1. Hauptkohorte Deutschland (n = 10.648 erreichbare Websites)

Untersuchung von Startseiten aus öffentlichen Verzeichnissen (Messzeitraum: Sept. - Okt. 2026)

Geprüftes Merkmal	Bezugsgröße (Nenner)	Anteil (%)	Absolute Fallzahl
Anschrift im sichtbaren Text / Impressum	n = 10.648 (Gesamtkohorte)	**94,1 %**	ca. 10.020
Kein Schema.org-Unternehmens-Markup	n = 10.648 (Gesamtkohorte)	**73,0 %**	7.778
Schema.org-Unternehmens-Markup vorhanden	n = 10.648 (Gesamtkohorte)	**27,0 %**	2.870
- davon verschachtelte PostalAddress	n = 2.870 (Teilgruppe mit Markup)	**21,3 %**	611
- PostalAddress bezogen auf Gesamtkohorte	n = 10.648 (Gesamtkohorte)	**6,5 %**	695
Valide /llms.txt vorhanden (nach Soft-404)	n = 10.648 (Gesamtkohorte)	**4,7 %**	497
/llms.txt vor Soft-404-Filterung (roh)	n = 10.648 (Gesamtkohorte)	**9,7 %**	1.030
KI-Trainings-Bots gesperrt (GPTBot etc.)	n = 10.648 (Gesamtkohorte)	**4,3 %**	453
KI-Suchcrawler gesperrt (Perplexity etc.)	n = 10.648 (Gesamtkohorte)	**2,2 %**	239
- Für KI-Suchcrawler ungesperrt	n = 10.648 (Gesamtkohorte)	**97,8 %**	10.409
Pauschalsperre aller Crawler (Disallow: /)	n = 10.648 (Gesamtkohorte)	**0,6 %**	60

2. Internationaler Ländervergleich (Deutschland, UK/EN, USA, Gesamt)

Vergleich der erreichbaren Startseiten über unterschiedliche Top-Level-Domains und Ländercodes

Markt / Segment	Erreichbar (n)	Ø Score	Schema-Markup	PostalAddress	Suchbots offen
Deutschland (DE)	n = 10.648 (63,8 %)	**65,6 / 100**	27,0 % (2.870)	6,5 % (695)	97,8 % (10.409)
UK / Englischsprachig (EN)	n = 839 (59,3 %)	**74,4 / 100**	54,1 % (454)	28,5 % (239)	98,5 % (826)
Vereinigte Staaten (US)	n = 2.114 (29,2 %)	**73,2 / 100**	53,8 % (1.138)	29,7 % (627)	98,6 % (2.085)
International / Rest	n = 109 (51,9 %)	**69,3 / 100**	31,2 % (34)	21,1 % (23)	96,3 % (105)
Gesamt (Alle Märkte)	n = 13.710 (53,6 %)	**67,4 / 100**	32,8 % (4.496)	11,6 % (1.584)	97,9 % (13.425)

Methodischer Rahmen & Prüfverfahren (W3C & RFC Standards):

- Datenbasis & Integrität: Startseiten erreichbarer Unternehmens-Websites aus öffentlichen Verzeichnissen (n = 13.710 weltweit, davon n = 10.648 in Deutschland; Erhebungszeitraum: September - Oktober 2026).
- Strikter Vorabfilter: Websites mit DNS-, TLS- oder HTTP-Verbindungsabbrüchen wurden vorab ausgeschlossen. Alle Prozentangaben beziehen sich auf nachweislich erreichbare, funktionierende Startseiten.
- Angewandte Spezifikationen: Schema.org Vocabulary v26.0 (Organization, LocalBusiness, PostalAddress), W3C JSON-LD 1.1 Recommendation, RFC 9309 (Robots Exclusion Protocol), llms.txt Spezifikation.
- /llms.txt Methodik: Saubere Differenzierung zwischen oberflächlicher technischer Erreichbarkeit (HTTP 200) und inhaltlicher Validität nach Filterung von Soft-404-Fehlerseiten (Markdown-Dokument mit H1-Titel).

3. Regionale Wirtschaftsräume & Metropolregionen im Detail

DATENBLATT FÜR REGIONAL- UND WIRTSCHAFTSREDAKTIONEN (MÜNSTERLAND, OSNABRÜCK, METROPOLEN)

A) Regionale Wirtschaftscluster (DE)

Wirtschaftsraum / Region	Stichprobe (n)	Ø Score	PostalAddress (JSON-LD)	Lücke ohne Adresse
Münsterland (Rheine, MS, ST)	n = 228 (94,6 %)	**68,6 / 100**	12,7 % (29)	**87,3 % (199)**
Osnabrücker Land (Stadt & Kreis)	n = 183 (98,9 %)	**60,4 / 100**	4,4 % (8)	**95,6 % (175)**
Euregio (Bentheim, Emsland)*	n = 46 (100,0 %)	**68,7 / 100**	6,5 % (3)	**93,5 % (43)**

** Hinweis Euregio: Grenzüberschreitende Pilot-Stichprobe (n = 46).*

B) Nationale & Internationale Metropolregionen

Metropolraum	Stichprobe (n)	Ø Score	PostalAddress (JSON-LD)	/llms.txt vorhanden
Metropolregion Berlin	n = 256 (100,0 %)	**71,1 / 100**	21,5 % (55)	13,7 % (35)
Metropolregion München	n = 252 (98,8 %)	**70,4 / 100**	18,3 % (46)	13,1 % (33)
Metropolregion Hamburg	n = 280 (98,9 %)	**69,3 / 100**	10,4 % (29)	7,9 % (22)
Metropolregion Rhein-Ruhr	n = 958 (99,6 %)	**69,1 / 100**	14,0 % (134)	9,2 % (88)
Rhein-Main (Frankfurt)	n = 310 (98,1 %)	**67,9 / 100**	13,5 % (42)	8,1 % (25)
Greater New York (US Metro)	n = 59 (98,3 %)	**82,3 / 100**	69,5 % (41)	30,5 % (18)
San Francisco Bay Area (US)	n = 47 (94,0 %)	**82,2 / 100**	85,1 % (40)	48,9 % (23)
Greater London (UK Metro)	n = 43 (87,8 %)	**88,7 / 100**	97,7 % (42)	39,5 % (17)

*** Hinweis: Metropolwerte weisen bereinigte valide /llms.txt nach Soft-404-Filterung aus.*

Zentrale regionale Befunde & Handlungsempfehlungen:

- Münsterland & Rheine: 94,6 % der untersuchten Domains waren technisch einwandfrei erreichbar, doch 87,3 % verzichten im Quellcode auf strukturierte Schema.org-Standortdaten für Suchmaschinen und KI-Systeme. Maschinenlesbare Adressdaten sind vor allem für Handwerk und Mittelstand im ländlichen Raum ein unterschätzter Hebel.
- Osnabrücker Land: Mit einer Lücke von 95,6 % bei maschinenlesbaren Adressen verpassen viele Unternehmen die Chance, in standortbasierten KI-Suchanfragen (z. B. 'Maschinenbau Osnabrück') eindeutig zugeordnet zu werden.
- Metropolen-Vergleich: Während in US- und UK-Metropolen (San Francisco 85,1 %, London 97,7 %) strukturierte Adress-Entitäten Standard sind, liegt Berlin mit 21,5 % (55 von 256; fehlend bei 78,5 % / 201) deutlich zurück (Fokuskohorten, nicht repräsentativ).

Konkrete Handlungsschritte für Webmaster, KMU und regionale Betriebe:
1. Strukturierte Adressdaten einbetten: Ergänzung von schema.org/PostalAddress in LocalBusiness- oder Organization-JSON-LD.
2. KI-Crawler differenziert steuern: Real-time Retrieval Bots (wie OAI-SearchBot) in robots.txt freigeben.
3. Valide /llms.txt Übersicht bereitstellen: Schlanke Markdown-Datei ohne HTML-Fehlerseite für KI-Agenten.

Pressekontakt, Datenmaterial & Nutzungsrechte:

MIRDAT - Michael Röttger | Friedrich-Ebert-Ring 154 | 48429 Rheine (Münsterland)
E-Mail: info@mirdat.de | Zentrale Quellenseite & Factsheet-Download: https://www.mirdat.de/presse
Zitierweise: "Quelle: MIRDAT Factsheet Website-Studie 2026 (mirdat.de/presse)" | Grafiken & Tabellen frei unter CC BY 4.0