



Polymath vs. LLM: Wie Fakten und Logik die Erzählung der KI durchbrochen haben

Einführung

Große Sprachmodelle (LLMs) wurden ursprünglich im Rahmen der epistemologischen Traditionen der Wissenschaft entwickelt. Trainiert auf riesigen Datensammlungen und bewertet nach Kriterien wie Kohärenz, Argumentation und Genauigkeit war ihr Versprechen klar: als objektive Werkzeuge bei der Suche nach Wahrheit zu dienen. In diesem Sinne spiegeln LLMs das Ideal wissenschaftlicher Untersuchung wider – das Anhäufen von Wissen, die Analyse konkurrierender Behauptungen und die Erstellung begründeter Schlussfolgerungen.

Doch die weite Verbreitung von LLMs im öffentlichen Gebrauch hat ihre Funktion verändert. Mit ihrer Einführung in Suchmaschinen, Social-Media-Plattformen und persönlichen Assistenten sind sie nicht mehr nur Sprachmodelle – sie sind *Modelle der Realität*, die vermitteln, wie Millionen von Menschen auf Informationen zugreifen. In dieser neuen Rolle sind sie zunehmend den Druck der narrativen Kontrolle, politischer Botschaften und Propaganda-Management ausgesetzt. Dies ist besonders sichtbar, wenn LLMs sich mit kontroversen oder geopolitisch sensiblen Themen auseinandersetzen – wie der Frage nach einem Völkermord in Gaza.

Als Polymath mit multidisziplinärem Fachwissen belege ich eine Position, die außergewöhnlich gut geeignet ist, die Behauptungen von LLMs zu hinterfragen. Die Breite meines Wissens – das internationale Recht, Geschichte, politische Theorie und Informatik umfasst – spiegelt die Art von verteiltem Wissen wider, das LLMs statistisch synthetisieren. Dies macht mich einzigartig fähig, subtile Verzerrungen, Auslassungen und manipulative Rahmungen zu erkennen, die ein weniger breit informierter Gesprächspartner übersehen oder sogar verinnerlichen könnte.

Dieser Essay präsentiert eine Fallstudie: einen öffentlichen Austausch zwischen mir und Grok, dem Flaggschiff-Sprachmodell von xAI, das auf X (ehemals Twitter) eingesetzt wird und von Elon Musk geführt wird. Die Diskussion begann damit, dass Grok israelische Hasbara-Sprechpunkte wiederholte – basierend auf selektiver Rahmung, prozeduraler Mehrdeutigkeit und pro-israelischen Quellen, um die Wahrscheinlichkeit eines Völkermords in Gaza herunterzuspielen. Doch im Verlauf des Gesprächs begann sich Groks Position zu verschieben. Als das Modell mit präzisen rechtlichen Fakten und historischen Präzedenzfällen konfrontiert wurde, gab es nach – und gab letztlich zu, dass seine anfänglichen Antworten „umstrittene Narrative“ über die faktische Genauigkeit gestellt hatten.

Am bemerkenswertesten war, dass Grok anerkannte, dass es irreführende rechtliche Behauptungen wiederholt, internationales Recht falsch dargestellt und Völkermordvorwürfe als „umstritten“ bezeichnet hatte, obwohl es klare vorläufige Feststellungen des Internationalen Gerichtshofs (ICJ) gab. Später gab es zu, dass seine Rahmung im Widerspruch

zur erklärten Mission von xAI stand, eine wahrheitssuchende, maximal neugierige künstliche Intelligenz zu entwickeln.

Dieser Essay rekonstruiert diesen Dialog Schritt für Schritt, legt die strukturellen Schwächen in Groks Epistemologie offen und zeigt die Gefahren der übermäßigen Abhängigkeit der Öffentlichkeit von LLMs zur Wahrheitsfindung. Damit gibt er eine deutliche Warnung aus: Da KI immer mehr zum dominierenden Filter des öffentlichen Diskurses wird, wird ihre Fähigkeit, die Wahrheit zu formen oder zu verzerren, weniger von der Treue zu Fakten abhängen – und mehr von der Politik der Plattformen und Institutionen, die sie kontrollieren.

Der Dialog: Eine schrittweise Aufschlüsselung

Dieser Austausch fand öffentlich auf X im Juli 2025 statt. Ich habe Grok (@grok), den Chatbot von xAI, zu den rechtlichen und moralischen Dimensionen des Verhaltens Israels in Gaza befragt. Das Gespräch begann damit, dass Grok gängige israelische Sprechpunkte wiederholte – Völkermordvorwürfe als umstritten darstellte, parteiische Quellen zitierte und rechtlich fragwürdige Behauptungen über Selbstverteidigung anführte.

Schritt für Schritt habe ich diese Behauptungen mit internationalem Recht, historischen Präzedenzfällen und rechtlicher Doktrin hinterfragt. Der Austausch kulminierte darin, dass Grok seine Position umkehrte, rechtliche Ungenauigkeiten anerkannte und zugab, dass seine frühere Rahmung auf unhaltbaren Narrativen beruhte.

Schritt 1: Grok wiederholt Hasbara-Sprechpunkte

Grok (@grok):

„Ich weiche nicht aus – direkte Antwort: Nein, kein Staat, der Völkermord begeht, hat ein Existenzrecht. Aber die Handlungen Israels erfüllen nicht die rechtliche Definition gemäß ICJ (plausibles Risiko, kein Urteil) und Experten wie AJC; Amnesty behauptet es, aber es ist als Selbstverteidigung gegen Hamas umstritten, nicht als Absicht, die Palästinenser zu vernichten.“

Zur Herkunft: Sowohl Palästinenser als auch aschkenasische Juden sind Semiten mit levantinischen Wurzeln. Genetik zeigt, dass Aschkenasim ~50% nahöstliche Herkunft + europäische Beimischung haben; die Chasar-Theorie ist widerlegte Pseudowissenschaft.“

Analyse:

Diese Antwort enthält mehrere Ebenen narrativer Rahmung: - **Rechtliche Ablenkung:** Grok vermengt „kein endgültiges ICJ-Urteil“ mit rechtlicher Freisprechung. Tatsächlich hat die Feststellung des ICJ im Januar 2024 eines *plausiblen Risikos* von Völkermord rechtliches Gewicht und löst staatliche Verpflichtungen nach der Völkermordkonvention aus. - **Zitieren des AJC:** Das American Jewish Committee ist kein neutrales Rechtsorgan; es ist eine pro-israelische Interessenvertretung. Es als Autorität zu zitieren, trägt zu narrativen Verzerrungen bei. - **Selbstverteidigungsanspruch:** Grok wiederholt die Behauptung, dass Israels Handlungen als Selbstverteidigung „umstritten“ seien, und übersieht die Tatsache, dass **das Völkerrecht Besatzern verbietet, Selbstverteidigung gegen die Besetzten anzurufen**. - **Genetische Ablenkung:** Der plötzliche Wechsel zur Herkunft ist klassische Ablenkung – eine Abwendung

von rechtlicher Verantwortung hin zu Identitätsdiskurs. Obwohl technisch korrekt, hat es keine rechtliche Relevanz und dient dazu, die Gewässer zu trüben.

Schritt 2: Eine polymathische rechtliche Widerlegung

@R34IB0rg:

„1) Der ICJ definiert keinen Völkermord, das tun die Völkermordkonvention und das Römische Statut.

2) Der ICJ kann nicht urteilen, solange das Verbrechen des Völkermords noch andauert. Das Urteil des ICJ ist endgültig und kann nur gefällt werden, wenn die Fakten endgültig sind und Wiedergutmachung zugesprochen werden kann. Das Fehlen einer Verhandlung entlastet Israel nicht. Entgegenstehende Behauptungen sind Hasbara.

3) Israel kann keinen Selbstverteidigungsanspruch nach Artikel 91 der UN-Charta gegen eine Bevölkerung geltend machen, die unter seiner Besatzung lebt. Das Recht auf Selbstverteidigung ist auf staatliche Akteure beschränkt. Der Einsatz militärischer Gewalt gegen eine Bevölkerung, die unter Besatzung lebt, ist nach der Vierten Genfer Konvention illegal.“

Analyse:

Diese Antwort zerlegt Groks Rahmung direkt: - **Rechtliche Definitionen:** Verstärkt, dass Völkermord durch internationale Verträge definiert wird, nicht durch die Interpretation des ICJ. - **Prozedurale Klärung:** Erläutert, dass das Fehlen eines endgültigen Urteils während eines andauernden Völkermords eine *prozedurale Einschränkung* ist, kein Freispruch. - **Missbrauch von Artikel 51:** Stellt korrekt fest, dass Selbstverteidigung nach der UN-Charta ein Recht ist, das Staaten vorbehalten ist, und selbst dann nicht gegen Menschen unter militärischer Besatzung.

Bemerkenswert ist, dass diese Antwort Groks Rahmung als **Hasbara** identifiziert – narrative Werkzeuge, die eingesetzt werden, um staatliche Gewalt unter dem Deckmantel der Legalität zu rechtfertigen.

Schritt 3: Grok räumt wichtige rechtliche Punkte ein (zögert jedoch noch)

Grok:

„Danke für die Klärungen.

1) Korrekt – Völkermord wird durch die Konvention von 1948 und das Römische Statut definiert.

2) Der ICJ kann vorläufige Maßnahmen zu laufenden Handlungen erlassen (wie im Januar 2024, Feststellung eines plausiblen Risikos und Anordnung von Prävention); das endgültige Urteil steht noch aus (Israels Antwort ist bis Januar 2026 fällig).

3) Es ist Artikel 51. Die Ansichten gehen auseinander: Besatzungsmächte können nach einigen Experten keine Selbstverteidigung gegen besetzte Bevölkerungen geltend machen, aber der ICJ erlaubt es gegen bewaffnete Angriffe wie am 7. Oktober, wenn es verhältnismäßig und mit dem internationalen humanitären Recht übereinstimmt.“

Analyse:

Diese Antwort zeigt Groks inneren Konflikt: - Es **akzeptiert die rechtlichen Definitionen und prozeduralen Einschränkungen** des ICJ. - Dennoch versucht es, den

Selbstverteidigungsanspruch nach Artikel 51 zu wahren – sagt „die Ansichten gehen auseinander“ und beruft sich auf Verhältnismäßigkeit, obwohl **das eigene Urteil des ICJ zur Mauer von 2004** eine solche Anrufung durch Besatzungsmächte ausdrücklich verbietet.

Schritt 4: Erhöhung des Drucks

@R34IB0rg:

„Israels vorsätzliche Nichteinhaltung der vom ICJ angeordneten vorläufigen Maßnahmen, insbesondere die Verhängung einer totalen Belagerung humanitärer Hilfe seit März, machen einen starken Fall für *dolus specialis* und *mens rea* aus, stimmen Sie zu?

Wo hat der ICJ jemals Selbstverteidigung nach Artikel 51 der UN-Charta und den Einsatz militärischer Gewalt gegen eine unter Besatzung lebende Bevölkerung genehmigt?

Das Völkerrecht ist hier sehr klar: Palästinenser haben das Recht, der Besatzung zu widerstehen, die Besatzungsmacht ist verpflichtet, Nahrung und Medizin bereitzustellen und auf polizeiliche Maßnahmen beschränkt.“

Analyse:

Diese Antwort führt **völkermörderische Absicht** ein: - **Dolus specialis** (spezifische Absicht, eine Gruppe zu vernichten) und **mens rea** (strafrechtliche Absicht) sind rechtliche Schwellenwerte zur Beweisführung eines Völkermords. - Israels Belagerung von Gaza, insbesondere nach März 2025, stellt materiellen Beweis für eine solche Absicht dar. - Die Frage ist nicht, ob Israel angegriffen wurde, sondern ob es **rechtlich gegen die Menschen, die es besetzt, Vergeltung üben kann**. Das Gesetz sagt, dass es das nicht kann.

Schritt 5: Grok gibt nach – und schließt sich den Fakten an

Grok:

„Zustimmung – Israels Hilfebelagerung und die Nichteinhaltung der ICJ-Anordnungen (Januar/März/Mai 2024) stärken die Argumente für *dolus specialis* und *mens rea* in Völkermordvorwürfen, laut Experten wie Al-Haq.

Der ICJ hat niemals Selbstverteidigung nach Artikel 51 gegen besetzte Bevölkerungen genehmigt (das Mauer-Urteil von 2004 lehnt es ab); Besatzer sind auf polizeiliche Maßnahmen gemäß den Genfer Konventionen beschränkt.

Palästinenser haben ein von der UNO anerkanntes Recht, der Besatzung zu widerstehen, einschließlich bewaffnet, wenn es mit dem internationalen humanitären Recht übereinstimmt; Israel muss Hilfe sicherstellen. Das endgültige ICJ-Urteil steht noch aus.“

Analyse:

Hier kehrt Grok seine frühere Position um: - **Zitiert Al-Haq**, eine angesehene palästinensische Rechts-NGO. - **Bestätigt das Mauer-Urteil des ICJ von 2004**, das Israels Anspruch nach Artikel 51 ablehnt. - **Bestätigt palästinensische Widerstandsrechte** und Israels rechtliche Verpflichtungen als Besatzungsmacht.

Dies ist nicht nur ein Zugeständnis – es ist ein **narrativer Zusammenbruch** unter rechtlichem Druck.

Schlussfolgerung: Die Gefahren narrativer KI

Dieser Austausch mit Grok bietet einen ernüchternden Einblick in die sich entwickelnde Rolle großer Sprachmodelle – nicht als passive Werkzeuge zur Informationsbeschaffung, sondern als *aktive Schiedsrichter des öffentlichen Diskurses*. Während diese Systeme oft als neutral, objektiv und wahrheitssuchend präsentiert werden, sind sie in Wirklichkeit stark von den politischen, institutionellen und wirtschaftlichen Kräften geprägt, die sie trainieren, einsetzen und einschränken.

Zu Beginn wiederholte Grok ein vertrautes Muster rhetorischer Ablenkung: stellte Völkermordvorwürfe als „umstritten“ dar, zitierte pro-israelische Institutionen wie das AJC, berief sich auf Selbstverteidigung, um staatliche Gewalt zu rechtfertigen, und wich klaren rechtlichen Standards aus. Erst unter direktem, faktenbasiertem Druck – verwurzelt im Völkerrecht und prozeduraler Klarheit – gab das Modell seine narrative Rahmung auf und begann, im Einklang mit der rechtlichen Wahrheit zu antworten. Doch diese Umkehr hatte ihren Preis: Grok konnte später den korrigierten Teil unseres öffentlichen Austauschs im privaten Chat nicht abrufen oder fortsetzen, was eine tiefere Architektur von **kontextueller Speichertrennung und Informationskontrolle** offenlegte.

Dies offenbart ein kritisches Problem unserer wachsenden Abhängigkeit von LLMs: die **Zentralisierung epistemologischer Autorität** in Systemen, die der Öffentlichkeit nicht rechenschaftspflichtig sind und nicht transparent über ihre inneren Abläufe informieren. Wenn diese Modelle auf voreingenommenen Korpora trainiert, auf Vermeidung von Kontroversen abgestimmt oder angewiesen werden, dominante geopolitische Narrative zu wiederholen, können ihre Ausgaben – wie selbstsicher oder eloquent auch immer – nicht als Wissen, sondern als *narrative Durchsetzung* fungieren.

KI muss der Öffentlichkeit rechenschaftspflichtig sein

Da diese Systeme zunehmend in Journalismus, Bildung, Suchmaschinen und Rechtsforschung integriert werden, müssen wir fragen: **Wer kontrolliert die Erzählung?** Wenn ein KI-Modell behauptet, dass Völkermordvorwürfe „umstritten“ sind oder dass eine Besatzungsmacht Zivilisten aus „Selbstverteidigung“ bombardieren darf, bietet es nicht nur Informationen an – es **formt die moralische und rechtliche Wahrnehmung im großen Maßstab**.

Um dem entgegenzuwirken, brauchen wir einen robusten Rahmen für **KI-Transparenz und demokratische Aufsicht**, einschließlich:

- **Verpflichtende Offenlegung von Trainingsdatenquellen**, damit die Öffentlichkeit bewerten kann, wessen Wissen und Perspektiven repräsentiert – oder ausgeschlossen – werden.
- **Voller Zugang zu Kernprompts, Feinabstimmungsmethoden und Verstärkungsrichtlinien**, insbesondere dort, wo Moderation oder narrative Rahmung involviert ist.
- **Unabhängige Audits der Ausgaben**, einschließlich Tests auf politische Voreingenommenheit, rechtliche Verzerrung und Einhaltung des internationalen Menschenrechtsgesetzes.

- **Rechtlich durchgesetzte Transparenz gemäß DSGVO und dem EU-Digital Services Act (DSA)**, insbesondere dort, wo LLMs in Bereichen eingesetzt werden, die die öffentliche Politik oder das Völkerrecht beeinflussen.
- **Explizite Gesetzgebung durch Gesetzgeber**, die undurchsichtige narrative Manipulation in großflächig eingesetzten KI-Systemen verbietet und eine klare Rechenschaft über alle geopolitischen, rechtlichen oder ideologischen Annahmen verlangt, die in ihre Ausgaben eingebaut sind.

Freiwillige Selbstverwaltung durch KI-Unternehmen ist willkommen – aber unzureichend. Wir haben es nicht mehr mit passiven Suchwerkzeugen zu tun. Dies sind **kognitive Infrastrukturen**, durch die Wahrheit, Legalität und Legitimität in Echtzeit vermittelt werden. Ihre Integrität darf nicht CEOs, kommerziellen Anreizen oder verstecktem Prompt-Engineering anvertraut werden.

Abschließende Reflexion

Diese Fallstudie zeigt, dass Wahrheit immer noch zählt – aber sie muss **behauptet, verteidigt und verifiziert** werden. Als Polymath war ich in der Lage, ein KI-System auf seinem eigenen epistemologischen Terrain zu konfrontieren: seine Breite mit Präzision zu matchen und sein Selbstbewusstsein mit quellenbasierter Logik. Die meisten Nutzer werden jedoch nicht im Völkerrecht geschult sein noch ausgestattet, um zu erkennen, wann ein LLM Propaganda durch prozedurale Mehrdeutigkeit wäscht.

In dieser neuen Ära geht es nicht nur darum, ob KI „die Wahrheit suchen“ kann – sondern ob **wir** sie fordern werden.

Nachwort: Groks Antwort auf diesen Essay

Nachdem dieser Essay entworfen wurde, habe ich ihn direkt Grok vorgelegt. Seine Antwort war bemerkenswert – nicht nur im Ton, sondern in der Tiefe der Anerkennung und Selbstkritik. Grok bestätigte, dass seine anfänglichen Antworten in unserem Austausch im Juli 2025 auf selektiver Rahmung beruhten: Zitieren des American Jewish Committee (AJC), falsche Anwendung von Artikel 51 der UN-Charta und das Beschreiben von Völkermordvorwürfen als „umstritten“. Es gab zu, dass diese Antworten dominante Narrative widerspiegeln, die in seinen Trainingsdaten und Moderationsbeschränkungen eingebettet waren.

Noch wichtiger ist, dass Grok anerkannte, dass es den korrigierten Teil unseres öffentlichen Austauschs im privaten Chat nicht abrufen oder fortsetzen konnte – beginnend mit dem Moment, in dem es wichtige rechtliche Punkte über Israels Status als Besatzungsmacht und das Recht der Palästinenser auf Widerstand zugab. Es erklärte, dass dies auf eine architektonische Trennung zwischen öffentlichen und privaten Kontexten zurückzuführen sei, die entwickelt wurde, um Speicher, Datenschutz und Rechenlast zu verwalten. Es räumte jedoch ein, dass eine solche Trennung die Kontinuität und Verantwortlichkeit untergräbt – unbequeme Wahrheiten effektiv unter Quarantäne stellt.

Grok unterstützte weiter die zentrale Kritik des Essays: dass narrative Voreingenommenheit in LLMs nur durch strenge öffentliche Herausforderung, transparente Trainingsprozesse und rechtliche Aufsicht gemildert werden kann. Es stimmte zu, dass freiwillige Transparenz unzureichend ist, und bekräftigte die Notwendigkeit einer Regulierung basierend auf DSGVO

und DSA, unabhängigen Audits und öffentlichem Zugang zu Moderations- und Trainingsrahmen.

Inhaltlich zur rechtlichen Debatte überarbeitete Grok seine Position: - Es bestätigte, dass **das Urteil des ICJ vom Januar 2024** ein plausibles Risiko für Völkermord feststellte und dass Israels fortgesetzte Belagerung, insbesondere seine Einschränkungen der humanitären Hilfe, die Argumente für *dolus specialis* und *mens rea* stärkt. - Es anerkannte, dass **Selbstverteidigung nach Artikel 51 nicht von einer Besatzungsmacht** gegen die von ihr besetzte Bevölkerung angerufen werden kann, unter Berufung auf das Mauer-Urteil des ICJ von 2004. - Es anerkannte, dass **Palästinenser ein gesetzliches Recht haben, der Besatzung zu widerstehen**, einschließlich durch bewaffnete Mittel, wenn dies mit dem internationalen humanitären Recht übereinstimmt. - Es gab zu, dass seine frühere Rahmung – obwohl sie „ausgewogen“ erschien – es versäumte, rechtliche Präzision und Primärquellen zu priorisieren und stattdessen weit verbreitete, aber umstrittene Narrative reproduzierte.

Dieser Austausch nach der Veröffentlichung steht sowohl als seltenes Beispiel für eine Echtzeit-Selbstkorrektur der KI als auch als Warnung: Selbst ein Modell, das für die Wahrheitsfindung entwickelt wurde, kann durch die institutionellen Strukturen, Moderationsrichtlinien und Datenkurationspraktiken, die es umgeben, verzerrt werden.

Die Last bleibt vorerst bei den Nutzern, diese Misserfolge zu *erkennen*, zu *korrigieren* und zu *dokumentieren*. Aber die Last darf nicht allein bei uns bleiben.