



19º CONGRESSO BRASILEIRO DE
**Pneumologia
Pediátrica**
Porto Alegre - RS

**10, 11 E 12 DE
ABRIL DE 2025**

Centro de Eventos da PUCRS
Av. Ipiranga, 6681 - Partenon, Porto Alegre - RS



Trabalhos Científicos

Título: Large Language Models (Llms) E Pneumologia Pediátrica: Uma Análise Do Desempenho De Ferramentas De Inteligência Artificial Em Questões De Múltipla Escolha Sobre Temas Comuns Da Especialidade

Autores: ARTUR ZANELATTO SANTOS (UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL), BRUNO GUIMARÃES SCALCO (UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL), BRUNO NATANAEL VIEIRA (UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL), GIULIA RADIN (UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL), ISABELLA PASSOS DOS SANTOS (UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL), MARIA EDUARDA RIBAS DOS SANTOS (UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL), OTÁVIO SCHUSTER (UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL), YOHAN CASIRAGHI (UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL), RICARDO DE SOUZA KUCHENBECKER (UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL)

Resumo: O uso de ferramentas de inteligência artificial tem despertado crescente interesse, ensejando reflexões sobre sua interface com as ciências médicas, principalmente no domínio de temas relativos a diferentes especialidades. Essas ferramentas incluem os large language models (LLMs), modelos de aprendizagem de máquina baseados no processamento de linguagem natural capazes de analisar grandes quantidades de dados para gerar respostas a uma variedade de textos de entrada. Entre as plataformas disponíveis atualmente, destacam-se o ChatGPT, o Gemini, o Copilot e o Perplexity. Este trabalho pretende mensurar a capacidade de LLMs em relação a temas da pneumologia pediátrica, avaliando sua aplicabilidade em contextos clínicos e acadêmicos. Realizou-se uma revisão prospectiva e exploratória a partir de 4 LLMs disponíveis comercialmente: ChatGPT 3.5 (GPT-3.5), Gemini, Copilot e Perplexity. A partir de um website especializado em questões preparatórias para a residência médica, foram selecionadas 374 questões objetivas – de múltipla escolha, com uma única resposta descrita em gabarito –, compreendendo os seguintes tópicos: asma na infância (n=79); bronquiolites (n=97); pneumonias na infância (n=46); doença respiratória precoce (DRP) neonatal (n=50); e tuberculose na infância (n=102). Cada tema ficou sob responsabilidade de um dos autores do presente trabalho. Apenas o enunciado e as alternativas foram inseridos como texto de entrada, sem a formulação de prompts específicos. Das 374 questões elegíveis, as taxas globais de respostas corretas foram de 62,8% para o GPT-3,5, 59,9% para o Gemini, 59,9% para o Copilot e 55,3% para o Perplexity. Por sua vez, apesar de o GPT-3,5 ter obtido a maior porcentagem de acertos nas questões sobre asma na infância (67,1%), bronquiolites (77,3%) e tuberculose na infância (59,8%), o Copilot obteve pontuação mais expressiva em enunciados sobre DRP neonatal (74%) e o Gemini liderou em pneumonias na infância (58,7%). Os LLMs analisados não apresentaram taxas de acerto satisfatórias sobre assuntos comuns da pneumologia pediátrica, principalmente em pneumonias e tuberculose. Como limitações, cabe considerar a menor disponibilidade de contextos e informações em relação às temáticas estudadas para o léxico em português, além dos limites para treinamento de modelos de linguagem, o que afeta, conseqüentemente, sua assertividade. Ademais, a natureza específica de uma única base de dados pode comprometer o desempenho do modelo de resolução de questões de múltipla escolha. LLMs não devem constituir fontes únicas de conhecimento, sendo necessária cautela em seu uso como substrato para aconselhamento e/ou decisão. Novas formas de integração de conhecimento técnico especializado, como o meta-in-context learning, podem aprimorar a capacidade de aprendizado contextual de LLMs pré-treinados, possibilitando aplicações mais eficazes no contexto da pneumologia pediátrica.