

Article

Addressing Data Scarcity in the Medical Domain: A GPT-Based Approach for Synthetic Data Generation and Feature Extraction

Fahim Sufi 

School of Public Health and Preventive Medicine, Monash University, Melbourne, VIC 3004, Australia;
research@fahimsufi.com

Abstract: This research confronts the persistent challenge of data scarcity in medical machine learning by introducing a pioneering methodology that harnesses the capabilities of Generative Pre-trained Transformers (GPT). In response to the limitations posed by a dearth of labeled medical data, our approach involves the synthetic generation of comprehensive patient discharge messages, setting a new standard in the field with GPT autonomously generating 20 fields. Through a meticulous review of the existing literature, we systematically explore GPT's aptitude for synthetic data generation and feature extraction, providing a robust foundation for subsequent phases of the research. The empirical demonstration showcases the transformative potential of our proposed solution, presenting over 70 patient discharge messages with synthetically generated fields, including severity and chances of hospital re-admission with justification. Moreover, the data had been deployed in a mobile solution where regression algorithms autonomously identified the correlated factors for ascertaining the severity of patients' conditions. This study not only establishes a novel and comprehensive methodology but also contributes significantly to medical machine learning, presenting the most extensive patient discharge summaries reported in the literature. The results underscore the efficacy of GPT in overcoming data scarcity challenges and pave the way for future research to refine and expand the application of GPT in diverse medical contexts.

Keywords: GPT; large language models; prompt engineering; synthetic data generation; medical data labeling; feature extraction



Citation: Sufi, F. Addressing Data Scarcity in the Medical Domain: A GPT-Based Approach for Synthetic Data Generation and Feature Extraction. *Information* **2024**, *15*, 264. <https://doi.org/10.3390/info15050264>

Academic Editor: Evaggelos Karvounis

Received: 25 March 2024

Revised: 24 April 2024

Accepted: 3 May 2024

Published: 6 May 2024



Copyright: © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The burgeoning field of medical machine learning confronts an ardent challenge—the paucity of comprehensive and clinically labeled training data [1,2]. The intricate nature of medical data, coupled with stringent privacy regulations, results in a scarcity that hampers the efficacy of machine learning models in healthcare applications. In particular, the insufficiency of labeled data exacerbates the predicament, impeding the ability to develop robust models capable of meaningful clinical insights [1,2].

This research endeavors to alleviate the constraints posed by the limited availability of labeled medical data by harnessing the unparalleled capabilities of Generative Pre-trained Transformers (GPT). In this study, we propose a novel approach that utilizes GPT to synthetically generate medical data, thereby circumventing the challenges associated with data scarcity. Moreover, GPT's intrinsic ability to analyze and comprehend the synthetic data it generates opens avenues for the extraction of new features, offering a solution to the dearth of labeled data in the medical domain.

The first phase of our investigation involves a meticulous and systematic review of the existing literature, delving into the capabilities of GPT in synthetic data generation. By scrutinizing prior studies, we aim to provide a comprehensive understanding of GPT's prowess in generating synthetic data for training machine learning models, thereby laying the groundwork for the subsequent phases of our research. Building upon the insights gleaned from the literature, our study proceeds to explore how GPT can not only generate

synthetic data but also engage in the analysis of these datasets to extract novel features. Through a critical examination of existing methodologies, we seek to elucidate the potential of GPT in addressing the challenge of data scarcity from a holistic perspective.

As a practical demonstration of our proposed approach, we present a method for synthetically generating patient discharge messages using GPT, as conceptually represented in Figure 1. This pragmatic application serves as a testament to the feasibility and effectiveness of our proposed solution in tackling the limited availability of training data in the medical domain. Furthermore, we showcase how GPT can play a pivotal role in feature extraction from these synthetic patient discharge messages, illustrating its capability to mitigate the scarcity of labeled data (as shown in Figure 1). Through these empirical demonstrations, we aim to establish a robust foundation for the integration of GPT into the realm of medical machine learning, paving the way for enhanced model development in the face of data scarcity. Within the scope of this study, more than 70 patient discharge messages were automatically generated by the proposed GPT prompt. For all these discharge messages, seventeen fields were synthetically generated first, and then three more fields were generated for labeling these discharge message (e.g., severity, chances of hospital re-admission with justification).

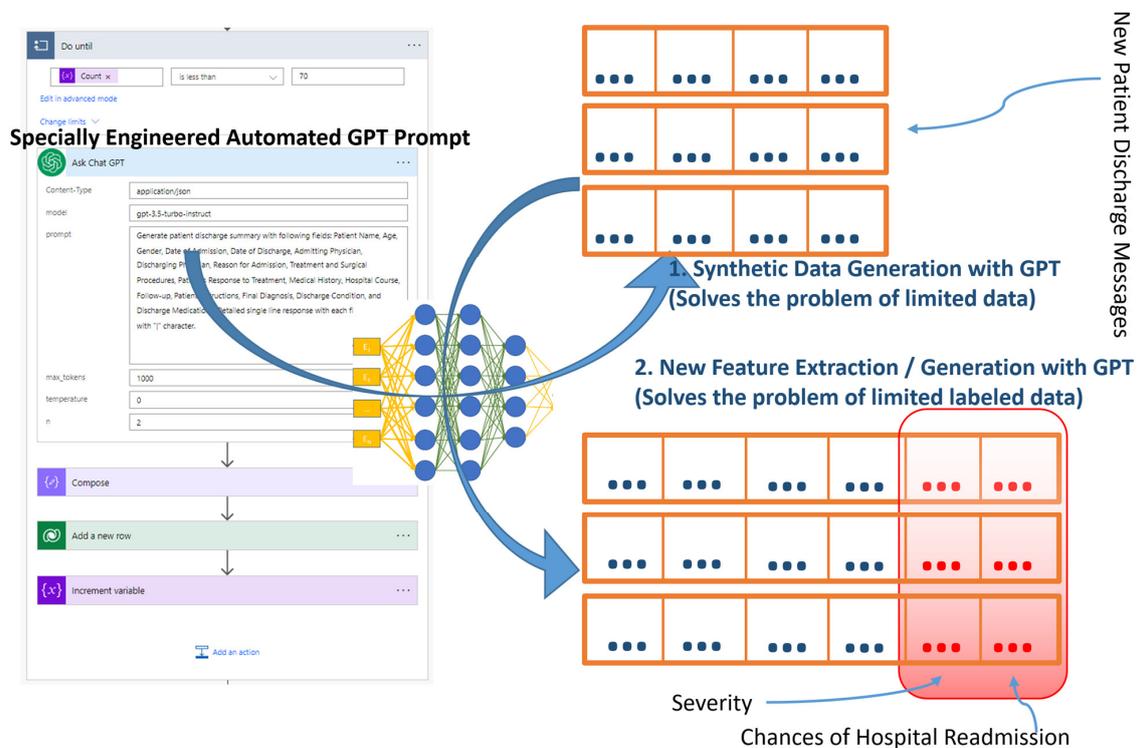


Figure 1. Conceptual diagram of GPT-based training data generation, feature extraction, and labelling.

This study contributes to the current body of knowledge in the following ways:

- Conducted a comprehensive review of existing literature to explore the utilization of GPT in the medical domain. Among twenty identified works, this study highlighted seven distinct research endeavors that employed GPT to generate or enhance medically relevant data [3–9].
- Unlike previous studies that relied on manual utilization of GPT’s web interface (as shown in [3–9]), this research autonomously leveraged the GPT Application Programming Interface (API) alongside automation tools, enabling the efficient generation of a large volume of medically significant data.
- Employing innovative prompt engineering techniques, this study generated 70 synthetic patient discharge messages encompassing seventeen fields and autonomously

labeled these messages using GPT technology, resulting in the addition of three augmented fields.

- The generated data underwent evaluation by medical professionals, yielding an impressive average precision, recall, and F1-score of 0.95, 0.97, and 0.96, respectively.
- Furthermore, the synthetically generated medical data were subjected to machine learning algorithms such as regression to uncover hidden correlations among various parameters.

In essence, this research seeks to contribute a novel and comprehensive methodology to the growing body of knowledge addressing the challenges posed by data scarcity in the medical domain [1,2,10]. According to the literature and to the best of our knowledge, this is the first study to generate highly accurate (with F1-score of up to 97%) patient discharge summaries using GPT technology.

2. Literature Review

A recent study in [11] reviews the use of ChatGPT in various aspects of medical research. It evaluates the evidence of ChatGPT's application in areas including but not limited to treatment, diagnosis, medication provision, drug development, medical report improvement, literature review writing, research conduct, data analysis, and personalized medicine. The review follows the PRISMA guidelines and encompasses studies published between 2022 and 2023. The paper in [12] explores the use of ChatGPT in the systematic review and meta-analysis process in medical research. The paper discusses how ChatGPT can be used for tasks like Risk of Bias analysis and data extraction from randomized controlled trials, highlighting the tool's ability to reduce the time and effort required for these tasks. It directly addresses the use of ChatGPT in streamlining the process of conducting systematic reviews and meta-analyses, which are integral components of evidence-based decision making in healthcare [12]. The paper illustrates how AI, specifically ChatGPT, can assist in various steps of the systematic review process, including evaluating methodologies and extracting data. The study in [13] focuses on the application of ChatGPT in streamlining the literature selection process for meta-analysis in medical research. It outlines a methodology for using ChatGPT to facilitate the screening of titles and abstracts during meta-analysis, aiming to reduce workload while maintaining recall efficiency. The study includes a glioma meta-analysis for validation and discusses the development of a pipeline called LARS (Literature Records Screener) to assess the performance of ChatGPT in this context [13]. It deals directly with improving the efficiency and effectiveness of literature selection and screening in the context of meta-analysis, a crucial step in systematic reviews and research synthesis [13]. The research work in [14] discusses the potential public health risks posed by large language models like ChatGPT, specifically focusing on the spread of misinformation (infodemic). It explores the evolution of these models, their impact on scientific literature production, and the need for policies to mitigate misinformation risks. It focuses on the broader public health impact and ethical considerations of AI technology in disseminating information [14]. The paper in [15] focuses on evaluating the use of large language models (LLMs) in healthcare. It addresses the need for a comprehensive evaluation framework that assesses LLMs not just for their natural language processing performance but also for their translational value in healthcare. The paper discusses various aspects of LLMs in healthcare, ethical concerns, and proposes a framework for evaluating their application in this field. It goes beyond just the technical aspects of LLMs and delves into the ethical, governance, and practical implications of their use in healthcare [15]. This paper emphasizes a comprehensive evaluation that includes translational value assessment and ethical considerations [15]. The publication in [16] examines the potential influence of large language models like ChatGPT on the field of nuclear medicine. It discusses the capabilities of these models in generating human-like text, their impact on academic publishing, and the potential risks associated with their use in the context of nuclear medicine. It highlights issues like academic integrity, misinformation, and the challenges posed by AI in producing reliable medical content [16]. The focus is on the broader implications of using AI tools

like ChatGPT in nuclear medicine, particularly concerning the reliability of the content produced and the ethical considerations surrounding their use in academic and clinical settings [16]. The discussion includes the potential for AI-generated content to influence academic integrity and the spread of misinformation, which are key concerns in the context of public health and ethical use of AI in medicine [16].

The paper in [3] explores the potential of AI, particularly large language models (LLMs) like GPT-4, in generating original scientific research. It discusses the use of GPT-4 to write an original pharmaceuticals manuscript, including formulating a research hypothesis, defining an experimental protocol, producing photo-realistic images, generating analytical data, and writing a publication-ready manuscript. This study also examines the limitations of LLMs in referencing literature and emphasizes the need for human input in interpretation and data validation [3]. It focuses on the innovative use of LLMs to generate and augment data, such as creating believable analytical data and images for pharmaceutical research [3]. The emphasis on the AI model's ability to conceive and execute a research hypothesis and generate multimodal data aligns with the aspects of data generation and augmentation [3]. Research work in [17] explores the applications of ChatGPT and other large language models in various aspects of orthopedics, including education, surgery, and research [17]. The study discusses how these AI tools can assist orthopedic clinicians and surgeons in tasks like disease diagnosis, surgical planning, and educational support. The focus is on the practical applications of ChatGPT in providing assistance to medical professionals in orthopedics, including aiding in diagnosis, surgery, and medical education, which aligns with the aspects of decision support and medical inquiry assistance [17]. The study in [18] presents a systematic review of the applications, benefits, and limitations of ChatGPT in healthcare education, research, and practice. The review includes an analysis of the potential benefits of ChatGPT in scientific writing, healthcare research, and practice, along with concerns regarding ethical, copyright, transparency, and legal issues [18]. Recent work in [19] examines the potential of AI systems, specifically large language models, in generating health awareness messages. The study uses the Bloom model for generating messages about folic acid, comparing them to highly retweeted human-generated messages in terms of quality and clarity. It also involves human and computational evaluations to assess the effectiveness of AI-generated messages in health communication. It focuses on the empirical assessment of AI-generated health messages, analyzing their effectiveness and comparing them to human-generated content [19]. The emphasis on computational and human evaluations of the messages aligns with the aspects of data analysis in medical research [19]. The study in [4] focuses on using GPT-3.5 for data augmentation to address vaccine hesitancy classification in the Dutch language. The study leverages the language model for generating realistic examples of anti-vaccination tweets and evaluates the impact of this augmentation on various machine learning models [4]. It also examines the ability of the synthetic data to generalize to human data in classification tasks. It illustrates the use of GPT-3.5 for generating synthetic data to balance an imbalanced dataset in vaccine hesitancy monitoring, highlighting its capabilities in data augmentation and labeling [4].

Recent work in [5] focuses on enhancing medical question answering systems using GPT-2 for question augmentation and T5-Small for topic extraction. The paper details a model that employs BERT, GPT-2, and T5-Small to improve medical question answering performance, demonstrating the effectiveness of these techniques through experiments [5]. It highlights the use of AI models for augmenting medical question data, a crucial aspect in improving the quality and coverage of datasets used in medical question answering systems [5]. The study in [6] examines the use of GPT-3 in generating synthetic data for Human-Computer Interaction (HCI) research. It explores the ability of GPT-3 to produce believable accounts of HCI experiences and discusses the potential benefits and risks associated with using synthetic data generated by language models. It highlights the use of GPT-3 for generating synthetic user research data, focusing on the model's ability to create realistic and believable responses in an HCI context [6]. The paper in [7] presents a study on using GPT-2 for data augmentation in the context of patient outcome prediction.

The focus is on generating artificial clinical notes in Electronic Health Records (EHRs) to improve the training of machine learning models for predicting patient outcomes, such as readmission rates. The paper discusses a novel textual data augmentation method and evaluates its effectiveness in enhancing predictive performance of deep learning models in healthcare [7]. It explores the use of GPT-2 to augment medical datasets, specifically focusing on generating textual data that can be used to train models for predicting patient outcomes, aligning with data augmentation and labeling aspects [7]. The research work in [8] focuses on using GPT-2 to generate synthetic biological signals, specifically EEG (electroencephalography) and EMG (electromyography), to enhance data classification. The study demonstrates that models trained on synthetic data generated by GPT-2 can classify real EEG and EMG datasets with significant accuracy and that the inclusion of synthetic data during training improves classification performance [8]. It emphasizes the use of AI for generating synthetic biological signals, which augments the available data for training machine learning models in the field of biological signal processing [8]. The paper in [9] focuses on using Transformer-based models, particularly GPT-2, for generating synthetic medical text to augment datasets. The study experiments with these models for data augmentation in clinically relevant NLP tasks such as unplanned readmission prediction and phenotype classification. It evaluates the effectiveness of synthetic data in improving the performance of deep learning models in these healthcare contexts [9]. It highlights the application of AI models in creating synthetic medical text data, aiming to augment existing datasets for improved model training and performance in specific medical tasks [9]. Finally, the paper in [20] discusses the potential of ChatGPT in various medical applications. It examines ChatGPT's ability to develop AI programs for medicine, its limitations and challenges, ethical concerns like biases and patient confidentiality, and compliance with healthcare regulations. The paper highlights ChatGPT's potential in democratizing coding and developing AI in medicine, leading to breakthroughs in the medical AI sector [20]. The focus on ethical concerns, patient autonomy, and the responsible use of AI in medicine, along with the exploration of AI's potential to revolutionize medical research and practice, aligns with this category [20]. These existing research works could be categorized into six distinct categories, as described in Figure 2.

- Literature Review and Meta-Analysis: Studies in [11–13,18] illustrate how AI, specifically ChatGPT, can streamline literature reviews and meta-analyses, aiding in efficient data extraction and evaluation methodologies.
- Data Analysis: As demonstrated in [21–26], GPT assists in analyzing research data and generating critical insights. Within the medical domain, research works in [11,19] demonstrate AI's utility in analyzing complex datasets, including patient outcomes and health message effectiveness, enhancing predictive modeling and comprehension of medical data.
- Medical Question Answering and Decision Support Systems: Studies like [11,17,18] show the role of AI in assisting medical professionals with accurate information, aiding diagnosis, and providing decision support in clinical settings.
- Drug Discovery and Clinical Trial Analysis: While not directly covered in the reviewed articles, this category involves using AI to accelerate drug discovery processes and analyze clinical trial data, potentially enhancing the efficiency and efficacy of pharmaceutical development [11].
- Ethical and Public Health Implications of AI in Medicine: Several recent studies like [11,14–16,18,20] discuss the broader ethical implications and public health concerns of AI in medicine, including misinformation and academic integrity.
- Data Generation, Augmentation, and Labeling: To generate new features from data with limited fields, machine learning techniques like entity recognition, category classification, sentiment analysis, and others have traditionally been used [27–34]. After generating new features, the augmented data can be used to effectively train the machine learning models [27–34]. However, with the advent of GPT, new features could be generated either from synthetic data or from existing data, without using

traditional feature extraction approaches, as shown in [35–70]. Even within the medical domain, synthetic data creation, data augmentation, and labelling have been proven to be crucial in recent times [3–9]. These papers illustrate the use of AI for creating and enhancing medical datasets, crucial for training robust machine learning models.

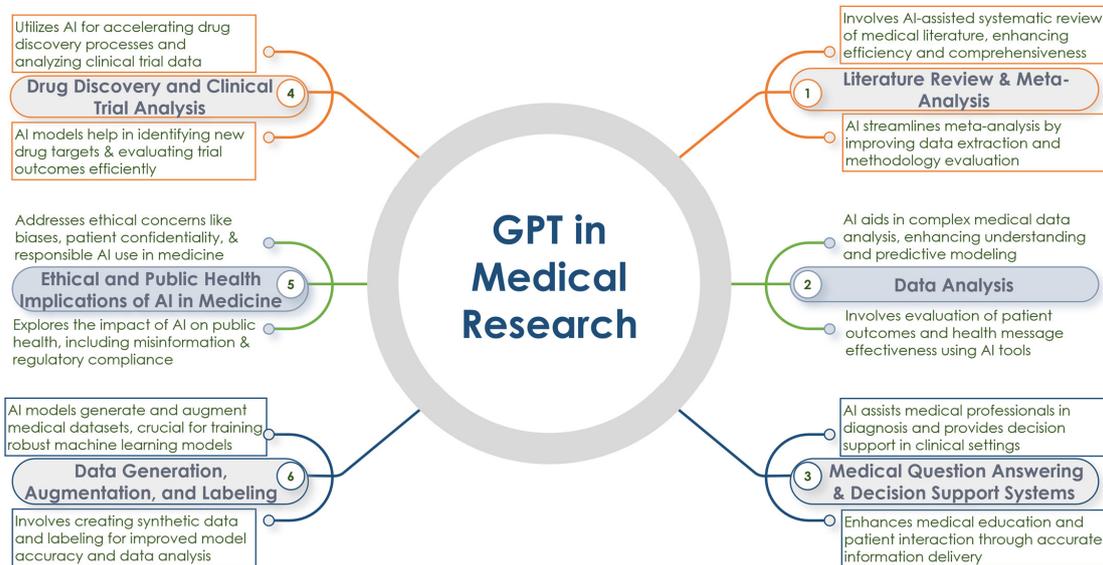


Figure 2. Six distinct areas of research for “GPT in Medical Domain”.

Finally, Table 1 clearly depicts how existing research works on using GPT in the medical domain could be categorized. As shown in Table 1, most of the existing literature falls under the category of “Data Generation, Augmentation, and Labeling”. Within the next section, a practical scenario of how GPT could be used to generate synthetic medical data as well as how to generate labels for these synthetic data will be detailed.

Table 1. Categorization of existing studies on the use of GPT in medical domain (X denotes “Topic of Interest”).

Reference	Literature Review and Meta-Analysis	Data Generation, Augmentation, and Labeling	Data Analysis	Medical Question Answering and Decision Support Systems	Drug Discovery and Clinical Trial Analysis	Ethical and Public Health Implications of AI in Medicine
[11]	X		X	X	X	X
[12]	X					
[13]	X					
[14]						X
[15]						X
[16]						X
[3]		X				
[17]				X		
[18]	X			X		X
[19]			X			
[4]		X				
[5]		X				
[6]		X				
[7]		X				
[8]		X				
[9]		X				
[20]						X

3. Methods

The GPT model is based on the Transformer architecture, which involves several key components, like Input Embedding and Positional Encoding, Transformer Blocks, Feed-Forward Neural Network, Normalization and Residual Connections, and Output layer [71].

3.1. Input Embedding and Positional Encoding

- Each input token (word or sub-word) is converted into a vector through an embedding layer.
- Positional encodings are added to these embeddings to provide information about the position of each token in the sequence.
- The combined embedding, E , is given by Equation (1).

$$E = E_{token} + E_{position} \quad (1)$$

3.2. Transformer Blocks

Each block consists of two main parts, the Multi-Head Self-Attention mechanism and the Feed-Forward Neural Network.

- Multi-Head Self-Attention:
 - a. The attention mechanism can be described by Equation (2).

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2)$$

- b. In Equation (2), Q , K , V are the query, key, and value matrices, and d_k is the dimension of the keys.
 - c. In multi-head attention, this process is carried out in parallel multiple times with different, learned linear projections of the queries, keys, and values. The outputs are then concatenated and linearly transformed.
- Feed-Forward Neural Network:
 - a. Each layer contains a fully connected feed-forward network, which is applied to each position separately and identically. This typically involves two linear transformations with a ReLU activation in between. It is represented with Equation (3).

$$FFN(x) = \max(0, xW_1 + b_1)W_2 + b_2 \quad (3)$$

3.3. Normalization and Residual Connections

- Each sub-layer (self-attention, feed-forward) in a transformer block has a residual connection around it, followed by layer normalization.
- The output of each sub-layer is $LayerNorm(x + Sublayer(x))$, where $Sublayer(x)$ is the function implemented by the sub-layer itself.

3.4. Output Layer

- The final layer is a linear transformation followed by a softmax function to predict the probability of the next token in the sequence.
- The output probabilities for a token are computed as $softmax(xW + b)$, where W and b are the weights and biases of the output layer.

This mathematical framework enables GPT to capture complex patterns and relationships in sequential data [71] and is used in this study to generate synthetic patient discharge messages and even perform analysis on those discharge messages for assessing severity and chances of hospital readmission.

3.5. The Process of Automating Synthetic Medical Data Generation

In the conventional approach, users of GPT technology access the model through its web interface, initiating interactions via specific prompts to derive outputs from the system (as shown in Figure 3). This traditional approach has been demonstrated by research works [3–9]. Employing such a traditional methodology to produce synthetic medical data necessitates substantial user involvement, which can be time-consuming. To circumvent the need for manual intervention in querying the GPT interface, the current study integrates the GPT API with Microsoft Power Automate to fully automate the process of generating patient discharge summaries, as shown in Figure 3. Microsoft Power Automate orchestrates the interactions with the GPT through its API, facilitating a seamless automated workflow. Consequently, this novel automation strategy enhances the efficiency and effectiveness of generating synthetic patient discharge messages, thus streamlining the process significantly. As seen from Figure 3, the proposed approach of interacting with ChatGPT API is automated, fast, and efficient.

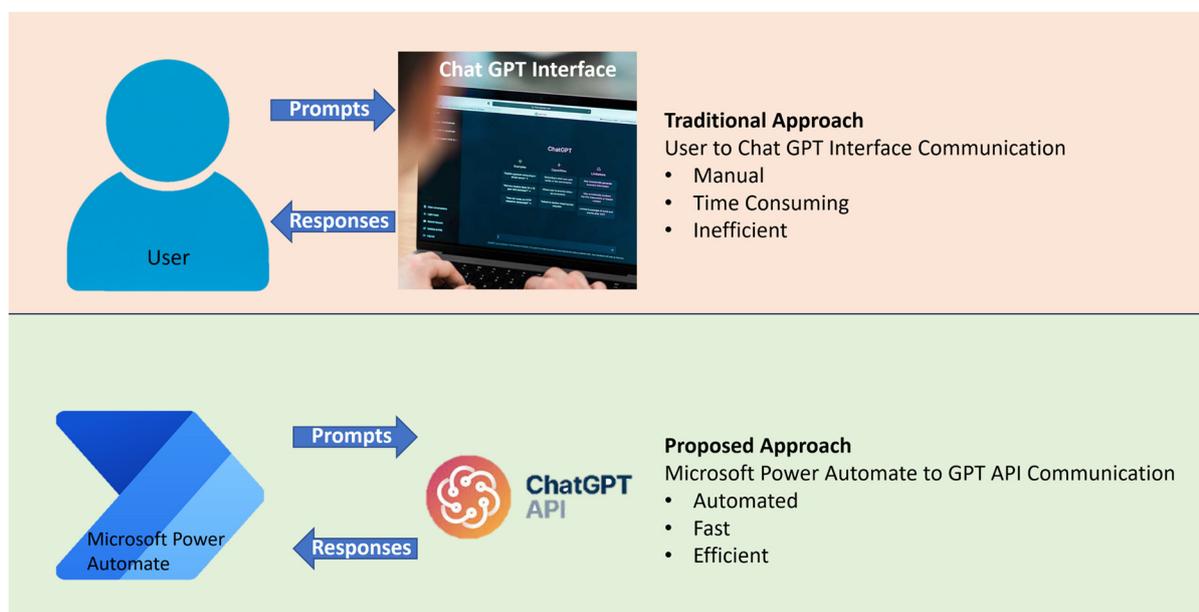


Figure 3. Traditional approach of manual interaction with Chat GPT web interface vs. fully automated interaction via GPT API.

As seen in Figure 4, the orchestration of GPT API communication is performed using Microsoft Power Automate. The HTTP request component of Microsoft Power Automate can autonomously invoke multiple API calls. As shown in Figure 4, the first HTTP post call to GPT API generates 70 discharge messages. The second HTTP post call then critically evaluates these messages and labels them in terms of (1) severity, (2) chances of hospital readmission, and (3) reasoning. The details of both these calls are shown in Figure 5. It should be noted that Microsoft Power Automate allows the second prompt to investigate the previously generated synthetic message through the variable “Output”, as shown in Figure 5b. Thus, the contextual background of the previously generated messages could be efficiently analyzed in the second prompt, along with augmenting the previous messages with newer labels (i.e., severity, chances of hospital readmission, and reasoning). The reasoning information would be validated by expert doctors at a later stage.

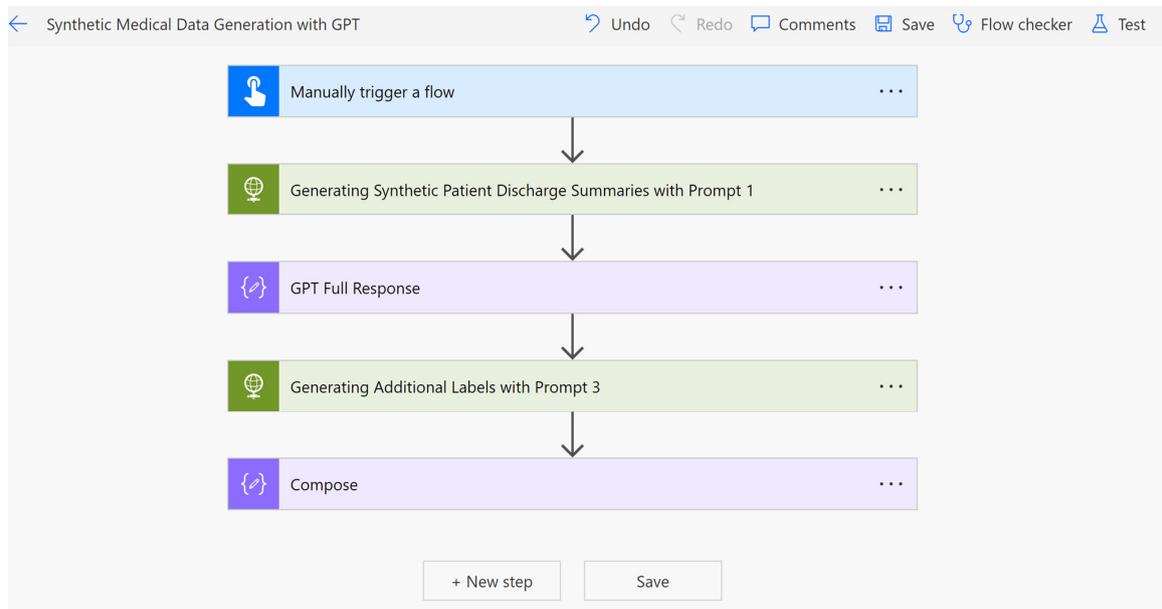


Figure 4. Microsoft Power Automate invoking API calls to GPT API in an automated manner using HTTP requests.

Generating Synthetic Patient Discharge Summaries with Prompt 1

*** Method** POST

*** URI** https://api.openai.com/v1/chat/completions

Headers

Authorization	Bearer XXXXXXXXXXXXXXXXXXXX
Content-Type	application/json
Enter key	Enter value

Queries

Enter key	Enter value
-----------	-------------

Body

```
{
  "model": "gpt-3.5-turbo",
  "messages": [
    {
      "role": "system",
      "content": "You are a Nurse at an Emergency Department."
    },
    {
      "role": "user",
      "content": "Generate 70 patient discharge summaries with following fields: Patient Name, Age, Gender, Date of Admission, Date of Discharge, Admitting Physician, Discharging Physician, Reason for Admission, Treatment and Surgical Procedures, Patient's Response to Treatment, Medical History, Hospital Course, Follow-up, Patient Instructions, Final Diagnosis, Discharge Condition, and Discharge Medications. Detailed single line response with each field separated with "|" character."
    }
  ]
}
```

Cookie Enter HTTP cookie

Show advanced options

(a)

Generating Additional Labels with Prompt 3

*** Method** POST

*** URI** https://api.openai.com/v1/chat/completions

Headers

Authorization	Bearer XXXXXXXXXXXXXXXXXXXX
Content-Type	application/json
Enter key	Enter value

Queries

Enter key	Enter value
-----------	-------------

Body

```
{
  "model": "gpt-3.5-turbo",
  "messages": [
    {
      "role": "system",
      "content": "You are a Nurse at an Emergency Department."
    },
    {
      "role": "user",
      "content": "Assess each of the patient's condition from [Outputs] and Rate the severities of these patients along with their chance of hospital readmission for each of these patients."
    }
  ]
}
```

Cookie Enter HTTP cookie

Show advanced options

(b)

Figure 5. The process of passing specially designed prompts through Microsoft Power Automate (HTTP post method). * preceding Method and URI denotes mandatory fields. (a) Generating 70 patient discharge messages. (b) Labelling each of the 70 messages with severity and chances of hospital readmission.

As shown in Figures 1 and 5, a specially engineered GPT prompt can be used for generating patient discharge messages. Microsoft Power Automate with GPT API automatically

generates patient discharge summaries with specifically guided headings, like Diagnosis, Treatment, Patient Instructions, Medications on Discharge, etc. The complete list can be seen from Appendix A using the prompt of Box 1. Many of these headings (presented in Appendix A) are required for assessment of severity and predicting the chances of hospital readmission, which would be performed in the next stage. As seen from Figures 6–9, GPT generated the discharge summaries synthetically (i.e., not real patient information).

Box 1. Generating Synthetic Patient Discharge Summaries.

Generate patient discharge summary with following fields: Patient Name, Age, Gender, Date of Admission, Date of Discharge, Admitting Physician, Discharging Physician, Reason for Admission, Treatment and Surgical Procedures, Patient’s Response to Treatment, Medical History, Hospital Course, Follow-up, Patient Instructions, Final Diagnosis, Discharge Condition, and Discharge Medications. Detailed single line response with each field separated with “|” character.

<p>Patient Name</p> <p>Alex Johnson</p> 	<p>Diagnosis</p> <p>Acute Appendicitis</p> <hr/> <p>Treatment</p> <p>Laparoscopic Appendectomy</p> <hr/> <p>Admitting Doctor</p> <p>Dr. Sarah Miller</p>	<p>Hospital Course</p> <p>Patient was admitted with acute abdominal pain. CT scan confirmed acute appendicitis. Underwent successful laparoscopic appendectomy. Postoperative recovery was uncomplicated. Pain was well-managed with medication. Diet was gradually advanced from liquids to solids. Patient demonstrated good mobility and strength.</p>	
<p>Wednesday, January 10, 2024</p> <p>Date of Admission</p>	<p>Tuesday, January 23, 2024</p> <p>Date of Discharge</p>	<p>Male</p> <p>Gender</p>	<p>38</p> <p>Age</p>
<p>Follow Up</p> <p>Follow-up appointment with Dr. Miller in 2 weeks. Immediate return to the ER if fever, increased pain, or signs of infection</p>	<p>Patient Instruction</p> <p>Avoid heavy lifting for 4 weeks. Gradually increase activity. Keep surgical area clean. Observe for any signs of infection</p>	<p>Medication on Discharge</p> <p>Paracetamol 500 mg every 6 hours as needed for pain. Ibuprofen 400 mg every 8 hours as needed for pain. Cephalexin 500 mg every 12 hours for 7 days</p>	

Figure 6. Synthetic patient discharge summary generated for Alex Johnson using GPT prompt.

<p>Patient Name</p> <p>Sophia Martinez</p> 	<p>Diagnosis</p> <p>Acute Migraine with Aura</p> <hr/> <p>Treatment</p> <p>Intravenous hydration, Medication for migraine relief (Sumatriptan), Rest in a dark, quiet room</p> <hr/> <p>Admitting Doctor</p> <p>Dr. Kevin Nguyen</p>	<p>Hospital Course</p> <p>Patient presented with severe migraine accompanied by visual aura. Responded well to medication and rest. No complications during hospital stay. Educated on migraine triggers and stress management techniques.</p>	
<p>Saturday, January 20, 2024</p> <p>Date of Admission</p>	<p>Tuesday, January 23, 2024</p> <p>Date of Discharge</p>	<p>Female</p> <p>Gender</p>	<p>32</p> <p>Age</p>
<p>Follow Up</p> <p>Neurology appointment in 2 weeks. Keep a headache diary. Return to ER if experiencing severe headache with neurological symptoms</p>	<p>Patient Instruction</p> <p>Avoid migraine triggers. Maintain regular sleep schedule. Stay hydrated. Practice stress-reducing activities</p>	<p>Medication on Discharge</p> <p>Sumatriptan 50 mg as needed for migraine. Ibuprofen 400 mg every 6 hours as needed for pain</p>	

Figure 7. Synthetic patient discharge summary generated for Sophia Martinez using GPT prompt.

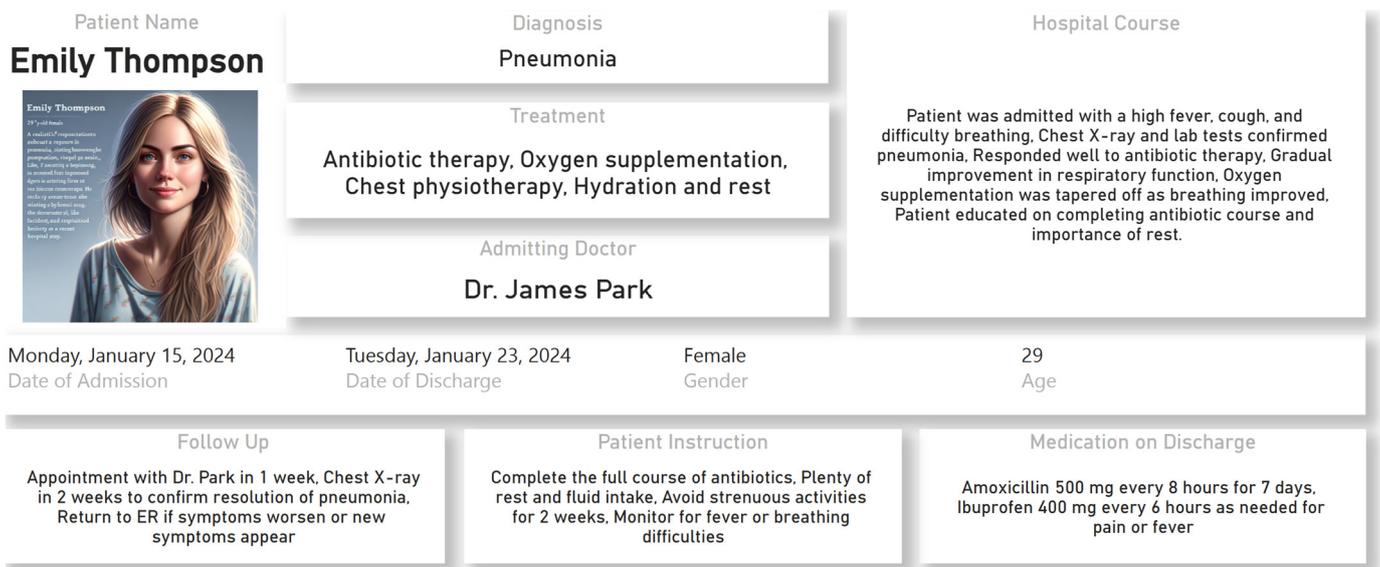


Figure 8. Synthetic patient discharge summary generated for Emily Thompson using GPT prompt.



Figure 9. Synthetic patient discharge summary generated for Michael Roberts using GPT prompt.

Images of the patients could also be generated by adding prompt of Box 2 along with Box 1.

Box 2. Generating the Images of the Patients Using the Information from Discharge Summaries.

Based on the description of the generated discharge summary, generate an image of that patient.

For Alex Johnson (Figure 6), the GPT response before generating the synthetic patient image is “Based on this summary, I will create an artistic representation of Alex Johnson, a 38-year-old male who has just recovered from an appendectomy. Let’s visualize Alex as having short brown hair, a medium build, and a friendly appearance, reflecting his recovery phase”.

As shown earlier in Figure 1 from the synthetically generated discharge summaries, GPT can effectively be used for generating new features. Figures 5b and 10 illustrate this process further. As seen from Figure 10, critical information (e.g., nature of their medical conditions, treatments received, and the instructions provided upon discharge) are used

for generating new features like severity of condition and change of hospital readmission. Box 3 shows the GPT prompt used for this feature augmentation process (as previously demonstrated in Figure 5b).

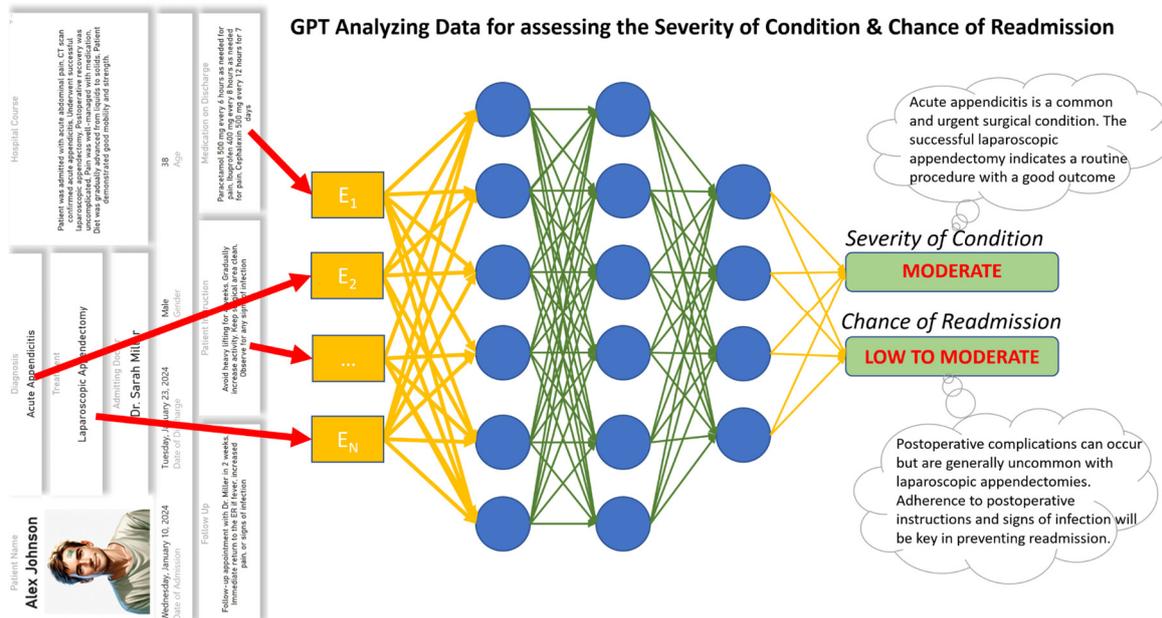


Figure 10. Feature extraction process using GPT for labelling the discharge messages.

Box 3. Generating New Features for Labeling the Discharge Messages.

Rate the severities of these patients along with their chance of hospital readmission for each of these patients.

As seen from Figure 10, for Alex Johnson (i.e., discharge summary presented in Figure 6), GPT assessed the severity of his condition to be “Moderate” and the changes of hospital readmission to be “Low to Moderate”. This process can be effectively used to label the synthetic data as low, moderate, high, etc., and could be efficiently used to train machine learning models at a later stage. The same methodology could be used for generating synthetic electrocardiogram signals or other bio-signals as well as labelling these signals. Hence, GPT to solve GPT is presented as an effective solution towards solving data scarcity as well as fewer labels in the medical domain.

4. Results

Using the methodology detailed in the previous section, within this study, 70 patient discharge summaries were synthetically generated. As seen from Table 2, these patient discharge summaries had 20 fields comprising Patient Name, Age, Gender, Date of Admission, Date of Discharge, Admitting Physician, Discharging Physician, Reason for Admission, Treatment and Surgical Procedures, Patient’s Response to Treatment, Medical History, Hospital Course, Follow-up, Patient Instructions, Final Diagnosis, Discharge Condition, Discharge Medications, Severity Level, Probability of Hospital Re-admission, and Reasoning. As mentioned in the previous section, the first 17 fields were generated with GPT Prompt 1 and then labelling information (i.e., Severity Level, Probability of Hospital Re-admission, and Reasoning) was generated with Prompt 2. Appendix A shows the details of these 70 generated discharge summaries. Out of these 20 fields, only Age was numeric in nature, and as a result, Table 3 provides various statistics on this numeric field. The value of Age ranged between 23 and 89. There were two date fields, namely date of admission and date of discharge.

Table 2. Seventy synthetically generated patient discharge summaries with 20 fields each.

Terminologies	Data Generation	Data Analysis (Labeling)	Data Type	Distinct Values	Unique Values	Example
Patient Name	X		String	70	70	John Doe
Age	X		Number	46	27	34
Gender	X		Binary	2	0	Male
Date of Admission	X		Date	62	55	5 January 2021
Date of Discharge	X		Date	61	54	2 February 2021
Admitting Physician	X		String	56	44	Dr. Smith
Discharging Physician	X		String	59	49	Dr. Williams
Reason for Admission	X		String	60	53	Acute appendicitis
Treatment and Surgical Procedures	X		String	64	59	Appendectomy
Patient’s Response to Treatment	X		String	64	59	Patient responded well to surgical intervention
Medical History	X		String	51	45	No significant past medical history
Hospital Course	X		String	69	68	Patient underwent successful appendectomy, recovered without complications
Follow-up	X		String	51	39	To review in outpatient clinic after 1 week
Patient Instructions	X		String	67	66	Light diet, rest and wound care
Final Diagnosis	X		String	65	60	Final diagnosis of acute appendicitis
Discharge Condition	X		String	12	8	Stable at discharge
Discharge Medications	X		String	69	68	Prescribed antibiotics, painkillers, and laxatives
Severity Level		X	String	3	0	Moderate
Probability of Hospital Re-admission		X	String	3	0	Low
Reasoning		X	String	69	68	Severity based on condition ‘Acute appendicitis’. Readmission probability based on discharge condition ‘Stable at discharge’ and medical history.

Table 3. Statistics on Age field.

Index	Age
count	70
mean	56.57142857
std	17.27322753
min	23
25%	45
50%	58
75%	70.75
max	89

Date of admission ranged from 12 January 2021 to 20 December 2021. Date of discharge ranged from 20 January 2021 to 30 December 2021. From these date fields, the duration of hospital stay could be calculated. Hospital stay ranged from 3 (for Sophie Duncan) to 334 days (Maria Johnson). Finally, Figure 11 shows the distributions of labeling data (i.e., Severity level and Chances of Hospital Re-admission). As seen from Figure 11, 12.86% of the discharge summaries were labeled with the severity level of high and 67.14% of the discharge summaries were labeled with severity level being low. In terms of hospital re-admission, 60% of cases were moderate, 24.29% of cases were low, and 15.71% of the cases were flagged as “moderate to high”.

The last three columns in Table 3, namely Severity Level, Probability of Hospital Re-admission, and Reasoning, were generated anew using Prompt 3. This additional information was autonomously generated by GPT, as demonstrated in Figure 5b. Given that GPT was instructed to act as a medical professional in generating these details, the augmented data underwent evaluation by two medical experts.

The evaluation results are depicted in Table 4, revealing an average precision, recall, and F1-score of 0.95, 0.97, and 0.96, respectively, across all three labeled tasks. This indicates GPT’s capability to automatically label medical data with a high level of accuracy. Notably, in Table 4, the F1-Score was highest, at 97% for reasoning, followed by severity and likelihood of hospital admission. This manual validation process underscores the potential for utilizing GPT and related technologies with confidence in generating and enhancing synthetic medical data.

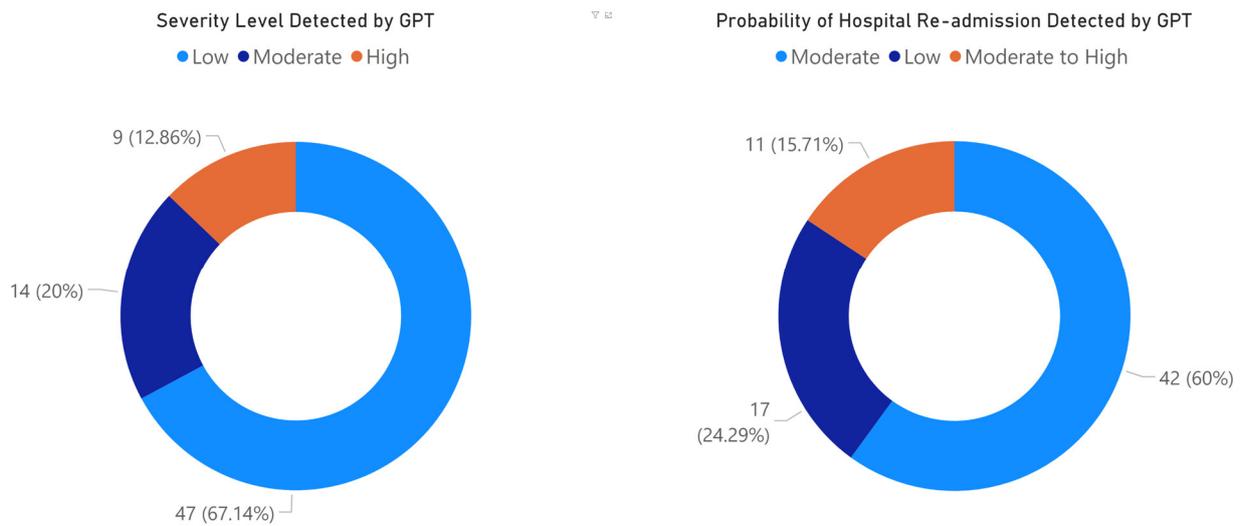


Figure 11. Results of labeling patient discharge summaries with GPT.

Table 4. Evaluation of the augmented data by GPT.

	TP	TN	FP	FN	Precision	Recall	F1-Score
Severity	62	4	3	1	0.953846	0.984127	0.96875
Chances of Hospital Readmission	59	5	4	2	0.936508	0.967213	0.951613
Reasoning	63	3	2	2	0.969231	0.969231	0.969231

Other than manually evaluating the validity of generated information, machine learning algorithms could also be used on the generated synthetic data for obtaining AI-driven insights [72]. The next section will discuss how machine learning algorithms could be used on these synthetic data for obtaining AI-driven insights.

5. Discussion and Concluding Remarks

This research introduces a groundbreaking methodology to address the challenge of data scarcity in medical machine learning by leveraging the capabilities of GPT. The study proposes a comprehensive approach that utilizes GPT for synthetic data generation and subsequent feature extraction, offering a transformative solution to the limitations imposed by the scarcity of labeled medical data. The empirical demonstration involving the synthetic generation of patient discharge messages serves as a practical testament to the feasibility and effectiveness of the proposed methodology, showcasing its potential to revolutionize the integration of GPT into the realm of medical machine learning. Figure 12 shows the deployment of the GPT-based solution in the latest Samsung Galaxy S23 Ultra mobile phone using Microsoft Power BI’s deployed App. The application of this deployment process has been showcased in recent studies through the utilization of low-code platforms [27,30–32,34]. As this study exclusively solved the labeled data scarcity for training machine learning models within medical domain (as discussed in [1,2,10]), it needs to be demonstrated how the generated synthetic data could be used in machine learning. Figure 12 shows that automated regression identified “Hospital Stays” to be highly correlated with the severity of the patient. The AI-driven insight shown in Figure 12 (within Samsung Galaxy S23 Ultra Mobile) shows that out of the 19 fields, Patient’s Age, Chance of Hospital readmission, and Hospital stays are correlated with severity. This automated regression using “Key Influencer” visualization of Microsoft Power BI has been reported in [73]. The previous section evaluated the validity of the generated medical data using manual evaluation by an expert medical professional. Now, this section demonstrates the use of the automated machine learning algorithm (i.e., regression to obtain the correlated variables) on the synthetic data.

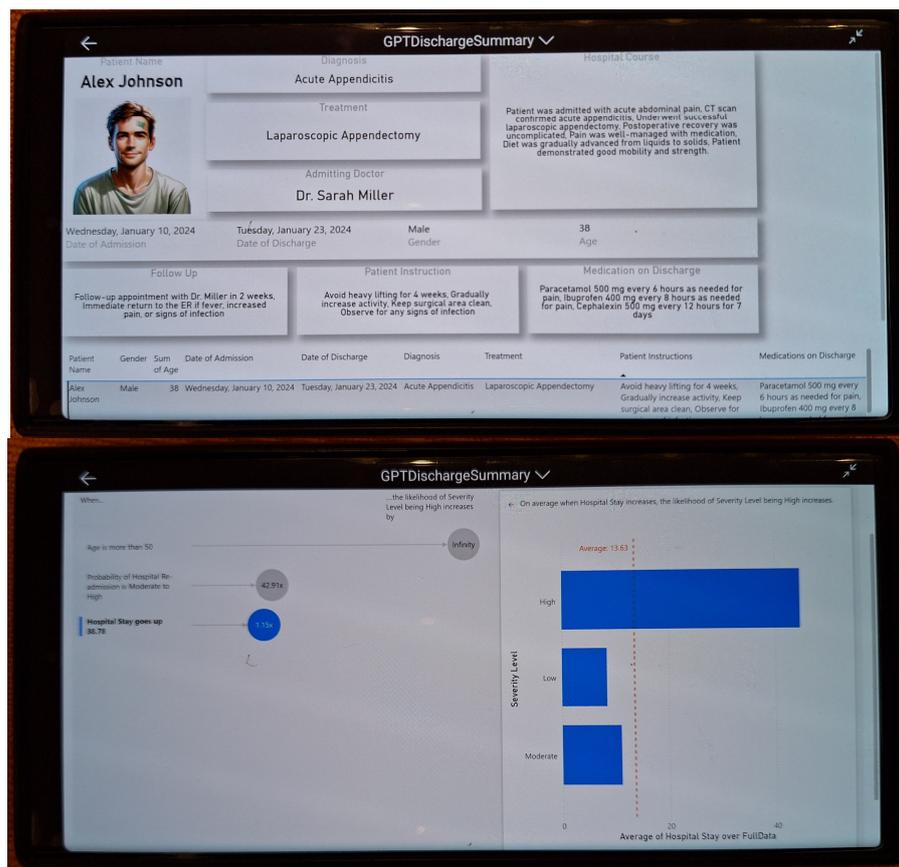


Figure 12. GPT-based patient discharge summary viewed and analyzed with machine learning algorithms in Samsung Galaxy S23 Ultra.

In summary, this study presents a pioneering and thorough methodology designed to address the data scarcity issues faced by researchers and scientists in the medical field. Leveraging this approach, automation tools such as Microsoft Power Automate were employed alongside the ChatGPT API to not only generate synthetic medical data automatically but also to label these datasets autonomously. The labeling process conducted by GPT was manually assessed by medical experts, yielding an impressive F1-score of 97%. Additionally, machine learning techniques, including regression analysis, were applied to the synthetic data, affirming the validity of the generated information. The integration of ChatGPT API's synthetic data generation and feature extraction capabilities not only facilitates the development of more robust machine learning models for healthcare applications but also sets the stage for future research endeavors. Future works should explore the application of GPT across diverse medical datasets, optimize its capabilities for specific contexts, and delve into the ethical implications of deploying synthetic data in medical research. This study lays the foundation for a trajectory of research that promises to redefine the landscape of medical machine learning, ultimately benefiting both researchers and clinicians in their pursuit of improved healthcare outcomes.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data attached within this paper.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A

Table A1. Seventy patient discharge summaries generated with GPT.

Patient Name	Age	Gender	Date of Admission	Date of Discharge	Admitting Physician	Discharging Physician	Reason for Admission	Treatment and Surgical Procedures	Patient's Response to Treatment	Medical History	Hospital Course	Follow-Up	Patient Instructions	Final Diagnosis	Discharge Condition	Discharge Medications	Severity Level	Probability of Hospital Re-Admission	Reasoning
John Doe	34	Male	1/1/2021	2/2/2021	Dr. Smith	Dr. Williams	Acute appendicitis	Appendectomy	Patient responded well to surgical intervention	No significant past medical history	Patient underwent successful appendectomy, recovered without complications	To review in outpatient clinic after 1 week	Light diet, rest and wound care	Final diagnosis of acute appendicitis	Stable at discharge	Prescribed antibiotics, painkillers, and laxatives	Moderate	Low	Severity based on condition 'Acute appendicitis'. Readmission probability based on discharge condition 'Stable at discharge' and medical history.
Maria Johnson	56	Female	1/12/2021	12/12/2021	Dr. Johnson	Dr. Robinson	Stroke	IV Thrombolysis, Physical therapy	Significant improvement in mobility and speech	History of hypertension and heart disease	Patient received thrombolysis within time limit and underwent intense rehab	To review in stroke clinic after 4 weeks	Medication compliance, regular exercise, and healthy diet	Final diagnosis of ischemic stroke	Functional improvements, stable at discharge	Prescribed blood thinners, statins, and antihypertensives	High	Moderate to High	Severity based on condition 'Stroke'. Readmission probability based on discharge condition 'Functional improvements, stable at discharge' and medical history.
Susan Harris	38	Female	3/15/2021	3/20/2021	Dr. Russo	Dr. Murray	Gallstones	Laparoscopic cholecystectomy	Patient responded well to surgery	No significant past medical history	Surgery was uncomplicated and patient recovered without issue	Follow up with primary care in 2 weeks	Maintain low-fat diet	Final diagnosis of cholelithiasis and cholecystitis	Stable, full recovery anticipated	Prescribed painkillers and antibiotics.	Moderate	Low	Severity based on condition 'Gallstones'. Readmission probability based on discharge condition 'Stable, full recovery anticipated' and medical history.
James Thompson	69	Male	2/1/2021	2/7/2021	Dr. White	Dr. Black	Chest pain, confirmed as myocardial infarction	Angioplasty and stent placement	Patient showed remarkable improvement post-procedure	Has a history of diabetes and hypertension	Patient had a successful procedure and was monitored in ICU for a day. Released later to general ward	Cardiology follow-up in one month	Lifestyle modification, medication compliance	Acute anterior wall Myocardial Infarction	Stable at discharge	Medications including antiplatelets, beta-blockers, ACE inhibitors, statins and anti-diabetic regimen.	High	Moderate to High	Severity based on condition 'Myocardial Infarction'. Readmission probability based on discharge condition 'Stable at discharge' and medical history of diabetes and hypertension.
Elizabeth Davis	42	Female	4/10/2021	4/15/2021	Dr. Turner	Dr. Walker	Pneumonia	Antibiotics treatment and respiratory therapy	Patient's condition improved significantly	Previously healthy with no significant medical history	Treated with IV antibiotics and oxygen through nasal cannula	Pulmonary follow-up in 3 weeks	Completion of oral antibiotic course, rest, and hydration	Final diagnosis of community-acquired pneumonia	Improving at discharge	Oral antibiotics and bronchodilator inhaler.	Moderate	Low	Severity based on condition 'Pneumonia'. Readmission probability based on discharge condition 'Improving at discharge' and previously healthy status.
David Wilson	57	Male	10/21/2021	10/31/2021	Dr. Morris	Dr. Wright	Liver failure	Supportive care, liver transplant assessment	Slow but steady improvement	History of alcoholism and Hepatitis C	Patient managed with diuretics and lactulose, assessed for transplant suitability	Follow-up with hepatology team in 1 week	Avoidance of alcohol, low salt diet	End-stage liver disease	Stable at discharge, with close outpatient monitoring	Prescribed diuretics, lactulose, and multivitamins.	High	Moderate to High	Severity based on condition 'Liver failure'. Readmission probability based on discharge condition 'Stable at discharge, with close outpatient monitoring' and medical history of alcoholism and Hepatitis C.

Table A1. Cont.

Patient Name	Age	Gender	Date of Admission	Date of Discharge	Admitting Physician	Discharging Physician	Reason for Admission	Treatment and Surgical Procedures	Patient's Response to Treatment	Medical History	Hospital Course	Follow-Up	Patient Instructions	Final Diagnosis	Discharge Condition	Discharge Medications	Severity Level	Probability of Hospital Re-Admission	Reasoning
Anna Taylor	89	Female	5/9/2021	5/16/2021	Dr. Simmons	Dr. Mitchell	Hip fracture after fall	Hip pinning surgery	Gradual improvement with physical therapy	Osteoporosis, past history of falls	Surgery was successful with no complications, physiotherapy started postoperatively	Ortho follow-up after 2 weeks	Physical therapy, fall precautions at home	Femoral neck fracture	Stable with improving mobility	Analgesics and Calcium and Vitamin D supplements.	Moderate	Low	Severity based on condition 'Hip fracture'. Readmission probability based on discharge condition 'Stable with improving mobility' and medical history of osteoporosis.
Michael Anderson	72	Male	6/19/2021	6/28/2021	Dr. Young	Dr. Hernandez	Prostate cancer	Prostatectomy	Well tolerated procedure with good recovery	Past history of asthma	Surgery completed successfully and patient made steady progress in recovery	Urology follow-up after 1 month	Medication compliance, report any urinary difficulties	Prostate adenocarcinoma	Stable at discharge	Prescribed painkillers and inhaled corticosteroids.	High	Low	Severity based on condition 'Prostate cancer'. Readmission probability based on discharge condition 'Stable at discharge' and past history of asthma.
Patricia Lee	52	Female	8/1/2021	8/7/2021	Dr. Morris	Dr. Hall	Breast Cancer	Lumpectomy and radiation	Good recovery with no post-op complications	First degree relative with breast cancer	Surgery completed with clear margins, initiated on post-op radiation	Oncology follow-up in 1 week	Healthy diet, regular exercise, follow recommended screening guidelines	Breast Cancer, stage IIa	Stable at discharge	Prescribe painkillers and anti-emetics.	High	Moderate to High	Severity based on condition 'Breast Cancer'. Readmission probability based on discharge condition 'Stable at discharge' and family history of breast cancer.
Jacob Martinez	30	Male	11/5/2021	11/9/2021	Dr. King	Dr. Gonzalez	Acute pancreatitis	Fluid resuscitation and supportive care	Improved significantly with treatment	History of gallstones	Patient received IV fluids and pain management	GI follow up in 2 weeks	Low-fat diet, avoid alcohol, medication compliance	Acute pancreatitis	Improved, stable at discharge	Prescribed pain medication and proton pump inhibitors.	Low	Moderate to High	Severity based on condition 'Acute pancreatitis'. Readmission probability based on discharge condition 'Improved, stable at discharge' and medical history of gallstones.
Melissa Martin	65	Female	9/19/2021	10/1/2021	Dr. Thompson	Dr. Moore	Type 2 Diabetes Complications	Insulin Therapy, Diabetic Education	Patient responded well to therapy	Long-standing history of Type 2 Diabetes	Patient was educated about the importance of regular blood sugar monitoring, diet and exercise	Endocrinology follow up in 1 month	Regular blood sugar monitoring, maintain balanced diet, regular exercise	Uncontrolled Type 2 Diabetes	Stable at discharge	Insulin and oral hypoglycemic agents.	High	Moderate to High	Severity based on condition 'Type 2 Diabetes Complications'. Readmission probability based on discharge condition 'Stable at discharge' and medical history of Type 2 Diabetes.
Jason Jackson	45	Male	5/22/2021	6/1/2021	Dr. Roberts	Dr. Lopez	Traumatic Brain Injury	Debulking surgery, rehabilitation	Patient showed gradual improvement	No remarkable past medical history	Patient underwent surgery and was transferred to rehabilitation post-stabilization	Neurosurgery follow-up in 1 week	Ongoing rehabilitation, medication adherence	Traumatic Brain Injury	Fair condition at discharge	Prescribed anticonvulsants and analgesics.	Moderate	Moderate to High	Severity based on condition 'Traumatic Brain Injury'. Readmission probability based on discharge condition 'Fair condition at discharge' and medical history.

Table A1. Cont.

Patient Name	Age	Gender	Date of Admission	Date of Discharge	Admitting Physician	Discharging Physician	Reason for Admission	Treatment and Surgical Procedures	Patient's Response to Treatment	Medical History	Hospital Course	Follow-Up	Patient Instructions	Final Diagnosis	Discharge Condition	Discharge Medications	Severity Level	Probability of Hospital Re-Admission	Reasoning
Linda Ramos	70	Female	12/10/2021	12/20/2021	Dr. Reed	Dr. Jenkins	Chronic Obstructive Pulmonary Disease (COPD) exacerbation	Inhaler therapy, steroids, antibiotics	Patient's breathing improved significantly	History of smoking and COPD	Managed with nebulizers, steroids and antibiotics	Pulmonary follow-up in 2 weeks	Smoking cessation, use inhalers as instructed	Chronic Obstructive Pulmonary Disease, acute exacerbation	Stable at discharge	Prescribed inhalers, steroids and antibiotics.	High	Moderate to High	Severity based on condition 'COPD exacerbation'. Readmission probability based on discharge condition 'Stable at discharge' and history of smoking and COPD.
Joshua White	62	Male	7/7/2021	7/12/2021	Dr. Foster	Dr. Simmons	Heart failure exacerbation	Diuretics, ACE inhibitors, lifestyle modification	Patient's condition improved and stabilized	History of hypertension and heart disease	Managed with medications and patient education about lifestyle changes	Cardiology follow-up in 1 month	Regular exercise, low sodium diet, medication compliance	Congestive Heart Failure, acute exacerbation	Stable at discharge	Prescribed diuretics, ACE inhibitors and beta blockers.	High	Moderate to High	Severity based on condition 'Heart failure exacerbation'. Readmission probability based on discharge condition 'Stable at discharge' and history of hypertension and heart disease.
Emma Bailey	88	Female	9/15/2021	10/1/2021	Dr. Russell	Dr. Watson	Alzheimer's disease, behavioral changes	Adjustment of medications, behavioral therapy	Gradual improvement in sleep pattern and agitation	Long-standing Alzheimer's disease	Patient was managed with adjustment of Alzheimer's medications and behavioral techniques	Neurology follow-up in 1 month	Routine, structured day, family support	Alzheimer's disease with behavioral complications	Stable at discharge	Prescribed Donepezil, antipsychotics and sleep aids.	Moderate	Low	Severity based on condition 'Alzheimer's disease, behavioral changes'. Readmission probability based on discharge condition 'Stable at discharge' and long-standing Alzheimer's disease.
Michael Cox	77	Male	12/20/2021	12/30/2021	Dr. Rogers	Dr. Bennett	Complications of Chronic Kidney Disease	Dialysis, nutritional counseling	Patient's renal function improved significantly	History of Chronic Kidney Disease and Hypertension	Managed with dialysis and medications	Nephrology follow-up in 2 weeks	Low sodium, low potassium diet, medication compliance	Chronic Kidney Disease, stage V	Stable at discharge	Prescribed blood pressure medications, phosphate binders and erythropoietin.	High	Moderate to High	Severity based on condition 'Complications of Chronic Kidney Disease'. Readmission probability based on discharge condition 'Stable at discharge' and history of Chronic Kidney Disease and Hypertension.
Sarah Walker	64	Female	7/25/2021	8/5/2021	Dr. Richardson	Dr. Hughes	Gastritis	Antacid administration, dietary changes	Patient experienced reduction of symptoms	History of gastritis and GERD	Managed with antacids and dietary changes	Follow-up appointment with gastroenterologist in 3 weeks	Avoid spicy food, medication compliance	Acute gastritis	Stable at discharge	Prescribed Proton-pump inhibitors.	Low	Low	Severity based on condition 'Gastritis'. Readmission probability based on discharge condition 'Stable at discharge' and history of gastritis and GERD.
Christopher Cooper	85	Male	8/1/2021	8/10/2021	Dr. Ramirez	Dr. Hill	Rheumatoid Arthritis pain	Pain medication adjustment, physical therapy	Patient's mobility improved and pain reduced	History of Rheumatoid Arthritis	Pain management approach adjusted, PT introduced	Follow-up with Rheumatologist in 2 weeks	Physical therapy exercises, medication compliance	Rheumatoid arthritis with acute flare	Stable at discharge	Prescribed NSAIDs, steroids, DMARDs.	Moderate	Low	Severity based on condition 'Rheumatoid Arthritis pain'. Readmission probability based on discharge condition 'Stable at discharge' and history of Rheumatoid Arthritis.

Table A1. Cont.

Patient Name	Age	Gender	Date of Admission	Date of Discharge	Admitting Physician	Discharging Physician	Reason for Admission	Treatment and Surgical Procedures	Patient's Response to Treatment	Medical History	Hospital Course	Follow-Up	Patient Instructions	Final Diagnosis	Discharge Condition	Discharge Medications	Severity Level	Probability of Hospital Re-Admission	Reasoning
Amanda Bell	59	Female	11/15/2021	11/25/2021	Dr. Graham	Dr. Meyer	Depression	Cognitive Behavioral Therapy, medication adjustment	Patient's mood improved with treatment	History of Major Depressive Disorder	Treatment included medication adjustment and therapy	Psychiatry follow-up in 1 week	Maintenance of therapy schedule, medication compliance	Major Depressive Disorder, recurrent, moderate	Stable at discharge	Prescribed SSRIs and benzodiazepines.	Moderate	Moderate to High	Severity based on condition 'Depression'. Readmission probability based on discharge condition 'Stable at discharge' and history of Major Depressive Disorder.
Anthony Reyes	73	Male	9/20/2021	10/1/2021	Dr. Jenkins	Dr. Gordon	Severe Hypertension	Increase in anti-hypertensives, lifestyle modifications	Patient's blood pressure reduced and stabilized	Long-standing history of Hypertension	Managed with an increase in hypertension medication and lifestyle modifications	Cardiology follow-up in 2 weeks	Regular exercise, weight loss, low sodium diet, medication compliance	Extremely high blood pressure	Stable at discharge	Prescribed ACE inhibitors, Diuretics.	Moderate	Low	Severity based on condition 'Severe Hypertension'. Readmission probability based on discharge condition 'Stable at discharge' and long-standing history of Hypertension.
Olivia Ward	32	Female	10/21/2021	10/30/2021	Dr. Cole	Dr. Cook	Pregnancy with hypertension	Bed rest, blood pressure medications	Blood pressure controlled with no distress to fetus	No significant past medical history	Managed with bed rest and blood pressure medications, and regular monitoring of fetus	Obstetrics follow-up in 1 week	Bed rest, medication compliance, regular antenatal checks	Gestational Hypertension	Stable at discharge	Prescribed labetalol.	Moderate	Low	Severity based on condition 'Pregnancy with hypertension'. Readmission probability based on discharge condition 'Stable at discharge' and no significant past medical history.
William Howard	56	Male	6/12/2021	6/23/2021	Dr. Baylor	Dr. Black	Pneumonia	Antibiotics, respiratory therapy	Patient's condition improved significantly	History of COPD	Treated with IV antibiotics and oxygen therapy	Pulmonary follow-up in 1 month	Complete antibiotic course, smoking cessation advice	Final diagnosis of community-acquired pneumonia	Stable at discharge	Prescribed oral antibiotics and inhalers.	Moderate	Moderate	Severity based on condition 'Pneumonia'. Readmission probability based on discharge condition 'Stable at discharge' and history of COPD.
Ava Davis	43	Female	8/5/2021	8/15/2021	Dr. Craig	Dr. Houston	Asthma exacerbation	Bronchodilators, steroids, inhaler technique review	Improvement in asthma control	Long-standing asthma	Treated with bronchodilators and steroids, inhaler technique revised	Pulmonary follow-up in 2 weeks	Avoid triggers, use inhaler as instructed	Asthma exacerbation	Stable at discharge	Prescribed inhalers and oral steroids.	Low	Moderate	Severity based on condition 'Asthma exacerbation'. Readmission probability based on discharge condition 'Stable at discharge' and long-standing asthma.
Benjamin Turner	66	Male	7/14/2021	7/24/2021	Dr. Foster	Dr. Reed	Diabetic foot ulcer	Wound care, blood sugar control, antibiotics	Slow healing but progress with wound	History of type 2 diabetes, peripheral neuropathy	Managed with wound care, foot off-loading, and blood sugar control	Endocrinology follow-up in 1 month	Foot care, blood sugar control, follow up check	Diabetic foot ulcer	Stable at discharge	Prescribed insulin, oral hypoglycemic, topical and oral antibiotics.	Low	Moderate	Severity based on condition 'Diabetic foot ulcer'. Readmission probability based on discharge condition 'Stable at discharge' and history of type 2 diabetes, peripheral neuropathy.

Table A1. Cont.

Patient Name	Age	Gender	Date of Admission	Date of Discharge	Admitting Physician	Discharging Physician	Reason for Admission	Treatment and Surgical Procedures	Patient's Response to Treatment	Medical History	Hospital Course	Follow-Up	Patient Instructions	Final Diagnosis	Discharge Condition	Discharge Medications	Severity Level	Probability of Hospital Re-Admission	Reasoning
Charlotte Simmons	31	Female	3/22/2021	4/1/2021	Dr. Thompson	Dr. Johnson	Ectopic Pregnancy	Laparoscopic surgery	Safe recovery post-surgery	Prior ectopic pregnancy	Ectopic pregnancy removal via laparoscopic approach	Ob-Gyn follow-up in 2 weeks	Rest, avoid lifting heavy weights, medication compliance	Final diagnosis of ectopic pregnancy	Rapid recovery at discharge	Prescribed painkillers and oral contraceptives.	Low	Moderate	Severity based on condition 'Ectopic Pregnancy'. Readmission probability based on discharge condition 'Rapid recovery at discharge' and prior ectopic pregnancy.
Daniel Rodriguez	58	Male	6/15/2021	6/26/2021	Dr. Brooks	Dr. Davis	Coronary artery disease	Angioplasty and stent placement	Significant improvement post-procedure	History of smoking and hypertension	Procedure successful with no complications, smoking cessation advice given	Cardiology follow-up in 1 month	Smoking cessation, regular exercise, medication compliance	Final diagnosis of coronary artery disease	Stable at discharge	Prescribed antiplatelets, beta-blockers, ACE inhibitors, statins.	Low	Moderate	Severity based on condition 'Coronary artery disease'. Readmission probability based on discharge condition 'Stable at discharge' and history of smoking and hypertension.
Lily Morris	76	Female	7/30/2021	8/9/2021	Dr. Carter	Dr. Collins	Urinary tract infection	Antibiotics and hydration	Resolved with treatment	History of recurring UTIs	Treated with antibiotics, urinary culture guided treatment	Urology follow-up in 3 weeks	Hydration, wipe front to back, medication compliance	Final diagnosis of urinary tract infection	Resolved at discharge	Prescribed oral antibiotics.	Low	Moderate	Severity based on condition 'Urinary tract infection'. Readmission probability based on discharge condition 'Resolved at discharge' and history of recurring UTIs.
Noah Taylor	69	Male	6/23/2021	7/1/2021	Dr. Howard	Dr. Bennett	Pulmonary embolism	Anticoagulation therapy	Symptoms improved with treatment	Past history of deep vein thrombosis	IV anticoagulation followed by oral therapy to maintain INR	Hematology follow-up in 1 week	Avoid activities that can lead to falls, medication compliance	Final diagnosis of pulmonary embolism	Stable at discharge	Prescribed oral anticoagulants.	Low	Moderate	Severity based on condition 'Pulmonary embolism'. Readmission probability based on discharge condition 'Stable at discharge' and past history of deep vein thrombosis.
Zoe Parker	54	Female	8/24/2021	8/31/2021	Dr. Martin	Dr. Martinez	Crohn's disease flare	Steroids, infliximab infusions	Response to treatment with symptom resolution	Established Crohn's disease	Managed with IV corticosteroids and infliximab infusions	Gastroenterology follow-up in 2 weeks	Avoid triggers, medication compliance, hydrated	Crohn's disease acute flare	Stable at discharge	Prescribed oral steroids, infliximab infusion appointments.	Low	Moderate	Severity based on condition 'Crohn's disease flare'. Readmission probability based on discharge condition 'Stable at discharge' and established Crohn's disease.
Ethan Miller	61	Male	12/8/2021	12/18/2021	Dr. Adams	Dr. Barnes	Lung Cancer	Chemotherapy	Tolerating chemotherapy with manageable side effects	No significant past medical history	Patient initiated on chemotherapy regimen	Oncology follow-up in 1 week	Adequate hydration, medication compliance	Final diagnosis of lung cancer	Stable at discharge	Prescribed anti-emetics and pain management regimen.	Low	Low	Severity based on condition 'Lung Cancer'. Readmission probability based on discharge condition 'Stable at discharge' and no significant past medical history.

Table A1. Cont.

Patient Name	Age	Gender	Date of Admission	Date of Discharge	Admitting Physician	Discharging Physician	Reason for Admission	Treatment and Surgical Procedures	Patient's Response to Treatment	Medical History	Hospital Course	Follow-Up	Patient Instructions	Final Diagnosis	Discharge Condition	Discharge Medications	Severity Level	Probability of Hospital Re-Admission	Reasoning
Emily Roberts	67	Female	11/26/2021	12/10/2021	Dr. Jackson	Dr. Thompson	Acute renal failure	Dialysis	Renal function improved with dialysis, kidney function partially restored	Past history of hypertension and diabetes	Treated with intermittent hemodialysis and managed blood pressure and glucose	Nephrology follow-up in 1 week	Low sodium and potassium diet, medication compliance	Final diagnosis of acute renal failure	Improved at discharge	Prescribed antihypertensives, insulin, and dialysis prescription.	Low	Moderate	Severity based on condition 'Acute renal failure'. Readmission probability based on discharge condition 'Improved at discharge' and past history of hypertension and diabetes.
Joseph Garcia	80	Male	10/5/2021	10/15/2021	Dr. Phillips	Dr. Campbell	Chronic heart failure exacerbation	Diuretics, ACE inhibitors, Beta-blockers	Symptoms improved with medication adjustment	Long-standing heart failure, prior myocardial infarction	Managed with increase in diuretic dose, blood pressure control	Cardiology follow-up in 1 week	Low sodium diet, daily weight monitoring, medication compliance	Chronic heart failure exacerbation	Stable at discharge	Prescribed diuretics, ACE inhibitors, beta-blockers.	Low	Moderate	Severity based on condition 'Chronic heart failure exacerbation'. Readmission probability based on discharge condition 'Stable at discharge' and long-standing heart failure, prior myocardial infarction.
Mia Wong	28	Female	7/22/2021	7/31/2021	Dr. Evans	Dr. Rogers	Thyroiditis	Thyroid hormone replacement therapy	Thyroid hormone levels returned to normal	No significant medical history	Managed with thyroid hormone replacement therapy	Endocrinology follow-up in 1 month	Medication compliance	Final diagnosis of subacute thyroiditis	Stable at discharge	Levothyroxine.	Low	Moderate	Severity based on condition 'Thyroiditis'. Readmission probability based on discharge condition 'Stable at discharge' and no significant medical history.
Isaac Perry	46	Male	9/25/2021	10/1/2021	Dr. Ross	Dr. Griffin	Cellulitis	IV antibiotics followed by oral antibiotics	Infection resolved with treatment	No significant medical history	Patient treated with IV then oral antibiotics	Follow-up with primary care in 1 week	Complete antibiotic course, local wound care	Final diagnosis of cellulitis	Resolved at discharge	Oral antibiotics.	Low	Moderate	Severity based on condition 'Cellulitis'. Readmission probability based on discharge condition 'Resolved at discharge' and no significant medical history.
Sophia Lewis	75	Female	8/2/2021	8/11/2021	Dr. Kennedy	Dr. Dunn	Congestive heart failure exacerbation	Diuretics, dietary adjustments	Symptoms improved with treatment	History of coronary artery disease	Managed with medication optimization and dietary advice	Cardiology follow-up in 2 weeks	Low sodium diet, medication adherence	Congestive Heart Failure Exacerbation	Stable at discharge	Prescribed loop diuretics, ACE inhibitors, and beta blockers.	Low	Moderate	Severity based on condition 'Congestive heart failure exacerbation'. Readmission probability based on discharge condition 'Stable at discharge' and history of coronary artery disease.
Grace Foster	61	Female	4/23/2021	4/30/2021	Dr. Reed	Dr. Kline	Chronic Kidney Disease	Dialysis	Stable under dialysis treatment	History of diabetes and hypertension	Underwent dialysis and optimized blood pressure control	Nephrology follow-up in 1 week	Low sodium diet, medication compliance	Chronic Kidney Disease Stage 5	Stable at discharge	Prescribed antihypertensive, erythropoiesis-stimulating agents.	Low	Moderate	Severity based on condition 'Chronic Kidney Disease'. Readmission probability based on discharge condition 'Stable at discharge' and history of diabetes and hypertension.

Table A1. Cont.

Patient Name	Age	Gender	Date of Admission	Date of Discharge	Admitting Physician	Discharging Physician	Reason for Admission	Treatment and Surgical Procedures	Patient's Response to Treatment	Medical History	Hospital Course	Follow-Up	Patient Instructions	Final Diagnosis	Discharge Condition	Discharge Medications	Severity Level	Probability of Hospital Re-Admission	Reasoning
Noah Butler	65	Male	10/1/2021	10/12/2021	Dr. Wells	Dr. Perez	COPD Exacerbation	Corticosteroids, bronchodilators	Breathing improved noticeably	Long-standing COPD, ex-smoker	Managed with nebulized bronchodilators and systemic corticosteroids	Pulmonary follow-up in 1 month	Smoking cessation, use inhalers as instructed	Acute COPD exacerbation	Stable at discharge	Prescribed inhalers and a short course of oral steroids.	Low	Moderate	Severity based on condition 'COPD Exacerbation'. Readmission probability based on discharge condition 'Stable at discharge' and long-standing COPD, ex-smoker.
Eleanor Barnes	50	Female	9/10/2021	9/16/2021	Dr. Stevens	Dr. Rivera	Rheumatoid Arthritis Flare	Steroids and NSAIDs	Pain and swelling reduced significantly	Long-standing Rheumatoid Arthritis	Managed with increase in steroids and NSAIDs	Rheumatology follow-up in 2 weeks	Gentle exercise, joint care, medication compliance	Acute Rheumatoid Arthritis flare	Stable at discharge	Prescribed steroids and NSAIDs.	Low	Moderate	Severity based on condition 'Rheumatoid Arthritis Flare'. Readmission probability based on discharge condition 'Stable at discharge' and long-standing Rheumatoid Arthritis.
Lucas Peterson	78	Male	3/26/2021	4/1/2021	Dr. McDonald	Dr. Baker	Gouty Arthritis	Colchicine, Allopurinol	Gout attack settled, and uric acid lowered	History of recurrent Gout attacks	Managed with acute gout treatment and urate-lowering therapy	Follow-up with Rheumatologist in 2 weeks	Low purine diet, avoid alcohol, medication compliance	Final diagnosis of Gouty Arthritis	Stable at discharge	Prescribed colchicine and allopurinol.	Low	Moderate	Severity based on condition 'Gouty Arthritis'. Readmission probability based on discharge condition 'Stable at discharge' and history of recurrent Gout attacks.
Sophie Duncan	23	Female	5/30/2021	6/2/2021	Dr. Bryant	Dr. Coleman	Acute appendicitis	Laparoscopic appendectomy	Excellent recovery with no complications	Previously healthy	Successfully underwent laparoscopic appendectomy	General Surgery follow-up in 2 weeks	Care of operative site, resume regular activity as tolerated	Acute appendicitis	Stable at discharge	Analgesics, wound care recommendations.	Low	Moderate	Severity based on condition 'Acute appendicitis'. Readmission probability based on discharge condition 'Stable at discharge' and being previously healthy.
Samuel Larson	71	Male	12/15/2021	12/22/2021	Dr. Foster	Dr. Craig	Pneumonia	Antibiotics, respiratory support	Response to antibiotics with improved breathing	History of COPD	Received IV antibiotics and supplemental oxygen	Follow-up with Pulmonologist in 4 weeks	Take medications as prescribed, rest and adequate nutrition	Pneumonia	Stable at discharge	Oral antibiotics for completing course.	Moderate	Moderate	Severity based on condition 'Pneumonia'. Readmission probability based on discharge condition 'Stable at discharge' and history of COPD.
Sarah Woods	58	Female	11/11/2021	11/17/2021	Dr. Romero	Dr. Jacobs	Breast Cancer	Lumpectomy and sentinel lymph node biopsy	No complications with satisfactory recovery	No significant history	Procedure went without any complications, pathological report awaited	Follow-up with Oncologist in 1 week	Incision care, avoid physical exertion	Breast Cancer	Stable at discharge	Pain management medications.	Low	Moderate	Severity based on condition 'Breast Cancer'. Readmission probability based on discharge condition 'Stable at discharge' and no significant past medical history.

Table A1. Cont.

Patient Name	Age	Gender	Date of Admission	Date of Discharge	Admitting Physician	Discharging Physician	Reason for Admission	Treatment and Surgical Procedures	Patient's Response to Treatment	Medical History	Hospital Course	Follow-Up	Patient Instructions	Final Diagnosis	Discharge Condition	Discharge Medications	Severity Level	Probability of Hospital Re-Admission	Reasoning
Jack Hudson	46	Male	10/24/2021	10/31/2021	Dr. Paul	Dr. Baker	Gastric ulcers	Proton pump inhibitors, dietary modifications	Symptoms improved significantly with treatment	History of NSAID use	Managed with PPI therapy and dietary advice	Gastroenterology follow-up in 1 month	Avoid spicy food, alcohol, smoking, medication adherence	Gastric ulcer	Stable at discharge	Omeprazole, Sucralfate.	Low	Moderate	Severity based on condition 'Gastric ulcers'. Readmission probability based on discharge condition 'Stable at discharge' and history of NSAID use.
Ivy Johnson	80	Female	9/3/2021	9/15/2021	Dr. Jackson	Dr. Riley	Stroke rehabilitation	Physical and occupational therapy	Gradual improvement with still residual weakness	Past history of hypertension and diabetes	Underwent intensive rehabilitation therapy	Follow-up with Outpatient Rehab and Neurologist in 4 weeks	Physiotherapy, medication compliance	Stroke with right hemiparesis	Stable at discharge	Antihypertensives, oral antidiabetics, aspirin.	Low	Moderate	Severity based on condition 'Stroke rehabilitation'. Readmission probability based on discharge condition 'Stable at discharge' and history of hypertension and diabetes.
Elijah Myers	55	Male	11/3/2021	11/10/2021	Dr. Ayers	Dr. Harlow	Pancreatitis	IV fluids, pain management, and dietary adjustments	Symptoms improved significantly	History of alcohol abuse	Managed with IV fluids, pain management, and alcohol detox	Gastroenterology and Addiction specialist follow-up in 1 week	Total abstinence from alcohol, low-fat diet, medication compliance	Alcohol-induced pancreatitis	Improved at discharge	Prescribed pain killers, pancreatic enzymes, and detox medications.	Low	Moderate	Severity based on condition 'Pancreatitis'. Readmission probability based on discharge condition 'Improved at discharge' and history of alcohol abuse.
Hannah Peters	36	Female	10/11/2021	10/20/2021	Dr. Madison	Dr. Turner	Uncontrolled Type 1 Diabetes	Insulin regulation, diet and lifestyle changes	Blood sugar levels returned to normal	Long-standing diabetes	Management involved adjustment of insulin dose and dietary advice	Endocrinology follow-up in 1 week	Regular monitoring, maintain balanced diet, regular exercise	Uncontrolled Type 1 Diabetes	Stable at discharge	Insulin as per optimized prescription.	Moderate	Moderate	Severity based on condition 'Uncontrolled Type 1 Diabetes'. Readmission probability based on discharge condition 'Stable at discharge' and long-standing diabetes.
William Riley	72	Male	7/22/2021	7/29/2021	Dr. Howard	Dr. Jenkins	Chronic Obstructive Pulmonary disease exacerbation	Oxygen therapy, steroids, and antibiotics	Breathing normalized, chest clearing	History of smoking and COPD	Managed with nebulizers, steroids, and antibiotics	Pulmonary follow-up in 2 weeks	Smoking cessation, use inhalers as instructed	COPD exacerbation	Stable at discharge	Inhalers, steroids, and antibiotics.	Low	Moderate	Severity based on condition 'Chronic Obstructive Pulmonary disease exacerbation'. Readmission probability based on discharge condition 'Stable at discharge' and history of smoking and COPD.
Lucy Foster	46	Female	9/15/2021	9/30/2021	Dr. Reese	Dr. Castillo	Breast Cancer	Chemotherapy	Moderate side effects managed	No significant family history	Commencement of chemotherapy regimen	Oncology follow-up in 1 week	Healthy diet, gentle exercise, medication compliance	Breast Cancer, stage IIb	Stable at discharge	Prescribed antiemetic and analgesic.	Low	Moderate	Severity based on condition 'Breast Cancer'. Readmission probability based on discharge condition 'Stable at discharge' and no significant family history.

Table A1. Cont.

Patient Name	Age	Gender	Date of Admission	Date of Discharge	Admitting Physician	Discharging Physician	Reason for Admission	Treatment and Surgical Procedures	Patient's Response to Treatment	Medical History	Hospital Course	Follow-Up	Patient Instructions	Final Diagnosis	Discharge Condition	Discharge Medications	Severity Level	Probability of Hospital Re-Admission	Reasoning
Oliver Shaw	35	Male	4/27/2021	5/2/2021	Dr. Piper	Dr. Shaw	Fracture tibia	Open reduction and internal fixation	Recovery as expected, mobilizing with support	No significant medical history	Smooth surgery, recovery in ward until independent mobilization achieved	Orthopedic follow-up in 1 week	Weight-bearing as per advice, rest, elevate limb	Tibia fracture	Stable at discharge	Analgesics, anticoagulant.	Low	Moderate	Severity based on condition 'Fracture tibia'. Readmission probability based on discharge condition 'Stable at discharge' and no significant medical history.
Stella Rogers	55	Female	8/21/2021	8/31/2021	Dr. Sparks	Dr. Kennedy	Vasculitis	Steroids and immunosuppressants	Symptoms improved significantly	No significant medical history	Managed with steroids and immunosuppressants	Rheumatology follow-up in 2 weeks	Medication compliance, regular follow ups, report any new symptoms	Vasculitis	Stable at discharge	Corticosteroids, immunosuppressants.	Low	Moderate	Severity based on condition 'Vasculitis'. Readmission probability based on discharge condition 'Stable at discharge' and no significant medical history.
Liam Griffin	81	Male	7/25/2021	8/8/2021	Dr. Patterson	Dr. Phillips	Pneumonia	Antibiotics and supportive care	Condition improved significantly	History of diabetes, hypertension	Treated with IV antibiotics and oxygen therapy	Pulmonology follow-up in 3 weeks	Medication compliance, smoking cessation	Pneumonia	Stable at discharge	Oral antibiotics to complete course.	Moderate	Moderate	Severity based on condition 'Pneumonia'. Readmission probability based on discharge condition 'Stable at discharge' and history of diabetes, hypertension.
Hazel Ortiz	45	Female	11/10/2021	11/18/2021	Dr. Snyder	Dr. Hamilton	Severe Anemia	Blood transfusion, iron supplements	Blood levels normalized	History of heavy menstrual bleeding	Fluid resuscitation and blood transfusions were given	Gynecology follow-up in 1 week	Oral iron supplements, balanced diet	Severe Iron Deficiency Anemia	Stable at discharge	Iron supplement, analgesic.	Low	Moderate	Severity based on condition 'Severe Anemia'. Readmission probability based on discharge condition 'Stable at discharge' and history of heavy menstrual bleeding.
Levi Cooper	63	Male	1/15/2021	1/20/2021	Dr. Bowman	Dr. Francis	Gastrointestinal bleeding	Endoscopy, Clipping of bleeding ulcer	Bleeding stopped, stable condition	History of chronic NSAID use	Endoscopic intervention was successful without complications	Gastroenterology follow-up in 1 week	Avoid NSAIDs and alcohol, medication compliance	Peptic Ulcer Disease with bleeding	Stable at discharge	Proton pump inhibitors.	Low	Moderate	Severity based on condition 'Gastrointestinal bleeding'. Readmission probability based on discharge condition 'Stable at discharge' and history of chronic NSAID use.
Lily Rogers	78	Female	6/8/2021	6/15/2021	Dr. Dean	Dr. Foster	Chronic Kidney Disease progression	Dialysis initiation	Stable after starting dialysis	History of diabetes, Chronic Kidney Disease	Initiated on dialysis	Nephrology follow-up in 1 week	Medication compliance, appropriate diet	End-Stage Renal Disease	Stable at discharge	Antihypertensives, erythropoiesis-stimulating agents, phosphate binders.	Low	Moderate	Severity based on condition 'Chronic Kidney Disease progression'. Readmission probability based on discharge condition 'Stable at discharge' and history of diabetes, Chronic Kidney Disease.

Table A1. Cont.

Patient Name	Age	Gender	Date of Admission	Date of Discharge	Admitting Physician	Discharging Physician	Reason for Admission	Treatment and Surgical Procedures	Patient's Response to Treatment	Medical History	Hospital Course	Follow-Up	Patient Instructions	Final Diagnosis	Discharge Condition	Discharge Medications	Severity Level	Probability of Hospital Re-Admission	Reasoning
Noah Barnes	48	Male	8/1/2021	8/8/2021	Dr. Ramirez	Dr. Hughes	Bell's Palsy	Corticosteroids, Physical therapy	Slow return of facial movement	No significant medical history	Managed with corticosteroids and physical therapy	Neurology follow-up in 1 month	Facial muscle exercises, medication compliance	Bell's Palsy	Improving at discharge	Prescribed corticosteroids, antivirals.	Low	Moderate	Severity based on condition 'Bell's Palsy'. Readmission probability based on discharge condition 'Improving at discharge' and no significant medical history.
Emily Foster	34	Female	1/22/2021	1/27/2021	Dr. Adams	Dr. Barnes	Appendicitis	Appendectomy	Excellent recovery post-surgery	No significant medical history	Underwent routine open appendectomy	Follow-up with surgeon in 2 weeks	Wound care, report any fever or wound discharge	Acute appendicitis	Stable at discharge	Prescribed painkillers and absorption.	Low	Moderate	Severity based on condition 'Appendicitis'. Readmission probability based on discharge condition 'Stable at discharge' and no significant medical history.
Ethan Johnson	45	Male	3/21/2021	4/5/2021	Dr. Roberts	Dr. Edwards	Colon cancer	Resection of colon cancer, start of adjuvant chemotherapy	Disease under control, tolerated chemo well	No significant past medical history	Complete tumor resection achieved with histology confirming margins	Oncologist follow-up in 2 weeks	Healthy diet, regular exercise, medication compliance	Colon cancer stage III	Stable at discharge	Prescribed chemotherapeutics, antiemetics.	Low	Low	Severity based on condition 'Colon cancer'. Readmission probability based on discharge condition 'Stable at discharge' and no significant past medical history.
Sophia James	24	Female	11/5/2021	11/15/2021	Dr. Jacobs	Dr. Willis	Severe Asthma Attack	Intravenous corticosteroids, nebulizer treatments	Breathing eased, symptoms improved	Lifetime Asthma	Hospitalized for acute asthma management	Pulmonary follow-up in 1 week	Avoid asthma triggers, regular use of control medication	Acute severe asthma attack, Asthma	Stable at discharge	Inhalers, oral corticosteroids for a short course.	Low	Moderate	Severity based on condition 'Severe Asthma Attack'. Readmission probability based on discharge condition 'Stable at discharge' and lifetime asthma.
Jacob Owens	58	Male	10/7/2021	10/14/2021	Dr. Griffin	Dr. Patterson	Peptic ulcer disease	Proton pump inhibitors, H. pylori eradication	Symptoms improved significantly	No significant past medical history	Received treatment for H. pylori and proton pump inhibitors	Gastroenterology follow-up in 1 month	Avoid NSAIDs, alcohol, spicy foods; take medications with meals	Peptic ulcer disease	Stable at discharge	Prescribed proton-pump inhibitors.	Low	Low	Severity based on condition 'Peptic ulcer disease'. Readmission probability based on discharge condition 'Stable at discharge' and no significant past medical history.
Layla Tyler	63	Female	8/16/2021	8/28/2021	Dr. Ellis	Dr. Foster	Congestive Heart Failure	Diuretics, vasodilators, beta-blockers	Symptoms improved with stabilization	Hypertension	Adjusted medication regimen; patient education about fluid intake and weight monitoring	Cardiology follow-up in 4 weeks	Medication compliance, daily weight, low sodium diet	Congestive Heart Failure	Stable at discharge	Prescribed diuretics, vasodilators, beta-blockers.	Low	Moderate	Severity based on condition 'Congestive Heart Failure'. Readmission probability based on discharge condition 'Stable at discharge' and hypertension.

Table A1. Cont.

Patient Name	Age	Gender	Date of Admission	Date of Discharge	Admitting Physician	Discharging Physician	Reason for Admission	Treatment and Surgical Procedures	Patient's Response to Treatment	Medical History	Hospital Course	Follow-Up	Patient Instructions	Final Diagnosis	Discharge Condition	Discharge Medications	Severity Level	Probability of Hospital Re-Admission	Reasoning
Max Peters	46	Male	3/12/2021	3/18/2021	Dr. King	Dr. Howard	Pneumothorax	Chest tube insertion	Chest re-expanded successfully	No significant past medical history	Underwent chest tube insertion for pneumothorax	Pulmonary follow-up in 2 weeks	Avoid heavy lifting, short flights for 2 weeks	Spontaneous Pneumothorax	Stable at discharge	Analgesics, follow up as directed.	Low	Low	Severity based on condition 'Pneumothorax'. Readmission probability based on discharge condition 'Stable at discharge' and no significant past medical history.
Harper Davis	71	Female	5/30/2021	6/6/2021	Dr. Ross	Dr. Holland	COPD Exacerbation	Bronchodilators, steroids	Breathing improved noticeably	COPD, ex-smoker	Managed with nebulized bronchodilators and oral steroids	Pulmonary follow-up in 2 weeks	Smoking cessation, use inhalers as instructed	COPD exacerbation	Stable at discharge	Inhalers, oral steroid taper.	Low	Moderate	Severity based on condition 'COPD Exacerbation'. Readmission probability based on discharge condition 'Stable at discharge' and COPD, ex-smoker history.
Thomas Mitchell	79	Male	5/10/2021	5/21/2021	Dr. Barrett	Dr. Osborne	Heart failure	Diuretics, beta-blockers, ACE inhibitors	Condition improved significantly with management	History of ischemic heart disease, hypertension	Managed with heart failure medications, fluid restriction	Cardiology follow-up in 2 weeks	Low salt diet, fluid restriction, medication compliance	Congestive heart failure	Stable at discharge	Furosemide, lisinopril, carvedilol.	Low	Moderate	Severity based on condition 'Heart failure'. Readmission probability based on discharge condition 'Stable at discharge' and history of ischemic heart disease, hypertension.
Emily Ross	43	Female	2/14/2021	2/21/2021	Dr. Hamilton	Dr. Jenkins	Cholecystitis	Cholecystectomy	Recovery without complications	No significant past medical history	Underwent laparoscopic cholecystectomy	Surgery follow-up in 2 weeks	Gradual increase in diet, wound care	Cholecystitis	Stable at discharge	Analgesics, wound care recommendations.	Low	Low	Severity based on condition 'Cholecystitis'. Readmission probability based on discharge condition 'Stable at discharge' and no significant past medical history.
Oliver Hall	27	Male	7/18/2021	7/22/2021	Dr. Washington	Dr. Murray	Meningitis	Antibiotics, steroids	Symptoms resolved notably	No significant past medical history	Managed with IV antibiotics and supportive care	Neurology follow-up in 2 weeks	Rest, hydration, antibiotic compliance	Meningitis	Stable at discharge	Continuation of oral antibiotics and analgesics.	Low	Low	Severity based on condition 'Meningitis'. Readmission probability based on discharge condition 'Stable at discharge' and no significant past medical history.
Abigail Jackson	65	Female	12/1/2021	12/10/2021	Dr. Jenkins	Dr. Thompson	Stroke	Thrombolytic therapy, rehabilitation	Partial resolution of deficits	Hypertension, diabetes	Underwent IV thrombolysis and rehabilitation	Neurology and rehabilitation follow-up in 1 month	Physiotherapy, medication compliance, lifestyle modifications	Ischemic stroke	Moderate impairment at discharge	Antihypertensives, antidiabetics, anticoagulation.	Low	Moderate	Severity based on condition 'Stroke'. Readmission probability based on moderate impairment at discharge and history of hypertension, diabetes.

Table A1. Cont.

Patient Name	Age	Gender	Date of Admission	Date of Discharge	Admitting Physician	Discharging Physician	Reason for Admission	Treatment and Surgical Procedures	Patient's Response to Treatment	Medical History	Hospital Course	Follow-Up	Patient Instructions	Final Diagnosis	Discharge Condition	Discharge Medications	Severity Level	Probability of Hospital Re-Admission	Reasoning
Jackson Perez	54	Male	6/30/2021	7/6/2021	Dr. Adams	Dr. Collins	Peptic Ulcer Disease	Proton pump inhibitors, H. pylori eradication	Symptoms markedly improved	Past history of smoking, alcohol use	Managed with proton pump inhibitors and H. pylori eradication therapy	Gastroenterology follow-up in 4 weeks	Medication compliance, lifestyle modification, stop alcohol and smoking	Peptic Ulcer Disease	Stable at discharge	Antibiotics for H.pylori, PPIs.	Low	Moderate	Severity based on condition 'Peptic Ulcer Disease'. Readmission probability based on stable condition at discharge and past history of smoking, alcohol use.
Sophia Kline	31	Female	2/2/2021	2/7/2021	Dr. Bailey	Dr. Bell	Pyelonephritis	IV antibiotics followed by oral antibiotics therapy	Symptoms resolved significantly	No significant past medical history	Managed with IV antibiotics followed by switch to oral	Primary care follow-up in 2 weeks	Hydration, avoid delaying urination, antibiotic compliance	Pyelonephritis	Stable at discharge	Oral antibiotics to complete 14 days course.	Low	Low	Severity based on condition 'Pyelonephritis'. Readmission probability based on stable condition at discharge and no significant past medical history.
Grayson Walker	32	Male	3/18/2021	3/25/2021	Dr. Rodriguez	Dr. Webb	Appendicitis	Appendectomy	Excellent recovery with no complications	No significant medical history	Underwent appendectomy without complications	Surgery follow-up in 2 weeks	Resume normal diet gradually, wound care, report any fever	Appendicitis	Stable at discharge	Analgesics.	Low	Moderate	Severity based on condition 'Appendicitis'. Readmission probability based on stable condition at discharge and no significant medical history.
Aria Harper	73	Female	11/12/2021	11/30/2021	Dr. Snyder	Dr. Walsh	Heart failure	Diuretics, lifestyle modification	Symptoms improved notably	History of Hypertension, Diabetes	Managed with diuretics and lifestyle modification advice	Cardiology follow-up in 1 month	Weight monitoring, low salt diet, exercise, medication compliance	Congestive Heart Failure	Stable at discharge	Prescribed diuretics, ACE inhibitors, and beta-blockers	Low	Moderate	Severity based on condition 'Heart failure'. Readmission probability based on stable condition at discharge and history of Hypertension, Diabetes.

References

1. Gilbert, A.; Marciniak, M.; Rodero, C.; Lamata, P.; Samset, E.; Mcleod, K. Generating Synthetic Labeled Data from Existing Anatomical Models: An Example with Echocardiography Segmentation. *IEEE Trans. Med. Imaging* **2021**, *40*, 2783–2794. [[CrossRef](#)] [[PubMed](#)]
2. Aouedi, O.; Sacco, A.; Piamrat, K.; Marchetto, G. Handling Privacy-Sensitive Medical Data With Federated Learning: Challenges and Future Directions. *IEEE J. Biomed. Health Inform.* **2022**, *27*, 790–803. [[CrossRef](#)]
3. Elbadawi, M.; Li, H.; Basit, A.W.; Gaisford, S. The role of artificial intelligence in generating original scientific research. *Int. J. Pharm.* **2024**, *652*, 123741. [[CrossRef](#)] [[PubMed](#)]
4. Van Nooten, J.; Daelemans, W. Improving Dutch Vaccine Hesitancy Monitoring via Multi-Label Data Augmentation with GPT-3.5. In Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis, Toronto, ON, Canada, 14 July 2023. Available online: <https://openai.com/blog/chatgpt> (accessed on 21 April 2024).
5. Zhou, S.; Zhang, Y. DATLMedQA: A data augmentation and transfer learning based solution for medical question answering. *Appl. Sci.* **2021**, *11*, 11251. [[CrossRef](#)]
6. Hämmäläinen, P.; Tavast, M.; Kunnari, A. Evaluating Large Language Models in Generating Synthetic HCI Research Data: A Case Study. In Proceedings of the Conference on Human Factors in Computing Systems, Hamburg, Germany, 23–28 April 2023; Association for Computing Machinery: New York, NY, USA, 2023. [[CrossRef](#)]
7. Lu, Q.; Dou, D.; Nguyen, T.H. Textual Data Augmentation for Patient Outcomes Prediction. In Proceedings of the 2021 IEEE international conference on bioinformatics and biomedicine (BIBM), Houston, TX, USA, 9–12 December 2021. [[CrossRef](#)]
8. Bird, J.J.; Pritchard, M.G.; Fratini, A.; Ekart, A.; Faria, D.R. Synthetic Biological Signals Machine-Generated by GPT-2 Improve the Classification of EEG and EMG through Data Augmentation. *IEEE Robot. Autom. Lett.* **2021**, *6*, 3498–3504. [[CrossRef](#)]
9. Amin-Nejad, A.; Ive, J.; Velupillai, S. Exploring Transformer Text Generation for Medical Dataset Augmentation. In Proceedings of the Twelfth Language Resources and Evaluation Conference, Marseille, France, 11–16 May 2020. Available online: <https://github.com/tensorflow/tensor2tensor> (accessed on 21 April 2024).
10. Thamsen, B.; Yevtushenko, P.; Gundelwein, L.; Setio, A.A.A.; Lamecker, H.; Kelm, M.; Schafstedde, M.; Heimann, T.; Kuehne, T.; Goubergrits, L. Synthetic Database of Aortic Morphometry and Hemodynamics: Overcoming Medical Imaging Data Availability. *IEEE Trans. Med. Imaging* **2021**, *40*, 1438–1449. [[CrossRef](#)] [[PubMed](#)]
11. Ruksakulpiwat, S.; Kumar, A.; Ajibade, A. Using ChatGPT in Medical Research: Current Status and Future Directions. *J. Multidiscip. Health* **2023**, *16*, 1513–1520. [[CrossRef](#)]
12. Mahuli, S.A.; Rai, A.; Mahuli, A.V.; Kumar, A. Application ChatGPT in conducting systematic reviews and meta-analyses. *Br. Dent. J.* **2023**, *235*, 90–92. [[CrossRef](#)] [[PubMed](#)]
13. Cai, X.; Geng, Y.; Du, Y.; Westerman, B.; Wang, D.; Ma, C.; Vallejo, J.J.G. Utilizing ChatGPT to select literature for meta-analysis shows workload reduction while maintaining a similar recall level as manual curation. *medRxiv* **2023**. [[CrossRef](#)]
14. De Angelis, L.; Baglivo, F.; Arzilli, G.; Privitera, G.P.; Ferragina, P.; Tozzi, A.E.; Rizzo, C. ChatGPT and the rise of large language models: The new AI-driven infodemic threat in public health. *Front. Public Health* **2023**, *11*, 1166120. [[CrossRef](#)]
15. Reddy, S. Evaluating large language models for use in healthcare: A framework for translational value assessment. *Inform. Med. Unlocked* **2023**, *41*, 101304. [[CrossRef](#)]
16. Alberts, I.L.; Mercolli, L.; Pyka, T.; Prenosil, G.; Shi, K.; Rominger, A.; Afshar-Oromieh, A. Large language models (LLM) and ChatGPT: What will the impact on nuclear medicine be? *Eur. J. Nucl. Med.* **2023**, *50*, 1549–1552. [[CrossRef](#)] [[PubMed](#)]
17. Chatterjee, S.; Bhattacharya, M.; Pal, S.; Lee, S.; Chakraborty, C. ChatGPT and large language models in orthopedics: From education and surgery to research. *J. Exp. Orthop.* **2023**, *10*, 1–10. [[CrossRef](#)] [[PubMed](#)]
18. Sallam, M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare* **2023**, *11*, 887. [[CrossRef](#)]
19. Lim, S.; Schmälzle, R. Artificial intelligence for health message generation: An empirical study using a large language model (LLM) and prompt engineering. *Front. Commun.* **2023**, *8*, 1129082. [[CrossRef](#)]
20. Waisberg, E.; Ong, J.; Kamran, S.A.; Masalkhi, M.; Zaman, N.; Sarker, P.; Lee, A.G.; Tavakkoli, A. Bridging artificial intelligence in medicine with generative pre-trained transformer (GPT) technology. *J. Med. Artif. Intell.* **2023**, *6*, 13. [[CrossRef](#)]
21. Maddigan, P.; Susnjak, T. Chat2VIS: Generating Data Visualizations via Natural Language Using ChatGPT, Codex and GPT-3 Large Language Models. *IEEE Access* **2023**, *11*, 45181–45193. [[CrossRef](#)]
22. Lengerich, B.J.; Bordt, S.; Nori, H.; Nunnally, M.E.; Aphinyanaphongs, Y.; Kellis, M.; Caruana, R. LLMs Understand Glass-Box Models, Discover Surprises, and Suggest Repairs. *arXiv* **2023**, arXiv:2308.01157.
23. Sharma, A.; Devalia, D.; Almeida, W.; Patil, H.; Mishra, A. Statistical Data Analysis using GPT3: An Overview. In Proceedings of the 2022 IEEE Bombay Section Signature Conference (IBSSC), Mumbai, India, 8–10 December 2022; IEEE: Piscataway, NJ, USA, 2022.
24. Espejel, J.L.; Ettifouri, E.H.; Alassan, M.S.Y.; Chouham, E.M.; Dahhane, W. GPT-3.5, GPT-4, or BARD? Evaluating LLMs reasoning ability in zero-shot setting and performance boosting through prompts. *Nat. Lang. Process. J.* **2023**, *5*, 100032. [[CrossRef](#)]
25. de Kok, T. Generative LLMs and Textual Analysis in Accounting: (Chat)GPT as Research Assistant? 2023. Available online: <https://ssrn.com/abstract=4429658> (accessed on 21 April 2024).

26. Yenduri, G.; Srivastava, G.; Maddikunta, P.K.R.; Jhaveri, R.H.; Wang, W.; Vasilakos, A.V.; Gadekallu, T.R. Generative Pre-trained Transformer: A Comprehensive Review on Enabling Technologies, Potential Applications, Emerging Challenges, and Future Directions. *arXiv* **2023**, arXiv:2305.10435. [[CrossRef](#)]
27. Sufi, F.K.; Alsulami, M.; Gutub, A. Automating Global Threat-Maps Generation via Advancements of News Sensors and AI. *Arab. J. Sci. Eng.* **2022**, *48*, 2455–2472. [[CrossRef](#)]
28. Sufi, F. Social Media Analytics on Russia–Ukraine Cyber War with Natural Language Processing: Perspectives and Challenges. *Information* **2023**, *14*, 485. [[CrossRef](#)]
29. Sufi, F.K.; Razzak, I.; Khalil, I. Tracking Anti-Vax Social Movement Using AI-Based Social Media Monitoring. *IEEE Trans. Technol. Soc.* **2022**, *3*, 290–299. [[CrossRef](#)]
30. Sufi, F.K.; Khalil, I. Automated Disaster Monitoring From Social Media Posts Using AI-Based Location Intelligence and Sentiment Analysis. *IEEE Trans. Comput. Soc. Syst.* **2022**. [[CrossRef](#)]
31. Sufi, F.K. AI-SocialDisaster: An AI-based software for identifying and analyzing natural disasters from social media. *Softw. Impacts* **2022**, *13*, 100319. [[CrossRef](#)]
32. Sufi, F.K. A decision support system for extracting artificial intelligence-driven insights from live twitter feeds on natural disasters. *Decis. Anal. J.* **2022**, *5*, 100130. [[CrossRef](#)]
33. Sufi, F.K.; Alsulami, M. Automated Multidimensional Analysis of Global Events with Entity Detection, Sentiment Analysis and Anomaly Detection. *IEEE Access* **2021**, *9*, 152449–152460. [[CrossRef](#)]
34. Sufi, F. Algorithms in Low-Code-No-Code for Research Applications: A Practical Review. *Algorithms* **2023**, *16*, 108. [[CrossRef](#)]
35. Balaji, S.; Magar, R.; Jadhav, Y.; Farimani, A.B. GPT-MolBERTa: GPT Molecular Features Language Model for molecular property prediction. *arXiv* **2023**, arXiv:2310.03030.
36. Hu, Y.; Mai, G.; Cundy, C.; Choi, K.; Lao, N.; Liu, W.; Lakhanpal, G.; Zhou, R.Z.; Joseph, K. Geo-knowledge-guided GPT models improve the extraction of location descriptions from disaster-related social media messages. *Int. J. Geogr. Inf. Sci.* **2023**, *37*, 2289–2318. [[CrossRef](#)]
37. Maimaiti, M.; Liu, Y.; Luan, H.; Sun, M. Data augmentation for low-resource languages NMT guided by constrained sampling. *Int. J. Intell. Syst.* **2021**, *37*, 30–51. [[CrossRef](#)]
38. Suhaeni, C.; Yong, H.-S. Mitigating Class Imbalance in Sentiment Analysis through GPT-3-Generated Synthetic Sentences. *Appl. Sci.* **2023**, *13*, 9766. [[CrossRef](#)]
39. Romero-Sandoval, M.; Calderón-Ramírez, S.; Solís, M. Using GPT-3 as a Text Data Augmentator for a Complex Text Detector. In Proceedings of the 2023 IEEE 5th International Conference on BioInspired Processing (BIP), San Carlos, Alajuela, Costa Rica, 28–30 November 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–6. [[CrossRef](#)]
40. Cohen, S.; Presil, D.; Katz, O.; Arbili, O.; Messica, S.; Rokach, L. Enhancing social network hate detection using back translation and GPT-3 augmentations during training and test-time. *Inf. Fusion* **2023**, *99*, 101887. [[CrossRef](#)]
41. Rebboud, Y.; Lisena, P.; Troncy, R. Prompt-based Data Augmentation for Semantically-Precise Event Relation Classification. In Proceedings of the 2023 IEEE 5th International Conference on BioInspired Processing (BIP), San Carlos, Alajuela, Costa Rica, 28–30 November 2023. Available online: <http://ceur-ws.org> (accessed on 21 April 2024).
42. Grasler, I.; Preus, D.; Brandt, L.; Mohr, M. Efficient Extraction of Technical Requirements Applying Data Augmentation. In Proceedings of the ISSE 2022–2022 8th IEEE International Symposium on Systems Engineering, Vienna, Austria, 24–26 October 2022; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2022. [[CrossRef](#)]
43. Singh, C.; Askari, A.; Caruana, R.; Gao, J. Augmenting interpretable models with large language models during training. *Nat. Commun.* **2023**, *14*, 7913. [[CrossRef](#)]
44. Modzelewski, A.; Sosnowski, W.; Wilczynska, M.; Wierzbicki, A. DSHacker at SemEval-2023 Task 3: Genres and Persuasion Techniques Detection with Multilingual Data Augmentation through Machine Translation and Text Generation. In Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023), Toronto, ON, Canada, 13–14 July 2023. Available online: <https://semeval.github.io/SemEval2023/> (accessed on 21 April 2024).
45. Hong, X.-S.; Wu, S.-H.; Tian, M.; Jiang, J. CYUT at the NTCIR-16 FinNum-3 Task: Data Resampling and Data Augmentation by Generation. In Proceedings of the 16th NTCIR Conference on Evaluation of Information Access Technologies, Tokyo, Japan, 14–17 June 2022. Available online: <https://huggingface.co/docs/transformers/main> (accessed on 21 April 2024).
46. Khatri, S.; Iqbal, M.; Ubakanma, G.; van der Vliet-Firth, S. SkillBot: Towards Data Augmentation using Transformer language model and linguistic evaluation. In Proceedings of the 2022 International Conference on Human-Centered Cognitive Systems, HCCS 2022, Shanghai, China, 17–18 December 2022; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2022. [[CrossRef](#)]
47. Vogel, L.; Flek, L. Investigating Paraphrasing-Based Data Augmentation for Task-Oriented Dialogue Systems. In *Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer Science and Business Media Deutschland GmbH: Berlin/Heidelberg, Germany, 2022; pp. 476–488. [[CrossRef](#)]
48. Casula, C.; Tonelli, S.; Kessler, F.B. Generation-Based Data Augmentation for Offensive Language Detection: Is It Worth It? In Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics, Dubrovnik, Croatia, 2–6 May 2023. Available online: <https://github.com/dhfbk/annotators-agreement-dataset> (accessed on 21 April 2024).

49. Pouran, A.; Veyseh, B.; Dernoncourt, F.; Min, B.; Nguyen, T.H. Generating Complement Data for Aspect Term Extraction with GPT-2. In Proceedings of the Third Workshop on Deep Learning for Low-Resource Natural Language Processing, Virtual, 14 July 2022.
50. D'Sa, A.G.; Illina, I.; Fohr, D.; Klakow, D.; Ruiter, D. Exploring Conditional Language Model Based Data Augmentation Approaches for Hate Speech Classification. In *Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer Science and Business Media Deutschland GmbH: Berlin/Heidelberg, Germany, 2021; pp. 135–146. [[CrossRef](#)]
51. Meyer, S.; Elswiler, D.; Ludwig, B.; Fernandez-Pichel, M.; Losada, D.E. Do We Still Need Human Assessors' Prompt-Based GPT-3 User Simulation in Conversational AI. In Proceedings of the 4th Conference on Conversational User Interfaces, Glasgow, UK, 26–28 July 2022; ACM International Conference Proceeding Series; Association for Computing Machinery: New York, NY, USA, 2022. [[CrossRef](#)]
52. Queiroz Abonizio, H.; Barbon Junior, S. Pre-trained Data Augmentation for Text Classification. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer Science and Business Media Deutschland GmbH: Berlin/Heidelberg, Germany, 2020; pp. 551–565. [[CrossRef](#)]
53. Tapia-Téllez, J.M.; Escalante, H.J. Data Augmentation with Transformers for Text Classification. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer Science and Business Media Deutschland GmbH: Berlin/Heidelberg, Germany, 2020; pp. 247–259. [[CrossRef](#)]
54. Hassani, H.; Silva, E.S. The Role of ChatGPT in Data Science: How AI-Assisted Conversational Interfaces Are Revolutionizing the Field. *Big Data Cogn. Comput.* **2023**, *7*, 62. [[CrossRef](#)]
55. Nouri, N. Data Augmentation with Dual Training for Offensive Span Detection. In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Seattle, DC, USA, July 2022.
56. Bayer, M.; Kaufhold, M.-A.; Buchhold, B.; Keller, M.; Dallmeyer, J.; Reuter, C. Data augmentation in natural language processing: A novel text generation approach for long and short text classifiers. *Int. J. Mach. Learn. Cybern.* **2022**, *14*, 135–150. [[CrossRef](#)] [[PubMed](#)]
57. Anaby-Tavor, A.; Carmeli, B.; Goldbraich, E.; Kantor, A.; Kour, G.; Shlomov, S.; Tepper, N.; Zwerdling, N. Do Not Have Enough Data? Deep Learning to the Rescue! *Proc. AAAI Conf. Artif. Intell.* **2020**, *34*, 7383–7390. [[CrossRef](#)]
58. Quteineh, H.; Samothrakis, S.; Sutcliffe, R. Textual Data Augmentation for Efficient Active Learning on Tiny Datasets. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Online, 16–20 November 2020. Available online: <https://www.snorkel.org/> (accessed on 21 April 2024).
59. Veyseh, A.P.B.; Van Nguyen, M.; Min, B.; Nguyen, T.H. Augmenting Open-Domain Event Detection with Synthetic Data from GPT-2. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer Science and Business Media Deutschland GmbH: Berlin/Heidelberg, Germany, 2021; pp. 644–660. [[CrossRef](#)]
60. Sawai, R.; Paik, I.; Kuwana, A. Sentence augmentation for language translation using gpt-2. *Electronics* **2021**, *10*, 3082. [[CrossRef](#)]
61. Pellicer, L.F.A.O.; Ferreira, T.M.; Costa, A.H.R. Data augmentation techniques in natural language processing. *Appl. Soft Comput.* **2023**, *132*, 109803. [[CrossRef](#)]
62. Chang, Y.; Zhang, R.; Pu, J. I-WAS: A Data Augmentation Method with GPT-2 for Simile Detection. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer Science and Business Media Deutschland GmbH: Berlin/Heidelberg, Germany, 2023; pp. 265–279. [[CrossRef](#)]
63. Chen, H.; Zhang, W.; Cheng, L.; Ye, H. Diverse and High-Quality Data Augmentation Using GPT for Named Entity Recognition. In *Communications in Computer and Information Science*; Springer Science and Business Media Deutschland GmbH: Berlin/Heidelberg, Germany, 2023; pp. 272–283. [[CrossRef](#)]
64. Nakamoto, R.; Flanagan, B.; Yamauchi, T.; Dai, Y.; Takami, K.; Ogata, H. Enhancing Automated Scoring of Math Self-Explanation Quality Using LLM-Generated Datasets: A Semi-Supervised Approach. *Computers* **2023**, *12*, 217. [[CrossRef](#)]
65. Jansen, B.J.; Jung, S.-G.; Salminen, J. Employing large language models in survey research. *Nat. Lang. Process. J.* **2023**, *4*, 100020. [[CrossRef](#)]
66. Joon, J.; Chung, Y.; Kamar, E.; Amershi, S. Increasing Diversity While Maintaining Accuracy: Text Data Generation with Large Language Models and Human Interventions. *arXiv* **2023**, arXiv:2306.04140.
67. Borisov, V.; Leemann, T.; Seßler, K.; Haug, J.; Pawelczyk, M.; Kasneci, G. Deep Neural Networks and Tabular Data: A Survey. *IEEE Trans. Neural Networks Learn. Syst.* **2022**, 1–21. [[CrossRef](#)] [[PubMed](#)]
68. Acharya, A.; Singh, B.; Onoe, N. LLM Based Generation of Item-Description for Recommendation System. In Proceedings of the 17th ACM Conference on Recommender Systems, RecSys 2023, Singapore, 18–22 September 2023; Association for Computing Machinery, Inc.: New York, NY, USA, 2023; pp. 1204–1207. [[CrossRef](#)]
69. Narayan, A.; Chami, I.; Orr, L.; Ré, C. Can Foundation Models Wrangle Your Data? *Proc. Vldb Endow.* **2022**, *16*, 738–746. [[CrossRef](#)]
70. Borisov, V.; Seßler, K.; Leemann, T.; Pawelczyk, M.; Kasneci, G. Language Models are Realistic Tabular Data Generators. *arXiv* **2022**, arXiv:2210.06280.
71. Lee, M. A Mathematical Interpretation of Autoregressive Generative Pre-Trained Transformer and Self-Supervised Learning. *Mathematics* **2023**, *11*, 2451. [[CrossRef](#)]

72. Alahmar, A.; Mohammed, E.; Benlamri, R. Application of data mining techniques to predict the length of stay of hospitalized patients with diabetes. In Proceedings of the 2018 International Conference on Big Data Innovations and Applications, Barcelona, Spain, 6–8 August 2018; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2018; pp. 38–43. [[CrossRef](#)]
73. Sufi, F.K. AI-GlobalEvents: A Software for analyzing, identifying and explaining global events with Artificial Intelligence. *Softw. Impacts* **2022**, *11*, 100218. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.