

Comparative evaluation of LLMs in clinical oncology.

May 8, 2024

Rydzewski NR, Dinakaran D, Zhao SG, et al. Comparative evaluation of LLMs in clinical oncology. NEJM AI. 2024;1(5):A10a2300151. doi:10.1056/aioa2300151.

<https://psnet.ahrq.gov/issue/comparative-evaluation-llms-clinical-oncology>

Large language models (LLM) are being developed to improve [diagnostic](#) accuracy. This study compared five LLMs on their accuracy of oncology diagnoses. Accuracy ranged from no better than random chance to similar to resident [physicians](#). Notably, all models exhibited poor performance on women-predominant malignancies, suggesting a [bias](#) in training materials. This highlights the importance of partnerships between developers and medical professionals to co-develop reliable training sets.