Reddit blocks Internet Archive to end sneaky AI scraping



“Until they’re able to defend their site and comply with platform policies (e.g., respecting user privacy, re: deleting removed content) we’re limiting some of their access to Reddit data to protect redditors,” Rathschmidt said.

A review of social media comments suggests that in the past, some Redditors have used the Wayback Machine to research deleted comments or threads. Those commenters noted that myriad other tools exist for surfacing deleted posts or researching a user’s activity, with some suggesting that the Wayback Machine was maybe not the easiest platform to navigate for that purpose.

Redditors have also turned to resources like IA during times when Reddit’s platform changes trigger content removals. Most

recently in 2023, when changes to Reddit’s public API threatened to kill beloved subreddits, archives stepped in to preserve content before it was lost.

IA has not signaled whether it’s looking into fixes to get Reddit’s restrictions lifted and did not respond to Ars’ request to comment on how this change might impact the archive’s utility as an open web resource, given Reddit’s popularity.

The director of the Wayback Machine, Mark Graham, told Ars that IA has “a longstanding relationship with Reddit” and continues to have “ongoing discussions about this matter.”

It seems likely that Reddit is financially motivated to restrict AI firms from taking advantage of Wayback Machine archives, perhaps hoping to spur more lucrative licensing deals like Reddit struck with OpenAI and Google. The terms of the OpenAI deal were kept quiet, but the Google deal was reportedly worth $60 million. Over the next three years, Reddit expects to make more than $200 million off such licensing deals.

Disclosure: Advance Publications, which owns Ars Technica parent Condé Nast, is the largest shareholder in Reddit.

https://arstechnica.com/tech-policy/2025/08/reddit-blocks-internet-archive-to-end-sneaky-ai-scraping/