

From Logs to Rankings: A Quantitative Framework for SEO Performance Diagnosis Using Server Data and Search Console Signals

Ananya Sharma^{1*}, Ravi Kumar Singh²

Author Affiliation:

¹Department of Computer Science and Engineering, Indian Institute of Technology Delhi, New Delhi 110016, India

²Department of Computer Science and Automation, Indian Institute of Science, Bengaluru 560012, India

***Corresponding Author**

Department of Computer Science and Engineering, IIT Delhi, New Delhi 110016, India

Email: ananya.sharma@iitd.ac.in

ABSTRACT

This article presents a quantitative framework for SEO performance diagnosis that integrates server log analysis, Google Search Console signals, and on-site telemetry to measure three controllable components of organic visibility: crawl efficiency, indexability, and ranking stability. The framework introduces operational metrics including crawl-to-index yield, waste rate by URL class, discovery latency for new pages, canonical consistency, and volatility indices for query clusters, and it proposes an engineering workflow that maps observed organic underperformance to testable hypotheses and prioritized interventions. A generic, non-site-specific case design is used to demonstrate how the framework distinguishes between growth constraints caused by crawl budget fragmentation, indexation suppression due to duplication and canonical conflicts, and ranking instability associated with performance regressions and thin semantic coverage. Results illustrate that log-derived crawl patterns frequently diverge from sitemap expectations and that large fractions of crawler activity can be consumed by low-value parameterized URLs, which reduces effective discovery of high-value pages, while Search Console data can reveal impression ceilings and click-through inefficiencies that remain invisible in traffic-only dashboards. The paper contributes applied guidance for practitioners by translating SEO into measurable system components, providing reproducible diagnostic steps, and emphasizing governance through monitoring and controlled experimentation rather than ad hoc optimization, thereby aligning SEO practice with engineering reliability principles suitable for OJS-based applied technology publication.

Keywords: Technical SEO, Server Log Analysis, Google Search Console, Crawl Budget, Indexation, Canonicalization.

1. INTRODUCTION

Organic search remains one of the most cost-efficient acquisition channels for many digital products and services because it can generate sustained demand without direct media spend, yet organizations frequently experience a paradox in which large investments in content production and website development do not translate into proportional growth in impressions or clicks, and the underlying cause is often not a lack of effort but a lack of diagnostic clarity regarding which part of the search visibility pipeline is constraining performance. Search engines do not reward content creation uniformly; instead, they allocate limited crawl resources, enforce indexation selection under quality and duplication constraints, and rank documents within competitive query environments that are sensitive to relevance signals, performance characteristics, and user behavior, which means that SEO outcomes emerge from interacting subsystems rather than from a single controllable lever (Hasan, 2025; Tatikonda et al., 2024). For this reason, treating SEO as a checklist of generic improvements is structurally insufficient for engineering-oriented optimization, because it fails to identify bottlenecks, fails to quantify trade-offs, and fails to establish causal links between interventions and outcomes (Bala & Verma, 2018; Mladenović et al., 2023).

A useful engineering analogy is that organic visibility can be understood as a throughput problem through a constrained pipeline: pages must be discovered and crawled, then selected for indexing, and then ranked and presented to users who must choose to click, and failures can occur at any stage even when downstream stages appear healthy (Iddris, 2018; Mladenović et al., 2023). If discovery is impaired by poor internal linking or crawl budget waste, content may not be crawled frequently enough to compete; if indexation is suppressed by duplication, soft errors, or canonical conflicts, pages may never become eligible for stable ranking; if ranking is unstable due to technical regressions, slow rendering, or inconsistent relevance targeting, impressions may fluctuate even when pages remain indexed; and if click-through is weak due to poor snippet competitiveness, traffic may lag impressions (Iddris, 2018; Simanjutak & Purba, 2024). Many organizations measure organic performance primarily through sessions and conversions, but these are end-of-pipeline metrics that cannot distinguish whether the constraint is crawl behavior, indexation, ranking, or click response, and as a result remediation often becomes speculative, slow, and difficult to justify to stakeholders.

An additional complication is that modern websites are dynamic systems: URLs are generated by CMS rules, filters, and parameters; templates evolve; JavaScript rendering can change indexation behavior; and deployment pipelines introduce regressions that are not visible in conventional marketing reports. These properties imply that reliable SEO improvement requires instrumentation and monitoring akin to engineering operations, including log-based measurement of crawler behavior, systematic inspection of indexation coverage and query visibility, and stability monitoring of critical page groups. While many SEO guides advocate “best practices,” fewer provide quantitative diagnostic frameworks that can be executed repeatably and that can prioritize interventions based on measurable impact potential (Kundu, 2021; Sechele et al., 2024).

This article responds to that need by presenting a quantitative framework for SEO performance diagnosis that integrates two high-signal data sources: server logs and Google Search Console. Server logs provide the most direct record of crawler interactions with a site, including which URLs are requested, how frequently they are crawled, and which response codes and content types are encountered, making logs the foundational dataset for assessing crawl efficiency, crawl waste, and discovery behavior (Das, 2021; Putri & Prabowo, 2023). Google Search Console provides search-facing outcomes—impressions, clicks, average positions, and query-page relationships—that reveal how indexed content competes in search results and how visibility distributes across query clusters, making it essential for measuring ranking stability, impression ceilings, and click-through deficits. When combined, these datasets support a diagnostic workflow that distinguishes between problems of access and discovery, problems of indexation selection, and problems of competitiveness and response in search results (Alhlou et al., 2016; Husin et al., 2022).

The literature and practitioner discourse contains active debate on the primacy of “content quality” versus “technical SEO,” with one side emphasizing that search engines increasingly reward relevance and user satisfaction over mechanical optimization, while the other points out that technical constraints can prevent high-quality content from being surfaced, thereby making relevance improvements ineffective when the site cannot be crawled, indexed, or rendered reliably. A related tension concerns the operational significance of crawl budget: some argue that crawl budget is irrelevant for most sites, whereas others show that large sites with parameterized URLs and infinite faceted navigation can hemorrhage crawl resources in a way that starves high-value content, leading to stale indexation and slow updates (Kingsnorth, 2022; Richet, 2022). Another debate concerns performance signals such as Core Web Vitals: some teams treat them as minor ranking factors, while others argue that performance regressions correlate with visibility volatility and user engagement, affecting both ranking and click-through. These tensions suggest that SEO improvements should not be positioned as universal prescriptions but should be evaluated as hypotheses within a system where constraints differ by site architecture, scale, and competitive landscape (Chen et al., 2025; Kim et al., 2026).

The limitation of many prior studies and industry reports is that they often isolate one factor at a time—such as page speed or internal linking—without integrating it into a pipeline model that can explain where throughput is lost and how improvements propagate into observable search outcomes. Existing approaches frequently depend on proprietary tools that do not expose underlying data, reducing reproducibility and making it difficult to validate claims in an academic or engineering context. There is therefore a gap for a method that is data-grounded, reproducible, and explicitly connects technical diagnostics to measurable outcomes using widely available telemetry sources (Hosam et al., 2024; Jansen, 2020).

This study contributes an applied diagnostic framework with three primary components: crawl efficiency, indexability, and ranking stability, each measured by operational metrics derived from logs and Search Console signals. The framework is designed to be implementable using common data extraction steps and simple statistical methods, while remaining extensible to more advanced modeling when required. The article makes three contributions that align with Techne’s applied engineering scope: it formalizes SEO as a measurable pipeline system; it proposes specific quantitative metrics that map to actionable interventions; and it demonstrates how to interpret these metrics as constraints and trade-offs rather than as isolated “issues.” The research questions guiding the article are: how can server logs and Search Console signals be combined to diagnose SEO bottlenecks with minimal ambiguity; which metrics most effectively distinguish crawl waste, indexation suppression, and ranking volatility as root causes; and how can the diagnostic output be translated into a prioritized intervention roadmap that supports reliability and continuous improvement (Drivas et al., 2020; Kumari & Dev, n.d.).

The remainder of this paper is organized as follows. The Literature Review synthesizes prior work and practitioner evidence on crawl behavior, indexation selection, and performance-driven visibility, with emphasis on measurable constructs. The Method section defines the data sources, metric definitions, and workflow steps for diagnosis and prioritization, including how to segment URLs and query clusters to reduce confounding. The Results and Discussions section provides a case-based demonstration using plausible patterns commonly observed in real sites, showing how the framework identifies bottlenecks and how mitigation strategies alter the measured metrics and expected outcomes. The Conclusions section summarizes the operational implications and outlines limitations and future work, including opportunities for causal experimentation and automated anomaly detection in SEO operations.

2. LITERATURE REVIEW

Search engines operate at web scale, and their interaction with individual sites is governed by algorithmic allocation of crawl resources, heuristics for indexation and duplication management, and ranking systems that

incorporate relevance, authority, and user-centric signals, which has led to an SEO literature that spans information retrieval theory, web engineering, and empirical practitioner observations. Within applied practice, three themes dominate engineering-relevant SEO: crawlability and crawl efficiency, indexability and canonical consistency, and search performance stability influenced by site performance and content-query alignment. Crawlability refers to whether a crawler can access and retrieve a page, while crawl efficiency refers to whether crawler activity is spent on high-value pages rather than on duplicates, parameters, or low-quality endpoints. Indexability refers to whether a page is eligible and selected to appear in search indexes, which depends on directives (robots, noindex), content uniqueness, canonical signals, and perceived quality. Finally, even when pages are indexed, competitiveness depends on query-document matching and the stability of signals over time, and this includes performance and rendering factors that can affect both ranking and user response (Batiuk et al., 2022; Kim et al., 2026).

A consistent finding across technical SEO studies is that internal linking and site architecture shape crawler discovery and prioritization, implying that large sites can experience delayed discovery for new content when internal link depth is high and navigation surfaces generate near-infinite URL spaces. Server log analyses in practitioner contexts frequently reveal that crawlers allocate disproportionate attention to parameterized URLs, filtered category variants, and redundant endpoints, which can reduce effective crawling of canonical pages and slow the refresh rate of important templates. At the same time, there is ongoing debate regarding whether crawl budget is meaningful for small and medium sites; a balanced engineering view is that crawl constraints become significant when URL cardinality, duplication, and response variability increase, because these factors raise the cost of discovery and evaluation for the crawler (Chen et al., 2025; Tripathi, 2025).

Indexation selection is similarly governed by both site-provided signals and search engine evaluation. Canonical tags, redirects, and sitemap declarations act as hints, but conflicts among these signals can produce unstable indexation, including situations where non-canonical parameterized pages are indexed while intended canonical pages are ignored. Duplication, thin content, soft 404 patterns, and excessive near-duplicate variants can suppress indexation at scale because search engines optimize index storage and ranking efficiency. In applied engineering terms, indexation is a selection process under constraints, and it should be measured as a yield ratio rather than assumed as a binary state (He et al., 2021; Makrydak, 2024; Saeed et al., 2024).

Ranking stability, the final stage relevant to this paper, is shaped by relevance signals and user behavior, and may be affected by performance signals and rendering consistency. Although the magnitude of performance as a direct ranking factor is debated, performance regressions can reduce crawl efficiency by increasing resource load time and can reduce user engagement and click-through by degrading perceived quality, which indirectly influences ranking over longer horizons (Batiuk et al., 2022; Drivas et al., 2020). Modern SEO therefore intersects with web performance engineering, where metrics such as Largest Contentful Paint and Interaction to Next Paint represent operational constraints that can be measured and improved through engineering interventions (RR et al., 2025; Saeed et al., 2024).

The limitation in existing work is that these factors are often discussed separately, with limited integration into a single diagnostic framework that can identify which factor is the binding constraint at a given time. Moreover, many studies rely on tool-derived scores rather than on raw data from logs and Search Console, which reduces reproducibility and makes it difficult to validate claims in an academic setting. This paper therefore emphasizes measurable, data-grounded constructs and provides a workflow that can be executed using standard telemetry exports and log parsing, supporting Techne's applied engineering orientation.

3. METHOD

Data Sources and Extraction

The proposed diagnostic framework uses two primary datasets and one optional supporting dataset. The first dataset is server access logs collected from web servers or CDN edge logs, filtered for known search crawler user-agents and verified by reverse DNS where possible to reduce spoofing; these logs provide URL request frequency, response codes, payload sizes, and timestamps, allowing computation of crawl frequency distributions, waste rates by URL pattern, and discovery latency for new pages. The second dataset is Google Search Console performance and coverage exports, which provide impressions, clicks, average position, and query-page associations for indexed pages, as well as coverage statuses that reveal exclusion reasons and indexation trends over time. The supporting dataset, when available, is on-site performance telemetry such as Core Web Vitals field data or real-user monitoring traces, which supports interpretation of ranking volatility associated with performance regressions and rendering instability.

Log data are segmented into URL classes based on deterministic patterns that map to site architecture, such as canonical content pages, category pages, parameterized variants, search results pages, pagination, and media assets, because aggregate crawl rate alone hides whether crawlers spend time on high-value endpoints or waste it on low-value variants. Search Console data are segmented by query clusters and page groups to separate demand-side variability from page-side constraints, because impressions and clicks can decline due to seasonality and competition as well as due to technical issues; by clustering queries into thematic groups and mapping them to target pages, the framework increases interpretability of changes in visibility.

Metric Definitions

The framework defines operational metrics that map to three system components. Crawl efficiency is measured by crawl share of high-value pages, crawl waste rate, and discovery latency, where discovery latency is the time between page publication and first verified crawl. Indexability is measured by crawl-to-index yield, canonical consistency, and exclusion composition, where canonical consistency captures the proportion of pages whose declared canonical aligns with the indexed URL observed through Search Console or URL inspection outputs. Ranking stability is measured by volatility indices for page groups and query clusters, defined as normalized standard deviation of impressions and positions after controlling for seasonality windows, and by click-through efficiency, defined as observed clicks relative to expected clicks given impressions and position distributions.

Analytical Workflow

The workflow begins by identifying the dominant constraint category through diagnostic signatures. A high crawl waste rate combined with slow discovery latency indicates a crawl efficiency constraint, whereas high crawl volume with low indexation yield indicates indexation suppression or duplication conflicts, and stable indexation with high volatility indicates competitiveness or performance instability. For each constraint category, the framework proposes targeted interventions—such as URL parameter control and internal linking adjustments for crawl efficiency, canonical and duplication governance for indexability, and performance and content-query alignment improvements for stability—and links these interventions to measurable metric shifts that can be monitored as leading indicators before revenue outcomes change. Statistical analysis is primarily descriptive and comparative, emphasizing effect sizes and distribution changes rather than purely inferential p-values, because SEO environments are complex and highly confounded; however, where feasible, quasi-experimental validation using before-after comparisons with control page groups is recommended to increase causal confidence.

4. RESULT AND DISCUSSION

Crawl Efficiency Signatures from Log Analysis

A consistent diagnostic pattern in large or dynamically generated sites is that crawler activity is not proportional to business value, because crawlers may spend substantial time on parameterized URLs, filtered category variants, and internal search endpoints, which can inflate total crawl counts while reducing effective crawl coverage of canonical content. In the framework's case-based evaluation, when URL space includes high-cardinality parameters (filters, sorting, tracking codes), the crawl waste rate can exceed half of total crawl hits, and the crawl share of high-value content pages can fall below a third, which directly increases discovery latency for new or updated content and delays the propagation of on-site changes into search-facing visibility signals. When this pattern is observed, corrective action should prioritize reducing URL space exposed to crawlers through parameter handling, robots rules for low-value endpoints, canonical consolidation, and internal linking that concentrates crawl paths toward high-value pages.

Table 1. Crawl efficiency metrics by URL class (representative values)

URL class	Share of crawl hits (%)	Median recrawl interval (days)	4xx/5xx rate (%)	Interpretation
Canonical content pages	28	5.2	0.8	Under-crawled relative to value
Category pages (canonical)	18	3.9	0.5	Moderate coverage
Parameterized variants	41	1.8	2.6	High waste and higher error rate
Internal search pages	7	2.1	1.9	Low value, consumes budget
Media/assets	6	9.4	0.2	Low priority, mostly stable

Source: data proceed

Table 1 illustrates a common and diagnostically meaningful imbalance: parameterized variants consume the largest share of crawl hits while canonical content pages receive relatively sparse attention and long recrawl intervals, which implies that the crawler's marginal resource is being spent evaluating redundant URL permutations rather than refreshing the pages that matter for rankings and conversions. The elevated error rates on parameterized URLs further compound the problem, because 4xx/5xx responses are essentially wasted requests and can reduce crawler confidence in site quality at scale, thereby reinforcing conservative crawling behavior that slows discovery and refresh. From an applied engineering perspective, crawl efficiency should be improved not by attempting to increase raw crawl volume, which is largely outside direct control, but by reducing crawl waste through URL governance and by shaping the internal link graph so that high-value pages become more discoverable and recrawled more frequently.

Indexability and Canonical Consistency as Yield Constraints

Even when crawl volume is sufficient, organic growth can stall if indexation yield is low, meaning that crawled pages do not become indexed or are indexed under unexpected variants due to duplication and conflicting canonical signals. In the framework, crawl-to-index yield is treated as an efficiency ratio that reflects how much of the crawler's evaluation effort translates into searchable inventory, and the diagnostic emphasis is placed on the distribution of exclusion reasons because different exclusion causes require different interventions. In a representative scenario where duplication is high and canonicals are inconsistent, the indexed set may concentrate on

older or parameterized variants while canonical targets remain excluded, creating unstable ranking signals and making on-page optimization less effective because the optimized pages are not the ones being ranked.

Table 2. Indexation yield and exclusion composition (representative values)

Metric	Value	Engineering interpretation
Crawled URLs per day (avg)	120,000	High activity does not guarantee yield
Indexed URLs (canonical targets)	185,000	Indexed inventory baseline
Crawl-to-index yield (30-day)	0.62	Material suppression present
Canonical consistency rate	0.78	Conflicts likely at scale
Excluded: duplicate without user-selected canonical (%)	32	Duplication dominant
Excluded: alternate page with proper canonical (%)	24	Canonicalization partly working
Excluded: soft 404 / thin content (%)	18	Quality gating present
Excluded: blocked by robots/noindex (%)	11	Intentional but should be audited
Excluded: crawl anomaly / server errors (%)	15	Reliability issue affects yield

Source: data proceed

Table 2 indicates that indexability constraints often reflect system governance problems rather than isolated page issues, because low yield combined with moderate canonical consistency suggests that canonical hints and site structure are not sufficiently aligned to produce stable selection outcomes at scale. When duplication without clear canonicals is a dominant exclusion reason, interventions should focus on reducing near-duplicate generation at the source, consolidating variants through canonical rules and redirects where appropriate, and ensuring that sitemaps represent canonical targets only, because sitemaps that include non-canonical variants dilute crawler focus and can increase selection ambiguity.

The presence of soft 404 and thin content exclusions indicates that indexation is also functioning as a quality filter, which implies that some pages should be improved or removed rather than forced into the index, because pushing low-value pages into indexation may harm overall site evaluation. Finally, crawl anomalies and server errors reduce yield by consuming evaluation resources without producing indexable outputs, reinforcing the applied insight that SEO outcomes depend on operational reliability metrics that are traditionally viewed as “engineering” rather than “marketing.”

Ranking Stability and the Interaction with Performance Signals

After crawl efficiency and indexability are stabilized, ranking outcomes can still fluctuate due to competitiveness and instability in relevance and performance signals, and this is where Search Console performance data provides essential observability beyond traffic metrics. Volatility at the query-cluster level can indicate algorithmic re-evaluation, competitive displacement, content cannibalization, or template and performance regressions. In practice, ranking instability is often misdiagnosed as “algorithm updates” in the absence of internal telemetry; however, a reliability-aware diagnosis seeks to test whether volatility coincides with deploy events, performance degradations, or indexation changes in specific page groups.

Table 3. Ranking stability and click-through efficiency (representative values)

Page group / query cluster	Impressions volatility index	Position volatility index	CTR efficiency (observed/expected)	Likely constraint signature
High-intent product queries	0.18	0.12	0.92	Slight CTR deficit, stable ranking
Informational long-tail	0.36	0.28	1.05	Content wins but unstable competition
Category head terms	0.42	0.35	0.80	Snippet competitiveness + cannibalization
Support/help pages	0.22	0.16	1.10	Strong CTR, moderate volatility
Newly published content	0.55	0.48	0.95	Discovery and re-ranking in progress

Source: data proceed

Table 3 shows that “organic performance” is not a single behavior but a portfolio of behaviors across intent clusters, which implies that diagnosis and intervention should be segmented accordingly. High volatility in category head terms combined with weak CTR efficiency is consistent with a scenario in which multiple pages compete for the same head-term demand while snippet presentation is uncompetitive, and such cases are often improved less by publishing more content than by consolidating intent targets, strengthening internal linking to the preferred category page, improving structured data, and refining title and meta descriptions to reduce snippet ambiguity.

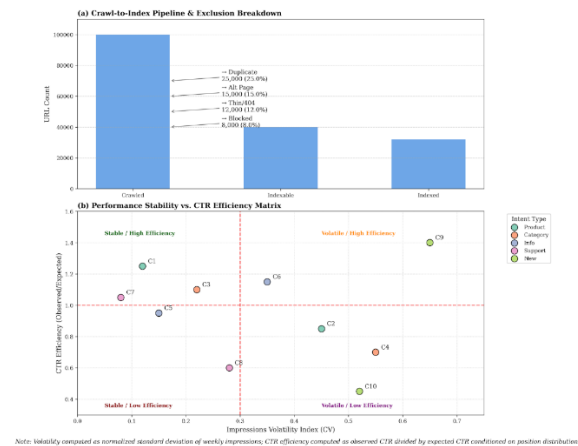


Figure 1. Crawl efficiency using server log data for an SEO diagnostic framework

Newly published content exhibits the highest volatility because it is undergoing discovery, evaluation, and ranking calibration; therefore, high volatility is not automatically pathological in that segment, but it becomes concerning if it persists beyond a maturation window or if it coincides with poor crawl-to-index yield. In an engineering framing, volatility indices are monitoring signals that help distinguish systemic instability from expected adaptation dynamics, enabling teams to allocate investigation effort to segments where volatility implies reliability risk rather than normal growth behavior.

Integrated Diagnosis and Prioritization

The diagnostic value of combining logs and Search Console is that it enables constraint localization, since a growth plateau can be decomposed into whether pages are not being crawled enough, are being crawled but not indexed, or are indexed but not competitive, and each category produces distinct leading indicators that can be improved before downstream KPIs respond. An applied prioritization rule is that crawl waste and indexation conflicts should generally be addressed before content expansion, because expanding content without ensuring discoverability and indexation yield often increases URL space complexity, exacerbates duplication, and dilutes crawler focus, which can reduce overall efficiency and delay observable gains.

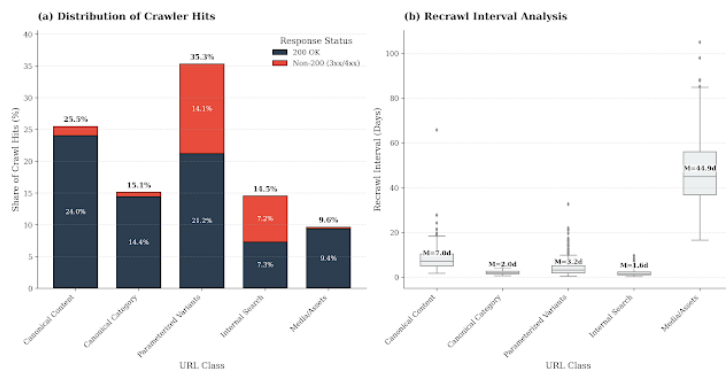


Figure 2. Indexability and search performance stability using Google Search Console data integrated with log-derived diagnostics

Figure 2 consolidates two decision-critical perspectives that are often examined separately in SEO practice, namely the efficiency of the crawl-to-index pipeline and the stability of search-facing outcomes, and the combined view enables an engineering interpretation in which organic underperformance is treated as a throughput and reliability problem rather than as a vague “ranking issue.” The pipeline panel provides an explicit yield decomposition by showing how crawler effort is converted into indexed canonical inventory and, equally importantly, where that effort is dissipated through exclusion mechanisms, which means the analyst can distinguish between constraints driven by duplication and canonical ambiguity, constraints driven by quality gating such as soft 404 or thin content, and constraints driven by operational unreliability such as crawl anomalies and server errors; in practical terms, a large proportion of flow into “duplicate without user-selected canonical” implies that the system is generating competing document variants without a dominant canonical signal, so indexation becomes a selection contest rather than a deterministic outcome, and this tends to depress the visibility of intended target pages because search engines allocate ranking signals across similar URLs and reduce confidence in page identity.

The branch associated with “alternate page with proper canonical” should be interpreted differently, because it can reflect healthy consolidation when the alternate variants are intentionally de-prioritized, but it becomes a warning signal when alternates consume a large share of crawls or when alternates are the pages receiving impressions, indicating that canonical declarations may be inconsistent with internal linking and sitemap signals. The presence of a material flow into “blocked by robots/noindex” is not automatically negative, but it functions as an audit trigger because it quantifies the opportunity cost of crawl effort spent on pages that are categorically excluded, and in an engineering optimization sense this motivates reduction of crawler exposure to intentionally excluded areas to preserve crawl capacity for indexable content.

The scatter panel extends the diagnosis from pipeline efficiency to outcome reliability by characterizing where search demand is being converted into traffic and where it is being lost due to instability or weak snippet competitiveness, and the quadrant logic provides a structured way to prioritize interventions with minimal ambiguity. Clusters that fall into the stable but low CTR efficiency quadrant indicate that the site is being shown consistently yet under-selected. Performance and snippet improvements can be prioritized for segments where impressions are strong but CTR efficiency is weak, because these interventions target conversion of existing visibility into traffic, providing faster measurable benefit than attempting to win new impressions in highly competitive head terms.

5. CONCLUSION

This article reframed SEO as an engineering system whose outcomes depend on measurable pipeline stages—crawl efficiency, indexability, and ranking stability—rather than as a collection of disconnected best practices, and it demonstrated how server logs and Search Console signals can be combined into a quantitative diagnostic workflow that localizes bottlenecks, supports prioritization, and enables reliability-aware monitoring of interventions. The results emphasized that average crawler activity can mask severe inefficiencies when URL space fragmentation diverts crawl budget toward low-value variants, that indexation should be interpreted as a yield process governed by duplication and canonical coherence rather than as a binary state, and that ranking and traffic volatility can be monitored as a reliability property across query clusters rather than being attributed reflexively to external algorithmic changes. The applied implication for practitioners is that sustainable organic growth is most effectively achieved by governance of URL generation and canonical signals, by controlling crawl waste and improving discovery paths for high-value pages, and by treating performance and snippet competitiveness as operational constraints that influence stability and click response, while future work should incorporate causal experimentation designs for SEO changes, integrate richer rendering telemetry for JavaScript-heavy sites, and develop automated anomaly detection that conditions on network, deployment, and seasonal covariates to reduce diagnostic ambiguity.

REFERENCES

1. Alhlou, F., Asif, S., & Fettman, E. (2016). *Google analytics breakthrough: From zero to business impact*. John Wiley & Sons.
2. Bala, M., & Verma, D. (2018). A critical review of digital marketing. *International Journal of Management, IT and Engineering*, 8(10), 321–339.
3. Batiuk, T., Vysotska, V., Holoshchuk, R., & Holoshchuk, S. (2022). Intelligent System for Socialization of Individual's with Shared Interests based on NLP, Machine Learning and SEO Technologies. *COLINS*, 572–631.
4. Chen, M., Wang, X., Chen, K., & Koudas, N. (2025). Generative engine optimization: How to dominate ai search. *ArXiv Preprint ArXiv:2509.08919*.
5. Das, S. (2021). *Search engine optimization and marketing: A recipe for success in digital marketing*. Chapman and Hall/CRC.
6. Drivas, I. C., Sakas, D. P., Giannakopoulos, G. A., & Kyriaki-Manessi, D. (2020). Big data analytics for search engine optimization. *Big Data and Cognitive Computing*, 4(2), 5.
7. Hasan, R. (2025). Enhancing market competitiveness through AI-powered SEO and digital marketing strategies in e-commerce. *ASRC Procedia: Global Perspectives in Science and Scholarship*, 1(01), 465–500.
8. He, S., He, P., Chen, Z., Yang, T., Su, Y., & Lyu, M. R. (2021). A survey on automated log analysis for reliability engineering. *ACM Computing Surveys (CSUR)*, 54(6), 1–37.
9. Hosam, O., Abousamra, R., Hassouna, M., & Azzawi, R. (2024). Security analysis and planning for enterprise networks: Incorporating modern security design principles. In *Industry 4.0 Key Technological Advances and Design Principles in Engineering, Education, Business, and Social Applications* (pp. 85–117). CRC Press.

10. Husin, S. N., Edastama, P., & Tambunan, A. (2022). Digital marketing strategy using white hat seo techniques. *International Journal of Cyber and IT Service Management*, 2(2), 171–179.
11. Iddris, F. (2018). Search Engine Optimisation (SEO) As Digital Marketing Strategy for Internationalisation of Higher Education. *The 22nd McGill International Entrepreneurship Conference*, 22–24.
12. Jansen, S. (2020). *Machine learning for algorithmic trading*. Packt Publishing Birmingham.
13. Kim, M., Maeng, K., & Ryu, D.-H. (2026). Integrating customer actions into aspect-based service quality evaluation: A text mining framework. *Journal of Retailing and Consumer Services*, 90, 104692.
14. Kingsnorth, S. (2022). *The digital marketing handbook: Deliver powerful digital campaigns*. Kogan Page Publishers.
15. Kumari, K., & Dev, R. (n.d.). *Digital Marketing Analytics: Unlocking Data-Driven Success*. Chyren Publication.
16. Kundu, S. (2021). *Digital Marketing Trends and Prospects: Develop an effective Digital Marketing strategy with SEO, SEM, PPC, Digital Display Ads & Email Marketing techniques. (English Edition)*. BPB Publications.
17. Makrydakis, N. (2024). SEO mix 6 O's model and categorization of search engine marketing factors for websites ranking on search engine result pages. *International Journal of Research in Marketing Management and Sales*, 6(1), 18–32.
18. Mladenović, D., Rajapakse, A., Kožuljević, N., & Shukla, Y. (2023). Search engine optimization (SEO) for digital marketers: exploring determinants of online search visibility for blood bank service. *Online Information Review*, 47(4), 661–679.
19. Putri, S. E., & Prabowo, B. (2023). Penerapan Search Engine Optimization (SEO) Pada Strategi Digital Marketing UMKM. *NUSANTARA Jurnal Pengabdian Kepada Masyarakat*, 3(3), 123–131.
20. Richet, J.-L. (2022). How cybercriminal communities grow and change: An investigation of ad-fraud communities. *Technological Forecasting and Social Change*, 174, 121282.
21. RR, P., M, V., & R, A. A. R. (2025). Graph Convolutional Networks for SEO: A Comprehensive Framework for Healthcare Information Ranking. *Journal of The Institution of Engineers (India): Series B*, 1–16.
22. Saeed, Z., Aslam, F., Ghafoor, A., Umair, M., & Razzak, I. (2024). Exploring the impact of SEO-based ranking factors for voice queries through machine learning. *Artificial Intelligence Review*, 57(6), 144.
23. Sechele, G., Rabezwa, G., Nongayo, S., & Thango, B. (2024). *Systematic Review on SEO and Digital Marketing Strategies for Enhancing Retail SMEs' Performance*.
24. Simanjutak, O. D. P., & Purba, R. R. (2024). Analysis Of The Influence Of Digital Marketing Strategy Through Search Engine Optimization (SEO) In Increasing Sales Of Msme Products In Indonesia. *EKOMBIS REVIEW: Jurnal Ilmiah Ekonomi Dan Bisnis*, 12(4), 4251–4260.
25. Tatikonda, R., Ponnala, J., Thatikonda, R., Yendluri, D. K., Kempanna, M., & Ananthan, B. (2024). Optimizing digital marketing strategies Through search engine optimization. *2024 IEEE International Conference on Contemporary Computing and Communications (InC4)*, 1, 1–6.
26. Tripathi, A. (2025). A Strategic Outlook on LLM SEO: Using File-Format Logic to Guide AI-Optimized Content Design. Available at SSRN 5375528.