Cloudflare’s AI Crawler Rules Can Block Googlebot via @sejournal, @MattGSouthern



Cloudflare is updating its method of identifying and blocking AI crawlers, which may result in Googlebot being blocked on sites that prevent AI training. The company announced the update as part of its second Content Independence Day.

The new controls let websites manage automated traffic based on three behaviors rather than a single “block AI bots” switch. They are live now for all customers, including the free tier. A separate set of default changes takes effect September 15.

Three Ways To Sort AI Crawlers

Cloudflare now sorts crawlers by what they do on a site rather than whether they count as “AI.” The company splits the AI use cases into three categories:

- Search indexes a site to answer questions later, and

Cloudflare ties this behavior to referral traffic.

- Agent, real-time bots acting for a person, such as ChatGPT-User or browser agents like Gemini or Claude operating Chrome.
- Training, crawling that pulls content to train or fine-tune a model.

Cloudflare says bot operators should run separate crawlers for each behavior so that websites can see why a bot is visiting and decide whether to allow or block it.

What Changes On September 15

Two default changes take effect on September 15. For new customers and new sites for existing customers, Training and Agent crawlers will be blocked by default on pages that display ads, while Search stays allowed. Cloudflare's press release also says existing free customers who have not changed their settings by September 15 will be moved to these defaults.

The second change goes even further. Cloudflare will start treating multi-purpose crawlers based on their overall behavior, applying the strictest rule that applies. For example, a crawler that performs both Search and Training will be blocked if a site blocks Training. Cloudflare uses Googlebot, Applebot, and Bingbot as examples, since each crawls for both search and AI training. If a site has already enabled the older "Block AI bots" setting, it will be covered by this new rule.

If you want to keep those crawlers, you can review or change these settings in your Cloudflare dashboard any time before September 15. Cloudflare says it will continue to notify customers ahead of the date.

New Signals For How Bots Use Content

Cloudflare is also testing a content-use signal that extends Content Signals in robots.txt. It carries three values, from most to least restrictive: immediate, which stores nothing; reference, which indexes and links back and is the new default; and full, which summarizes and reproduces. Cloudflare says these state a preference and do not block on their own.

The company has revised the definition of “Verified” for bots. Now, a verified bot isn’t automatically permitted everywhere; instead, its access depends on its category. Additionally, bots that replicate content in its entirety are ineligible for verification. Cloudflare introduced a searchable directory, BotBase, for Enterprise Bot Management users, which displays each tracked bot’s classification and a copyable detection ID for security rules.

The Report Behind The Changes

The update arrived with a Cloudflare report marking the one-year anniversary of the first Content Independence Day. According to the report, AI training now accounts for the majority of crawler requests on its network, a rise from roughly 20% in spring 2025. It also notes that daily AI agent requests increased by more than 1,700% over the year. These statistics are based on Cloudflare’s network traffic and do not represent the entire web.

Why This Matters

The September 15 rule links AI training blocks to search crawling on Cloudflare’s network. If a site blocks Training to protect its content from AI models, it might also unintentionally block Googlebot, since a Cloudflare block operates at the network level, making it harder to bypass than a simple robots.txt line that Google can ignore since a Cloudflare block operates at the network level, since

robots.txt is an advisory instruction to crawlers. Losing Googlebot's access means the site won't be crawled as effectively, which could eventually impact its visibility in search results.

I've tracked publishers moving to default-deny setups and blocking both retrieval and training bots over the past year. The exposure is the same each time. Blocking the training layer can also block the search layer that keeps a site findable.

Looking Ahead

Websites using Cloudflare should review their AI blocking settings by September 15, decide whether to keep Search crawlers enabled. The combined-crawler rule mainly affects those who turned on "Block AI bots" previously and haven't adjusted their settings since. Free users who do not change their settings will have them updated to the new defaults on that date.

Cloudflare wants operators of mixed-purpose crawlers to separate those bots by behavior over the coming year. Whether major operators differentiate their bots by behavior will determine whether this becomes a real choice, rather than a compromise between blocking AI training and maintaining search visibility.

Featured Image: jackpress/Shutterstock

https://www.searchenginejournal.com/cloudflares-ai-crawler-rules-can-block-googlebot/581385/