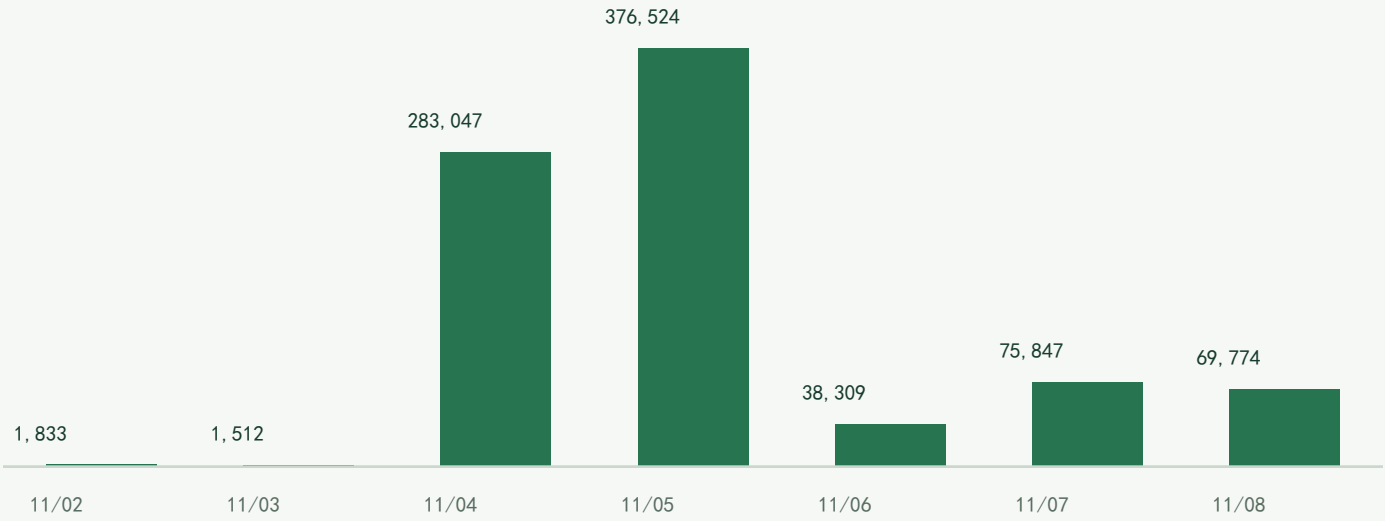
本周概览 / THE WEEKLY CRAWL

AI 与搜索爬虫，正在访问哪些品牌？

从 846,846 次本站请求中，识别出 190,043 份被访问的品牌、公司与组织档案。我们关联访问来源、具体档案与内容类型，观察品牌信息如何被访问。

全周请求量	独立档案	来源公司	错误率
846,846	190,043	15	0.122%
含 GET、HEAD 与失败响应	品牌、公司及组织 / 2xx	25 种爬虫	HTTP 状态码 >=400

每日请求量 / 次



659,571 次请求集中于 11 月 4–5 日。 周一与周二的访问量均不足 2,000 次，周三开始明显增加，周四达到峰值，随后回落。

最繁忙的一小时：11 月 5 日 14:00–15:00，共 39,248 次请求。峰值不代表全天持续处于相同水平。

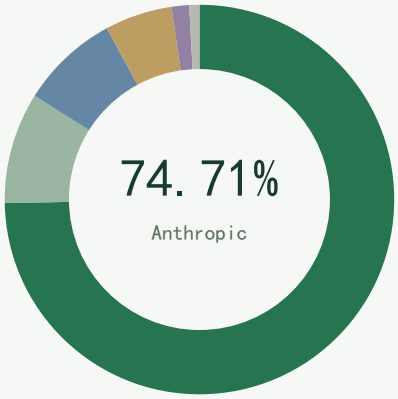
本期没有周环比。 日志始于北京时间 10 月 28 日 23:17，前一周数据不完整。本期作为完整自然周的观察基线。

访问来源 / WHO IS VISITING

Anthropic 占 74.71%，请求量与覆盖广度需分开看。

公司总量反映访问规模；爬虫名称与独立档案数，帮助识别不同的访问方式。

全周来源构成 / 846,846 次请求



Anthropic	632,680	74.71%
Amazon	78,186	9.23%
OpenAI	69,190	8.17%
Google	47,354	5.59%
Perplexity	12,301	1.45%
其他10家	7,135	0.84%

OpenAI 共 69,190 次：GPTBot 50,764 次，OAI-SearchBot 18,365 次，ChatGPT-User 61 次。Anthropic 中，ClaudeBot 为 632,668 次，Claude-User 为 12 次。

来源	匹配的独立档案数
Anthropic	189,178
Amazon	33,757
Google	8,533
Perplexity	3,734
OpenAI	1,562

档案数仅统计返回 2xx 且可匹配资料库的访问。同一档案可被多家公司访问，各来源的覆盖数不能相加作为全站总数。

爬虫的名称或用途分类，不代表内容实际进入训练或被引用。

档案覆盖 / BREADTH & OVERLAP

19 万份档案被访问，五分之一获得多来源访问。

40,692 份档案被至少两家公司访问，占已覆盖档案的 21.41%。



仅1家公司 149,351　　恰好2家公司 35,706　　至少3家公司 4,986

热门档案 x 来源公司 / 圆点表示本周有 2xx 访问

	Anthropic	Amazon	OpenAI	Google	Perplexity	Meta	Microsoft	ByteDance	Exa	Shape	请求数
亚马逊	●	●	●	●	●	●	●	●	●	●	498
谷歌	●	●	●	●	●	●	●	●	●	●	380
特斯拉	●	●	●	●	●	●	●	●	·	●	309
三星	●	●	●	·	●	·	●	·	●	●	293
苹果公司	●	●	●	●	●	●	●	●	·	●	260
三星电子	●	●	●	·	●	·	●	·	·	●	232
Amazon Web Services	●	●	●	●	●	·	●	·	·	●	204
环球音乐集团	●	●	●	·	●	●	●	●	·	●	197
微软	●	●	●	●	●	●	●	●	·	●	193
脸书	●	●	●	●	●	·	●	●	●	●	187

亚马逊、谷歌各获得十家公司爬虫访问。访问频次与来源多样性是不同指标，多来源访问并不意味着多家公司都引用或推荐该档案。

188,317 份档案本周首次出现于现有日志。 这一指标受采集初期影响，不等于首次被 AI 发现。三星与三星电子等条目保留资料库中的独立实体关系。

内容与行业 / WHAT GETS FETCHED

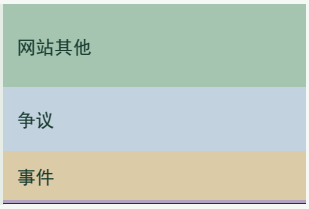
争议与事件内容，合计获得 10.50 万次请求。

争议路径 59,566 次，近期事件路径 45,419 次。路径分类包含失败响应，与档案分析的 2xx 口径不同。

其他 Brand Wiki 路径：661,393 次，占 78.10%，包含档案主页面及部分站点入口。下图单独比较余下 21.90% 的请求，使用独立线性刻度。

其他 Brand Wiki

78.10% / 661,393 次

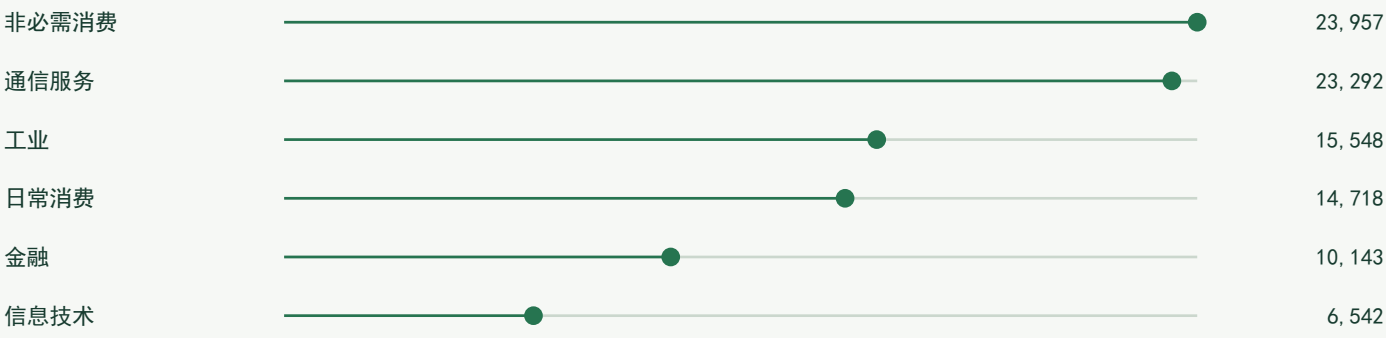


其余内容路径 / 请求数



Blog 涉及 28 条不同路径；/json、/md 涉及 97 条路径。路径数量不等于文章或档案数量；显式格式请求少，不代表爬虫不读取普通网页。

六个主要已分类行业 / 独立档案数



45.09% 的已覆盖档案归为“其他”或缺少分类。 各行业在库基数不同，本表仅描述访问分布，不能用于判断行业偏好。行业采用提取时资料分类。

访问质量与方法 / QUALITY & METHODS

错误率 0.122%，
错误请求几乎都指向不存在的页面。

1,034 次错误响应中，404 占约 99.42%。后续应排查旧链接、路径写法与不存在的地址。

错误组成 / 仅 1,034 次错误响应



404: 1,028 次 / 99.42%　　403: 6 次 / 0.58%

HTTP状态	含义	请求数
200	正常响应	845,705
308	重定向	106
304	未修改	1
404	未找到	1,028
403	禁止访问	6

下一周继续观察：集中抓取是否持续；本周覆盖的档案是否再次被访问；多家公司共同访问的档案数与占比是否增加。

统计方法与数据边界

时间与来源：北京时间 11/02 00:00 至 11/09 00:00，结束时点不计入。网站日志按 User-Agent 识别并去重，排除 SEO 工具。全站统计包含 GET、HEAD 及失败响应。

档案匹配：仅计返回 2xx 且匹配提取时资料库的访问，按档案标识合并路径。匹配档案请求共 747,883 次。另有 1,650 个未匹配路径键、涉及 4,101 次请求，未纳入档案与行业分析。

历史限制：日志始于北京时间 10/28 23:17，前周不完整，不计算周环比。缺少周末在库总量快照，不计算全库覆盖率。行业分类基于当前资料，不是历史快照。

解释边界：未保存来源身份校验结果。请求不是独立用户、AI 引用、训练采用或业务转化；首次记录不代表首次被发现。

快照与修订：数据于北京时间 2026/11/10 12:21 左右提取。后续去重与资料修订可能使历史数字小幅变化。