procedure undertaken
- Colonoscopy procedures performed on multiple occasions
- Suture open wound repair

****Medications****

- Simvastatin 20 MG Oral Tablet (currently active)
- Amlodipine 5 MG Oral Tablet (currently active)
- Clopidogrel 75 MG Oral Tablet (previously active , now stopped)
- Nitroglycerin 0.4 MG/ACTUAT Mucosal Spray (previously active , now
- Acetaminophen 325 MG Oral Tablet (previous use)

- Atorvastatin 80 MG Oral Tablet (previous use)
- Captopril 25 MG Oral Tablet (previous use)

****Allergies****

No allergies listed.

A.0.6 Structured Output Result

Patient summary:

Summary

****Demographics:****

- Name: Jimmy Ankunding
- Gender: Male
- DOB: 1932-12-18
- Marital Status: Married (M)
- Deceased: Yes

****Observations:****

- Height: 165.79 cm
- Weight: 83.72 kg
- BMI: 30.46 kg/m²
- Blood Pressure: Systolic 134.27 mmHg / Diastolic 82.43 mmHg
- Glucose: 73.51 mg/dL
- Hemoglobin A1c: 6.34 %
- Cholesterol: Total 175.49 mg/dL, LDL 91.30 mg/dL, HDL 59.83 m

****Conditions:****

- Chronic sinusitis: 1959/10/30
- Prediabetes: 1972/12/24
- Anemia: 1977/01/02
- Stroke: 1995/04/23
- Coronary Heart Disease: 2004/04/18
- Viral sinusitis (multiple occurrences)
- Laceration of forearm: 2005/08/25
- Myocardial Infarction: 2011/10/09

****Procedures:****

- Medication Reconciliation
- Echocardiography
- Percutaneous coronary intervention
- Coronary artery bypass grafting
- Colonoscopy (multiple occurrences)
- Suture open wound

****Medications:****

- ****Active:****
 - Simvastatin 20 MG Oral Tablet
 - Amlodipine 5 MG Oral Tablet
- ****Stopped:****
 - Clopidogrel 75 MG Oral Tablet
 - Nitroglycerin 0.4 MG/ACTUAT Mucosal Spray
 - Acetaminophen 325 MG Oral Tablet

- Atorvastatin 80 MG Oral Tablet
- Captopril 25 MG Oral Tablet

****Allergies:****

- None reported

****Smoking Status:****

- Never smoker

Synth report:

****[ANONYMIZED MEDICAL REPORT]****

****Patient Summary****

- ****Name:**** [Anonymized]
- ****Gender:**** Male
- ****Age:**** Deceased in Late Adulthood
- ****Marital Status:**** Married

****Clinical Observations****

- Height: Approximately 166 cm
- Weight: Around 84 kg
- Body Mass Index (BMI): About 30.5 kg/m²
- Blood Pressure (BP): Systolic ~134 mmHg, Diastolic ~82 mmHg
- Glucose Level: Around 74 mg/dL
- Hemoglobin A1c (HbA1c): Approximately 6.3%
- Cholesterol Levels:
 - Total: About 175 mg/dL
 - LDL: Around 91 mg/dL
 - HDL: Approximately 60 mg/dL

****Relevant Medical History****

- Chronic sinusitis, diagnosed in Young Adulthood (1959)
- Prediabetes, diagnosed in Early Middle Age (1972)
- Anemia, diagnosed in Early Middle Age (1977)
- Stroke, occurred in Mid-Adulthood (1995)
- Coronary Heart Disease, diagnosed in Late Middle Age (2004)
- Viral sinusitis: Multiple occurrences
- Laceration of forearm, treated in Late Middle Age (2005)
- Myocardial Infarction, occurred in Late Middle Age (2011)

****Recent Procedures****

- Medication Reconciliation
- Echocardiography
- Percutaneous coronary intervention
- Coronary artery bypass grafting
- Colonoscopy: Multiple occurrences

- Suture open wound

****Medications****

- ****Active:****
 - Simvastatin 20 MG Oral Tablet
 - Amlodipine 5 MG Oral Tablet
- ****Stopped:****
 - Clopidogrel 75 MG Oral Tablet
 - Nitroglycerin 0.4 MG/ACTUAT Mucosal Spray
 - Acetaminophen 325 MG Oral Tablet
 - Atorvastatin 80 MG Oral Tablet
 - Captopril 25 MG Oral Tablet

****Allergies****

- None reported

****Smoking Status****

- Never smoker