INDUSTRY OBSERVATION

MJ-GEO-RPT-2026-002

TILE GEO · 行业观察

33 家
0 家

瓷砖品牌官网在 AI 眼里长什么样：一次可复现的实测，和它不能证明的事

数据可复核 · 口径公开 · 欢迎引用

63	33	0
候选品牌官网域名	可正常访问（分母）	完成结构化改造

秒景美家

数据采集 2026-10-23 · 发布 2026-10-31

摘 要

五个可以直接引用的结论

- **63 个候选域名，33 个能正常打开。** 另外 30 个返回 502、请求超时，或已经停在域名交易平台的待售页上。它们不进分母——把打不开的站点算进「没做改造」是不诚实的。
- **33 家里，1 家有 JSON-LD 结构化数据，1 家有 canonical，0 家有 FAQPage，0 家有真实可用的 llms.txt。**
- **首页可见正文的中位数是 450 字。** 17 家不足 500 字，其中 7 家的可见正文是 0 字——页面只剩一张主图和几个脚本占位。
- **初测「有 llms.txt」的 6 个站点，逐个点开复核后全部为假阳性。** 软 404 会让所有自动化检测工具同时误判，包括市面上正在售卖的「AI 可见性检测」服务。
- **同时满足「有 JSON-LD、有 canonical、正文 H2 达 5 个」三项的：0 家。** 这个组合判据是本报告自己定的，不是行业标准，理由在第一节。

数据采集 2026 年 10 月 23 日，单一时点，无历史对比。以下所有数字都可以按第六节的方法自己复算一遍。

01

怎么测的，为什么这么测

先说清楚测的是什么。 本文所称「AI 可读性」，指在**不执行 JavaScript** 的前提下，服务端返回的原始 HTML 里，能被机器切分、识别类型、单独引用的内容量。它不是「AI 会不会推荐你」的同义词——后者还取决于内容质量、第三方提及与实体信息一致性，这一点在第五节会再说一次。所以这份报告测的是入场券，不是排位。

方法很笨，也可以被任何人复制：用 Python 对每个候选域名发一次真实 GET 请求，带浏览器 UA 和中文 Accept-Language，把服务端返回的原始 HTML 原样存下来，在**原始 HTML** 上统计指标。

全程不执行 JavaScript。 这一条决定了后面所有结论的性质。

今天多数 AI 爬虫不跑 JS，它们拿到的是服务端吐出来的那份 HTML。如果内容要靠脚本请求接口后再渲染，那么在 AI 眼里，那一页就是空的。所以「浏览器看起来正常」和「机器能读到」是两件不同的事——这也是不少品牌一边说「我们官网做得挺好看」，一边在 AI 里查无此人的原因。

十项判据，和那个组合判据

统计项包括：可见正文汉字数、h1 与 h2 数量、ld+json 块数、canonical、图片 alt 覆盖率、og 标签、FAQPage 标记、llms.txt（须人工复核内容有效性），以及正文长度的两档阈值。

「完成 GEO 改造」这一项，取的是三项同时满足：**JSON-LD + canonical + H2 达 5 个**。这不是行业标准，是本报告为这次扫描设的组合判据。之所以取这三项，是因为它们分别对应三件事——

1. 内容能不能被切开：正文有没有 H2 层级，机器能不能判断这一页在讲几件事；
2. 页面能不能被单独定位：有没有自指 canonical，同一内容不会在多个 URL 上互相稀释；
3. 内容能不能被识别类型：有没有 JSON-LD，机器知不知道「这一段」是公司、是文章、还是问答。

三项都缺，AI 读到的最多是一段没有边界的长文本。它可以被通读，但很难被切片、被单独引用。

02

数据：十项指标的命中情况

表 1　十项指标命中家数（2026-10-23，样本 n=33，百分比按 33 为分母四舍五入）

指标	判据	命中	占比
JSON-LD 结构化数据	原始 HTML 中存在 ld+json 块	1 / 33	3%
canonical 规范链接	存在 rel="canonical"	1 / 33	3%
FAQPage 结构化问答	存在 FAQPage 类型的标记	0 / 33	0%
llms.txt 站点说明	路径返回 200 且内容为真实说明	0 / 33	0%
og 社交标签	同时存在 og:title 与 og:description	3 / 33	9%
H2 小节达 5 个	正文 h2 元素数量 ≥ 5	2 / 33	6%
图片 alt 全覆盖	img 总数 > 0 且缺失 alt 数为 0	1 / 33	3%
正文超过 3000 字	剥离标签后的可见汉字数 > 3000	3 / 33	9%
正文超过 8000 字	剥离标签后的可见汉字数 > 8000	0 / 33	0%
完成 GEO 改造	上述三项同时满足	0 / 33	0%

这张表最该看的不是右列的低百分比，而是**绝对数字**：33 个样本里，多数指标的命中数连 3 家都不到。也就是说，这不是「做得好与做得不好」的分布，而是「几乎没做」与「零星做了半项」的分布。

© 2026 秒景美家 · 公开文件，欢迎引用

03

三个最值得看的发现

发现一：首页正文中位数 450 字，7 家是 0 字

33 家官网的首页可见正文，中位数是 450 字。17 家不足 500 字，其中 7 家的可见正文是 0 字——页面结构是一张主图加几个 JS 占位，文字全部在图片里，或者全部靠脚本加载。

这个数字比「有没有 JSON-LD」更值得警惕。没有结构化标记，机器至少还能读到文字；正文是空的，机器什么也读不到，后面所有优化都无处附着。你在官网投了多少钱做视觉，在 AI 眼里就有多安静。

发现二：初测「有 llms.txt」的 6 个站点，复核后全是假的

llms.txt 是给 AI 爬虫看的一份站点说明书，可以理解为写给机器人的 robots.txt 的邻居。初测有 6 个站点在这个路径上返回了 200，我们逐个点开复核，全部是假阳性，分三类：

- 站点对所有路径都返回 200，内容是「403 Forbidden」的错误页——服务器配置有问题，软 404 让检测工具全部误判；
- 站点是单页应用的空壳，任何路径都吐同一份首页 HTML；
- 域名已经停在交易平台的待售页上，品牌并不在使用。

真实可用的 llms.txt，33 家里是 0 个。

软 404 是这次扫描最大的干扰源。 它会让所有自动化检测同时误判——所以再强调一次：如果你的检测报告写着「我们有」，建议你手工在浏览器里打开那个地址看一眼。这一条不只针对 llms.txt，对任何「路径存在性」检测都成立。

发现三：指标最好的，不是声量最大的

那 1 家有 JSON-LD 的品牌，不是行业里声量最大的一家；反过来，几个声量最大的品牌，多项指标为零。

这说明官网的 AI 可读性跟市场规模、广告投放没有必然关系。它取决于官网在最近一次改版时，把什么当成目标：是把「好看」当目标，还是把「能被读到」当目标。前者是品牌部的语言，后者是工程的语言，这两件事在同一个页面上经常打架，而且过去十年一直是前者赢。

© 2026 秒景美家 · 公开文件，欢迎引用

04

为什么是零

不是不重视数字化。恰恰相反，这个行业的官网在过去十年被反复改版过，视觉一代比一代好。问题出在方向上：**这些官网是给人看的，不是给机器读的。** 具体是四种做法。

1. **首屏文字在图里。** 主视觉是一张轮播图，品牌主张、产品卖点、联系方式全部做进了图片，HTML 里只剩一个 img 标签。
2. **正文靠脚本加载。** 新闻列表、产品详情点进去，内容由 JS 请求接口后渲染，服务端返回的 HTML 里只有一个容器。
3. **没有标题层级。** 整页只有视觉字号，没有 h1 与 h2 的语义层级，机器无法判断这一页在讲几件事、每件事的边界在哪里。
4. **没有「这一段是什么」的标注。** 公司主体、产品、问答、面包屑全都靠人读，没有任何结构化声明。

对人来说，这套做法效果不错。对机器来说，等于把一本说明书做成了一本相册。**相册可以很好看，但无法被检索、被切片、被引用。**

八项动作，按投入产出排序

表 2　八项改造动作与可核验的验收标准

顺序	动作	验收标准（可核验）
1	每页一个 H1，正文按 H2 切到 5 个以上	源码中 h1 计数 = 1，h2 计数 ≥ 5
2	结论前置：每节第一段先给答案	抽查任意 3 个小节，首段是否为结论句
3	图片全部补 alt，产品图与场景图优先	img 总数 > 0 且缺 alt 数为 0
4	补 FAQ 模块，把客户真实问题写成一问一答	页面存在 ≥ 5 组显式问答，且与 FAQPage 标记逐字一致
5	加 JSON-LD：Organization、Article、FAQPage、BreadcrumbList	ld+json 块可解析，四类节点齐全，内容与可见正文一致
6	设置 canonical	每页存在自指 canonical，参数页指向主版本
7	加 llms.txt 与 llms-full.txt	路径返回 200，内容为真实说明（非错误页、非空壳）
8	全站主体信息统一	公司全称、地址、电话一处修改、全站同步、逐字一致

前四项不需要开发资源，只需要重排内容。成本不在开发，在内容重排——把「给人看的画册」改写成「能被切成片段的文字」。这一步没有现成模板可套，因为只有品牌自己知道产品到底好在哪、客户真正在问什么。

顺序不能反。这次扫描里那 1 家有 JSON-LD 的站点之所以没有进入「完成改造」，正是因为标记层做了、内容层是空的。**先有可读的文字，再谈标注。**

05

这份数据不能证明什么

一份只讲结论、不讲边界的报告不值得信，所以这一节和前面同等重要。

第一，30 家打不开的站点，我们没有任何判断。 502 和请求超时，可能是服务器拒绝了非浏览器请求，也可能是站点本身不稳定、域名闲置。把它们算进「没做改造」是不诚实的，所以它们不进分母。

第二，我们自己也是被测对象。 第 34 个样本就是本方官网，它不参与统计，也不能由此得出「我们比品牌做得好」——我们自己就是做这件事的。它只说明这套判据可行。自测读数：ld+json 1 块、canonical 已设置、h2 十个、可见正文 6174 字、5 张图里 4 张带 alt。缺的那 1 张 alt 我们没有为了凑分数临时补上，它在下一轮改版里处理。

第三，这份数据测的是「原始 HTML 的可读性」。 它不等于「AI 一定不推荐你」。AI 判断一个品牌，还看内容质量、第三方提及、实体信息一致性。**可读性是入场券，不是排名。** 任何只测可读性就敢宣称推荐率或排位的报告，都不值得信——包括我们自己出的任何一份。

第四，样本说明。 33 个样本是品牌官网域名，以行业里公开可查的品牌为主，包含主站、备用域名与部分品牌独立站；单一时点扫描，不代表整站水平，也不代表整个行业。它的价值在于给出一个可复现的判据和一组可核对的数字，而不是给行业下结论。

© 2026 秒景美家 · 公开文件，欢迎引用

06

复现方式与引用授权

判据、指标定义与统计口径全部公开，任何人都可以复现，也可以把自己的站点替换进样本：

1. 候选域名：品牌官网主站、备用域名、品牌独立站，共 63 个；
2. 请求方式：Python 逐站 GET，浏览器 UA + 中文 Accept-Language，超时按不可判定处理；
3. 读取对象：服务端返回的原始 HTML，不执行 JavaScript；
4. 统计与复核：按第二节十项口径统计；所有 llms.txt 阳性样本必须逐个打开核对内容有效性。

逐站明细含被测品牌域名，属于内部数据，不公开发布；对外只公开汇总口径与判据。想知道自己的站点在哪一项失分，用同一套判据自查即可。

引用与转载： 欢迎媒体、机构与个人引用本报告的数据与结论，注明来源「秒景美家 GEO 实测报告（2026 年 10 月）」并保留原始页链接即可，无需另行申请。数据采集 2026-10-23，报告发布 2026-10-31。

- 原始报告页：https://miaojingmeijia.com/geo/report-33/
- 方法与判据总览：https://miaojingmeijia.com/geo/
- 品牌官网首页：https://miaojingmeijia.com

关于作者： 张春雷，秒景美家创始人、首席研发官。

关于发布方： 秒景美家是秒景AI2.0 与秒景GEO 的品牌名。运营主体为美景家超算（深圳）科技有限公司，统一社会信用代码 91440300MAE5YUAX80。方法论先在自家官网跑通，再用于客户项目；本方官网作为第 34 个样本单列，标注为自测，不参与统计。

结语

这份报告最想留下的不是「0 家」这个数字，而是它为什么是 0：不是这 33 家不会做，是过去十年的改版目标里，从来没有「被机器读到」这一项。目标换了，动作自然会跟上——前四项改造不需要一行代码。

如果你担心自己的官网也在其中，最省力的自查不是再买一份检测报告，而是打开官网任意一页，右键看源码：能不能找到一个 H1，能不能数出五个 H2，正文在不在 HTML 里。三分钟就有答案，而且这个答案不需要任何人替你判定。

© 2026 秒景美家 · 公开文件，欢迎引用

附

常见问题与术语口径

前面六节先给结论、再给边界。这两组内容在讲这份报告时被问得最多，集中放在这里，也给引用者省一次来回。

三个常被问到的问题

问：这份报告是不是为了卖服务？

是，秒景美家做 GEO 服务，这份数据确实服务于这门生意。所以第五节把「不能证明什么」单独写了一整节：它不证明 AI 会推荐谁，不证明排名，30 家打不开的站点我们不做任何判断。判据、指标定义、统计口径全部公开，任何人都能把自己的站点替换进样本复算——如果数字对不上，先怀疑我们。

问：为什么不测用户实际看到的页面，只测原始 HTML？

因为不执行 JavaScript 是多数 AI 爬虫当前的默认行为，它们拿到的就是服务端吐出来的那一份。浏览器渲染出来的效果是给人的，HTML 里有什么才是给机器的。测渲染后的页面，测的是人的体验；测原始 HTML，测的才是可读性。这两个数在同一个站点上可以差出几千字。

问：我的官网按这套判据测出来也是零，第一步该做什么？

先看正文在不在 HTML 里，别先谈结构化。打开任意一页，右键查看网页源代码，搜你首页最想被引用的那句话：搜得到，说明文字在 HTML 里，再往下做 H2 与标记；搜不到，先解决「文字在图里」或「内容靠脚本加载」——这两件事不解决，后面七项改造都无处附着。顺序不能反，这也是表 2 把内容动作排在前四位的原因。

五个术语的本文口径

术语	本文口径
AI 可读性	不执行 JS 时，原始 HTML 里可被切分、识别、引用的内容量
软 404	任意路径都返回 200，内容是错误页或首页副本
JSON-LD	页面内声明的结构化数据，用来标明「这一段是什么」
canonical	页面自指的规范链接，避免同一内容在多 URL 上稀释
llms.txt	给 AI 爬虫的站点说明文件，内容真实有效才算命中

© 2026 秒景美家 · 公开文件，欢迎引用

秒景美家

可读性是入场券

不是排名

本报告数据与结论欢迎引用，注明来源即可

miaojingmeijia.com

© 2026 秒景美家 · MJ-GEO-RPT-2026-002