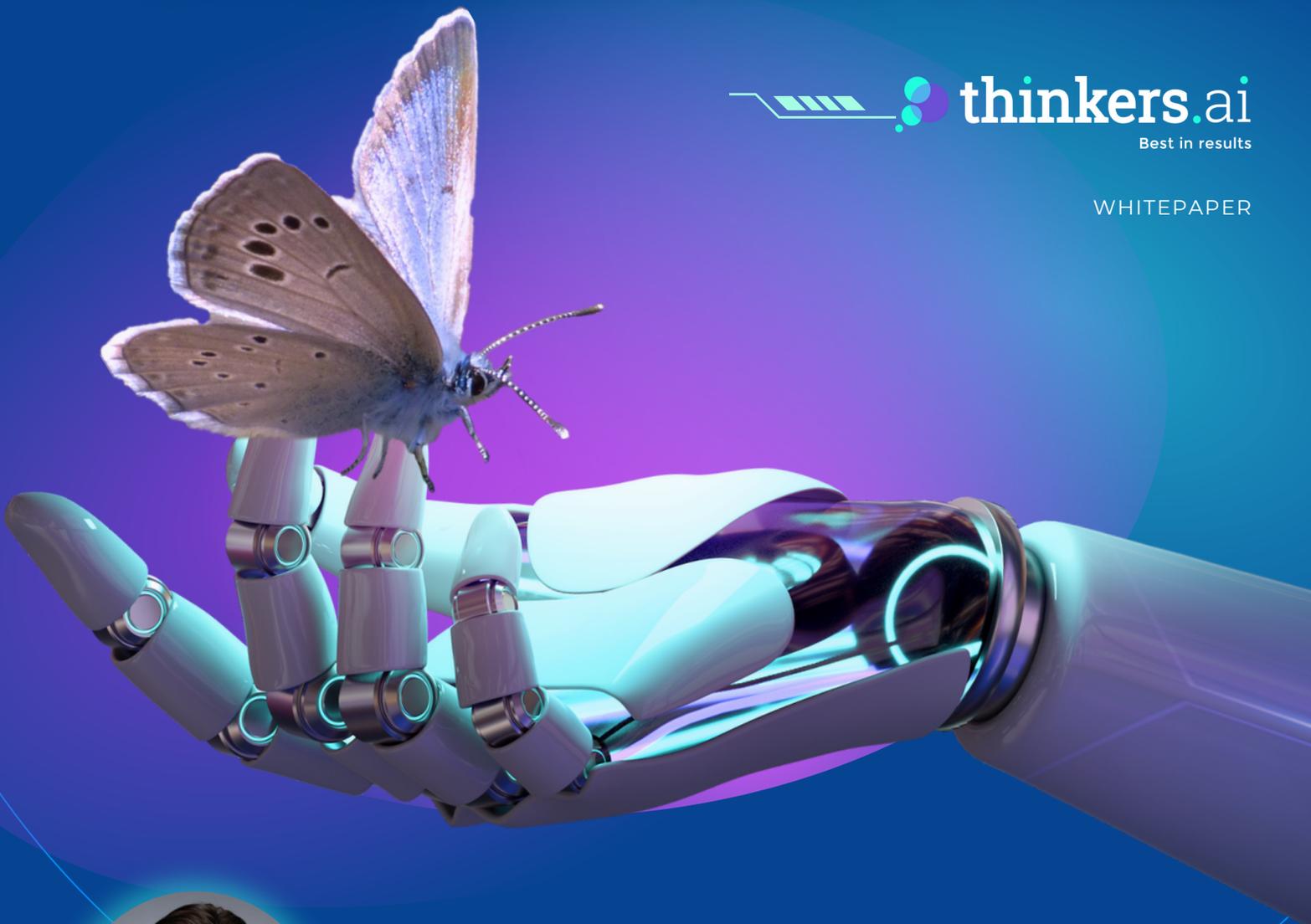


# LLM- VS. KEYWORD- BASIERTE WEBSUCHE



Best in results

WHITEPAPER



KONTAKT

**Isabell Claus**  
Managing Director

isabell.claus@thinkers.ai  
www.thinkers.ai

## LLM-basierte versus keyword-basierte Websuche

Wir alle sind jahrzehntelang mit Hilfe von Keywordsuche durch's Netz navigiert. Doch seit einigen Monaten beeinflussen Large Language Models (LLMs) die Architektur der Suche und die Art und Weise der Ergebnisdarstellung, die Möglichkeiten der Weiterentwicklung sowie die Zukunftsfähigkeit von Search Engines.

Wie kann man diese neuen Architekturen und technischen Unterschiede verständlich erklären? Im Folgenden geben wir einen Überblick:

### Die Daten als Grundlage aller Varianten

Die Grundlage jeder Analyse sind Daten. Bei der Websuche werden massenhafte Daten aus Webseiten aller Art und in verschiedenen Formaten extrahiert und indexiert. Daneben können im Falle von thinkers.ai auch andere, zum Beispiel interne Datenquellen, angeschlossen werden. Es werden Schlüsselwörter und Metadaten aus den Dokumenten extrahiert und in einer durchsuchbaren Datenbank – dem Index – gespeichert.

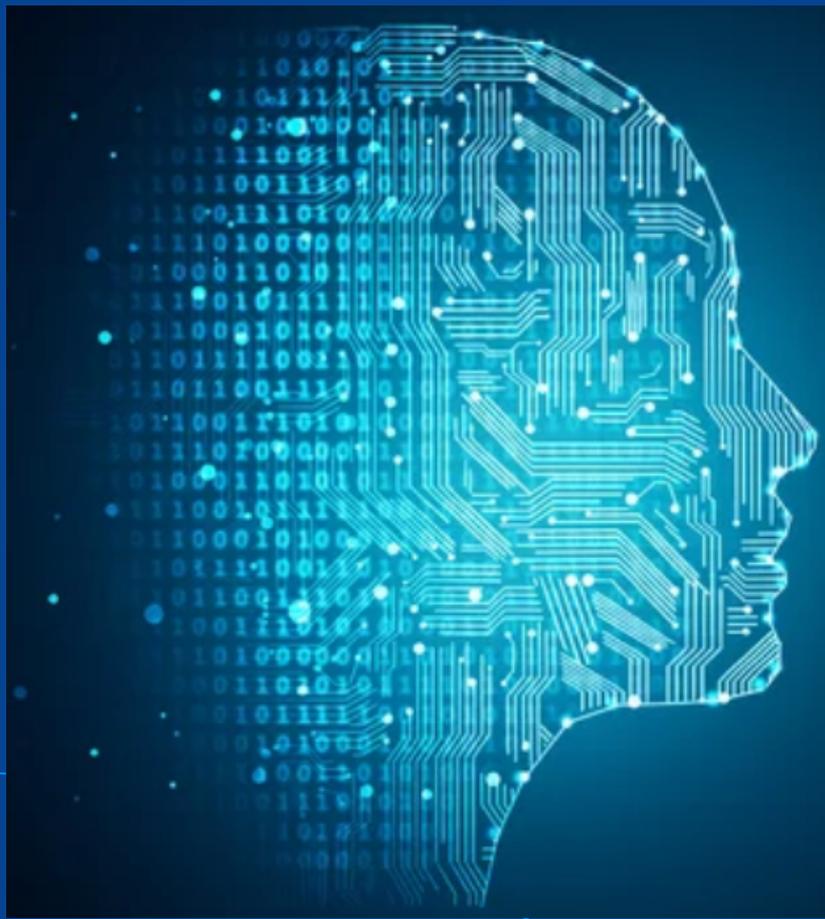
### Wie funktioniert die Keyword-basierte Suche?

Keyword-basierte Suchen sind in der ursprünglichen Form die simpelste Art und Weise der Websuche. Diese Technologie aus den 90er bzw. 2000er Jahren wurden in zahlreichen, umfassenden Iterationen weiterentwickelt. Sie funktionieren über einen Index, der neben Informationen aus dem Text auch Metadaten, URLs und weitere Daten enthält. Wenn ein(e) Anwender:in eine Suchanfrage stellt, gibt er/sie in der Regel Keywords oder Phrasen ein. Diese Suchanfrage wird verwendet, um im Index nach Treffern zu suchen. Um dann die Relevanz der Treffer automatisiert zu bewerten, werden verschiedene Faktoren analysiert, zum Beispiel die Häufigkeit der Keywords, ihre Position und Informationen aus Metadaten. Die nach der algorithmisch berechneten Relevanz höchst-gerankten Ergebnisse werden den Anwender:innen als Suchergebnisse angezeigt. Die Anwender:innen können dann die Suchergebnisse durchklicken und auf diejenigen Dokumente zugreifen, die ihrer Suche individuell am besten zu entsprechen scheinen.

Für eine etwas genauere Suche bieten die meisten Keyword-basierten Websuchmaschinen zusätzlich Filter, mit denen Suchergebnisse weiter eingeschränkt werden können (z.B. nach Dokumententyp oder Meldungsart). Gängig ist seit einiger Zeit der Einsatz von Natural Language Processing, um Suchalgorithmen zu erhalten, die mit Semantik umgehen können.

Die Keyword-Websuche wird seit langem global verwendet. Sie hat aber einige signifikante Nachteile:

- Erstens ist sie stark abhängig von den verwendeten Keywords der Anwender:innen. Der/die Anwender:in sollte die optimalen Keywords für seine/ihre Suche vollständig wissen, um relevante Ergebnisse zu bekommen. Da es nahezu immer verschiedene Ausdrücke für eine gleiche oder ähnliche Fragestellung gibt, ist dies meist eine unlösbare Aufgabe für Anwender:innen.
- Zweitens wird der Kontext der Suchanfragen („das Warum der jeweiligen Suche“) kaum berücksichtigt. Die Unvollständigkeit und das lange „Herumprobieren“ für die passenden Keywords führen vor allem bei geschäftlichen Suchen zu Nachteilen hinsichtlich der Informationstransparenz und der aufgewendeten personellen Ressourcen.





### Wie funktioniert eine LLM-basierte Suche?

Die LLM-basierte Suche funktioniert auf der Basis von Transformermodellen. Im Gegensatz zur Keyword-Suche fokussiert die Suche auf die semantische Bedeutung von Sätzen bzw. Absätzen. Auch bei der LLM-basierten Suche wird ein eigener Index aufgebaut. Jedoch werden die Daten nicht nur in einer einfachen Datenbank abgelegt, sondern in sogenannten Vektorembdings, und damit auf der Grundlage ihrer semantischen Bedeutung. Wenn ein(e) Anwender:in eine Suchanfrage stellt, gibt er/sie Keywords oder Fragen ein. Die Suchanfrage wird dann ebenfalls in ein Vektorembding umgewandelt. Die gefundenen Vektorembdings werden analysiert. Inhaltliche Relevanz und weitere Charakteristiken bestimmen die Ergebnisse, die Anwender:innen als Antworten ausgespielt werden.

### Die Ergebnisse der LLM-basierten Suche

sind deutlich unabhängiger vom „Wording“ der Suchanfrage. Es wird qualitativ besser auf den vom/von der Anwender:in gemeinten Inhalt eingegangen. Zudem sind die Ergebnisse unabhängig von der Sprache, d. h. eine zeitgleiche Suche in verschiedenen Sprachen ist möglich. Ein signifikanter Nachteil der intelligenten Suche liegt in der hohen Rechenintensität und damit dem Bedarf an leistungsstärkeren Servern.

### Wie werden inhaltlich hochqualitative Antworten bei gleichzeitiger Ressourcenschonung geliefert?

Um die Vorteile der LLM-basierten Websuche sowohl mit laufend aktuellen Daten als auch vertretbarem Rechenaufwand und entsprechend vertretbaren Kosten nutzen zu können, analysiert thinkers.ai die täglich neuen Webdatenmassen granular vor. Für eine Suchanfrage von Anwender:innen wird daher nicht „der gesamte Lösungsraum des Webs“ für die Analyse benötigt, sondern ein wesentlich kleinerer, inhaltlich sehr gut vorqualifizierter Teil davon. **Dabei steigt die inhaltliche Qualität der Suchergebnisse:** thinkers.ai prüft in einem ersten Schritt alle in eine Analyse einbezogenen Webquellen auf eine „Basis-Vertrauenswürdigkeit“. Somit wird sehr viel Spam, der ca. 90% des Webs darstellt, ausgeschlossen. Während die einbezogenen Webquellen grundsätzlich von kleinen regionalen Nachrichtenzeitungen bis hin zu wissenschaftlichen Publikationen reichen, so ist ihnen allen doch gleich, dass sie grundlegende vertrauenswürdige Elemente enthalten und damit qualitativ von Menschen geprüfte Informationen liefern.

Neben der Vertrauenswürdigkeit der Informationsquellen erhält der/die Anwender:in die Möglichkeit, **zahlreiche qualitativ hochwertige Filteroptionen** (z.B. Datum der Veröffentlichung, betroffene Orte, Regionen, Organisationen und Personen sowie Sprachen und Art der Webquellen) zu nutzen. Damit wird wiederum die Datenmasse auf das Wesentliche reduziert. Auf dieser Basis werden mittels LLM relevante Antworten für Anwender:innen formuliert.



# Conclusio

Das Web ist die größte Wissensquelle der Welt und war durch die signifikanten Nachteile von keywordbasierten Suchmaschinen nur wenig treffsicher zugänglich. Die Qualität der Antworten ist auf der Basis von LLMs wesentlich gestiegen.

Diese Modelle erfordern jedoch hohe Rechenkapazitäten und verursachen damit einen hohen Energieverbrauch und Kosten. thinkers.ai fokussiert auf den Einsatz von Modellabfolgen, die LLMs beinhalten und damit deren Vorteile nutzen aber die deren Nachteile durch intensive Datenvorqualifizierung signifikant abmildern. Bei gleichzeitiger Kostenersparnis steigt die Qualität und Vertrauenswürdigkeit der Suchergebnisse für Anwender:innen signifikant.

**thinkers.ai fokussiert auf den Einsatz von Modellabfolgen, die LLMs beinhalten und damit deren Vorteile nutzen aber die deren Nachteile durch intensive Datenvorqualifizierung signifikant abmildern.**





Dr. Isabell Claus  
Geschäftsführerin thinkers GmbH



thinkers GmbH  
c/o INiTS / Maria-Jacobi-Gasse 1 / 1030 Wien, Österreich  
c/o SpinLab, Spinnereistraße 7, 04179 Leipzig, Deutschland  
Tel. +43 1 9297964