

Review of: "A Comparative Study of Large Language Models in Explaining Intrinsically Disordered Proteins"

Scythia Marrow

Potential competing interests: No potential competing interests to declare.

This work is an incredible example of a singular publon. The authors laid out a clear thesis, comparing the performance of different large language models (LLMs) on answering natural language questions about Intrinsically Disordered Proteins (IDPs). They showed strong evidence for their thesis, providing clear statistical evidence of the superiority of GPT-4 over other LLMs. They even found an unexpected effect, where models allowed to browse the web performed worse than non-browser-enabled models, and hypothesized several mechanisms for why this effect may occur.

The text was clear, concise, and extremely well-cited. All evaluation methods and statistical techniques were standard scientific practice.

This article is ready for publication as is.