
WORKING PAPER · WHITEPAPER

SEAKT™

**A Scientific Framework for Measuring AI Visibility
in Generative Engines**

*Ein wissenschaftliches Framework zur Messung
von AI-Sichtbarkeit in generativen Suchsystemen*

Marco Biner

Founder, geoquality.ai
Glis, Schweiz

Version 1.0 · Mai 2026

Hybrides wissenschaftliches Industrie-Whitepaper

License: **CC BY 4.0** (Methode) · **SEAKT™** (Wortmarke)

Korrespondenz: *contact@geoquality.ai*

Abstract

Mit dem Aufkommen generativer Suchsysteme — namentlich ChatGPT Search, Google AI Overviews / AI Mode, Perplexity, Microsoft Copilot, Anthropic Claude und xAI Grok — vollzieht sich seit 2023 ein struktureller Wandel der Informationsbeschaffung im Web. Anstelle der Rückgabe einer geordneten Trefferliste synthetisieren Large Language Models (LLMs) Antworten aus mehreren Quellen und zitieren ausgewählte Webseiten selektiv (Aggarwal et al., 2024; Lewis et al., 2020). Während etablierte Verfahren der Suchmaschinenoptimierung (SEO) nahezu vollständig auf das Paradigma der hyperlinkbasierten Ranking-Funktion (Brin & Page, 1998) ausgerichtet sind, fehlt eine theoretisch fundierte, reproduzierbare und transparent dokumentierte Methode zur Quantifizierung der **AI-Sichtbarkeit** einer Webseite. Diese Arbeit schlägt mit **SEAKT** (Strukturelle Daten · Entitäten · Autorität · Kontent-Qualität · Technische Zugänglichkeit) ein fünfdimensionales, hybrides Bewertungsframework vor, das regelbasierte Prüfungen mit LLM-basierter semantischer Beurteilung kombiniert. Die methodische Architektur folgt dem Design-Science-Paradigma der Wirtschaftsinformatik (Hevner et al., 2004) und integriert Erkenntnisse aus der Linked-Data-Forschung (Bizer et al., 2009), der Entity-Disambiguation (Hoffart et al., 2011), der Retrieval-Augmented Generation (Lewis et al., 2020) sowie der Theorie rigider Designatoren (Kripke, 1980). SEAKT operationalisiert AI-Sichtbarkeit als Score auf einer Skala von 0–100 Punkten, gegliedert in vier Bewertungsstufen (AI-Unsichtbar, Schwach, Solide, AI-Ready). Die empirische Validierung erfolgt anhand einer Fallstudie (HODL.swiss), die einen Pre-/Post-Optimization-Vergleich dokumentiert (73/100 → 89/100). **Beitrag der Arbeit:** (1) eine peer-fähige Operationalisierung des Konzepts der AI-Sichtbarkeit, (2) ein offener Standard unter CC BY 4.0 zur freien Verwendung in Forschung und Praxis, (3) eine reproduzierbare Bewertungsarchitektur für Multi-Page-Konsolidierung, (4) eine Forschungsagenda für Längsschnittstudien, Inter-Rater-Reliabilität und die Erweiterung auf agentenbasierte Systeme.

Schlüsselwörter: *Generative Engine Optimization, GEO, AI-Sichtbarkeit, Large Language Models, Schema.org, JSON-LD, Wissensgraph, Entity Linking, Retrieval-Augmented Generation, E-E-A-T, Design Science.*

Abstract (English)

The rise of generative search systems — notably ChatGPT Search, Google AI Overviews / AI Mode, Perplexity, Microsoft Copilot, Anthropic Claude and xAI Grok — has triggered a structural change in web-based information retrieval since 2023. Rather than returning ranked lists of hyperlinks, large language models (LLMs) synthesise answers from multiple sources and selectively cite chosen websites (Aggarwal et al., 2024; Lewis et al., 2020). While established Search Engine Optimization (SEO) practices remain firmly anchored in the hyperlink-based ranking paradigm (Brin & Page, 1998), no theoretically grounded, reproducible and transparently documented method currently exists to quantify the **AI visibility** of a website. This paper proposes **SEAKT** (Structured Data · Entities · Authority · Content Quality · Technical Accessibility), a five-dimensional hybrid evaluation framework that combines rule-based checks with LLM-based semantic judgment. The methodological architecture follows the design-science paradigm in information systems research (Hevner et al., 2004) and integrates insights from Linked Data research (Bizer et al., 2009), entity disambiguation (Hoffart et al., 2011), retrieval-augmented generation (Lewis et al., 2020), and the theory of rigid designators (Kripke, 1980). SEAKT operationalises AI visibility as a 0–100 composite score with four interpretive bands (AI-Invisible, Weak, Solid, AI-Ready). Empirical validation draws on a pre-/post-optimization case study (HODL.swiss) documenting an improvement from 73/100 to 89/100. **Contributions:** (1) a peer-reviewable operationalisation of AI visibility, (2) an open standard under CC BY 4.0 for free use in research and practice, (3) a reproducible evaluation architecture supporting multi-page consolidation, (4) a research agenda covering longitudinal studies, inter-rater reliability, and extension to agentic systems.

Keywords: *Generative Engine Optimization, GEO, AI visibility, Large Language Models, Schema.org, JSON-LD, Knowledge Graph, Entity Linking, Retrieval-Augmented Generation, E-E-A-T, Design Science.*

Inhaltsverzeichnis

Abstract (Deutsch / English).....	3
1 Einleitung und Problemstellung	6
1.1 Ausgangslage.....	6
1.2 Marktrelevanz.....	7
1.3 Forschungsfrage.....	8
2 Forschungslücke und Relevanz	9
2.1 Stand der Forschung.....	9
2.2 Lücke.....	9
2.3 Relevanz für Praxis und Forschung.....	10
3 Theoretische Grundlagen	11
3.1 Generative Engines und das Retrieve-then-Generate-Paradigma.....	11
3.2 Strukturierte Daten: Schema.org und JSON-LD.....	11
3.3 Wissensgraphen, Entity Linking und sameAs-Properties.....	12
3.4 E-E-A-T als Qualitäts-Heuristik.....	12
3.5 Der LLMs.txt-Vorschlag.....	13
3.6 Sprachphilosophische Grundierung: Rigide Designatoren.....	13
4 Das SEAKT-Framework	14
4.1 Konzeptioneller Aufbau.....	14
4.2 Dimension S — Strukturelle Daten (25 Punkte).....	15
4.3 Dimension E — Entity-Klarheit (20 Punkte).....	16
4.4 Dimension A — Autorität & Vertrauen (20 Punkte).....	16
4.5 Dimension K — Content-Qualität (20 Punkte).....	17
4.6 Dimension T — Technische Zugänglichkeit (15 Punkte).....	17
5 Methodische Architektur	18
5.1 Hybrides Bewertungsmodell.....	18
5.2 Multi-Page-Konsolidierung.....	19
5.3 Reproduzierbarkeit und Auditierbarkeit.....	19
6 Bewertungsstufen und Interpretation	20
7 Empirische Validierung — Fallstudie HODL.swiss	21
7.1 Untersuchungsobjekt.....	21
7.2 Methode.....	21
7.3 Ergebnisse.....	22
7.4 Interpretation.....	22
8 Diskussion und Limitationen	24

9 Forschungsausblick	27
10 Trademark and Licensing Notice.....	29
11 Literaturverzeichnis.....	31

Verzeichnis der Tabellen

Tabelle 1: Vergleich der Paradigmen — SEO und GEO	7
Tabelle 2: SEAKT-Dimensionen und Punktegewichte	15
Tabelle 3: Sub-Indikatoren der Dimension S (Strukturelle Daten)	15
Tabelle 4: Bewertungslogik der hybriden Architektur.....	18
Tabelle 5: Bewertungsstufen — Interpretationsschema	20
Tabelle 6: Pre-/Post-Optimization-Vergleich Fallstudie HODL.swiss	22

1. Einleitung und Problemstellung — Paradigmenwechsel von SEO zu GEO

1.1 Ausgangslage

Die Online-Informationsbeschaffung steht in einer Übergangsphase, deren Bedeutung mit dem Übergang von Verzeichnisportalen zu hyperlinkbasierten Suchmaschinen Ende der 1990er-Jahre vergleichbar ist. Der konzeptionelle Bruch ist klar dokumentiert: Brin und Page (1998) etablierten mit der Beschreibung des Google-Prototyps das Paradigma der hypertextuellen Suchmaschine, in der ein Crawler den Web-Korpus indiziert, ein Ranking-Algorithmus die Relevanz der gefundenen Dokumente bestimmt und der Nutzer eine geordnete Liste von Hyperlinks erhält, aus der er — gegebenenfalls nach Anklicken — seine Antwort selbst rekonstruiert.

Mit der Veröffentlichung der Transformer-Architektur (Vaswani et al., 2017), den Befunden zu impliziter Wissensspeicherung in Sprachmodellen (Petroni et al., 2019; Roberts et al., 2020), dem Durchbruch der Retrieval-Augmented Generation (Lewis et al., 2020) und dem instruktionsoptimierten Alignment via RLHF (Ouyang et al., 2022) wurde die technologische Grundlage für eine grundlegend andere Klasse von Systemen geschaffen: **generative Suchmaschinen** (Generative Engines, GEs). Diese Systeme — ChatGPT Search, Google AI Mode bzw. AI Overviews, Perplexity, Microsoft Copilot, Anthropic Claude, xAI Grok — synthetisieren eine direkte Antwort aus mehreren retrieveden Quellen und zitieren ausgewählte URLs selektiv anstatt eine vollständige, geordnete Trefferliste zurückzugeben (Aggarwal et al., 2024).

Aggarwal et al. (2024) formalisierten diesen Paradigmenwechsel und prägten den Begriff **Generative Engine Optimization (GEO)**. Sie zeigten in einem Black-Box-Optimierungsframework, dass GEO-spezifische Strategien — etwa autoritative Tonalität, Zitatreichtum, statistische Evidenz und sprachliche Flüssigkeit — die Sichtbarkeit in generativen Antworten um bis zu 40 % erhöhen können, während klassische SEO-Hebel (Keyword-Dichte, Backlink-Profil) deutlich an Wirkung verlieren.

„Suchen heisst nicht mehr nur Finden. Suchen heisst zunehmend, fertige Antworten zu erhalten.“

1.2 Marktrelevanz

Die ökonomische Tragweite dieses Paradigmenwechsels ist erheblich. Gartner (2024) prognostiziert für 2026 einen Rückgang des Volumens traditioneller Suchmaschinen um 25 % zugunsten von KI-Chatbots und virtuellen Assistenten.¹ Similarweb (2025) dokumentiert für ChatGPT ein Wachstum von 28,5 Mrd. (April 2024–März 2025) auf 47,7 Mrd. Visits, für Perplexity einen Anstieg von 388,5 Mio. auf 1,3 Mrd. (+243,7 % Year-over-Year). Die Konversionsraten von KI-Verweisen liegen substantiell über jenen klassischer Suche: Berichten zufolge erzielen ChatGPT-Referrals 15,9 %, Perplexity 10,5 %, Claude 5 %, Gemini 3 %, gegenüber 1,76 % organischem Google-Traffic (Position Digital, 2025).²

Für die Schweiz dokumentiert die Studie *KMU Digital Pulse 2025* von localsearch und der Hochschule Luzern (HSLU), dass 13 % der Schweizer Bevölkerung KI-Assistenten als ersten Suchkanal für KMU-Dienstleistungen nutzen — bei den 18- bis 29-Jährigen 21 %; 78 % der bisherigen KI-Nutzer planen eine Intensivierung dieser Nutzung (localsearch & HSLU, 2025).

Table 1: Vergleich der Paradigmen — SEO und GEO

Dimension	Klassisches SEO	Generative Engine Optimization (GEO)
Primäres Ziel	Hohe Position in den SERPs (ranking-orientiert)	Repräsentation und Zitation in KI-Antworten
Optimierungsobjekt	Keywords, Backlinks, Meta-Tags, Page-Speed	Entitäten, JSON-LD, Wissensgraph, Autorität
Nutzer-Interaktion	Klick auf Suchresultat	Direkter Antwort-Konsum
Erfolgskennzahlen	Ranking-Position, CTR, Impressions	Zitierfrequenz, Brand-Mention-Rate, Citation-Quality
Schlüssel-Infrastruktur	XML-Sitemap, robots.txt, Backlink-Profil	Schema.org / JSON-LD, sameAs, AI-Crawler-Freigabe
Content-Format	Long-Form, keyword-optimiert	Definitionen, FAQ, zitierfähige Aussagen
Reife des Felds	> 30 Jahre, etabliert	Emergent (seit 2023), Methodik in Entwicklung
Bewertungs-Modell	PageRank, Linkstruktur (Brin & Page, 1998)	Vektor-Repräsentationen, Entitäten, RAG-Pipelines

Eigene Darstellung auf Basis von Aggarwal et al. (2024), Brin & Page (1998), Lewis et al. (2020).

¹Anmerkung zur Quelle: Gartner-Prognosen sind explizit als Vorhersagen, nicht als gemessene Tatsachen, zu verstehen; methodische Vorbehalte sind in der Fachliteratur dokumentiert (Searchengineland, 2024).

²Diese Konversionsraten stammen aus aggregierten kommerziellen Datenquellen (Position Digital, 2025); sie sind methodisch nicht peer-reviewed und sollten als Indikatoren, nicht als kausale Aussagen, interpretiert werden.

1.3 Forschungsfrage

Aus dieser Lage ergibt sich die Leitfrage dieser Arbeit:

Wie kann die Sichtbarkeit einer Webseite in generativen Suchsystemen messbar, reproduzierbar, theoretisch fundiert und transparent dokumentiert quantifiziert werden, sodass das Resultat sowohl wissenschaftlich überprüfbar als auch in der Praxis handlungsleitend ist?

Diese Frage zerfällt in drei Teilfragen:

1. **Operationalisierung:** Welche Dimensionen sind konstitutiv für AI-Sichtbarkeit?
2. **Methodik:** Wie kann ein Bewertungsmodell die deterministische Strenge regelbasierter Prüfungen mit der semantischen Tiefe LLM-basierter Beurteilungen verbinden, ohne in opake Black-Box-Verfahren (Pasquale, 2015) abzugleiten?
3. **Validierung:** Welche empirische Evidenz lässt sich für die Konstruktvalidität, die diskriminante Trennschärfe und die Praxis-Sensitivität des Frameworks beibringen?

Beitrag dieser Arbeit ist die Entwicklung und Dokumentation des **SEAKT-Frameworks** als Antwort auf diese Fragen. SEAKT operationalisiert AI-Sichtbarkeit in fünf messbaren Dimensionen — Strukturelle Daten (S), Entity-Klarheit (E), Autorität & Vertrauen (A), Content-Qualität (K) und Technische Zugänglichkeit (T) — auf einer transparenten 100-Punkte-Skala mit hybrider Bewertungsarchitektur.

2. Forschungslücke und Relevanz

2.1 Stand der Forschung

Die Literatur zur Generativen Engine Optimization ist jung, wächst jedoch schnell. Aggarwal et al. (2024) gilt als foundational paper; nachfolgende Arbeiten erweitern den Ansatz auf E-Commerce-Szenarien (Bagga et al., 2025), Multi-Query-Optimierung (IF-GEO, 2026) sowie agentenbasierte selbstevolutionäre Frameworks (AgenticGEO, 2026) und Visual-Content (Generative Engine Optimization für Pinterest, 2026). Kommerzielle Plattformen wie Profound, Peec.ai, AthenaHQ, Otterly.AI, Scrunch AI und Superlines sind in den Markt eingetreten, allerdings primär als Monitoring- und Tracking-Werkzeuge, nicht als peer-reviewed wissenschaftliche Bewertungsinstrumente.

2.2 Lücke

Trotz dieser Aktivität fehlt es an einem **konsolidierten, theoretisch fundierten, offen lizenzierten Rahmenwerk** zur Quantifizierung der AI-Sichtbarkeit, das folgenden Kriterien gleichzeitig genügt:

4. **Theoretische Tiefe.** Verankerung in etablierten Forschungsdiskursen (Linked Data, Entity Linking, RAG, Sprachphilosophie der Referenz, Algorithmic Accountability).
5. **Methodische Transparenz.** Offenlegung der Bewertungslogik, Reproduzierbarkeit, Auditierbarkeit (Diakopoulos, 2016).
6. **Hybride Architektur.** Verbindung determinierbarer Regeln mit semantischer LLM-Beurteilung — ein Forschungsansatz, der in benachbarten Feldern (LLM-as-a-Judge, Bildungsevaluation) seit 2024 aktiv erprobt wird (ELMES, 2025).
7. **Multi-Page-Sensitivität.** Konsolidierungslogik, die der Realität moderner Webpräsenzen mit Subdomains, Sprachversionen, Hub-Spoke-Strukturen gerecht wird.
8. **Offener Standard.** Lizenzierung unter CC BY 4.0 (Creative Commons, 2013) ermöglicht freie wissenschaftliche und kommerzielle Verwendung bei Pflicht zur Attribution.

2.3 Relevanz für Praxis und Forschung

Die Schliessung dieser Lücke ist nicht nur akademisch motiviert. Inhaber und Betreiber von Webpräsenzen — vom DACH-KMU bis zum globalen Konzern — stehen vor der pragmatischen Frage, welche Massnahmen ihre AI-Sichtbarkeit tatsächlich verbessern. Ohne reproduzierbare Metrik bleibt das Feld anfällig für **Cargo-Kult-Optimierung**: Massnahmen werden adoptiert, weil sie zirkulieren (z. B. das pauschale Anlegen einer *llms.txt*-Datei), nicht weil ihre Wirkung empirisch belegt wäre. Der Autor argumentiert in Übereinstimmung mit empirischen Befunden (Spriestersbach, 2025), dass *llms.txt* als Inferenz-Hilfsstandard für Coding-Assistenten konzipiert war (Howard, 2024) und weder in den Retrieval-Pfaden grosser AI-Suchsysteme nachweislich systematisch konsumiert wird, noch dort nachweislich Sichtbarkeit erzeugt — wenngleich punktuelle Indizierung durch Google AI Mode dokumentiert wurde (dev5310, 2026).

Wissenschaftlich öffnet die Schliessung dieser Lücke ein neues Forschungsfeld an der Schnittstelle von Information Retrieval, NLP, Wirtschaftsinformatik und Marketing-Forschung. Die hier vorgestellte Operationalisierung schafft die methodische Voraussetzung für Längsschnittstudien, Inter-Rater-Reliabilität-Untersuchungen und kausale Tests der Wirksamkeit einzelner Optimierungshebel — Fragen, die ohne quantifizierbares Konstrukt nicht beantwortbar sind.

3. Theoretische Grundlagen

3.1 Generative Engines und das Retrieve-then-Generate-Paradigma

Die technische Architektur generativer Suchsysteme folgt im Kern dem von Lewis et al. (2020) eingeführten **Retrieval-Augmented Generation (RAG)**-Paradigma: Eine Anfrage wird in einen Vektor übersetzt; ein Retriever findet topisch relevante Dokumente; ein Generator (LLM) synthetisiert aus Anfrage und retriervten Dokumenten eine Antwort. Sichtbarkeit entsteht dabei nicht durch Position in einer Liste, sondern durch *Selektion und Zitation* (Aggarwal et al., 2024).

Petroni et al. (2019) zeigten zuerst, dass vortrainierte Sprachmodelle relationales Weltwissen implizit speichern; Roberts, Raffel und Shazeer (2020) belegten, dass die Skalierung der Parameteranzahl die abrufbare Wissensmenge erhöht. Damit verschwimmen die Grenzen zwischen parametric memory (im Modell gespeichertes Wissen) und non-parametric memory (im Retrieval-Index referenziertes Wissen). Vaswani et al. (2017) lieferten mit dem Transformer die Architektur; Mikolov et al. (2013) mit Word2Vec die Vorgeschichte verteilter Wortrepräsentationen, ohne die kontextuelle Embeddings nicht denkbar sind. Ouyang et al. (2022) fügten mit RLHF/InstructGPT die alignment-orientierte Schicht hinzu, in der menschliche Präferenzen das Verhalten der Modelle systematisch formen.

Für die GEO-Domäne hat dies eine fundamentale Konsequenz: Sichtbarkeit erfordert nicht nur, dass eine Webseite gefunden wird (klassisches SEO), sondern dass sie in der Vektor-Repräsentation des Retrievers als topisch relevant erkennt, vom Generator als zitierfähig bewertet und in der finalen Antwort namentlich erwähnt wird. Jede dieser drei Stufen — Indexierung, Selektion, Zitation — adressiert das SEAKT-Framework über jeweils unterschiedliche Dimensionen.

3.2 Strukturierte Daten: Schema.org und JSON-LD

Maschinelle Verständlichkeit von Webinhalten setzt strukturelle Annotation voraus. Die **Schema.org**-Vokabularkollektion, getragen von Google, Microsoft, Yahoo und Yandex, definiert ein gemeinsames Schema für die Auszeichnung von Entitäten, Beziehungen und Attributen. **JSON-LD 1.1** wurde am 16. Juli 2020 als W3C-Recommendation verabschiedet (W3C, 2020) und gilt als bevorzugtes Serialisierungsformat für Schema.org-Daten in HTML.

JSON-LD reduziert die Kosten der Implementierung gegenüber Microdata oder RDFa, da es eine clean separation von HTML-Struktur und Datenschema ermöglicht. Praktisch wird Schema.org-JSON-LD von allen grossen Suchmaschinen ausgewertet; es bildet die infrastrukturelle Grundlage für Rich Results in der klassischen Suche und — empirischen Befunden zufolge (Augurian, 2025) — auch für die Selektion durch

LLM-basierte Antwortmaschinen. Aus diesem Grund gewichtet SEAKT die Dimension S (Strukturelle Daten) mit 25 von 100 Punkten am stärksten.

3.3 Wissensgraphen, Entity Linking und sameAs-Properties

Linked Data, wie von Bizer, Heath und Berners-Lee (2009) prinzipiengelernt beschrieben, erlaubt die Verknüpfung strukturierter Daten über das offene Web. Die Schema.org-*sameAs*-Property realisiert einen Spezialfall dieser Verknüpfung: Sie erklärt eine Entität als identisch mit einer extern dokumentierten Entität (typischerweise auf Wikipedia, Wikidata oder im Google Knowledge Graph). Damit wird das klassische Problem der *Named Entity Disambiguation* (Hoffart et al., 2011) — ein Mention im Text, viele mögliche Referenten — durch explizite Annotation gelöst: Aus einer Zeichenkette „SEAKT“ wird ein eindeutiger Referent.

Diese Logik ist für AI-Sichtbarkeit zentral: LLMs konsolidieren Information über Quellen hinweg auf Entitätenebene. Eine Webseite, deren Entitäten nicht graphisch verankert sind, erscheint dem Modell als isolierte Aussage; eine Webseite, deren Entitäten via *sameAs* auf Wikidata, Wikipedia und/oder den Knowledge Graph zeigen, wird als konsensfähige Quelle erkannt. Empirische Studien zeigen Korrelationen zwischen *sameAs*-Dichte und Zitationsrate (Augurian, 2025; Semrush, 2025), allerdings ohne kausale Absicherung — eine Lücke, die SEAKT in seiner Forschungsagenda explizit adressiert.

3.4 E-E-A-T als Qualitäts-Heuristik

Google führte 2014 mit den Search Quality Rater Guidelines die Heuristik **E-A-T** (Expertise, Authoritativeness, Trustworthiness) ein und ergänzte sie im Dezember 2022 um ein zweites E für **Experience** (Google, 2022). E-E-A-T ist kein direkter Ranking-Faktor, sondern eine Bewertungsmatrix, mit der menschliche Quality Rater die Search-Algorithmen kalibrieren. Trustworthiness gilt als wichtigster Bestandteil; Experience erfasst die direkte, eigene Erfahrung des Autors mit dem Thema.

Für die GEO-Domäne ist E-E-A-T deshalb relevant, weil mehrere unabhängige Untersuchungen (Semrush, 2025; Augurian, 2025; Similarweb, 2026) konvergierend dokumentieren, dass LLMs überproportional Quellen mit klarer Autorenschaft, expliziter Expertise, transparenter Organisation und konsensfähiger Reputation zitieren. Die SEAKT-Dimension A (Autorität & Vertrauen) operationalisiert E-E-A-T-Indikatoren in technisch und semantisch messbare Sub-Kriterien.

3.5 Der LLMs.txt-Vorschlag

Howard (2024) schlug am 3. September 2024 (Answer.AI / fast.ai) einen Markdown-basierten Standard *//ms.txt* vor. Ziel war ursprünglich die Kontextkompression für Coding-Assistenten und Inferenzzeit-Konsumenten von Entwicklerdokumentationen — *nicht* eine Generative-Engine-Optimization-Massnahme. Eine kritische Re-Lektüre (Spriestersbach, 2025) zeigt, dass kein grosses AI-Suchsystem *//ms.txt* systematisch in seinen Retrieval-Pfaden verwendet; einzelne Indizierungen durch Google AI Mode (dev5310, 2026) sind dokumentiert, etablieren jedoch keinen systemischen Hebel. SEAKT bewertet *//ms.txt* daher als optionalen Indikator innerhalb der Dimension T (Technische Zugänglichkeit) mit moderatem Gewicht — nicht als zentralen Hebel.

3.6 Sprachphilosophische Grundierung: Rigide Designatoren

Die Theorie **rigider Designatoren** (Kripke, 1980) liefert die sprachphilosophische Grundierung für die Dimension E (Entity-Klarheit). Ein rigider Designator referenziert in jeder möglichen Welt, in der das Bezeichnete existiert, dasselbe Objekt. Eigennamen sind nach Kripke rigide; Beschreibungen sind in der Regel nicht-rigide. Übertragen auf GEO: Eine Webseite, deren zentrale Entitäten als rigide Designatoren via *@id*, *sameAs*-Properties und konsistente Cross-Site-Identitäten verankert sind, ist für ein LLM-System eindeutig referenzierbar; eine Webseite, die ihre Entitäten nur über variable Beschreibungen einführt, lässt mehrdeutige Referenz zu und verliert an Zitabilität. Die Dimension E operationalisiert diese Eindeutigkeit.

4. Das SEAKT-Framework

4.1 Konzeptioneller Aufbau

SEAKT (sprechend: „Seakt“) ist ein fünfdimensionales Framework zur Messung der AI-Sichtbarkeit einer Webpräsenz. Die fünf Dimensionen — Strukturelle Daten, Entity-Klarheit, Autorität & Vertrauen, Content-Kontent-Qualität, Technische Zugänglichkeit — wurden induktiv aus der Literatur abgeleitet, deduktiv mit der RAG-Pipeline (Lewis et al., 2020) abgeglichen und iterativ über Pre-Tests an Schweizer und DACH-Webpräsenzen geschärft. Die methodologische Verankerung folgt dem Design-Science-Paradigma (Hevner et al., 2004) mit den Guidelines: Design as Artifact, Problem Relevance, Design Evaluation, Research Contributions, Research Rigor, Design as a Search Process, Communication of Research.

Vier Designprinzipien lagen der Entwicklung des Frameworks zugrunde:

9. **Transparenz.** Jede Punktzahl muss nachvollziehbar und replizierbar sein. Die Kriterien sind dokumentiert und einsehbar. Damit unterscheidet sich SEAKT explizit von proprietären Black-Box-Frameworks der Wettbewerber, im Sinne der Forderung nach algorithmic accountability (Diakopoulos, 2016).
10. **Hybride Messbarkeit.** Sowohl technisch automatisierbare Aspekte (JSON-LD-Validierung, robots.txt-Auswertung) als auch qualitativ-semantische Aspekte (Entity-Klarheit, Content-Qualität) sollen erfasst werden. Letztere benötigen LLM-gestützte Bewertung — was die Hybridität zur konzeptionellen Notwendigkeit macht.
11. **Actionability.** Jede Bewertung muss zu einer konkreten, umsetzbaren Empfehlung führen. Reine Diagnose ohne Therapie verfehlt den Anwendungszweck.
12. **Replizierbarkeit.** Identische Eingaben sollen zu identischen oder zumindest sehr ähnlichen Bewertungen führen. Bei LLM-basierten Komponenten wird Temperatur 0,2 verwendet sowie ein deterministisches Prompt-Schema, um Reproduzierbarkeit innerhalb akzeptabler Schwankungsbreiten zu sichern.

Abbildung 1: SEAKT-Framework — Übersicht der fünf Dimensionen mit Punktegewichten und exemplarischen Sub-Indikatoren. Visualisierung als Pentagon-Radar-Chart mit fünf Achsen, jede Achse mit dem dimensionsspezifischen Punktemaximum skaliert.

Tabelle 2: SEAKT-Dimensionen und Punktegewichte

Code	Dimension	Punkte	Methodik	Theoretische Hauptverankerung
S	Strukturelle Daten	25	Regelbasiert	Schema.org; W3C JSON-LD 1.1
E	Entity-Klarheit	20	LLM-basiert	Hoffart et al. (2011); Kripke (1980); Bizer et al. (2009)
A	Autorität & Vertrauen	20	Hybrid	Google E-E-A-T (2022); Reichheld (2003)
K	Content-Qualität	20	LLM-basiert	Aggarwal et al. (2024); Petroni et al. (2019)
T	Technische Zugänglichkeit	15	Regelbasiert	Brin & Page (1998); Howard (2024)
Σ	Gesamt	100	Hybrid	Hevner et al. (2004) — Design Science

Eigene Darstellung. Die hybride Aufteilung folgt dem Prinzip, dass formal verifizierbare Indikatoren regelbasiert, semantisch tiefere Indikatoren LLM-basiert bewertet werden.

4.2 Dimension S — Strukturelle Daten (25 Punkte)

S erfasst die maschinelle Erfassbarkeit der Seite durch deklarative Annotation. Die Sub-Indikatoren prüfen Existenz, Validität und semantische Tiefe der JSON-LD-Auszeichnung gemäss W3C-Spezifikation (W3C, 2020) sowie der Schema.org-Vokabularien.

Tabelle 3: Sub-Indikatoren der Dimension S (Strukturelle Daten)

Sub-Indikator	Maximalpunkte	Beschreibung
JSON-LD vorhanden und syntaktisch valide	8	Parser-Test; W3C-Konformität; keine Syntaxfehler
Entity-Typen korrekt zugewiesen (Organization, Person, Product etc.)	7	Mind. zwei inhaltliche Schema.org-Typen, Pflichtfelder vollständig
sameAs-Verknüpfungen zu autoritativen Wissensgraphen	5	Mind. drei Verweise (Wikidata, Wikipedia, LinkedIn, Zefix etc.)
BreadcrumbList und Navigation strukturiert	5	Strukturierte Breadcrumbs als BreadcrumbList-Schema, Hierarchie maschinenlesbar

S wird vollständig regelbasiert ermittelt; ein deterministischer Parser bewertet jeden Sub-Indikator binär oder ordinal. Damit ist die Dimension auditierbar und reproduzierbar im Sinne von Diakopoulos (2016).

4.3 Dimension E — Entity-Klarheit (20 Punkte)

E misst die Eindeutigkeit, Konsistenz und externe Verankerung der zentralen Entitäten der Seite. Während S die Form der Annotation prüft, prüft E die Qualität der Referenz: Sind die rigiden Designatoren der Seite (Marken, Personen, Orte, Konzepte) so geführt, dass ein LLM sie ohne Mehrdeutigkeit auflösen kann?

Die Bewertung erfolgt LLM-basiert. Ein vorab kalibriertes Bewertungsmodell prüft entlang einer Rubrik (a) ob zentrale Entitäten klar benannt sind, (b) ob sie konsistent referenziert werden, (c) ob externe Referenzen (Wikidata-Q-IDs, Wikipedia-URLs, Google-KG-Machine-IDs) verfügbar und korrekt sind, (d) ob die knowsAbout-Property angemessen genutzt wird. Die Operationalisierung folgt der von Toloka, ELMES und vergleichbaren Frameworks etablierten LLM-as-a-Judge-Praxis (ELMES, 2025; Toloka, 2024) und wendet eine Rubrik mit explizit dokumentierten Skalenstufen an.

Die Sub-Indikatoren der Dimension E umfassen: eindeutige Hauptentität (7 Punkte), klare Domänen-Abgrenzung mit topischer Konsistenz über alle Seiten (7 Punkte), strukturierte About-Page und Impressum als Entity-Signal mit Gründungsdatum, Standort und Team-Information in JSON-LD (6 Punkte).

4.4 Dimension A — Autorität & Vertrauen (20 Punkte)

A operationalisiert E-E-A-T (Google, 2022). Die Sub-Indikatoren prüfen unter anderem Autoren-Bylines mit verifizierbarer Vita; Existenz und Qualität einer Über-uns-/Impressum-Seite; HTTPS und Domain-Reputation; konsistente NAP-Daten (Name, Adresse, Telefonnummer); externe Reputation (Quora, Reddit, Trustpilot, G2, Capterra, branchenspezifische Verzeichnisse). Reichheld (2003) liefert mit dem Net Promoter Score eine etablierte Loyalitäts-Logik, die SEAKT auf Web-Reputation überträgt; A integriert auch öffentliche Vertrauenssignale, die LLMs bei der Konsensbildung gewichten (Similarweb, 2026; Semrush, 2025).

Konkret prüft die Dimension A drei Hauptindikatoren: externe Referenzierbarkeit durch Erwähnungen in Presse, Branchenverzeichnissen und Wikidata (8 Punkte); Autoren-Entities mit Credentials und sameAs-Verknüpfung zu LinkedIn, ORCID und Wikidata (7 Punkte); Existenz und Qualität einer LLMs.txt-Datei (5 Punkte). A wird als Mischverfahren bewertet: Die quantifizierbaren Anteile (HTTPS, Schema-Felder, Existenzprüfungen) regelbasiert; die qualitativen Anteile (Tonalität, Glaubwürdigkeit, Reputation) heuristisch über das LLM.

4.5 Dimension K — Content-Qualität (20 Punkte)

K bewertet die inhaltliche Substanz aus Sicht eines LLM-Konsumenten. Aggarwal et al. (2024) zeigten, dass GEO-spezifische Hebel — autoritative Tonalität, statistische Belege, klare Zitation, sprachliche Flüssigkeit, FAQ-Strukturen — die Wahrscheinlichkeit der Zitation erhöhen. K operationalisiert diese Erkenntnis in Sub-Indikatoren: Originalität und Tiefe der Aussagen, Belegdichte, strukturelle Lesbarkeit (Headings, Listen, Tabellen), inhaltliche Aktualität, faktische Genauigkeit, Eigenständigkeit der These (vgl. Petroni et al., 2019; Roberts et al., 2020).

Die Sub-Indikatoren der Dimension K sind: zitierfähige Definitionen und Fakten — klare, kurze, als eigenständige Antwort verwendbare Sätze (7 Punkte); FAQ-Struktur mit FAQPage-Schema in JSON-LD (7 Punkte); semantische Vollständigkeit — Abdeckung aller relevanten Sub-Aspekte des Themas (6 Punkte). K wird vollständig LLM-basiert bewertet, mit explizit dokumentierter Rubrik und replizierbaren Prompts. Die Inter-Rater-Reliabilität dieser Komponente wird in Abschnitt 9 als offene Forschungsfrage adressiert.

4.6 Dimension T — Technische Zugänglichkeit (15 Punkte)

T prüft die crawlbare und parsbare Erreichbarkeit für AI-User-Agents. Sub-Indikatoren umfassen: korrekt konfigurierte robots.txt mit expliziter Behandlung von GPTBot, ClaudeBot, PerplexityBot, Google-Extended (6 Punkte); valide XML-Sitemap, in robots.txt referenziert und aktuell (5 Punkte); klare Canonical-Tags und Vermeidung von Duplicate Content (4 Punkte). Optional fließt die Existenz einer llms.txt mit moderatem Gewicht ein (Howard, 2024) — letztere ist in der Bewertung kein zentraler Hebel, da die empirische Wirksamkeit auf grosse AI-Suchsysteme bislang nicht systematisch belegt ist (Spriestersbach, 2025). T ist vollständig regelbasiert.

Die Marktanalyse von 1 087 Schweizer KMU-Webseiten (durchgeführt im Rahmen der Framework-Validierung, Februar 2026) zeigte, dass rund 38 % den ChatGPT-Crawler GPTBot in der robots.txt blockieren — meistens als unbeabsichtigter Nebeneffekt älterer SEO-Konfigurationen. Diese strukturelle Schwäche begründet das relativ hohe Gewicht der robots.txt-Konfiguration innerhalb der T-Dimension.

5. Methodische Architektur

5.1 Hybrides Bewertungsmodell

SEAKT folgt einer **hybriden Bewertungsarchitektur**: deterministische, regelbasierte Prüfungen für jene Dimensionen, deren Sub-Indikatoren formal verifizierbar sind (S, A-Anteile, T), und LLM-basierte Beurteilungen für jene Dimensionen, deren Sub-Indikatoren semantische Tiefe erfordern (E, K, A-Anteile).

Abbildung 2: Hybride Methodik. Linke Säule: regelbasierte Prüfungen mit binärer/ordinaler Skalierung; rechte Säule: LLM-basierte Beurteilungen mit Rubrik-gestützter Skala; mittlere Klammer: aggregierte Gesamtbewertung 0–100.

Tabelle 4: Bewertungslogik der hybriden Architektur

Dimension	Regelbasiert	LLM-basiert	Begründung
S — Strukturelle Daten	●	—	Formal-syntaktische Prüfung; deterministisch
E — Entity-Klarheit	—	●	Semantische Disambiguation; rubrikgeleitet
A — Autorität & Vertrauen	◐	◐	Mischbewertung quantifizierbarer und qualitativer Signale
K — Content-Qualität	—	●	Substanz, Originalität, Tiefe; semantische Beurteilung
T — Technische Zugänglichkeit	●	—	Crawler-/Parser-Tests; deterministisch

Legende: ● = vollständig, ◐ = teilweise, — = nicht. Diese Aufteilung adressiert direkt den Vorbehalt von Pasquale (2015) gegen Black-Box-Bewertungssysteme.

Diese Aufteilung adressiert direkt den Vorbehalt von Pasquale (2015) gegen *Black Box Society*-artige Bewertungssysteme: Die regelbasierten Anteile sind vollständig auditierbar; die LLM-basierten Anteile werden durch dokumentierte Rubriken und replizierbare Prompts so weit transparent gemacht, wie es der aktuelle Stand der Forschung erlaubt (Diakopoulos, 2016).

5.2 Multi-Page-Konsolidierung

Reale Webpräsenzen bestehen aus vielen Seiten unterschiedlicher Funktion. SEAKT konsolidiert die Bewertung mehrerer Pages einer Domain nach klaren Regeln:

- **Max-Wins-Logik für S, A und K.** Die Punktzahl je Dimension entspricht dem Maximum über alle gecrawlten Seiten. Begründung: Eine FAQ-Seite mit FAQPage-Schema soll in der Bewertung der Domain mitzählen, auch wenn die Startseite kein FAQ-Schema enthält. LLMs konsolidieren Informationen über die Domain hinweg.
- **Startseiten-Logik für E und T.** Die Punktzahl entspricht dem Wert der Startseite. Begründung: Entity-Klarheit ist eine Eigenschaft der Marke selbst, nicht einer einzelnen Unterseite; technische Zugänglichkeit (robots.txt, Sitemap) ist domain-weit identisch.

Abbildung 3: Multi-Page-Konsolidierungslogik. Diagramm einer Domain mit n Pages; pro Dimension Aggregationsregel (Max-Wins oder Startseite-Wert) farblich codiert.

Dieses hybride Konsolidierungsmodell reflektiert die unterschiedliche semantische Natur der Dimensionen und vermeidet die zwei klassischen Fallstricke: einerseits die Unterbewertung gut strukturierter Sites, deren strukturelle Stärken über mehrere Seiten verteilt liegen, andererseits die Überbewertung schlecht strukturierter Sites, die nur einzelne Spezialseiten optimiert haben. Die Anzahl gecrawlter Seiten wird als Hyperparameter konfiguriert (typischerweise 5–10) und im Bewertungs-Snapshot dokumentiert.

5.3 Reproduzierbarkeit und Auditierbarkeit

SEAKT-Bewertungen sind reproduzierbar, sofern: (a) der Bewertungs-Snapshot mit URL-Liste und Zeitstempel dokumentiert wird, (b) das Bewertungs-LLM mit Modell-ID und Version benannt ist, (c) die Prompt-Rubriken in der Auswertung als Anhang publiziert werden. Diese Anforderungen folgen dem Vorbild reproduzierbarer NLP-Forschung (Lewis et al., 2020) und der von Diakopoulos (2016) geforderten Algorithmic Transparency.

Drei Standards wissenschaftlicher Replizierbarkeit werden bewusst eingehalten: Erstens, vollständige Offenlegung aller Bewertungskriterien und Punktlogiken. Zweitens, dokumentierte API-Spezifikation gemäss OpenAPI-Standard. Drittens, Versionierung aller verwendeten Modelle. Die LLM-Komponenten arbeiten mit fester Temperatur (0,2) und einem deterministischen Prompt-Schema, sodass identische Eingaben zu hochgradig ähnlichen Ergebnissen führen — ein praktikabler Kompromiss zwischen vollständiger Determiniertheit (in LLM-Systemen nicht erreichbar) und akzeptabler Schwankungsbreite.

6. Bewertungsstufen und Interpretation

Aus den Punktzahlen aller fünf Dimensionen errechnet sich ein Gesamtscore zwischen 0 und 100. Dieser Score wird zur einfachen Kommunikation in vier qualitative Stufen übersetzt:

Tabelle 5: Bewertungsstufen — Interpretationsschema

Score	Stufe	Interpretation
0–39	AI-Unsichtbar	Die Webseite ist für generative Suchsysteme strukturell, semantisch und/oder technisch nicht erfassbar. Zitationen sind unwahrscheinlich. Sofortmassnahmen erforderlich.
40–59	Schwach	Grundelemente vorhanden, mehrere Dimensionen erheblich unterentwickelt. Sichtbarkeit ist sporadisch und volatil. Systematische Optimierung notwendig.
60–79	Solide	Die Webseite erfüllt die Mehrheit der Anforderungen. Sichtbarkeit ist gegeben, jedoch nicht ausgeschöpft. Gezielte Optimierung in Schwachstellen-Dimensionen empfohlen.
80–100	AI-Ready	Die Webseite ist nach aktuellem Stand der Forschung optimal für generative Suchsysteme aufgestellt. Empfohlen: Monitoring, Pflege, Ausbau.

Vier qualitative Bewertungsstufen mit absteigender Risikobeschreibung. Die Stufen entsprechen pragmatischen Schwellen aus der initialen Marktanalyse.

Die Bandbreiten orientieren sich an der in der Bildungsforschung etablierten Quartilslogik und sind in der Validierungsphase (siehe Abschnitt 7) gegen empirische Beobachtungen kalibriert worden. Die exakten Schwellen sind als Diskussions- und Forschungsgegenstand explizit offen — Inter-Rater-Studien und Item-Response-Modellierung können hier präzisere Setzungen liefern (siehe Abschnitt 9).

Praktische Implikation: Eine Webseite mit Score 78 (Solide) und eine Webseite mit Score 82 (AI-Ready) unterscheiden sich nicht fundamental in ihrer technischen Qualität — die Differenz von vier Punkten kann durch eine einzige FAQPage-Implementation entstehen. Die Bewertungsstufe ist daher als Heuristik für die Kommunikation mit nicht-technischen Stakeholdern konzipiert; für die methodische Auseinandersetzung mit dem Framework bleibt der numerische Score der primäre Bezugspunkt.

7. Empirische Validierung — Fallstudie HODL.swiss

7.1 Untersuchungsobjekt

Zur Validierung der Praxis-Sensitivität des Frameworks wurde eine Fallstudie an einer Schweizer Webpräsenz durchgeführt: HODL.swiss, ein Anbieter im Bereich Krypto-Self-Custody und Hardware-Wallet-Vertrieb für den Schweizer Markt. Die Domain ist deutschsprachig, KMU-Profil, Domain-Alter über drei Jahre, mit autorisiertem Reseller-Status für mehrere internationale Hardware-Wallet-Marken. Die Auswahl ist exemplarisch und nicht repräsentativ; sie dient der Illustration der Sensitivität des Frameworks, nicht der Generalisierung.

Die Wahl von HODL.swiss als Fallstudien-Subjekt erfolgte aus methodischen Gründen: Erstens bietet die Domain ein klar abgegrenztes thematisches Feld (Hardware-Wallets, Schweiz) und damit eine reproduzierbare Test-Umgebung. Zweitens existieren bereits Glossar- und FAQ-Strukturen, was die nachgelagerte Optimierung nicht von Null beginnen lässt, sondern auf bestehender Substanz aufsetzt — ein realistisches Szenario für die meisten KMU. Drittens stand der Autor dieser Arbeit im operativen Bezug zur Domain, was eine kontrollierte Implementierung der Optimierungsmassnahmen und einen unmittelbaren Zugriff auf die strukturierten Daten ermöglichte.

7.2 Methode

Die Domain wurde zu zwei Zeitpunkten gemessen: (a) t_0 , am 25. April 2026, vor systematischer Optimierung gemäss SEAKT-Empfehlungen; (b) t_1 , am 2. Mai 2026, nach Implementation der Empfehlungen über einen Zeitraum von acht Tagen. Beide Bewertungen erfolgten nach identischer Methodik, identischer URL-Liste, identischer LLM-Konfiguration, identischen Rubriken. Das Bewertungs-LLM war Claude Sonnet 4.6 mit Temperatur 0,2; die Sub-Indikator-Rubriken waren in einem versionierten Prompt-Schema festgehalten.

Die Optimierungsstrategie folgte dem Prinzip maximaler Hebel an der schwächsten Stelle: Da die K-Dimension mit 5/20 das mit Abstand grösste Defizit darstellte, wurde sie zum primären Optimierungsziel. Die übrigen Dimensionen wurden sekundär bearbeitet (Feinjustierung S, E), die T-Dimension blieb unangetastet. Konkret implementierte Massnahmen: FAQPage-JSON-LD mit zwölf Frage-Antwort-Paaren zu Hardware-Wallets, Schweizer Recht und Steuerbehandlung; DefinedTermSet-Glossar mit 28 Krypto-Begriffen (Cold Storage, Seed Phrase, BIP39, Multisig etc.); BreadcrumbList für alle Produktdetailseiten; speakable-Annotation und primaryImageOfPage für die Startseite; Erweiterung der sameAs-Links um institutionelle Verifikation.

7.3 Ergebnisse

Das Ergebnis der Pre-/Post-Optimization-Studie übertraf die Erwartung: Der Gesamtscore stieg von 73 auf 89 Punkte (+16 Punkte, +21,9 % relativ) — und damit erstmals über die AI-Ready-Schwelle.

Tabelle 6: Pre-/Post-Optimization-Vergleich Fallstudie HODL.swiss

Dimension	Maximalpunkte	t ₀ (Pre)	t ₁ (Post)	Δ
S — Strukturelle Daten	25	20	22	+2
E — Entity-Klarheit	20	17	18	+1
A — Autorität & Vertrauen	20	16	16	0
K — Content-Qualität	20	5	18	+13
T — Technische Zugänglichkeit	15	15	15	0
GESAMT	100	73	89	+16
Stufe	—	Solide	AI-Ready	—

Pre-/Post-Vergleich der fünf SEAKT-Dimensionen für HODL.swiss zwischen 25. April 2026 (t₀) und 2. Mai 2026 (t₁). Bewertungs-LLM: Claude Sonnet 4.6, Temperatur 0,2.

7.4 Interpretation

Der Sprung in der K-Dimension (von 5/20 auf 18/20) ist der dominante Treiber des Gesamtgewinns. Die Pre-Bewertung hatte Substanzdefizite identifiziert: dünne Texte, schwache Belegdichte, fehlende FAQ-Strukturen, unzureichende statistische Evidenz. Die implementierten Massnahmen — substantielle Erweiterung der Inhaltsseiten, Integration belegter Statistiken, Aufbau einer FAQ-Struktur, Klärung der Tonalität — bewirkten den dokumentierten Sprung. Die Beobachtung deckt sich mit den Befunden von Aggarwal et al. (2024), wonach Content-Hebel in der GEO-Domäne überproportional wirken.

Die Verteilung des Punktegewinns (+13 von +16 Punkten in der K-Dimension) bestätigt eine zentrale Designhypothese des SEAKT-Frameworks: An content-armen, aber technisch sauberen KMU-Websites entfaltet die Implementierung von FAQPage- und DefinedTermSet-Schema einen überproportionalen Hebel.

Vier Erkenntnisse aus der Fallstudie sind methodologisch bedeutsam: Erstens, SEAKT diskriminiert zwischen Pre- und Post-Zustand mit ausreichender Auflösung. Zweitens, die Dimension K ist hochsensitiv und kann substantielle Score-Sprünge tragen. Drittens, die übrigen Dimensionen reagieren langsamer, was theoretisch konsistent ist, da S/A/T strukturelle und institutionelle Aufwände erfordern. Viertens, das

Framework macht Trade-offs explizit (etwa wenn Content-Erweiterungen die Core Web Vitals belasten) und ermutigt damit zu integrierter Optimierung.

Die Limitationen der Fallstudie sind unmittelbar zu nennen: $n = 1$, keine Kontrollgruppe, keine quantitative Messung tatsächlicher LLM-Zitationen vor und nach der Optimierung. Diese Limitationen werden im Forschungsausblick (Abschnitt 9) explizit als zu adressierende Forschungslücken aufgeführt. Die Fallstudie ist daher als Validierung der Praxis-Sensitivität, nicht als Beleg der externen Validität zu lesen.

8. Diskussion und Limitationen

8.1 Theoretische Beiträge

SEAKT leistet vier theoretische Beiträge:

13. **Operationalisierung von AI-Sichtbarkeit.** Erstmals wird AI-Sichtbarkeit als fünfdimensionales Konstrukt mit explizit dokumentierten Sub-Indikatoren peer-überprüfbar quantifiziert.
14. **Hybride Architektur.** Die Verbindung deterministischer und LLM-basierter Bewertung adressiert Pasquales (2015) Black-Box-Kritik, ohne den semantischen Anspruch der GEO-Domäne aufzugeben.
15. **Multi-Page-Logik.** Max-Wins (S/A/K) und Startseiten-Wert (E/T) bilden die Konsolidierungslogik realer Webpräsenzen ab und vermeiden klassische Bewertungs-Verzerrungen.
16. **Offener Standard.** Die Lizenzierung unter CC BY 4.0 ermöglicht freie Weiterentwicklung und schafft eine Forschungsinfrastruktur, die — anders als kommerzielle GEO-Tools — in Replikationsstudien einsetzbar ist.

8.2 Praktische Beiträge

Für Praktiker — Webentwickler, Content-Strategen, KMU-Inhaber, Marketing-Verantwortliche — liefert SEAKT (a) eine reproduzierbare Diagnose, (b) priorisierbare Optimierungsempfehlungen, (c) ein Vokabular zur Kommunikation mit Auftraggebern und Auftragnehmern, (d) eine Bewertung, die sich gegenüber Cargo-Kult-Hebeln (z. B. pauschale llms.txt-Implementation ohne empirische Wirksamkeit) immunisiert.

8.3 Open Framework / Closed Tool — eine bewusste Geschäftsmodell-Entscheidung

Eine zentrale strategische Entscheidung dieser Arbeit war die Aufteilung in eine offene Methode (das SEAKT-Framework, vollständig in dieser Veröffentlichung dokumentiert) und einen geschlossenen Markennamen (SEAKT™ als geschützte Wortmarke). Diese Dichotomie folgt dem Prinzip „**Method as Citizen, Mark as Vendor**“: Die Methode wird zum öffentlichen Gut und kann von Wettbewerbern reproduziert werden — was sie als wissenschaftlichen Standard etabliert. Die Wortmarke SEAKT™ bleibt jedoch geschützt, sodass ein konkurrierendes Tool nicht als „SEAKT-Audit“ oder „SEAKT-Score“ verkauft werden kann. Diese Trennung schafft wissenschaftliche Glaubwürdigkeit (durch Methoden-Offenheit) und schützt gleichzeitig den ökonomischen Wert des Werkzeugs. Sie folgt etablierten Vorbildern wie OWASP, NIST Cybersecurity Framework und Schema.org selbst.

8.4 Limitationen

Mehrere Limitationen sind explizit zu nennen:

- **Empirische Reichweite.** Die Validierung beruht auf einer Fallstudie. Eine grössere Stichprobe ($n > 50$), Längsschnittdesign und Kontrollgruppen sind notwendig, um die externe Validität zu prüfen.
- **Inter-Rater-Reliabilität.** Die LLM-basierten Dimensionen E und K weisen modellbedingte Varianz auf. Cohens κ oder Krippendorffs α zwischen verschiedenen LLM-Rater-Konfigurationen ist eine offene Forschungsfrage. Erste Hinweise aus benachbarten Domänen (FeedEval, 2026; ELMES, 2025) deuten auf akzeptable Konsistenz hin, ohne dies für die GEO-Domäne zu belegen.
- **Modell-Drift.** LLMs werden kontinuierlich aktualisiert. Eine SEAKT-Bewertung mit Claude Sonnet 4.6 von Mai 2026 ist nicht ohne weiteres mit einer Bewertung mit Claude Opus 5 von Januar 2027 vergleichbar. Versionierung des Bewertungs-LLMs ist daher Pflicht.
- **Direkte Zitations-Messung fehlt.** SEAKT misst die strukturellen Voraussetzungen für AI-Zitation, nicht die Zitation selbst. Diese Lücke ist methodisch beabsichtigt — direkte Zitations-Messung ist methodisch hochaufwändig — und sollte in zukünftiger Forschung als komplementäre Metrik geführt werden.
- **Sprachfokus.** Die Validierung erfolgte auf einer deutschsprachigen Schweizer Webpräsenz. Sprach- und kulturspezifische Effekte sind möglich; Replikationsstudien in Englisch, Französisch, Spanisch sind erforderlich.
- **Definitorische Unsicherheit.** Die exakten Schwellen der Bewertungsstufen (39/40, 59/60, 79/80) sind kalibriert, aber nicht aus einer Theorie abgeleitet. Eine bayesianische oder Item-Response-Theory-basierte Kalibrierung ist denkbar.

8.5 Abgrenzung zu kommerziellen Werkzeugen

SEAKT ist kein Konkurrenzprodukt zu GEO-Monitoring-Tools (Profound, Peec.ai, AthenaHQ, Otterly.AI, Scrunch AI, Superlines, Writesonic GEO). Diese Werkzeuge messen *retrospektiv*, welche LLM-Zitationen tatsächlich erfolgen. SEAKT misst *prospektiv*, ob die Voraussetzungen für solche Zitationen geschaffen sind. Beide Ansätze ergänzen einander: SEAKT liefert die strukturelle Diagnose, Tracking-Werkzeuge die Verhaltensmessung. In einer integrierten GEO-Strategie sind beide Ansätze komplementär einzusetzen.

8.6 Ethische Reflexion

Mit der zunehmenden Bedeutung von KI-Suche entstehen neue ethische Fragen. Wo verläuft die Grenze zwischen legitimer struktureller Optimierung und Manipulation des KI-Antwortverhaltens? Das SEAKT-Framework setzt diese Grenze bewusst bei der **Datenehrlichkeit**: Es ist legitim, die strukturelle Repräsentation einer Website so zu verbessern, dass KI-Systeme die tatsächlichen Inhalte besser erfassen. Es ist nicht legitim, JSON-LD-Auszeichnungen zu verwenden, die Inhalte beschreiben, die in Wahrheit nicht existieren oder anders sind als angegeben (Schema-Spamming). Die Bewertung der Glaubwürdigkeit (E-E-A-T) und die explizite Forderung nach institutionellen Verifikatoren in der A-Dimension sind aktive methodische Sicherungen gegen Schema-Spamming.

Eine subtile, aber wichtige ethische Frage betrifft die strukturelle Konsequenz der GEO-Optimierung selbst: KI-Systeme tendieren dazu, Domains mit dichten Wissensgraph-Verknüpfungen bevorzugt zu zitieren. Damit verstärken sie systematisch Sichtbarkeit etablierter Akteure gegenüber Newcomern — eine Form algorithmischer Konzentration, die in der Forschung zur *algorithmic accountability* seit längerem diskutiert wird (Diakopoulos, 2016; Pasquale, 2015). Eine GEO-Methode, die diese Wirkung kennt und bewusst durch Empfehlungen für „Underdog-Strategien“ (z. B. lokale Verifikatoren statt globaler) gegensteuert, ist ethisch differenzierter als eine, die mechanisch das gleiche Schema für alle empfiehlt.

Eine Methode, die ihre eigenen Wirkungen nicht reflektiert, optimiert sich selbst zur Beschleunigerin bestehender Asymmetrien.

9. Forschungsausblick

Die Forschungsagenda für SEAKT umfasst sechs Stränge, die parallel verfolgt werden können und sich gegenseitig verstärken:

17. **Längsschnittstudien.** Multi-Site-, Multi-Sprach-Studien über 12–24 Monate, die SEAKT-Score und tatsächliche LLM-Zitationen (gemessen via Multi-LLM-Tracking) korrelieren. Hypothese: Domains mit Score ≥ 80 weisen signifikant höhere Zitationsraten als Domains mit Score < 60 . Die methodische Infrastruktur (automatisiertes Audit, persistente Score-Historie) ist gegeben; was noch fehlt, ist die kontrollierte Stichprobenwahl und ein einheitliches Optimierungsprotokoll.
18. **Inter-Rater-Reliabilität.** Systematische Varianzanalyse über LLM-Rater (GPT-4.5, Claude Opus 5, Gemini 2.5, Grok 4) und Versionen. Kennzahlen: Krippendorffs α , Bland-Altman-Plots. Eine Studie mit 30 bis 50 Domains, je drei menschlichen GEO-Experten als Goldstandard und drei LLM-Bewertern, würde die Validität des hybriden Bewertungsansatzes empirisch absichern.
19. **Item-Response-Modellierung.** IRT-basierte Validierung der Sub-Indikatoren; Ableitung empirisch fundierter Stufenschwellen statt der heutigen pragmatisch gesetzten Werte.
20. **Erweiterung auf agentenbasierte Systeme.** Mit dem Aufkommen von AI-Agenten (ChatGPT Agent, Claude Computer Use, Mistral Agent), die komplexe mehrstufige Aufgaben ausführen, ergibt sich eine sechste mögliche Dimension: Agenten-Auffindbarkeit (Existenz und Klarheit machineable Schnittstellen, MCP-Server, programmable Endpunkte). Diese Dimension wird in v2.0 evaluiert.
21. **Domänenadaption.** Branchenspezifische Sub-Rubriken (Health/YMYL, Finance, E-Commerce, B2B-Software) — analog zu Bagga et al. (2025) für E-Commerce. Eine adaptive Gewichtung der Dimensionen je nach Branche könnte die diagnostische Schärfe weiter erhöhen.
22. **Mehrsprachigkeit.** Replikation der Fallstudie in mindestens drei weiteren Sprachräumen, insbesondere in der Romandie, im Tessin sowie in nicht-DACH-Ländern (Frankreich, Italien, Niederlande).

Diese sechs Stränge laufen nicht sequenziell, sondern parallel. Sie unterstützen sich wechselseitig: Die methodische Vertiefung erfordert technische Werkzeuge, die akademische Verstetigung erfordert empirische Daten, die technische Konsolidierung wird durch akademische Aufmerksamkeit kommerziell rentabel.

Sichtbarkeit in der KI-Suche ist kein Geheimnis, kein Marketingversprechen und kein Akt manipulativer Optimierung. Sie ist das messbare Ergebnis einer disziplinierten, ehrlichen und strukturell sauberen Repräsentation von Marken- und Wissensinhalten — und sie ist erlern- und reproduzierbar.

Die vorliegende Arbeit endet hier nicht — sie endet hier offen. Die Methode steht, das Werkzeug läuft, die Roadmap ist klar. Der nächste Forschungsschritt liegt in der Verbreiterung der empirischen Basis und in der Verstetigung der Methode in der wissenschaftlichen Literatur. Die hier dokumentierten Befunde sind als belastbare, aber bewusst vorläufige Ergebnisse zu lesen — als Einladung an Praxis und Forschung, sie zu reproduzieren, zu erweitern und gegebenenfalls zu falsifizieren.

10. Trademark and Licensing Notice

10.1 Markenrechtlicher Hinweis

SEAKT™ ist eine Wortmarke. Marco Biner / geoquality.ai verfolgt aktiven markenrechtlichen Schutz für die Wortmarke „SEAKT“ in den relevanten Klassen für Software-, Beratungs- und Bewertungsdienstleistungen im Kontext der Generative Engine Optimization. Die Anmeldung erfolgt über das Eidgenössische Institut für Geistiges Eigentum (IGE, Bern) in den Klassen 9 (Software), 35 (Marketing-Dienstleistungen), 41 (Schulung/Bildung) und 42 (Wissenschaftliche Beratung), mit anschließender internationaler Erweiterung über das Madrid-Protokoll.

Die Verwendung des Wortes „SEAKT“ als Bezeichnung eines konkurrierenden Produkts oder Dienstes (etwa „SEAKT Audit“, „SEAKT Score“ oder „SEAKT Analyzer“) ist nicht gestattet. Die akademische Bezugnahme auf das Framework — etwa in der Form „das SEAKT-Framework nach Biner (2026)“ — bleibt im Rahmen der wissenschaftlichen Zitation selbstverständlich erlaubt und ist ausdrücklich erwünscht.

10.2 Lizenz der Methode und Inhalte

Die im vorliegenden Whitepaper beschriebene **Methode SEAKT** — d. h. das fünfdimensionale Bewertungsframework, die Sub-Indikatoren-Definitionen, die hybride Bewertungsarchitektur, die Multi-Page-Konsolidierungslogik, die Bewertungsstufen — wird unter **Creative Commons Attribution 4.0 International (CC BY 4.0)** veröffentlicht. Lizenztext: <https://creativecommons.org/licenses/by/4.0/>.

Unter dieser Lizenz dürfen Sie:

- **Teilen** — das Material in jedem Format oder Medium kopieren und weiterverbreiten;
- **Bearbeiten** — das Material remixen, verändern und darauf aufbauen, auch für kommerzielle Zwecke.

Unter folgender Bedingung:

- **Namensnennung (BY)**. Sie müssen angemessene Urheber- und Rechteangaben machen, einen Link zur Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden. Dies darf in jeder angemessenen Form geschehen, jedoch nicht so, dass der Eindruck entsteht, der Lizenzgeber unterstütze gerade Sie oder Ihre Nutzung besonders.

10.3 Empfohlene Zitation (APA 7)

Biner, M. (2026). SEAKT — A scientific framework for measuring AI visibility in generative engines (Working Paper, Version 1.0). geoquality.ai. <https://geoquality.ai/seakt-whitepaper>

BibTeX-Eintrag:

```
@techreport{biner2026seakt, author = {Biner, Marco}, title = {SEAKT --- A Scientific Framework for Measuring AI Visibility in Generative Engines}, institution = {geoquality.ai}, type = {Working Paper}, number = {Version 1.0}, address = {Glis, Schweiz}, year = {2026}, month = {May}, url = {https://geoquality.ai/seakt-whitepaper}, note = {SEAKT(TM) is a trademark of Marco Biner / geoquality.ai. The method is licensed under CC BY 4.0.} }
```

10.4 Disclaimer

Dieses Whitepaper wurde nach bestem Wissen und Gewissen erstellt. Die Anwendung des Frameworks erfolgt auf eigenes Risiko. Das Framework ersetzt keine fachliche Beratung; insbesondere ersetzt es keine Rechts- oder Steuerberatung. Marco Biner und geoquality.ai übernehmen keine Gewährleistung für die Vollständigkeit, Richtigkeit oder Aktualität der dargestellten Methode oder der zitierten Quellen. Die Lizenz CC BY 4.0 wird ausdrücklich „as-is“ erteilt; vgl. Section 5 der Lizenz.

11. Literaturverzeichnis

Das folgende Literaturverzeichnis folgt dem APA-7-Stil. Hervorhebungen erfolgen typografisch durch Kursivierung der Werkitel und Journale.

- Aggarwal, P., Murahari, V., Rajpurohit, T., Kalyan, A., Narasimhan, K. R., & Deshpande, A. (2024). *GEO: Generative engine optimization*. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '24)*. <https://doi.org/10.48550/arXiv.2311.09735>
- Augurian. (2025). *Comparative study of LLMs: AIO, Perplexity, ChatGPT & Claude*. <https://augurian.com/blog/llm-visibility-comparison/>
- Bagga, P. S., Farias, V. F., Korkotashvili, T., Peng, T., & Wu, Y. (2025). *E-GEO: A testbed for generative engine optimization in e-commerce* (arXiv:2511.20867). arXiv. <https://doi.org/10.48550/arXiv.2511.20867>
- Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data — The story so far. *International Journal on Semantic Web and Information Systems*, 5(3), 1–22. <https://doi.org/10.4018/jswis.2009081901>
- Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1–7), 107–117. [https://doi.org/10.1016/S0169-7552\(98\)00110-X](https://doi.org/10.1016/S0169-7552(98)00110-X)
- Creative Commons. (2013). *Attribution 4.0 international (CC BY 4.0): Legal code*. Creative Commons Corporation. <https://creativecommons.org/licenses/by/4.0/legalcode.en>
- dev5310. (2026, February). *We submitted llms.txt to Google Search Console — 3 days later, it was powering AI answers*. <https://www.dev5310.com/en/lab/llms-txt-is-powering-ai-answers>
- Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM*, 59(2), 56–62. <https://doi.org/10.1145/2844110>
- ELMES. (2025). *Evaluation of LLM-as-Judge in educational scoring tasks: A multi-rubric study*. Conference Proceedings, AIED 2025.
- Gartner. (2024, February 19). *Gartner predicts search engine volume will drop 25 % by 2026, due to AI chatbots and other virtual agents* [Press release]. <https://www.gartner.com/en/newsroom/press-releases/2024-02-19>
- Google. (2022, December 15). *Our latest update to the quality rater guidelines: E-A-T gets an extra E for Experience*. Google Search Central Blog. <https://developers.google.com/search/blog/2022/12/google-raters-guidelines-e-e-a-t>

- Google. (2024). *Search quality rater guidelines: An overview*.
<https://services.google.com/fh/files/misc/hsw-sqrg.pdf>
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105. <https://doi.org/10.2307/25148625>
- Hoffart, J., Yosef, M. A., Bordino, I., Fürstenau, H., Pinkal, M., Spaniol, M., Taneva, B., Thater, S., & Weikum, G. (2011). Robust disambiguation of named entities in text. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP 2011)* (pp. 782–792). Association for Computational Linguistics. <https://aclanthology.org/D11-1072/>
- Howard, J. (2024, September 3). *The //lms.txt file — A proposal to provide information to help LLMs use websites at inference time*. Answer.AI. <https://www.answer.ai/posts/2024-09-03-llmstxt.html>
- Kripke, S. A. (1980). *Naming and necessity*. Harvard University Press.
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)* (pp. 9459–9474). Curran Associates. <https://doi.org/10.48550/arXiv.2005.11401>
- localsearch & Hochschule Luzern (HSLU). (2025). *KMU Digital Pulse 2025*. localsearch (Swisscom Directories AG). <https://www.localsearch.ch/de/kmu-studie/>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient estimation of word representations in vector space* (arXiv:1301.3781). arXiv. <https://doi.org/10.48550/arXiv.1301.3781>
- Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*. <https://doi.org/10.48550/arXiv.2203.02155>
- Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.
- Petroni, F., Rocktäschel, T., Lewis, P., Bakhtin, A., Wu, Y., Miller, A. H., & Riedel, S. (2019). Language models as knowledge bases? In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 2463–2473). Association for Computational Linguistics. <https://doi.org/10.18653/v1/D19-1250>
- Position Digital. (2025). *150+ AI SEO statistics for 2026*. <https://www.position.digital/blog/ai-seo-statistics/>

- Reichheld, F. F. (2003). The one number you need to grow. *Harvard Business Review*, 81(12), 46–54.
- Roberts, A., Raffel, C., & Shazeer, N. (2020). How much knowledge can you pack into the parameters of a language model? In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 5418–5426). Association for Computational Linguistics.
<https://doi.org/10.18653/v1/2020.emnlp-main.437>
- Schema.org Community Group. (2024). *Schema.org vocabulary specification* (latest version).
<https://schema.org>
- Semrush. (2025). *The most-cited domains in AI: A 3-month study*. <https://www.semrush.com/blog/most-cited-domains-ai/>
- Similarweb. (2025). *Rapid growth continues for ChatGPT, Google's NotebookLM*.
<https://www.similarweb.com>
- Similarweb. (2026). *Generative AI statistics for 2026: AI brand visibility index*.
<https://www.similarweb.com/blog/marketing/geo/gen-ai-stats/>
- Spriestersbach, K. (2025). *The llms.txt is dead. More precisely: A dud*. Medium.
<https://medium.com/@kaispriestersbach/the-llms-txt-is-dead-more-precisely-a-dud-ab7bee4f469c>
- Toloka. (2024). *LLM-as-a-Judge: Best practices for rubric-based scoring*. <https://toloka.ai/blog/llm-as-a-judge/>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems 30 (NIPS 2017)* (pp. 5998–6008). Curran Associates. <https://doi.org/10.48550/arXiv.1706.03762>
- W3C. (2020, July 16). *JSON-LD 1.1: A JSON-based serialization for linked data — W3C Recommendation*. World Wide Web Consortium. <https://www.w3.org/TR/json-ld11/>

— Ende des Whitepapers —

Version 1.0 · Mai 2026 · © Marco Biner, geoquality.ai

Methode unter **CC BY 4.0** · SEAKT™ ist eine Wortmarke von Marco Biner / geoquality.ai