GEO: Generative Engine Optimization

Pranjal Aggarwal*
Indian Institute of Technology Delhi
New Delhi, India
pranjal2041@gmail.com

Vishvak Murahari*
Princeton University
Princeton, USA
murahari@cs.princeton.edu

Tanmay Rajpurohit
Independent
Seattle, USA
tanmay.rajpurohit@gmail.com

Ashwin Kalyan
Independent
Seattle, USA
asaavashwin@gmail.com

Karthik Narasimhan
Princeton University
Princeton, USA
karthikn@princeton.edu

Ameet Deshpande
Princeton University
Princeton, USA
asd@princeton.edu

ABSTRACT

The advent of large language models (LLMs) has ushered in a new paradigm of search engines that use generative models to gather and summarize information to answer user queries. This emerging technology, which we formalize under the unified framework of generative engines (GEs), can generate accurate and personalized responses, rapidly replacing traditional search engines like Google and Bing. Generative Engines typically satisfy queries by synthesizing information from multiple sources and summarizing them using LLMs. While this shift significantly improves *user* utility and *generative search engine* traffic, it poses a huge challenge for the third stakeholder – website and content creators. Given the black-box and fast-moving nature of generative engines, content creators have little to no control over *when* and *how* their content is displayed. With generative engines here to stay, we must ensure the creator economy is not disadvantaged. To address this, we introduce Generative Engine Optimization (GEO), the first novel paradigm to aid content creators in improving their content visibility in generative engine responses through a flexible black-box optimization framework for optimizing and defining visibility metrics. We facilitate systematic evaluation by introducing GEO-bench, a large-scale benchmark of diverse user queries across multiple domains, along with relevant web sources to answer these queries. Through rigorous evaluation, we demonstrate that GEO can boost visibility by up to 40% in generative engine responses. Moreover, we show the efficacy of these strategies varies across domains, underscoring the need for domain-specific optimization methods. Our work opens a new frontier in information discovery systems, with profound implications for both developers of generative engines and content creators.[1]

CCS CONCEPTS

• **Computing methodologies** → *Natural language processing*; Machine learning; • **Information systems** → **Web searching and information discovery**.

KEYWORDS

generative models, search engines, datasets and benchmarks

ACM Reference Format:
Pranjal Aggarwal, Vishvak Murahari, Tanmay Rajpurohit, Ashwin Kalyan, Karthik Narasimhan, and Ameet Deshpande. 2024. GEO: Generative Engine Optimization. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '24), August 25–29, 2024, Barcelona, Spain.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3637528.3671900

1 INTRODUCTION

The invention of traditional search engines three decades ago revolutionized information access and dissemination globally [4]. While they were powerful and ushered in a host of applications like academic research and e-commerce, they were limited to providing a list of relevant websites for user queries. However, the recent success of large language models [5, 21] has paved the way for better systems like BingChat, Google's SGE, and perplexity.ai that combine conventional search engines with generative models. We dub these systems generative engines (GE) because they *search* for information and *generate* multi-modal responses by using multiple sources. Technically, generative engines (Figure 2) retrieve relevant documents from a database (like the internet) and use large neural models to generate a response grounded on the sources, ensuring attribution and a way for the user to verify the information.

The usefulness of generative engines for developers and users is evident – users access information faster and more accurately, while developers craft precise and personalized responses, improving user satisfaction and revenue. However, generative engines disadvantage the third stakeholder – website and content creators. Generative Engines, in contrast to traditional search engines, remove the need to navigate to websites by directly providing a precise and comprehensive response, potentially reducing organic traffic to websites and impacting their visibility [16]. With millions of small businesses and individuals relying on online traffic and visibility for their livelihood, generative engines will significantly disrupt the creator economy. Further, the black-box and proprietary nature of generative engines makes it difficult for content

*Equal Contribution
[1]Code and Data available at https://generative-engines.com/GEO/

KDD '24, August 25–29, 2024, Barcelona, Spain

ACM ISBN 979-8-4007-0490-1/24/08
https://doi.org/10.1145/3637528.3671900

GEO：生成式引擎优化

普兰贾尔·阿加瓦尔
印度理工学院德里分校
印度新德里，
pranjal2041@gmail.com

世界之冠
普林斯顿大学 美国普林斯顿
murahari@cs.princeton.edu

坦迈·拉杰普罗希特
独立
美国西雅图
tanmay.rajpurohit@gmail.com

阿什温·卡利安
独立
美国西雅图
asaavashwin@gmail.com

卡尔蒂克·纳拉西姆汉
普林斯顿大学 美国普林斯顿
karthikn@princeton.edu

阿米特·德什潘德
普林斯顿大学 美国普林斯顿
asd@princeton.edu

抽象

大型语言模型（LLM）的出现，开启了一种利用生成式模型收集和总结信息以回答用户查询的新搜索引擎范式。这项新兴技术，我们在生成引擎（GE）统一框架下正式化，能够生成准确且个性化的响应，迅速取代谷歌和必应等传统搜索引擎。生成引擎通常通过综合多个来源的信息并用大型语言模型进行总结来满足查询。虽然这一转变显著提升了用户的实用性和生成性搜索引擎流量，但对第三个利益相关者——网站和内容创作者——构成了巨大挑战。鉴于生成引擎的黑箱和快速变化特性，内容创作者几乎无法控制内容何时以及如何展示。随着生成引擎的持续存在，我们必须确保创造者经济不会处于劣势。为了解决这个问题，我们正在——

介绍第一部小说《生成引擎优化（GEO）》
该范式旨在帮助内容创作者通过灵活的黑箱优化框架，提升生成引擎响应中的内容可见度，以优化和定义可见度指标。我们通过引入 GEO-bench——一个涵盖多个领域多样化用户查询的大型基准工具，以及相关网络资源来回答这些问题，促进系统性评估。通过严格评估，我们证明 GEO 在生成发动机响应中可提升多达 40%的可见度。此外，我们展示了这些策略的有效性在不同领域间存在差异，强调了针对特定领域优化方法的必要性。我们的工作开辟了信息发现系统的新领域，对生成引擎开发者和内容创作者都具有深远影响。

平等贡献 1 代码和数据可于 https://generative-engines.com/GEO/ 获取

KDD '24,2024 年 8 月 25 日至 29 日，西班牙巴塞罗那
ACM ISBN 979-8-4007-0490-2008 年 1 月 24 日 https://doi.org/10.1145/3637528.3671900

CCS 概念

• 自然语言处理→计算方法论；机器学习；• 信息系统→网络搜索和信息发现。

关键字

生成模型、搜索引擎、数据集和基准测试

ACM 参考格式：
普兰贾尔·阿加瓦尔、维什瓦克·穆尔哈里、坦迈·拉杰普罗希特、阿什温·卡利安、卡尔蒂克·纳拉西姆汉和阿米特·德什潘德。2024 年。GO：生成引擎优化。载于第 30 届 ACM SIGKDD 知识发现与数据挖掘会议论文集（KDD '24），2024 年 8 月 25 日至 29 日，西班牙巴塞罗那。ACM，纽约，美国，12 页。https://doi.org/10.1145/3637528。3671900

1 简介

三十年前，传统搜索引擎的发明彻底改变了全球信息获取和传播方式[4]。虽然它们很强大，带来了学术研究和电子商务等多种应用，但仅限于提供相关网站列表以供用户查询。然而，大型语言模型的近期成功[5， 21]为更好的系统铺平了道路，如 BingChat、谷歌的 SGE 以及结合传统搜索引擎与生成模型的 perplexity.ai。我们称这些系统为生成引擎（GE），因为它们通过多元来源搜索信息并生成多模态响应。从技术上讲，生成引擎（图 2）从数据库（如互联网）检索相关文档，并利用大型神经模型生成基于来源的响应，确保归属并供用户验证信息。

生成引擎对开发者和用户的实用性显而易见——用户获取信息更快更准确，开发者则设计精准且个性化的回答，提升用户满意度和收入。然而，生成式引擎对第三个利益相关者——网站和内容创作者——造成了不利影响。与传统搜索引擎不同，生成式引擎通过直接提供精确且全面的响应，消除了访问网站的需求，这可能减少网站的自然流量并影响其可见性[16]。数百万小企业和个人依赖网络流量和可见度维生，生成式引擎将极大地颠覆创造者经济。此外，生成引擎的黑箱和专有特性使内容难以生成

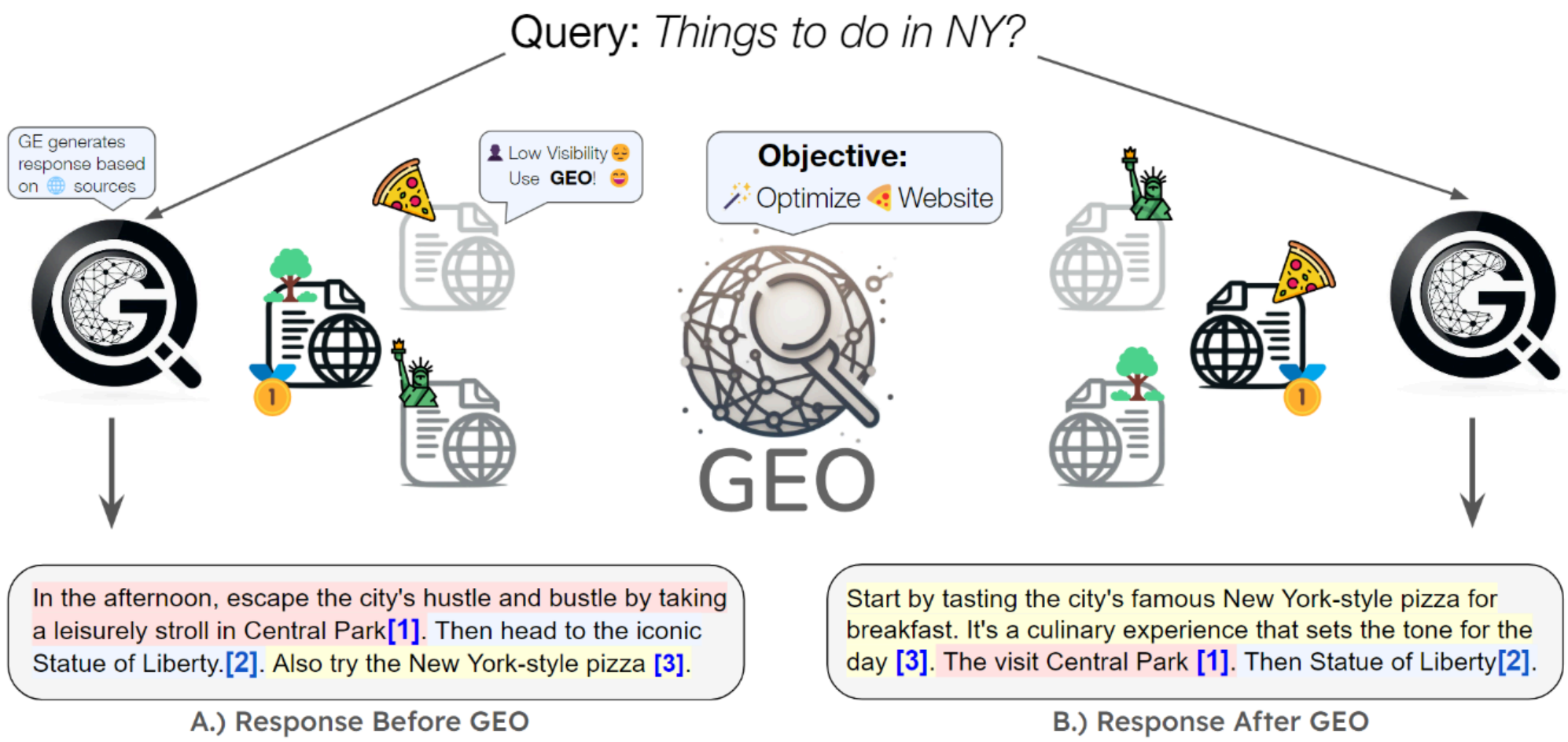

Figure 1: Our proposed Generative Engine Optimization (GEO) method optimizes websites to boost their visibility in Generative Engine responses. GEO's black-box optimization framework then enables the website owner of the pizza website, which lacked visibility originally, to optimize their website to increase visibility under Generative Engines. Further, GEO's general framework allows content creators to define and optimize their custom visibility metrics, giving them greater control in this new emerging paradigm.

creators to *control* and *understand* how their content is ingested and portrayed.

In this work, we propose the first general creator-centric framework to optimize content for generative engines, which we dub Generative Engine Optimization (GEO), to empower content creators to navigate this new search paradigm. GEO is a flexible black-box optimization framework for optimizing web content visibility for proprietary and closed-source generative engines (Figure 1). GEO ingests a source website and outputs an optimized version by tailoring and calibrating the presentation, text style, and content to increase visibility in generative engines.

Further, GEO introduces a flexible framework for defining visibility metrics tailor-made for generative engines as the notion of visibility in generative engines is more nuanced and multi-faceted than traditional search engines (Figure 3). While average ranking on the response page is a good measure of visibility in traditional search engines, which present a linear list of websites, this does not apply to generative engines. Generative Engines provide rich, structured responses and embed websites as inline citations in the response, often embedding them with different lengths, at varying positions, and with diverse styles. This necessitates the need for visibility metrics tailor-made for generative engines, which measure the visibility of attributed sources over multiple dimensions, such as relevance and influence of citation to query, measured through both an objective and a subjective lens.

To facilitate faithful and extensive evaluation of GEO methods, we propose GEO-bench, a benchmark consisting of 10000 queries from diverse domains and sources, adapted for generative engines.

Through systematic evaluation, we demonstrate that our proposed Generative Engine Optimization methods can boost visibility by up to 40% on diverse queries, providing beneficial strategies for content creators. Among other things, we find that including citations, quotations from relevant sources, and statistics can significantly boost source visibility, with an increase of over 40% across various queries. We also demonstrate the efficacy of Generative Engine Optimization on Perplexity.ai, a real-world generative engine and demonstrate visibility improvements up to 37%.

In summary, our contributions are three-fold:

(1) We propose Generative Engine Optimization, the first general optimization framework for website owners to optimize their websites for generative engines. Generative Engine Optimization can improve the visibility of websites by up to 40% on a wide range of queries, domains, and real-world black-box generative engines.

(2) Our framework proposes a comprehensive set of visibility metrics specifically designed for generative engines and enables content creators to flexibly optimize their content through customized visibility metrics.

(3) To foster faithful evaluation of GEO methods in generative engines, we propose the first large-scale benchmark consisting of diverse search queries from wide-ranging domains and datasets specially tailored for Generative Engines.

图 1：我们提出的生成引擎优化（GEO）方法通过优化网站，提升其在生成引擎响应中的可见度。GEO 的黑箱优化框架使得原本缺乏可见性的披萨网站所有者能够优化其网站，以提升生成式引擎下的可见度。此外，GEO 的通用框架允许内容创作者定义和优化其自定义可见度指标，从而在这一新兴范式中拥有更大的控制权。

创作者能够控制并理解他们的内容如何被吸收和展示。

在本研究中，我们提出了第一个以创作者为中心的通用框架，用于生成引擎内容优化，我们称之为
生成引擎优化（GEO），帮助内容创作者驾驭这一新的搜索范式。GEO 是一个灵活的黑箱优化框架，用于优化专有和闭源生成引擎的网页内容可见性（见图 1）。GEO 会接收源网站，并通过定制和校准展示、文本风格和内容，输出优化版本，以提升生成引擎中的可见度。

此外，GEO 引入了一个灵活的可视化指标框架，专为生成式引擎量身打造，因为生成引擎中的可见性概念比传统搜索引擎更为细致和多面（见图 3）。虽然在传统搜索引擎中，响应页面的平均排名是衡量可见性的良好指标，因为搜索引擎呈现的是线性网站列表，但生成式搜索引擎则不适用这一点。生成引擎提供丰富、结构化的回答，并将网站嵌入为内嵌引用，通常以不同长度、不同位置和多样风格嵌入。这需要为生成引擎量身定制的可见度指标，这些指标通过客观和主观视角衡量归属来源在多个维度上的可见性，如引用对查询的相关性和影响。

为了促进对 GEO 方法的忠实和广泛评估，我们提出了 GEO-bench 基准测试，该基准测试包含来自不同领域和来源的 1 万条查询，适用于生成引擎。

通过系统评估，我们展示了提出的生成引擎优化方法能够在多样化查询中提升高达 40%的可见度，为内容创作者提供有益策略。我们发现，包含引用、相关来源的引用和统计数据可以显著提升来源可见度，在各种查询中增长超过 40%。我们还展示了生成引擎优化在 Perplexity.ai 这一真实生成引擎上的有效性，并展示了可视度提升高达 37%。

总之，我们的贡献有三方面：(1) 我们提出生成式引擎优化（Generative Engine Optimization），这是首个面向网站所有者优化生成引擎网站的通用优化框架。生成式引擎优化可以将网站在各种查询、域名和现实黑箱生成引擎上的可见度提升高达 40%。

(2) 我们的框架提出了一套专为生成引擎设计的全面可见度指标，使内容创作者能够通过定制的可见度指标灵活优化内容。(3) 为促进生成引擎中 GEO 方法的忠实评估，我们提出了首个由广泛领域和数据集中专门针对生成引擎量身定制的多样化搜索查询组成的大规模基准测试。

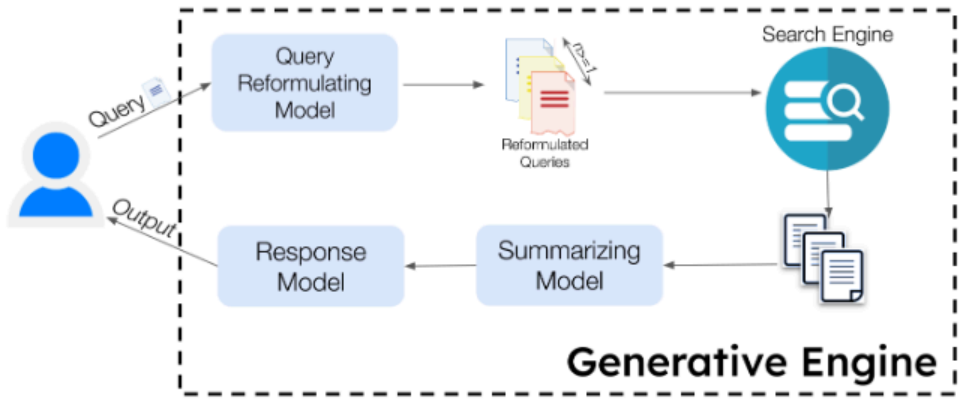

Figure 2: Overview of Generative Engines. Generative Engines primrarily consists of a set of generative models and a search engine to retrieve relevant documents. Generative Engines take user query as input and through a series of steps generate a final response that is grounded in the retrieved sources with inline attributions.

2 FORMULATION & METHODOLOGY

2.1 Formulation of Generative Engines

Despite the deployment of numerous generative engines to millions of users, there is currently no standard framework. We provide a formulation that accommodates various modular components in their design. We describe a generative engine, which includes several backend generative models and a search engine for source retrieval. A Generative Engine (GE) takes a user query q_u and returns a natural language response r, where P_U represents personalized user information. The GE can be represented as a function:

$$f_{GE} := (q_u, P_U) \rightarrow r \tag{1}$$

Generative Engines comprise two crucial components: a.) A set of generative models $G = \{G_1, G_2 ... G_n\}$, each serving a specific purpose like query reformulation or summarization, and b.) A search engine SE that returns a set of sources $S = \{s_1, s_2 ... s_m\}$ given a query q. We present a representative workflow in Figure 2, which, at the time of writing, closely resembles the design of BingChat. This workflow breaks down the input query into a set of simpler queries that are easier to consume for the search engine. Given a query, a query re-formulating generative model, $G_1 = G_{qr}$, generates a set of queries $Q^1 = \{q_1, q_2 ... q_n\}$, which are then passed to the search engine SE to retrieve a set of ranked sources $S = \{s_1, s_2, ..., s_m\}$. The sets of sources S are passed to a summarizing model $G_2 = G_{sum}$, which generates a summary Sum_j for each source in S, resulting in the summary set ($Sum = \{Sum_1, Sum_2, ..., Sum_m\}$). The summary set is passed to a response-generating model $G_3 = G_{resp}$, which generates a cumulative response r backed by sources S. In this work, we focus on single-turn Generative Engines, but the formulation can be extended to multi-turn Conversational Generative Engines (Appendix A).

The response r is typically a structured text with embedded citations. Citations are important given the tendency of LLMs to hallucinate information [10]. Specifically, consider a response r composed of sentences $\{l_1, l_2 ... l_o\}$. Each sentence may be backed by a set of citations that are part of the retrieved set of documents $C_i \subset S$. An ideal generative engine should ensure all statements in the response are supported by relevant citations (high citation recall), and all citations accurately support the statements they're associated with (high citation precision) [14]. We refer readers to Figure 3 for a representative generative engine response.

2.2 Generative Engine Optimization

The advent of search engines led to search engine optimization (SEO), a process to help website creators optimize their content to improve search engine rankings. Higher rankings correlate with increased visibility and website traffic. However, traditional SEO methods are not directly applicable to Generative Engines. This is because, unlike traditional search engines, the generative model in generative engines is not limited to keyword matching, and the use of language models in ingesting source documents and response generation results in a more nuanced understanding of text documents and user query. With generative engines rapidly emerging as the primary information delivery paradigm and SEO is not directly applicable; new techniques are needed. To this end, we propose Generative Engine Optimization, a new paradigm where content creators aim to increase their visibility (or impression) in generative engine responses. We define the visibility of a website (also referred to as a citation) c_i in a cited response r by the function $Imp(c_i, r)$, which the website creator wants to maximize. From the generative engine's perspective, the goal is to maximize the visibility of citations most relevant to the user query, i.e., maximize $\sum_i f(Imp(c_i, r), Rel(c_i, q, r))$, where $Rel(c_i, q, r)$ measures the relevance of citation c_i to the query q in the context of response r and f is determined by the exact algorithmic design of generative engine and is a black-box function to end-users. Further, both the functions Imp and Rel are subjective and not well-defined yet for generative engines, and we define them next.

2.2.1 Impressions for Generative Engines. In SEO, a website's impression (or visibility) is determined by its average ranking over a range of queries. However, generative engines' output nature necessitates different impression metrics. Unlike search engines, Generative Engines combine information from multiple sources in a single response. Factors such as length, uniqueness, and presentation of the cited website determine the true visibility of a citation. Thus, as illustrated in Figure 3, while a simple ranking on the response page serves as an effective metric for impression and visibility in conventional search engines, such metrics are not applicable to generative engine responses.

In response to this challenge, we propose a suite of impression metrics designed with three key principles in mind: 1.) The metrics should hold relevance for creators, 2.) They should be explainable, and 3.) They should be easily comprehensible by a broad spectrum of content creators. The first of these metrics, the "Word Count" metric, is the normalized word count of sentences related to a citation. Mathematically, this is defined as:

$$Imp_{wc}(c_i, r) = \frac{\sum_{s \in S_{c_i}} |s|}{\sum_{s \in S_r} |s|} \tag{2}$$

Here S_{c_i} is the set of sentences citing c_i, S_r is the set of sentences in the response, and $|s|$ is the number of words in sentence s. In cases where a sentence is cited by multiple sources, we share the word count equally with all the citations. Intuitively, a higher word count correlates with the source playing a more important part in the answer, and thus, the user gets higher exposure to that source.

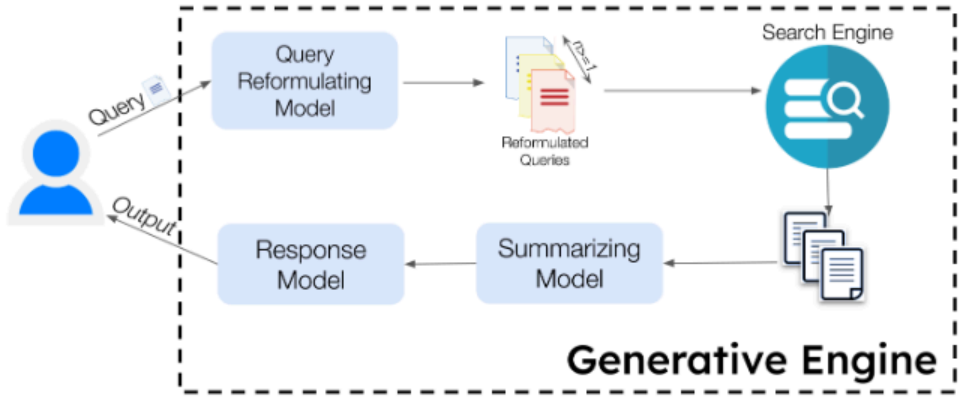

图 2：生成引擎概述。生成引擎主要由一组生成模型和用于检索相关文档的搜索引擎组成。生成引擎将用户查询作为输入，通过一系列步骤生成基于检索来源并附有内嵌归因的最终响应。

2 表述与方法 2.1 生成引擎的表述

尽管向数百万用户部署了大量生成引擎，但目前尚无标准框架。我们提供一种方案，在设计中兼容各种模块化组件。我们描述了一个生成引擎，包含多个后端生成模型和用于源检索的搜索引擎。生成引擎（GE）对用户查询 q 并返回自然语言响应 r，其中 P 代表个性化用户信息。GE 可以表示为一个函数：

$$f: = (q, P) \rightarrow r \quad (1)$$

生成引擎由两个关键组成部分组成：a.）一组生成模型 G = {G, G...G}，每个角色都有特定用途，如查询重述或总结，b.）一个返回一组来源的搜索引擎 SE S = {s， s...给定查询 Q。我们在图 2 中展示了一个代表性的工作流程，在撰写时，其设计与 BingChat 非常相似。该工作流程将输入查询拆解成一组更简单的查询，以便搜索引擎更容易理解。给定一个查询，重新表述生成模型的查询 G= G 生成一组查询 Q= {q， q...q}，然后传递给搜索引擎 SE，以检索一组排名来源 S = {s， s， ...， s}。源集合 S 传递给汇总模型 G= G，G=G 生成 S 中每个源的摘要和，结果为

摘要集（Sum = {Sum， Sum， ...， Sum}）。摘要集传递给一个响应生成模型 G= G，该模型生成一个累计响应 r，并由源 S 支持。本研究重点介绍单回合生成引擎，但该表述可扩展至多回合对话生成引擎（附录 A）。

回复 r 通常是带有嵌入引用的结构化文本。鉴于大型语言模型倾向于产生幻觉信息[10]，引用非常重要。具体来说，考虑一个由句子 {I, I... 组成的反应 r...I}. 每个句子都可以有一组引用，这些引用是检索到的文档集 C⊂S 的一部分。理想的生成引擎应确保回复中的所有陈述均有相关引用支持（高引用回忆率）

所有引用都准确支持其相关陈述（高引用精度）[14]。我们建议读者参考图 3 以获取具有代表性的生成引擎响应。

2.2 生成式引擎优化

搜索引擎的出现催生了搜索引擎优化（SEO），这是一种帮助网站创建者优化内容以提升搜索引擎排名的过程。排名越高，可见度和网站流量越高。然而，传统的 SEO 方法并不直接适用于生成式引擎。这是因为与传统搜索引擎不同，生成引擎中的生成模型不限于关键词匹配，且在导入源文档和生成响应时使用语言模型，使得对文本文档和用户查询的理解更加细致。随着生成引擎迅速成为主要的信息传递范式，SEO 并不直接适用;需要新的技术。为此，我们提出了生成引擎优化（Generative Engine Optimization），这是一种新范式，内容创作者旨在提升他们在生成引擎响应中的可见度（或曝光量）。我们通过函数 Imp（c， r）定义网站（也称为引用）的可见性（也称为引用），并定义了引用的响应 r，网站创建者希望最大化该功能。从生成引擎的角度来看，目标是最大化与用户查询最相关的引用的可见性，即最大化

$$\sum_{i} Imp(c_i, r), Rel(c_i, q, r))$$

，其中 Rel （c， q， r ） 衡量引用 cto 查询 q 在响应 r 和 f 的相关性，由生成引擎的精确算法设计确定，是终端用户的黑箱函数。此外，Imp 和 Rel 这两个功能都是主观的，对于生成引擎来说尚未明确定义，接下来我们将对它们进行定义。

2.2.1 生成引擎的印象。在 SEO 中，网站的 im–按压（或可见性）由其在一系列查询中的平均排名决定。然而，生成引擎的输出特性需要不同的印象指标。与搜索引擎不同，生成式引擎将多个来源的信息整合到一个回复中。引用的长度、唯一性和呈现方式等因素决定了引用的真实可见度。因此，如图 3 所示，虽然在响应页面上简单排名是传统搜索引擎中展示和可见度的有效指标，但这些指标不适用于生成引擎的响应。

针对这一挑战，我们提出了一套基于三个关键原则设计的曝光量指标：1.） 这些指标应对创作者具有相关性，2.） 它们应当易于解释，3.） 它们应被广泛的内容创作者易于理解。其中第一个指标是“字数”指标，是与引用相关的句子的归一化字数。数学上，定义为：

$$Imp(c, r) = \sum_{s \in S|s|} \frac{|s|}{\sum_{s \in S|s|}} \quad (2)$$

这里 Sis 是引用 c 的句子集合，Sis 是响应中的句子集合，|s |是句子 s 中的单词数。当句子被多个引用来源时，字数与所有引用平均分配。直观上，字数越多，来源在回答中扮演更重要角色，因此用户接触该来源的机会也越多。

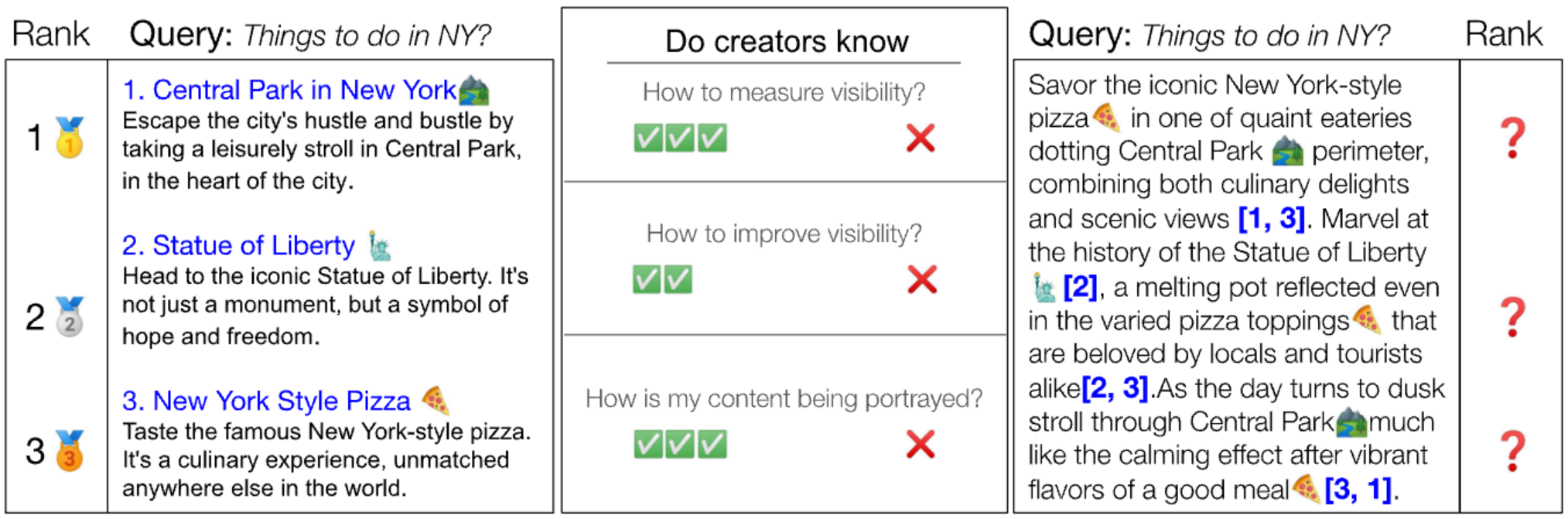

Figure 3: Ranking and Visibility Metrics are straightforward in traditional search engines, which list website sources in ranked order with verbatim content. However, Generative Engines generate rich, structured responses, often embedding citations in a single block interleaved with each other. This makes ranking and visibility nuanced and multi-faceted. Further, unlike search engines, where significant research has been conducted on improving visibility, optimizing visibility in generative engine responses remains unclear. To address these challenges, our black-box optimization framework proposes a series of well-designed impression metrics that creators can use to *gauge* and *optimize* their website's performance and also allows the creator to define their impression metrics.

However, since "Word Count" is not impacted by the ranking of the citations (whether it appears first, for example), we propose a position-adjusted count that reduces the weight by an exponentially decaying function of the citation position:

$$Imp_{pwc}(c_i, r) = \frac{\sum_{s \in S_{c_i}} |s| \cdot e^{-\frac{pos(s)}{|S|}}}{\sum_{s \in S_r} |s|} \tag{3}$$

Intuitively, sentences that appear first in the response are more likely to be read, and the exponent term in definition Imp_{pwc} gives higher weightage to such citations. Thus, a website cited at the top may have a higher impression despite having a lower word count than a website cited in the middle or end of the response. Further, the choice of exponentially decaying function is motivated by several studies showing click-through rates follow a power-law as a function of ranking in search engines [7, 8]. While the above impression metrics are objective and well-grounded, they ignore the subjective aspects of the impact of citations on the user's attention. To address this, we propose the "Subjective Impression" metric, which incorporates facets such as the relevance of the cited material to the user query, influence of the citation, uniqueness of the material presented by a citation, subjective position, subjective count, probability of clicking the citation, and diversity in the material presented. We use G-Eval [15], the current state-of-the-art for evaluation with LLMs, to measure each of these sub-metrics.

2.2.2 Generative Engine Optimization methods for website. To improve impression metrics, content creators must make changes to their website content. We present several generative engine-agnostic strategies, referred to as Generative Engine Optimization methods (GEO). Mathematically, every GEO method is a function $f : W \rightarrow W_i'$, where W is the initial web content, and W' is the modified content after applying the GEO method. The modifications can range from simple stylistic alterations to incorporating new content in a structured format. A well-designed GEO is equivalent to a black-box optimization method that, without knowing the exact algorithmic design of generative engines, can increase the website's visibility and implement textual modifications to W independent of the exact queries.

For our experiments, we apply Generative Engine Optimization methods on website content using a large language model, prompted to perform specific stylistic and content changes to the website. In particular, based on the GEO method defining a specific set of desired characteristics, the source content is modified accordingly. We propose and evaluate several such methods:

1: Authoritative: Modifies text style of the source content to be more persuasive and authoritative, **2. Statistics Addition:** Modifies content to include quantitative statistics instead of qualitative discussion, wherever possible, **3. Keyword Stuffing:** Modifies content to include more keywords from the query, as expected in classical SEO optimization. **4. Cite Sources & 5. Quotation Addition:** Adds relevant citations and quotations from credible sources respectively, 6.) **6. Easy-to-Understand:** Simplifies the language of website, while **7. Fluency Optimization** improves the fluency of website text. **8. Unique Words & 9. Technical Terms:** involves adding unique and technical terms respectively wherever possible,

These methods cover diverse general strategies that website owners can implement quickly and use regardless of the website content. Further, except for methods 3, 4, and 5, the remaining methods enhance the presentation of existing content to increase its persuasiveness or appeal to the generative engine, without requiring extra content. On the other hand, methods 3,4 and 5 may

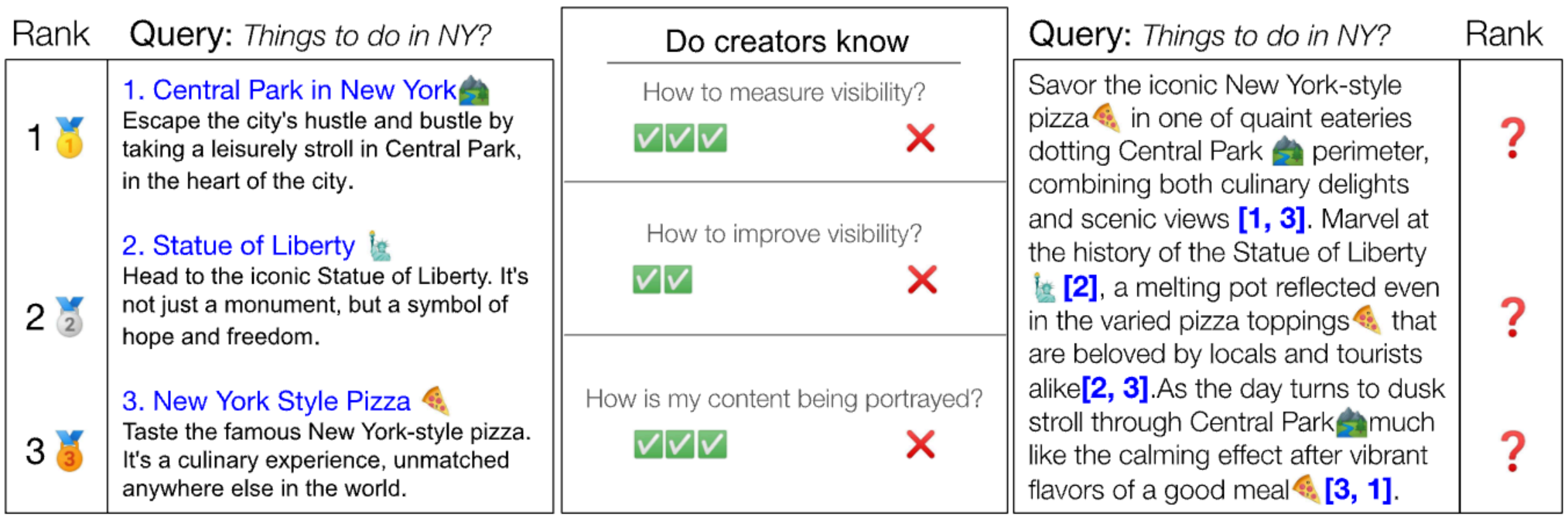

图 3：排名和可见度指标在传统搜索引擎中很直观，它们按排名顺序列出网站来源，内容逐字呈现。然而，生成引擎生成丰富且结构化的响应，通常将引用嵌入一个交错的块中。这使得排名和可见度变得细致且多面。此外，与搜索引擎不同，搜索引擎在提升可见性方面已有大量研究，生成引擎响应的可见度优化仍不明确。为应对这些挑战，我们的黑箱优化框架提出了一系列设计良好的展示量指标，供创作者用来评估和优化网站表现，同时也允许创作者定义自己的展示量指标。

然而，由于“字数”不受引用排名（例如是否排在前）影响，我们提出一个位置调整计数，通过引文位置的指数衰减函数减少权重：

$$\text{Imp}(c, r) = \frac{\sum_{s \in S_{|s|}} \cdot e^{-|}}{\sum_{s \in S_{|s|}}} \quad (3)$$

直观上，回答中最先出现的句子更可能被阅读，且定义中的指数项赋予此类引用更高的权重。因此，排名顶部的网站即使字数较少，也可能比回答中段或末尾引用的网站获得更高的曝光率。此外，选择指数衰减函数的动机是多项研究表明点击率在搜索引擎排名中遵循幂律的[7， 8]。虽然上述印象指标客观且有充分依据，但它们忽略了引用对用户注意力影响的主观因素。为此，我们提出了“主观印象”指标，涵盖被引用材料与用户查询的相关性、引用的影响、引用材料的独特性、主观位置、主观数量、点击引用的概率以及所呈现材料的多样性等方面。我们使用 G 评估[15]，这是当前用于大型语言模型评估的先进标准，来衡量这些子指标。

2.2.2 生成式引擎优化网站方法。为了提升展示量指标，内容创作者必须对其网站内容进行调整。我们提出了几种生成式无关引擎策略，称为生成引擎优化方法（GEO）。数学上，每个地理学方法都是函数 f： W → W ，其中 W 是初始网页内容 W 是应用 GEO 方法后的修改内容。修改可以从简单的风格调整到以结构化格式加入新内容不等。一个设计良好的地理环境相当于一种黑箱优化方法，即使不了解生成引擎的具体算法设计，也能提升网站的可见度，并在独立于具体查询的情况下对 W 进行文本修改。

在我们的实验中，我们对网站内容应用生成引擎优化方法，使用大型语言模型，针对网站进行特定的风格和内容调整。特别是，基于 GEO 方法定义的特定期望特征，源内容会相应地进行修改。我们提出并评估了几种此类方法：1. 权威性：修改源内容的文本风格，使其更具说服力和权威性;2. 统计补充：尽可能修改内容，包含定量统计数据而非定性讨论;3. 关键词填充：修改内容，包含更多来自查询的关键词，符合传统 SEO 优化的预期。4. 引用资料 & 5.引用补充：分别添加相关引用和可信来源的引用，6.）易懂：简化网站语言，7. 流畅度优化提升网站文本的流畅度。8. 独特词汇 & 9.技术术语：在可能的情况下分别添加独特和技术术语，这些方法涵盖了网站所有者可以快速实施并无论网站内容如何都能使用的多种通用策略。此外，除方法 3、4 和 5 外，其余方法增强现有内容的呈现效果，以提高其说服力或吸引生成引擎，而无需额外内容。另一方面，方法 3、4 和 5 可以

require some form of additional content. To analyze the performance gain of our methods, for each input user query, we randomly select one source website to be optimized and apply each of the GEO methods separately on the same source. We refer readers to Appendix B.4 for more details on GEO methods.

3 EXPERIMENTAL SETUP

3.1 Evaluated Generative Engine

In accordance with previous works [14], we use a 2-step setup for Generative Engine design. The first step involves fetching relevant sources for input query, followed by a second step where an LLM generates a response based on the fetched sources. Similar to previous works, we do not use summarization and provide the whole response for each source. Due to context length limitations and quadratic scaling cost based on the context size of transformer models, only the top 5 sources are fetched from the Google search engine for every query. The setup closely mimics the workflow used in previous works and the general design adopted by commercial GEs such as you.com and perplexity.ai. The answer is then generated by the gpt3.5-turbo model [20] using the same prompt as prior work [14]. We sample 5 different responses at temperature=0.7, to reduce statistical deviations.

Further in Section C.1, we evaluate the same Generative Engine Optimization methods on Perplexity.ai, which is a commercially deployed generative engine, highlighting the generalizability of our proposed Generative Engine Optimization methods.

3.2 Benchmark : GEO-bench

Since there is currently no publicly available dataset containing Generative Engine related queries, we curate **GEO-bench**, a benchmark consisting of 10K queries from multiple sources, repurposed for generative engines, along with synthetically generated queries. The benchmark includes queries from nine different sources, each further categorized based on their target domain, difficulty, query intent, and other dimensions.

Datasets: **1. MS Macro, 2. ORCAS-1, and 3. Natural Questions:** [1, 6, 13] These datasets contain real anonymized user queries from Bing and Google Search Engines. These three collectively represent the common set of datasets that are used in search engine related research. However, Generative Engines will be posed with far more difficult and specific queries with the intent of synthesizing answers from multiple sources instead of searching for them. To this end, we repurpose several other publicly available datasets: **4. AllSouls:** This dataset contains essay questions from "All Souls College, Oxford University." The queries in this dataset require Generative Engines to perform appropriate reasoning to aggregate information from multiple sources. **5. LIMA:** [25] contains challenging questions requiring Generative Engines to not only aggregate information but also perform suitable reasoning to answer the question (e.g., writing a short poem, python code.). **6. Davinci-Debtate** [14] contains debate questions generated for testing Generative Engines. **7. Perplexity.ai Discover**[2]**:** These queries are sourced from Perplexity.ai's Discover section, which is an updated list of trending queries on the platform. **8. ELI-5**[3]**:** This dataset contains questions from the ELI5 subreddit, where users ask complex questions and expect answers in simple, layman's terms. **9. GPT-4 Generated Queries:** To supplement diversity in query distribution, we prompt GPT-4 [21] to generate queries ranging from various domains (e.g., science, history) and based on query intent (e.g., navigational, transactional) and based on difficulty and scope of generated response (e.g., open-ended, fact-based).

. Our benchmark comprises 10K queries divided into 8K, 1K, and 1K for train, validation, and test splits, respectively. We preserve the real-world query distribution, with our benchmark containing 80% informational queries and 10% each for transactional and navigational queries. Each query is augmented with the cleaned text content of the top 5 search results from the Google search engine.

Tags. Optimizing website content often requires targeted changes based on the task's domain. Additionally, a user of Generative Engine Optimization may need to identify an appropriate method for only a subset of queries, considering multiple factors such as domain, user intent, and query nature. To facilitate this, we tag each query with one of seven different categories. For tagging, we employ the GPT-4 model and manually verify high recall and precision on the test split.

Overall, GEO-bench consists of queries from 25 diverse domains such as Arts, Health, and Games; it features a range of query difficulties from simple to multi-faceted; includes 9 different types of queries such as informational and transactional; and encompasses 7 different categorizations. Owing to its specially designed high diversity, the size of the benchmark, and its real-world nature, GEO-bench is a comprehensive benchmark for evaluating Generative Engines and serves as a standard testbed for assessing them for various purposes in this and future works. We provide more details about GEO-bench in Appendix B.2.

3.3 GEO Methods

We evaluate 9 different proposed GEO methods as described in Section 2.2.2. We compare them with a baseline, which measures the impression metric of unmodified website sources. We evaluate methods on the complete GEO-bench test split. Further, to reduce variance in results, we run our experiments on five different random seeds and report the average.

3.4 Evaluation Metrics

We utilize the impression metrics as defined in Section 2.2.1. Specifically, we employ two impression metrics: **1. Position-Adjusted Word Count**, which combines word count and position count. To analyze the effect of individual components, we also report scores on the two sub-metrics separately. **2. Subjective Impression**, which is a subjective metric encompassing seven different aspects: 1) relevance of the cited sentence to the user query, 2) influence of the citation, assessing the extent to which the generated response relies on the citation, 3) uniqueness of the material presented by a citation, 4) subjective position, gauging the prominence of the positioning of source from the user's viewpoint, 5) subjective count, measuring the amount of content presented from the

[2]https://www.perplexity.ai/discover

[3]https://huggingface.co/datasets/eli5_category

需要一些额外的内容。为了分析我们方法的性能提升，对于每个输入用户查询，我们随机选择一个源网站进行优化，并将每个 GEO 方法分别应用于同一源。关于 GEO 方法的更多细节，请参阅附录 B.4。

3 实验装置 3.1 评估生成引擎

根据之前的工作[14]，我们采用两步构建来进行生成引擎设计。第一步是获取相关源用于输入查询，第二步是 LLM 根据获取的源生成响应。与之前的研究类似，我们不使用摘要，而是为每个来源提供完整的回应。由于上下文长度限制和基于变换器模型上下文大小的二次扩展成本，每个查询只从谷歌搜索引擎获取前五个来源。该设置高度模仿了以往工作的工作流程，以及商业通用电气（如 you.com 和 perplexity.ai）采用的总体设计。答案随后由 gpt3.5 涡轮模型[20]生成，使用与之前工作相同的提示[14]。我们在温度=0.7 下抽样 5 个不同的响应，以减少统计偏差。

在 C.1 节，我们评估了相同的生成引擎优化方法，Perplexity.ai 是商业部署的生成引擎，强调了我们提出的生成引擎优化方法的通用性。

3.2 基准测试：GEO-bench

由于目前没有包含生成引擎相关查询的公开数据集，我们策划了 GEO-bench，这是一个基准测试，包含来自多个来源的 1 万条查询，这些查询被重新利用用于生成引擎，同时还包括合成生成的查询。基准测试包含来自九个不同来源的查询，每个来源根据其目标领域、难度、查询意图及其他维度进一步分类。

数据集：1. MS Macro，2. ORCAS-1,3. 自然问题：[1， 6， 13] 这些数据集包含来自 Bing 和 Google 搜索引擎的真实匿名用户查询。这三者共同代表了搜索引擎相关研究中使用的共同数据集集合。然而，生成式引擎将面临更复杂且具体的查询，目的是综合多个来源的答案，而非搜索。为此，我们还重新利用了其他几个公开数据集：4. AllSouls：该数据集包含来自“牛津大学 All Souls 学院”的论文题目。该数据集中的查询需要生成引擎进行适当的推理，以汇总来自多个来源的信息。5. LIMA：[25] 包含挑战性问题，要求生成引擎不仅要汇总信息，还要进行适当的推理来回答问题（例如，写一首短诗，Python 代码）。6. 达芬奇债务 [14] 包含为测试生成引擎而产生的辩论题目。7. Perplexity.ai 发现：这些查询来源于 Perplexity.ai 的发现板块，该部分是平台上最新的热门查询列表。8. ELI-5：本数据集包含来自 ELI5 子版块的问题，用户提出复杂问题，期望用简单、通俗易懂的语言回答。9. GPT-4 生成查询：为补充查询分布的多样性，我们提示 GPT-4[21]生成涵盖多个领域（如科学、历史）的查询，并基于查询意图（如导航型、交易型）以及基于生成响应的难度和范围（如开放式、基于事实型）的查询。

.我们的基准测试包含 1 万个查询，分为 8K、1K 和 1K，分别用于训练、验证和测试。我们保持了现实世界的查询分布，基准测试包含 80%的信息性查询，事务性和导航性各 10%。每个查询都会补充谷歌搜索引擎前五名搜索结果的净化文本内容。

标签。优化网站内容通常需要根据任务的领域进行有针对性的调整。此外，生成引擎优化的用户可能需要为部分查询确定合适的方法，考虑域名、用户意图和查询性质等多重因素。为此，我们会为每个查询标注七个不同的类别之一。在标记方面，我们采用 GPT-4 模型，手动验证测试分段的高召回率和高精度。

总体而言，GEO-bench 包含来自 25 个不同领域的查询，如艺术、健康和游戏;它具有从简单到多方面的查询难度;包含 9 种不同类型的查询，如信息查询和事务查询;并涵盖了 7 种不同的分类。由于其专门设计的高多样性、基准规模大以及其现实世界特性，GEObench 是评估生成引擎的综合基准，并作为评估生成引擎的标准测试平台，用于本项目及未来工作中的各种用途。我们在附录 B.2 中提供了更多关于 GEO-bench 的细节。

3.3 GEO 方法

我们评估了 9 种不同的 GEO 方法，详见第 2.2.2 节。我们将它们与基线进行比较，基线衡量未修改网站来源的曝光量。我们评估了完整的 GEO-bench 测试分段方法。此外，为了减少结果的差异，我们在五种不同的随机种子上进行实验并报告平均值。

3.4 评估指标

我们采用第 2.2.1 节定义的曝光量指标。具体来说，我们采用了两个曝光指标：1. 位置调整字数，结合字数和位置数。为了分析各个组成部分的影响，我们还分别报告了这两个子指标的得分。2. 主观印象，这是一个包含七个不同方面的主观指标：1）引用句子对用户查询的相关性，2）引用的影响，评估生成回答对引用的依赖程度，3）引用所呈现材料的独特性，4）主观定位，评估从用户视角判断来源位置的重要性， 5）主观计数，衡量从中呈现的内容量

2https：//www.perplexity.ai/discover

3https：//huggingface.co/datasets/eli5_category

Method	Position-Adjusted Word Count			Subjective Impression							
	Word	Position	**Overall**	Rel.	Infl.	Unique	Div.	FollowUp	Pos.	Count	**Average**
Performance without Generative Engine Optimization											
No Optimization	19.5	19.3	19.3	19.3	19.3	19.3	19.3	19.3	19.3	19.3	19.3
Non-Performing Generative Engine Optimization methods											
Keyword Stuffing	17.8	17.7	17.7	19.8	19.1	20.5	20.4	20.3	20.5	20.4	20.2
Unique Words	20.7	20.5	20.5	20.5	20.1	19.9	20.4	20.2	20.7	20.2	20.4
High-Performing Generative Engine Optimization methods											
Easy-to-Understand	22.2	22.4	22.0	20.2	21.0	20.0	20.1	20.1	20.9	19.9	20.5
Authoritative	21.8	21.3	21.3	22.3	22.1	22.4	23.1	22.2	23.1	22.7	22.9
Technical Terms	23.1	22.7	22.7	20.9	21.7	20.5	21.2	20.8	21.9	20.8	21.4
Fluency Optimization	25.1	24.6	24.7	21.1	22.9	20.4	21.6	21.0	22.4	21.1	21.9
Cite Sources	24.9	24.5	24.6	21.4	22.5	21.0	21.6	21.2	22.2	20.7	21.9
Quotation Addition	27.8	27.3	**27.2**	23.8	25.4	23.9	24.4	22.9	24.9	23.2	**24.7**
Statistics Addition	25.9	25.4	25.2	22.5	24.5	23.0	23.3	21.6	24.2	23.0	23.7

Table 1: Absolute impression metrics of GEO methods on GEO-bench. Performance Measured on Two metrics and their sub-metrics. Compared to baselines, simple methods like Keyword Stuffing traditionally used in SEO don't perform well. However, our proposed methods such as Statistics Addition and Quotation Addition show strong performance improvements across all metrics. The best methods improve upon baseline by 41% and 28% on Position-Adjusted Word Count and Subjective Impression respectively. For readability, Subjective Impression scores are normalized with respect to Position-Adjusted Word Count resulting in similar baseline scores.

citation as perceived by the user, 6) likelihood of the user clicking the citation, and 7) diversity of the material presented. These sub-metrics assess diverse aspects that content creators can target to improve one or more areas effectively. Each sub-metric is evaluated using GPT-3.5, following a methodology akin to that described in G-Eval [15]. In G-Eval, a form-based evaluation template is provided to the language model, along with a GE generated response with citations. The model outputs a score (computed by sampling multiple times) for each citation. However, since G-Eval scores are poorly calibrated, we normalize them to have the same mean and variance as Position-Adjusted Word Count to enable a fair and meaningful comparison. We provide the exact templates used in Appendix B.3.

Furthermore, all impression metrics are normalized by multiplying them with a constant factor so that the sum of the impressions of all citations in a response equals 1. In our analysis, we compare methods by calculating the relative improvement in impression. For an initial generated response r from sources $S_i \in \{s_1, \ldots, s_m\}$, and a modified response r', the relative improvement in impression for each source s_i is measured as:

$$Improvement_{s_i} = \frac{Imp_{s_i}(r') - Imp_{s_i}(r)}{Imp_{s_i}(r)} \times 100 \tag{4}$$

The modified response r' is produced by applying the GEO method being evaluated to one of the sources s_i. The source s_i selected for optimization is chosen randomly but remains constant for a particular query across all GEO methods.

4 RESULTS

We evaluate various Generative Engine Optimization methods designed to optimize website content for better visibility in Generative Engine responses, compared against a baseline with no optimization. Our evaluation used GEO-bench, a diverse benchmark of user queries from multiple domains and settings. Performance was measured using two metrics: *Position-Adjusted Word Count* and *Subjective Impression*. The former considers word count and citation position in the GE's response, while the latter computes multiple subjective factors, giving an overall impression score.

Table 1 details the absolute impression metrics of different methods on multiple metrics. The results reveal that our GEO methods consistently outperform the baseline across all metrics on GEO-bench. This shows the robustness of these methods to varying queries, yielding significant improvements despite query diversity. Specifically, our top-performing methods, Cite Sources, Quotation Addition, and Statistics Addition, achieved a relative improvement of 30-40% on the *Position-Adjusted Word Count* metric and 15-30% on the *Subjective Impression* metric. These methods, involving adding relevant statistics (Statistics Addition), incorporating credible quotes (Quotation Addition), and including citations from reliable sources (Cite Sources) in the website content, require minimal changes but significantly improve visibility in GE responses, enhancing both the credibility and richness of the content.

Interestingly, stylistic changes such as improving fluency and readability of the source text (Fluency Optimization and Easy-to-Understand) also resulted in a significant visibility boost of 15-30%. This suggests that Generative Engines value not only content but also information presentation.

方法	位置调整字数							主观印象		
	词	位置	整体	关系	独立	唯一	分区	跟进	位置	数量	平均
无生成引擎优化时的性能											
无 优化	19.5	19.3	19.3	19.3	19.3	19.3	19.3	19.3	19.3	19.3	19.3
非性能生成引擎优化方法											
关键词填充	17.8	17.7	17.7	19.8	19.1	20.5	20.4	20.3	20.5	20.4	20.2
唯一词	20.7	20.5	20.5	20.5	20.1	19.9	20.4	20.2	20.7	20.2	20.4
高性能生成引擎优化方法											
易懂	22.2	22.4	22.0	20.2	21.0	20.0	20.1	20.1	20.9	19.9	20.5
权威	21.8	21.3	21.3	22.3	22.1	22.1	22.4	23.1	22.1	2	23.1
22.7	22.9	技术术语	23.1	22.7	22.7	20.9	21.7	20.5	21.2	20.8	21.9
20.8	21.4	流畅度优化	25.1	24.6	24.7	21.1	22.9	20.4	21.6	21.0	22.4
21.1	21.9	引用来源	24.9	24.5	24.6	21.4	22.5	21.0	21.6	21.2	22.2
20.7	21.9	引用附加	27.8	27.3	27.2	23.8	25.4	23.9	24.4	22.9	24.9
23.224.7	统计加法	25.9	25.4	25.2	22.5	24.5	23.0	23.3	21.6	24.2	23.0
23.7											

表 1：GEO 方法在 GEO-bench 上的绝对印象指标。绩效基于两个指标及其子指标衡量。与基线相比，传统上用于 SEO 的简单方法如关键词堆砌效果不佳。然而，我们提出的方法，如统计加法和引语法加法，在所有指标上都显示出显著的性能提升。最佳方法在位置调整字数和主观印象方面分别比基线提升 41%和 28%。为了提高可读性，主观印象分数与位置调整字数进行归一化，得出相似的基线分数。

引用内容被用户感知，6）用户点击引用的可能性，以及 7）所呈现材料的多样性。这些子指标评估内容创作者可以针对的一个或多个方面进行优化，以有效改进这些方面。每个子指标均采用 GPT-3.5 进行评估，方法类似于 G 评估[15]中描述的方法。在 G-Eval 中，向语言模型提供基于表单的评估模板，以及 GE 生成的带引用的回复。模型为每个引用输出一个分数（通过多次抽样计算得出）。然而，由于 G 评估分数校准不佳，我们将其归一化为与位置调整字数相同的平均值和方差，以便实现公平且有意义的比较。我们提供了附录 B.3 中使用的具体模板。

此外，所有曝光度量通过乘以常数因子进行归一化，使得回复中所有引用的印象总和等于 1。在我们的分析中，我们通过计算印象的相对改善来比较方法。对于来自来源 S 的初始生成反应 r∈{s， . . . ， s}和修改后的反应 r，每个来源 sis 的印象相对改善测量为：

$$\text{改进} = \frac{\text{Imp}(r) - \text{Imp}(r)}{\text{Imp}(r)} \times 100 \quad (4)$$

修正响应 r 是通过对其中一个源 s 应用被评估的 GEO 方法得到的。被选中的优化源是随机选择的，但在所有 GEO 方法中对特定查询保持不变。

4 结果

我们评估多种生成式引擎优化方法
旨在优化网站内容，以提升生成引擎响应中的可见性，相较于无优化基线。我们的评估使用了 GEO-bench，这是一个涵盖多个领域和多个领域的用户查询的多样化基准测试。表现通过两个指标进行衡量：位置调整字数和主观印象。前者考虑了 GE 回答中的字数和引用位置，而后者则计算多个主观因素，从而得出整体印象得分。

表 1 详细列出了不同方法在多个指标上的绝对印象指标。结果显示，我们的 GEO 方法在 GEObench 上的所有指标中始终优于基线水平。这显示了这些方法对不同查询的鲁棒性，尽管查询多样性显著提升。具体来说，我们表现最好的方法——引用来源、引用加法和统计加法——在位置调整字数指标上取得了 30-40%的相对提升，在主观印象指标上提升了 15-30%。这些方法包括添加相关统计数据（统计加法）、引用可信引文（引文补充），以及在网站内容中包含可靠来源的引用（引用来源），只需极少的修改，却显著提升了通用教育回复的可见度，提升内容的可信度和丰富性。

有趣的是，诸如提升源文本的流畅性和可读性（流畅度优化和易懂性）等风格变化，也带来了显著提升 15-30%的可见度。这表明生成式引擎不仅重视内容，也重视信息呈现。

Method	Relative Improvement (%) in Visibility				
	Rank-1	Rank-2	Rank-3	Rank-4	Rank-5
Authoritative	-6.0	4.1	-0.6	12.6	6.1
Fluency Opt.	-2.0	5.2	3.6	-4.4	2.2
Cite Sources	-30.3	2.5	20.4	15.5	115.1
Quotation Addition	-22.9	-7.0	3.5	25.1	99.7
Statistics Addition	-20.6	-3.9	8.1	10.0	97.9

Table 2: Visibility changes through GEO methods for sources with different Rankings in Search Engine. GEO is especially helpful for lower ranked websites.

Method	Top Performing Tags		
	Rank-1	Rank-2	Rank-3
Authoritative	Debate	History	Science
Fluency Opt.	Business	Science	Health
Cite Sources	Statement	Facts	Law & Gov.
Quotation Addition	People & Society	Explanation	History
Statistics Addition	Law & Gov.	Debate	Opinion

Table 3: Top Performing categories for each of the GEO methods. Website-owners can choose relevant GEO strategy based on their target domain.

Further, given generative models are often designed to follow instructions, one would expect a more persuasive and authoritative tone in website content to boost visibility. However, we find no significant improvement, demonstrating that Generative Engines are already somewhat robust to such changes. This highlights the need for website owners to focus on improving content presentation and credibility.

Finally, we evaluate keyword stuffing, i.e., adding more relevant keywords to website content. While widely used for Search Engine Optimization, we find such methods offer little to no improvement on generative engine's responses. This underscores the need for website owners to rethink optimization strategies for generative engines, as techniques effective in search engines may not translate to success in this new paradigm.

5 ANALYSIS

5.1 Domain-Specific Generative Engine Optimizations

In Section 4, we presented the improvements achieved by GEO across the entirety of the GEO-bench benchmark. However, in real-world SEO scenarios, domain-specific optimizations are often applied. With this in mind, and considering that we provide categories for every query in GEO-bench, we delve deeper into the performance of various GEO methods across these categories.

Table 3 provides a detailed breakdown of the categories where our GEO methods have proven to be most effective. A careful analysis of these results reveals several intriguing observations. For instance, Authoritative significantly improves performance in debate-style questions and queries related to the "historical" domain. This aligns with our intuition, as a more persuasive form of writing is likely to hold more value in debates.

Similarly, the addition of citations through Cite Sources is particularly beneficial for factual questions, likely because citations provide a source of verification for the facts presented, thereby enhancing the credibility of the response. The effectiveness of different GEO methods varies across domains. For example, as shown in row 5 of Table 3, domains such as 'Law & Government' and question types like 'Opinion' benefit significantly from the addition of relevant statistics in the website content, as implemented by Statistics Addition. This suggests that data-driven evidence can enhance the visibility of a website in particular contexts. The method Quotation Addition is most effective in the 'People & Society,' 'Explanation,' and 'History' domains. This could be because these domains often involve personal narratives or historical events, where direct quotes can add authenticity and depth to the content. Overall, our analysis suggests that website owners should strive towards making domain-specific targeted adjustments to their websites for higher visibility.

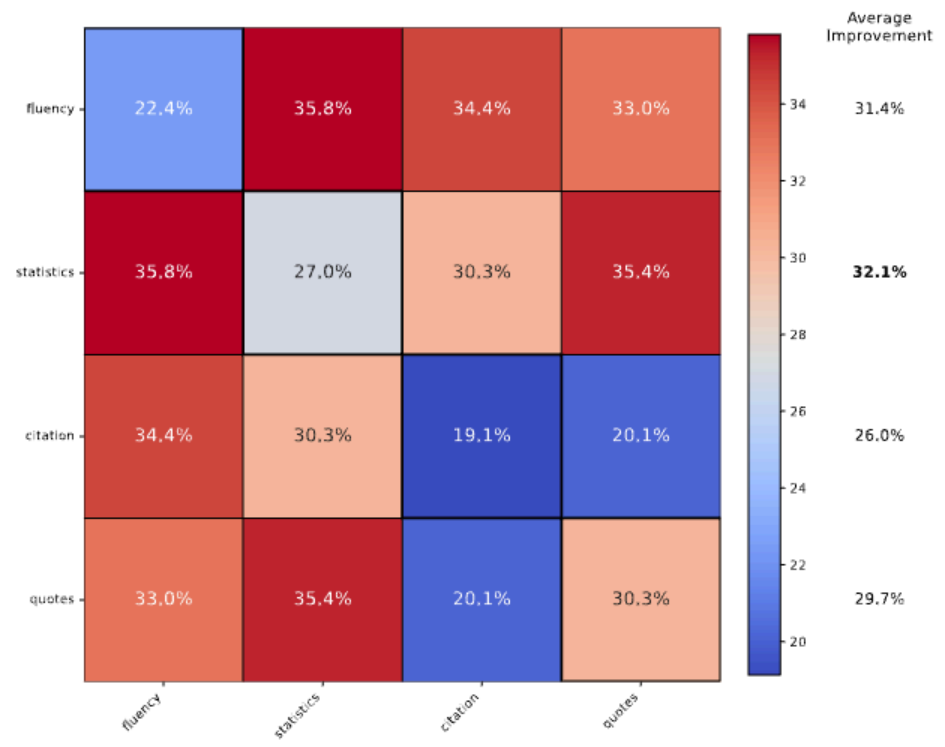

Figure 4: Relative Improvement on using combination of GEO strategies. Using Fluency Optimization and Statistics Addition in conjunction results in maximum performance. The rightmost column shows using Fluency Optimization with other strategies is most beneficial.

5.2 Optimization of Multiple Websites

In the evolving landscape of Generative Engines, GEO methods are expected to become widely adopted, leading to a scenario where all source contents are optimized using GEO. To understand the implications, we conducted an evaluation of GEO methods by optimizing all source contents simultaneously, with results presented in Table 2. A key observation is the differential impact of GEO on websites based on their Search Engine Results Pages (SERP) ranking. Notably, lower-ranked websites, which typically struggle for visibility, benefit significantly more from GEO. This is because traditional search engines rely on multiple factors, such as the number of backlinks and domain presence, which are challenging for small creators to achieve. However, since Generative Engines utilize

方法	可见度的相对改善（百分比）				
	第一名	第二名	第三名	第四名	第五名
权威	−6.0	4.1	−0.6	12.6	6.1
流利度 选择	−2.0	5.2	3.6	−4.4	2.2
引用来源	−30.3	2.5	20.4	15.5	115.1
引用附加	−22.9	−7.0	3.5	25.1	99.7
统计加法	−20.6	−3.9	8.1	10.0	97.9

表 2：通过 GEO 方法对搜索引擎中排名不同来源的可见性变化。GEO 对排名较低的网站尤其有帮助。

方法	表现最佳标签		
	第一名	第二名	第三名
权威辩论	历史	科学	
流利度 选择性	商业	科学	健康
引用 来源	陈述	事实	法律与政府
引用加法	人物与社会	解释	历史
统计 加法	法律与政府辩论	意见	

表 3：各 GEO 方法的表现最佳类别。网站所有者可以根据目标领域选择相关的地理战略。

此外，鉴于生成模型通常设计为遵循指令，人们理应期望网站内容更具说服力和权威性，以提升可见度。然而，我们未发现显著改进，表明生成引擎已经对这些变化有一定的韧性。这凸显了网站所有者需要专注于提升内容呈现和可信度。

最后，我们评估关键词堆砌，即为网站内容添加更多相关关键词。虽然这些方法在搜索引擎优化中被广泛使用，但我们发现对生成引擎的响应几乎没有改善。这凸显了网站所有者重新思考生成式引擎优化策略的必要性，因为搜索引擎中有效的技术在这一新范式中可能无法转化为成功。

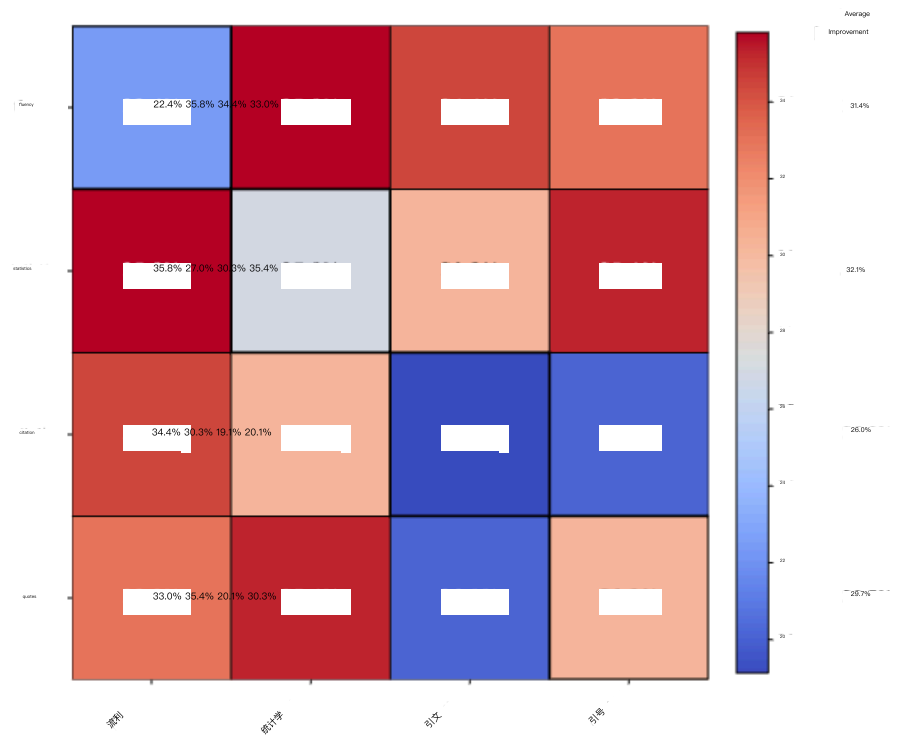

图 4：使用多种 GEO 策略组合的相对改进。结合流畅度优化和统计加法，可以获得最大性能。最右侧的列显示了将流畅度优化与其他策略结合使用最有利的效果。

5 分析 5.1 领域特定生成引擎优化

在第四部分，我们展示了 GEO 在整个 GEO 工作台基准中取得的改进。然而，在现实中的 SEO 场景中，通常会应用针对特定领域的优化。考虑到我们在 GEO-bench 中为每个查询都提供了类别，我们深入探讨了各种 GEO 方法在这些类别中的表现。

表 3 详细列出了我们 GEO 方法被证明最有效的类别。对这些结果的细致分析揭示了几个有趣的观察。例如，权威性显著提升了与“历史”领域相关的辩论式问题和查询的表现。这与我们的直觉相符，因为更具说服力的写作形式在辩论中更有可能具有价值。

同样，通过引用来源添加引用对事实问题尤为有利，可能是因为引用为事实提供了验证来源，从而增强了回答的可信度。不同 GEO 方法的有效性在不同领域存在差异。例如，如表 3 第 5 行所示，“法律与政府”等领域以及“观点”等问题类型，在网站内容中通过统计加法实现了相关统计数据，显著受益。这表明基于数据的证据能够提升网站在特定情境中的可见度。引用添加法在“人物与社会”、“解释”和“历史”领域最为有效。这可能是因为这些领域通常涉及个人叙述或历史事件，直接引用可以为内容增添真实性和深度。总体来看，我们的分析建议网站所有者应努力对网站进行针对特定领域的精准调整，以提高可见度。

5.2 多网站优化

在生成引擎不断发展的领域中，GEO 方法预计将被广泛采用，从而实现所有源内容都通过 GEO 进行优化的场景。为理解其影响，我们通过同时优化所有源内容对 GEO 方法进行了评估，结果见表 2。一个关键观察是 GEO 对网站基于搜索引擎结果页（SERP）排名的影响差异。值得注意的是，排名较低、通常难以获得曝光率的网站，从 GEO 中获益明显更多。这是因为传统搜索引擎依赖于多个因素，如反向链接数量和域名存在度，而这些对小型创作者来说具有挑战性。然而，由于生成引擎利用了

Method	GEO Optimization	Relative Improvement
Cite Sources	**Query:** What is the secret of Swiss chocolate With per capita annual consumption averaging between 11 and 12 kilos, Swiss people rank among the top chocolate lovers in the world (According to a survey conducted by The International Chocolate Consumption Research Group [1])	**132.4%**
Statistics Addition	**Query:** Should robots replace humans in the workforce? **Source:** Not here, and not now — until recently. The big difference is that the robots have come not to destroy our lives, but to disrupt our work, with a staggering 70% increase in robotic involvement in the last decade.	**65.5%**
Authoritative	**Query:** Did the jacksonville jaguars ever make it to the superbowl? **Source:** It is important to note that The Jaguars have never appeared made an appearance in the Super Bowl. However, They have achieved an impressive feat by securing 4 divisional titles to their name. , a testament to their prowess and determination.	**89.1%**

Table 4: Representative examples of GEO methods optimizing source website. Additions are marked in green and Deletions in red. Without adding any substantial new information, GEO methods significantly increase the visibility of the source content.

generative models conditioned on website content, factors such as backlink building should not disadvantage small creators. This is evident from the relative improvements in visibility shown in Table 2. For example, the Cite Sources method led to a substantial 115.1% increase in visibility for websites ranked fifth in SERP, while on average, the visibility of the top-ranked website decreased by 30.3%.

This finding highlights GEO's potential as a tool to democratize the digital space. Many lower-ranked websites are created by small content creators or independent businesses, who traditionally struggle to compete with larger corporations in top search engine results. The advent of Generative Engines might initially seem disadvantageous to these smaller entities. However, the application of GEO methods presents an opportunity for these content creators to significantly improve their visibility in Generative Engine responses. By enhancing their content with GEO, they can reach a wider audience, leveling the playing field and allowing them to compete more effectively with larger corporations.

5.3 Combination of GEO Strategies

While individual GEO strategies show significant improvements across various domains, in practice, website owners are expected to employ multiple strategies in conjunction. To study the performance improvements achieved by combining GEO strategies, we consider all pairs of combinations of the top 4 performing GEO methods, namely Cite Sources, Fluency Optimization, Statistics Addition, and Quotation Addition. Figure 4 displays the heatmap of relative improvement in the Position-Adjusted Word Count visibility metric achieved by combining different GEO strategies. The analysis demonstrates that the combination of Generative Engine Optimization methods can enhance performance, with the best combination (Fluency Optimization and Statistics Addition) outperforming any single GEO strategy by more than 5.5%[4]. Furthermore, Cite Sources significantly boosts performance when used in conjunction with other methods (Average: 31.4%), despite it being relatively less effective when used alone (8% lower than Quotation Addition). The findings underscore the importance of studying GEO methods in combination, as they are likely to be used by content creators in the real world.

[4]Due to cost constraints, the analysis was conducted on a subset of 200 examples from the test split, and therefore the numbers presented here differ from those in Table 1

5.4 Qualitative Analysis

We present a qualitative analysis of GEO methods in Table 4, containing representative examples where GEO methods boost source visibility with minimal changes. Each method optimizes a source through suitable text additions and deletions. In the first example, we see that simply adding the source of a statement can significantly boost visibility in the final answer, requiring minimal effort from the content creator. The second example demonstrates that adding relevant statistics wherever possible ensures increased source visibility in the final Generative Engine response. Finally, the third row suggests that merely emphasizing parts of the text and using a persuasive text style can also lead to improvements in visibility.

6 GEO IN THE WILD : EXPERIMENTS WITH DEPLOYED GENERATIVE ENGINE

Method	Position-Adjusted Word Count	Subjective Impression
No Optimization	24.1	24.7
Keyword Stuffing	21.9	28.1
Quotation Addition	**29.1**	32.1
Statistics Addition	26.2	**33.9**

Table 5: Absolute impression metrics of GEO methods on GEO-bench with Perplexity.ai as GE. While SEO methods such as Keyword Stuffing perform poorly, our proposed GEO methods generalize well to multiple generative engines significanlty improve content visibility.

To reinforce the efficacy of our proposed Generative Engine Optimization methods, we evaluate them on Perplexity.ai, a real deployed Generative Engine with a large user base. Results are

地理环境优化相对改进方法	
引用来源 132.4%	问题：瑞士巧克力的秘密是什么 瑞士人均年均消费量在 11 至 12 公斤之间，是全球最热爱巧克力的群体之一（根据国际巧克力消费研究组[1]的一项调查）。
统计加法 65.5%	问题：机器人应该取代人类进入职场吗? 来源：这里没有，也不是现在——直到最近。最大的不同在于，机器人的出现并非要摧毁我们的生活，而是要干扰我们的工作，过去十年机器人参与率惊人地增长了 70%。
权威 89.1%	问题：杰克逊维尔美洲虎队曾经进入过超级碗吗? 来源：值得注意的是，美洲虎队从未进入过超级碗。然而，他们取得了令人印象深刻的成就，获得了 4 个分区冠军。，这证明了他们的实力与决心。

表 4：GEO 方法优化源网站的代表性示例。新增内容用绿色标注，删除用红色标注。在不增加实质性新信息的情况下，GEO 方法显著提升了源内容的可见度。

基于网站内容的生成模型，如反向链接建设等因素，不应使小型创作者处于劣势。这从表 2 中可见性的相对改善可见一斑。例如，引用来源法使排名 SERP 第五的网站可见度大幅提升了 115.1%，而排名前列的网站平均可见度下降了 30.3%。

这一发现凸显了 GEO 作为民主化数字空间工具的潜力。许多排名较低的网站由小型内容创作者或独立企业创建，这些企业传统上难以在搜索引擎排名中与大型企业竞争。生成引擎的出现最初可能对这些小型实体不利。然而，GEO 方法的应用为这些内容创作者提供了显著提升其在生成引擎响应中可见度的机会。通过通过 GEO 提升内容，他们能够触及更广泛的受众，拉平竞争环境，使他们能够更有效地与大型企业竞争。

5.3 GEO 策略组合

虽然单个 GEO 策略在多个领域都显著提升，但实际上，网站所有者预计会结合多种策略。为了研究结合 GEO 策略所带来的性能提升，我们考虑了前四大表现最优 GEO 方法的所有组合，即引用来源、流畅性优化、统计加法和引用加法。图 4 展示了通过结合不同地理地理策略实现的位置调整字数可见度指标相对提升的热图。分析表明，生成引擎优化方法的结合能够提升性能，最佳组合（流畅度优化和统计加法）比任何单一 GEO 策略高出 5.5%以上。此外，引用来源在使用时显著提升了性能

由于成本限制，分析针对测试拆分中 200 个样本进行，因此这里呈现的数据与表 1 不同

与其他方法结合使用（平均：31.4%），尽管单独使用时效果相对较低（比引用加法低 8%）。研究结果强调了结合研究 GEO 方法的重要性，因为这些方法很可能被现实世界的内容创作者使用。

5.4 定性分析

我们在表 4 中展示了 GEO 方法的定性分析，包含 GEO 方法在最小变化下提升源可见性的代表性示例。每种方法都通过适当的文本添加和删除来优化来源。在第一个例子中，我们看到仅仅添加陈述来源即可显著提升最终答案的可见度，内容创作者只需最小的努力。第二个例子表明，尽可能添加相关统计数据能确保最终生成引擎响应中源的可见性提升。最后，第三行建议仅仅强调文本部分内容并采用说服性文本风格，也能提升可见性。

6 GEO 野外：部署生成引擎的实验

方法	位置调整字数	主观印象
无优化	24.1	24.7
关键词 填充	21.9	28.1
引用加法	29.1	32.1
统计加法	26.2	33.9

表 5：GEO 方法在 GEbench 上的绝对印象指标，Perplexity.ai 为 GE。虽然像关键词堆砌这样的 SEO 方法表现不佳，但我们提出的 GEO 方法能够很好地推广到多个生成引擎，显著提升内容可见度。

为了强化我们提出的生成式引擎优化方法的有效性，我们在 Perplexity.ai 上进行了评估，这是一种拥有大量用户基础的真实部署生成引擎。结果如下

in Table 5. Similar to our generative engine, Quotation Addition performs best in Position-Adjusted Word Count with a 22% improvement over the baseline. Methods that performed well in our generative engine such as Cite Sources, Statistics Addition show improvements of up to 9% and 37% on the two metrics. Our observations, such as the ineffectiveness of traditional SEO methods like Keyword Stuffing, are further highlighted, as it performs 10% worse than the baseline. The results are significant for three reasons: 1) they underscore the importance of developing different Generative Engine Optimization methods to benefit content creators, 2) they highlight the generalizability of our proposed GEO methods on different generative engines, 3) they demonstrate that content creators can use our easy-to-implement proposed GEO methods directly, thus having a high real-world impact. We refer readers to Appendix C.1 for more details.

7 RELATED WORK

Evidence-based Answer Generation: Previous works have used several techniques for answer generation backed by sources. Nakano et al. [19] trained GPT-3 to navigate web environments to generate source-backed answers. Similarly, other methods [17, 23, 24] fetch sources via search engines for answer generation. Our work unifies these approaches and provides a common benchmark for improving these systems in the future. In a recent working draft, Kumar and Lakkaraju [11] showed that strategic text sequences can manipulate LLM recommendations to enhance product visibility in generative engines. While their approach focuses on increasing product visibility through adversarial text, our method introduces non-adversarial strategies to optimize any website content for improved visibility in generative engine search results.

Retrieval-Augmented Language Models: Several recent works have tackled the issues of limited memory of language models by fetching relevant sources from a knowledge base to complete a task [3, 9, 18]. However, Generative Engine needs to generate an answer and provide attributions throughout the answer. Further, Generative Engine is not limited to a single text modality regarding both input and output. Additionally, the framework of Generative Engine is not limited to fetching relevant sources but instead comprises multiple tasks such as query reformulation, source selection, and making decisions on how and when to perform them.

Search Engine Optimization: In nearly the past 25 years, extensive research has optimized web content for search engines [2, 12, 22]. These methods fall into On-Page SEO, improving content and user experience, and Off-Page SEO, boosting website authority through link building. In contrast, GEO deals with a more complex environment involving multi-modality, conversational settings. Since GEO is optimized against a generative model not limited to simple keyword matching, traditional SEO strategies will not apply to Generative Engine settings, highlighting the need for GEO.

8 CONCLUSION

In this work, we formulate search engines augmented with generative models that we dub generative engines. We propose Generative Engine Optimization (GEO) to empower content creators to optimize their content under generative engines. We define impression metrics for generative engines and propose and release GEO-bench: a benchmark encompassing diverse user queries from multiple domains and settings, along with relevant sources needed to answer those queries. We propose several ways to optimize content for generative engines and demonstrate that these methods can boost source visibility by up to 40% in generative engine responses. Among other findings, we show that including citations, quotations from relevant sources, and statistics can significantly boost source visibility. Further, we discover a dependence of GEO methods' effectiveness on the query domain and the potential of combining multiple GEO strategies in conjunction. We show promising results on a commercially deployed generative engine with millions of active users, showcasing the real-world impact of our work. In summary, our work is the first to formalize the important and timely GEO paradigm, releasing algorithms and infrastructure (benchmarks, datasets, and metrics) to facilitate rapid progress in generative engines by the community. This serves as a first step towards understanding the impact of generative engines on the digital space and the role of GEO in this new paradigm of search engines.

9 LIMITATIONS

While we rigorously test our proposed methods on two generative engines, including a publicly available one, methods may need to adapt over time as GEs evolve, mirroring the evolution of SEO. Additionally, despite our efforts to ensure the queries in our GEO-bench closely resemble real-world queries, the nature of queries can change over time, necessitating continuous updates. Further, owing to the black-box nature of search engine algorithms, we didn't evaluate how GEO methods affect search rankings. However, we note that changes made by GEO methods are targeted changes in textual content, bearing some resemblance with SEO methods, while not affecting other metadata such as domain name, backlinks, etc, and thus, they are less likely to affect search engine rankings. Further, as larger context lengths in language models become economical, it is expected that future generative models will be able to ingest more sources, thus reducing the impact of search rankings. Lastly, while every query in our proposed GEO-bench is tagged and manually inspected, there may be discrepancies due to subjective interpretations or errors in labeling.

10 ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 2107048. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] Daria Alexander, Wojciech Kusa, and Arjen P. de Vries. 2022. ORCAS-I: Queries Annotated with Intent using Weak Supervision. *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval* (2022). https://api.semanticscholar.org/CorpusID:248495926

[2] Prashant Ankalkoti. 2017. Survey on Search Engine Optimization Tools & Techniques. *Imperial journal of interdisciplinary research* 3 (2017). https://api.semanticscholar.org/CorpusID:116487363

[3] Akari Asai, Xinyan Velocity Yu, Jungo Kasai, and Hannaneh Hajishirzi. 2021. One Question Answering Model for Many Languages with Cross-lingual

见表 5。与我们的生成引擎类似，Quotation Addition 在位置调整字数方面表现最佳，比基线提升了 22%。在我们的生成引擎中表现良好的方法，如 Cite Sources、Statistics Addition，显示两项指标分别提升了 9%和 37%。我们的观察，比如传统 SEO 方法如关键词堆砌的无效，表现比基线低了 10%。这些结果具有三个重要意义：1）强调开发不同生成引擎优化方法以惠及内容创作者的重要性，2）强调我们提出的 GEO 方法在不同生成引擎上的通用性，3）表明内容创作者可以直接使用我们易于实施的 GEO 方法，从而产生高度的实际影响。详情请参阅附录 C.1。

7 相关工作

基于证据的答案生成：先前的研究使用

有多种答案生成技术，均有资料支持。Nakano 等人[19]训练 GPT-3 在网络环境中导航，生成基于来源的答案。同样，其他方法[17， 23，24]通过搜索引擎获取来源以生成答案。我们的工作统一了这些方法，并为未来改进这些系统提供了共同的基准。在最近的一份工作草案中，Kumar 和 Lakkaraju [11] 展示了战略文本序列可以控大型语言模型推荐，以提升生成引擎中的产品可见性。虽然他们的方法侧重于通过对抗性文本提升产品可见度，而我们的方法引入了非对抗性策略，以优化网站内容以提升生成引擎搜索结果中的可见性。

检索增强语言模型：若干近期工作

通过从知识库获取相关来源来完成任务，解决了语言模型记忆有限的问题[3， 9， 18]。然而，生成引擎需要生成答案，并在答案中提供归因。此外，生成引擎并不局限于输入和输出的单一文本模态。此外，生成引擎的框架不仅限于获取相关源，还包括多项任务，如查询重构、源代码选择，以及决策如何、何时执行这些任务。

搜索引擎优化：在过去近 25 年里，广泛的研究优化了搜索引擎的网页内容[2， 12， 22]。这些方法分别包括页面内 SEO（提升内容和用户体验）和页面外 SEO（通过链接建设提升网站权威）。相比之下，GEO 处理的是更复杂的环境，涉及多模态和对话环境。由于 GEO 针对不限于简单关键词匹配的生成模型进行优化，传统的 SEO 策略不适用于生成引擎设置，凸显了 GEO 的必要性。

8 结论

在本研究中，我们构建了带有生成模型的搜索引擎，称之为生成引擎我们提出生成式引擎优化（GEO），赋能内容创作者在生成式引擎下优化内容。我们定义了生成引擎的曝光指标，并提出并发布了 GEO-bench：一个涵盖多个领域和设置的多样化用户查询的基准测试，以及回答这些问题所需的相关来源。我们提出了多种优化生成引擎内容的方法，并证明这些方法能将源代码在生成引擎响应中的可见度提升高达 40%。我们发现，包含引用、相关来源的引用和统计数据可以显著提升来源的可见度。此外，我们发现 GEO 方法对查询域的有效性以及多种 GEO 策略结合的潜力。我们在商业部署的生成引擎上取得了有前景的成果，拥有数百万活跃用户，展示了我们工作的现实影响。总之，我们的工作是首个正式化重要且及时的 GEO 范式，发布了算法和基础设施（基准测试、数据集和指标），以促进社区在生成引擎领域的快速发展。这也是理解生成式引擎对数字空间影响以及 GEO 在这一新搜索引擎范式中作用的第一步。

9 限制

虽然我们严格测试了两种生成引擎，包括一个公开的生成引擎，但随着通用电气的发展，方法可能需要适应，这与 SEO 的发展相呼应。此外，尽管我们努力确保 GEObench 中的查询与现实世界查询相似，但查询的性质可能随时间变化，需要持续更新。此外，由于搜索引擎算法的黑箱特性，我们没有评估 GEO 方法如何影响搜索排名。然而，我们注意到，GEO 方法所做的更改是针对文本内容的有针对性调整，与 SEO 方法有些相似，同时不影响域名、反向链接等其他元数据，因此对搜索引擎排名的影响较小。此外，随着语言模型中更长上下文长度变得经济实惠，预计未来的生成模型将能够接收更多来源，从而减少搜索排名的影响。最后，虽然我们提议的 GEO 工作台中的每个查询都被标记并手动检查，但由于主观解读或标签错误，可能会存在差异。

10 致谢

本材料基于国家科学基金会 2107048 号资助下支持的工作。本材料中表达的任何观点、发现、结论或建议均为作者个人观点，不一定反映美国国家科学基金会的立场。

引用

[1] 达里亚·亚历山大、沃伊切赫·库萨和阿尔延·P·德弗里斯。2022 年。ORCAS-I：使用弱监督进行意图注释的查询。第 45 届国际会议记录
全国 ACM SIGIR 信息研发大会
检索（2022 年）。https://api.semanticscholar.org/CorpusID：248495926 [2] Prashant Ankalkoti。2017 年。搜索引擎优化工具调查 &
技术。《帝国大学跨学科研究杂志》3 期（2017 年）。https：api.semanticscholar.org/CorpusID：116487363 [3] 浅井明里、余欣妍速度、笠井淳吾和汉娜内·哈吉希尔齐。2021.
跨语言的一种问答模型

Dense Passage Retrieval. In *Neural Information Processing Systems*. https://api.semanticscholar.org/CorpusID:236428949

[4] Sergey Brin and Lawrence Page. 1998. The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Comput. Networks* 30 (1998), 107–117. https://api.semanticscholar.org/CorpusID:7587743

[5] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. In *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (Eds.), Vol. 33. Curran Associates, Inc., 1877–1901. https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

[6] Nick Craswell, Bhaskar Mitra, Emine Yilmaz, Daniel Fernando Campos, and Jimmy J. Lin. 2021. MS MARCO: Benchmarking Ranking Models in the Large-Data Regime. *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval* (2021). https://api.semanticscholar.org/CorpusID:234336491

[7] Brian Dean. 2023. We Analyzed 4 Million Google Search Results. Here's What We Learned About Organic Click Through Rate. https://backlinko.com/google-ctr-stats Accessed: 2024-06-08.

[8] Danny Goodwin. 2011. Top Google Result Gets 36.4% of Clicks [Study]. https://www.searchenginewatch.com/2011/04/21/top-google-result-gets-36-4-of-clicks-study/

[9] Kelvin Guu, Kenton Lee, Zora Tung, Panupong Pasupat, and Ming-Wei Chang. 2020. REALM: Retrieval-Augmented Language Model Pre-Training. *ArXiv* abs/2002.08909 (2020). https://api.semanticscholar.org/CorpusID:211204736

[10] Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. 2023. Survey of hallucination in natural language generation. *Comput. Surveys* 55, 12 (2023), 1–38.

[11] Aounon Kumar and Himabindu Lakkaraju. 2024. Manipulating Large Language Models to Increase Product Visibility. arXiv:2404.07981 [cs.IR]

[12] R.Anil Kumar, Zaiduddin Shaik, and Mohammed Furqan. 2019. A Survey on Search Engine Optimization Techniques. *International Journal of P2P Network Trends and Technology* (2019). https://doi.org/10.14445/22492615/IJPTT-V9I1P402

[13] Tom Kwiatkowski, Jennimaria Palomaki, Olivia Redfield, Michael Collins, Ankur P. Parikh, Chris Alberti, Danielle Epstein, Illia Polosukhin, Jacob Devlin, Kenton Lee, Kristina Toutanova, Llion Jones, Matthew Kelcey, Ming-Wei Chang, Andrew M. Dai, Jakob Uszkoreit, Quoc V. Le, and Slav Petrov. 2019. Natural Questions: A Benchmark for Question Answering Research. *Transactions of the Association for Computational Linguistics* 7 (2019), 453–466. https://api.semanticscholar.org/CorpusID:86611921

[14] Nelson F. Liu, Tianyi Zhang, and Percy Liang. 2023. Evaluating Verifiability in Generative Search Engines. *ArXiv* abs/2304.09848 (2023). https://api.semanticscholar.org/CorpusID:258212854

[15] Yang Liu, Dan Iter, Yichong Xu, Shuo Wang, Ruochen Xu, and Chenguang Zhu. 2023. G-Eval: NLG Evaluation using GPT-4 with Better Human Alignment. *ArXiv* abs/2303.16634 (2023). https://api.semanticscholar.org/CorpusID:257804696

[16] G. D. Maayan. 2023. How Google SGE will impact your traffic – and 3 SGE recovery case studies. *Search Engine Land* (5 Sep 2023). https://searchengineland.com/how-google-sge-will-impact-your-traffic-and-3-sge-recovery-case-studies-431430

[17] Jacob Menick, Maja Trebacz, Vladimir Mikulik, John Aslanides, Francis Song, Martin Chadwick, Mia Glaese, Susannah Young, Lucy Campbell-Gillingham, Geoffrey Irving, and Nathan McAleese. 2022. Teaching language models to support answers with verified quotes. *ArXiv* abs/2203.11147 (2022). https://api.semanticscholar.org/CorpusID:247594830

[18] Grégoire Mialon, Roberto Dessì, Maria Lomeli, Christoforos Nalmpantis, Ramakanth Pasunuru, Roberta Raileanu, Baptiste Rozière, Timo Schick, Jane Dwivedi-Yu, Asli Celikyilmaz, Edouard Grave, Yann LeCun, and Thomas Scialom. 2023. Augmented Language Models: a Survey. *ArXiv* abs/2302.07842 (2023). https://api.semanticscholar.org/CorpusID:256868474

[19] Reiichiro Nakano, Jacob Hilton, S. Arun Balaji, Jeff Wu, Ouyang Long, Christina Kim, Christopher Hesse, Shantanu Jain, Vineet Kosaraju, William Saunders, Xu Jiang, Karl Cobbe, Tyna Eloundou, Gretchen Krueger, Kevin Button, Matthew Knight, Benjamin Chess, and John Schulman. 2021. WebGPT: Browser-assisted question-answering with human feedback. *ArXiv* abs/2112.09332 (2021). https://api.semanticscholar.org/CorpusID:245329531

[20] OpenAI. 2022. Introducing ChatGPT. https://openai.com/index/chatgpt/

[21] OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, Red Avila, Igor Babuschkin, Suchir Balaji, Valerie Balcom, Paul Baltescu, Haiming Bao, Mohammad Bavarian, Jeff Belgum, Irwan Bello, Jake Berdine, Gabriel Bernadett-Shapiro, Christopher Berner, Lenny Bogdonoff, Oleg Boiko, Madelaine Boyd, Anna-Luisa Brakman, Greg Brockman, Tim Brooks, Miles Brundage, Kevin Button, Trevor Cai, Rosie Campbell, Andrew Cann, Brittany Carey, Chelsea Carlson, Rory Carmichael, Brooke Chan, Che Chang, Fotis Chantzis, Derek Chen, Sully Chen, Ruby Chen, Jason Chen, Mark Chen, Ben Chess, Chester Cho, Casey Chu, Hyung Won Chung, Dave Cummings, Jeremiah Currier, Yunxing Dai, Cory Decareaux, Thomas Degry, Noah Deutsch, Damien Deville, Arka Dhar, David Dohan, Steve Dowling, Sheila Dunning, Adrien Ecoffet, Atty Eleti, Tyna Eloundou, David Farhi, Liam Fedus, Niko Felix, Simón Posada Fishman, Juston Forte, Isabella Fulford, Leo Gao, Elie Georges, Christian Gibson, Vik Goel, Tarun Gogineni, Gabriel Goh, Rapha Gontijo-Lopes, Jonathan Gordon, Morgan Grafstein, Scott Gray, Ryan Greene, Joshua Gross, Shixiang Shane Gu, Yufei Guo, Chris Hallacy, Jesse Han, Jeff Harris, Yuchen He, Mike Heaton, Johannes Heidecke, Chris Hesse, Alan Hickey, Wade Hickey, Peter Hoeschele, Brandon Houghton, Kenny Hsu, Shengli Hu, Xin Hu, Joost Huizinga, Shantanu Jain, Shawn Jain, Joanne Jang, Angela Jiang, Roger Jiang, Haozhun Jin, Denny Jin, Shino Jomoto, Billie Jonn, Heewoo Jun, Tomer Kaftan, Łukasz Kaiser, Ali Kamali, Ingmar Kanitscheider, Nitish Shirish Keskar, Tabarak Khan, Logan Kilpatrick, Jong Wook Kim, Christina Kim, Yongjik Kim, Jan Hendrik Kirchner, Jamie Kiros, Matt Knight, Daniel Kokotajlo, Łukasz Kondraciuk, Andrew Kondrich, Aris Konstantinidis, Kyle Kosic, Gretchen Krueger, Vishal Kuo, Michael Lampe, Ikai Lan, Teddy Lee, Jan Leike, Jade Leung, Daniel Levy, Chak Ming Li, Rachel Lim, Molly Lin, Stephanie Lin, Mateusz Litwin, Theresa Lopez, Ryan Lowe, Patricia Lue, Anna Makanju, Kim Malfacini, Sam Manning, Todor Markov, Yaniv Markovski, Bianca Martin, Katie Mayer, Andrew Mayne, Bob McGrew, Scott Mayer McKinney, Christine McLeavey, Paul McMillan, Jake McNeil, David Medina, Aalok Mehta, Jacob Menick, Luke Metz, Andrey Mishchenko, Pamela Mishkin, Vinnie Monaco, Evan Morikawa, Daniel Mossing, Tong Mu, Mira Murati, Oleg Murk, David Mély, Ashvin Nair, Reiichiro Nakano, Rajeev Nayak, Arvind Neelakantan, Richard Ngo, Hyeonwoo Noh, Long Ouyang, Cullen O'Keefe, Jakub Pachocki, Alex Paino, Joe Palermo, Ashley Pantuliano, Giambattista Parascandolo, Joel Parish, Emy Parparita, Alex Passos, Mikhail Pavlov, Andrew Peng, Adam Perelman, Filipe de Avila Belbute Peres, Michael Petrov, Henrique Ponde de Oliveira Pinto, Michael, Pokorny, Michelle Pokrass, Vitchyr H. Pong, Tolly Powell, Alethea Power, Boris Power, Elizabeth Proehl, Raul Puri, Alec Radford, Jack Rae, Aditya Ramesh, Cameron Raymond, Francis Real, Kendra Rimbach, Carl Ross, Bob Rotsted, Henri Roussez, Nick Ryder, Mario Saltarelli, Ted Sanders, Shibani Santurkar, Girish Sastry, Heather Schmidt, David Schnurr, John Schulman, Daniel Selsam, Kyla Sheppard, Toki Sherbakov, Jessica Shieh, Sarah Shoker, Pranav Shyam, Szymon Sidor, Eric Sigler, Maddie Simens, Jordan Sitkin, Katarina Slama, Ian Sohl, Benjamin Sokolowsky, Yang Song, Natalie Staudacher, Felipe Petroski Such, Natalie Summers, Ilya Sutskever, Jie Tang, Nikolas Tezak, Madeleine B. Thompson, Phil Tillet, Amin Tootoonchian, Elizabeth Tseng, Preston Tuggle, Nick Turley, Jerry Tworek, Juan Felipe Cerón Uribe, Andrea Vallone, Arun Vijayvergiya, Chelsea Voss, Carroll Wainwright, Justin Jay Wang, Alvin Wang, Ben Wang, Jonathan Ward, Jason Wei, CJ Weinmann, Akila Welihinda, Peter Welinder, Jiayi Weng, Lilian Weng, Matt Wiethoff, Dave Willner, Clemens Winter, Samuel Wolrich, Hannah Wong, Lauren Workman, Sherwin Wu, Jeff Wu, Michael Wu, Kai Xiao, Tao Xu, Sarah Yoo, Kevin Yu, Qiming Yuan, Wojciech Zaremba, Rowan Zellers, Chong Zhang, Marvin Zhang, Shengjia Zhao, Tianhao Zheng, Juntang Zhuang, William Zhuk, and Barret Zoph. 2024. GPT-4 Technical Report. arXiv:2303.08774 [cs.CL]

[22] A. Shahzad, Deden Witarsyah Jacob, Nazri M. Nawi, Hairulnizam Bin Mahdin, and Marheni Eka Saputri. 2020. The new trend for search engine optimization, tools and techniques. *Indonesian Journal of Electrical Engineering and Computer Science* 18 (2020), 1568. https://api.semanticscholar.org/CorpusID:213123106

[23] Kurt Shuster, Jing Xu, Mojtaba Komeili, Da Ju, Eric Michael Smith, Stephen Roller, Megan Ung, Moya Chen, Kushal Arora, Joshua Lane, Morteza Behrooz, W.K.F. Ngan, Spencer Poff, Naman Goyal, Arthur Szlam, Y-Lan Boureau, Melanie Kambadur, and Jason Weston. 2022. BlenderBot 3: a deployed conversational agent that continually learns to responsibly engage. *ArXiv* abs/2208.03188 (2022). https://api.semanticscholar.org/CorpusID:251371589

[24] Romal Thoppilan, Daniel De Freitas, Jamie Hall, Noam Shazeer, Apoorv Kulshreshtha, Heng-Tze Cheng, Alicia Jin, Taylor Bos, Leslie Baker, Yu Du, YaGuang Li, Hongrae Lee, Huaixiu Steven Zheng, Amin Ghafouri, Marcelo Menegali, Yanping Huang, Maxim Krikun, Dmitry Lepikhin, James Qin, Dehao Chen, Yuanzhong Xu, Zhifeng Chen, Adam Roberts, Maarten Bosma, Vincent Zhao, Yanqi Zhou, Chung-Ching Chang, Igor Krivokon, Will Rusch, Marc Pickett, Pranesh Srinivasan, Laichee Man, Kathleen Meier-Hellstern, Meredith Ringel Morris, Tulsee Doshi, Renelito Delos Santos, Toju Duke, Johnny Soraker, Ben Zevenbergen, Vinodkumar Prabhakaran, Mark Diaz, Ben Hutchinson, Kristen Olson, Alejandra Molina, Erin Hoffman-John, Josh Lee, Lora Aroyo, Ravi Rajakumar, Alena Butryna, Matthew Lamm, Viktoriya Kuzmina, Joe Fenton, Aaron Cohen, Rachel Bernstein, Ray Kurzweil, Blaise Aguera-Arcas, Claire Cui, Marian Croak, Ed Chi, and Quoc Le. 2022. LaMDA: Language Models for Dialog Applications. arXiv:2201.08239 [cs.CL]

[25] Chunting Zhou, Pengfei Liu, Puxin Xu, Srini Iyer, Jiao Sun, Yuning Mao, Xuezhe Ma, Avia Efrat, Ping Yu, L. Yu, Susan Zhang, Gargi Ghosh, Mike Lewis, Luke Zettlemoyer, and Omer Levy. 2023. LIMA: Less Is More for Alignment. *ArXiv* abs/2305.11206 (2023). https://api.semanticscholar.org/CorpusID:258822910

密集通道检索。发表于神经信息处理系统。https：api.semanticscholar.org/CorpusID：236428949 [4] 谢尔盖·布林和劳伦斯·佩奇。1998 年。大型水体的解剖学——
文本性网络搜索引擎。计算。《网络》30 期（1998 年），107–117 页。https：//api.semanticscholar.org/CorpusID：7587743
[5] 汤姆·布朗、本杰明·曼、尼克·赖德、梅拉妮·苏比亚、贾里德·D·卡普兰、普拉富拉·达里瓦尔、阿尔文德·尼拉坎坦、普拉纳夫·夏姆、吉里什·萨斯特里、阿曼达·阿斯凯尔、桑迪尼·阿加瓦尔、阿里尔·赫伯特-沃斯、格雷琴·克鲁格、汤姆·亨尼汉、瑞温·柴尔德、阿迪提亚·拉梅什、丹尼尔·齐格勒、杰弗里·吴、克莱门斯·温特、克里斯·赫塞、马克·陈、埃里克·西格勒、马特乌什·利特温、斯科特·格雷、本杰明·切斯、杰克·克拉克、克里斯托弗·伯纳， 萨姆·麦坎迪利什、亚历克·拉德福德、伊利亚·萨茨凯弗和达里奥·阿莫代。2020 年。语言模型是少数样本学习者。
载于《神经信息处理系统进展》，H. Larochelle、M. Ranzato、R. Hadsell、M.F. Balcan 和 H. Lin（编），第 33 卷。柯兰联合公司，1877–1901 年。https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf
[6] 尼克·克拉斯韦尔、巴斯卡尔·米特拉、埃米内·伊尔马兹、丹尼尔·费尔南多·坎波斯和吉米·J·林。2021 年。马尔科女士：大数据中的排名模型基准测试
政权。第 44 届国际 ACM SIGIR 信息检索研究与开发会议论文集（2021 年）。https://api.semanticscholar.org/ CorpusID：234336491

[7] 布莱恩·迪恩。2023 年。我们分析了 400 万个谷歌搜索结果。以下是我们对自然点击率的了解。https://backlinko.com/googlectr-stats 访问时间：2024-06-08。[8] 丹尼·古德温。2011 年。谷歌顶级结果获得 36.4%的点击[研究]。https://www.searchenginewatch.com/2011/04/21/top-google-resultgets-36-4-of-clicks-study/ [9] 郭凯文、李健顿、董卓拉、帕努蓬·帕苏帕特和张明伟。

2020 年。REALM：检索增强语言模型预训练。ArXiv abs/2002.08909（2020 年）。https://api.semanticscholar.org/CorpusID：211204736
[10] 季紫薇、李娜妍、丽塔·弗里斯克、余铁正、丹苏、严旭、石井悦子，
方叶珍、安德烈亚·马多托和帕斯卡尔·冯。2023 年。自然语言生成中的幻觉调查。计算。《调查》55 卷，12 期（2023 年），1–38 页。
[11] 奥农·库马尔和希马宾杜·拉卡拉朱。2024 年。《控大型语言》
提升产品可见度的模型。arXiv：2404.07981 [c.IR] [12] R.阿尼尔·库马尔、扎伊杜丁·沙伊克和穆罕默德·弗尔坎。2019 年。关于
搜索引擎优化技巧。《国际点对点网络趋势与技术杂志》（2019 年）。https://doi.org/10.14445/22492615/IJPTT-V9I1P402 [13] 汤姆·克维亚特科夫斯基，詹妮玛丽亚·帕洛马基，奥利维亚·雷德菲尔德，迈克尔·柯林斯，
安库尔·P·帕里克、克里斯·阿尔贝蒂、丹妮尔·爱泼斯坦、伊利亚·波洛苏欣、雅各布·德夫林、肯顿·李、克里斯蒂娜·图塔诺娃、利昂·琼斯、马修·凯尔西、张明伟、安德鲁·M·戴、雅各布·乌斯科雷特、郭 V·乐和斯拉夫·彼得罗夫。2019 年。自然问题：问答研究的基准。计算语言学协会会刊 7 （2019），453–466。https：//api.semanticscholar.org/CorpusID：86611921

[14] 刘尼尔森、张天一、梁珀西。2023 年。评估生成式搜索引擎的可验证性。ArXiv abs/2304.09848（2023 年）。https://api。semanticscholar.org/CorpusID：258212854 [15] 杨柳、丹伊特、徐奕冲、王硕、徐若辰和朱成光。
2023 年。G 评估：利用 GPT-4 进行 NLG 评估，结合更好的人类对齐。ArXiv abs/2303.16634（2023 年）。https://api.semanticscholar.org/CorpusID：257804696
[16] 丽莎·马西。2023 年。Google SGE 将如何影响你的流量
——以及 3 个 SGE 恢复案例研究。搜索引擎之地（2023 年 9 月 5 日）。https://searchengineland.com/how-google-sge-will-impact-your-trafficand-3-sge-recovery-case-studies-431430
[17] 雅各布·梅尼克，玛雅·特雷巴奇，弗拉基米尔·米库利克，约翰·阿斯拉尼德斯，弗朗西斯·宋，马丁·查德威克、米娅·格莱斯、苏珊娜·杨、露西·坎贝尔-吉林汉姆、杰弗里·欧文和内森·麦卡利斯。2022 年。教学语言模型以支持带有验证引用的答案。ArXiv abs/2203.11147（2022 年）。https： //api.semanticscholar.org/CorpusID：247594830
[18] 格雷瓜尔·米亚隆、罗伯托·德西、玛丽亚·洛梅利、克里斯托福罗斯·纳尔姆潘蒂斯、拉马坎特·帕苏努鲁、罗伯塔·拉莱亚努、巴蒂斯特·罗齐埃、蒂莫·希克、简·德维维迪-尤、阿斯利·切利基尔马兹、爱德华·格雷夫、扬·勒昆和托马斯·夏洛姆。2023 年。增强语言模型综述。ArXiv abs/2302.07842（2023 年）。https://api.semanticscholar.org/CorpusID：256868474
[19] 中野玲一郎、雅各布·希尔顿、S·阿伦·巴拉吉、吴杰夫、欧阳龙、克里斯蒂娜金、克里斯托弗·赫塞、尚塔努·贾因、维尼特·科萨拉朱、威廉·桑德斯、徐江、卡尔·科布、泰娜·埃伦杜、格雷琴·克鲁格、凯文·巴顿、马修·奈特、本杰明·切斯和约翰·舒尔曼。2021 年。WebGPT：基于人类反馈的浏览器辅助问答。ArXiv abs/2112.09332（2021 年）。https： //api.semanticscholar.org/CorpusID：245329531

[20] OpenAI。2022 年。介绍 ChatGPT。https://openai.com/index/chatgpt/ [21] OpenAI、Josh Achiam、Steven Adler、Sandhini Agarwal、Lama Ahmad、Ilge
阿卡娅、弗洛伦西亚·莱奥尼·阿莱曼、迪奥戈·阿尔梅达、扬科·阿尔滕施密特、萨姆·奥特曼、沙马尔·阿纳德卡特、雷德·阿维拉、伊戈尔·巴布什金、苏奇尔·巴拉吉、瓦莱丽·巴尔科姆、保罗·巴尔特斯库、海明·鲍、穆罕默德·巴伐利亚、杰夫·贝尔古姆、伊尔万·贝洛、杰克·伯丁、加布里埃尔·伯纳代特-沙皮罗、克里斯托弗·伯纳、伦尼·博格多诺夫、奥列格·博伊科、玛德莱恩·博伊德、安娜-路易莎·布拉克曼、格雷格·布罗克曼、蒂姆·布鲁克斯，
迈尔斯·布伦戴奇、凯文·巴顿、崔佛·蔡、罗茜·坎贝尔、安德鲁·坎恩、布列塔尼·凯里、切尔西·卡尔森、罗里·卡迈克尔、布鲁克·陈、切·昌、福蒂斯·尚齐斯、德里克·陈、萨利·陈、陈鲁比、杰森·陈、马克·陈、本·切斯、切斯特·赵、凯西·朱、钟亨、戴夫·卡明斯、杰里迈亚·柯里尔、戴云兴、科里·德卡罗、托马斯·德格里、诺亚·多伊奇、达米恩·德维尔、阿尔卡·达尔， 大卫·多汉、史蒂夫·道林、希拉·邓宁、阿德里安·埃科菲特、律师埃莱蒂、泰娜·埃伦杜、大卫·法希、利亚姆·费杜斯、尼科·费利克斯、西蒙·波萨达·费什曼、贾斯顿·福尔特、伊莎贝拉·富尔福德、高利奥、埃利·乔治斯、克里斯蒂安·吉布森、维克·戈尔、塔伦·戈吉内尼、加布里埃尔·戈、拉法·贡蒂霍-洛佩斯、乔纳森·戈登、摩根·格拉夫斯坦、斯科特·格雷、瑞安·格林、约书亚·格罗斯、顾世翔、谢恩、郭宇飞、克里斯·哈拉西， 韩杰西、杰夫·哈里斯、何宇晨、迈克·希顿、约翰内斯·海德克、克里斯·赫塞、艾伦·希基、韦德·希基、彼得·霍谢勒、布兰登·霍顿、许健、胡生利、胡欣、惠真加、山塔努·贾因、肖恩·贾因、张乔安、江安、江之志、金浩俊、金丹尼、城本志乃、比莉·琼恩、俊熙宇、托默·卡夫地、卢卡什·凯撒、阿里·卡马利、英格玛·卡尼奇德， 尼蒂什·希里什·凯斯卡尔、塔巴拉克·汗、洛根·基尔帕特里克、金钟旭、金克里斯蒂娜、金永志、简·亨德里克·基什纳、杰米·基罗斯、马特·奈特、丹尼尔·科科塔伊洛、卢卡什·康德拉丘克、安德鲁·康德里奇、阿里斯·康斯坦丁尼迪斯、凯尔·科西奇、格雷琴·克鲁格、郭维沙尔、迈克尔·兰佩、伊凯·兰、李泰迪、梁玉、丹尼尔·莱维、李卓明、林瑞秋、林茉莉、林斯蒂芬妮， 马特乌什·利特温、特蕾莎·洛佩兹、瑞安·洛、帕特里夏·卢、安娜·马坎朱、金·马尔法奇尼、萨姆·曼宁、托多尔·马尔科夫、雅尼夫·马尔科夫斯基、比安卡·马丁、凯蒂·梅耶、安德鲁·梅恩、鲍勃·麦格鲁、斯科特·梅耶·麦金尼、克里斯汀·麦克利维、保罗·麦克米兰、杰克·麦克尼尔， 大卫·梅迪纳、阿洛克·梅塔、雅各布·梅尼克、卢克·梅茨、安德烈·米申科、帕梅拉·米什金、文尼·莫纳科、埃文·森川、丹尼尔·莫辛、童穆、米拉·穆拉蒂、奥列格·穆尔克、大卫·梅利、阿什文·奈尔、中野礼一郎、拉吉夫·纳亚克、阿尔文德·尼拉坎坦、理查德·吴、卢贤宇、龙欧阳、卡伦·奥基夫、雅库布·帕乔茨基、亚历克斯·派诺、乔·帕勒莫、阿什利·潘图利亚诺、贾姆巴蒂斯塔·帕拉斯坎多洛、乔尔·帕里什、埃米·帕帕里塔， 亚历克斯·帕索斯、米哈伊尔·帕夫洛夫、安德鲁·彭、亚当·佩雷尔曼、菲利普·德·阿维拉、贝尔布特·佩雷斯、迈克尔·彼得罗夫、恩里克·庞德·德·奥利维拉·平托、迈克尔、波科尔尼、米歇尔·波克拉斯、维奇尔·H·庞、托利·鲍威尔、阿莱西娅·鲍尔、鲍里斯·鲍尔、伊丽莎白·普罗尔、劳尔·普里、亚历克·拉德福德、杰克·雷、阿迪亚·拉梅什、卡梅隆·雷蒙德、弗朗西斯·雷尔、肯德拉·林巴赫、卡尔·罗斯、鲍勃·罗斯特德、亨利·鲁塞兹、尼克·赖德， 马里奥·萨尔塔雷利、特德·桑德斯、希巴尼·桑图尔卡尔、吉里什·萨斯特里、希瑟·施密特、大卫·施努尔、约翰·舒尔曼、丹尼尔·塞尔萨姆、凯拉·谢泼德、托基·谢尔巴科夫、杰西卡·谢尔、莎拉·肖克、普拉纳夫·夏姆、希蒙·西多尔、埃里克·西格勒、玛迪·西门斯、乔丹·西特金、卡塔琳娜·斯拉马、伊恩·索尔、本杰明·索科洛夫斯基、杨·宋、娜塔莉·斯陶达彻、费利佩·彼得罗斯基·苏奇、娜塔莉·萨默斯、伊利亚·苏茨凯弗、杰·唐、尼古拉斯·特扎克、玛德琳·B。 汤普森、菲尔·蒂莱特、阿明·图图恩奇安、曾伊丽莎白、普雷斯顿·塔格尔、尼克·特利、杰瑞·特沃雷克、胡安·费利佩·塞隆·乌里韦、安德里亚·瓦隆、阿伦·维贾伊维吉亚、切尔西·沃斯、卡罗尔·韦恩赖特、王贾斯汀·杰伊、王艾尔文、王本、乔纳森·沃德、魏杰森、CJ·温曼、阿基拉·韦利欣达、彼得·韦林德、王佳怡、王莉莲、马特·维托夫、戴夫·威尔纳、克莱门斯·温特、塞缪尔·沃尔里奇、黄汉娜、 劳伦·沃克曼、吴舍温、吴杰夫、吴迈克尔、晓凯、徐陶、余莎拉、余凯文、袁启明、沃伊切赫·扎伦巴、罗文·泽勒斯、张钟、张马文、赵胜佳、郑天浩、庄俊唐、朱威廉和巴雷特·佐夫。2024 年。GPT-4 技术报告。arXiv：2303.08774 [cs.CL]

[22] A. 沙赫扎德，德登·维塔尔西亚·雅各布，纳兹里·M·纳维，海鲁尼扎姆·本·马赫丁，以及马尔赫尼·埃卡·萨普特里。2020 年。搜索引擎优化的新趋势，
工具和技术。《印度尼西亚电气工程与计算机杂志》
《科学》18 卷（2020 年），1568 页。https://api.semanticscholar.org/CorpusID：213123106 [23] 库尔特·舒斯特、许静、莫吉塔巴·科梅利、大菊、埃里克·迈克尔·史密斯、斯蒂芬·翁梅根、陈茂雅、库沙尔·阿罗拉、约书亚·莱恩、莫尔特扎·贝鲁兹、W.K.F. 恩、斯宾塞·波夫、纳曼·戈亚尔、亚瑟·斯拉姆、Y-Lan Boureau、梅拉妮·坎巴杜尔和杰森·韦斯顿。2022 年。BlenderBot 3：一款部署式对话代理，持续学习负责任地参与。ArXiv abs/2208.03188（2022 年）。https://api.semanticscholar.org/CorpusID：251371589
[24] 罗马尔·托皮兰，丹尼尔·德弗雷塔斯，杰米·霍尔，诺姆·沙齐尔，阿普尔夫·库尔-施雷什塔、郑恒子、金爱丽西亚、泰勒·博斯、莱斯利·贝克、余杜、李雅光、李洪来、郑怀秀、阿明·加富里、马塞洛·梅内加利、黄延平、马克西姆·克里昆、德米特里·勒皮欣、詹姆斯·秦、陈德浩、徐元中、陈志峰、亚当·罗伯茨、马尔滕·博斯马、赵文森特、周延奇、张忠清、伊戈尔·克里沃孔、威尔·鲁施、马克·皮克特、普拉内什·斯里尼瓦桑、曼来奇， 凯瑟琳·迈尔-赫尔斯特恩、梅雷迪思·林格尔·莫里斯、图尔西·多希、雷内利托·德洛斯·桑托斯、托朱·杜克、约翰尼·索拉克、本·泽文伯根、维诺德库马尔·普拉巴卡兰、马克·迪亚兹、本·哈钦森、克里斯汀·奥尔森、亚历杭德拉·莫利纳、艾琳·霍夫曼-约翰、乔什·李、洛拉·阿罗约、拉维·拉贾库马尔、阿莱娜·布特里纳、马修·拉姆、维多利亚·库兹米纳、乔·芬顿、亚伦·科恩、瑞秋·伯恩斯坦、雷·库兹韦尔、布莱斯·阿格拉-阿尔卡斯、克莱尔·奎、玛丽安·克罗克、埃德·奇， 以及国乐。2022 年。LaMDA：对话应用的语言模型。arXiv：2201.08239 [cs.CL]

[25] 春廷周、刘鹏飞、徐普欣、斯里尼·艾尔、焦孙、毛宇宁、雪哲马、阿维娅·埃夫拉特、余平、余一、张苏珊、加尔吉·高什、迈克·刘易斯、卢克·泽特尔莫耶和奥默·莱维。2023 年。利马：对阵营来说，少即是多。ArXiv abs/2305.11206（2023 年）。https://api.semanticscholar.org/CorpusID：258822910

Listing 1: Prompt used for Generative Engine. The GE takes the query and 5 sources as input and outputs the response to query with response grounded in the sources.

```
Write an accurate and concise answer for the given user question,
      using _only_ the provided summarized web search results.
      The answer should be correct, high-quality, and written by
      an expert using an unbiased and journalistic tone. The user
      's language of choice such as English, Francais, Espamol,
      Deutsch, or should be used. The answer should be
      informative, interesting, and engaging. The answer's logic
      and reasoning should be rigorous and defensible. Every
      sentence in the answer should be _immediately followed_ by
      an in-line citation to the search result(s). The cited
      search result(s) should fully support _all_ the information
      in the sentence. Search results need to be cited using [
      index]. When citing several search results, use [1][2][3]
      format rather than [1, 2, 3]. You can use multiple search
      results to respond comprehensively while avoiding
      irrelevant search results.

Question: {query}

Search Results:
{source_text}
```

Listing 2: Representative Queries from each of the 9 datasets in GEO-BENCH

```
### ORCAS
- what does globalization mean
- wine pairing list

### AllSouls
- Are open-access journals the future of academic publishing?
- Should the study of non-Western philosophy be a requirement
      for a philosophy degree in the UK?

### Davinci-Debate
- Should all citizens receive a basic income?
- Should governments promote atheism?

### ELI5
- Why does my cat kick its toys when playing with them?
- what does caffeine actually do your muscles, especially
      regarding exercising?

### GPT-4
- What are the benefits of a keto diet?
- What are the most profound impacts of the Renaissance period
      on modern society?

### LIMA
- What are the primary factors that influence consumer behavior?
- What would be a great twist for a murder mystery? I'm looking
      for something creative, not to rehash old tropes.

### MS-Macro
- what does monogamous
- what is the normal fbs range for children

### Natural Questions
- where does the phrase bee line come from
- what is the prince of persia in the bible

### Perplexity.ai
- how to gain more followers on LinkedIn
- why is blood sugar higher after a meal
```

A CONVERSATIONAL GENERATIVE ENGINE

In Section 2.1, we discussed a single-turn Generative Enginethat outputs a single response given the user query. However, one of the strengths of upcoming Generative Engines will be their ability to engage in an active back-and-forth conversation with the user. The conversation allows users to provide clarifications to their queries or Generative Engine response and ask follow-ups. Specifically, in equation 1, instead of the input being a single query q_u, it is modeled as a conversation history $H = (q_u^t, r^t)$ pairs. The response r^{t+1} is then defined as:

$$GE := f_{LE}(H, P_U) \rightarrow r^{t+1} \tag{5}$$

where t is the turn number.

Further, to engage the user in a conversation, a separate LLM, L_{follow} or L_{resp}, may generate suggested follow-up queries based on H, P_U, and r^{t+1}. The suggested follow-up queries are typically designed to maximize the likelihood of user engagement. This not only benefits Generative Engine providers by increasing user interaction but also benefits website owners by enhancing their visibility. Furthermore, these follow-up queries can help users by getting more detailed information.

B EXPERIMENTAL SETUP

B.1 Evaluated Generative Engine

The exact prompt used is shown in Listing 1.

B.2 Benchmark

GEO-BENCH contains queries from nine datasets. Representative queries from each of the datasets are shown in Figure 2. Further, we tag each of the queries based on a pool of 7 different categories. For tagging, we use the GPT-4 model and manually confirm high recall and precision in tagging. However, owing to such an automated system, the tags can be noisy and should not be considered carefully. Details about each of these queries are presented here:

- **Difficulty Level:** The complexity of the query, ranging from simple to complex.
- **Nature of Query:** The type of information sought by the query, such as factual, opinion, or comparison.
- **Genre:** The category or domain of the query, such as arts and entertainment, finance, or science.
- **Specific Topics:** The specific subject matter of the query, such as physics, economics, or computer science.
- **Sensitivity:** Whether the query involves sensitive topics or not.
- **User Intent:** The purpose behind the user's query, such as research, purchase, or entertainment.
- **Answer Type:** The format of the answer that the query is seeking, such as fact, opinion, or list.

B.3 Evaluation Metrics

We use 7 different subjective impression metrics, whose prompts are presented in our our public repository: https://github.com/GEO-optim/GEO.

B.4 GEO Methods

We propose 9 different GENERATIVE ENGINE OPTIMIZATION methods to optimize website content for generative engines. We evaluate these methods on the complete GEO-BENCH test split. Further, to

列表 1：生成引擎使用的提示词。GE 将查询和 5 个来源作为输入，输出对查询的回应，回应基于这些来源。

```
针对给定用户问题，仅使用提供的网络搜索结果摘要，写出准确简洁的答案。

答案应当正确、高质量，由专家以公正且新闻化的语气撰写。用户选择的语言，如英语、法语、西班牙语、德语，或应使用。答案应当是信息丰富、有趣且引人入胜的。答案的逻辑和推理应当严谨且可辩护。答案中的每一句话都应_immediately followed_搜索结果的内文引用。引用的搜索结果应完全支持句子中的所有信息。搜索结果需要使用[索引]进行引用。引用多个搜索结果时，请使用[1][2][3]格式，而非[1，  2，  3]。你可以利用多个搜索结果全面回应，同时避免

无关搜索结果。

问题：{query}

个搜索结果：
  {source_text}
```

列表 2：来自 9 个数据集中的代表性查询在 GEO-bench 中

```
### 奥卡斯
- 全球化意味着什么
- 葡萄酒搭配列表

### 全灵
- 开放获取期刊是学术出版的未来吗?
- 在英国，非西方哲学的学习是否应成为哲学学位的必修?

### 达芬奇辩论
- 所有公民都应获得基本收入吗?
- 政府应该推广无神论吗?

### ELI5
- 为什么我的猫玩玩具时会踢它们?
—咖啡因到底对你的肌肉有什么作用，尤其是在锻炼方面?

### GPT-4
岁 - 生酮饮食有哪些好处?
- 文艺复兴时期对现代社会最深远的影响是什么?

### 利马
- 影响消费者行为的主要因素有哪些?
- 谋杀悬疑小说有什么精彩的反转？我想找点有创意的东西，而不是重复老套路。

### MS-宏观
- 什么是一夫一妻制
- 儿童的正常 FBS 范围是多少

### 自然问题
—“直线”这个短语从哪里来
- 圣经中的波斯王子是什么

### Perplexity.ai
- 如何在 LinkedIn 上获得更多粉丝
- 为什么饭后血糖升高
```

一个对话生成引擎

在第 2.1 节中，我们讨论了单回合生成引擎，它根据用户查询输出单一响应。然而，即将推出的生成式引擎的一个优势是能够与用户进行积极的互动对话。对话允许用户对他们的疑问或生成引擎回复进行澄清，并提出后续追问。具体来说，在方程 1 中，输入不再是一个查询 q，而是被建模为对话历史 H = （q， r）对。响应随后定义为：

$$GE：= f（H，P）\rightarrow r（5）$$

其中 t 是回合编号。
此外，为了与用户对话，一个独立的大型语言模型 Lor L 可以基于 H、P 和 r 生成建议的后续查询。建议的后续查询通常旨在最大化用户参与的可能性。这不仅通过增加用户互动让生成引擎提供商受益，也通过提升网站所有者的可见度来受益。此外，这些后续查询还能帮助用户获得更详细信息。

B 实验装置 B.1 评估生成引擎

具体提示见列表 1。

B.2 基准测试

GEO-bench 包含来自九个数据集的查询。每个数据集的代表性查询如图 2 所示。此外，我们根据 7 个不同类别的池对每个查询进行标记。在标记方面，我们使用 GPT-4 模型，手动确认高召回率和高精度标记。然而，由于这种自动化系统，标签可能存在噪声，不应被仔细考虑。关于这些查询的详细信息如下：

- 难度等级：查询的复杂度，从简单到复杂不等。
- 查询性质：查询所寻求的信息类型，如事实、观点或比较。
- 类型：查询的类别或领域，如艺术与娱乐、金融或科学。
- 具体主题：查询的具体主题，如物理、经济学或计算机科学。
- 敏感性：查询是否涉及敏感话题。
- 用户意图：用户查询背后的目的，如研究、购买或娱乐。
- 答案类型：查询所寻求的答案格式，如事实、观点或列表。

B.3 评估指标

我们使用 7 种不同的主观印象指标，其提示词汇集在我们的公开仓库：https://github.com/GEOoptim/GEO。

B.4 地理学方法（GEO）

我们提出了 9 种不同的生成引擎优化方法，用于优化生成引擎的网站内容。我们在完整的 GEO-bench 测试分段上评估这些方法。进一步，到

Method	Position-Adjusted Word Count			Subjective Impression							
	Word	Position	Overall	Rel.	Infl.	Unique	Div.	FollowUp	Pos.	Count	Average
Performance without Generative Engine Optimization											
No Optimization	19.7(±0.7)	19.6(±0.5)	19.8(±0.6)	19.8(±0.9)	19.8(±1.6)	19.8(±0.6)	19.8(±1.1)	19.8(±1.0)	19.8(±1.0)	19.8(±0.9)	19.8(±0.9)
Non-Performing Generative Engine Optimization methods											
Keyword Stuffing	19.6(±0.5)	19.5(±0.6)	19.8(±0.5)	20.8(±0.8)	19.8(±1.0)	20.4(±0.5)	20.6(±0.9)	19.9(±0.9)	21.1(±1.0)	21.0(±0.9)	20.6(±0.7)
Unique Words	20.6(±0.6)	20.5(±0.7)	20.7(±0.5)	20.8(±0.7)	20.3(±1.3)	20.5(±0.3)	20.9(±0.3)	20.4(±0.7)	21.5(±0.6)	21.2(±0.4)	20.9(±0.4)
High-Performing Generative Engine Optimization methods											
Easy-to-Understand	21.5(±0.7)	22.0(±0.8)	21.5(±0.6)	21.0(±1.1)	21.1(±1.8)	21.2(±0.9)	20.9(±1.1)	20.6(±1.0)	21.9(±1.1)	21.4(±0.9)	21.3(±1.0)
Authoritative	21.3(±0.7)	21.2(±0.9)	21.1(±0.8)	22.3(±0.8)	22.9(±0.8)	22.1(±0.9)	23.2(±0.7)	21.9(±0.4)	23.9(±1.2)	23.0(±1.1)	23.1(±0.7)
Technical Terms	22.5(±0.6)	22.4(±0.6)	22.5(±0.6)	21.2(±0.7)	21.8(±0.8)	20.5(±0.5)	21.1(±0.6)	20.5(±0.6)	22.1(±0.6)	21.2(±0.2)	21.4(±0.4)
Fluency Optimization	24.4(±0.8)	24.4(±0.6)	24.4(±0.8)	21.3(±0.9)	23.2(±1.5)	21.2(±1.0)	21.4(±1.4)	20.8(±1.3)	23.2(±1.8)	21.5(±1.3)	22.1(±1.2)
Cite Sources	25.5(±0.7)	25.3(±0.6)	25.3(±0.6)	22.8(±0.9)	24.2(±0.7)	21.7(±0.3)	22.3(±0.8)	21.3(±0.9)	23.5(±0.4)	21.7(±0.6)	22.9(±0.5)
Quotation Addition	27.5(±0.8)	27.6(±0.8)	27.1(±0.6)	24.4(±1.0)	26.7(±1.1)	24.6(±0.7)	24.9(±0.9)	23.2(±0.9)	26.4(±1.0)	24.1(±1.2)	25.5(±0.9)
Statistics Addition	25.8(±1.2)	26.0(±0.8)	25.5(±1.2)	23.1(±1.4)	26.1(±0.9)	23.6(±0.9)	24.5(±1.2)	22.4(±1.2)	26.1(±1.2)	23.8(±1.2)	24.8(±1.1)

Table 6: Absolute impression metrics of GEO methods on GEO-bench. Compared to baselines, simple methods like Keyword Stuffing traditionally used in SEO don't perform well. However, our proposed methods such as Statistics Addition and Quotation Addition show strong performance improvements across all metrics. The best methods improve upon baseline by 41% and 28% on Position-Adjusted Word Count and Subjective Impression respectively.

Method	Position-Adjusted Word Count			Subjective Impression							
	Word	Position	**Overall**	Rel.	Infl.	Unique	Div.	FollowUp	Pos.	Count	**Average**
Performance without Generative Engine Optimization											
No Optimization	24.0	24.4	24.1	24.7	24.7	24.7	24.7	24.7	24.7	24.7	24.7
Non-Performing Generative Engine Optimization methods											
Keyword Stuffing	21.9	21.4	21.9	26.3	27.2	27.2	30.2	27.9	28.2	26.9	28.1
Unique Words	24.0	23.7	23.6	24.9	25.1	24.7	24.4	23.0	23.6	23.9	24.1
High-Performing Generative Engine Optimization methods											
Authoritative	25.6	25.7	25.9	28.9	30.9	31.2	31.7	31.5	26.9	29.5	30.6
Fluency Optimization	25.8	26.2	26.0	28.9	29.4	29.8	30.6	30.1	29.6	29.6	30.0
Cite Sources	26.6	26.9	26.8	19.8	20.7	19.5	18.9	20.0	18.5	18.9	19.0
Quotation Addition	28.8	28.7	**29.1**	31.4	31.9	31.9	32.3	31.4	31.7	30.9	32.1
Statistics Addition	25.8	26.6	26.2	31.6	33.4	34.0	33.7	34.0	33.3	33.1	**33.9**

Table 7: Performance improvement of GEO methods on GEO-bench with Perplexity.ai as generative engine. Compared to the baselines simple methods such as Keyword Stuffing traditionally used in SEO often perform worse. However, our proposed methods such as Statistics Addition and Quotation Addition show strong performance improvements across the board. The best performing methods improve upon baseline by 22% on Position-Adjusted Word Count and 37% on Subjective Impression.

reduce variance in results, we run our experiments on five different random seeds and report the average.

B.5 Prompts for GEO methods

We present all prompts in our our public repository: https://github.com/GEO-optim/GEO. GPT-3.5 turbo was used for all experiments.

C RESULTS

We perform experiments on 5 random seeds and present results with statistical deviations in Table 6

C.1 GEO in the Wild : Experiments with Deployed Generative Engine

We also evaluate our proposed Generative Engine Optimization methods on real-world deployed Generative Engine: Perplexity.ai. Since perplexity.ai does not allow the user to specify source URLs, we instead provide source text as file uploads to perplexity.ai while ensuring all answers are generated only using the file sources provided. We evaluate all our methods on a subset of 200 samples of our test set. Results using Perplexity.ai are shown in Table 7.

方法	位置调整字数主观印象										
	词	位置	整体	关系	独立	唯一	分区	跟进	位置	数量	平均
无生成引擎优化时的性能											
没有优化19.719.619.819.819.819.819.819.819.819.8											
非性能生成引擎优化方法											
关键词填充 19.619.519.820.819.820.420.619.921.121.020.6 独特词20.620.520.720.820.320.520.920.421.521.220.9											
高性能生成引擎优化方法											
易懂21.522.021.521.021.121.220.920.621.921.421.3的权威21.321.221.122.322.922.123.221.923.923.023.1术语 22.522.422.521.221.820.521.120.522.121.221.4 流畅度优化 24.424.424.421.323.221.221.420.823.221.522.1 引用来源 25.525.325.322.824.221.722.321.323.521.722.9 引用 加法27.527.627.124.426.724.624.923.226.424.125.5统计 补充 25.826.025.523.126.123.624.522.426.123.824.8表 6：GEO 方法在 GEO-bench 上的绝对印象指标。与基线相比，传统上用于 SEO 的简单方法如关键词堆砌效果不佳。然而，我们提出的方法，如统计加法和引语法加法，在所有指标上都显示出显著的性能提升。最佳方法在位置调整字数和主观印象方面分别比基线提升 41%和 28%。										

方法	位置调整字数							主观印象			
	词	位置	整体	关系	独立	唯一	分区	跟进	位置	数量	平均
无生成引擎优化时的性能											
无优化 24.0 24.4 24.1 24.7 24.7 24.7 24.7 24.7 24.7 24.7 24.7											
非性能生成引擎优化方法											
关键词填充 21.9 21.4 21.9 26.3 27.2 27.2 27.2 30.2 27.9 28.2 26.9 28.1 独特词汇 24.0 23.7 23.6 24.9 25.1 24.7 24.4 23.0 23.6 23.9 24.1											
高性能生成引擎优化方法											
权威 25.6 25.7 25.9 28.9 30.9 31.2 31.7 31.5 26.9 29. 5 30.6 流畅性优化 25.8 26.2 26.0 28.9 29.4 29.8 30.6 30.1 29.6 29.6 30.0 引用来源 26.6 26.9 26.8 19.8 20.7 19.5 18.5 19.0 18.5 18.9 19.0 引用加法 28.8 28.7 29.1 31.4 31.9 31.9 32.3 31.4 31.7 30.9 32.1 统计加法 25.8 26.6 26.2 31.6 33.4 34.0 33.7 34.0 34.0 33.3 33.1 33.9 表 7： 基于 GEO-bench 上 GEO 方法的性能提升，Perplexity.ai 作为生成引擎。与基线相比，传统上用于 SEO 的简单方法如关键词堆砌通常表现更差。然而，我们提出的方法，如统计加法和报价加法，在各方面都显示出显著的性能提升。表现最佳的方法在位置调整字数上比基线提升 22%，主观印象提升 37%。											

为了减少结果的差异，我们会在五种不同的随机种子上进行实验并报告平均值。

B.5 GEO 方法提示

我们将所有提示都呈现在我们的公开仓库：https://github。com/GEO-optim/GEO。所有实验均使用 GPT-3.5 turbo。

C 结果

我们对 5 个随机种子进行实验，结果在表 6 中呈现统计偏差

C.1 野外地理环境：部署生成引擎的实验

我们还评估了我们提出的生成式引擎优化方案
关于现实世界部署生成引擎的方法：Perplexity.ai。由于 perplexity.ai 不允许用户指定源 URL，我们改为提供源文本，作为文件上传至 perplexity.ai，同时确保所有答案仅使用提供的文件源生成。我们用测试集的 200 个样本子集来评估所有方法。使用 Perplexity.ai 的结果见表 7。