Hugging Face clones OpenAI's Deep Research in 24 hours

On Tuesday, Hugging Face researchers released an open source AI research agent called "Open Deep Research," created by an in-house team as a challenge 24 hours after the launch of OpenAI's Deep Research feature, which can autonomously browse the web and create research reports. The project seeks to match Deep Research's performance while making the technology freely available to developers.

"While powerful LLMs are now freely available in open-source, OpenAI didn't disclose much about the agentic framework underlying Deep Research," writes Hugging Face on its announcement page. "So we decided to embark on a 24-hour mission to reproduce their results and open-source the needed framework along the way!"

Similar to both OpenAI's Deep Research and Google's implementation of its own "Deep Research" using Gemini (first introduced in December—before OpenAI), Hugging Face's solution adds an "agent" framework to an existing AI model to allow it to perform multi-step tasks, such as collecting information and building the report as it goes along that it presents to the user at the end.

The open source clone is already racking up comparable benchmark results. After only a day's work, Hugging Face's Open Deep Research has reached 55.15 percent accuracy on the General AI Assistants (GAIA) benchmark, which tests an AI model's ability to gather and synthesize information from multiple sources. OpenAI's Deep Research scored 67.36 percent accuracy on the same benchmark.

As Hugging Face points out in its post, GAIA includes complex multi-step questions such as this one:

> *Which of the fruits shown in the 2008 painting "Embroidery from Uzbekistan" were served as part of the October 1949 breakfast menu for the ocean liner that was later used as a floating prop for the film "The Last Voyage"? Give the items as a comma-separated list, ordering them in clockwise order based on their arrangement in the painting starting from the 12 o'clock position. Use the plural form of each fruit.*

To correctly answer that type of question, the AI agent must seek out multiple disparate sources and assemble them into a coherent answer. Many of the questions in GAIA represent no easy task, even for a human, so they test agentic AI's mettle quite well.

https://arstechnica.com/ai/2025/02/after-24-hour-hackathon-hugging-faces-ai-research-agent-nearly-matches-openais-solution/