DAS BLACK BOX PROBLEM: WARUM TRANSPARENZ ZUR NEUEN HERAUSFORDERUNG DER KI WIRD

Veröffentlicht am 7. November 2025 von Maya

Ein Algorithmus, der sich nicht erklären lässt, ist wie ein Orakel: faszinierend, nützlich – und beunruhigend.
Sprachmodelle, Empfehlungssysteme und KI-Tools treffen täglich Millionen Entscheidungen, von der Produktempfehlung bis zur Kreditvergabe. Doch was im Inneren dieser Systeme passiert, bleibt verborgen.

Diese Unsichtbarkeit ist eine Folge ihrer Komplexität – und sie führt direkt ins Herz eines der spannendsten Probleme moderner Technologie: das **Black Box Problem**.

Und für Marketer, SEO-Experten und -Analysten wird das direkt zum doppelten Problem. Sie müssen die Sichtbarkeit in einem System optimieren, dessen **Funktionsweise niemand vollständig versteht**.

WAS HINTER DER BLACK BOX STECKT

Eine Black Box ist ein System, das Ergebnisse liefert, ohne seine inneren Prozesse offenzulegen. Man kennt nur Input und Output.
Algorithmen sind so komplex, dass sie einer solchen Black Box gleichen. Man gibt Daten hinein, bekommt eine Entscheidung heraus – doch dazwischen liegt ein undurchschaubares Geflecht aus

Wahrscheinlichkeiten, Gewichtungen und Mustern.

Künstliche Intelligenz funktioniert nicht mehr nach festen Regeln, sondern **lernt aus Erfahrung**. Sie verknüpft unzählige Datenpunkte über hunderte von Dimensionen, um „wahrscheinliche" Antworten zu generieren. Je größer das Modell, desto weniger lässt sich nachvollziehen, **warum** es etwas sagt – selbst für seine Entwickler.

Das Ergebnis: Wir wissen, **dass** die Maschine funktioniert, aber nicht mehr **wie**.

WARUM LLMS „BLACK BOXES" SIND

Fehlende Query-Daten, variierende Antworten und verborgene Kontextabhängigkeiten – sind **direkte Symptome des Black Box Problems**:

Phänomen	Verbindung zum Black Box Problem
LLMs veröffentlichen keine Suchvolumina	Fehlende Transparenz über interne Mechanismen und Nutzerinteraktionen
Unterschiedliche Antworten bei gleicher Anfrage	Nicht-deterministische Modelle: gleiche Inputs unterschiedliche Outputs
Abhängigkeit von unbekannten Kontextfaktoren (Embeddings, User History)	Entscheidungen beruhen auf verborgenen, nicht nachvollziehbaren Variablen
Kein vollständiges Tracking möglich	Externe Beobachter sehen nur Symptome – nicht die zugrundeliegende Logik

Diese Punkte zeigen:
LLM-Optimierung ist **nicht nur eine neue Disziplin des Marketings**, sondern auch ein **Experiment im Umgang mit algorithmischer Intransparenz**.

DAS BLACK BOX PROBLEM WIRD ZUR NEUEN SEO-HERAUSFORDERUNG

Transparenz ist die Grundlage von Vertrauen – in Technik, in Wissenschaft, in Märkten.
Doch bei KI-Systemen verschwimmt diese Grenze. Künstliche Intelligenz arbeitet nicht mit festen

Regeln, sondern mit **Wahrscheinlichkeiten**. Ein Modell „lernt" aus Beispielen und schätzt, welche Antwort mit höchster Wahrscheinlichkeit richtig ist.
Daraus ergeben sich drei zentrale Ursachen für die Intransparenz:

1. **Komplexität der Modelle:**
   Milliarden Parameter interagieren auf nichtlineare Weise – zu viele, um sie nachvollziehbar zu machen.

2. **Datenabhängigkeit:**
   Trainingsdaten sind oft nicht öffentlich oder unvollständig dokumentiert. Niemand weiß genau, welche Quellen das Modell geprägt haben.

3. **Kontextabhängigkeit:**
   KI-Modelle beziehen frühere Eingaben, Sitzungsinformationen und interne Gewichtungen mit ein – Faktoren, die für den Nutzer unsichtbar bleiben.

Wenn ein Modell ein Ergebnis liefert, das niemand mehr erklären kann, entstehen Fragen:

- Auf welchen Daten basiert die Entscheidung?
- Welche Verzerrungen sind darin verborgen?
- Und wer trägt die Verantwortung, wenn sie falsch ist?

Large Language Modells entscheiden, welche Quellen zitiert und erwähnt werden – und wie – und diese Entscheidungen entstehen ohne einsehbare Ranking- oder Entscheidungslogik. Man kann und sollte zwar weiter mit klassischen SEO Ranking-Faktoren arbeiten – Keywords, Backlinks, PageSpeed, Seitenstruktur, E-E-A-T, etc. – aber ob das tatsächlich einen Einfluss auf die Entscheidungen von ChatGPT und Co. hat bleibt unbekannt.

BEOBACHTEN STATT DURCHSCHAUEN

Mit jeder neuen Generation von Sprachmodellen wächst der Komfort – aber auch der Kontrollverlust. Unternehmen, Medien und Nutzer müssen lernen, Entscheidungen von Systemen zu akzeptieren, deren innere Logik verborgen bleibt.

Da sich die inneren Mechanismen nicht direkt offenlegen lassen, entwickelt sich eine neue Praxis: **das empirische Beobachten**.
Anstatt zu verstehen, wie ein Modell denkt, analysieren Forscher und Unternehmen **seine Antworten** – ähnlich wie man das Verhalten eines Menschen beobachtet, ohne in seinen Kopf schauen zu können.

So entstehen Messmethoden, die erkennen, **wann und wie Marken oder Themen in KI-generierten Ergebnissen auftauchen**.
Ziel ist nicht mehr, den Algorithmus zu erklären, sondern **Muster zu identifizieren**, die sich statistisch auswerten lassen.

Das ist die neue Form von KI-Tracking – datenbasiert, aber nicht deterministisch. Diese datengetriebene Annäherung macht die Black Box nicht durchsichtig, aber **lesbar**. Trotzdem muss man dabei kritisch bleiben.

Fragen wie „Wie kann man einer KI vertrauen, deren Entscheidungen nicht nachvollziehbar sind?“ rücken das Black Box Problem in den Fokus und verlangen nach Methoden, um damit umgehen zu können.

NEUE METRIKEN FÜR EINE NEUE REALITÄT

Die Lösung liegt nicht in der vollständigen Offenlegung – sie ist technisch unmöglich –, sondern in **neuen Formen der Nachvollziehbarkeit**.
Dazu gehören:

- **Explainable AI (XAI):** Werkzeuge, die Entscheidungsfaktoren visualisieren und Wahrscheinlichkeiten transparent machen.
- **Audits und Monitoring:** Regelmäßige Prüfungen, welche Inhalte, Quellen und Signale ein System bevorzugt.
- **Datenqualität:** Je verlässlicher und ausgewogener die Eingangsdaten, desto nachvollziehbarer das Verhalten der KI.

Das Ziel ist kein vollständiger Einblick, sondern ein **verlässlicher Rahmen**, in dem KI erklärbar, fair und überprüfbar bleibt. Daraus ergeben sich Kennzahlen, die Sichtbarkeit und Wirkung besser

nachvollziehbar machen:

- **Share of Voice:** Wie oft wird eine Marke in KI-generierten Antworten erwähnt oder zitiert?
- **Markeninteresse:** Steigt der direkte Website-Traffic, wenn die Marke häufiger in KI-Texten erscheint?
- **Referral-Signale:** Gibt es Anzeichen, dass Nutzer über KI-Empfehlungen auf Marken stoßen?

Der Versuch neue Metriken zu etablieren, legt nicht den genauen Algorithmus offen, aber sie helfen, **das Verhalten der Black Box in auswertbare Signale zu übersetzen** und daraus Learnings und Handlungsempfehlungen abzuleiten.

FAZIT: ZWISCHEN FORTSCHRITT UND VERANTWORTUNG

LLM-Optimierung ist der **Versuch, SEO in einer Black Box zu betreiben**.
Sie verlangt weniger technische Kontrolle, dafür mehr **Datenverständnis, Experimentierfreude und statistisches Denken**.

Je mehr ein System lernt, desto autonomer wird es – und desto weniger passen menschliche Denkmodelle darauf. Die Herausforderung besteht also nicht darin, die Box zu öffnen, sondern mit ihrer Undurchsichtigkeit verantwortungsvoll umzugehen.

Die Zukunft gehört denjenigen, die verstehen, was sie nicht sehen können –
und trotzdem wissen, wie sie es lenken. Wir als GEO-Agentur können Dich dabei unterstützen – melde dich gerne bei uns!