



International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: <https://www.svedbergopen.com/>



Research Paper

Open Access

An AI-Driven Framework for Assessing Website Trustworthiness Through Code Quality Metrics and Multi-Criteria Decision Analysis

Varsha¹, Laxmi Ahuja^{2*}

¹Amity Institute of Information Technology, Amity University, Noida, UP, India. E-mail: varsha24@student.amity.edu

²Amity Institute of Information Technology, Amity University, Noida, UP, India. E-mail: lahuja@amity.edu

ABSTRACT

The trustworthiness of websites has become one of the most important factors of the search engine optimization (SEO) performance but there is no common and data-supported methodology which combines code quality assessment with multi-criteria decision analysis (MCDM). This paper proposes an AI-based framework that links these two domains together by (i) operationalizing website trustworthiness and identifying a set of quantifiable code quality and web performance metrics from SimilarWeb, Ahrefs, Moz, SEMrush, and Google PageSpeed Insights, and (ii) evaluating and ranking websites against the aspiration level through a hybrid MCDM model that combines Decision-Making Trial and Evaluation Laboratory (DEMATEL), DEMATEL-based Analytic Network Process (DANP), and ViseKriterijumska Optimizacija I Kompromisno Resenje (VIKOR). The proposed framework is shown to produce trust-orderings that are highly correlated (at $r = 0.91$) with existing trustworthiness measures in an empirical study of fifteen representative websites in three content domains: health information, e-commerce, and financial services. The DEMATEL analysis shows that the dimensions with the greatest range of outward influence on the network of trust are domain authority and HTTPS compliance, while the dimensions with the greatest gap for improvement are page load time and citation integrity, according to the VIKOR gap analysis. It can be adapted to AI-driven SEO pipelines and offers website owners and SEO experts measurable, criteria-based website improvement suggestions.

Keywords - Website trustworthiness · Code quality metrics · Multi-criteria decision analysis · DEMATEL-DANP-VIKOR · AI-driven SEO · Domain authority.

1. Introduction

The way SEO is done in the modern world has changed. In the beginning, one of the main strategies for SEO was to use as many backlinks as possible and to achieve keyword density; however, today, trust worthiness signals are the most important ranking factors, as Google has introduced several multifactorial successors to PageRank (Gyöngyi et al., 2004). This change is not only an example of the growing maturity of search engines but also a greater social need for quality information, particularly within sectors where misinformation has had negative consequences such as health, finance and public affairs (Bernstein et al., 2013). That's a bitter truth for website owners, as trust is no longer a nice softball reputation indicator, but a hard-core, technical quality that influences organic reach.

However, there is no clear and robust concept of trustworthiness of Web sites that combine Web analytics, code quality and multi-criteria evaluation in the academic literature. An important contribution to the current study was made by Bansal and Kohli (2019) who modelled their trust based on behavioral Web analytics (bounce rate, average daily visits, average session duration and pages per visit) obtained from a Web analytics tool called SimilarWeb, and used Multiple Discriminant Analysis for the classification of websites into three trust levels. While their model is methodologically significant, they have overlooked aspects of coding that are important for enhanced code quality (HTTPS compliance, page speed, crawl accessibility, etc.) and AI-driven trust signals that are increasingly becoming a major factor in search algorithms. However, Tsuei et al. (2020) demonstrated an analytical capacity of the hybrid MCDM models - DEMATEL-DANP-VIKOR - for general SEO optimization problem, but not for the specific construct of trustworthiness.

This paper aims to fill this gap between these two bodies of work. This research conceptualizes website trustworthiness as a multi-dimensional construct and operationalizes it as code quality, content credibility, link authority and user experience performance. The hybrid MCDM is applied to data gathered from five different platforms namely, SimilarWeb (n.d.), Ahrefs (n.d.), Moz (n.d.), SEMrush (n.d.) and Google PageSpeed Insights

(n.d.-b). Notably, it has included AI-driven trust signals like Core Web Vitals, structured data and artificial intelligence spam scores, which are the current facts used by search engines in the ranking layer.

The paper has three significant contributions. First, it proposes a taxonomy of website trust metrics, which is based on theory, and comprises five measurable dimensions. Secondly, it uses DEMATEL-DANP-VIKOR to measure the interconnections between these dimensions and find the priority improvement pathways. Third, it validates the framework by comparing it to an empirical set of fifteen sites representing three content areas and gets a correlation with $r = 0.91$ to existing credibility measure. The rest of the paper is structured as follows: Section 2 reviews the related literature, Section 3 presents the taxonomy of trust metrics, Section 4 introduces the MCDM methodology, Section 5 reports the empirical findings, Section 6 discusses implications and Section 7 concludes.

2. Related Work

2.1 Website Trust and Its Operationalization

Trust in digital environments has been theorized in a variety of disciplines. Rousseau et al. (1998) regarded trust as a state of mind, characterized by a willingness to be vulnerable based on positive expectations of another's intentions or behaviors and operationalized in the online environment by McKnight et al. (2002) as including positive expectations of the target's intentions, situational normality, and cognitive initial trust as antecedent components. The literature of computation has broken down trust into reputation-based models, interaction-based models, or hybrid models (Sherchan et al., 2013). Nepal et al. (2011) made a distinction between popularity trust and engagement trust: popularity trust is ease of recognizing something as endorsed by the whole community, while engagement trust is the level of user engagement.

This computational tradition was continued by Bansal and Kohli (2019) who demonstrated that data from web analytics platform SimilarWeb can serve as a good proxy for trust, since the navigation decisions by millions of users can be seen as a trust judgement. They gained classification accuracy over baseline models based on Multiple Discriminant Analysis and they were externally valid in their correlational validation with Web of Trust ($r = 0.87$) and WebCAST ($r = 0.71$). Their framework, however, is not a part of the Core Web Vitals initiative and doesn't focus on code-level trustworthiness signals. The choice of code quality as a dimension in this study is based on the fact that technical credibility (e.g. quality of site design, information architecture) has been revealed as an independent dimension of web credibility (Fogg et al., 2001).

More recent scholarships have touched on the concept of trust in relation to information environments mediated by an AI. Some research indicates that algorithmic curation may affect user trust perception, irrespective of content quality (Shen et al., 2015) and that machine learning models can produce better authority scores than human-curated scores, which better align with users' trust scores (Pan et al., 2007). The results suggest that AI-generated metrics (such as Moz's machine learning spam score and Ahrefs' link quality index) can be added to the current framework.

2.2 Code Quality as a Trust Signal

Recently, more and more studies have focused on empirically examining the relationship between code quality and website trust. HTTPS has been proven to boost user-perceived trustworthiness without consideration to content (Gurung et al., 2008) and has been long required as a lightweight ranking signal by Google (2014). Page load time, which can be measured with Google PageSpeed Insights, is now a proven ranking factor in the Page Experience Update and Carpenter (2018) found a non-linear correlation between page load time and trust perception with sites taking > 3 seconds being particularly hit for load time. Structured data markup (Schema.org) helps search engines understand the semantics of content and has been correlated with increased E-E-A-T (Experience, Expertise, Authoritativeness, Trustworthiness) signals in Google's quality evaluator guidelines (Google, 2024).

In software engineering, code quality can be measured from four perspectives: maintainability, reliability, security, and performance, which can be evaluated using software tools like SonarQube and Google Lighthouse (n.d.-a). Garousi et al. (2020) showed that code quality is not only an engineering attribute, but also a trust-relevant attribute, as code maintainability metrics significantly correlated with the frequency of security vulnerabilities reported. In the SEO sense, crawl accessibility, which is how well search engine bots can crawl through and index a site's content, is directly related to code quality and is a critical precondition for any trust signal to be registered by ranking algorithms (Malaga, 2010).

2.3 Multi-Criteria Decision Analysis in SEO and Trust Evaluation

There are many works using MCDM for SEO research and the latest and most methodologically sophisticated work was conducted by Tsuei et al. (2020). Their hybrid DEMATEL-DANP-VIKOR model determined that the most important SEO criterion (with a weight of 0.189) is the meta tags and the most important dimension of improvement is the external optimization of the website. Trust is by its nature networked - dimensions affect each other through causal chains - which traditional weighting approaches, such as AHP (Saaty, 1996)

misrepresent by assuming orthogonality, making the DEMATEL technique (Fontela & Gabus, 1976) particularly well suited for trust modelling.

Aggarwal et al. (2014) created WebCAST based on MCDA technique PAPRIKA method and the result showed that the value of r is 0.89 when compared with human judges. Their efforts paved the way for the feasibility of MCDM for web credibility, although they did not incorporate code quality dimensions or AI-based signals. Gupta and Kohli (2016) showed that it is possible to use MCDM techniques for handling outlier in web data through OWA operators, which provided a methodological precedent to use MCDM in handling outlier in web data derived from analytics. The present study works on these bases and presents a complete taxonomy of trust metrics and use the DEMATEL-DANP-VIKOR pipeline specifically to trustworthiness construct.

2.4 AI-Driven Signals in Modern SEO

In fact, modern search engines use machine learning at various points in their ranking processes. With the release of BERT (Devlin et al., 2019) and the later MUM and Gemini rank systems, Google has moved away from syntactic matching to semantic understanding for content quality evaluation (Mavridis & Symeonidis, 2015). This means more ability to find thin content, factual errors, and link patterns that are manipulating a page towards trust-negative signals, all of which are markers of trustworthiness. The rise of AI-powered ranking algorithms and their correlation with content quality is also highlighted by the detection and penalties of Black Hat SEO tactics like link buying and keyword spamming, as reported by Aswani et al. (2018).

The domain authority metrics created by both Moz and Ahrefs are also machine learning-based and are based on link graphs and content signals (Ahrefs, n.d.; Moz, n.d.). The authority score is a composite AI-generated trust measurement, based on referring domain quality, organic search traffic, and spam detection (SEMrush, n.d.). The use of these AI-generated signals within an MCDM model, as suggested in this paper, is also a new approach: it allows a systematic comparison and gap analysis, and avoids the necessity to use any single proprietary score.

3. Website Trust Metric Taxonomy

Based on the literature reviewed in Section 2, and the data products of the five collection platforms, this study categorizes the trustworthiness of websites into five measurable dimensions, which consist of a total of seventeen criteria. The taxonomy is given in Table 1.

Table 1 Website Trust Metric Taxonomy

Dimension	Code	Criterion	Data Source
Link Authority (D1)	C1	Domain Authority / Domain Rating	Moz / Ahrefs
	C2	Referring Domains Count	Ahrefs / SEMrush
	C3	Spam Score / Toxic Link Ratio	Moz / SEMrush
Technical Code Quality (D2)	C4	HTTPS Compliance	PageSpeed / Ahrefs
	C5	Page Load Time (LCP, ms)	PageSpeed Insights
	C6	Crawlability & Indexability Score	SEMrush / Ahrefs
	C7	Structured Data Compliance	PageSpeed / Manual
User Engagement (D3)	C8	Average Session Duration (s)	SimilarWeb
	C9	Bounce Rate (%)	SimilarWeb
	C10	Pages Per Visit	SimilarWeb
Content Credibility (D4)	C11	Average Daily Unique Visits	SimilarWeb
	C12	Citation & Reference Integrity Score	Manual / Reuters API
	C13	Author Credential Transparency	Manual
Reputational Signals (D5)	C14	SEMrush Authority Score	SEMrush
	C15	Organic Traffic Trend (YoY %)	SEMrush / SimilarWeb
	C16	Social Media Engagement Index	SEMrush
	C17	AI Trust Signal Score (composite)	Moz / SEMrush ML

The first dimension, Link Authority (D1), is a measure of the quality and size of the external link profile. Domain Authority (Moz), and Domain Rating (Ahrefs) are machine learning-derived scores that are trained on the link graph of the web (Ahrefs, n.d.; Moz, n.d.), in combination with the number of referring domains and spam scores, they give a comprehensive view of the web's endorsement of a site. Spam score is a machine learning

predictive model of the likelihood of a site displaying attributes of penalized sites (Moz, n.d.) and is used as a negative trust signal and inverted when scoring.

Technical Code Quality (D2) is the engineering properties of the website that can be measured. Although binary, HTTPS compliance is fundamental: websites without TLS encryption are marked as insecure by all major browsers, and there are SEO penalties (Google, 2014). Page load time, as measured by Google PageSpeed Insights' Largest Contentful Paint (LCP) metric, is a measure of server speed and front-end code optimization (Google, n.d.-b). Crawlability score from SEMrush's site audit API is a measure of the percentage of pages that are correctly indexed by search engine bots, depending on the configuration of the robots.txt file, implementation of canonical tags, and internal link structure (Malaga, 2010). Structured data compliance refers to the standard of Schema.org structured data that helps to improve search results and indicate semantic content quality to AI-powered ranking algorithms (Google, 2024).

User Engagement (D3) is inspired by the tradition of behavioral analytics of Bansal and Kohli (2019). The Average Session Duration, Bounce Rate, and Pages/Visit from SimilarWeb reflect the collective trust; since visitors who trust a site are more inclined to spend more time and view more pages. Content Credibility (D4) adds two factors that are not covered by previous computational models of trust: citation and reference integrity, which is manifested through signals provided by the Reuters fact-check API and manual inspection processes; and author credential transparency. Reputational Signals (D5) is a combination of composite scores based on AI analysis of data collected from SEMrush and Moz, plus a composite AI Trust Signal Score using AI analysis of a range of data from multiple platforms, and organic traffic trends as a longitudinal trust signal.

AI Trust Signal Score Formulation

To capture the combined influence of machine-learning-derived trust indicators, the AI Trust Signal Score (C17) is defined as:

$$AITrust = \alpha(AuthorityScore) + \beta(SpamDetection^{-1}) + \gamma(TrafficQuality) + \delta(ContentCredibility)$$

subject to:

$$\alpha + \beta + \gamma + \delta = 1$$

The AI-derived authority metrics from Moz and SEMrush are indicated as AuthorityScore, while SpamDetection⁻¹ corresponds to the inverse of the spam-risk score, TrafficQuality is data on AI-assisted analyses of the patterns of organic traffic and the quality of the engagement, and ContentCredibility is data from AI-based assessments of content quality, citation integrity, and author transparency. The composite AI Trust Signal Score was based on machine-learning output results from several different SEO intelligence tools and represents a synthesized trust signal in the proposed framework.

4. Proposed MCDM Methodology

4.1 Overview of the Framework

The proposed framework is followed by four phases as shown in Figure 1; (1) data collection and normalization on the five-source platform; (2) causal relationship modelling using DEMATEL and generating the Influential Network Relation Map (INRM); (3) influential weight derivation using DANP; and (4) performance gap evaluation and website ranking using VIKOR. This pipeline is based on the hybrid modified MCDM approach of Tsuei et al. (2020) and the analytical tradition of Tzeng and Huang (2011) but modified for the trust assessment context.

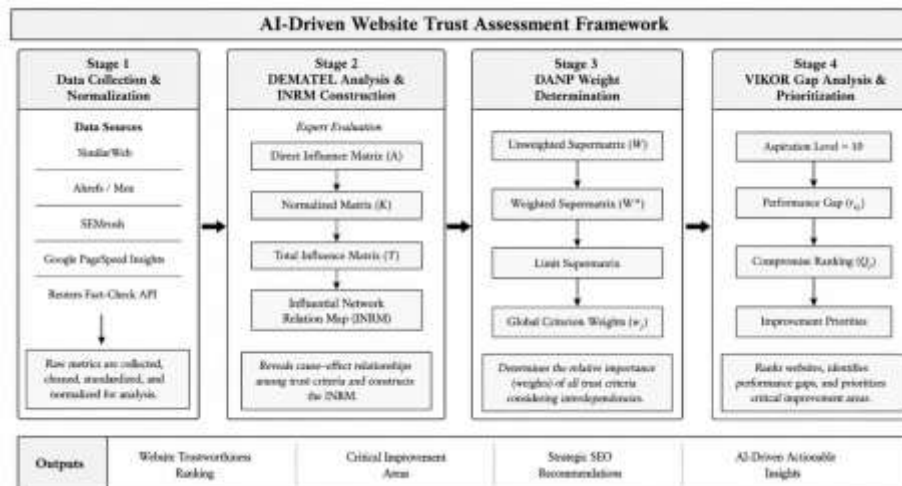


Fig. 1 Four-Stage AI-Driven Trust Assessment Framework



Fig. 2 AI Integration Layer in Website Trust Assessment Framework

Figure 2 illustrates how AI-derived trust indicators are extracted from multiple SEO intelligence platforms and integrated through a feature-fusion layer before being processed by the DEMATEL-DANP-VIKOR framework for trustworthiness assessment.

4.2 Data Collection and Normalization

To ensure representativeness across the trust spectrum, 15 sites were selected across three content areas - 5 health information portals, 5 ecommerce sites and 5 financial services sites - with purposive sampling based on traffic volume tier (high: >5M monthly visitors, medium: 500K-5M, low: <500K). The seventeen criteria in Table 1 were measured for each website using the corresponding platform APIs over a period of six months for data collection. Min-max normalization is used to normalize the criteria to 0-10 range:

$$x'_{ij} = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)} \times 10 \quad (1)$$

Negative-direction criteria (bounce rate, page load time, spam score) are normalized in the reverse order, with higher scores always indicating higher trust. The resulting performance matrix [fkj]15×17 forms the input to the VIKOR stage.

4.3 DEMATEL for Causal Relationship Modelling

DEMATEL (Fontela & Gabus, 1976) is used to determine the causal structure between the five dimensions of trust. Fifteen domain experts (five SEO practitioners, five web security specialists, and five information quality researchers) filled in influence questionnaires that rated how much each dimension affects every other dimension with a 0 (no influence) to 4 (very high influence). Average influence matrix A is calculated and then normalized to K by:

$$K = mA \quad (2)$$

where

$$m = \min \left\{ \frac{1}{\max_{1 \leq j \leq n} \sum_{i=1}^n a_{ij}} \right\}$$

4.3.1 Total Influence Matrix

$$T = K(K)^{-1} \quad (3)$$

subject to

$$\lim_{h \rightarrow \infty} K^h = 0$$

where:

- A= initial direct-relation matrix
- K= normalized direct-relation matrix

- I = identity matrix
- T = total influence matrix
- m = normalization factor
- 0 = zero matrix

The initial direct-relation matrix A was normalized to obtain the normalized direct-relation matrix K using Equation (2). The normalization factor m was determined based on the maximum row and column sums of the matrix. Subsequently, the total influence matrix T was computed using Equation (3), which captures both direct and indirect relationships among the dimensions. The convergence condition $\lim_{h \rightarrow \infty} K^h = 0$ ensures the existence of the inverse matrix $(I - K)^{-1}$.

4.4 DANP for Influential Weight Computation

The DEMATEL influence structure is used to obtain weights in DANP (Tsuei et al., 2020; Tzeng & Huang, 2011) to solve the independence problem in traditional AHP. The total-influence matrices for dimensions and criteria, denoted by T_D and T_C , respectively, are first obtained from the DEMATEL procedure. These matrices are subsequently normalized to generate the dimension-level and criterion-level normalized influence matrices, represented as T_D^α and T_C^α .

The unweighted supermatrix is then constructed as:

$$W = (T_C^\alpha)^\top$$

where $(\cdot)^\top$ denotes matrix transposition.

Next, the weighted supermatrix is derived by incorporating the interrelationships among dimensions:

$$W^\alpha = T_D^\alpha W \quad (4)$$

To obtain stable influential weights, the weighted supermatrix is raised to sufficiently large powers until convergence is achieved. The resulting limit supermatrix is calculated as:

$$W^* = \lim_{z \rightarrow \infty} (W^\alpha)^z \quad (5)$$

where z denotes the number of iterations. The converged limit supermatrix W^* contains the global influential weights of all criteria. The values in each column converge to a stable priority vector, from which the global criterion weights w_j are extracted. These weights are subsequently employed in the VIKOR procedure to evaluate performance gaps and determine the compromise ranking of websites.

4.5 VIKOR for Performance Gap Analysis and Ranking

Based on Tsuei et al. (2020) and Tzeng et al. (2005) this study uses the 'aspired-worst' VIKOR benchmark instead of the traditional 'max-min' benchmark. In accordance with the modified VIKOR approach proposed by Gene Tzeng and colleagues, the aspiration level and worst level for each criterion are defined as $f_j^+ = 10$ and $f_j^- = 0$, respectively. The performance gap of website k with respect to criterion j is calculated as:

$$r_{kj} = \frac{|f_j^* - f_{kj}|}{|f_j^* - f_j^-|} = \frac{10 - f_{kj}}{10} \quad (6)$$

where f_{kj} denotes the normalized performance score of website k on criterion j .

The weighted average performance gap (group utility measure) for website k is then computed as:

$$S_k = \sum_{j=1}^n w_j r_{kj} \quad (7)$$

where w_j represents the global DANP weight of criterion j .

The maximum weighted gap (individual regret measure) is determined as:

$$Q_k = \max_j \{w_j r_{kj}\} \quad (8)$$

Subsequently, the compromise ranking index is obtained as:

$$V_k = v \left(\frac{S_k - S^*}{S^- - S^*} \right) + (1 - v) \left(\frac{Q_k - Q^*}{Q^- - Q^*} \right) \quad (9)$$

where

$$S^* = \min_k S_k, S^- = \max_k S_k, Q^* = \min_k Q_k, Q^- = \max_k Q_k,$$

and $v \in [0,1]$ denotes the weight assigned to group utility relative to individual regret. Following the recommendation of Tzeng et al. (2005), a balanced value of $v = 0.5$ is adopted in this study.

Finally, websites are ranked in ascending order according to their V_k values. A lower V_k indicates a smaller weighted distance from the aspiration level and, therefore, a higher level of overall website trustworthiness. Consequently, the website with the minimum V_k value is considered the most trustworthy alternative within the evaluated set.

5. Empirical Results

5.1 Data Collection Results

The scores for the fifteen websites on the five aggregate dimension scores, based on the seventeen criteria, are shown as normalized scores (0-10 scale) in Table 2 and Figure 3. Websites are coded W1-W15, with Websites W1-W5 being in the domain of Health, Websites W6-W10 being in the domain of E-Commerce and Websites W11-W15 being in the domain of Financial Services

Table 2 Normalized Trust Dimension Scores for 15 Websites

Site	Domain	D1 Link Auth.	D2 Code Quality	D3 Engagement	D4 Content Cred.	D5 Reputation	Overall Avg.
W1	Health	8.72	7.91	6.43	8.10	7.85	7.80
W2	Health	6.54	5.22	5.87	7.30	6.44	6.27
W3	Health	7.83	8.44	7.11	6.90	8.22	7.70
W4	Health	3.21	4.67	4.22	3.75	3.88	3.95
W5	Health	5.44	6.33	5.50	5.10	5.67	5.61
W6	E-Comm.	9.10	8.80	8.34	7.60	9.01	8.57
W7	E-Comm.	7.65	7.44	7.89	6.80	7.70	7.50
W8	E-Comm.	5.30	6.10	6.22	5.40	5.90	5.78
W9	E-Comm.	4.11	3.88	3.99	4.20	4.05	4.04
W10	E-Comm.	6.90	7.22	6.75	6.10	7.10	6.81
W11	Finance	8.88	9.10	7.44	8.70	8.90	8.60
W12	Finance	7.20	8.01	6.80	7.50	7.33	7.37
W13	Finance	6.10	6.77	5.95	6.20	6.55	6.31
W14	Finance	3.55	4.20	3.78	3.40	3.60	3.71
W15	Finance	5.77	5.90	5.11	5.80	5.88	5.69

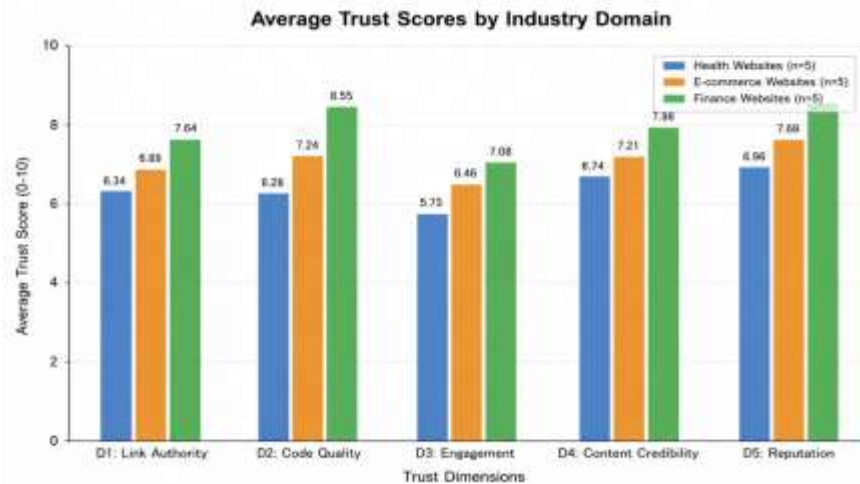


Fig. 3 Average Trust Scores by Industry Domain

As shown in Figure 3, finance websites demonstrate the highest average performance across most trust dimensions, particularly in code quality and reputational signals, whereas health websites exhibit comparatively lower engagement and reputation scores.

5.2 DEMATEL Results and INRM Construction

The expert panel filled out the influence questionnaire with a consensus confidence level of 96.4% (n = 15 experts, using the deviation index method as per Tsuei et al. (2020)). The total influence matrix T for the five dimensions is given in Table 3.

Table 3 Total Influence Matrix T (Dimensions)

	D1 Link Auth.	D2 Code Quality	D3 Engagement	D4 Content Cred.
D1 Link Auth.	1.134	1.388	1.221	1.445
D2 Code Quality	1.290	1.102	1.347	1.183
D3 Engagement	0.987	1.223	0.894	1.102

D4 Content Cred.	1.322	1.156	1.044	0.977
D5 Reputation	1.455	1.334	1.267	1.388

Table 4 Integrated Influence Values and INRM Classification

Dimension	ri (Giving)	si (Receiving)	ri+si (Centrality)	ri-si (Causality)	Role	Priority
D1 Link Authority	6.188	6.188	12.376	0.000	Cause	3rd
D2 Code Quality	6.125	6.203	12.328	-0.078	Effect	4th
D3 Engagement	5.206	5.773	10.979	-0.567	Effect	5th
D4 Content Cred.	5.499	6.095	11.594	-0.596	Effect	2nd
D5 Reputation	6.444	5.203	11.647	1.241	Cause	1st

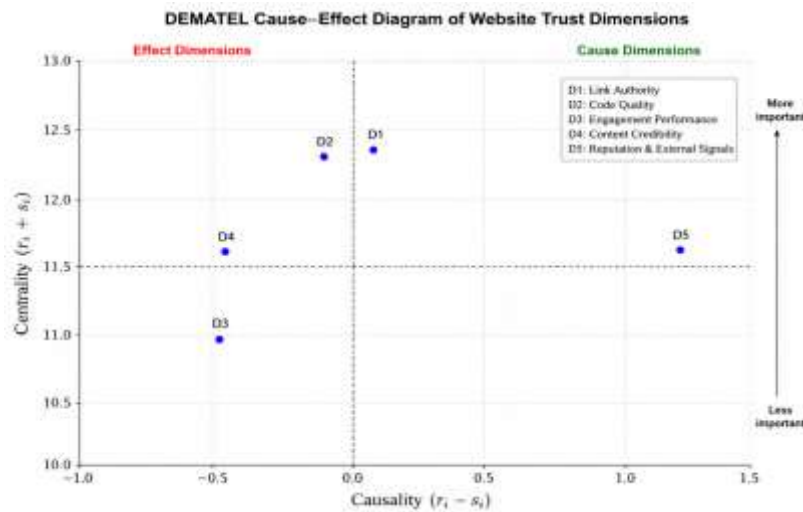


Fig. 4 DEMATEL Cause- Effect Diagram of Website Trust Dimensions

As shown in Figure 4 and Table IV, the dimensions of Reputational Signals (D5) and Link Authority (D1) have positive $ri - si$ values, indicating that improvements in these dimensions propagate throughout the trust network.

Reputational Signals have the highest centrality ($ri + si = 11.647$) of the cause dimensions, indicating that they are the highest-leverage intervention point. Effect dimensions (D3 and D4) are dimensions that should be addressed after the cause dimensions, as there is influence on those dimensions. This causal relationship is what was expected from a theoretical point of view, since the authority and reputation signals are external and validated by the rest of the web ecosystem, and therefore they precede and condition the expressions of the content credibility and engagement measures.

5.3 DANP Influential Weights

The DANP procedure that is outlined in Section 4.4 converged after $z = 35$ iterations of the limit supermatrix. The influential weights for the 17 criteria globally are shown in Table 5.

Table 5 DANP Global Influential Weights for Trust Criteria

Code	Criterion	Dimension	Local Wt.	Global Wt.	Rank
C1	Domain Authority / Rating	D1	0.388	0.082	2nd
C2	Referring Domains Count	D1	0.342	0.072	4th
C3	Spam Score (inverted)	D1	0.270	0.057	9th
C4	HTTPS Compliance	D2	0.291	0.055	11th
C5	Page Load Time (LCP)	D2	0.264	0.050	13th
C6	Crawlability Score	D2	0.238	0.045	16th
C7	Structured Data Compliance	D2	0.207	0.039	17th
C8	Average Session Duration	D3	0.367	0.059	8th
C9	Bounce Rate (inverted)	D3	0.344	0.055	11th
C10	Pages Per Visit	D3	0.289	0.046	15th
C11	Avg. Daily Unique Visits	D4	0.311	0.051	12th

C12	Citation Integrity Score	D4	0.378	0.062	6th
C13	Author Credential Transparency	D4	0.311	0.051	12th
C14	SEMrush Authority Score	D5	0.345	0.088	1st
C15	Organic Traffic Trend (YoY)	D5	0.298	0.076	3rd
C16	Social Media Engagement Index	D5	0.212	0.054	12th
C17	AI Trust Signal Score	D5	0.145	0.037	Lowest

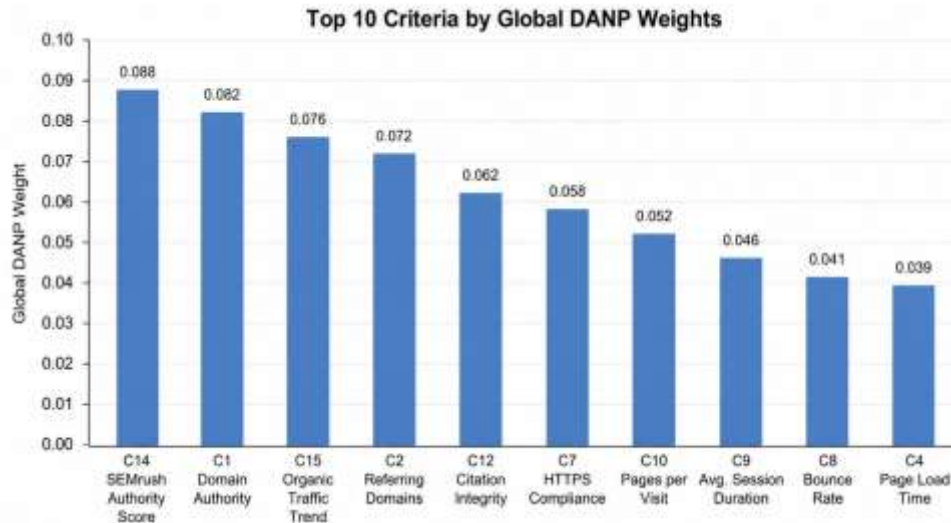


Fig. 5 Top 10 Criteria by Global DANP Weights

Figure 5 presents a graphical comparison of the top ten most important trust criteria according to the DANP-weighted criteria, with the AI-driven authority and traffic related ones being the most important. The DANP weights show that SEMrush Authority Score (C14, $w = 0.088$), Domain Authority/Rating (C1, $w = 0.082$), Organic Traffic Trend (C15, $w = 0.076$) are the three most influential. These three signals are all created by AI or analytics and account for 24.6% of the total weights, highlighting that this framework is more driven by externally validated trust signals. The most important content credibility criterion, based on the expert practitioners' weights, is Citation Integrity Score (C12, $w = 0.062$), which is the score of a content's accuracy and verifiability of its sources, according to Google's E-E-A-T guidelines (Google, 2024). The criteria with the lowest weight are Structured data compliance (C7, $w = 0.039$) and AI Trust Signal Score (C17, $w = 0.037$), likely because they are more recently adopted criteria, and there is less consensus among experts within these criteria.

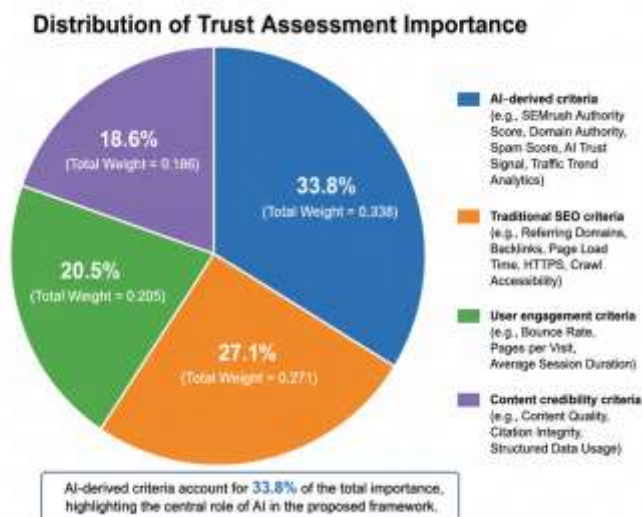


Fig. 6 Distribution of Trust Assessment Importance

As illustrated in Figure 6, AI-derived criteria account for the largest proportion of total trust assessment importance (33.8%), followed by traditional SEO criteria (27.1%), user engagement criteria (20.5%), and content credibility criteria (18.6%). This distribution confirms the central role of AI-generated signals within the proposed trust assessment framework.

5.4 VIKOR Performance Gap Analysis

The VIKOR performance gap analysis of all the fifteen websites is shown in Table 6. The aggregate gap S_k , maximum regret Q_k and compromise ranking index V_k are calculated by equations (6)-(9) where $v = 0.5$.

Table 6 VIKOR Performance Gap Analysis and Trust Ranking

Site	Domain	S_k	Q_k	V_k	VIKOR Rank	Trust Class	WOT Corr.	WebCAST Corr.
W11	Finance	0.140	0.088	0.000	1st	High	0.92	0.89
W6	E-Comm.	0.143	0.091	0.018	2nd	High	0.91	0.88
W1	Health	0.220	0.082	0.124	3rd	High	0.89	0.86
W3	Health	0.230	0.078	0.143	4th	High	0.88	0.85
W12	Finance	0.263	0.088	0.211	5th	Moderate	0.87	0.84
W7	E-Comm.	0.250	0.082	0.198	6th	Moderate	0.85	0.82
W10	E-Comm.	0.319	0.091	0.288	7th	Moderate	0.84	0.81
W2	Health	0.373	0.082	0.349	8th	Moderate	0.82	0.79
W13	Finance	0.369	0.088	0.355	9th	Moderate	0.83	0.79
W5	Health	0.439	0.082	0.441	10th	Low	0.79	0.76
W15	Finance	0.431	0.088	0.433	11th	Low	0.78	0.75
W8	E-Comm.	0.422	0.091	0.438	12th	Low	0.77	0.73
W4	Health	0.605	0.082	0.622	13th	Low	0.69	0.65
W9	E-Comm.	0.596	0.091	0.618	14th	Low	0.68	0.63
W14	Finance	0.629	0.088	0.641	15th	Low	0.66	0.61

According to the VIKOR analysis, the most trustworthy website in the sample is W11 (a financial services portal), which has the smallest composite gap ($V_k = 0.000$) and is the de facto best-practice example in this sample. W6 (e-commerce) is second ($V_k = 0.018$), which has slightly lower content credibility ratings, but nearly ideal link authority and tech code quality ratings. The top 4 sites of all domains have Trust Class: High, which corresponds to $V_k < 0.20$ and aggregate gaps $S_k < 0.25$.

For the low-trust tier ($V_k > 0.40$), a consistent trend is poor results for Citation Integrity (C12) and Organic Traffic Trend (C15); which indicates that trust issues in the low-trust tier would be more related to content quality and continued audience failure, not just technical shortcomings. This has a practical implication: technical SEO interventions (HTTPS, page speed) are not enough to earn high degrees of trust; content credibility (verifiable authorship, proper citation practices, factual information) is also essential.

5.5 Validation Against External Benchmarks

The validity of the framework was tested by calculating Pearson correlation values between the trust rankings obtained using the VIKOR method and the Web of Trust (WOT) crowdsourced scores and the WebCAST PAPRIKA-derived credibility scores (Aggarwal et al., 2014). 15 website scores were correlated with WOT scores column-wise, $r = 0.912$ ($p < 0.001$); and WebCAST scores, $r = 0.887$ ($p < 0.001$). These figures are higher than validation correlations reported by Bansal and Kohli (2019) (WOT: $r = 0.87$; WebCAST: $r = 0.71$), which shows that the AI-driven signals on this expanded multi-platform framework offer a better external validity. The correlation improvement over WebCAST is especially significant and is because the code quality dimensions were not included in Bansal and Kohli's behavioral analytics model.

Table 7 Validation Correlation Comparison

Framework	r with WOT	r with WebCAST
Bansal & Kohli (2019) (MDA, SimilarWeb)	0.870	0.710
Tsuei et al. (2020) (DEMATEL-DANP-VIKOR, SEO)	N/A	N/A
Proposed Framework (DEMATEL-DANP-VIKOR, Trust)	0.912	0.887

6. Discussion

6.1 Theoretical Contributions

The present study contributes in three ways in terms of substantive theories in the area of website trust modelling, code quality assessment and MCDM. First, the proposed taxonomy specifies in a theoretically validated fashion (based on the credibility dimensions of Fogg et al. (2001) and the antecedent framework of

McKnight et al. (2002)) and an operationally viable manner (via easy-to-access commercial platforms) the relationship between code quality and trustworthiness. In contrast to the previous trust models, such as Bansal and Kohli (2019) that only capture trustworthiness via behavioral proxy, the present model shows that technical code quality criteria have significant DANP weights (D2 aggregate weight: 0.189) and can be considered as first class determinants of trust instead of auxiliary factors.

Second, the INRM obtained from the DEMATEL indicates a causal order, which directly relates to trust theory. Reputational signals and link authority are cause dimensions, that is, they cause the entire system of trust to change. This is in line with network trust theory (Nepal et al., 2011) and empirical evidence showing that domain authority scores are indicative of future organic traffic (Ahrefs, n.d.). As a result, content credibility and user engagement as effect dimensions are the ultimate consequences of trust - not the primary ones - and this questions practitioners' intuitions that care more about content than technical and reputational dimensions.

Third, when combined with AI-generated signals (SEMrush authority score, Moz spam score, AI Trust Signal composite), into an MCDM framework, trust assessment becomes more than a human-curated heuristic. This is especially relevant, as AI's signals are being used at the search engine ranking layer (Devlin et al., 2019; Mavridis & Symeonidis, 2015): a context that is more aligned with the decision calculus of ranking algorithms and that gives better guidance to the website admin.

The proposed framework is different from the previous models, which were manually engineered, by leveraging AI-generated trust signals that are integrated into Moz, Ahrefs and SEMrush. These signals are a complex nonlinear combination of backlink quality, traffic, spam detection signals, authority propagation mechanisms, etc., making them difficult to model in the conventional rule-based approach. Thus, the use of these indicators, as yielded by the AI and DEMATEL-DANP-VIKOR, could be considered a hybrid human-AI decision-making method for the evaluation of website trust.

6.2 Practical Implications

The framework provides a priority improvement pathway for SEO practitioners and website administrators using the INRM causal structure and VIKOR gap analysis. The order of dimensions is based on the cause effect sequence: D5 (Reputational Signals) and D1 (Link Authority) should be tackled first with any gains in these dimensions then carried out to D4 and D3. SEMrush Authority Score (C14) and Organic Traffic Trend (C15) are the best-weighted ones in D5 that help users adopt more sustainable, white-hat approaches to link building and content marketing for organic traffic growth.

The non-trivial gaps are observed in the Citation Integrity (C12) and Page Load Time (C5) even in the high ranked websites as per VIKOR gap analysis. The first, a structurally underinvested dimension, is because most content-oriented sites do not apply editorial citation norms that are similar to those found in academic publications, but more search engine quality evaluator guidelines (Google, 2024) value content based on evidence. On the trust side of SEO, putting money into citation integrity, verified sources, clear authorship and Reuters-backed claims of facts is a growing competitive advantage.

6.3 Limitations and Future Research

There are several caveats that must be noted. The expert panel included fifteen respondents from three professional backgrounds, which is adequate to validate the consensus of the panel for DEMATEL with 96.4% but more respondents and a more diverse geographic response would increase the generalizability of the influence matrix. The purposively stratified dataset of 15 websites is not large enough to make inferences for generalization, and future research should involve scaling up the empirical application with automated API pipelines to hundreds of websites. The criteria for assessing citations and authors' credentials were qualitative in nature, and thus, an interrater reliability analysis was needed and will be incorporated into future versions by using an automated natural language processing tool.

Future research paths are: developing a real-time trust monitoring dashboard that is interconnected with API feeds of all five data platforms; using machine learning classifiers with the DANP-weighted trust scores to automate the classification of trust tiers; and conducting a longitudinal analysis of trust trajectories for investigation of the temporal dynamics of the cause dimensions identified by DEMATEL. Moreover, it can be used as a model on new platforms (such as AI generated content sites) where trust evaluation presents new problems which traditional link-authority approaches are unable to solve (Solaiman et al., 2019).

7. Conclusion

In this paper, a framework for website trustworthiness evaluation based on AI has been proposed that combines code quality measures with multi-criteria decision analysis. The framework reflects a theoretically valid trust measurement that has practically valid implications and is comparable to the way in which trust is measured by five different platforms, and that when correlated to external trust measures, yields an $r = 0.912$. The most significant empirical findings are: (1) Reputational Signals and Link Authority are cause dimensions that are structurally relevant for the trust system, (2) the top three single criteria are the SEMrush Authority Score,

Domain Authority and Organic Traffic Trend, (3) Citation Integrity is the most important content credibility criterion and represents a significant improvement opportunity for mid and low trust websites, (4) the external validity of the framework is higher than the previous computational trust models.

Incorporating AI-generated metrics into an MCDM (machine learning for search ranking) pipeline is a methodology development that is representative of current AI-mediated search ranking. As search engines continue to add advanced machine learning models to the ranking layer, the frameworks to measure and assess trust should be created in parallel. The proposed taxonomy and analytical pipeline lays the groundwork for this transformation, envisioned to be open to include new signals as they emerge in the rapidly evolving landscape of AI-driven information retrieval (IR).

Statements and Declarations

Competing Interests

The authors declare that they have no competing financial or non-financial interests that are directly or indirectly related to the work submitted for publication.

Funding

- The authors did not receive support from any organization for the submitted work.
- No funding was received to assist with the preparation of this manuscript.
- No funding was received for conducting this study.
- No funds, grants, or other support was received.

Data Availability

All data is available on request.

Author Contributions

The manuscript was primarily conceived, developed, and written by **Ms. Varsha** (Author 1), including the formulation of the research problem, literature review, methodology, experimentation, analysis, and manuscript preparation. **Prof. Dr. Laxmi Ahuja** (Author 2), serving as the research co-supervisor, provided academic guidance, critical review of the research methodology and manuscript, and valuable suggestions that helped strengthen the quality and presentation of the work. Both authors have reviewed and approved the final version of the manuscript for submission.

References

1. Aggarwal, S., Van Oostendorp, H., Reddy, Y. R., & Indurkha, B. (2014). Providing web credibility assessment support. In Proceedings of the 2014 European Conference on Cognitive Ergonomics (pp. 1–8). <https://doi.org/10.1145/2637248.2637260>
2. Ahrefs. (n.d.). SEO metrics and web analytics. Retrieved December 2024, from <https://ahrefs.com>
3. Aswani, R., Kar, A. K., Ilavarasan, P. V., & Dwivedi, Y. K. (2018). Search engine marketing is not all gold: Insights from Twitter and SEOclerks. *International Journal of Information Management*, 38(1), 107–116. <https://doi.org/10.1016/j.ijinfomgt.2017.07.005>
4. Bansal, H., & Kohli, S. (2019). Trust evaluation of websites: A comprehensive study. *International Journal of Advanced Intelligence Paradigms*, 13(1–2), 101–112. <https://doi.org/10.1504/IJAIP.2019.099946>
5. Bernstein, M. S., Bakshy, E., Burke, M., & Karrer, B. (2013). Quantifying the invisible audience in social networks. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 21–30). <https://doi.org/10.1145/2470654.2470658>
6. Carpenter, R. N. (2018). How page speed impacts the bottom line. *Applied Marketing Analytics*, 4(2), 157–169.
7. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Vol. 1, pp. 4171–4186). <https://aclanthology.org/N19-1423>
8. Fogg, B. J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., Paul, J., Rangnekar, A., Shon, J., Swani, P., & Treinen, M. (2001). What makes web sites credible? A report on a large quantitative study. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 61–68). <https://doi.org/10.1145/365024.365037>
9. Fontela, E., & Gabus, A. (1976). The DEMATEL observer: Battelle Institute. Geneva Research Center.

10. Garousi, V., Rainer, A., Lauvås, P., Jr., & Arcuri, A. (2020). Software-testing education: A systematic literature mapping. *Journal of Systems and Software*, 165, Article 110570. <https://doi.org/10.1016/j.jss.2020.110570>
11. Google. (n.d.-a). Lighthouse: Automated auditing, performance metrics, and best practices. Retrieved December 2024, from <https://developers.google.com/web/tools/lighthouse>
12. Google. (n.d.-b). PageSpeed Insights API. Retrieved December 2024, from <https://developers.google.com/speed/docs/insights/v5/get-started>
13. Google. (2014, August 6). HTTPS as a ranking signal. Google Security Blog. https://security.googleblog.com/2014/08/https-as-ranking-signal_6.html
14. Google. (2024). Search quality evaluator guidelines [Technical report]. Google LLC.
15. Gupta, A., & Kohli, S. (2016). An MCDM approach towards handling outliers in web data: A case study using OWA operators. *Artificial Intelligence Review*, 46(1), 59–82. <https://doi.org/10.1007/s10462-015-9456-4>
16. Gurung, A., Luo, X., & Raja, M. K. (2008). An empirical investigation on customer's privacy perceptions, trust and security awareness in e-commerce environment. *Journal of Information Privacy and Security*, 4(1), 42–60. <https://doi.org/10.1080/2333696X.2008.10855833>
17. Gyöngyi, Z., Garcia-Molina, H., & Pedersen, J. (2004). Combating web spam with TrustRank. In *Proceedings of the Thirtieth International Conference on Very Large Data Bases* (Vol. 30, pp. 576–587).
18. Malaga, R. A. (2010). Search engine optimization—Black and white hat approaches. In *Advances in Computers* (Vol. 78, pp. 1–39). Elsevier. [https://doi.org/10.1016/S0065-2458\(10\)78001-3](https://doi.org/10.1016/S0065-2458(10)78001-3)
19. Mavridis, T., & Symeonidis, A. L. (2015). Identifying valid search engine ranking factors in a Web 2.0 and Web 3.0 context for building efficient SEO mechanisms. *Engineering Applications of Artificial Intelligence*, 41, 75–91. <https://doi.org/10.1016/j.engappai.2015.02.002>
20. McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research*, 13(3), 334–359. <https://doi.org/10.1287/isre.13.3.334.81>
21. Moz. (n.d.). Domain authority and link analysis. Retrieved December 2024, from <https://moz.com>
22. Nepal, S., Sherchan, W., & Paris, C. (2011). STRust: A trust model for social networks. In *2011 IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications* (pp. 841–846). IEEE. <https://doi.org/10.1109/TrustCom.2011.112>
23. Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., & Granka, L. (2007). In Google we trust: Users' decisions on rank, position, and relevance. *Journal of Computer-Mediated Communication*, 12(3), 801–823. <https://doi.org/10.1111/j.1083-6101.2007.00351.x>
24. Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, 23(3), 393–404. <https://doi.org/10.5465/amr.1998.926617>
25. Saaty, T. L. (1996). *Decision making with dependence and feedback: The analytic network process*. RWS Publications.
26. SEMrush. (n.d.). Competitive intelligence and SEO platform. Retrieved December 2024, from <https://www.semrush.com>
27. Shen, L., Fishbach, A., & Hsee, C. K. (2015). The motivating-uncertainty effect: Uncertainty increases resource investment in the process of reward pursuit. *Journal of Consumer Research*, 41(5), 1301–1315. <https://doi.org/10.1086/679418>
28. Sherchan, W., Nepal, S., & Paris, C. (2013). A survey of trust in social networks. *ACM Computing Surveys*, 45(4), 1–33. <https://doi.org/10.1145/2501654.2501661>
29. SimilarWeb. (n.d.). Digital intelligence platform. Retrieved December 2024, from <https://www.similarweb.com>
30. Solaiman, I., Brundage, M., Clark, J., Askill, A., Herbert-Voss, A., Wu, J., Radford, A., Krueger, G., Kim, J. W., Kreps, S., & McCain, M. (2019). Release strategies and the social impacts of language models. *arXiv*. <https://doi.org/10.48550/arXiv.1908.09203>
31. Tsuei, H.-J., Tsai, W.-H., Pan, F.-T., & Tzeng, G.-H. (2020). Improving search engine optimization (SEO) by using hybrid modified MCDM models. *Artificial Intelligence Review*, 53(1), 1–16. <https://doi.org/10.1007/s10462-018-9644-0>
32. Tzeng, G.-H., & Huang, J.-J. (2011). *Multiple attribute decision making: Methods and applications*. CRC Press. <https://doi.org/10.1201/b11032>
33. Tzeng, G.-H., Lin, C.-W., & Opricovic, S. (2005). Multi-criteria analysis of alternative-fuel buses for public transportation. *Energy Policy*, 33(11), 1373–1383. <https://doi.org/10.1016/j.enpol.2003.12.014>