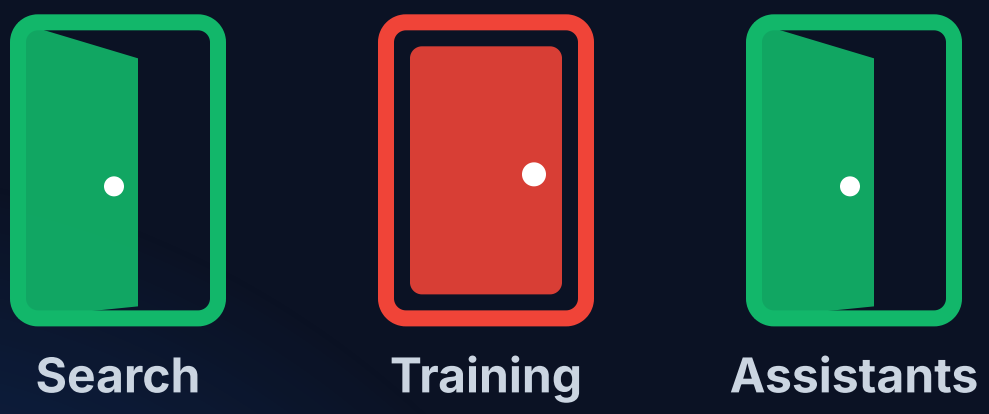
Original research · November 2026

# Who lets AI in?

We read the robots.txt of the world's top 1,000 websites, crawler by crawler, to see who keeps AI out of training, out of search, or both.




Swipe for the findings

GEOFLAKES

Finding 1

# 23%

**of top sites block at least one AI training crawler.**

- **GPTBot** 15%
- **CCBot** 17%
- **ClaudeBot** 16%
- **Google-Extended** 13%

Finding 2

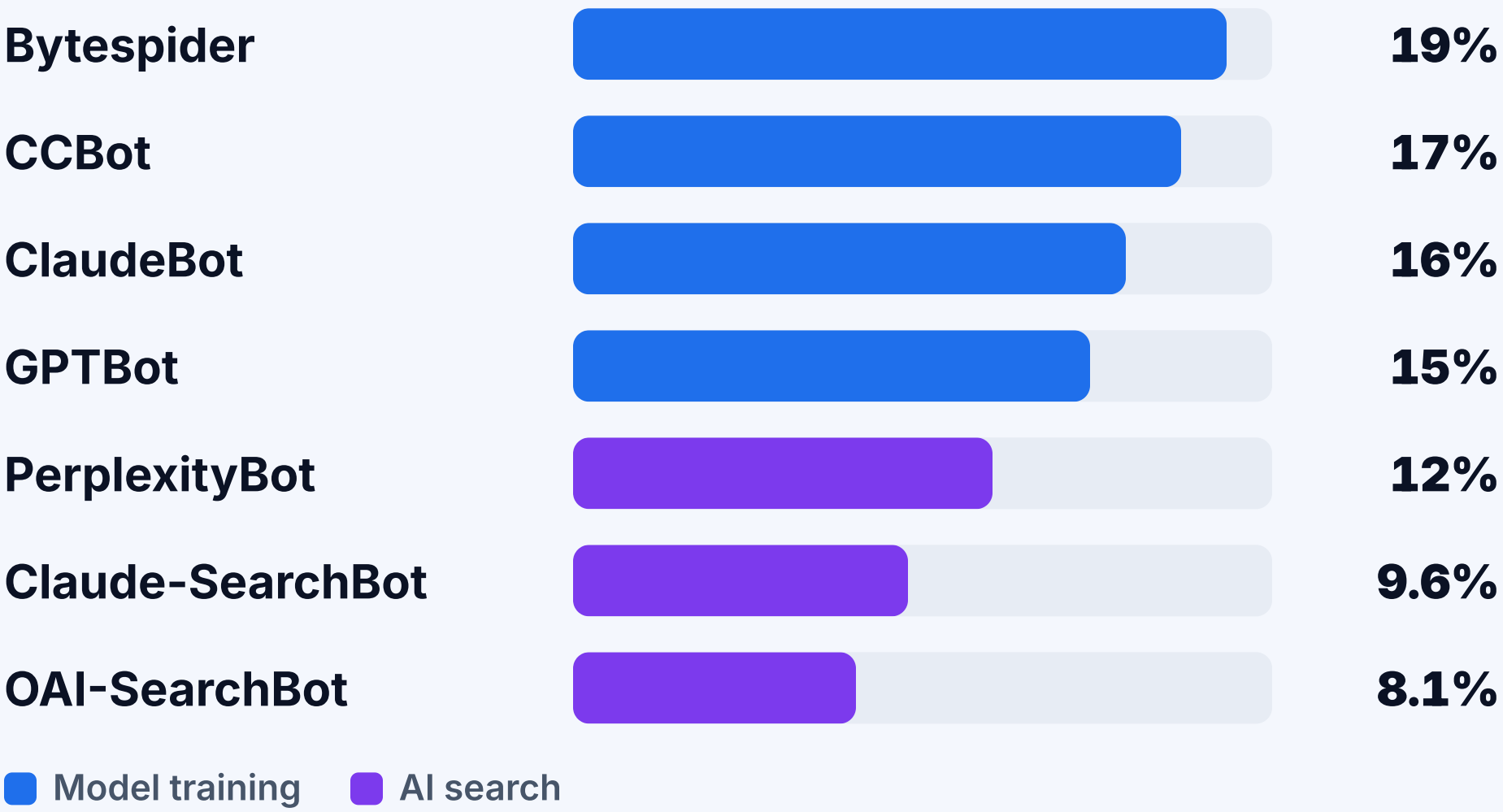
# 13%

**keep at least one AI search crawler out, and with it that engine's answers.**

- **OAI-SearchBot** 8.1%
- **PerplexityBot** 12%
- **Claude-SearchBot** 9.6%

Finding 3

# Training crawlers are blocked more often than AI search ones



Finding 4

# 45% of sites that block OpenAI's training crawler still let ChatGPT search in



**GPTBot**

Training: blocked

**OAI-SearchBot**

ChatGPT search: allowed

# 1 in 3 AI search blocks comes from one blanket rule

```
User-agent: *
Disallow: /
```

Usually written for scrapers years ago. Today it decides whether ChatGPT, Claude and Perplexity can cite you.

Your turn

# Check your robots.txt in five minutes

1. Allow OAI-SearchBot, Claude-SearchBot and PerplexityBot
2. Decide on training crawlers separately
3. Look for a blanket `User-agent: *` block
4. Check your CDN's AI bot setting

Full study and method: geoflakes.ai/resources

GEOFLAKES