

Combining multiple large language models improves diagnostic accuracy.

November 20, 2024

Barabucci G, Shia V, Chu ES, et al. Combining multiple large language models improves diagnostic accuracy. NEJM AI. 2024;1(11):A1cs2400502. doi:10.1056/aics2400502.

<https://psnet.ahrq.gov/issue/combining-multiple-large-language-models-improves-diagnostic-accuracy>

[Collective intelligence](#) (e.g., collaboration of multiple providers to come to a final diagnosis) has been shown to produce a more accurate diagnosis than even the group's most senior member. This study applied methods of collective intelligence to four large language models (LLM). The collective [diagnosis](#) was more accurate than individual LLMs, even when the highest performing LLM was removed. The authors suggest aggregating diagnoses from multiple LLMs may increase clinician trust in the response and mitigate reliance on a sole LLM or vendor.