

Ein offener Webindex: Anwendungen, rechtlicher Rahmen, Akzeptanz

Leopold Beer, Paul C. Johannes, Huda Koulani, Christian L. Geminn, Matthias Söllner und Stefan Voigt

Zusammenfassung

Ein frei zugänglicher europäischer offener Webindex (OWI) könnte als Grundlage für eine Vielzahl von Suchmaschinen dienen und das Entstehen neuer webdatenbasierter Anwendungen, einschließlich Systemen der künstlichen Intelligenz (KI), fördern. Damit würde ein solcher offener Index wesentlich zur Vielfalt in digitalen Märkten beitragen und Nutzern die Möglichkeit geben, zwischen verschiedenen Diensten zu wählen. Auf diese Weise würde der Index zum Pluralismus auf dem aktuell monopolisierten Suchmaschinenmarkt beitragen. Der OWI könnte auch für Innovationen im Bereich der KI von entscheidender Bedeutung sein, da er die notwendige Datenbasis für das Training und die Weiterentwicklung von Retrieval Augmented Generation (RAG) und anderen KI-Systemen schaffen würde. Der vorliegende Beitrag befasst sich mit dem rechtlichen Rahmen für einen solchen offenen Webindex und erörtert Herausforderungen. Nach einer Einleitung (1) wird zunächst das Potenzial eines offenen Webindex (2) und seiner Anwendungen dargestellt. Sodann werden die verschiedenen Akteure (3) erörtert und Anwendungsfälle (4) beleuchtet, darunter vertikale Suchmaschinen (4.1), das Training von Large Language Models (LLMs) (4.2), die Verbesserung lokaler Suchdienste (4.3), Business Intelligence und Analytics (4.4) sowie die Aufdeckung von Desinformation und Bot-Netzwerken (4.5). Chancen und Risiken für die Grundrechte und -freiheiten (5) werden besprochen, gefolgt von einem Überblick über die rechtlichen Anforderungen an einen OWI (6) in Bezug auf Datenschutz (6.1), digitale Dienste (6.2), Rechte des geistigen Eigentums (6.3) und künstliche Intelligenz (6.4). Erste Überlegungen zur Untersuchung von Akzeptanz und Vertrauen der Nutzer werden angestellt (7). Abschließend werden die Erkenntnisse zusammengefasst und ein Ausblick in die Zukunft gegeben (8).

1. Einleitung

In der digitalen Welt zu navigieren, ist ohne Suchmaschinen kaum möglich. Ihre Nutzung wird als vermeintlich kostenlose Dienstleistung angeboten, die sich durch die Auswertung von Nutzerdaten nicht nur finanziert, sondern auch eine immense Gewinnschöpfung ermöglicht.¹ Zunehmend werden Suchmaschinen jedoch durch LLM-basierte Systeme (Large Lan-

1 Alphabet erwirtschaftete laut eigenen Angaben (<https://abc.xyz/investor/>) im ersten Quartal 2025 einen Gesamtumsatz von 90,23 Milliarden US-Dollar, wovon 50,7 Milliarden US-Dollar auf den Bereich Google Search entfielen. Der Nettogewinn des Konzerns betrug in dem Zeitraum 34,54 Milliarden US-Dollar. Der Nettogewinn für das Segment Google-Suche ist nicht separat ausgewiesen.

guage Models), die auf künstlicher Intelligenz basieren, ergänzt oder ersetzt. Diese Systeme, die in verschiedene Dienste integriert werden, bieten eine intuitivere Schnittstelle für die Recherche zu verschiedenen Themen. Dennoch greifen auch diese LLM-basierten Systeme letztlich auf die Daten und Strukturen von Suchmaschinen zurück, um ihre Antworten zu generieren und zu optimieren.

Mit dem Ziel, die Funktionalität und die Dienstleistungsqualität sicherzustellen, erfassen und verarbeiten Anbieter vieler Suchmaschinen und LLMs große Mengen personenbezogener Daten, aus denen vielfältige Rückschlüsse auf Interessen und Lebensumstände der Nutzenden möglich sind. Gerade die individualisierte Werbung wird in der Folge zu einer zentralen Einnahmequelle. Hinzu tritt die Möglichkeit zur Monetarisierung der Anzeigereihenfolge von Suchergebnissen (z. B. durch an erster Stelle präsentierte Anzeigen oder eine bezahlte Besserstellung in der Anzeigereihenfolge). Die daraus resultierenden Probleme haben die Google-Gründer Brin und Page bereits Ende der 90er Jahre selbst benannt und vorausgesagt: „[W]e expect that advertising funded search engines will be inherently biased towards the advertisers and away from the needs of the consumers.”² In anderen Worten: „[T]he search engines we have are not neutral or benevolent. They determine what data to collect from us and how to use it in algorithms. They are not open and transparent. So, it is hard to spot any bias or manipulation. They profile us. They divide us. This is not the search we want. We need search algorithms tuned for humanity, not individual greed.”³

Alle Suchmaschinen basieren auf einem sogenannten Webindex, den man sich wie die Kartei einer Bibliothek vorstellen kann, nur eben um ein Vielfaches größer und viel dynamischer. Das Internet muss ständig nach neuen oder geänderten Inhalten durchsucht werden, damit diese über eine Suchmaschine abgerufen werden können. Dies erfordert eine hohe Rechenleistung und große Speicherkapazitäten, was die Erstellung und den Betrieb eines Webindex komplex und kostspielig macht. Neue Suchmaschinen haben es daher schwer, sich zu etablieren. Zwar gibt es alternative Suchanbieter wie Ecosia oder DuckDuckGo, diese benötigen aber in der Regel Zugriff auf einen oder sogar mehrere bestehende Webindizes und

2 Brin/Page, Anatomy of Large-Scale Hypertextual Web Search Engine, Computernetzwerke und ISDN-Systeme 1998, S. 107.

3 Smith, 2023.

entsprechenden Rankings. Dabei müssen sie sich den Bedingungen des Betreibers unterwerfen.⁴

Bislang gibt es weltweit nur vier umfassende Indizes des Internets: Google, Microsoft (beide USA), Baidu (China) und Yandex (Russland). Diese Anbieter diktieren die Bedingungen für den Zugang zu ihren Indizes und nehmen damit die Rolle von Torwächtern (Gatekeeper) auf dem Suchmaschinenmarkt ein. Das schränkt die Wahlmöglichkeiten der Verbraucher ein. So können Nutzende beispielsweise keine besonders datenschutzfreundlichere Suchmaschine wählen. Datenschutz und andere Belange der Privatsphäre, denen in der Europäischen Union (EU) besondere Bedeutung beigemessen wird, dürften bei den Gestaltungsentscheidungen für bestehende Indizes aufgrund des Sitzes der jeweiligen Unternehmen eine untergeordnete Rolle spielen. Durch die jetzige Ausgestaltung der vorherrschenden Suchindizes und das Fehlen von datenschutzfreundlichen Wettbewerbern kann das Grundrecht auf Privatsphäre und Datenschutz von den Nutzenden der Suchmaschinen kaum effektiv wahrgenommen werden. Das umfangreiche Sammeln und Auswerten von Nutzerdaten durch Suchmaschinen⁵ erfüllt das Bedürfnis nach Privatsphäre⁶ der Bevölkerung keinesfalls.

Der Marktanteil von Googles Suchmaschine in Deutschland liegt derzeit bei rund 90 %;⁷ der weltweite Marktanteil auf allen Geräten (Desktop und mobil) beträgt 89,6 %.⁸ Einer der Hauptgründe für diese Dominanz ist der über Jahrzehnte aufgebaute Webindex von Google. Ein Blick auf den Marktanteil von Microsofts Suchmaschine Bing (weltweit: 4,0 %⁹) zeigt jedoch, dass ein hochwertiger Webindex nicht der einzige Faktor für die Dominanz von Google in diesem Bereich sein kann. Vielmehr spielen auch Gewöhnungseffekte, Vertrauen,¹⁰ Benutzerfreundlichkeit, andere angebundene Dienste, Voreinstellungen, Hindernisse beim Wechsel auf ande-

4 Z.B. basiert die deutsche Suchmaschine Ecosia auf dem Index von Bing.

5 Ein Blick in Google's Datenschutzerklärung (abrufbar unter <https://policies.google.com/privacy?hl=de#infocollect>) zeigt die umfangreiche Erhebung personenbezogener Daten durch das Unternehmen, etwa von Sprach- und Audiodaten oder Standortdaten.

6 Braun/Trepte, Privatheit und informationelle Selbstbestimmung, 2017, S. 6.

7 StatCounter, Marktanteile der führenden Suchmaschinen in Deutschland, 2024.

8 StatCounter, Marktanteil der führenden Suchmaschinen weltweit, 2025.

9 StatCounter, Marktanteil der führenden Suchmaschinen weltweit, 2025.

10 Schultheiß/Lewandowski, Misplaced trust?, Journal of Information Science 2021.

re Suchdienste und die subjektiv empfundene Qualität der Dienste eine wichtige Rolle.¹¹

Ein offener, frei zugänglicher und großmaßstäbiger Webindex könnte dazu beitragen diese Probleme lösen, indem er anderen Suchmaschinenbetreibern als Grundlage für die Bereitstellung ihrer Dienste dient und so das Entstehen neuer, nutzer- und datenschutzfreundlicherer Suchmaschinen fördert.¹² So könnte ein offener Index einen wichtigen Beitrag zur Vielfalt und Wahlfreiheit im Bereich der Internetsuche leisten. Crawling und Indexierung würden vom eigentlichen Betrieb der Suchmaschine getrennt werden. Kleine und aufstrebende Anbieter können sich den Aufbau und Betrieb eines eigenen Indexes meist nicht leisten; ein offener Webindex würde ihnen langfristig eine Perspektive bieten. Der offener Webindex könnte auch für Wissenschaft und Forschung zugänglich gemacht werden. Darüber hinaus wäre ein solcher Index in hohem Maße innovationsfördernd für den Bereich der Künstlichen Intelligenz, da er die notwendige Datenbasis für das Lernen, Trainieren und Weiterentwickeln von Retrieval Augmented Generation (RAG) und anderen KI-Systemen schaffen würde. Insbesondere die aktuelle Entwicklung großer KI-basierter Sprachmodelle (z.B. ChatGPT von OpenAI/Microsoft oder Gemini von Google) unterstreicht die hohe Relevanz umfassender und qualitativ hochwertiger vorgefilterter Web-Datenbanken, wie sie ein offener Webindex bieten würde. Solche Datenbanken sind sowohl für die digitale Wirtschaft als Ganzes als auch für den Suchmaschinenmarkt der Zukunft im Besonderen von größter Bedeutung.¹³

Möglich wäre das alles nur durch eine kollaborative Anstrengung, bei der viele Institutionen bei Crawling und Indexing sowie bei der Bereitstellung und Pflege des Index zusammenwirken. Sogar *Brin* und *Page* betonten anfangs die Vorteile eines kollektiven Zusammenwirkens, bevor sie mit Google Search eine zentralisierte und profit- und werbegetriebene Lösung etablierten. Ein offener Web-Index könnte die Websuche revolutionieren: „[...] presenting the results from querying that index could be done completely unfiltered – simply clustered by source such as researched and

11 *Liaw/Huang*, An investigation of user attitudes toward search engines, *Computers in Human Behavior* 2003, 751.

12 *Lewandowski*, in: König/Rasch (Hrsg.), *Reflections on Web Search*, 2014, S. 49; *Lewandowski*, *The Web is missing an essential part of infrastructure*, *Communications of the ACM* 2019, 24.

13 *Wilson*, *Microsoft Bing vs. Google Bard*, 2023.

reviewed, mainstream media, social media trends, independent content. And you could have the controls to dive in and explore and filter by any criteria you chose – rather than being blinkered to a subset prepared to match you. And the rating from various fact checker organizations could be used to red-flag content. The validated would no longer get submerged under the popular. Fact would be distinct from fiction. That would be revolutionary.”¹⁴ Ein wirkmächtiges Instrument der informationellen und kommunikativen Selbstbestimmung könnte entstehen, das zugleich dem strategischen Ziel der Stärkung von Demokratie und Gesellschaft dient. Eine offene Suchinfrastruktur ermöglicht es, den gerade in Suchmaschinen präsenten Möglichkeiten der Kontrolle und Überwachung zu begegnen.

Die Etablierung eines offenen, europäischen Web-Index würde zudem einen wesentlichen Beitrag leisten, Standortnachteile für Deutschland und Europa abzubauen und wäre ein wichtiger Schritt hin zur Realisierung von digitaler Souveränität. Zugleich würde die dadurch ermöglichte Pluralisierung des Marktes für Suchmaschinen Privatheit und Datenschutz stärken, sofern sichergestellt würde, dass die Gestaltung der entstehenden Suchinfrastruktur an der bestmöglichen Verwirklichung von Grundrechten und -prinzipien sowie des geltenden Datenschutzrechts ausgerichtet würde. Diese Pluralisierung würde zudem die individuelle Handlungsfreiheit der BürgerInnen fördern und die Qualität der Internetsuche insgesamt verbessern: „When geographic map data was opened, competition ensured we got the better road navigation systems we all love in our cars. So similarly now, we need to open the web map data, the search indexes, so that competition can get us better web navigation systems.“¹⁵ Diese erwünschten positiven Auswirkungen auf Privatheit und Selbstbestimmung können sich aber nur entfalten, wenn sie eine Unterstützung erfahren, die sowohl eine Ausrichtung an Grundrechten und am einfachen Datenschutzrecht als auch Fragen der Nutzerakzeptanz neuer Technologien betrifft. Insgesamt wäre ein offener Web-Index ein wichtiger Schritt, um im zentralen Bereich der Internetsuche Informationelle Selbstbestimmung als langfristigen Wettbewerbsvorteil und Katalysator für Innovationen zu stärken.

Die Aussagekraft von im Kontext der Internetsuche anfallenden Daten und die daraus resultierenden Bedrohungen von Privatheit und Selbstbestimmung soll ein Beispiel illustrieren:¹⁶ Im Jahr 2006 veröffentlichte die

14 Smith, 2023.

15 Smith, 2023.

16 Beispiel entnommen aus *Geminn, Deus ex machina?*, 2023, 25.

Forschungsabteilung des Internetanbieters AOL eine Textdatei, die Suchanfragen der unternehmenseigenen Suchmaschine über die AOL Client Software enthielt. Enthalten waren die Anfragen von 657.000 U.S.-Amerikanern über einen Zeitraum von drei Monaten (März bis Mai 2006), wobei deren Identität durch zufällig generierte Zahlen ersetzt wurde. Die Suchanfragen betrafen die Interessen und Sorgen der Nutzer, und dabei auch dunkle Facetten der Persönlichkeit; so fanden sich auch zahlreiche Anfragen zu Suizidmethoden oder etwa auch solche wie „how to kill your wife“. Die New York Times deckte in der Folge anhand der Suchanfragen die Identität von Nutzerin Nr. 4417749 auf; es handelte sich um die damals 62-jährige Thelma Arnold aus Gilburn, Georgia. Die Identifizierung war in diesem Fall besonders einfach, denn die Nutzerin hatte zahlreiche Suchanfragen gestellt, die den Nachnamen Arnold oder ihren Wohnort enthielten. In einem Interview mit der Zeitung fand Frau Arnold deutliche Worte: „We all have a right to privacy. Nobody should have found this all out.“¹⁷ Das Bedürfnis nach Privatheit und Selbstbestimmung wird von den aktuell dominierenden Suchmaschinen indes nicht erfüllt, sondern ganz im Gegenteil erheblich bedroht. Eine offene Suchinfrastruktur ermöglicht es indes, den gerade in Suchmaschinen präsenten Möglichkeiten der Kontrolle und Überwachung zu begegnen.

Einen ersten praktischen Schritt hin zu einem offenen Webindex geht das Horizon Europe-Projekt OpenWebSearch.EU (2022-2026). In diesem Projekt wird von einem technisch-wissenschaftlich orientierten Konsortium aus 14 europäischen Großrechenzentren, Wissenschaftszentren, Universitäten sowie der Open Search Foundation eine offene und verteilte Suchinfrastruktur für das Internet konzipiert und prototypisch implementiert. Dabei werden sowohl signifikante Webdatencrawls als auch die weiterführende Webdatenanalyse und Webindexerstellung für eine Nutzung in Suchdiensten für Europa durchgeführt. Das vom Bundesministerium für Bildung und Forschung (BMBF) geförderte Projekt „Privatsphäre fördernde Digitale Infrastrukturen – PriDI“ unterstützt die Erstellung eines offenen Webindex durch rechtswissenschaftliche und wirtschaftsinformatische Expertise.¹⁸

17 *Barbaro/Zeller Jr.*, A Face Is Exposed for AOL Searcher No. 4417749, The New York Times v. 9. August 2006.

18 Beitrag basiert auf und erweitert Whitepaper Geminn u.a., Anwendungsfälle eines offenen Webindex, 2025; abrufbar unter <https://pridi-projekt.de>.

2 Der offene Webindex

Ähnlich wie die Indizes von Google oder Bing wird der OWI durch systematisches Crawlen des Webs, die Analyse der gecrawlten Inhalte und deren Speicherung mit Metadaten in einer Datenbank erstellt.¹⁹ Der OWI soll die digitale Souveränität der EU stärken, indem er die Abhängigkeit von außereuropäischen Monopolisten im Suchmaschinenmarkt durch einen nachhaltigen, frei zugänglichen Webindex verringert.

Entwickelt wurde eine verteilte Crawling-, Indexierungs- und Hosting-Architektur für den OWI. Diese besteht in der Kombination eines Frontier-Crawlers, der im Wesentlichen das Web entlang eingebetteter Links kartiert und URLs sammelt, mit verteilten Worker-Crawlern, die später die Websites abrufen und deren Inhalte in sogenannten Web-Archivdateien (WARC) speichern. Später werden die "rohen" Webdaten weiterverarbeitet, bereinigt, gefiltert, mit Metadaten angereichert, nach Sprache und Genre klassifiziert und als Webindex-Charts im Common Index File Format (CIFF) gespeichert.

Das System ist so konzipiert, dass es Speicher- und Rechenkapazitäten in mehreren Hochleistungsrechenzentren in ganz Europa bündelt und dynamisch erweitert werden kann, indem weitere Rechenzentren in den Verbund aufgenommen werden. Um auf den Index zuzugreifen, können sich Suchmaschinenanbieter oder andere wissenschaftliche Nutzer des OWI über ein öffentliches System authentifizieren und über die Befehlszeile auf Teile des Index zugreifen und diese abrufen. Derzeit werden die Webdaten unter einer Forschungslizenz zur Verfügung gestellt. Das Forschungsteam wird den Zugang zum System jedoch auch für kommerzielle Akteure ermöglichen.

Die öffentliche Zugänglichkeit des Index soll die Freiheit der Internetsuche stärken und eine Grundlage für Innovationen in Wissenschaft und Wirtschaft bilden. Die Forscher haben inzwischen rund 2,23 Milliarden URLs in 185 verschiedenen Sprachen gecrawlt. Der offene Webindex hat derzeit ein Volumen von rund 14 TB und steht interessierten Entwicklern für Tests zur Verfügung. Der Index von Google ist mit einem Volumen von rund 100 000 TB allerdings deutlich größer.²⁰

19 Weitere Einzelheiten über die Erstellung und den Betrieb des OWI in Hendriksen u.a., Open Web Index, *Advances in Information Retrieval* 2024.

20 Google, *How Google Search organizes information*, 2025.

3 Akteure

Bei der Bewertung des OWI aus rechtlicher Sicht und mit Blick auf die Nutzerakzeptanz wird im Folgenden zwischen den verschiedenen Interessengruppen unterschieden. Zum einen sind das die *Betroffenen*. Diese Rolle beschreibt Personen oder Unternehmen, deren persönliche Daten (betroffene Personen) oder geistiges Eigentum (Rechteinhaber) im OWI gespeichert sind und von Tools oder Suchmaschinen, die auf dem OWI basieren, genutzt oder gefunden werden können. Der Index selbst wird vom *OWI-Entwickler* entwickelt und gepflegt. Der OWI-Entwickler hat sich zu einer unabhängigen juristischen Person, dem Betreiberkonsortium, zusammengeschlossen. Die Abrufer oder Konsumenten der im OWI enthaltenen Daten und Informationen werden als *Anwendungsentwickler* bezeichnet. Es handelt sich um Personen oder Systeme, die Webdaten aus dem Index abrufen, um auf Grundlage der abgerufenen Daten verschiedene Anwendungen und Modelle zu erstellen und zu entwickeln.²¹ Dabei kann es sich um Einzelpersonen, Unternehmen, öffentliche Einrichtungen oder Wissenschaftler:innen handeln, die die offenen Daten des OWI nutzen, um ihre eigenen Anwendungen und Dienste zu entwickeln. Schließlich kommen auch die *Endnutzer* mit dem Index in Berührung. Sie sind die natürlichen oder juristischen Personen, die von den Anwendungsentwicklern entwickelte Werkzeuge und Systeme nutzen.

4 Anwendungsfälle

Der offene Web-Index lässt sich in einer Vielzahl von Anwendungsbereichen²² nutzen, die im nachfolgenden überblicksartig dargestellt werden.²³

21 Z.B. im Rahmen von Trainingsprozessen für KI-Anwendungen, siehe 4.2 und 4.4.

22 Siehe auch Studie Nowakowski u.a., Market potential assessment, 2024.

23 Aufstellung basiert auf Geminn u.a., Anwendungsfälle eines offenen Webindex, 2025, welche hier in Teilen wiedergegeben wird.

4.1 Entwicklung vertikaler Suchmaschinen

Der OWI kann als Grundlage für die Entwicklung neuer (vertikaler) Suchmaschinen²⁴ dienen. Es könnten etwa Suchmaschinen für bestimmte Nutzergruppen, Lokalisationen oder Themen entstehen. Der OWI schafft die Grundlage dafür, indem er eine umfangreiche Datenbasis zur Verfügung stellt. Auf diese können Anwendungsentwickler von spezialisierten Suchmaschinen zugreifen und genau die Daten herausfiltern, die für den jeweiligen Zweck relevant sind. So können auch kleinere Organisationen und Unternehmen Suchmaschinen entwickeln und die Vielfalt im Suchmaschinenmarkt erhöhen.²⁵ Denn sie müssen keine enormen Ressourcen für das Crawlen und Indexieren des gesamten oder eines Teils des Internets aufbringen, sondern können auf Grundlage des OWI präzise und relevante Ergebnisse für spezifische Themengebiete liefern. Da der Quellcode des OWI öffentlich einsehbar ist, können die Anwendungsentwickler zudem die zugrundeliegenden Algorithmen verstehen und auf ihre spezifischen Bedürfnisse anpassen. Mithilfe des offenen Webindex können neue Suchmaschinen für Nachrichten entstehen, die Privatpersonen Nachrichten über sowohl aktuelle als auch vergangene Ereignisse zur Verfügung stellen. Darüber hinaus können Pressespiegel zu verschiedenen Themen, Ereignissen oder auch Unternehmen zusammengestellt werden.

4.2 Training von Large Language Models (LLMs)

Für das Training von Large Language Models (LLMs)²⁶ sind umfassende und qualitativ sehr gut vorgefilterte Webdatenbestände von hoher Relevanz. Ein Beispiel für ein LLM ist GPT-4 von OpenAI, worauf das Sprachmodell ChatGPT basiert. Solche Modelle werden mit großen Mengen an Textdaten trainiert, die überwiegend aus online verfügbaren Quellen stammen. Mithilfe von maschinellem Lernen unter Nutzung von neuronalen Netzwerken und Deep Learning-Methoden lernen LLMs, statistische Zusammenhänge zwischen Wörtern und Sätzen zu erkennen, um auf der Grundlage statistischer Häufigkeiten Sprach-Artefakte zu vervollständigen oder zu ge-

24 Curran/McGlinchey, Vertical search engines, The ITB Journal 2003, S. 22.

25 Vgl. Geminn u.a., Anwendungsfälle eines offenen Webindex, 2025, S. 11f.

26 Dakhel u.a., in: Nguyen-Duc u.a. (Hrsg.), Generative AI for Effective Software Development, 2024, S. 3f.

nerieren. Der OWI trägt dazu bei, neue LLMs zu entwickeln, indem er auch kleineren Unternehmen kostengünstig eine hinreichende Menge an Trainingsdaten zur Verfügung stellt. Dadurch weisen die Produkte solcher Start-Ups im Vergleich zu den derzeit auf dem Markt angebotenen Sprachmodellen eine konkurrenzfähige Qualität auf.²⁷

RAG, oder *Retrieval Augmented Generation*, ist ein Ansatz in der KI, bei dem Textgenerierungsmodelle um eine Komponente zur Informationsbeschaffung erweitert werden.²⁸ Dabei kombiniert RAG die Fähigkeit eines Sprachmodells, Antworten zu generieren, mit einer Retrieval-Komponente, die relevante Informationen aus externen Wissensquellen wie Datenbanken oder Dokumenten abrufen. Die erste Komponente von RAG, das Sprachmodell, wird auf großen Textmengen trainiert, um Sprache zu verstehen und Texte zu generieren. Es bildet das Grundgerüst für die Generierung von Antworten. Die zweite Komponente, das Retrieval, greift auf eine Wissensbasis oder eine Dokumentensammlung zu, um bei jeder Anfrage relevante Informationen abzurufen. Diese Kombination ermöglicht präzisere und aktuellere Antworten, da das Modell nicht nur auf statisches Wissen beschränkt ist, sondern gezielt aktuelle Informationen in die Generierung integrieren kann. Anwendungsfelder für RAG sind unter anderem Kundensupport, medizinische Beratung und Wissensmanagement, bei denen es auf spezifische und aktuelle Antworten ankommt.

Mithilfe der Daten im OWI könnten auch andere webdatenbasierte KI-Anwendungen trainiert werden, etwa Systeme des *Knowledge Representation and Reasoning* (KRR).²⁹ Hier geht es im Gegensatz zu LLMs, die mit Statistik Texte produzieren, darum, Informationen so darzustellen, dass sie ein Computer verarbeiten und gleich einem Menschen komplexe Probleme lösen kann. Ziel sind „intelligente“ Maschinen, die vom menschlichen Wissen lernen und gleich diesem handeln. KRR-Systeme werden beispielsweise im Qualitätsmanagement zur Überwachung der Produktqualität eingesetzt oder zur Betrugsprävention im Versicherungswesen.

27 Vgl. Geminn u.a., Anwendungsfälle eines offenen Webindex, 2025, S. 13f.

28 Arslan u.a., A survey on RAG with LLMs, *Procedia Computer Science* 2024, S. 3781.

29 Sharma/Garg, *Artificial Intelligence*, 2021.

4.3 Verbesserung lokaler Suchdienste

Google verknüpft via Google Maps Nutzer mit Unternehmen. Im Sinne dieses Dienstes könnte der OWI dazu genutzt werden neue Kartendienste zu verbessern. Denn die Webdaten aus dem Index enthalten katalogisiert Informationen zu Städtenamen, Hotels, Restaurants, Geschäften und Sehenswürdigkeiten. Diese Daten können mit den jeweiligen Orten auf einer Karte verknüpft werden. Zudem ist es möglich, weitere Hintergrunddaten aus dem OWI automatisiert zu den jeweiligen Orten hinzuzufügen, wie etwa die Historie von Städten oder Preise von Hotels, Restaurants und Geschäften. Auf diese Weise sind Anbieter von Kartendiensten nicht mehr auf eine umfangreiche Mitarbeit von Nutzern angewiesen, sondern können ihre Inhalte auf Grundlage der Webdaten initial aufstellen und laufend aktualisieren. Dies ist ein Kostenvorteil, der dazu führen könnte, dass neue Anbieter von Kartendiensten auf den Markt eintreten.³⁰

Zudem ist es mithilfe des OWI möglich, das lokale Web für Bewohner oder Touristen in einer bestimmten Region in eine separate Suchmaschine zu überführen und so eine lokale Suche zu ermöglichen. In den Suchergebnissen werden dann automatisch nur die Inhalte lokaler Anbieter angezeigt. Der Nutzer findet so zielgerichtet die Informationen, die aufgrund seiner Lokalität für ihn relevant sind. Es ergäbe sich der Effekt einer Art Gemeindeplattform, in der sich die Bürger untereinander austauschen können und lokale Ökosysteme gestärkt werden. Das würde auch dazu führen, dass lokale Unternehmen gestärkt und ihre Angebote vor Ort sichtbarer gemacht werden. Auf diese Weise würden durch Suchmaschinen die günstigsten und besten lokalen Angebote angezeigt und nicht diejenigen, die mit dem Suchmaschinenanbieter verbunden sind.³¹

4.4 Business Intelligence und Analytics

Business Intelligence (BI) und Analytics spielen eine zentrale Rolle in modernen Unternehmen, da sie datenbasierte Entscheidungsprozesse ermöglichen.³² BI umfasst die Sammlung, Verarbeitung und Darstellung von

30 Vgl. *Geminn u.a.*, Anwendungsfälle eines offenen Webindex, 2025, S. 17.

31 Siehe auch das Urteil gegen Google wegen unlauterer Selbstreferenzierung in Google Shopping: EuGH, Urteil vom 10. September 2024 – Ref. C-48/22 P.

32 *Baars/Kemper*, Business Intelligence & Analytics, 2021.

Geschäftsdaten, um historische Trends und aktuelle Leistungskennzahlen zu visualisieren. Analytics geht einen Schritt weiter und nutzt fortschrittliche statistische Methoden sowie maschinelles Lernen, um tiefergehende Einblicke zu gewinnen und Vorhersagen zu treffen. Unternehmen können dadurch nicht nur Entwicklungen aus der Vergangenheit analysieren, sondern auch zukünftige Geschäftsstrategien optimieren. Diese datengetriebenen Ansätze führen zu verbesserten Geschäftsentscheidungen, effizienteren Prozessen und folglich zu Wettbewerbsvorteilen in einem zunehmend datenintensiven Markt. Die Ansätze von Business Intelligence und Analytics erfordern operative Daten, Kundendaten, Finanzdaten, Web- und Social-Media-Daten, logistische Daten, Marktdaten, externe Daten sowie Maschinendaten, um fundierte Analysen durchzuführen und datengetriebene Entscheidungen zu unterstützen.³³ Es lassen sich Produkte entwickeln, die die Ansätze aus Business Intelligence und Analytics implementieren und geeignete Webdaten aus den OWI-Subindizes für diese Berechnungen heranziehen. Nicht alle genannten Datenarten wären im OWI verfügbar, da zum Beispiel Finanz- und Kundendaten von Unternehmen nicht im Internet zur Verfügung stehen. Daher können Anwendungsentwickler relevante Daten aus den verschiedenen Subindizes und dem OWI ziehen, um den individuell geeigneten Datensatz für die Berechnungen zusammenstellen.

4.5 Aufdeckung von Desinformation und Bot-Netzwerken

Der offene Webindex kann dabei helfen, Desinformation³⁴ und Bot-Netzwerke aufzudecken. Auf Grundlage des umfangreichen Datenpools können plattformübergreifende Lösungen entwickelt werden, um Social Media Accounts zu erkennen und zu verfolgen, die allein oder in Netzwerken Desinformation in der digitalen Welt verbreiten. Das gilt ebenso für Accounts, die künstlich ein bestimmtes (politisches) Narrativ verbreiten wollen, was sich zuletzt im Zusammenhang mit dem Angriff Russlands auf die Ukraine als Herausforderung dargestellt hat. Systematische Verzerrungen der öffentlichen Meinungsbildung können dann entsprechend geflaggt oder aus dem jeweiligen Netzwerk herausgenommen werden. Der OWI wäre für die Entwicklung eines Tools zur Aufdeckung von Desinformation und Bot-Netzwerken hilfreich, da mithilfe des OWI-Inhalte über verschiedenen

33 Vgl. *Geminn u.a.*, Anwendungsfälle eines offenen Webindex, 2025, S. 20.

34 *Shams u.a.*, Web search engine misinformation notifier extension, Healthcare 2021.

Plattformen und Webseiten verfolgt werden können, um auf diese Weise den Ursprung der Nachrichten verifizieren zu können.³⁵

5 Chancen und Risiken für Grundrechte und -freiheiten

Gerade in Rechtsgebieten, die einem schnellen Wandel unterliegen, birgt die Betrachtung grundrechtlicher Anforderungen gewisse Vorteile im Vergleich zu einer Analyse einfachgesetzlicher Anforderungen. Grundrechte und -freiheiten dienen als beständige, allgemein gültige Leitplanken. Durch Erstellung und Kuratierung eines offenen Web-Index sind potenziell eine Vielzahl von Grundrechten und Grundfreiheiten aus dem Grundgesetz (GG) und der EU-Grundrechte-Charta (GRCh) betroffen:

Art. 2 Abs. 1 iVm. Art. 1 GG beinhaltet das Recht auf informationelle Selbstbestimmung, welches durch das Crawling und Indexieren persönlicher Daten betroffen sein kann. So besteht durch den Index insbesondere das Risiko einer zweckunbestimmten Datenvorratshaltung.

Eine solche Datenvorratshaltung stünde zunächst im Konflikt mit dem Recht auf Achtung der Kommunikation aus Art. 7 GRCh und dem Recht auf Schutz personenbezogener Daten aus Art. 8 GRCh. Gleichzeitig kann durch den offenen Webindex allerdings auch Transparenz über die Verarbeitung personenbezogener Daten im Netz geschaffen und den betroffenen Personen eine Kontrollmöglichkeit über die Verbreitung solcher Informationen gewährt werden.

Art. 15 GRCh und Art. 11 GG schützen die Berufsfreiheit. Das Crawling und die Indexierung von Daten kann dazu führen, dass Daten gegen den Willen oder in Widerspruch mit den Interessen des Urhebers oder Rechteinhabers genutzt werden und so Geschäftsmodelle gestört werden. Auch die in Art. 16 GRCh geschützte unternehmerischen Freiheit kann dadurch betroffen sein. Es ergeben sich für die berufsbezogenen Grundrechte und -freiheiten jedoch auch erhebliche Chancen durch das Aufbrechen bestehender Quasi-Monopole.

Das Eigentum wird in Art. 17 GRCh und Art. 14 GG geschützt. Insbesondere das sogenannte geistige Eigentum ist von einem Kontrollverlust durch den offenen Webindex bedroht, etwa wenn Werke gegen den Willen des Rechteinhabers in den Index aufgenommen und verbreitet werden. Andererseits kann der Index auch dazu beitragen, urheberrechtlich geschützte

³⁵ Vgl. *Geminn u.a.*, Anwendungsfälle eines offenen Webindex, 2025, S. 22f.

Werke weiter zu verbreiten und besser auffindbar zu machen, sodass kreative Urheber mehr Kunden erreichen können und die Verwertungsrechte an wirtschaftlichem Wert gewinnen. Der Index könnte den Rechteinhabern auch helfen, überhaupt unrechtmäßige Verwendungen ihres geistigen Eigentums online aufzudecken.

Durch die Zugänglichmachung von Webinhalten durch den Index können auch das Recht auf freie Meinungsäußerung, die Informationsfreiheit (Art. 11 GrCh und Art. 5 Abs. 1 GG) sowie die Freiheit von Kunst und Wissenschaft (Art. 13 GRCh und Art. 5 Abs. 3 GG) betroffen sein. Denn der Index birgt grundsätzlich die Gefahr, unwahre Informationen zu verbreiten oder Meinungen zu manipulieren. Zugleich kann ein offener Webindex den Zugang zu Informationen im Netz erleichtern sowie vielfältige Meinungsäußerungen und wissenschaftliche Publikationen einem breiten Publikum zugänglich machen.

Art. 36 und Art. 22 GRCh beinhalten das Recht auf Zugang zu Dienstleistungen von allgemeinem wirtschaftlichem Interesse sowie den Schutz der Vielfalt der Sprachen. Abhängig davon, welche Inhalte in den Index aufgenommen werden und für wen der offene Webindex zugänglich ist, können auch diese Grundfreiheiten betroffen sein.

Zuletzt wird der Webindex offen ausgestaltet sein und derzeit durch europäische Akteure finanziert und umgesetzt. Vor diesem Hintergrund sind Risiken von zu großer oder missbräuchlicher politischer Einflussnahme mit den Art. 20 und 23 GG in Einklang zu bringen.

6 Einfachgesetzlicher Rechtsrahmen

Die Komplexität des OWI wirft viele rechtliche Fragen auf, so dass der Rechtsrahmen nicht einfach zu bestimmen ist. Das europäische Recht für Daten und Online-Dienste hat sich in den letzten Jahren stark verändert. Während zunächst vor allem die Datenschutz-Grundverordnung (DSGVO, Verordnung (EU) 2016/679) den Umgang mit personenbezogenen Daten detailliert regelte, ist inzwischen ein Netz von mehr oder weniger spezialisierten, unmittelbar geltenden Rechtsakten entstanden. Der Digital Services Act (DSA, Verordnung (EU) 2022/2065), der Data Governance Act (DGA, Verordnung (EU) 2022/868), der Data Act (DA, Verordnung (EU) 2022/868), der Digital Markets Act (DMA, Verordnung (EU) 2022/1925) und die KI-Verordnung (KI-VO, Verordnung (EU) 2024/1689) bilden

einen weiten rechtlichen Rahmen für digitale Dienstleistungen und Geschäftsmodelle, die auch einen OWI und Anwendungsentwickler betreffen können.³⁶ Diese Verordnungen gelten unmittelbar und einheitlich in allen Mitgliedsstaaten.

Gleichzeitig gibt es in der Europäischen Union einen harmonisierten Rahmen für das Urheberrecht. Er wird in erster Linie durch eine Kombination aus EU-Richtlinien, internationalen Verträgen und nationalen Gesetzen der Mitgliedstaaten geregelt. Obwohl die Union bestrebt ist, die Urheberrechtsgesetze in ihren Mitgliedstaaten zu harmonisieren, gibt es immer noch Unterschiede auf nationaler Ebene.

Im Folgenden werden das Datenschutzrecht (6.1), der Digital Services Act (6.2), das Urheberrecht (6.3) und die KI-Verordnung (6.4) beleuchtet.

6.1 Datenschutzrecht

Der OWI wird, absichtlich oder unabsichtlich, personenbezogene Daten von natürlichen Personen (betroffene Personen) im Sinne von Art. 4 Nr. 1 DSGVO in den gecrawlten Daten enthalten.³⁷ Es kann sich dabei sogar um Daten besonderer Kategorien im Sinne von Art. 9 Abs. 1 DSGVO handeln. Personenbezogene Daten können auch Teil der im OWI gesammelten und angereicherten Metadaten sein (z. B. Eigentümer von Webseiten, Kontaktangaben). Die Tätigkeit von Dritten ins Internet gestellte oder dort veröffentlichte Informationen zu finden, automatisch zu indexieren und (ggf. auch nur vorübergehend) zu speichern und schließlich Anderen zur Verfügung zu stellen ist insofern eine Verarbeitung personenbezogener Daten iSv Art. 4 Nr. 1 und 2 der DSGVO.³⁸

Die Institution, welche den OWI betreibt, ist in Bezug auf diese personenbezogenen Daten verantwortlich im Sinne des Art. 4 Nr. 7 DSGVO.³⁹ Wird der OWI von einem Joint-Venture verschiedener Organisationen oder

36 Dazu *Geminn/Johannes*, Handbuch europäisches Datenrecht, 2025.

37 *Geminn u.a.*, Legal Challenges of an Open Web Index, *International Cybersecurity Law Review* 2021, 183; *Geminn*, Rechtsfragen eines offenen Web Index, *Multimedia und Recht* 2021, 18.

38 EuGH, Urteil vom 8.12.2022 – Az. C-460/20 (TU, RE/Google LLC), Rn. 49 – ECLI:EU:C:2022:962.

39 EuGH, Urteil vom 13.5.2014 – Az. C-131/12 (Google Spain), Rn. 41 – ECLI:EU:C:2014:317; EuGH, Urteil vom 27.9.2019 – Az. C-136/17 (Google vs. CNIL), Rn. 35 – ECLI:EU:C:2019:773.

öffentlicher Einrichtungen angeboten, sind die beteiligten Stellen gemeinsam für die Verarbeitung verantwortlich.

Die Verarbeitung personenbezogener Daten ist nur rechtmäßig, wenn sie auf eine entsprechende Rechtsgrundlage im Sinne von Art. 6 DSGVO gestützt werden kann. Die Erstellung und der Betrieb eines unabhängigen, frei zugänglichen Webindex durch öffentlich-rechtliche Einrichtungen (oder in deren Auftrag) liegt im öffentlichen Interesse, z. B. um die Abhängigkeit von Unionsbürgern von ausländischen Suchmaschinen zu verringern. Für die Rechtmäßigkeit der Verarbeitung im öffentlichen Interesse gemäß Art. 6 Abs. 1 UAbs. 1 lit. e DSGVO bedarf es einer Rechtsgrundlage im Unionsrecht oder im Recht der Mitgliedstaaten gemäß Abs. 3. Eine solche Grundlage ist bislang nicht ersichtlich, sodass die öffentliche(n) Stelle(n), die den OWI entwickeln, auf eine allgemeine Ermächtigung im mitgliedstaatlichen Recht zurückgreifen müssten. Beispielsweise dürfte ein Rechenzentrum, welches als öffentliche Stelle in Bayern organisiert ist, personenbezogene Daten für einen OWI, z. B. das Crawlen, Kopieren und Sortieren von Webseiten und deren Indexierung, aufgrund seiner eigenen Zweckbestimmung (d. h. Satzung und/oder Errichtungsgesetz) in Verbindung mit Art. 4 Bayerisches Datenschutzgesetz (BayDSG) auf der Grundlage von und in Verbindung mit Art. 6 Abs. 1 Nr. 1. lit. e und Abs. 2 und 3 DSGVO verarbeiten.

Bei privatwirtschaftlichen Einrichtungen können die Erstellung und der Betrieb des Open Web Index (OWI) nur auf die Wahrung der berechtigten Interessen des für die Verarbeitung Verantwortlichen oder eines Dritten gemäß Art. 6 Abs. 1 UAbs. 1 lit. f DSGVO gestützt werden. Dies erfordert eine sorgfältige Abwägung der Interessen des OWI-Betreibers mit den Grundrechten der betroffenen Personen.

Das berechtigte Interesse des OWI-Entwicklers als verantwortliche Stelle besteht darin, ein OWI zu erstellen und zu betreiben, um das Internet zu indexieren und Informationen im Internet auffindbar zu machen. Dies dient nicht nur den wirtschaftlichen Interessen des Betreibers, sondern auch einem berechtigten Interesse der Allgemeinheit, Zugang Informationen zu erhalten. Zur Erstellung des Index muss der OWI-Entwickler möglichst viele Internetseiten crawlen und indexieren. Dabei verarbeitet er zwangsläufig personenbezogene Daten einer Vielzahl von betroffenen Personen. Diese Verarbeitung ist aus Sicht des OWI-Entwicklers erforderlich, um das berechtigte Interesse zu verwirklichen, da es keine gleichwertigen Mittel gibt, die weniger stark in die Grundrechte und Grundfreiheiten der betroffenen Personen eingreifen.

Fraglich bleibt, ob das Interesse betroffener Personen, nicht in einen Webindex aufgenommen zu werden und damit ihre informationelle Selbstbestimmung und ihr Recht auf Datenschutz zu schützen, das berechnigte Interesse des OWI-Entwicklers überwiegen kann. Dies erscheint unwahrscheinlich, insbesondere wenn man bedenkt, dass die gleichen Daten bereits mehrfach von anderen Indexbetreibern wie Google, Bing, Yandex oder Baidu gecrawlt wurden. Da ein Indexdienst Daten crawlt, die für jedermann frei zugänglich sind, erscheinen diese Daten nicht besonders schutzbedürftig. Der EuGH hat gefestigter Rechtsprechung zwar die datenschutzrechtliche Verantwortlichkeit eines Suchmaschinen- und Indexbetreibers ausdrücklich festgestellt,⁴⁰ jedoch nie die grundsätzliche Ermächtigung des Betreibers in Frage gestellt, personenbezogene Daten zu crawlen und zu indexieren. Dies ist aus Praktikabilitätsabwägungen einleuchtend, da die Indexierung an sich wohl alternativlos ist. Hervorzuheben ist auch, dass die betroffenen Personen nicht schutzlos gestellt sind. Sie können sich bereits ex ante gegen eine Datenverarbeitung durch den OWI schützen, indem sie geeignete Maßnahmen ergreifen, damit ihre Daten nicht im Internet veröffentlicht sind – also z. B. direkt gegen Webseitenbetreiber vorgehen. Ex post kann die betroffene Person der Verarbeitung durch den OWI-Entwickler gemäß Art. 21 Abs. 1 DSGVO widersprechen und nach Art. 17 Abs. 1 lit. c DSGVO die Löschung der sie betreffenden Daten fordern. Zusammenfassend überwiegen die Interessen der betroffenen Personen nicht die des OWI-Betreibers und der Allgemeinheit an der Erstellung eines OWI.

Dies gilt im Ergebnis auch für die Verarbeitung besonderer Kategorien personenbezogener Daten. Diese unterliegt den Anforderungen von Art. 9 DSGVO. Die Verarbeitung solcher Daten ist grundsätzlich untersagt, es sei denn, die Voraussetzungen des Art. 9 Abs. 2 DSGVO sind erfüllt. Solange der OWI nur öffentlich zugängliche Daten crawlt, ist davon auszugehen, dass diese Daten in der Regel von der betroffenen Person veröffentlicht wurden und daher gemäß Abs. 2 lit. e verarbeitet werden dürfen. Es bleibt jedoch ein Restrisiko, dass sensible Daten gecrawlt und verarbeitet werden, die nicht von der betroffenen Person, sondern von einem Dritten veröffentlicht wurden (z.B. bei Datenpannen). Diesem Risiko könnte der OWI-Entwickler zumindest durch technische und organisatorische Maßnahmen

40 EuGH, Urteil vom 8.12.2022 – Az. C-460/20 (TU, RE/Google LLC) - ECLI:EU:C:2022:962. EuGH, Urteil vom 13.5.2014 – Az. C-131/12 (Google Spain) - ECLI:EU:C:2014:317; EuGH, Urteil vom 27.9.2019 – Az. C-136/17 (Google vs. CNIL) - ECLI:EU:C:2019:773.

begegnen, um die Verarbeitung von personenbezogenen Daten besonderer Kategorien zu minimieren. Zu denken wäre hier an Filtertechniken und Anonymisierung.

6.2 Digital Services Act

In seiner jetzigen Form wird der OWI auch unter die Regeln über digitale Dienstleistungen fallen, insbesondere unter den DSA. Diese Verordnung regelt im Wesentlichen die Haftung und den Umfang der Sorgfaltspflichten für Anbieter digitaler Vermittlungsdienste. Ziel ist, einen sicheren Rechtsrahmen für den digitalen Raum zu schaffen, um so ein sicheres, berechenbares und vertrauenswürdiges Online-Umfeld zu schaffen und die Grundrechte zu schützen, vgl. Art. 1 Abs. 1 DSA. Vermittlungsdienste "umfassen ein breites Spektrum wirtschaftlicher Tätigkeiten, die online stattfinden und sich ständig weiterentwickeln, um eine schnelle, sichere und geschützte Übertragung von Informationen zu ermöglichen und bequeme Lösungen für alle Beteiligten im Online-Ökosystem zu bieten."⁴¹

Die Einordnung des OWI unter die einzelnen Merkmale des DSA ist schwierig, insbesondere wenn es um die Qualifikation als Hosting-Dienst, Online-Plattform oder Suchmaschine geht. Die Einordnung des OWI als Vermittlungsdienst i.S.d. Art. 3 lit. g Nr. iii DSA und als Online-Suchmaschine i.S.d. Art. 3 lit. j DSA ist fraglich.⁴² Es ist offensichtlich, dass der Gesetzgeber bei der Ausgestaltung des DSA von Anwendungsfällen wie Social Media und Online-Marktplätzen ausgegangen ist, während übergeordnete Anwendungen wie ein Webverzeichnis oder ein Index gar nicht erst in Betracht gezogen wurden. Auch die Regelungen zu Online-Suchmaschinen wurden nicht zu Ende gedacht. Dennoch sind die Funktionen und Möglichkeiten eines OWI häufig mit typischen Anwendungsfällen des DSA vergleichbar, sodass nach dem Schutzzweck des DSA der OWI darunterfallen sollte.

Sowohl im Hinblick auf den Zweck des DSA als auch aus allgemeinen Erwägungen wäre es sinnvoll, OWI-Betreibern die Haftungsbefreiung des DSA oder vergleichbare Privilegien zuzugestehen. Gleichzeitig sollten ihnen aber auch entsprechende Sorgfaltspflichten auferlegt werden, die

41 Erwägungsgrund 29 DSA.

42 *Nebel/Johannes*, Open Web Index im Lichte des Digital Services Act, Multimedia und Recht 2024, 1013.

gerade bei einer sehr großen Online-Suchmaschine erheblich sind. Hier sollte der Gesetzgeber nachbessern und sowohl Online-Suchmaschinen als auch den OWI in die Definition des Art. 3 lit. g DSGVO einbeziehen und die korrespondierenden Pflichten und Haftungsbefreiungen entsprechend benennen. Trotz Ermangelung einer solchen eindeutigen Regelung sollte der OWI-Entwickler davon ausgehen, dass es sich bei dem OWI um eine Suchmaschine und einen Hosting-Provider im Sinne des DSA handelt und entsprechende Maßnahmen ergreifen.⁴³

6.3 Urheberrecht

Für das Urheberrecht hat der unionale Gesetzgeber mit der Urheberrechtsrichtlinie (2001/29/EG) und der Richtlinie über das Urheberrecht im digitalen Binnenmarkt (2019/790) Bestimmungen erlassen, die von den Mitgliedstaaten in nationales Recht umgesetzt wurden. Im Folgenden wird die Gesetzgebung in Deutschland als Beispiel herangezogen.

Es gibt zwei Perspektiven: Zum einen ist zu prüfen, inwieweit der OWI selbst rechtlich geschützt ist. Insbesondere könnte der Index als Datenbankwerk nach § 4 Abs. 2 UrhG geschützt sein, wonach der Index als solcher unabhängig vom Urheberrecht an den indizierten Inhalten geschützt wäre und nur unter bestimmten Voraussetzungen nach § 55a UrhG, insbesondere durch Einräumung von Nutzungsrechten, von Dritten genutzt werden könnte.

Zum anderen ist zu klären, inwieweit urheberrechtlich geschützte Inhalte im OWI indexiert und von Anwendungen, die auf dem Index basieren, genutzt werden dürfen. Der OWI enthält eine große Menge an Daten, bei denen es sich größtenteils um nach § 2 UrhG geschützte Werke handelt, die im Rahmen der Aufnahme in den Index vervielfältigt werden. Das Vervielfältigungsrecht nach § 16 UrhG steht jedoch nach § 15 Abs. 1 Nr. 1 UrhG grundsätzlich dem Urheber des Werkes zu und nicht einem Nutzer, wie dem OWI-Entwickler. Ähnliche Probleme bestehen auch im Zusammenhang mit Anwendungen, die auf dem OWI basieren, etwa wenn der Index zum Training von KI-Modellen verwendet wird. Bei der Erstellung des Index und der Entwicklung von Anwendungen lassen sich in Urheberrechte eingreifende Vervielfältigungen nicht vermeiden.

43 *Nebel/Johannes*, Open Web Index im Lichte des Digital Services Act, Multimedia und Recht 2024, 1016.

Das Urheberrechtsgesetz enthält jedoch verschiedene Ausnahmen, die Vervielfältigungshandlungen rechtfertigen können. Im Jahr 2021 hat der deutsche Gesetzgeber in Umsetzung von Art. 4 der Richtlinie Nr. 2019/790 die Ausnahmeregelung des § 44b UrhG für allgemeines Text- und Data-Mining geschaffen. Diese Norm soll es ermöglichen, große Mengen digitaler Informationen zu analysieren. Nach Ansicht der Wissenschaft⁴⁴ kann das Training von KI-Modellen in der Regel durch § 44b UrhG gerechtfertigt werden; und auch die Rechtsprechung zeigt eine leichte Tendenz in diese Richtung⁴⁵. Allerdings sind etwaige Vorbehalte des Urhebers nach § 44b Abs. 3 UrhG zu berücksichtigen. Bei Vervielfältigungshandlungen, die durch Crawler bei der Indexierung von Webinhalten entstehen, kommt auch § 44a UrhG in Betracht. Hier sieht der Gesetzgeber eine Ausnahme für Vervielfältigungshandlungen vor, die nur vorübergehender Natur und Teil eines technischen Verfahrens sind, keine eigenständige wirtschaftliche Bedeutung haben und einem Zweck des § 44a UrhG dienen.

Neben diesen Ausnahmen haben die deutschen Gerichte auch entschieden, dass Suchmaschinen urheberrechtlich geschützte Inhalte indexieren dürfen: Mit der Veröffentlichung urheberrechtlich geschützter Werke im Internet willigt der Rechteinhaber ein (konkludente Einwilligung), dass seine Inhalte von Suchmaschinen gefunden werden können.⁴⁶ Voraussetzung dafür ist die Indexierung, sodass sich die Suchmaschinenbetreiber auf die konkludente Einwilligung der Rechteinhaber berufen können, um einen Index aufzubauen und zu pflegen. Das kann unter Umständen sogar dann der Fall sein, wenn das urheberrechtlich geschützte Werk gegen den Willen des Rechteinhabers ins Internet gestellt wurde.⁴⁷ Anders verhält es sich jedoch, wenn der Rechteinhaber Beschränkungen auferlegt, z. B.

44 Z.B. Bomhard/Siglmüller, *Trilogieergebnis*, *Recht Digital* 2024, 50; Dregelies, *KI-Training*, *Gewerblicher Rechtsschutz und Urheberrecht* 2024, 1484; Heine, *Nutzungsrechte*, *Gewerblicher Rechtsschutz und Urheberrecht in der Praxis* 2024, 88; Hofmann, *Geschäftsmodelle*, *Zeitschrift für Urheber- und Medienrecht* 2024, 166; Käde, *Training generativer KI-Modelle*, *Künstliche Intelligenz und Recht* 2024, 162; Maamar, *Urheberrechtliche Fragen*, *Zeitschrift für Urheber- und Medienrecht* 2023, 481; Wagner, *Blackbox*, *Zeitschrift für IT-Recht und Recht der Digitalisierung* 2024, 298; Gegenmeinung: Dornis/Stober, *Urheberrecht und Training generativer KI-Modelle*, 2024; Dornis, *KI-Training und Text- und Data-Mining*, *Künstliche Intelligenz und Recht* 2024, 156.

45 LG Hamburg, Urteil vom 27. September 2024 – Az. 310 O 227/23.

46 BGH, Urteil vom 19. Oktober 2011 – Az. I ZR 140/10; Klass, *Schlichte Einwilligung*, *Zeitschrift für Urheber- und Medienrecht* 2013, 1.

47 BGH, Urteil vom 21. September 2017 – Az. I ZR 11/16.

in Form von robots.txt-Dateien oder Anmeldeanforderungen. Der Europäische Gerichtshof vertritt eine ähnliche Linie und argumentierte, dass Suchmaschinen zur Verbreitung von Informationen und zum Funktionieren des Internets beitragen.⁴⁸ Aus diesem Grund stelle die Verlinkung auf öffentlich zugängliche Inhalte keine Urheberrechtsverletzung dar. Diese Argumentation auf den OWI übertragbar.

6.4 KI-Verordnung

Der KI-Verordnung zielt auf die Regulierung von Systemen und Praktiken im Bereich der künstlichen Intelligenz ab. Sie soll einen robusten und flexiblen Rechtsrahmen schaffen, der die Nutzung von KI und automatisierten Systemen vertrauenswürdig und sicher macht. Die Verordnung verfolgt einen risikobasierten Ansatz.⁴⁹ Je höher das Risiko, desto umfangreicher sind die Verpflichtungen, die den Entwicklern und Betreibern⁵⁰ von KI-Systemen auferlegt werden. KI-Systeme mit inakzeptablen Risiken, z. B. Systeme, die ein "Social Scoring" durch Regierungen oder Unternehmen ermöglichen, werden als Bedrohung für die Grundrechte der Menschen angesehen und sind daher verboten.⁵¹ Um dem spezifischen Transparenzrisiko zu begegnen, müssen bestimmte KI-Systeme, etwa Chatbots, die Nutzer eindeutig darüber informieren, dass sie mit einer Maschine interagieren. KI-generierte Inhalte müssen zudem als solche gekennzeichnet werden.⁵² Nur für KI-Systeme mit geringem Risiko gelten keine Verpflichtungen gemäß der KI-VO. KI-Systeme mit hohem Risiko⁵³ hingegen, wie etwa KI-basierte medizinische Software oder KI-Systeme, die für die Personalbeschaffung eingesetzt werden, müssen strenge Anforderungen erfüllen, also zum Beispiel Systeme zur Risikominderung einführen, hochwertige Datensätze nutzen, klare Nutzerinformationen geben und menschliche Aufsicht gewährleisten.

48 EuGH, Urteil vom 13. Februar 2014 – Az. C-466/12.

49 Siehe Erwägungsgrund 26 KI-VO.

50 Siehe Art. 3 Nr. 8 KI-VO.

51 Siehe Art. 5 KI-VO.

52 Siehe Art. 50 KI-VO.

53 Siehe Art. 6 KI-VO und Anhang I und II KI-VO.

6.4.1 Unmittelbare Anwendbarkeit der KI-VO auf den OWI

Der OWI-Entwickler muss prüfen, inwieweit die Bestimmungen der KI-VO unmittelbar auf die zur Erstellung des OWI eingesetzten Technologien anwendbar sind. Dem risikobasierten Ansatz der KI-VO folgend werden KI-Systeme, "die erhebliche nachteilige Auswirkungen auf die Gesundheit, die Sicherheit und die Grundrechte von Personen haben können"⁵⁴ in Art. 6 KI-VO als Hochrisikosysteme eingestuft, wobei zwischen hochriskanten KI-Systemen im Zusammenhang mit der Produktregulierung (Abs. 1) und eigenständigen hochriskanten KI-Systemen (Abs. 2) unterschieden wird.

Angenommen werden kann, dass die Algorithmen und Systeme, die vom OWI zur Unterstützung und Koordinierung des Web-Crawling verwendet werden, zunächst grundsätzlich um KI-Systeme im Sinne der KI-VO handelt. Diese beziehen sich auf den ersten Blick auch nur auf interne Vorgänge. Dennoch ist anhand einer Risikobewertung zu prüfen, welcher Risikostufe diese KI-Systeme zugeordnet werden müssen.

Als Hochrisikosysteme gelten KI-Systeme, die Sicherheitsbauteile von Produkten sind (Art. 3 Nr. 14 KI-VO) oder selbst Produkte sind, die unter die Rechtsvorschriften des Anhangs II der KI-VO fallen (z. B. Maschinen, medizinische Geräte, Kraftfahrzeuge und Flugzeuge). Der OWI würde demnach wohl nicht als Hochrisikosystem nach Art. 6 Abs. 1 KI-VO einzustufen sein.

In Art. 6 KI-VO werden außerdem die in Anhang III der KI-VO aufgeführten KI-Systeme als besonders risikoreich bewertet. Dazu gehören KI-Systeme in den Bereichen Biometrie, kritische Infrastrukturen, Bildung, Beschäftigung, Grundversorgung, Strafverfolgung, Migration, Asyl und Grenzkontrolle sowie Justizverwaltung und demokratische Prozesse. KI, die bei der Indexierung sowie bei der Entscheidung über den Ausschluss oder die Aufnahme bestimmter Webinhalte eingesetzt wird, hat im Allgemeinen spürbare Auswirkungen auf die Indexnutzung durch Dritte. Dies ist jedoch eine Frage des konkreten Einsatzumfangs. Nach jetzigem Funktionsumfang des Index ist nicht davon auszugehen, dass der Index einem der in Anhang III aufgeführten Sektoren zuzuordnen ist. Perspektivisch könnte jedoch Einordnung unter der in Nr. 2 aufgeführten "kritischen Infrastrukturen" erfolgen.

⁵⁴ Siehe Erwägungsgrund 46 KI-VO.

6.4.2 Mittelbare Auswirkungen der KI-VO auf den Index

Darüber hinaus stellt sich die Frage, ob die KI-Verordnung Regelungen enthält, die die Nutzung von OWI-Daten für bestimmte KI-Anwendungen beeinflussen. Zum Beispiel ist der Betrieb bestimmter KI-Systeme nach Art. 5 KI-VO verboten. Der OWI-Entwickler könnte deswegen erwägen, die Verwendung des Index für das Training solcher KI-Systeme zu verhindern. Das könnte über eine Indexlizenz oder die Bedingungen für die Nutzung des OWI erreicht werden. Der OWI-Betreiber könnte dabei einen Einschätzungsspielraum nutzen. Verboten ist nach Art. 5 KI-VO in der Regel das Inverkehrbringen und die Inbetriebnahme verbotener KI-Systeme. Die bloße wissenschaftliche Erforschung und Entwicklung des Einsatzes von KI-Systemen, die lediglich auch zu verbotenen Praktiken fähig wären, ist deswegen nicht verboten.⁵⁵ Darüber hinaus ist nach Art. 2 Abs. 6 KI-VO die KI-Verordnung nicht anwendbar für KI-Systeme oder KI-Modelle einschließlich ihrer Ergebnisse, die ausschließlich zum Zwecke der wissenschaftlichen Forschung und Entwicklung entwickelt und in Betrieb genommen werden, vgl. Art. 3 Nr. 11 KI-VO. Das dient der Innovationsförderung und dem Schutz der Wissenschaftsfreiheit.⁵⁶ Wissenschaftliche Forschung umfasst nach Art. 13 GRCh Tätigkeiten mit dem Ziel, "methodisch, systematisch und nachprüfbar neue Erkenntnisse zu gewinnen". Dazu gehören die Grundlagenforschung und die angewandte Forschung im öffentlichen (z. B. Universitäten) und privaten (z. B. industrielle Forschung) Sektor.

Ein anderes Beispiel für eine mögliche mittelbare Auswirkungen der KI-VO auf den Index oder den OWI-Betreiber, dass KI-Systeme mit hohem Risiko mit Trainings-, Validierungs- und Testdatensätzen entwickelt werden müssen, die bestimmte Qualitätskriterien erfüllen.⁵⁷ Art. 10 Abs. 3 KI-VO schreibt vor, dass Trainings-, Validierungs- und Testdatensätze relevant, hinreichend repräsentativ und im Hinblick auf den vorgesehenen Zweck möglichst fehlerfrei und vollständig sein müssen. Fraglich ist u.a., ob aus Datenschutzgründen anonymisierte oder pseudonymisierte Daten (z. B. durch Hinzufügen von Rauschen) noch als fehlerfrei und vollständig angesehen werden können.⁵⁸ Die Datensätze müssen zudem die entsprechenden statistischen Merkmale aufweisen, ggf. auch im Hinblick auf die Personen oder Personengruppen, durch die das Hochrisiko-KI-System bestimmungs-

55 Siehe auch Erwägungsgrund 25 KI-VO.

56 Siehe Erwägungsgrund 25 KI-VO.

57 Siehe Art. 10 Abs. 2 - 5 KI-VO in Übereinstimmung mit Art. 10 Abs. 1 KI-VO.

58 Vogel u. a., in: Demmler u.a. (Hrsg.), INFORMATIK 2022, 2022, 659.

gemäß eingesetzt werden soll. OWI-Daten stellen eine umfangreiche und vielfältige Informationsquelle dar, die Texte und Links enthält. Diese Vielfalt ist für das Training von KI-Modellen entscheidend. Um die Qualitätskriterien für KI-Systeme mit hohem Risiko erfüllen zu können, sollte der OWI-Entwickler Techniken anwenden, um die Datenqualität sicherzustellen, z. B. Datenbereinigung, -anreicherung, -abgleich oder -kommentierung. Gleichzeitig könnte er versuchen, seine Datensätze so zu erstellen und Maßnahmen zu ergreifen, die nachfolgende Anwendungsentwickler nutzen oder auf denen sie aufbauen können, um die Datenqualität für ihren spezifischen Anwendungsfall sicherzustellen.

7 Nutzerakzeptanz und Vertrauen

Die rechtliche Untersuchung des OWI sollte mit Überlegungen zur Nutzerakzeptanz und zum Vertrauen einhergehen, um eine positive Nutzerwahrnehmung und letztlich den Markterfolg zu fördern. Obwohl eine rechtskonforme Gestaltung technischer Innovationen sowie Datenschutz und Privatsphäre in Europa von großer Bedeutung sind, zeigt die Praxis, dass Nutzer bei der Wahl digitaler Dienste oft andere Faktoren in ihre Entscheidung einbeziehen – oder diese sogar als wichtiger einstufen. Daher ist es essenziell, Schlüsselaspekte der Nutzerakzeptanz frühzeitig in den Designprozess des OWI und den darauf basierenden Tools zu integrieren. Ein leistungsfähiger europäischer Webindex muss nicht nur Datenschutz- und Privatsphäre-Konzepte berücksichtigen, sondern auch zentrale Kriterien der Nutzerakzeptanz. Nur so kann er auf dem globalen Markt konkurrieren oder zumindest mit internationalen Anbietern in Europa mithalten.

Die Wahrnehmung der Nutzer und die Absicht der Anwender, neue technologische Lösungen zu nutzen, sind entscheidende Faktoren für deren erfolgreiche Entwicklung. Diese Faktoren werden im Rahmen des Forschungsgebiets Nutzerakzeptanz untersucht.⁵⁹ Ein weiterer Schlüsselfaktor für Technologieakzeptanz ist das Vertrauen in IT-Artefakte.⁶⁰ Diese Überlegungen sind für den Kontext des OWI wichtig, um das Vertrauen der

59 Davis, User acceptance of information technology, *International Journal of Man-Machine Studies* 1993, S. 475.

60 Gefen u.a., Trust and tam in online shopping, *MIS Quarterly* 2003, S. 51; Söllner u.a., Why different trust relationships matter, *European Journal of Information Systems* 2016, S. 274.

Nutzer in den OWI und in die darauf aufbauenden Tools zu stärken. Mit zunehmender technologischer Komplexität wird Vertrauen immer essenzieller, da Nutzer nicht in der Lage sind, die zugrunde liegenden Strukturen ständig zu durchdringen.⁶¹

Aus der Theorie lassen sich einige Modelle vorschlagen, die unterstützende Faktoren zu Vertrauen und Nutzerakzeptanz bereitstellen. Zur praktischen Anwendung für den OWI eignet sich eine Kombination aus zwei etablierten Modellen: Trust-TAM⁶² und UTAUT⁶³. Relevante Faktoren können zusammengestellt werden, um die Grundlagen für den Aufbau und die Aufrechterhaltung von Nutzerakzeptanz und Vertrauen in den OWI und seine zukünftigen Produkte zu schaffen. Außerdem sind diese Faktoren in Bezug auf die einzelnen Akteure zu betrachten und weiterzuentwickeln, um Nutzerakzeptanz und Vertrauen in den jeweiligen Kontexten gezielt zu fördern.

8 Schlussfolgerungen und Ausblick

Diese Analyse soll das enorme Potenzial eines frei zugänglichen und umfassenden OWI zur Umgestaltung der digitalen Landschaft verdeutlichen. Durch die Förderung eines vielfältigeren und wettbewerbsfähigeren Suchmaschinenmarktes kann der OWI den Anwendungsentwicklern und in der Folge den Endnutzern mehr Auswahl und Kontrolle ermöglichen. Darüber hinaus kann der OWI als entscheidende Ressource für die KI-Entwicklung dienen, innovative Technologien vorantreiben und deren verantwortungsvolle Nutzung unterstützen.

Gleichzeitig bringt die Entwicklung und Implementierung des OWI spezifische Herausforderungen und Chancen für den Schutz der Grundrechte mit sich. Während der OWI das Potenzial hat, den Zugang zu Informationen (Art. 11 GRCh) zu verbessern, den Wettbewerb auf dem digitalen Markt zu fördern und die Innovation voranzutreiben, wirft er auch Fragen des Datenschutzes (Art. 8 GRCh), der unternehmerischen Freiheit (Art. 16 GRCh) und des Schutzes des geistigen Eigentums (Art. 17 GRCh) auf.

Die erfolgreiche Umsetzung des OWI erfordert eine sorgfältige Prüfung der rechtlichen und nutzerbezogenen Anforderungen. Die Gewährleistung

61 Muir, Trust in Automation, Ergonomics 1994, S. 1905.

62 Gefen u.a., Trust and tam in online shopping, MIS Quarterly 2003, S. 64.

63 Venkatesh u.a., Consumer acceptance, MIS Quarterly 2012, S. 157.

des Datenschutzes, die Wahrung der Rechte an geistigem Eigentum und die Abmilderung potenzieller Risiken für Grundrechte sind von größter Bedeutung. Die Einhaltung des europäischen Rechtsrahmens muss die Entwicklung und Nutzung des OWI leiten, begleitet von Überlegungen zu Nutzerakzeptanz und Vertrauen, um verantwortungsvolle technologische Innovationen zu unterstützen.

Literatur

- Arslan, Muhammad; Ghanem, Hussam; Munawar, Saba und Cruz, Christophe (2024): A Survey on RAG with LLMs. *Procedia Computer Science*, 246, S. 3781-90. ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2024.09.178>.
- Baars, Henning und Hans-Georg Kemper (2021): *Business Intelligence & Analytics – Grundlagen und praktische Anwendungen: Ansätze der IT-basierten Entscheidungsunterstützung*. 4te Auflage. Wiesbaden: Springer Fachmedien Wiesbaden. ISBN 9783834823441
- Bomhard, David und Siglmüller, Jonas (2024): KI-Gesetz - das Trilogieergebnis, *Recht Digital*, 5(2), S. 45-55.
- Braun, Max und Trepte, Sabine (2017): Privatheit und informationelle Selbstbestimmung: Trendmonitor zu den Einstellungen, Meinungen und Perspektiven der Deutschen. Stuttgart: Universität Hohenheim. URL: https://medienpsychologie.uni-hohenheim.de/fileadmin/einrichtungen/psych/Dateien/Laufende_Projekte/Trendmonitor_Privatheit_Hohenheim.pdf (besucht am 30.07.2025).
- Brin, Sergey und Page, Lawrence (1998): The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Computernetzwerke und ISDN-Systeme*, 30(1-7). S. 107.
- Curran, Kevin und McGlinchey, Jude (2007): Vertical Search Engines. *The ITB Journal*, 8(2). S. 22-27. doi:10.21427/D7MT83
- Dakhel, Arghavan Moradi; Nikanjam, Amin; Khomh, Foutse; Desmarais, Michel C. und Washizaki, Hironori (2024): An Overview on Large Language Models. In: Nguyen-Duc, Anh; Abrahamsson, Pekka und Khomh, Foutse (Hrsg.): *Generative AI for Effective Software Development*. Cham: Springer, S. 3-21. https://doi.org/10.1007/978-3-031-55642-5_1.
- Davis, Fred D. (1993): User acceptance of information technology: System characteristics, user perceptions and behavioral impacts. *International Journal of Man-Machine Studies*, 38(3), S. 475-489. <https://doi.org/10.1006/imms.1993.1022>.
- Dornis, Tim W (2024): Generatives KI-Training und Text- und Data-Mining. Eine funktionale Unterscheidung. *Künstliche Intelligenz und Recht*, 1(5), S. 156-61.
- Dornis, Tim W. und Stober, Sebastian (Hrsg.): (2024): *Urheberrecht und Training generativer KI-Modelle: Technologische und juristische Grundlagen*, 1. Aufl. Baden-Baden: Nomos. <https://doi.org/10.5771/9783748949558>.
- Dregelies, Max (2024): KI-Ausbildung nach dem KI-Gesetz. *Gewerblicher Rechtsschutz und Urheberrecht*, 126(20), S. 1484-1493.
- Gefen, David; Karahanna, Elena und Straub, Detmar W. (2003): Trust and tam in online shopping: An integrated model. *MIS Quarterly*, 27(1), S. 51-90. 10.2307/30036519.

- Geminn, Christian L.; Voigt, Stefan; Söllner, Matthias; Johannes, Paul C.; Koulani, Huda; Beer, Leopold (2025): Anwendungsfälle eines offenen Webindex - Eine Handreichung für Wirtschaft und Wissenschaft. https://pridi-projekt.de/wp-content/uploads/2025/05/PriDI-Paper_Anwendungsfaelle.pdf.
- Geminn, Christian L.; Erenli, Kai und Pfeiffer, Leon (2021): Legal Challenges of an Open Web Index. *International Cybersecurity Law Review*, (2), S. 183-94. <https://doi.org/10.1365/s43439-021-00017-8>.
- Geminn, Christian L. (2021): Rechtsfragen eines offenen Web-Index - Infrastrukturen für die digitale Gesellschaft. *Multimedia und Recht*, (12), S. 16-19.
- Geminn, Christian L. (2023): Deus ex machina? – Grundrechte und Digitalisierung, Tübingen: Mohr Siebeck.
- Geminn, Christian L. und Johannes, Paul C. (Hrsg.) (2025): *Handbuch europäisches Datenrecht*. Baden-Baden: Nomos (in Vorbereitung).
- Heine, Robert (2024): Generative KI: Nutzungsrechte und Nutzungsvorbehalt. *Gewerblicher Rechtsschutz und Urheberrecht in der Praxis*, 16(4), S. 87-90.
- Google: How Google Search organizes information. Google.com. URL:https://www.google.com/intl/en_us/search/howsearchworks/how-search-works/organizing-information/ (besucht am 26.02.2025).
- Hendriksen, G. u.a. (2024): The Open Web Index. In: Goharian, N. u.a. (Hrsg.): *Advances in Information Retrieval. ECIR 2024. Lecture Notes in Computer Science*, Cham: Springer, S. 130-43. https://doi.org/10.1007/978-3-031-56069-9_10
- Hofmann, Franz (2024): Retten Schranken Geschäftsmodelle generativer KI-Systeme?. *Zeitschrift für Urheber- und Medienrecht*, 68(3), S. 166-74.
- Kaede, Lisa (2024): Training generativer KI-Modelle ist (auch) Text- und Data-Mining. Anwendbarkeit der TDM-Schranke des § 44b UrhG. *Künstliche Intelligenz und Recht*, 1(5), S. 162-69.
- Klass, Nadine (2013): Neue Internettechnologien und das Urheberrecht: Die schlichte Einwilligung als Rettungsanker?. *Zeitschrift für Urheber- und Medienrecht*, 57(1), S. 1-11.
- Lewandowski, Dirk (2014): Why We Need an Independent Index of the Web. In: König, René und Rasch, Miriam (Hrsg.): *Society of the Query Reader: Reflections on Web Search*. Amsterdam: Institute of Network Cultures. S. 49-58.
- Lewandowski, Dirk (2019): The Web is missing an essential part of infrastructure: an Open Web Index. *Communications of the ACM*, 62(4), S. 24-27. <https://doi.org/10.1145/3312479>.
- Liaw, Shu-Sheng; Huang, Hsiu-Mei (2003): An investigation of user attitudes toward search engines as an information retrieval tool. *Computers in Human Behavior*, 19(6), S. 751-65. [https://doi.org/10.1016/S0747-5632\(03\)00009-8](https://doi.org/10.1016/S0747-5632(03)00009-8).
- Maamar, Niklas (2023): Urheberrechtliche Fragen beim Einsatz von generativen KI-Systemen. *Zeitschrift für Urheber- und Medienrecht*, 67(7), S. 481-91.
- Muir, Bonnie M. (1994): Trust in Automation: Part I. Theoretical Issues in the Study of Trust and Human Intervention in Automated Systems, *Ergonomics* 37(11), S. 1905-1922. [10.1080/00140139408964957](https://doi.org/10.1080/00140139408964957).

- Nebel, Maxi und Johannes, Paul. C. (2024): Open Web Index im Lichte des Digital Services Act - Voraussetzungen - Grenzen - Rechtsfolge. *Multimedia und Recht*, (12), S. 1010-16.
- Nowakowski, Daniel; Kerner, Lara und Zimmermann, Nils (2024): Market potential assessment of OpenWebSearch.eu. OpenWebSearch.EU, Final project report. URL: <https://openwebsearch.eu/wp-content/uploads/2024/09/MarketAssessmentOfOWI-Report-V1.pdf>.
- Schultheiß, Sebastian und Lewandowski, Dirk: Misplaced trust? The relationship between trust, ability to identify commercially influenced results and search engine preference. *Journal of Information Science* 49 (3). <https://doi.org/10.1177/0165551521101415>.
- Shams, Abdullah Bin; Hoque Apu, Ehsanul; Rahman, Ashiqur; Sarker Raihan, Md. Mohsin; Siddika, Nazeeba; Preo, Rahat Bin; Hussein, Molla Rashied; Mostari, Shabnam und Kabir, Russell (2021): Web Search Engine Misinformation Notifier Extension (SEMiNext): A Machine Learning Based Approach during COVID-19 Pandemic. *Healthcare*, 9(2), ArtNr. 156. <https://doi.org/10.3390/healthcare9020156>.
- Sharma, Lavanya und Garg, Prageep Kumar (Hrsg.): (2021): *Artificial Intelligence: Technologies, Applications, and Challenges* (1. Auflage). New York: Chapman and Hall/CRC. <https://doi.org/10.1201/9781003140351>.
- Smith, Tim (2023): Navigating the Entangled Web, TedxVerbier, abrufbar unter https://www.ted.com/talks/tim_smith_navigating_the_entangled_web
- Söllner, Matthias; Hoffmann, Axel und Leimeister, Jan Marco (2016): Why different trust relationships matter for information systems users. *European Journal of Information Systems*, 25(3), S. 274-87. <https://doi.org/10.1057/ejis.2015.17>.
- StatCounter (23. January 2025): Marktanteil der führenden Suchmaschinen weltweit von Januar 2015 bis Januar 2025. Statista.com. URL:<https://www.statista.com/statistics/1381664/worldwide-all-devices-market-share-of-search-engines/> (besucht am 29.01.2025)
- StatCounter (30. April 2024): Marktanteile der führenden Suchmaschinen in Deutschland von Januar 2018 bis März 2024. Statista.com. URL:<https://www.statista.com/statistics/445974/search-engines-market-share-of-desktop-and-mobile-search-germany/> (besucht am 29.01.2025)
- Venkatesh, Viswanath; Thong, James Y. und Xu, Xin (2012): Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, 36(1), S. 157-78. JSTOR.
- Wagner, Kristina (2024): Generative KI: Eine "Blackbox" urheberrechtlicher Haftungsrisiken? Balanceakt zwischen Innovationsförderung und effektivem Rechtsschutz für Werke Dritter. *Zeitschrift für IT-Recht und Recht der Digitalisierung*, 27(4), S. 298-304.
- Wilson, Mark (10. Mai 2023): Microsoft Bing gegen Google Bard: Wer gewinnt den Kampf der KI-Chatbots? TechRadar.com. URL:<https://www.techradar.com/features/microsoft-bing-vs-google-bard-whos-winning-ai-powered-search> (besucht am 26.02.2025).

Vogel, Inna; Setz, Tahireh; Choi, Jeong-Eun und Steinebach, Martin (2022): Natural Language Processing (NLP) und der Datenschutz - Chancen und Risiken für den Schutz der Privatheit. In: Daniel Demmler, Daniel Krupka und Hannes Federrath (Hrsg.): *Recht und Technik: Datenschutz im Diskurs (RuT)*. INFORMATIK 2022, Gesellschaft für Informatik (GI). Hamburg, Deutschland, 26.-30. September, 2022. *Proceedings*. Bonn: Gesellschaft für Informatik, S. 651-664. PISSN: 1617-5468. ISBN: 978-3-88579-720-3. DOI: 10.18420/inf2022_52

