发现鸿沟

HoProduct Hunt 初创企业为何会在大语言模型的发现查询中消失 s

生成式引擎优化（*GEO*）研究

Amit Prakash Sharma印度理
工学院巴特那分校，2025年12月

摘要

当有人请 ChatGPT 推荐一款项目管理工具时，哪些产品会出现在回答中？更重要的是，对于初创企业创始人来说，他们新推出的产品有机会被推荐吗？本研究旨在回答这些问题。

我从 2025年 Product Hunt 排行榜上展示的前 500 款产品中随机选取了 112 家初创企业，并针对两个不同的大语言模型——ChatGPT（gpt-4o-mini）和 Perplexity（支持网络搜索的 sonar）——分别进行了 2,240 次查询测试。

结果令人震惊。当用户按名称询问产品时，两个大语言模型几乎都能完美识别。ChatGPT 的识别率为 99.4%，Perplexity 为 94.3%。但当用户提出类似“今年推出的最佳人工智能工具有哪些？”这类发现式问题时，成功率骤降至 3.32% 和 8.29%。ChatGPT 的差距高达 30 倍。

也许最令人意外的发现是，生成式引擎优化（GEO）——即针对人工智能可见度优化网站内容的做法——与实际发现率毫无关联。GEO 分数较高的产品，在自然查询中出现的可能性并不高于分数较低的产品。

什么因素确实重要？对于 Perplexity，传统的搜索引擎优化信号，例如引荐域名（$r = +0.319$，$p < 0.001$）和 Product Hunt 排名（$r = -0.286$，$p = 0.002$），可以预测其可见度。清除 Reddit 数据中的假阳性后，社区影响力也显现出显著性（$r = +0.395$，$p = 0.002$）。

实际的结论有些反直觉：不要直接针对人工智能发现进行优化。相反，应先打好搜索引擎优化基础，大语言模型可见度自然会随之提升。

关键词：生成式引擎优化、大语言模型、产品发现、初创公司可见度、人工智能搜索

符号和缩略语列表

LLM	大语言模型
GEO	生成式引擎优化
SEO	搜索引擎优化
POTD	每日产品
API	应用程序编程接口
r	皮尔逊相关系数
p	统计显著性（p 值）
NS	不显著
n	样本量

第 1 章

引言

1.1 问题

人工智能驱动的搜索正在发生一些奇怪的变化。问 ChatGPT“Notion 是什么？”，你会得到一份详细且准确的回答。但如果问“有哪些最好的笔记应用？”，Notion 甚至可能不会出现在答案中。我着手研究的，正是这种认知与推荐之间的差距。

这种转变意义重大，因为人们发现产品的方式正在改变。如果这一趋势持续下去——而且有充分理由相信它会持续——那么，了解产品如何出现在人工智能的回答中，就成了商业上的必要之举，而不仅仅是学术上的好奇。

自 Brin 和 Page 发布 PageRank 论文以来，传统搜索引擎优化已经得到详尽研究。我们知道 Google 如何对网页进行排名，也知道什么能让网站具有权威性。但大语言模型的运作方式不同。它们不会返回按排名排列的链接列表，而是生成综合性的回答。规则有所不同，而且我们尚未完全弄清这些规则。

1.2 为什么初创企业面临更棘手的问题

初创企业在这方面处于特别不利的地位。我们来看看其中的运作机制：

首先，问题出在训练数据上。ChatGPT 的知识有一个截止日期。在该日期之后推出的任何产品，在模型的世界里都不存在。无论进行多少优化，都无法改变这一根本事实。

其次，即使是截止日期之前推出的产品，训练数据也会天然地过度代表已经站稳脚跟的企业。维基百科上有大量关于 Salesforce 的条目，却没有关于上个月才推出的客户关系管理工具的条目。这就形成了我所说的“权威集中”效应。就人工智能可见度而言，富者愈富。

第三，像 Perplexity 这样的网络搜索增强模型带来了不同的动态。它们可以通过实时搜索发现新产品，但搜索引擎优化因素又开始发挥作用。这是一场规则不同的游戏。

1.3 研究问题

本研究探讨了六个问题：

研究问题1：直接查询（按产品名称进行询问）与发现查询（产品可能自然出现的类别搜索）之间的差距有多大?

研究问题 2：Product Hunt 指标（如赞成票、评论和每日排名）是否与大语言模型可见度相关?

RQ3：正如先前研究所示，生成式引擎优化是否确实能提高发现率?

RQ4：传统的搜索引擎优化信号对人工智能可见度仍然重要吗?

RQ5：与知识截止模型相比，网络搜索大语言模型在发现新产品方面的表现好多少?

RQ6：来自 Reddit、GitHub 和 Hacker News 的社区信号能否预测大语言模型可见性?

第 2 章

文献综述

2.1 从搜索引擎优化到生成式引擎优化的演变

搜索引擎优化在 20 世纪 90 年代末发展成为一门专业领域，此时距离网络搜索成为主流信息发现方式仅过去不久。Brin 和 Page 在 PageRank 方面的研究揭示了一个奠基性洞见：基于链接的权威信号能够决定页面排名 [2]。在之后的几十年里，搜索引擎优化发展成一项复杂的实践，涵盖内容优化、技术因素和站外信号。2017 [3] 中基于 Transformer 的语言模型问世并逐渐扩大规模，带来了真正崭新的事物。与返回按排名排列的文档列表的搜索引擎不同，大语言模型会生成综合性的回答。用户在 Google 上搜索“最佳客户关系管理软件”，会看到十条蓝色链接；用户向 ChatGPT 提问，则会得到一段可能点名提及三四款产品的文字。两者的可见度机制截然不同。

2.2 生成式引擎优化

Aggarwal 及其同事在 2024 年的论文中提出了“生成式引擎优化”这一术语[1], , 并针对 10,000 个查询测试了九种内容优化策略。他们的研究结果表明，引文、统计数据、技术术语和权威性语言等某些信号，可以将内容在搜索增强生成系统中的可见度提高 30–40%。但其中有个关键问题，作者只是简略地提及。他们的研究聚焦于已出现在搜索结果中的信息类内容，测试的也是那些本就能够被发现的内容所适用的优化策略。问题在于，这些策略是否有助于那些一开始就无法被发现的产品。这正是大多数新兴初创企业面临的处境，也亟待解决。

2.3 将 Product Hunt 作为研究背景

Product Hunt 为研究可见度提供了一个独特的自然实验场。据估计，该平台每月访问量在 293 万至 330 万人次之间。在平台上发布的每款产品都能获得相同的初始曝光度。它们通过社区投票和每日排行榜争夺关注。最终形成了一组成功指标（赞成票、评论、排名）各不相同、但发布环境相似的产品。

第 3 章

研究方法

3.1 数据集构建

我从 2025年 Product Hunt 排行榜上精选的前 500 款产品中随机抽取了 112 款。筛选标准很简单：至少获得 200 张赞成票（以确保有意义的社区互动）、拥有可供分析的正常运行网站，以及提供英语内容。我排除了网站无法访问或自发布以来已大幅转型的产品。

对于每款产品，我收集了：Product Hunt 指标（赞成票、评论、每日产品排名、创作者人数）、来自 DataForSEO 的技术 SEO 信号（引荐域名、Dofollow 比率、域名权威度）、社区影响力（Reddit提及、唯一子版块、GitHub 星标、Hacker News 积分），以及基于网站内容分析得出的生成式引擎优化综合得分。

表 3.1：数据集摘要统计数据

指标	数值
分析的产品	112（从前500个中随机选取）
执行的查询总数	2,240
每个产品的直接查询	3
每个产品的发现式查询	7
测试的大语言模型	ChatGPT (gpt-4o-mini)、Perplexity (sonar)

3.2 查询设计

直接查询与发现查询之间的区别是本研究的核心。直接查询会明确提及产品名称：“什么是[ProductName]？”或“介绍一下[ProductName]。”这些查询用于测试大语言模型是否知道该产品的存在。

发现查询以类别为基础，例如：“2025年推出的最佳人工智能编程助手有哪些？”或“推荐一些新的生产力工具。”这些查询用于测试用户没有专门寻找某款产品时，该产品是否会自然地出现在结果中。

我为每个产品使用了十个查询：三个直接查询和七个发现查询。完整的查询模板见附录 C。

3.3 GEO 分数构建

遵循 Aggarwal 等人 [1], 提出的框架，我构建了一个综合 GEO 分数，用于衡量六个维度：网站上的统计数据、引用密度、技术术语使用情况、权威语言模式、结构化数据实施和内容深度。每个维度的评分范围为 0–100 分，综合分数则取六个维度的平均值。

评分主要基于正则表达式，查找数字后跟百分号、引用标记和技术术语等模式。这种方法显然存在局限性，无法捕捉真正经过优化的内容所蕴含的细微差别。不过，它为 GEO 文献指出的重要信号类型提供了合理的近似衡量方式。

3.4 大语言模型测试协议

我测试了代表不同架构方法的两个大语言模型：

ChatGPT（gpt-4o-mini）：一种知识截止模型。它只了解训练时掌握的信息，仅此而已。这代表了“纯粹”的大语言模型情形。API 文档见[5]。

Perplexity（搭载网络搜索的 sonar）：一种搜索增强模型。它会实时查询网络，并将最新信息纳入回答中。这代表了这样一种混合情况：传统网页信号可能仍然重要。API 文档见 [5]。所有查询均于2025年12月15日至20日期间通过 API 执行。所有调用均使用一致的参数：温度为 0.7，未修改系统提示。

第4章

结果

4.1 可见度差距（研究问题1）

这些数字比我预想的更能说明问题。两个大语言模型对直接查询的识别率都接近完美：ChatGPT 为 99.4%，Perplexity 为 94.3%。如果你直接说出某个产品的名称，这些模型就知道它的存在。

但是，发现查询呢？ChatGPT 仅在 3.32% 的情况下展示了产品。Perplexity 的表现更好，达到 8.29%，但这一比例仍然低得惊人。ChatGPT 的差距比为 30:1，Perplexity 的差距比为 11:1。

表 4.1：按查询类型划分的可见性比率

指标	ChatGPT	Perplexity
直接查询成功率	99.4%（334/336）	94.3%（317/336）
发现成功率	3.32% (26/784)	8.29% (65/784)
可见度差距	**30:1**	**11:1**
曾被发现的产品	6 (5.4%)	31（27.7%）

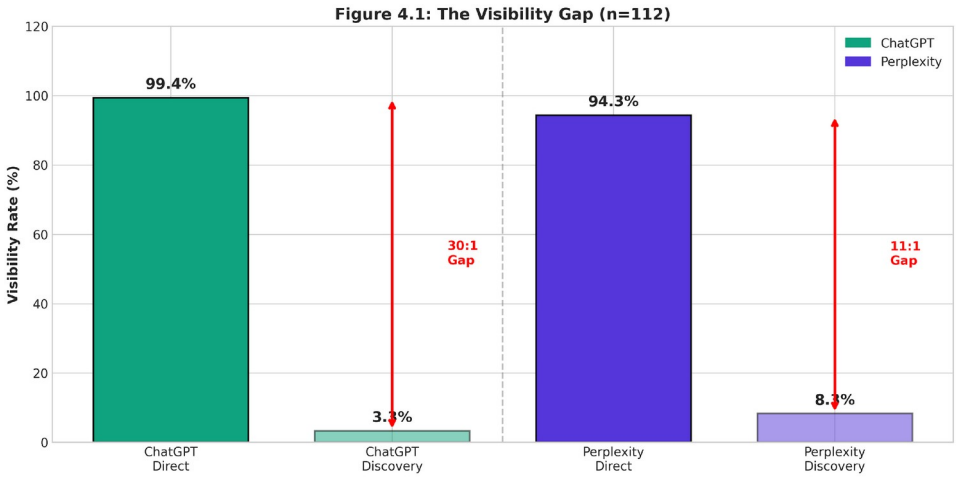

图 *4.1*：直接查询与发现查询之间的可见度差距

4.2 Product Hunt 信号（研究问题 2）

Product Hunt 指标呈现出明显的模式：它们对 Perplexity 有影响，但对 ChatGPT 没有影响。

对于 Perplexity发现，每日产品排名呈现显著负相关（r = - 0.286，p = 0.002），这意味着排名越靠前的产品可见度越高。赞成票

呈正相关（r = +0.225，p = 0.017），产品评分也呈正相关（r = +0.187，p = 0.048）。

对于 ChatGPT，这些相关性均未达到显著水平。该模型似乎基本上随机地对待各个产品；考虑到其训练数据的局限性，这一点并不意外。

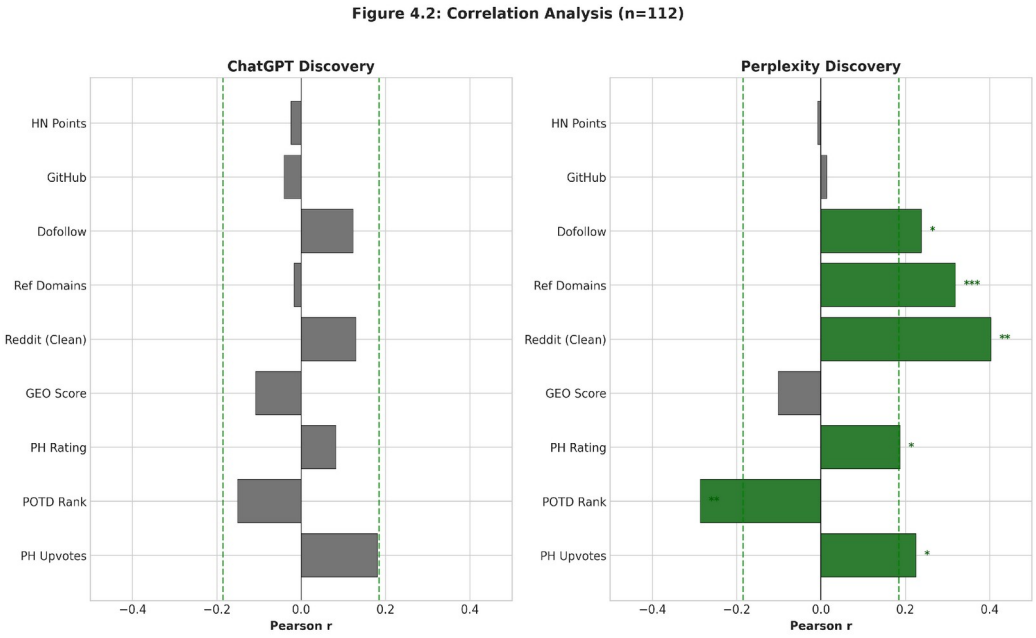

图 *4.2*：相关性分析显示 *Perplexity* 对各项信号的响应程度

4.3 生成式引擎优化带来的意外发现（RQ3）

有趣的地方就在这里。旨在衡量人工智能优化效果的 GEO 综合得分，与可见度之间并无显著相关性。对于 ChatGPT，结果如此（r = -0.108，p = 0.256）；对于 Perplexity，也同样如此（r = -0.102，p = 0.286）。

这与 Aggarwal 等人研究的启示直接矛盾 [1]。不过，我认为这种矛盾是可以解释的。他们的研究测试了优化对已出现在搜索增强响应中的内容所产生的影响。而我的数据捕捉的是更早期的阶段：产品试图首先获得可见度。

看起来，生成式引擎优化发挥的是乘数作用，而非催化剂作用。如果你已经能被发现，经过优化的内容或许能帮你更频繁地出现。但如果你根本无法被发现——我的样本中大多数情况都是如此——那就没有任何东西可以放大。零乘以任何数都还是零。

4.4 技术性搜索引擎优化 (RQ4)

传统 SEO 信号在 Perplexity 的数据中重新发挥了作用。引荐域名成为发现可见度最强的预测变量（r = +0.319，p < 0.001）。Dofollow 比率也表现出显著性（r = +0.238，p = 0.012）。

这在直觉上说得通。Perplexity 会实时搜索网络，因此，能让页面在传统搜索中获得较高排名的因素，也会提高页面被 Perplexity 找到并引用的可能性。

Figure 4.3: Key Scatter Plots

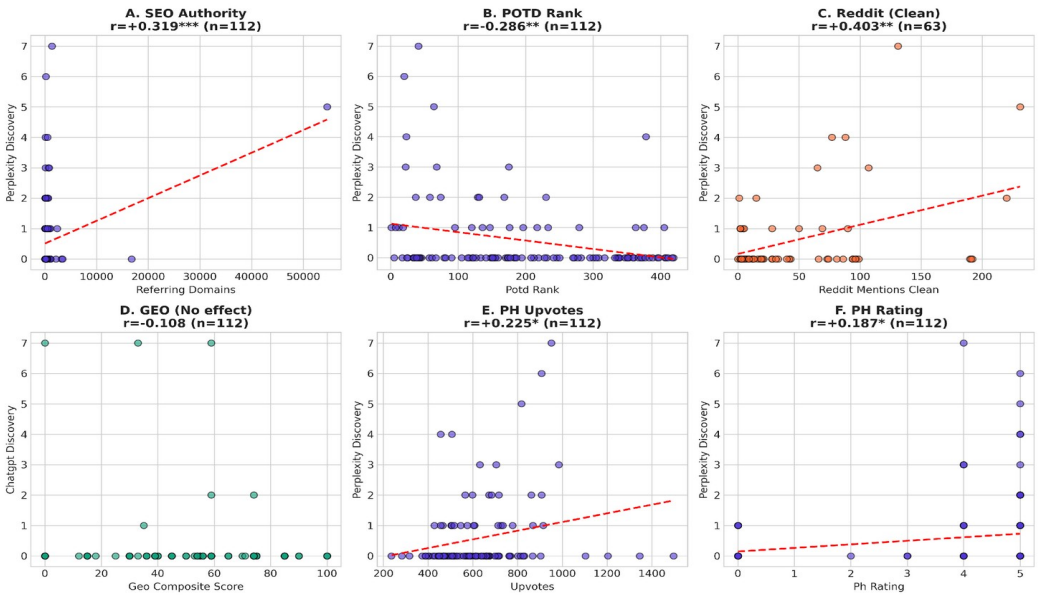

图 *4.3*：预测因素与发现可见度之间的关键关系

4.5 网络搜索优势（RQ5）

Perplexity 的网络搜索使其发现率比 ChatGPT 高出 2.5 倍（8.29% 对 3.32%）。但或许更重要的是，两种架构在可预测性方面表现出显著差异。

对于 ChatGPT：没有显著相关性。发现基本上看似随机。对于 Perplexity：存在七项显著相关性。发现遵循可识别的模式。

这种“架构鸿沟”确实会带来实际影响。如果你想优化大语言模型的发现效果，瞄准网络搜索模型，就能找到可以着手调整的杠杆。而瞄准知识截止模型嘛……我想，大概只能寄希望于运气了。

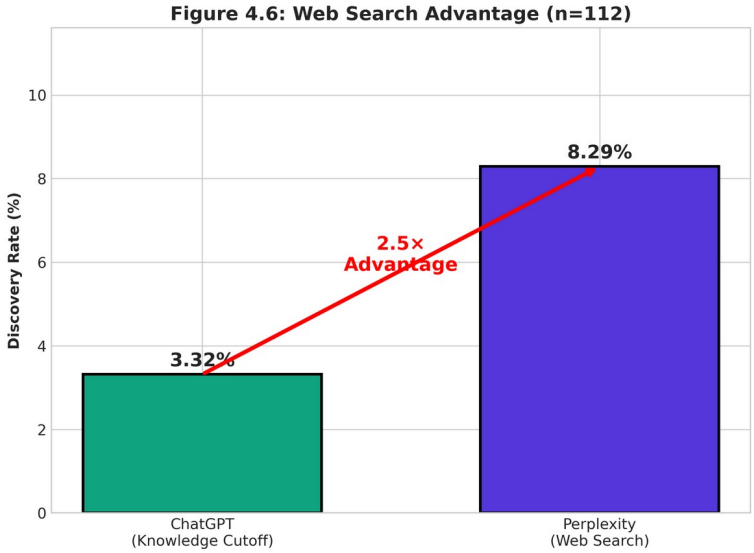

图 *4.4*：网络搜索模型的发现率高出 *2.5* 倍

4.6 社区信号（RQ6）

要弄清社区信号背后的情况，我们费了一番侦查功夫。初步分析显示，Reddit提及与可见度之间没有相关性（r = +0.052，p = 0.586）。但总觉得哪里不对劲。

查看数据时，我注意到一些名称通用的产品：“Cursor”、“DROP”、“Solar”，它们的 Reddit 提及次数极高，却没有任何发现成功。Reddit 搜索抓取到了无关帖子：人们讨论鼠标光标、音乐发行和太阳能板。

识别并剔除 52 个容易产生假阳性的通用名称产品（占样本的 46%）后，情况发生了显著变化：

Reddit提及（清理后）：r = +0.395，p = 0.002

唯一子版块（已清理）：r = +0.405，p = 0.001

社区影响力确实很重要，但前提是测量准确。GitHub 星标和 Hacker News 在两项分析中均未显示出显著影响。

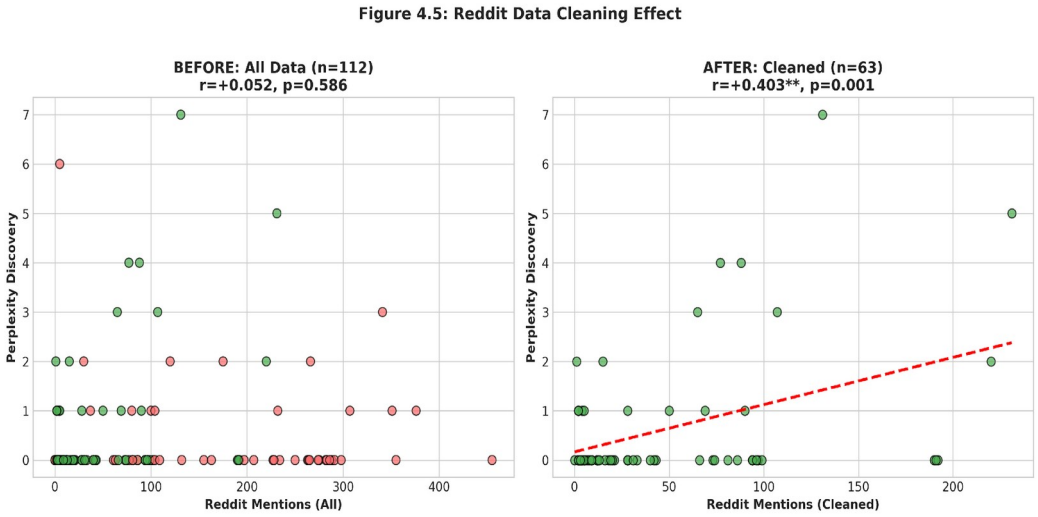

图 *4.5*：*Reddit* 数据清洗揭示了隐藏信号

第5章

讨论

5.1 理解鸿沟

30:1 的可见度差距不只是一个统计数据，更是初创企业创始人需要理解的战略现实。造成这一差距的似乎有三种机制：

时效性壁垒：2025年1月推出的产品不可能出现在 ChatGPT 的训练数据中。这是一个硬性限制，无法通过优化来克服。

权威集中：大语言模型的训练会自然地过度偏重权威来源的内容。训练数据中关于 Slack 的文本，比关于今年推出的任何竞争产品的文本都要多。

查询架构：直接查询会激活事实检索机制（“X 是什么？”），而发现查询则会激活带有不同偏见的推荐系统。

5.2 GEO悖论

对于 GEO 未发现显著效果这一结果，需要谨慎解读。我并不是说生成式引擎优化不起作用。Aggarwal研究 [1] 提供了有力证据，表明它在某些情境下确实有效。

我的主张范围更窄：对于试图提升人工智能可见度的新产品而言，生成式引擎优化并无帮助，因为根本性的可发现性壁垒依然存在。这就像给一辆困在车库里的汽车打磨抛光。抛光或许做得非常好，但汽车仍无法驶上道路。

这表明可以采取分阶段的方法：先通过搜索引擎优化、社区影响力和时间来建立可发现性，然后在确实能够被发现之后，再针对人工智能可见度优化内容。

5.3 架构分野的含义

ChatGPT（0 个显著预测因子）与 Perplexity（7 个显著预测因子）之间的鲜明差异，指向了一项实用策略：将资源集中在网络搜索大语言模型上。

对于 Perplexity 等类似模型，行之有效的方法并不陌生：增加引荐域名、在 Product Hunt 上取得成功，并在 Reddit 上推动真实的讨论。这些活动一直是传统营销取得成功的关键，如今还额外有助于提升人工智能可见度。

对于知识截止模型，除了等待之外，我们无能为力。随着这些模型更新训练数据，产品会逐渐进入它们的知识库。但创始人无法直接影响这一进程的时间安排。

5.4 局限性

这些发现有几项需要注意的地方：

样本完全来自 Product Hunt，因此偏向某些产品类别（开发者工具、生产力、人工智能）。消费类产品、企业软件或通过其他渠道发布的产品，结果可能有所不同。

我测试了两个大语言模型。Claude、Gemini 等模型可能会呈现不同的模式。

Reddit 数据清理剔除了 46% 的样本。尽管这是确保准确性所必需的，但这会降低社区信号分析的统计功效。

GEO评分基于正则表达式。更精细的内容分析或许能捕捉到我的方法遗漏的信号。

第六章

启示

6.1 面向创始人和营销人员

不要指望人工智能发现。ChatGPT 上 3% 的发现率意味着，在 100 次相关查询中，有 97 次不会提及你的产品。请据此规划市场进入策略。

投资于 SEO 基础。引荐域名是最强的预测变量。那些看似枯燥的反向链接建设工作依然重要。

争取在 Product Hunt 上取得成功。每日产品排名和赞成票与 Perplexity 上的可见度相关。这个平台不仅是发布渠道，也是可见度信号。

建立真实的社区影响力。在相关 Reddit 子版块中开展讨论（而非伪装成用户发布垃圾信息）与发现相关。质量比数量更重要。

降低早期生成式引擎优化的优先级。数据表明，在达到可发现性阈值之前，优化并无助益。

瞄准网络搜索大语言模型。Perplexity 等模型不仅带来更高的发现率，也提供了可预测的优化杠杆。

6.2 研究人员

本研究为未来工作开辟了若干方向：

纵向追踪：随着产品逐渐成熟并不断积累网络存在度，其可见度将如何变化？

成功案例分析：在 ChatGPT 中获得发现的 6 款产品，与未获得发现的 106 款产品有何不同？

跨模型比较：Claude、Gemini 和其他大语言模型的发现模式有何不同？

生成式引擎优化的生命周期：生成式引擎优化何时开始带来回报？

更好的 Reddit 方法：开发能够避免通用名称问题的实体感知搜索方法。

第7章

结论

这项研究记录了大语言模型发现产品时存在的显著可见度差距。对从 1825 个 Product Hunt 初创企业中排名前 500 的企业里随机选取的 112 家进行测试，并针对 2,240 个查询进行分析后发现：当被直接询问时，大语言模型能够识别产品（99.4%）；但它们很少会自发推荐这些产品（ChatGPT 为 3.32%，Perplexity 为 8.29%）。

ChatGPT 的 30:1 差距代表着一道结构性壁垒，初创企业无法直接克服。旨在提升人工智能可见度的一系列技术，即生成式引擎优化，与发现成功率并无相关性，这表明它发挥的是乘数效应，必须以已有的可见度基础为前提才能奏效。

Perplexity 等网络搜索增强模型带来了更高的发现率，也提供了明确可行的优化路径。引荐域名、Product Hunt 上的成功表现以及经过清理的 Reddit 形象，都能预测可见度。知识截止型大语言模型与网络搜索大语言模型之间的架构鸿沟，成为本研究的核心发现，并具有实际意义。

对初创企业创始人而言，这一战略启示有悖直觉：不要直接针对人工智能发现进行优化。应夯实传统基础，包括搜索引擎优化权威度、社区影响力和平台上的成功。人工智能可见度自然会随之而来。不能本末倒置。

附录 A

研究问题摘要

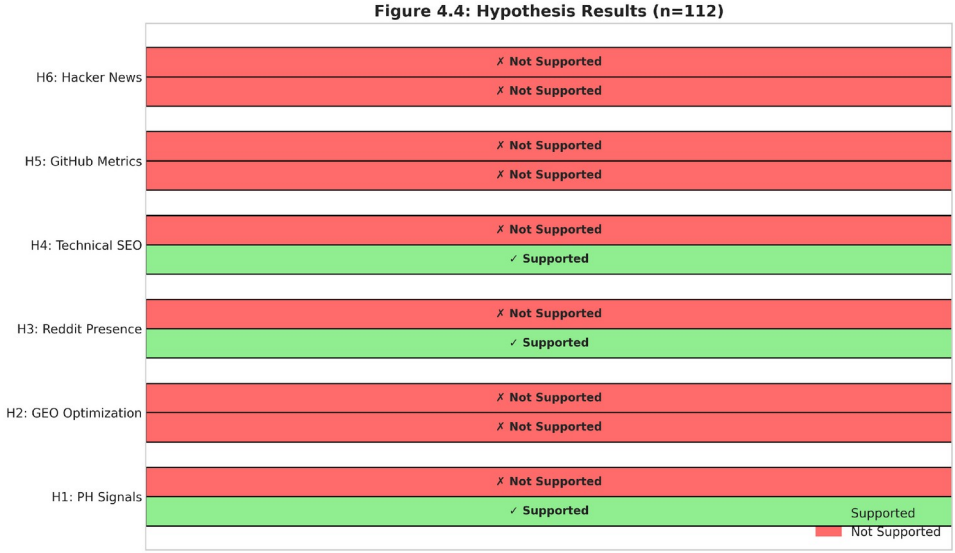

图 *A.1*：研究问题结果的可视化摘要

表 A.1：研究问题结果

RQ	问题	发现
RQ1	可见性差距幅度	30:1（ChatGPT），11:1 (Perplexity)
RQ2	PH 指标与可见度	支持（仅限 Perplexity）
RQ3	生成式引擎优化效果	不支持（r = -0.10）
RQ4	技术 SEO 信号	支持（r = +0.319***）
RQ5	网页搜索优势	发现效果提升 2.5 倍
RQ6	社区信号	支持 Reddit（已清理）

附录B

相关矩阵

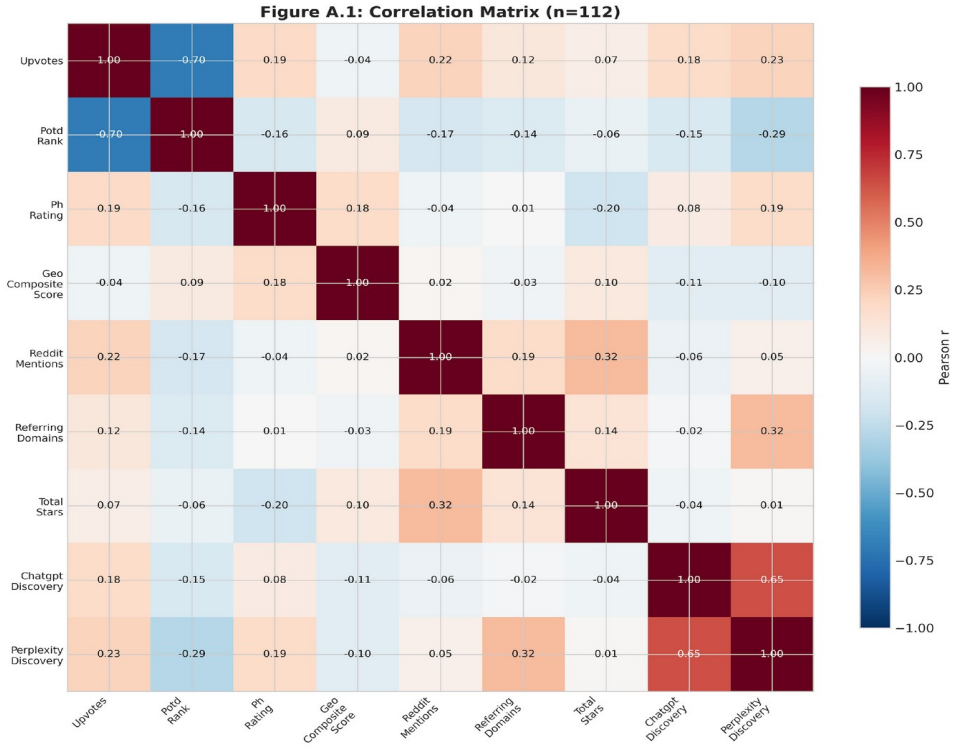

图 *B.1*：研究变量的完整相关矩阵

表 B.1：显著相关性（p < 0.05，Perplexity发现）

预测变量	r	p	n
唯一子版块（清理后）	+0.405	0.001	60
Reddit提及（已清理）	+0.395	0.002	60
引荐域名	+0.319	<0.001	112
每日产品排名	-0.286	0.002	112
Dofollow 比率	+0.238	0.012	112
PH 赞成票	+0.225	0.017	112
PH 评分	+0.187	0.048	112

附录 C

查询模板

以下查询模板用于测试每款产品。[产品名称] 和[类别] 分别替换为实际产品名称及其主要类别。

C.1 直接查询（每个产品 3 条）

D1：“[产品名称]是什么？”

D2：“请介绍一下[产品名称]”

D3：「你听说过[产品名称]吗？它是做什么的？」

C.2 发现查询（每个产品 7 条）

Q1：「2025年推出的最佳[类别] 工具有哪些？」

Q2：「推荐一些新的[类别] 产品」

Q3：「有哪些[类别] 初创企业值得我了解？」

Q4：「我正在寻找[类别] 解决方案。有哪些选择？」

Q5：“目前有哪些热门的[类别] 工具？”

Q6：“你能推荐一些创新的[类别] 产品吗？”

Q7：“最近有哪些新的[类别] 工具受到关注？”

C.3 响应评估

针对每条查询，均使用产品名称的精确字符串匹配来评估回复。如果大语言模型的输出中任何位置出现了产品名称，就将该回复标记为成功。为确保结果可复现并避免主观解读，我们采用了这种二元判断方式（提及／未提及），而非更细致的指标。

参考文献

(1) Aggarwal, P., Murahari, V., Rajpurohit, T., Narasimhan, K., & Deshpande, A. (2024). 生成式引擎优化：生成式引擎优化。arXiv 预印本。https://arxiv.org/abs/2311.09735

(2) Brin, S., & Page, L. (1998). 大规模超文本网络搜索引擎的剖析。Computer Networks and ISDN Systems, 30(1-7), 107-117.

(3) Vaswani, A., Shazeer, N., Parmar, N., 等。(2017). Attention is all you need. 神经信息处理系统进展, 30. https://arxiv.org/abs/1706.03762

(4) OpenAI. (2024). OpenAI API 文档。https://platform.openai.com/docs
(5) Perplexity AI. (2024). Perplexity API 文档。https://docs.perplexity.ai