

# OpenResearcher: An Open-Source Agent That Beats GPT-4.1 at Deep Research, Trained Entirely Offline

Kabui, Charles

2026-03-29

---

[Read at ToKnow.ai](#)

---

**OpenResearcher Beats GPT-4.1, Trained Entirely Offline**

Open-source deep research agent from TIGER-Lab

- 54.8%** #1 open-source on BrowseComp-Plus
- 97K** Offline training trajectories synthesized
- 15M** Document offline corpus  
No live web APIs needed

March 29, 2026

ToKnow.ai

TIGER-Lab at the University of Waterloo released [OpenResearcher](#), a fully open pipeline for training deep research agents, the kind that can search, read, and reason across dozens of web pages to answer hard questions. The model is a [30B-parameter mixture-of-experts](#)

architecture with only 3B active parameters. What makes it unusual: the entire training loop runs offline. The team bootstrapped a [15M-document corpus](#) from FineWeb, then used three browser primitives (search, open, find) to synthesize [97K training trajectories](#) without ever calling a live web API. Some trajectories stretch to 100+ tool calls. After supervised fine-tuning, the model scores [54.8% on BrowseComp-Plus](#), ranking #1 among open-source models and beating GPT-4.1, Claude Opus 4, Gemini 2.5 Pro, DeepSeek-R1, and Alibaba’s Tongyi DeepResearch.

Training deep research agents has been gated by expensive, rate-limited web APIs. Every search query and page fetch costs money, and the results are hard to reproduce. OpenResearcher removes that dependency entirely: the offline corpus and trajectory synthesis pipeline can scale to arbitrary dataset sizes at near-zero marginal cost. Any research lab can now train a competitive deep research agent on standard hardware. NVIDIA has already [adopted the approach](#) for its Nemotron model family.

A while ago, [OpenSeeker showed](#) that 11,700 high-quality samples could match proprietary search agents. Now OpenResearcher proves you don’t even need live web access to collect those samples. The bottleneck for training research agents has shifted from data access to trajectory design.

Sources:

- [OpenResearcher Paper \(arXiv\)](#)
- [OpenResearcher GitHub \(585 stars\)](#)
- [OpenResearcher-30B-A3B Model \(HuggingFace\)](#)
- [OpenResearcher Dataset: 97K Trajectories \(HuggingFace\)](#)
- [HuggingFace Daily Papers, March 24 \(#2 Paper of the Day\)](#)

---

*Disclaimer: For information only. Accuracy or completeness not guaranteed. Illegal use prohibited. Not professional advice or solicitation. **Read more:** [/terms-of-service](#)*