



Large language model para geração de prontuários eletrônicos sintéticos

Large language model to generate synthetic electronic medical records

Large language model para gerar registros médicos electrónicos sintéticos

Gabriel Constantin da Silva¹, Silvio César Cazella²

1 Mestrando, PPGTIG Saúde, UFCSPA, Porto Alegre (RS), Brasil.

2 Professor Doutor, PPGTIG Saúde, UFCSPA, Porto Alegre (RS), Brasil.

Autor correspondente: Mestrando Gabriel Constantin da Silva

E-mail: gabriel.constantin@ufcspa.edu.br

Links: <https://zenodo.org/records/10472522>

Resumo

Introdução: A utilização de dados de saúde em pesquisas é limitada por questões éticas. Isso desafia os pesquisadores a encontrarem formas de obter o material necessário para desenvolverem seu trabalho. **Método:** Usou-se uma ferramenta de Large Language Model (LLM) para gerar prontuários eletrônicos (PE) sintéticos de pacientes cardiológicos utilizando-se as técnicas "*few-shot prompting*" e "*chain-of-thought prompting*". **Objetivo:** criar um conjunto de dados abrangente e acessível para auxiliar no treinamento de algoritmos de classificação de texto em cenários médicos. **Resultados:** Foram gerados 103 PE sintéticos, abrangendo diagnósticos cardíacos distintos. **Conclusão:** A geração de PE sintéticos através de LLM apresentaram qualidade esperada, sendo condizentes com o conteúdo encontrado em PE reais. O conjunto de dados está disponível no repositório Zenodo para uso irrestrito pela comunidade de pesquisa, seguindo o conceito de ciência aberta.

Descritores: Prontuários eletrônicos sintéticos; Large Language Model; Ciência Aberta

Abstract

Introduction: The use of health data in research is limited by ethical issues. This challenges researchers to find ways to obtain the necessary material to carry out their work. **Method:** A Large Language Model (LLM) tool was used to generate synthetic electronic health records (EHR) for cardiology patients, employing the techniques of "*few-shot prompting*" and "*chain-of-thought prompting*". **Objectives:** Create a



comprehensive and accessible dataset to aid in training text classification algorithms in medical scenarios. **Results:** 103 synthetic EHR were generated, covering different cardiac diagnoses. **Conclusion:** The generation of synthetic EHR through LLM presented the expected quality, being consistent with the content found in real EHR. The dataset is available in the Zenodo repository for unrestricted use by the research community, following the concept of open science.

Keywords: Electronic Health Records; Large Language Model; Open Science

Resumen

Introducción: El uso de datos en la investigación está limitado por cuestiones éticas. Esto desafía a los investigadores a obtener el material para realizar su trabajo.

Métodos: Se utilizó una herramienta de Large Language Model (LLM) para generar registros electrónicos de salud (EHR) sintéticos de pacientes cardiológicos, empleando las técnicas de "few-shot prompting" y "chain-of-thought prompting". **Objetivo:** crear un conjunto de datos completo y accesible para ayudar a entrenar algoritmos de clasificación de textos en escenarios médicos. **Resultados:** se generaron 103 RES sintéticos, que cubren diagnósticos cardíacos diferentes. **Conclusión:** La generación de RES sintéticos a través del LLM presentó la calidad esperada, consistente con el contenido encontrado en los RES reales. El conjunto de datos está disponible en el repositorio de Zenodo para uso por parte de la comunidad de investigación, siguiendo el concepto de ciencia abierta.

Descriptor: Registros Electrónicos de Salud; Large Language Model; Ciencia Abierta

Introdução

Um problema muito comum em pesquisas com modelos de Inteligência Artificial é a obtenção de conjuntos de dados para realizar treinamento, testes e validação. Obter ou gerar dados experimentais pode ser caro e muitas vezes apresenta um desafio intransponível¹. O uso de dados reais corre o risco de expor informações confidenciais². Essa preocupação é ainda maior em relação à coleta de dados de saúde, limitando o uso eficiente desses dados³.



A geração de dados sintéticos por meio de modelos de IA generativa pode ser uma estratégia eficaz para melhorar o treinamento de algoritmos de classificação de texto, especialmente quando a pesquisa envolve dados sensíveis como da área da saúde. A geração de dados sintéticos envolve a criação artificial de dados que refletem as características e distribuições de dados reais, sem incluir informações confidenciais⁴.

A utilização de dados sintéticos oferece uma alternativa promissora para o uso de dados reais em pesquisa, experimentação e treinamento, protegendo a privacidade e mitigando os riscos de exposição de informações sensíveis⁵.

A geração de dados sintéticos por meio de modelos de IA generativa pode ser uma estratégia viável na área da saúde. No contexto de criação de Prontuários Eletrônicos (PE) sintéticos pode se beneficiar com a capacidade dos modelos de linguagem de gerar textos quase idênticos àqueles escritos por seres humanos⁶.

Large Language Models

Large Language Models (LLM), como o GPT (*Generative Pre-trained Transformer*), são baseados em uma série de avanços teóricos e práticos no campo do processamento de linguagem natural (PLN) e aprendizado profundo. É um modelo de aprendizado profundo pré-treinado com uma enorme quantidade de textos⁷.

Introduzidos no artigo "*Attention is All You Need*" por Vaswani (2017)²¹, os *Transformers* são o núcleo dos LLMs modernos. Eles são uma arquitetura de rede neural que se baseia principalmente no mecanismo de atenção para ponderar a importância relativa de diferentes partes da entrada de dados. Isso permite que os modelos processem sequências de entrada, como texto, de maneira eficiente e flexível, capturando dependências de longo alcance⁸.

Devido ao seu tamanho e complexidade, LLMs têm uma capacidade notável de generalizar a partir de seus dados de treinamento e gerar respostas coerentes e contextuais em uma ampla gama de tópicos, mesmo que não sejam explicitamente treinados para esses tópicos⁹.

ChatGPT



ChatGPT é um modelo de linguagem de última geração que utiliza técnicas de aprendizagem profunda para gerar texto semelhante ao humano utilizando uma arquitetura de rede neural profunda baseada no modelo *Transformer*, e como modelo generativo, pode gerar textos inéditos com base na entrada que recebe.

O modelo é pré-treinado em grandes quantidades de dados de texto, o que lhe permite aprender os padrões e estruturas subjacentes da linguagem. Uma das principais vantagens do ChatGPT é a sua capacidade de gerar dados coerentes e contextualizados, tornando-o uma ferramenta popular para aplicações como chatbots, tradução de idiomas e geração de texto. ChatGPT é um produto da OpenAI¹, uma empresa dedicada ao desenvolvimento de tecnologias de ponta¹⁰.

Engenharia de *Prompts*

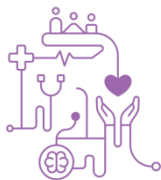
Prompts e a engenharia de *prompts* são elementos fundamentais no uso eficaz de Large Language Models (LLM) como o GPT. Esses *prompts* são projetados para interagir de forma eficiente com o modelo de forma semelhante à comunicação humana, podendo gerar respostas satisfatórias a intenção do usuário, e realizar uma ampla gama de tarefas de processamento de linguagem natural¹¹.

Eles desempenham um papel crucial na forma como esses modelos geram respostas e interagem com os usuários. Um *prompt* é uma instrução ou pergunta fornecida ao modelo de linguagem, que serve como ponto de partida para a geração de texto. Os *prompts* são a forma de comunicação principal entre o usuário e o LLM, indicando ao modelo o tipo de informação ou resposta desejada¹².

Prontuários Médicos Eletrônicos

Prontuários Eletrônicos (PE), também conhecidos como Registros Eletrônicos de Saúde, são sistemas digitais que armazenam de maneira sistemática informações de saúde dos pacientes em formato eletrônico. Esses sistemas representam uma evolução digital dos registros de saúde tradicionais em papel, oferecendo uma série de vantagens em termos de acessibilidade, eficiência e qualidade do atendimento ao paciente¹³.

¹ <https://openai.com/>



A adoção dos PE teve um crescimento exponencial tanto em hospitais quanto consultórios particulares. Apesar de terem sido criados para melhorar a eficiência operacional das atividades relacionadas ao cuidado com a saúde, os PE abrem possibilidades de aproveitamento para a área de informática médica como extração de conceitos médicos, suporte de tomada de decisões, entre outras aplicações voltadas para a saúde¹⁴.

Dados Sintéticos

A geração de dados sintéticos imita dados do mundo real, mas não os replica diretamente, garantindo a privacidade e aumentando a diversidade de conjuntos de dados utilizados para treinamento em IA (Inteligência Artificial). No entanto, para estruturas de dados complexas, como registros médicos ou dados técnicos detalhados, a integração de conhecimentos e restrições específicas do domínio é crucial para a geração de dados sintéticos realistas¹⁵.

A geração de dados sintéticos, segundo Clarke (1976), consiste em um processo multifacetado que se divide em duas etapas cruciais. A primeira etapa concentra-se na criação de dados realistas e consistentes internamente, utilizando técnicas como modelos probabilísticos e algoritmos de geração adversarial.

Na segunda etapa, a validação lógica garante a coerência das características geradas, utilizando regras predefinidas e modelos de inferência lógica. Abordar esses desafios de forma eficaz permite a criação de dados sintéticos de alta qualidade para diversas aplicações, como treinamento de modelos de inteligência artificial, simulações computacionais e criação de ambientes virtuais¹⁶.

A geração de dados sintéticos usando ChatGPT envolve treinar o modelo de linguagem em um grande corpus de dados de texto e, em seguida, usá-lo para gerar novos dados sintéticos baseados no padrão e estruturas que aprendeu com os dados de treinamento. Isto pode ser feito fornecendo um *prompt* ou um texto inicial para o modelo, que ele usa para gerar um novo texto semelhante em estilo e conteúdo aos dados originais.

No entanto, também existem algumas limitações e desafios associados, sendo um dos principais a garantia que os dados sejam representativos dos dados do mundo



real e que capturem os mesmos padrões e estruturas que estão presentes nos dados originais¹⁰.

Plataforma Zenodo

O Zenodo é uma plataforma de repositório aberto multidisciplinar que permite que pesquisadores, projetos e instituições compartilhem e disseminem seus resultados científicos, em qualquer área do conhecimento. É uma iniciativa do CERN, a organização europeia de pesquisa nuclear, com o objetivo de promover a ciência aberta e facilitar o acesso à informação científica².

Metodologia

O estudo apresenta o uso de tecnologia aplicada à geração de dados sintéticos, por isso não há necessidade de aprovação em comitê de ética.

Para atingir os objetivos propostos, utilizou-se estratégias como "*few-shot prompting*", onde exemplos específicos são fornecidos ao modelo, e "*chain-of-thought prompting*", que guia o modelo por etapas^{17,18}. Essas abordagens modelaram os prontuários de forma a refletir o processo de pensamento e tomada de decisão dos profissionais de saúde, tornando os dados simulados mais alinhados com os prontuários reais.

A versão do ChatGPT utilizada foi a 3.5 e a conversa com o modelo ocorreu a partir de um diálogo limpo (*New chat*), ou seja, sem ter recebido instruções anteriores. Isso é importante para garantir que as respostas não sejam influenciadas por contextos não desejados. O contexto não afeta apenas a forma como os LLMs interpretam e geram texto, mas também oferece caminhos para meta-aprendizagem e adaptações específicas de tarefas, enfatizando a interação dinâmica entre os LLMs e os seus contextos de entrada¹⁹.

Geração de Dados Sintéticos

As etapas foram seguidas para solicitar ao ChatGPT a criação de prontuários médicos em cardiologia que estão descritos no Quadro 1. A gera prontuários para os 8 diagnósticos mais comuns em cardiologia (totalizando 103 prontuários):

² <https://zenodo.org/>



Quadro 1 - Etapas seguidas na criação dos prontuários sintéticos.

Descrição das etapas seguidas para gerar os prompts no ChatGPT
Etapa 1: Escolha de 14 diagnósticos de cardiologia, por exemplo, insuficiência cardíaca, hipertensão arterial, infarto do miocárdio, arritmias, cardiopatia isquêmica, valvopatias.
Etapa 2: Criação de modelo para PE com base na estrutura encontrada em exemplos reais, incluindo seções como Identificação do Paciente, Histórico Médico, Sintomas, Diagnóstico, Plano de Tratamento e Notas Adicionais.
Etapa 3: Criação de prompt para cada diagnóstico afim de guiar o ChatGPT a gerar um prontuário completo. Exemplo de Prompt: "Gerar um prontuário médico simulado para um paciente com [diagnóstico]. Incluir informações básicas do paciente, histórico médico relevante, sintomas que levaram ao diagnóstico atual. Usar terminologia médica apropriada e garantir uma descrição realista e detalhada."
Etapa 4: Solicitação ao ChatGPT para gerar 10 prontuários únicos, modificando os detalhes do paciente a cada vez para garantir variedade. Isso resultou em 10 prontuários distintos para cada um dos diagnósticos, totalizando 103 prontuários.
Etapa 5: Revisão dos prontuários gerados para garantir que estavam realistas e livres de informações incoerentes ou imprecisas.

Resultados e Discussão

O conjunto de dados final consistiu em 103 PE sintéticos, abrangendo oito diagnósticos cardíacos diferentes.

Quadro 2 – Amostras com CID, descrição e quantidade.

Descrição	CID	Quantidade
Transtornos não-reumáticos da valva mitral	I34	21
Aneurisma e dissecação da aorta	I71	19
Infarto agudo do miocárdio	I21	18
Angina pectoris	I20	11
Neoplasia benigna de outros órgãos intratorácicos	D15	10
Transtornos não-reumáticos da valva aórtica	I35	10



Arritmia cardíaca	149	9
Flutter e fibrilação atrial	148	5

Quadro 3 - Exemplo de um PE sintético gerado pelo ChatGPT.

Queixa Principal: O paciente apresentou-se na unidade de saúde com queixa de dor no peito intensa e falta de ar.
Histórico Médico:
Antecedentes Pessoais: Hipertensão arterial controlada.
Histórico Familiar: Pai com histórico de doença cardíaca.
Hábitos de Vida: Ex-fumante, sem consumo de álcool, pratica atividades físicas regularmente.
Exame Físico:
Sinais Vitais: Pressão arterial: 140/90 mmHg, Frequência cardíaca: 95 bpm, Frequência respiratória: 20 rpm, Temperatura: 36,7 °C.
Estado Geral: Paciente ansioso, sudorese.
Exame Cardiovascular: Bulhas cardíacas rítmicas, pulso regular.
Exames Complementares:
Eletrocardiograma (ECG): Apresenta alterações compatíveis com Infarto Agudo do Miocárdio.
Outros Exames Cardiológicos: Testes de sangue com troponina elevada.
Diagnóstico: Com base nos sintomas apresentados e nos resultados dos exames, o paciente foi diagnosticado com Infarto Agudo do Miocárdio.

O conjunto total de dados está disponível no repositório Zenodo (<https://zenodo.org/records/10472522>).

A utilização do ChatGPT para gerar PE sintéticos mostrou-se uma alternativa viável para a criação de conjuntos de dados abrangentes e acessíveis para pesquisa em cardiologia. Diante dos resultados obtidos e apresentados sugere-se que o uso de PE sintéticos é uma alternativa possível e eficaz para contornar as limitações éticas e legais para uso em pesquisa. O uso de modelos de IA para leitura e interpretação de textos escritos por humanos, especificamente prontuários médicos escritos pelos profissionais de saúde, depende do acesso e uso dos prontuários reais, que é limitado.

Uma das razões por trás do acesso limitado decorre do fato de que os dados de PE são compostos por identificadores pessoais, que em combinação com informações médicas potencialmente sensíveis, suscitam preocupações com a privacidade. Como resultado, o acesso a esses dados para fins secundários (por exemplo, pesquisa) é regulamentado e controlado²⁰.

A geração de dados sintéticos é uma possível alternativa para isso porque garante que não haverá problemas de vazamento de dados sensíveis de pacientes. A



privacidade e a segurança dos dados de saúde são questões fundamentais em qualquer sistema de saúde eletrônico²¹.

Algumas tentativas têm sido feitas para gerar PE relacionados ao surto de doenças específicas, onde os padrões de cuidado descritos nos PE foram aplicados para gerar conjuntos de dados sintéticos. Muitos desses métodos dependem fortemente do conhecimento específico do domínio, juntamente com dados reais para gerar PE sintéticos²².

Construir um sistema capaz de sintetizar dados médicos realistas implica modelar os processos que geram essas informações. Por isso, a modelagem correta e eficiente representa o primeiro passo em direção ao desenvolvimento de uma abordagem para criar sistemas preditivos em ambientes médicos²³.

As principais vantagens desta abordagem incluem eficiência, pois a geração de PE sintéticos é um processo rápido e automatizado, em comparação com a coleta manual de dados de pacientes reais. Além disso, possuem flexibilidade, pois o modelo escrito com uso de ChatGPT pode ser facilmente adaptado para gerar PE com diferentes diagnósticos e níveis de detalhamento. Por último, há a vantagem de consistência, uma vez que os PE sintéticos são gerados de forma consistente, evitando erros e inconsistências que podem ocorrer na coleta manual de dados¹⁸.

No entanto, a geração de dados sintéticos não está livre de problemas. Uma preocupação gerada com o aumento do tamanho e da capacidade dos LLMs é sobre o viés dos dados sintéticos gerados. Os modelos podem perpetuar ou amplificar vieses presentes nos dados de treinamento, levantando questões importantes sobre seu uso e governança, principalmente se o modelo tiver algum impacto nas condutas clínicas adotadas²⁴.

Outro problema muito comum no uso de LLM é conhecido como alucinação (*hallucination*, em inglês), onde os modelos geram conteúdos incorretos, imprecisos ou que não são baseados em fatos reais. Isso representa um desafio significativo na confiabilidade no uso de LLM em aplicações de mundo real²⁵.

Algumas formas de entender e avaliar as alucinações estão sendo propostas para servir de guia para entender que tipos de conteúdos são mais propícios a apresentar esse fenômeno. Isso passa pela análise humana, através do uso de profissionais especializados para anotação das informações equivocadas geradas



como resposta a demandas factuais (e.g. Qual a data da declaração de independência dos Estados Unidos da América?)²⁶.

No contexto de uso de LLM voltados para área de saúde isso é mais importante ainda. As alucinações apresentadas pelo modelo podem gerar informações que terão impacto real na saúde de pessoas, caso elas adotem recomendações sugeridas. A desinformação no domínio médico pode levar graves consequências para a saúde no atendimento ao paciente e resultados, a precisão e a confiabilidade das informações fornecidas pelos modelos de linguagem podem ser uma questão de vida ou morte²⁷.

Mesmo na situação apresentada por este trabalho, onde usamos um LLM para criação de um prontuário, é importante que os textos gerados sejam analisados por profissional da área da saúde com conhecimento específico na área de cardiologia, a fim de identificar possíveis inconsistências nos PE sintéticos.

A criação de pontos de referência para o uso de LLM no contexto da saúde foi proposto para amenizar este tipo de problema. Um exemplo disso é o Med-HALT (*Medical Domain Hallucination Test*), que fornece uma grande base de dados com informações médicas oriundas de diferentes países. Ele abarca duas categorias de teste: racionalidade e baseados em memória²⁷.

Outro desafio é garantir que os dados sintéticos não introduzam quaisquer preconceitos ou artefatos que possam impactar o desempenho ou a imparcialidade do modelo de aprendizado de máquina. Por exemplo, se os dados sintéticos forem gerados com base em um conjunto de dados tendencioso ou incompleto, isso pode levar a um modelo tendencioso ou impreciso em determinados contextos³⁰.

Além disso, a dificuldade para diferenciar entre conteúdo criado por máquinas ou por humanos torna-se mais difícil na medida em que a tecnologia evolui. Isso abre brechas para novas formas de plágio, pois pode-se usar trechos de textos de pessoas, que foram usados para treinar o LLM, como se fossem de autoria da máquina. Existe o risco de informações erradas serem espalhadas como se fossem reais, e para tentar evitar isso, uma forma de marca d'água tem sido proposta para deixar claro que os dados foram criados por máquina²⁸.

Apesar destes desafios, a utilização de dados sintéticos é cada vez mais popular em muitas aplicações de aprendizado de máquina, especialmente em áreas como visão computacional e processamento de linguagem natural³⁰.



Conclusão

Este trabalho abordou a geração de prontuários eletrônicos sintéticos utilizando o modelo GPT-3.5 e os resultados demonstraram que o modelo é capaz de gerar prontuários com alta qualidade e fidelidade, capturando a estrutura e o conteúdo dos prontuários reais. No entanto, foi observado que o modelo não seguiu com exatidão o comando de gerar 10 amostras para cada diagnóstico.

Uma das principais vantagens do uso de dados sintéticos é que eles podem ajudar a resolver problemas de escassez de dados, que é um problema comum em muitas aplicações de aprendizado de máquina. Ao gerar novos dados semelhantes aos dados originais, os modelos de aprendizado de máquina podem ser treinados em um conjunto de dados maior e mais diversificado, o que pode levar a um melhor desempenho e generalização³⁰.

Embora este estudo tenha apresentado resultados promissores, é importante considerar que a geração de dados sintéticos é um campo de pesquisa em constante evolução. Novos modelos e técnicas estão sendo desenvolvidos continuamente, e é importante acompanhar esses avanços para garantir que os prontuários sintéticos gerados sejam de alta qualidade e confiáveis.

Por fim, deixamos registrado que existe um trabalho em andamento que utilizou dados sintéticos em conjunto com dados reais para avaliar a classificação automatizada de prontuários médicos eletrônicos.

Referências

1. Bozkurt M, Harman M. Automatically generating realistic test input from web services. In: Service Oriented System Engineering (SOSE), IEEE 6th International Symposium; 2011.
2. Rubin D. Discussion: Statistical disclosure limitation. J Off Stat. 1993;9(2):461-8.
3. Van Panhuis WG, Paul P, Emerson C, Grefenstette J, Wilder R, Herbst AJ, Burke DS. A systematic review of barriers to data sharing in public health. BMC Public Health. 2014;14:1-9.
4. Weston J, Bordes A, Chopra S, Rush A, Merrienboer B, Joulin A, Mikolov T. Towards AI-complete question answering: A set of prerequisite toy tasks.
5. Gargiulo F, Ternes S, Huet S, Deffuant G. An iterative approach for generating statistically realistic populations of households. PLoS ONE. 2010;5(1).



6. Eigenschink P, Reutterer T, Vamosi S, Vamosi R, Sun C, Kalcher K. Deep Generative Models for Synthetic Sequential Data: A Survey. *IEEE Access*. 2023.
7. Ahmad A, Waseem M, Liang P, Fahmideh M, Aktar MS, Mikkonen T. Towards human-bot collaborative software architecting with chatgpt. In: *Proceedings of the 27th International Conference on Evaluation and Assessment in Software Engineering*; 2023. p. 279-85.
8. Wang J, Liu Z, Zhao L, Wu Z, Ma C, Yu S, Zhang S. Review of large vision models and visual prompt engineering. *Meta-Radiol*. 2023;100047.
9. Meskó B. Prompt engineering as an important emerging skill for medical professionals: tutorial. *J Med Internet Res*. 2023;25:e50638.
10. Hassani H, Silva ES. The Role of ChatGPT in Data Science: How AI-Assisted Conversational Interfaces Are Revolutionizing the Field. *Big Data Cogn Comput*. 2023;7(2):62. [doi.org]10.3390/bdcc7020062.
11. Janik R. Aspects of human memory and Large Language Models. *ArXiv [Preprint]*. 2023:abs/2311.03839.
12. Heston TF, Khun C. Prompt engineering in medical education. *Int Med Educ*. 2023;2(3):198-205.
13. Cawsey AJ, Webber BL, Jones RB. Natural language generation in health care. *J Am Med Inform Assoc*. 1997;4(6):473-482. doi:10.1136/jamia.1997.0040473.
14. Shickel B, Tighe P, Bihorac A, Rashidi P. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE J Biomed Health Inform*. 2018;22(5):1589-604.
15. Chen, R. J., Lu, M. Y., Chen, T. Y., Williamson, D. F., & Mahmood, F. (2021). Synthetic data in machine learning for medicine and healthcare. *Nature Biomedical Engineering*, 5(6), 493-497.
16. Clarke LA. A system to generate test data and symbolically execute programs. *IEEE Trans Softw Eng*. 1976;SE-2(3):215-222.
17. Wei J, Wang X, Schuurmans D, Bosma M, Xia F, Chi E, Zhou D. Chain-of-thought prompting elicits reasoning in large language models. *Adv Neural Inf Process Syst*. 2022;35:24824-37.
18. Shao Z, Gong Y, Shen Y, Huang M, Duan N, Chen W. Synthetic prompting: Generating chain-of-thought demonstrations for large language models. In: *International Conference on Machine Learning*. PMLR; 2023 Jul. p. 30706-75.
19. Li, J., Cheng, X., Zhao, W., Nie, J., & Wen, J. HaluEval: A Large-Scale Hallucination Evaluation Benchmark for Large Language Models. *ArXiv*. 2023;abs/2305.11747. <https://doi.org/10.48550/arXiv.2305.11747>.
20. Choi E, Biswal S, Malin B, Duke J, Stewart WF, Sun J. Generating multi-label discrete patient records using generative adversarial networks. *ArXiv [Preprint]*. 2017:1703.06490.



21. Scott McLachlan, Kudakwashe Dube, and Thomas Gallagher. Using the caremap with health incidents statistics for generating the realistic synthetic electronic ealthcare record. In Healthcare Informatics (ICHI), 2016 IEEE International Conference on, pages 439–448. IEEE, 2016.
22. Lombardo JS, Moniz LJ. A method for generation and distribution of synthetic medical record data for evaluation of disease-monitoring systems. Johns Hopkins APL Tech Dig. 2008;27(4):356.
23. Esteban C, Hyland SL, Rätsch G. Real-valued (medical) time series generation with recurrent conditional GANs. arXiv preprint arXiv:1706.02633. 2017.
24. Strasser A. On pitfalls (and advantages) of sophisticated large language models. ArXiv [Internet]. 2023 Mar [citado 2024 Mar 27]; abs/2303.17511. Disponível em: <https://doi.org/10.48550/arXiv.2303.17511>
25. Luo J, Xiao C, Ma F. Zero-Resource Hallucination Prevention for Large Language Models. ArXiv. 2023;abs/2309.02654. <https://doi.org/10.48550/arXiv.2309.02654>.
26. Baowaly MK, Lin C-C, Liu C-L, Chen KT. Synthesizing electronic health records using improved generative adversarial networks. J Am Med Inform Assoc. 2019;26(3):228-241.
27. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Polosukhin I. Attention is all you need. Adv Neural Inf Process Syst. 2017;30.
28. Wiggers K. OpenAI's attempts to watermark AI text hit limits. TechCrunch. 2022 Dec 10. Disponível em: <https://techcrunch.com/2022/12/10/openais-attempts-to-watermark-ai-text-hit-limits>
29. Kumar UL, Pal A, Sankarasubbu M. Med-HALT: Medical Domain Hallucination Test for Large Language Models. In: Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP); 2023:314-334. <https://doi.org/10.48550/arXiv.2307.15343>.
30. Coda-Forno J, Binz M, Akata Z, Botvinick M, Wang JX, Schulz E. Meta-in-context learning in large language models. ArXiv abs/2305.12907 [Internet]. 2023 [cited 2024 Apr 07]. Available from: doi: 10.48550/arXiv.2305.12907.