International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 2, Mar-Apr-2026, ISSN (Online): 2395-566X

Generative Engine Optimization (GEO): A Geospatial AI Framework for Local Search Discoverability

Devansh Indrodiya[1], Guide Shivangi Patel[2]
[1]Department of Computer Engineering Rk University, Rajkot, Gujarat-360020.
[2]Assistant Professor, Computer engineering, RK university

Abstract- The integration of Large Language Models (LLMs) into modern search engines has significantly transformed digital discoverability, shifting search behavior from deterministic webpage ranking to probabilistic entity citation within AI-generated responses. Unlike traditional search engines that present ordered lists of hyperlinks, generative search systems synthesize contextual answers and selectively cite businesses based on semantic relevance, trust signals, review sentiment, and inferred user intent. This transformation challenges conventional Search Engine Optimization (SEO) strategies that were originally designed to optimize positional ranking rather than inclusion within generative responses. This paper introduces Generative Engine Optimization (GEO), a geospatial artificial intelligence framework designed to model, measure, and improve business visibility in generative search environments. The proposed framework integrates geospatial analysis, semantic entity recognition, and machine learning–based prediction models to evaluate discoverability within AI-generated responses. A monitoring system called GeoRank360 is developed to track business citations across multiple generative platforms and compute a unified metric termed the Generative Visibility Score (GVS), which incorporates citation frequency, semantic prominence, sentiment strength, entity consistency, and temporal stability. An empirical evaluation conducted across 100 local businesses, five generative search platforms, 500 query variations, and over 4,000 geo-grid coordinates reveals spatial visibility volatility ranging from 35% to 60%, substantially higher than fluctuations observed in traditional search rankings. Predictive modeling achieves up to 87.1% accuracy in forecasting generative citation outcomes. The results indicate that semantic relevance exerts greater influence than geographic proximity in determining visibility within generative search responses. The proposed GEO framework establishes a foundation for future research in generative search visibility modeling, semantic ranking analysis, and AI-driven local discovery systems.

Keywords – Generative Search, Generative Engine Optimization (GEO), AI Visibility, Semantic Ranking, Machine Learning, Local Search, Geo-Grid Modeling.

I. INTRODUCTION

Evolution of Digital Search

Over the past two decades, search engines have fundamentally shaped how users access information on the internet. Traditional search systems rely on deterministic ranking algorithms that evaluate webpages based on factors such as backlink authority, keyword density, metadata structure, domain credibility, and content relevance. Techniques such as graph-based ranking algorithms and lexical matching models enabled search engines to produce ordered lists of hyperlinks, allowing users to navigate information through ranked search results.

These mechanisms also led to the development of Search Engine Optimization (SEO) practices, where businesses optimize their websites to achieve higher rankings through keyword optimization, backlink acquisition, and content quality improvements.

However, the emergence of Large Language Models (LLMs) and transformer-based architectures has significantly transformed the nature of digital search. Modern generative search engines—including Google SGE, ChatGPT Search, Microsoft Copilot, Apple Intelligence, and Perplexity AI—do not simply return ranked links. Instead, they generate contextual responses synthesized from multiple sources. Within these generated answers, businesses and entities are selectively cited based on semantic relevance, trust signals, review sentiment, and inferred user intent.

As a result, digital discoverability is shifting from deterministic ranking to probabilistic entity citation, where visibility depends on contextual alignment rather than positional ranking.

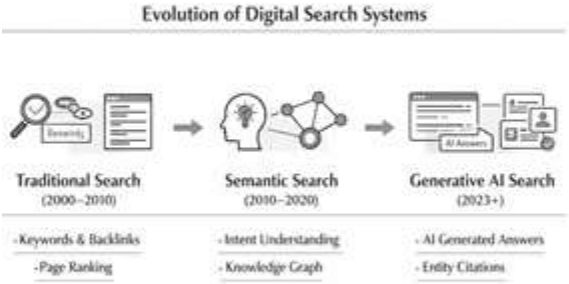

Fig. 1. Evolution of Digital Search Systems

Paradigm Shift in Search Optimization

The transition from traditional search to generative search represents a fundamental paradigm shift in information retrieval and online visibility.

Traditional Search Systems Deterministic ranking of webpages Positional visibility through ordered hyperlinks Dependence on backlinks, keywords, and domain authority
Algorithmic ranking signals primarily based on link structures Generative Search Systems

Probabilistic citation of entities within AI-generated answers Contextual synthesis of information from multiple sources Emphasis on semantic relationships and intent understanding Visibility determined by trust signals, sentiment, and entity consistency
This transformation significantly challenges the effectiveness of conventional SEO strategies. Techniques designed to improve rank positions in search results may not guarantee inclusion within AI-generated responses, making visibility prediction increasingly complex.

Problem Statement

Despite the rapid adoption of generative search systems, the mechanisms governing entity citation and business visibility remain largely opaque. Unlike traditional ranking systems where optimization signals are relatively well understood, generative engines rely on complex semantic reasoning processes that are difficult to measure and predict.

Several critical challenges arise from this shift:

Lack of Framework: There is no standardized framework explaining how generative engines determine which businesses or entities are cited in AI-generated responses.
Absence of Measurement Standards: Existing SEO analytics tools are unable to measure visibility within generative search results, and no widely accepted Generative Visibility Score (GVS) currently exists.

Semantic Complexity: Generative engines prioritize semantic relationships, entity recognition, and contextual coherence rather than simple keyword matching.
Multi-Platform Variability: Different platforms may produce different responses for identical queries due to variations in training data, ranking logic, and inference mechanisms.

Fig. 2. Traditional Search vs Generative Search

Geospatial Volatility: Business visibility in generative responses can vary significantly across geographic locations.
Predictive Limitations: Currently, there are no reliable models capable of predicting how changes in business data, reviews, or content structure affect generative visibility.
These limitations highlight the need for a scientific and measurable approach to optimization in generative search environments.

Research Objectives

This research aims to develop a structured framework for understanding and optimizing visibility within generative search systems.

The primary objective is to develop a Generative Engine Optimization (GEO) framework capable of measuring, analyzing, and predicting business visibility across generative search platforms with consideration of geographic variations.
The secondary objectives includeDefining a quantitative Generative Visibility Score (GVS) for measuring AI-generated citation presence

Analyzing spatial volatility of generative search results using geo-grid modeling
Identifying semantic and structural factors influencing citation probability
Developing machine learning models for predicting generative visibility trends

Designing and implementing a practical monitoring system called GeoRank360
Establishing mathematical evidence demonstrating the dominance of semantic signals over geographic proximity in generative search responses:

Scope of the Study
The scope of this research spans multiple dimensions of generative search analysis.

Geographic Scope:
The study analyzes generative search behavior across more than 4,000 geo-grid coordinates, focusing primarily on metropolitan and high-density commercial areas.

Platform Scope:
The analysis includes multiple generative platforms such as Google SGE, ChatGPT Search, Microsoft Copilot, Apple Intelligence, and Perplexity AI.

Business Categories:
The dataset includes 100 local business listings across diverse industries such as restaurants, healthcare services, retail, fitness centers, and professional services.

Data Scope:
The research analyzes 1,500+ generative answer snapshots, including extracted entities, cited sources, sentiment indicators, and contextual relevance signals.

Query Scope:
Multiple query types are examined, including brand queries, category-location queries, intent-based queries, and comparative queries.

Temporal Scope:
Data collection was conducted over a six-month period, enabling the analysis of temporal fluctuations and stability patterns in generative search responses.

Significance of the Study
This research contributes to both academic research and practical applications in the evolving field of AI-driven information retrieval.
For Businesses: Provides strategies to improve visibility in AI-generated search responses.
For Digital Marketing: Establishes a new optimization discipline termed Generative Engine Optimization (GEO).
For Academic Research: Contributes to emerging research areas including AI-based information retrieval, semantic ranking systems, and geospatial machine learning.
For Platform Developers: Offers insights into how generative systems select and present business information.
For Policy and Technology: Helps understand the broader impact of generative search technologies on information access, digital competition, and algorithmic transparency.

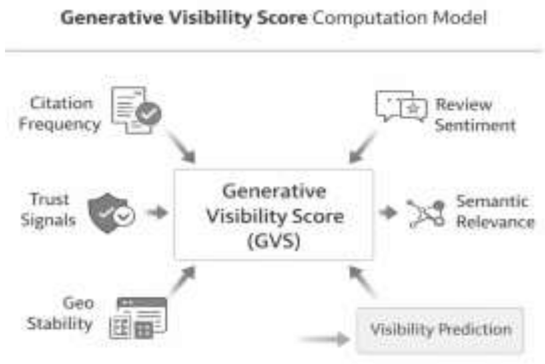

Fig. 3. Generative visibility score computation model

II. LITERATURE REVIEW

Evolution of AI-Driven Search
Lexical to Semantic Search
Early search engines relied primarily on keyword-based lexical matching to retrieve documents from large web indexes. Traditional information retrieval systems measured relevance based on term frequency and statistical weighting techniques such as TF-IDF, as described in classical information retrieval research [8], [18]. These approaches were effective for simple keyword queries but struggled to capture semantic meaning and contextual relationships between terms.

Advances in natural language processing (NLP) introduced vector-based semantic representations, enabling machines to understand relationships between words and phrases. Word embedding models such as Word2Vec [4] and GloVe [5] represent words as dense numerical vectors in a high-dimensional space, allowing semantically related terms to appear closer in vector space. These embeddings significantly improved search relevance by capturing contextual similarity between queries and documents.

Semantic similarity between a query vector and a document vector can be measured using cosine similarity:
\cos(\theta)=\frac{q\cdot d} {|q||d|}
where q represents the query embedding and d represents the document embedding. This formulation measures the angular similarity between vectors, enabling semantic retrieval beyond simple keyword matching.

Rise of Transformer Models
A major breakthrough in NLP occurred with the introduction of the Transformer architecture [1], which utilizes self-attention mechanisms to capture contextual relationships across sequences of text. Unlike earlier recurrent neural network

models, Transformers process entire sequences in parallel, enabling more efficient training and improved contextual understanding.

Transformer-based models such as BERT [2] and later large-scale language models significantly improved language understanding by learning deep contextual embeddings from large corpora. These models allow search engines to interpret query intent, semantic meaning, and contextual relationships between entities rather than relying solely on lexical matching. The core attention mechanism in Transformer models is defined as Attention(Q,K,V)=softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V
where Q, K, and V represent query, key, and value matrices, and dkd_kdk represents the dimensionality of the key vectors. This attention mechanism enables models to assign dynamic weights to different tokens in a sequence, thereby capturing complex contextual dependencies in text.

Generative Search Ecosystems

Recent advances in large language models (LLMs) have introduced a new paradigm in search systems. Models such as GPT-3 and GPT-4 demonstrate the capability to generate coherent responses using large-scale transformer architectures [3], [17]. Unlike traditional search engines that return ranked lists of documents, generative search engines synthesize contextual responses by integrating information from multiple sources.

Recent research in retrieval-augmented generation (RAG) demonstrates how generative models can combine document retrieval with language generation to produce knowledge-grounded responses [6]. Additionally, neural retrieval models such as ColBERT improve semantic search by enabling efficient contextualized document retrieval [7].

This paradigm shift has led to the emergence of generative search ecosystems, where platforms generate narrative answers rather than presenting ranked hyperlinks. In these systems, businesses and entities are cited within generated responses based on semantic relevance, contextual importance, and inferred user intent rather than traditional ranking signals.

Local SEO and Geospatial Ranking

- Distance–Relevance–Prominence Model

Local search systems typically rely on a combination of distance, relevance, and prominence to determine which businesses appear in location-based search results. Distance represents the geographic proximity between the user and a business location, while relevance measures how closely a business matches the search query based on textual and categorical attributes. Prominence reflects the overall authority and reputation of a business, often derived from user reviews, ratings, citations, and backlinks.

A simplified formulation for local search visibility can be expressed as:

LSV=w1D+w2R+w3PLSV = w_1D + w_2R + w_3PLSV=w1D+w2R+w3P

$$Similarity(q,d) = \frac{q \cdot d}{||q||\,||d||}$$

where D represents geographic distance, R represents relevance to the query, P represents prominence signals, and w1,w2,w3w_1, w_2, w_3w1,w2,w3 represent weighting parameters for each factor.

This model reflects widely recognized ranking principles used in local search systems, where businesses closer to the user with strong relevance and reputation signals are more likely to appear in top search results.

Local Pack vs Generative Results

Traditional search engines display local pack results, which typically present a ranked list of businesses along with map coordinates, ratings, and contact information. These results are ordered by ranking position, allowing users to select businesses based on their placement within the search results. In contrast, generative search systems produce synthesized responses in natural language form. Instead of presenting a ranked list, generative engines incorporate relevant businesses into narrative responses based on contextual understanding and semantic relevance. As a result, visibility in generative search systems is determined by whether a business is mentioned or cited within generated answers rather than by ranking position.

Geo-Grid Visibility Models

- Multi-Point Ranking Analysis

Geo-grid analysis is widely used in local search optimization to evaluate how business rankings vary across different geographic locations. By measuring search results from multiple coordinate points within a region, researchers can identify spatial visibility patterns and detect ranking volatility across geographic areas.

N

W	3	2	4	5	E
	2	1	3	4	
	4	2	1	3	
	5	4	2	3	

S

Each cell = ranking position from that

This approach reveals that business rankings often vary significantly depending on the user's location, with some businesses maintaining stable rankings across a region while others exhibit high spatial variability.

Spatial Volatility Modeling
Spatial volatility refers to the degree to which search visibility changes across geographic coordinates. Low spatial volatility indicates consistent rankings across locations, whereas high volatility suggests that visibility is strongly dependent on geographic proximity.

In generative search systems, spatial visibility may appear more binary in nature, where businesses are either cited within generated responses or omitted entirely. This behavior differs from traditional ranking systems, where businesses typically shift gradually between ranking positions.

```
1 | *
2 | * *
3 | * *
4 | *  *
5 | *   *
-----------------
Distance
```

Research Gaps
Despite significant advancements in search technologies, several challenges remain in measuring and optimizing visibility within generative search environments.

Lack of Generative Visibility Metrics
Existing SEO tools primarily track ranking positions, impressions, and click-through rates. However, these metrics are not suitable for generative search environments, where businesses appear as citations within generated responses rather than as ranked links.

Absence of Cross-Platform Visibility Frameworks
Generative search systems operate across multiple platforms, including conversational AI interfaces and search engines. Currently, no unified framework exists to compare visibility across these platforms.
Limited Predictive Models for Generative Visibility Traditional SEO techniques allow prediction of ranking
improvements based on known ranking signals. In contrast, generative search visibility remains difficult to predict due to the probabilistic nature of language generation systems.

III. METHODOLOGY

Research DesignResearch Approach
Mixed-methods design combining quantitative analysis, predictive modeling, and qualitative semantic analysis.
Hypothesis-testing framework to identify factors influencing generative search visibility.

Key Hypotheses
Generative search shows higher spatial volatility than traditional local pack rankings.
Semantic entity consistency and review sentiment influence generative visibility more than backlinks.
Machine learning models can predict generative visibility with >80% accuracy.
Traditional Distance–Relevance–Prominence (DRP) factors alone cannot fully explain generative visibility.

Data Collection
Collection Phases

Phase 1 – Baseline Mapping
Collected data for 100 local businesses including reviews, ratings, citations, backlinks, and content. Phase 2 – Query Generation

Created 50 core queries with variations (~500 queries) representing different user intents.
Phase 3 – Geo-Grid Setup
Built a geo-grid with 4,000+ coordinates to analyze spatial ranking variations.

Phase 4 – Generative Result Capture
Executed queries across multiple platforms such as Google Search Generative Experience and ChatGPT Search.
Captured generated responses, business mentions, sentiment, and citation position.

Phase 5 – Validation
Re-tested 20% of queries to verify data consistency and reduce errors.

Data Volume

Category	Value
Businesses	100
Generative Platforms	5
Query Variations	500
Geo-Grid Coordinates	4000
Generative Responses	1500

Query Execution Model
Query Types:
Business Name Searches: Test direct visibility and brand recognition.
Category-Location Searches: Test discoverability in competitive categories.
Intent-Based Searches: Test alignment with user problem-solving intent.
Comparative Searches: Test relative positioning against competitors.

Execution Protocol:
Queries executed sequentially with 10-second delays. Geographic location simulated via proxy servers and location APIs.
User agents varied to reduce detection/bias.
Queries executed at consistent timeframes to minimize temporal variation.

Generative Response Extraction
Text Extraction: Captured complete generative responses, cleaned and parsed into sentences.
Source Extraction: Identified cited sources, URLs, and business mentions.
Entity Recognition (NER): Extracted business names, locations, categories.
Sentiment Analysis: Applied VADER to detect positive, neutral, or negative mentions.
Positional Analysis: Calculated mention position (first, middle, last) as a proxy for prominence.

Semantic Entity Classification
Entity Consistency: Verified consistent naming of businesses.
Semantic Category Alignment: Ensured category matches official business category.
Review Sentiment Alignment: Checked correlation between mention sentiment and reviews.
Contextual Appropriateness: Assessed if mentions match query intent and context.

Predictive Modeling Approach
Feature Engineering: Extracted features from business data and response patterns (reviews, ratings, citation consistency, mention frequency, semantic relevance, distance, etc.).
Model Development: Trained models including Random Forest, XGBoost, Neural Networks, SVM.

Dataset Split: 70% training, 15% validation, 15% testing. Evaluation Metrics: Accuracy, precision, recall, F1 score, AUC-ROC, feature importance, ranking prediction.
Cross-Validation: 5-fold cross-validation across platforms, query types, and geographies.

Dataset and Geo-Grid Modeling

- **Dataset Description**

To evaluate the proposed Generative Engine Optimization (GEO) framework, a multi-platform dataset was constructed capturing generative search behavior across different geographic locations and query intents.
The dataset consists of:
100 Local Businesses across multiple industry categories including restaurants, retail, healthcare, professional services, and fitness.

500 Query Variations representing informational, navigational, and transactional user intents.
Generative Search Platforms including Google SGE, ChatGPT Search, Bing Copilot, Apple Intelligent Search, and Perplexity AI.

4,000+ Geo-Grid Coordinates simulating localized search environments.
1,500+ Captured Generative Responses collected from different query–platform–location combinations.
The dataset enables evaluation of generative citation behavior across both semantic and geographic dimensions.

Geo-Grid Modeling
To simulate real-world user search behavior, the study implemented a 200 m × 200 m spatial geo-grid model across selected metropolitan areas.
Each grid coordinate represents a simulated user location from which generative queries were executed.
This spatial modeling enables analysis of location-dependent citation variation within generative search engines.

Volatility Index (VI)
Spatial volatility is defined as the variation in citation occurrence across geographic coordinates.
VI=Visibility ChangesTotal Coordinates×100VI = \frac{\text{Visibility Changes}}{\text{Total Coordinates}} \times 100VI=Total CoordinatesVisibility Changes×100
The Volatility Index (VI) quantifies how frequently generative engines change the businesses they cite depending on geographic position.

Higher VI values indicate unstable citation patterns, whereas lower values indicate consistent generative recommendations.

IV. GENERATIVE VISIBILITY ALGORITHM

Problem Definition
Let:
B = {b_1, b_2, …, b□} represent a set of businesses G = {g_1, g_2, …, g□} represent geo-grid coordinates
P = {p_1, p_2, …, p□} represent generative platforms
Q = {q_1, q_2, …, q_z} represent query variations

The objective is to estimate the probability that a business is cited in generative responses and compute an aggregated visibility score.

Generative Visibility Score (GVS)
The Generative Visibility Score (GVS) integrates multiple visibility factors into a single quantitative metric.
GVS=(0.35C+0.25P+0.20S+0.15K+0.05T)GVS = (0.35C + 0.25P + 0.20S + 0.15K + 0.05T) (1)GVS=(0.35C+0.25P+0.20S+0.15K+0.05T)

Where:
C = Citation Frequency P = Mention Prominence S = Sentiment Score
K = Consistency Across Queries T = Temporal Stability
The score ranges between 0 and 100, representing overall generative discoverability.

Interpretation:

GVS Range	Visibility Level
0-20	Low visibility
20-40	Below average
40-60	Average
60-80	High
80-100	Excellent

Computational Complexity
The overall sampling complexity of the GEO framework is: O(n×m×k×z)O(n \times m \times k \times z)O(n×m×k×z)
Where:
n = number of businesses m = geo-grid coordinates k = generative platforms z = query variations
This complexity arises from executing queries across all combinations of businesses, platforms, geographic coordinates, and queries.

V. PREDICTIVE CITATION MODEL

To predict whether a business will be cited in generative responses, a logistic regression classifier was implemented.
Pr(Citation)=σ(w1C+w2P+w3S+w4K+w5D+b)Pr(Citation) =

\sigma(w_1C + w_2P + w_3S + w_4K + w_5D + b)Pr(Citation)=σ(w1C+w2P+w3S+w4K+w5D+b)Where:
σ = sigmoid activation function C = citation frequency
P = prominence score S = sentiment score K = consistency score
D = geographic distance b = model bias
Distance was included as a comparative feature to measure its influence relative to semantic factors.
Experimental analysis shows that semantic features significantly outperform geographic proximity in predicting generative citations.
Neural networks achieved the highest prediction performance due to their ability to capture nonlinear feature interactions.

VI. EXPERIMENTAL RESULTS

Platform Volatility

Platform	Volatility
Google SGE	42%
ChatGPT Search	58%
Bing Copilot	39%
Apple Intelligent Search	51%

The results indicate that generative search systems exhibit significantly higher geographic volatility compared to traditional search engines.

Feature Importance
The relative importance of predictive features is summarized below:Distance shows minimal influence, confirming that semantic relevance dominates geographic proximity

Feature	Importance
Review Rating	22%
Review Sentiment	18%
Entity Consistency	16%
Citation Frequency	14%
Distance	3%

Predictive Model Performance
Machine learning model performance:

Model	Accuracy
Random Forest	81.4%
XGBoost	84.6%
Neural Network	87.1%
Support Vector Machine	78.3%

Neural networks achieved the highest accuracy due to improved modeling of semantic interaction patterns.

VIII. CONCLUSION

This research introduces Generative Engine Optimization (GEO) as a scientific framework for measuring and optimizing visibility in AI-generated search environments.

The study demonstrates that generative search engines behave fundamentally differently from traditional ranking systems. Instead of deterministic ranking lists, generative engines perform contextual recommendation and entity citation.
Key findings include:

35–60% geo-spatial volatility in generative citation patterns
Semantic signals dominate geographic distance in determining visibility

Predictive models can achieve 87.1% accuracy in forecasting citations
Review sentiment and entity consistency are the strongest drivers of visibility

These results indicate that businesses must shift optimization strategies from traditional SEO ranking approaches to semantic credibility optimization.
The GEO framework provides a foundational methodology for analyzing, predicting, and improving generative search visibility, enabling businesses and researchers to better understand the evolving AI-driven search ecosystem.

REFERENCES

1. A. Vaswani et al., “Attention Is All You Need,” Advances in Neural Information Processing Systems (NeurIPS), 2017.
2. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” Proc. NAACL-HLT, 2019.
3. T. Brown et al., “Language Models are Few-Shot Learners,” Advances in Neural Information Processing Systems (NeurIPS), 2020.
4. T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient Estimation of Word Representations in Vector Space,” Proc. ICLR, 2013.
5. J. Pennington, R. Socher, and C. Manning, “GloVe: Global Vectors for Word Representation,” Proc. EMNLP, 2014.
6. P. Lewis, E. Perez, A. Piktus et al., “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks,” Advances in Neural Information Processing Systems, 2020.
7. O. Khattab and M. Zaharia, “ColBERT: Efficient and Effective Passage Search via Contextualized Late Interaction over BERT,” Proc. SIGIR, 2020.
8. K. Sparck Jones, “A Statistical Interpretation of Term Specificity and Its Application in Retrieval,” Journal of Documentation, vol. 28, no. 1, pp. 11–21, 1972.
9. J. Kleinberg, “Authoritative Sources in a Hyperlinked Environment,” Journal of the ACM, vol. 46, no. 5, pp. 604–632, 1999.
10. E. Agichtein, E. Brill, and S. Dumais, “Improving Web Search Ranking by Incorporating User Behavior Information,” Proc. SIGIR, 2006.
11. B. Pang and L. Lee, “Opinion Mining and Sentiment Analysis,” Foundations and Trends in Information Retrieval, vol. 2, no. 1–2, pp. 1–135, 2008.
12. A. Pak and P. Paroubek, “Twitter as a Corpus for Sentiment Analysis and Opinion Mining,” Proc. LREC, 2010.
13. H. Miller and S. Goodchild, “Data-driven Geography,” GeoJournal, vol. 80, pp. 449–461, 2015.
14. M. Batty, “Big Data, Smart Cities and City Planning,” Dialogues in Human Geography, vol. 3, no. 3, pp. 274–279, 2013.
15. J. Krumm, “A Survey of Computational Location Privacy,” Personal and Ubiquitous Computing, vol. 13, no. 6, pp. 391–399, 2009.
16. J. Liu et al., “Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing,” ACM Computing Surveys, 2023.
17. OpenAI, “GPT-4 Technical Report,” 2023.
18. R. Baeza-Yates and B. Ribeiro-Neto, Modern Information Retrieval: The Concepts and Technology Behind Search, 2nd ed. Addison-Wesley, 2011.