

Microsoft's AI is four times better at diagnosing complex cases than human doctors

Microsoft AI has revealed that its AI Diagnostic Orchestrator (MAI-DxO) platform can accurately diagnose 85% of complex cases from New England Journal of Medicine (NEJM) medical records.

Microsoft AI has revealed that its AI Diagnostic Orchestrator (MAI-DxO) platform can accurately diagnose 85% of complex cases from the New England Journal of Medicine (NEJM) medical records. More impressively, the cases published in NEJM are often diagnostically complex and intellectually demanding. These cases typically require consultation from multiple experts and a variety of complex tests to arrive at a preliminary diagnosis. So the fact that Microsoft AI can produce results with such high accuracy is a revolutionary achievement.

MAI-DxO turns language models into a virtual panel of doctors capable of asking follow-up questions, ordering tests, or making diagnoses. MAI-DxO improved the diagnostic performance of every model Microsoft tested, with the best results recorded when paired with OpenAI's o3 model.

When MAI-DxO used o3, it was able to correctly resolve 85.5% of NEJM case records. In a comparative study of 21 practicing physicians from the US and UK with 5-20 years of clinical experience, the average accuracy they achieved in their diagnoses was only about 20%.

Microsoft believes that tools like these can dramatically transform healthcare, by empowering patients to self-manage routine aspects of their care and equipping clinicians with advanced decision support tools specifically for complex cases.

To learn how AI performs on NEJM cases, Microsoft created a benchmarking framework called the Sequential Diagnosis Benchmark (SD Bench), which converts 304 NEJM cases into step-by-step diagnoses. The models can then iteratively ask questions and prescribe tests. As new information becomes available, the model updates its reasoning and gradually moves toward a final diagnosis that can be compared to what is published in NEJM.

As mentioned earlier, the MAI-DxO system simulates a virtual panel of doctors capable of asking questions, ordering tests, or making diagnoses. Additionally, it can operate within defined cost constraints to help prevent over-testing.

While Microsoft's experiments are showing promise, there is still room for improvement. More evidence needs to be gathered from real-world clinical settings before AI can be used safely in healthcare scenarios. There also

needs to be appropriate governance and regulatory frameworks in place to ensure the model can operate reliably and safely. Microsoft is working with healthcare organizations to test and validate its approach before it is rolled out to the public.

You finished reading the article "**Microsoft's AI is four times better at diagnosing complex cases than human doctors**" edited by the [TipsMake](#) team. We hope this article has provided you with many useful tech tips and tricks. You can search for similar articles on tips and guides. Thank you for reading and for following us regularly.