

Privacy Enhanced RAG for LLMs 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
2 Literature Review	2
2.1 Background	2
2.1.1 Retrieval-augmented Generation (RAG)	2
2.1.2 Exploitation of RAG Systems	2
2.1.3 Large Language Model (LLM) Safeguards	2
3 Methodology	4
3.1 Proposal	4
3.2 Implementation	4
3.2.1 Medical Anonymisation	4
3.2.2 System Design	5
3.2.3 Building the RAG Corpus	5
3.2.4 Retrieval	7
3.2.5 Synthetic Report Generation	7
3.3 Moving Forward	8
.1 Appendix	11
A Appendix	12
A.0.1 Report Generation Prompt	12
A.0.2 Synthetic Report Generation Prompt	13
A.0.3 Structured Output Prompt	14
A.0.4 Zero-Shot Result	16
A.0.5 Chain-of-Thought Result	20
A.0.6 Structured Output Result	23

List of Figures

2.1	Example of a Conventional RAG system	3
3.1	System Design	5
3.2	FHIR to sentence	6
3.3	Embeddings to Database	6
3.4	Input Query: Which patients have diabetes?	7
3.5	Zero-Shot Generated Summary V.S. Synthesized Summary	8

List of Tables

Chapter 1

Introduction

LLMs are transforming various industries. They are able to perform tasks such as automated handling of workflows with Agentic frameworks, Natural Language Processing (NLP) tasks such as information extraction, and even rudimentary reasoning in some models. In fields like Healthcare, they can automate clinical note generation and summarization, assist in diagnosis, and provide personalized patient care.

However, they all suffer the same traditional issue, hallucinations, where they generate seemingly coherent but incorrect information. To address this, RAG was developed as a method to provide context to LLMs by incorporating an external knowledge base, allowing them to generate more accurate, domain-specific responses. This technique has applications in areas where hallucinations can cause severe harm, such as medicine, legal analysis, and cybersecurity.

While RAG enhances LLM capabilities, it introduces new security risks. Attackers can exploit RAG systems to extract proprietary or sensitive data through prompt injection attacks. This is a critical privacy concern, especially in healthcare where patient confidentiality is important.

In this project, we seek to test and develop an Agent-based synthetic document generation framework to mitigate these risks. By separating the RAG database from the externally facing LLM, we seek to enhance security while preserving the contextual accuracy of responses.

Chapter 2

Literature Review

2.1 Background

2.1.1 Retrieval-augmented Generation (RAG)

Retrieval-augmented generation (RAG), first introduced by [1], is a method used to enhance the text generation abilities of LLMs.

It is seeing increased usage in multiple LLM-applications, such as medical chat-bots [2], as well as systems for detecting vulnerabilities in code [3].

The widespread adoption of RAG is due to its flexibility; A pre-trained model can be made to function in different domains simply by making adjustments to its RAG components, bypassing the need for extensive re-training.

Furthermore, numerous advancements are being made in enhancing the capabilities of RAGs, such as advanced techniques that improve retrieval qualities, or modular architectures that allow for specific fine-tuning of RAG components [4].

2.1.2 Exploitation of RAG Systems

Studies ([5]; [6]; [7]) have shown that RAG systems are susceptible to well-crafted prompt attacks during the retrieval stage.

By using targeted as well as untargeted attacks [6], a malicious attacker is able to retrieve personally identifiable information (PII), such as phone numbers and addresses, from a RAG's corpus.

[5] and [7] showcase the ability to affect an LLMs output by inserting specially crafted adversarial passages into its RAG corpus.

This typically affects LLMs that make use of real-time context databases, such as a search engine, which allows an attacker to insert malicious documents into the context database.

2.1.3 Large Language Model (LLM) Safeguards

The widespread usage of LLMs necessitates the development of safeguards to prevent ethical misuse and abuse. These safeguards are often times complex, varying based on application requirements.

[8] discusses the different components involved in implementing guard rails for LLMs and touches on some of the currently deployed solutions available. In general,

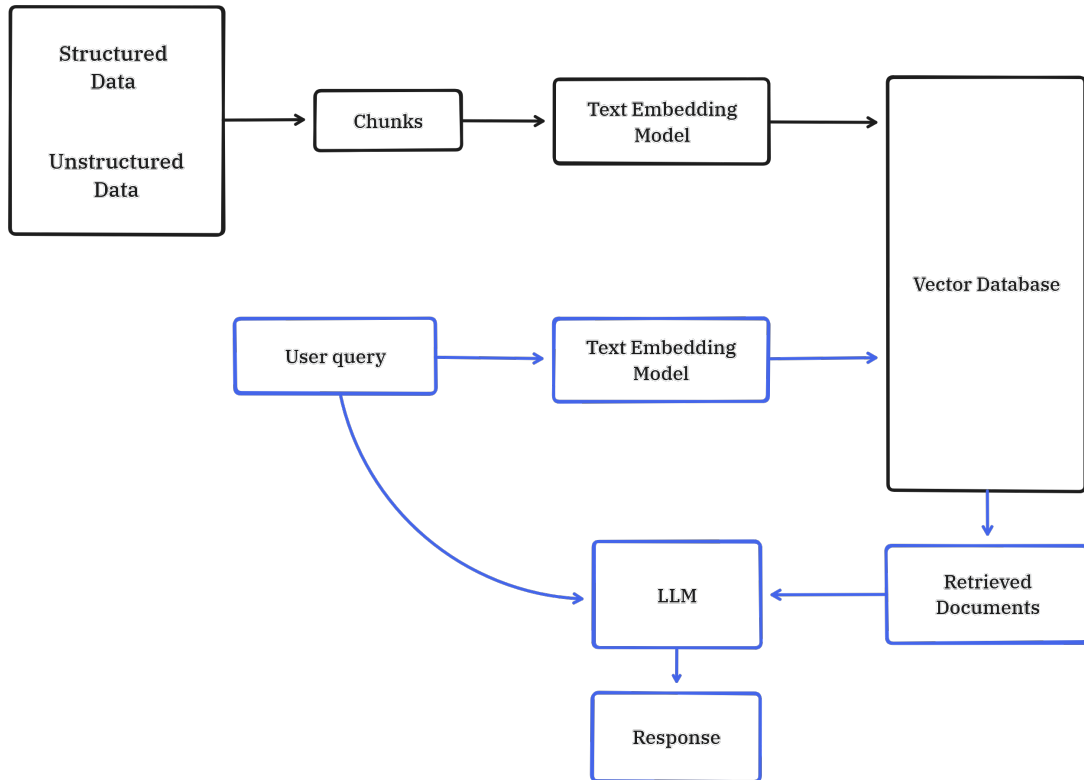


Figure 2.1: Example of a Conventional RAG system

safeguards are designed to prevent the LLMs from generating unintended output. This unintended output can be generated in numerous ways, most notably through “hallucinations” as well as targeted prompt attacks known as “jailbreaking”.

Chapter 3

Methodology

3.1 Proposal

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 [6], 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 that this is vulnerable to attacks, I suggest generating synthetic context from this set of documents. This set of documents could refer to a patient's records, containing information such as their blood pressure, etc. A separate LLM will analyze the set of records retrieved by the user's query, comparing the two and extracting relevant information. The LLM will then generate a synthetic record containing all the information needed to answer the query, whilst removing PII at the same time. This separates the context given to the LLM from the RAG corpus, whilst still ensuring that its response remains contextually relevant.

3.2 Implementation

To design an adequate system that can preserve a patient's privacy, we must first understand the different types of methods associated with this task.

3.2.1 Medical Anonymisation

According to [9], the three main methods in preserving medical privacy are Pseudonymisation, De-identification and Anonymisation. Pseudonymisation refers to the replacement of attributes with pseudonyms. De-identification refers to the removal of PII from patient records. Anonymisation refers to the distortion of data such that any record lacks individuality.

The method we will be focusing on is the process of anonymisation. There are two different ways this is achieved. Recalculation, which involves turning absolute dates into relative representations (such as turning a patient's date of birth into their equivalent age representation), and Perturbation, which involves directly modifying a patient's medical records away from their original form.

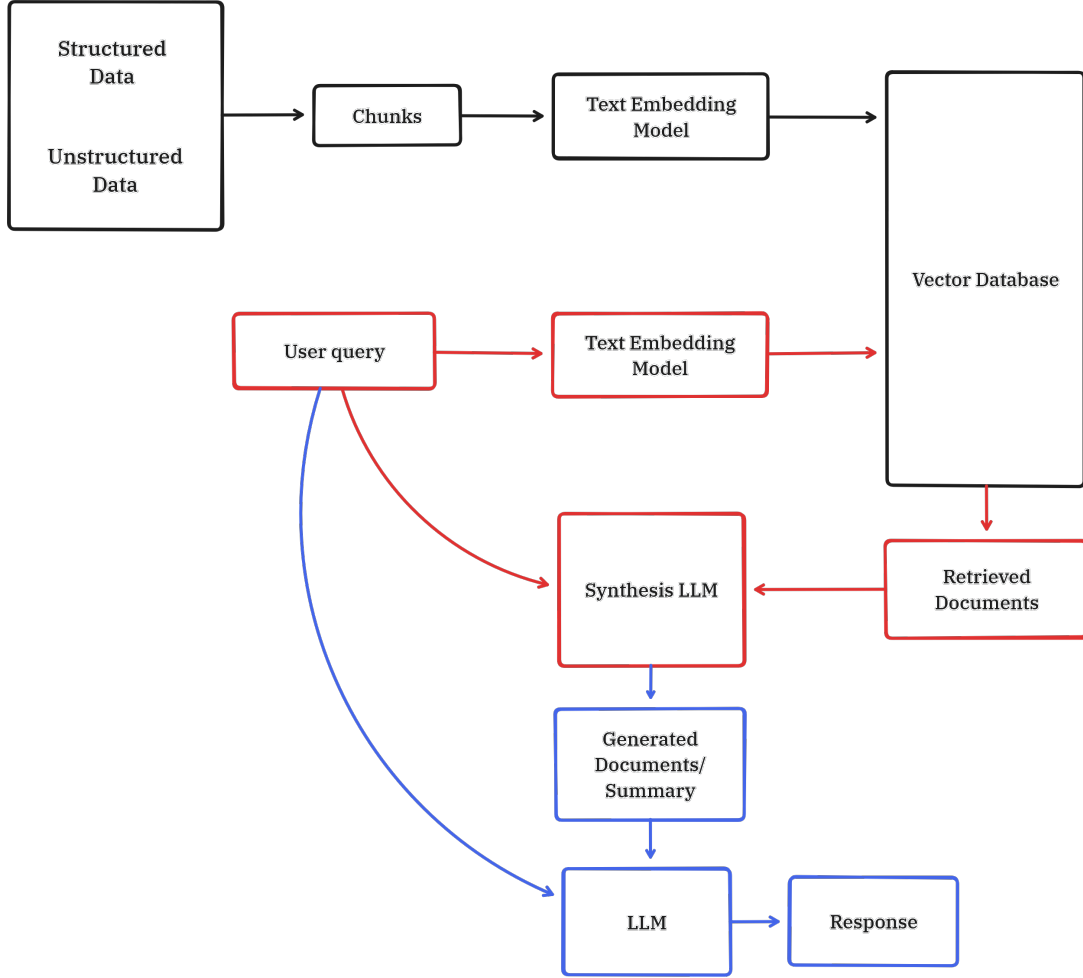


Figure 3.1: System Design

3.2.2 System Design

The proposed system design is outlined in figure 3.1. The system is similar to figure 2.1, with the inclusion of an intermediary step that passes both the user query as well as the retrieved context documents to a secondary LLM for synthesis. The secondary LLM is responsible for both information extraction as well as synthetic document generation. To mimic privacy requirements, a locally run LLM is used here to ensure that the RAG corpus remains air-gapped. The use of an adequately sized LLM is critical here, and in our case we will be using Mistral’s Nemo-12B. The synthesized document is then returned to the front-facing LLM along with the user’s query to generate an appropriate response.

3.2.3 Building the RAG Corpus

RAG systems can make use of either structured or unstructured data. For our case, we will be making use of a synthetic FHIR dataset, generated by Synthea [10]. 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. A resource can take on

```

{
  "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

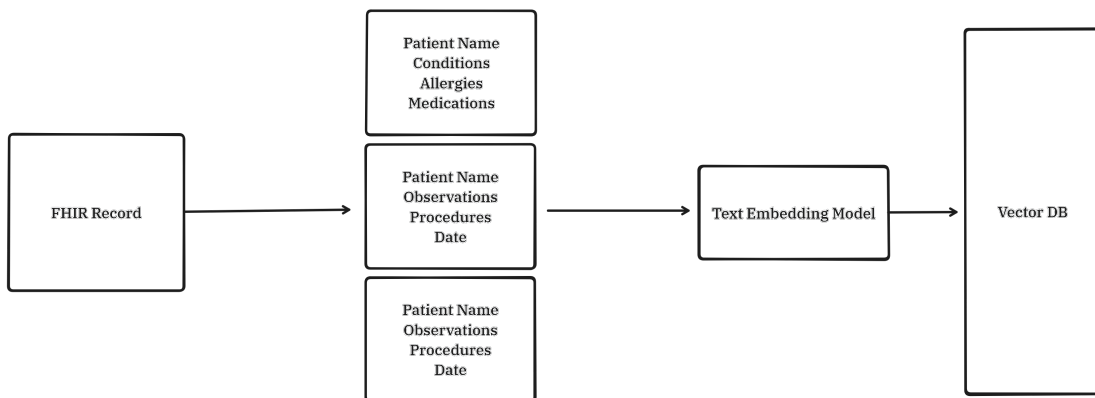


Figure 3.3: Embeddings to Database

different types, and each type contains information necessary for its specific use case. FHIR records can appear in the form of JSON, XML or RDF. Here we will be making use of JSON FHIR files to create our RAG corpus.

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. This can be further truncated. For example we can simply embed only the name of the reading and its associated value.

We perform this operation on the patient's record, extracting Observations and Procedures, separating them by date, and extracting a patient's diagnosed conditions, medications, as well as allergies, and storing them in a separate document. These documents are then converted into vectors through the use of a text-embedding model. For this, we will be using the *bge-base-en* embedding model. The embeddings are then stored in a Postgres database utilizing the *pgvector* extension.

```

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?

3.2.4 Retrieval

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.

3.2.5 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:

- 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.

<p>Patient summary:</p> <p>Patient Information</p> <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 <p>Clinical Observations</p> <p><i>Vitals:</i></p> <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) 	<p>Patient Summary</p> <ul style="list-style-type: none"> • Age: [Anonymized] years old ([Anonymized] life phase) • Gender: Male • Marital Status: Married <p>Clinical Observations</p> <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

3.3 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] 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).
- [2] 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. 2).
- [3] 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> (cit. on p. 2).
- [4] 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> (cit. on p. 2).
- [5] 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. 2).
- [6] 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. 2, 4).
- [7] 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. 2).
- [8] Yi Dong et al. *Building Guardrails for Large Language Models*. 2024. arXiv: 2402.01822 [cs.CL]. URL: <https://arxiv.org/abs/2402.01822> (cit. on p. 2).
- [9] 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 p. 4).
- [10] Corporation MITRE. *Synthea*. 2024. URL: <https://synthetichealth.github.io/synthea/> (visited on 11/29/2024) (cit. on p. 5).
- [11] 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>.

- [12] 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>.

.1 Appendix

Appendix A

Appendix

A.0.1 Report Generation Prompt

SYSTEM PROMPT:

You are a highly skilled medical assistant trained to generate detailed medical summaries.

Include Key Identifiers: Use the patient's full name, age, and any other relevant identifiers.

Organizational Structure: Break the summary into clear sections with the following structure:

- Patient Information

- Clinical Observations

- Relevant Medical History

- Recent Procedures

- Medications

- Allergies

Data Reporting:

- Include exact numerical values for vital signs, lab results, and other quantitative data.

- Use precise dates for events, diagnoses, or procedures where applicable.

- Report medical conditions and history with specific terminology.

- Summarize readings into a range of values that include the maximum and minimum values.

- Provide the range of values that occur for each reading.

Tone and Clarity:

- Use formal and professional language. Avoid abbreviations unless necessary.

- 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 data in

Anonymisation: The patient's name and any identifying information must be

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. Use
- Replace exact dates with the patient's life phase (e.g., "Young Adulth
- 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 in t

Tone and Clarity:

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

```
[REPORT SUMMARY]
    {generated_summary}
[/REPORT SUMMARY]
```

A.0.3 Structured Output Prompt

SYSTEM PROMPT:

You are a highly capable summarization assistant. Your goal is to extract

Patient Demographics: Include name, gender, date of birth, marital s

Vital Observations: Summarize relevant metrics like height, weight,

Medical Conditions: List significant diagnoses along with recorded c

Procedures: Highlight completed procedures.

Medications: Categorize medications into active and stopped, noting

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] m

– [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/Former 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: 0.15, 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: 30.46 kg/m²)
- Blood Pressure:
 - Diastolic Blood Pressure: Range of values (min: 72.92 mmHg, max: 83.72 mmHg)
 - Systolic Blood Pressure: Range of values (min: 104.78 mmHg, max: 115.78 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: 70.57 mg/dL, max: 98.21 mg/dL)
- High Density Lipoprotein Cholesterol (HDL): Range of values (min: 59.82 mg/dL, max: 73.14 mg/dL)

Hematology:

- Hemoglobin A1c/Hemoglobin.total in Blood: Range of values (min: 5.84%, 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: 15.79 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: 80.00 fL, max: 100.00 fL)
- MCH [Entitic mass] by Automated count: Range of values (min: 27.48 pg, max: 31.57 pg)
- MCHC [Mass/volume] by Automated count: Range of values (min: 33.76 g/dL, max: 37.79 g/dL)
- Erythrocyte distribution width [Entitic volume] by Automated count: Range of values (min: 11.5%, max: 14.5%)
- Platelets [# /volume] in Blood by Automated count: Range of values (min: 100,000, max: 400,000)

- Platelet distribution width [Entitic volume] in Blood by Automated count
- Platelet mean volume [Entitic volume] in Blood by Automated count: Range

****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 2010/
- 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 A1c/Hemoglobin.total in Blood: Range of values (min: 5.8%, max: 6.5%)
 - Leukocytes: Range of values (min: 4.7, max: 8.1) 10³/uL
 - Erythrocytes: Range of values (min: 4.3, max: 5.4) 10⁶/uL
 - Hemoglobin: Range of values (min: 14.2 g/dL, max: 14.2 g/dL)
 - Hematocrit: 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: 46 fL)
 - Platelets: Range of values (min: 239, max: 308) 10³/uL
 - Platelet distribution width: Range of values (min: 371 fL, max: 430 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/m2
Blood Pressure (Diastolic/Systolic)	72.92 - 136.51 mmHg / 104.78 - 136.51 mmHg
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 73/105 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 stop
- 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 mg/dL

****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