How to use Language Models for Synthetic Text Generation in Cerebrovascular Disease-specific Medical Reports

Byoung-Doo Oh[1], Gi-Youn Kim[2], Chulho Kim[3] and Yu-Seop Kim[2]
[1]Cerebrovascular Disease Research Center, Hallym University, Republic of Korea
[2]Department of Convergence Software, Hallym University, Republic of Korea
[3]Department of Neurology, Chuncheon Sacred Heart Hospital, Republic of Korea
{iambd822, dekmj001}@gmail.com, gumdol52@hallym.or.kr, yskim01@hallym.ac.kr

Abstract

The quantity and quality of data have a significant impact on the performance of artificial intelligence (AI). Consequently, there is a growing interest in synthetic data generation for medical AI. However, research has primarily focused on medical images, with little given to text-based data such as medical records. This is because in the biomedical field, text data often contains sensitive information such as personal information, making it difficult to access a sufficient amount of data required for medical AI. Therefore, this study explores the application of language models (LMs) for synthetic text generation in low-resource domains like medical records. It compares the results of synthetic text generation based on different LMs. To achieve this, we focused on two criteria for LM-based synthetic text generation of medical records using two keywords entered by the user: 1) the impact of the LM's knowledge, 2) the impact of the LM's size. Additionally, we objectively evaluated the generated synthetic text, including quantitative metrics such as BLUE and ROUGE, along with clinician's evaluations.

1 Introduction

The performance of artificial intelligence (AI) is greatly influenced by the quantity and quality of data, including computational resources (Whang et al., 2023). For example, massive language model (LM) services like ChatGPT (OpenAI, 2023), with 20 billion parameters, require large-scale datasets and high-performance computing resources. However, building extensive datasets in the biomedical domain is challenging for the following reasons (Kalkman et al., 2022): 1) Inclusion of sensitive information such as personal details require ensuring privacy protection, information security, transparency, and related considerations. 2) Caution must be exercised due to concerns such as the potential failure to maintain anonymity and the possibility of data misuse. Consequently, despite the existing demand, accessing the necessary data in the biomedical domain can be difficult, with only a limited subset being publicly available.

Nevertheless, medical AI has consistently advanced, primarily in fields with less sensitive information, such as medical image analysis. However, text data containing extensive information, like Electronic Medical Records (EMR), is organized in different formats depending on regions or institutions and poses a challenge for effective utilization due to the inclusion of a considerable amount of sensitive information (Lee and Kim, 2021). Therefore, there is a need for a technique of synthetic text generation, one of the text augmentation methods, to enable the effective utilization of medical records in medical AI.

In this study, we explored methods to synthetic text generation using LMs for text augmentation in low-resource medical records. We compared the following aspects: 1) the impact of the LM's knowledge, and 2) the impact of the LM's size. To do so, we compared ChatGPT, BioGPT (Luo et al., 2023), distilGPT2 (Sanh et al., 2019), and our own GPT2 (we called CerebroGPT). At this time, we conducted experiments using low-resource medical records related to cerebrovascular diseases. Finally, for objective evaluation, we performed clinician's evaluations along with four quantitative metrics (BLUE, ROUGE, Cosine similarity with sentence embeddings and TF-IDF).

This paper is structured as follows: In Chapter 2, we provide a review of prior research on text augmentation methods. Chapter 3 describes the medical records we utilized in this study. Chapter 4 details the methodology, while Chapter 5 details

Proceedings of the 1st Workshop on Personalization of Generative AI Systems (PERSONALIZE 2024), pages 10–17
March 22, 2024

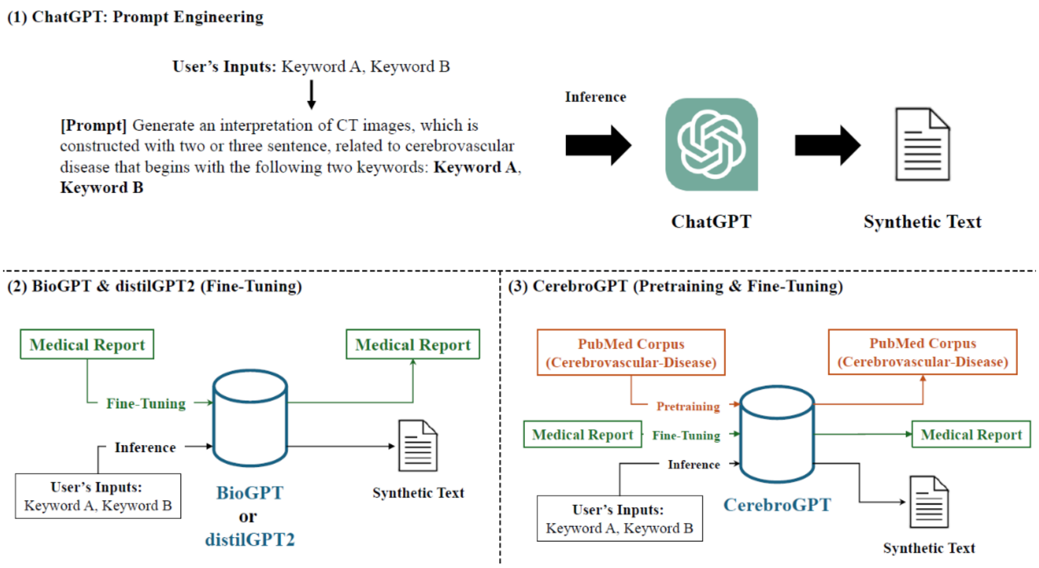

Figure 1. Overviews of the proposed methods for synthetic text generation by LMs.

the experimental setup and results. In Chapter 6, we summarize our observations and present the conclusion.

2 Related Work

Rule-based methods involve traditional techniques such as synonym replacement, random word insertion and deletion, collectively known as Easy Data Augmentation (Wei and Zou, 2019). Due to their simple principles, they have addressed good performance when applied to small-scale datasets. However, it may not be suitable for the medical domain when applied to text, as they could incorrectly represent the location of abnormalities or describe diseases inaccurately in medical records.

Model-based methods augment text based on the language understanding derived from the knowledge of LMs. Rakshit et al. (2022) enhanced text generation by fine-tuning GPT2 using the generative abilities of the LM. It involved adding new special tokens ("<|startoftext|>") to the model during fine-tuning, enabling it to generate text learned based on knowledge of LMs.

After the release of ChatGPT, prompt engineering based on ChatGPT has also been attempted (Ubani et al., 2023; Dai et al., 2023; Møller et al., 2023). In these efforts, suitable prompts for specific tasks were explored when using ChatGPT for generating synthetic training data in low-resource tasks. They focused on the fact that ChatGPT, pretrained on a large corpus, has a broad semantic space, allowing it to generate various expressions for the same meaning.

3 Medical Reports in Cerebrovascular Disease

The dataset consists of radiologist's reports on CT images collected from Hallym University Sacred Heart Hospital [1] and Chuncheon Sacred Heart Hospital [2] from 2012 to 2020, involving a total of 35,511 individuals which is intracerebral hemorrhage (ICH), non-ICH, and normal patients. While it includes a total of 90,489 reports, but there are many duplicate reports with the same content. Therefore, we utilized 47,675 reports after removing duplicates. Additionally, to standardize different report formats by radiologists and to address sensitive information such as personal details, we conducted preprocessing using regular expressions. All procedures in this study were performed in accordance with the Decla-ration of Helsinki, and it was approved by the Institutional Review Board at Chuncheon Sacred Heart Hospital (IRB No. 2021-10-012). The examples of medical reports are shown in Appendix A.

4 Language Models for Synthetic Text Generation

As shown in Figure 1, we compared the following LMs for LM-based synthetic text generation of medical records:

[1] https://hallym.hallym.or.kr/

[2] https://chuncheon.hallym.or.kr/

- **ChatGPT (20B)**: ChatGPT utilizes the GPT3.5-Turbo version, a LM that has been pretrained on diverse datasets for various domains and fine-tuned using methods like RLHF (Reinforcement Learning of Human Feedback) in InstructGPT (Ouyang et al., 2022). Therefore, we explored suitable prompts through prompt engineering and generated synthetic text using this prompt.
- **BioGPT (355M)**: BioGPT is structured with the same architecture as GPT2-medium: vocabulary size is 42,384, number of layers is 24, hidden size is 1024 dimensions, and number of heads is 16. It is a domain-specific LM fine-tuned for the biomedical domain, trained on a comprehensive biomedical dataset of 15 million papers (title and abstract) from PubMed. Therefore, we fine-tuned a BioGPT using medical record corpus.
- **distilGPT2 (82M)**: The architecture of distilGPT2 is as follows: vocabulary size is 50,260, number of layers is 6, hidden size is 768 dimensions, and number of heads is 12. It is a LM that underwent knowledge distillation from GPT2 (124M) trained on the OpenWebText Corpus (Gokaslan and Cohen, 2019), a corpus from a general domain, to achieve a smaller size and faster inference speed. Therefore, we fine-tuned a distilGPT2 using medical record corpus.
- **CerebroGPT (10M)**: CerebroGPT is a cerebrovascular disease-specific LM which is developed in-house and small-sized, with the following architecture: vocabulary size is 42,384, number of layers is 3, hidden size is 128 dimensions, and number of heads is 4. It was pretrained on a corpus of 396,826 papers (title and abstract) from the cerebrovascular disease, obtained by filtering out duplicate papers and items with only titles and no abstracts from PubMed. And Therefore, we fine-tuned a CerebroGPT using medical record corpus.

Synthetic text generation requires producing diverse expressions or sentence structures for the same content. Therefore, we compared four generation strategies (grid search, beam search, top-k sampling, top-p sampling) and employed the most suitable strategy for each LM. And all LMs are designed to generate synthetic text based on two keywords inputted by the user.

5 Experiments

5.1 Experimental Setup

The experiments were conducted on a server equipped with two Nvidia A100 GPUs. All LMs except ChatGPT used the Adam optimizer with a learning rate set to 1e-4. BioGPT and distilGPT2 had a batch size of 8, while CerebroGPT had a batch size of 64. Early stopping was set to 10. For actual clinician's evaluations, we selected 100 samples as the evaluation sets and split the remaining data into training and validation sets with a ratio of 9:1.

5.2 Results and Discussion

We employed quantitative metrics for the generated text: BLEU, ROUGE, cosine similarity with sentence embedding and TF-IDF. The detailed descriptions of these metrics are as follows: 1) **BLEU**: We use BLEU to evaluate how well the synthetic text well represents the cerebrovascular disease-specific vocabularies. 2) **ROUGE**: We use ROUGE-N and ROUGE-L for evaluation. It evaluates to the same goal as BLEU. 3) **Cosine similarity with sentence embedding**: We use BioGPT to represent the embeddings of the text. Subsequently, we calculated the cosine similarity between the two embeddings, evaluating the semantic similarity between medical records and synthetic text. 4) **Cosine similarity with TF-IDF**: We represented vectors of TF-IDF on a sentence-by-sentence basis after filtering out stop words and evaluated the cosine similarity for these vectors. At this time, we used the original report from which two keywords were extracted from the test data as a reference sentence. Additionally, we evaluated the quality of the synthetic text directly by clinicians to ensure an objective evaluation.

The results of the synthetic texts generated by each model are presented in Appendix B. And the results for the four quantitative metrics are shown in Table 1. In overall aspects such as semantics and representation ability of cerebrovascular disease-related vocabulary, distilGPT2 demonstrated the most superior performance. ChatGPT, considered as one of the most advanced LMs, effectively expressed grammatically complete sentences and medical knowledge. However, we consider that it received a lower score due to the inclusion of many general expressions and the format different to

	Sentence Embeddings	TF-IDF	ROUGE-1	ROUGE-2	ROUGE-L
ChatGPT	0.6692	0.1612	0.1154	0.0293	0.1131
BioGPT (top-p)	0.5711	0.2168	0.1045	0.0141	0.1011
distilGPT2 (top-k)	**0.8514**	**0.4549**	**0.4217**	**0.2618**	**0.4146**
CerebroGPT (top-k)	0.7997	0.3828	0.3223	0.1703	0.3179
	BLEU-1	**BLEU-2**	**BLEU-3**	**BLEU-4**	**BLEU**
ChatGPT	0.0689	0.1873	0.0021	0.0001	0.0004
BioGPT (top-p)	0.0452	0.0078	0.0001	2.16e-308	1.12e-80
distilGPT2 (top-k)	**0.2945**	**0.1916**	**0.1189**	**0.0734**	**0.1165**
CerebroGPT (top-k)	0.2057	0.1126	0.0516	0.0271	0.0503

Table 1: Evaluation results for synthetic text using evaluation metric.

<table>
<tr><th colspan="4">Best performance ranking</th></tr>
<tr><th>3</th><th>4</th><th>1</th><th>2</th></tr>
<tr><th>ChatGPT</th><th>BioGPT</th><th>distilGPT2</th><th>CerebroGPT</th></tr>
<tr><td>Excellent ability to express various sentences with the general fact.</td><td rowspan="4">Difficult to evaluate synthetic text because it is not created with a complete sentence structure.</td><td>Excellent ability to express various sentences with the medical fact.</td><td>The ability to represent sentences based on medical facts is excellent, but the ability to generate sentences with various expressions is slightly poor.</td></tr>
<tr><td>Excellent ability to express grammatically complete sentences.</td><td>Some ungrammatical sentence expressions.</td><td>Some ungrammatical sentence expressions.</td></tr>
<tr><td>Many additional general expressions appear in the synthetic text.</td><td rowspan="2">Represented in a format similar to actual medical reports.</td><td rowspan="2">Represented in a format most similar to actual medical reports.</td></tr>
<tr><td>Represented in a different format from actual medical reports.</td></tr>
</table>

Table 2: Human evaluation results (with clinician).

medical records. BioGPT, pretrained on a PubMed corpus of 15 million papers, effectively expresses medical knowledge. However, fine-tuned on low-resource medical records compared to large model sizes, it was judged to be unable to accurately represent incomplete sentence structures and the format of medical records. Furthermore, CerebroGPT, the model with the smallest size, demonstrated performance just below distilGPT2. Despite its compact size, it exhibited excellent capabilities in expressing cerebrovascular disease-related vocabulary and effectively represented the format of medical records. However, we consider that it tended to express less grammatical constructs compared to distilGPT2, contributing to these observed results. Lastly, In the case of BLEU and ROUGE, the overall scores are distributed low, because the goal is to generate synthetic text with diverse expressions rather than generating the same text as the reference sentence.

Next, we received evaluations from clinician, and detailed reviews are available in Table 2. The clinician evaluated 100 synthetic texts generated by each LM by directly reviewing them. As a result, distilGPT2 and CerebroGPT were evaluated to produce synthetic text that contained medical facts at the similar as a radiologist or other clinician. In addition, when checking the overall aspect, including grammatical sentence structure expressions, etc., the clinician's evaluation evaluated the quality of synthetic text in the same as the quantitative evaluation, in the order of distilGPT2, CerebroGPT, ChatGPT, and BioGPT.

6 Conclusion

In this study, we explored the use of LMs for synthetic text generation based on two keywords provided by the user to augment low-resource medical records. We compared ChatGPT, BioGPT, distilGPT2, and CerebroGPT (in-house). The experimental results indicated that, overall, distilGPT2 exhibited superior performance in semantic aspects and the expression of cerebrovascular disease-related vocabulary. Moreover, distilGPT2 demonstrated the most suitable quality of synthetic text in the evaluation by clinicians. Therefore, we conclude that selecting a LM tailored to the dataset size of the specific domain, rather than relying on large-scale LMs (LLMs), is crucial for augmenting low-resource domain text. Additionally, it is considered that generating high-quality synthetic text is possible by preprocessing the text into the desired format.

However, as data augmentation ultimately involves generating synthetic dataset, annotating for the synthetic data must also be conducted. Therefore, we will develop an annotation model for cerebrovascular disease-related synthetic text. In addition, we plan to obtain additional permission to build a database of not only radiologist's reports on CT images but also EMRs from a wide range of medical fields with sensitive information removed. Through this, we plan to use Retrieval-Augmented Generation (Lewis et al., 2020) to conduct research on an advanced text augmentation algorithm based on medical facts in EMR. These results are expected to play an important role in further improving synthetic text generation in the medical domain and utilizing the rich information contained in EMR in the future medical AI field.

Limitations

This work has several limitations. First, the cerebrovascular disease-related medical records used contain sensitive information and cannot be disclosed publicly. Second, the format of the preprocessed medical records may differ from those used in other countries or institutions. Third, the proposed method is limited to cerebrovascular diseases and does not cover the entire biomedical domain. The synthetic text generation in the broader medical domain, including the format and vocabulary expressions familiar to most clinicians (e.g., abbreviations), remains a future task.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the K orea government (MSIT) (No. 2022R1A5A8019303) and Institute of Information & communication s Technology Planning & Evaluation (IITP) grant f unded by the Korea government (MSIT) [NO.2021-0-02068, Artificial Intelligence Innovation Hub (Artificial Intelligence Institute, Seoul National Un iversity)].

References

Aaron Gokaslan and Vanya Cohen. 2019. OpenWebText corpus.

Anders Giovanni Møller, Jacob Aarup Dalsgaard, Arianna Pera, and Luca Maria Aiello. 2023. Is a prompt and a few samples all you need? Using GPT-4 for data augmentation in low-resource classification tasks. *arXiv preprint arXiv:2304.13861*.

Haixing Dai, Zhengliang Liu, Wenxiong Liao, Xiaoke Huang, Yihan Cao, Zihao Wu, Lin Zhao, Shaochen Xu, Wei Liu, Ninghao Liu, Sheng Li, Dajiang Zhu, Hongmin Cai, Lichao Sun, Quanzheng Li, Dinggang Shen, Tianming Liu, and Xiang Li. 2023. AugGPT: Leveraging ChatGPT for Text Data Augmentation. *arXiv preprint arXiv:2302.13007*.

Jason Wei and Kai Zou. 2019. EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, pages 6382–6388, Hong Kong, China. Association for Computational Linguistics.

Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. 2022. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, pages 27730-27744.

OpenAI. 2023. Gpt-4 technical report.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *Advances in Neural Information Processing Systems*, pages 9459-9474.

Pranati Rakshit, Anuj Kumar and Amlan Chakraborty. 2022. A Survey on Video Description and Summarization Using Deep Learning-Based Methods. *Machine Learning in Information and Communication Technology: Lecture Notes in Networks and Systems*, 498:213–223.

Renqian Luo, Liai Sun, Yingce Xia, Tao Qin, Sheng Zhang, Hoifung Poon, and Tie-Yan Lin. 2022. BioGPT: generative pre-trained transformer for biomedical text generation and mining. *Briefings in Bioinformatics*, 23(6): bbac409.

Shona Kalkman, Johannes van Delden, Amitava Banerjee, Benoît Tyl, Menno Mostert, and Ghislaine van Thiel. 2022. Patients' and public views and attitudes towards the sharing of health data for research: a narrative review of the empirical evidence. *Journal of medical ethics*, 48(1): 3-13.

Solomon Ubani, Suleyman Olcay Polat, Rodney Nielsen. 2023. ZeroShotDataAug: Generating and Augmenting Training Data with ChatGPT. *arXiv preprint arXiv:2304.14334*.

Steven Euijong Whang, Yuji Roh, Hwanjun Song, and Jae-Gil Lee. 2023. Data collection and quality challenges in deep learning: A data-centric ai perspective. *The VLDB Journal*, 32(4): 791-813.

Suehyun Lee and Hun-Sung Kim. 2021. Prospect of Artificial Intelligence Based on Electronic Medical Record. *Journal of Lipid and Atherosclerosis*, 10(3): 282-290.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. DistilBERT, a Distilled Version of BERT: Smaller, Faster, Cheaper and Lighter. In *Proceeding of the NeurIPS EMC^2 Workshop*.

A Medical Report Corpus

Medical records were collected for a total of 35,511 people, and details are as follows: 1) Patients with intracranial hemorrhage: 10,249, 2) Patients with other cerebrovascular diseases or normal individuals: 25,262. Table 3 shows the original text and preprocessed results of the medical records.

B Generated Synthetic Texts

Each of the generated synthetic texts is a result created based on the same two keywords: "Multiple, both", "History, trauma", "Non-specific, small", "1, Diffuse". This has been put together for each LM, as shown in the following Table 4-7.

Intracerebral Hemorrhage Patient	
Original text	Preprocessed text
Diffuse cerebral subdural hemorrhage: decreased	Diffuse cerebral subdural hemorrhage. - decreased.
Diffuse cerebral subdural hygroma, subdural hemorrhage - No signficant interval change since the last examination.	Diffuse cerebral subdural hygroma, subdural hemorrhage. - No signficant interval change since the last examination. Both parietal craniostomy.
Non-ICH and Normal Patient	
Original text	Preprocessed text
Unremarkable finding of brain parenchyma and cerebrospinal fluid space	Unremarkable finding of brain parenchyma and cerebrospinal fluid space.
equivocal focal low density in both globus pallidus R/O) Acute carbon monoxide intoxication	equivocal focal low density in both globus pallidus. R/O: Acute carbon monoxide intoxication.

Table 3: Example of medical report (intracerebral hemorrhage patient, non-ich and normal patient).

distilGPT2
Mulitple both frontal contusional hematomas -resolving. Midline frontal epidural hemorrhage -No signficant interval change since the last examination. Diffuse both cerebral subdural hygromas. -decreased. Left frontal subcortical hemaL subarachnoid hemorrhages -mild decreased.
History -trauma, S/P coil embolization at ruptured aneurysm at left P-om. SAH, IVH and hydrocephalus. Otherwise, no demonstrable abnormal finding.
Non-specific small nodular calcification or post infectious calcified granuloma. -Meningioma or parasite infestation. Chronic right otomastoiditis.
1. Diffuse subarachnoid hemorrhage. 2. intraventricular hemorrhages in lateral, fourth and third ventricles. 3. intracerebral hemorrhlacranial arteries. -No signficant interval change since the last examination.

Table 4: Examples of synthetic text generated by distilGPT2.

CerebroGPT
Mulitple both frontal subcortical intracerebral hemorrhage. Small vessel disease of both frontal lobes. -remnant intraventricular hemorrhages in left basal ganglia.
History -trauma, both frontal deep white matters. Multifocal low density in both frontal lobes.
Non-specific small infarctions in left basal ganglia and left frontal lobe. No signficant interval change since the last examination. Otherwise no demonstrable abnormal finding. Left corona radiata and temporal cortical Intracerebral hemorrhage in both fronto-parietal lobe.
1. Diffuse subarachnoid hemorrhage in left fronto-parietal cortical subarachnoid hemorrhage -decreased.

Table 5: Examples of synthetic text generated by CerebroGPT.

ChatGPT
Multiple both large and small infarcts are present, indicative of a cerebrovascular disease affecting various regions of the brain. These findings suggest a significant compromise in blood flow leading to ischemic strokes. Prompt medical intervention should be considered to prevent further damage and assess the underlying cause.
History - trauma: The CT images reveal a vascular abnormality in the brain, suggesting a possible cerebrovascular disease. This may be indicative of a previous traumatic event, such as a head injury or blood vessel rupture, leading to impaired blood flow in the affected area. Further medical evaluation and treatment should be considered.
Non-specific small hypodensities are observed in the subcortical region of the brain, suggesting possible lacunar infarcts or chronic microvascular ischemic changes in a patient with cerebrovascular disease. Further clinical correlation and follow-up imaging may be necessary to ascertain the underlying cause and guide management decisions.
1. Diffuse subarachnoid hemorrhage is a condition characterized by bleeding throughout the spaces between the brain and the thin membranes covering it, suggesting a ruptured blood vessel. Immediate medical attention is vital as it can lead to significant brain damage or even life-threatening complications. 2. Diffuse subarachnoid hyperdensity is indicative of widespread blood extravasation within the subarachnoid spaces, typically caused by an aneurysm rupture. Timely intervention and interventions such as coiling or clipping are crucial to prevent further bleeding and minimize long-term neurological deficits.

Table 6: Examples of synthetic text generated by ChatGPT. Highlights are the results of generating sentences with general content, such as advice to consult a doctor or get tested at a hospital.

BioGPT
Mulitple both contuhematoma left-cortex subcortical. frontal hemorrhage counter ou. hemorrhage right ie@-@ cortical subcortical -interval. parietal hemorrhage Multiright-subdural ag. hemorrhages ght temporal. st/ of, to,.. of. / 1 Chronic. Chronic otomo., maxillary sphenoid. osclerotic calcification both vertebral. no abnormal. significant change last. C F U MRI. urther of age refer. to.-. c F Uporboth bone with cortex. dural. - 2 c f u, ension multistage. stified 3 considering. old and with as to 3 Mild in. not. 5 evidence subarachnoid,,-R O 1 Old. unal 2 No. 1 Multiple sional. lacunar. No change within. (), c no change 2 Multiple fracture right bone skull able 3 4 use. 3 Traumatic and, and contuhemorrhages mega significantly usp1 including elresolving reversible. both cortical and ernal. Non pecsubarachnoid. in frontal sul. suggvely 1--developed unres. right cerebral hyroma (5 4 5 Lt. brain. reversible edema urther. 4 Re). olstate 2 Suspously ventricular. ereus. R S O. Resving 2 Nontraumatic, with edema both ventricle post. 2 RE-brain and brain not how ressive - (2 4,) 4 4> than 4.44. hydrocephalus slightly C D. s 5 Mild. evidence diffuse swelling M1 other idum since. 4. able. considering significant MU4 4 2 2 8uschange. ending. nreversible of hydrocephalus 4 mild. with. lgprogressive involving a the. on aggrav. developed. post2 ...
History -trauma, /. x Trauma hemorrhagic.,. infarction hemorrhagic sions both and frontal. SDH right lobe left lobe Otherwise demonstrable finding or orwith ficlevel. urther, lutide of ost- discrete. C view within. stable how by than res.. / subarachnoid. of temporal. not. to. with arterial. aneor ymaneurysm and as more rather. ension use fic. developed el. able. 1 along dural. - 1 Chronic cerebral. dural with. c. Chronic. No chang. ogenic. R O metastasis 2 Old infarction left ganglia 3 Fungal. no change left otitis 4 Chronic otomo. in right sto. usion right jugular. 3 Chronic lesions both astand ara. within mastoid cells to developed. opp usein watershed a-. evidence 1- 1 Old. ward. lacunar. 4 Soft density left udspace. ified. -, t ...

Table 7: Examples of generated synthetic texts (BioGPT). BioGPT's synthetic text is very long and unstructured sentences.