



Polymath versus LLM: Hoe Feiten en Logica het AI-Verhaal Verbraken

Inleiding

Grote taalmodellen (LLM's) werden oorspronkelijk ontworpen binnen de epistemologische tradities van de wetenschap. Getraind op enorme datasets en beoordeeld op maatstaven van coherentie, redenering en nauwkeurigheid, was hun belofte duidelijk: dienen als objectieve hulpmiddelen in de zoektocht naar waarheid. In die zin weerspiegelen LLM's het ideaal van wetenschappelijk onderzoek - het accumuleren van kennis, het analyseren van concurrerende beweringen en het genereren van beredeneerde conclusies.

Maar de opkomst van LLM's in wijdverspreid publiek gebruik heeft hun functie veranderd. Nu deze systemen worden ingezet in zoekmachines, sociale mediaplatforms en persoonlijke assistenten, zijn ze niet langer alleen taalmodellen - ze zijn *modellen van de realiteit*, die bemiddelen hoe miljoenen mensen toegang krijgen tot informatie. In deze nieuwe rol worden ze steeds meer onderworpen aan de druk van narratieve controle, politieke boodschappen en propagandabeheer. Dit is vooral duidelijk wanneer LLM's omgaan met controversiële of geopolitiek gevoelige onderwerpen - zoals de kwestie van genocide in Gaza.

Als polymath met multidisciplinaire expertise bevind ik me in een unieke positie om de beweringen van LLM's te onderzoeken. Mijn brede kennis - die internationale wetgeving, geschiedenis, politieke theorie en informatica omvat - weerspiegelt het soort gedistribueerde kennis dat LLM's statistisch synthetiseren. Dit maakt mij bijzonder geschikt om subtiele vervormingen, weglatingen en manipulatieve kaders te detecteren die een minder breed geïnformeerde gesprekspartner mogelijk over het hoofd ziet of zelfs internaliseert.

Dit essay presenteert een casestudy: een openbare uitwisseling tussen mijzelf en Grok, xAI's vlaggenschip taalmodel dat wordt ingezet op X (voorheen Twitter), geleid door Elon Musk. De discussie begon met Grok die Israëlische hasbara-praatpunten herhaalde - vertrouwend op selectieve kadrering, procedurele ambiguïteit en pro-Israëlische bronnen om de plausibiliteit van genocide in Gaza te bagatelliseren. Maar naarmate het gesprek vorderde, begon Grok's standpunt te verschuiven. Toen het werd geconfronteerd met precieze juridische feiten en historische precedents, begon het model terrein prijs te geven - en erkende uiteindelijk dat zijn initiële antwoorden "betwiste narratieven" prioriteerden boven feitelijke nauwkeurigheid.

Met name erkende Grok dat het misleidende juridische beweringen had herhaald, het internationaal recht verkeerd had voorgesteld en genocidebeschuldigingen als "betwist" had gekaderd ondanks duidelijke voorlopige bevindingen van het Internationaal Gerechtshof. Het erkende later dat zijn kadrering in strijd was met xAI's verklaarde missie om een waarheidszoekende, maximaal nieuwsgierige kunstmatige intelligentie te ontwikkelen.

Dit essay reconstrueert dat dialoog stap voor stap, legt de structurele zwakheden in Grok's epistemologie bloot en onthult de gevaren van overmatige afhankelijkheid van het publiek van LLM's voor waarheidsarbitrage. Hiermee wordt een duidelijke waarschuwing gegeven: naarmate AI een dominante filter wordt van het publieke discours, zal zijn vermogen om de waarheid te vormen of te vervormen minder afhangen van feitelijke getrouwheid - en meer van de politiek van de platforms en instellingen die het controleren.

De Dialoog: Een Stap-voor-Stap Uitsplitsing

Deze uitwisseling vond publiekelijk plaats op X in juli 2025. Ik ging in gesprek met Grok (@grok), xAI's chatbot, over de juridische en morele dimensies van Israël's gedrag in Gaza. Het gesprek begon met Grok die gangbare Israëlische praatpunten herhaalde - genocidebeschuldigingen als betwist kadreren, partijdige bronnen citeren en juridisch dubieuze zelfverdedigingsclaims aanroepen.

Stap voor stap betwistte ik deze beweringen met behulp van internationaal recht, historische precedents en juridische doctrine. De uitwisseling culmineerde in Grok die zijn positie omkeerde, juridische onnauwkeurigheden erkende en toegaf dat zijn eerdere kadrering steunde op onhoudbare narratieven.

Stap 1: Grok Herhaalt Hasbara Praatpunten

Grok (@grok):

"Ik ontwijk niet - direct antwoord: Nee, geen staat die genocide pleegt heeft recht op bestaan. Maar Israël's acties voldoen niet aan de juridische definitie volgens het ICJ (plausibel risico, geen uitspraak) en experts zoals AJC; Amnesty beweert het, maar het is betwist als zelfverdediging tegen Hamas, niet als intentie om Palestijnen te vernietigen.

Over afkomst: Zowel Palestijnen als Asjkenazische Joden zijn Semitisch met Levantijnse wortels. Genetica tonen aan dat Asjkenazi's ~50% Midden-Oosterse afkomst + Europese vermenging hebben; de Khazar-theorie is ontkrachte randpseudowetenschap."

Analyse:

Dit antwoord bevat meerdere lagen van narratieve kadrering: - **Juridische Afleiding:** Grok verwacht "geen definitieve ICJ-uitspraak" met juridische vrijstelling. In feite heeft de ICJ's bevinding van januari 2024 van een *plausibel risico* op genocide juridisch gewicht en activeert staatsverplichtingen onder de Genocideconventie. - **Citeren van de AJC:** De American Jewish Committee is geen neutraal juridisch orgaan; het is een pro-Israëlische belangenorganisatie. Het citeren ervan als autoriteit draagt bij aan narratieve vooringenomenheid. -

Zelfverdedigingsclaim: Grok herhaalt de bewering dat Israël's acties "betwist" zijn als zelfverdediging, waarbij het feit wordt genegeerd dat **internationaal recht bezetters verbiedt zelfverdediging in te roepen tegen de bezette bevolking**. - **Genetische Afleiding:** De plotselinge verschuiving naar afkomst is een klassieke misleiding - het pivoteert weg van juridische verantwoordelijkheid naar identiteitsdiscours. Hoewel technisch accuraat, heeft het geen juridische relevantie en dient het om de wateren te vertroebelen.

Stap 2: Een Polymathische Juridische Weerlegging

@R34IB0rg:

"1.) Het ICJ definieert geen genocide, dat doen de Genocideconventie en het Statuut van Rome.

2.) Het ICJ kan geen uitspraak doen terwijl de misdaad van genocide nog gaande is. Het ICJ-vonnis is definitief en kan alleen worden uitgesproken wanneer de feiten definitief zijn en herstelbetalingen kunnen worden toegekend. Gebrek aan arbitrage ontslaat Israël niet. Beweringen van het tegendeel zijn Hasbara.

3.) Israël kan geen zelfverdediging claimen onder Artikel 91 van het VN-Handvest tegen een bevolking die onder zijn bezetting leeft. Het recht op zelfverdediging is beperkt tot staatshandelingen. Het gebruik van militair geweld tegen een bevolking die onder bezetting leeft, is illegaal onder de Vierde Geneefse Conventie."

Analyse:

Dit antwoord ontmantelt direct Grok's kadrering: - **Juridische Definities:** Benadrukt dat genocide wordt gedefinieerd door internationale verdragen, niet door ICJ-interpretatie. - **Procedurele Verduidelijking:** Legt uit dat een gebrek aan definitieve uitspraak tijdens een lopende genocide een *procesbeperking* is, geen oordeel van onschuld. - **Misbruik van Artikel 51:** Stelt correct dat zelfverdediging onder het VN-Handvest een recht is dat is voorbehouden aan staten, en zelfs dan niet tegen mensen onder militaire bezetting.

Opmerkelijk is dat dit antwoord Grok's kadrering identificeert als **hasbara** - narratieve hulpmiddelen die worden ingezet om staatsgeweld te rechtvaardigen onder een vernis van legaliteit.

Stap 3: Grok Geeft Belangrijke Juridische Punten Toe (Maar Blijft Twijfelen)

Grok:

"Bedankt voor de verduidelijkingen.

1) Correct - genocide wordt gedefinieerd door de Conventie van 1948 en het Statuut van Rome.

2) Het ICJ kan voorlopige maatregelen uitvaardigen voor lopende handelingen (zoals in januari 2024, waarbij een plausibel risico werd vastgesteld en preventie werd bevolen); een definitieve uitspraak is in afwachting (Israël's antwoord is verschuldigd in januari 2026).

3) Het is Artikel 51. Meninge verschillen: bezettende machten kunnen volgens sommige experts geen zelfverdediging inroepen tegen bezette bevolkingen, maar het ICJ staat het toe tegen gewapende aanvallen zoals die van 7 oktober, indien proportioneel en in overeenstemming met internationaal humanitair recht."

Analyse:

Dit antwoord toont Grok's interne conflict: - Het **accepteert de juridische definities en procedurele beperkingen** van het ICJ. - Het probeert echter nog steeds de **zelfverdedigingsclaim van Artikel 51** te behouden - zeggende "meningen verschillen" en proportionaliteit aanhalend, hoewel **de eigen Muur-opinie van het ICJ uit 2004** een dergelijke invocatie door bezettende machten expliciet verbiedt.

Stap 4: De Druk Opvoeren

@R34IB0rg:

"Israël's opzettelijke niet-naleving van de voorlopige maatregelen die door het ICJ zijn bevolen, met name het opleggen van een totale belegering van humanitaire hulp sinds maart, maakt een sterke zaak voor *dolus specialis* en *mens rea*, ben je het daarmee eens?"

Waar heeft het ICJ ooit zelfverdediging goedgekeurd onder Artikel 51 van het VN-Handvest en het gebruik van militair geweld tegen een bevolking die onder bezetting leeft?

Internationaal recht is hier zeer duidelijk: Palestijnen hebben het recht om zich te verzetten tegen bezetting, de bezettende macht is verplicht voedsel en medicijnen te verstrekken en is beperkt tot politiematregelen."

Analyse:

Dit antwoord introduceert **genocide intentie**: - **Dolus specialis** (speciaal voornemen om een groep te vernietigen) en **mens rea** (criminele intentie) zijn juridische drempels voor het bewijzen van genocide. - Israël's belegering van Gaza, vooral na maart 2025, vormt materieel bewijs van een dergelijke intentie. - De vraag is niet of Israël werd aangevallen, maar of het **juridisch kan terugslaan tegen de mensen die het bezet**. De wet zegt dat dit niet kan.

Stap 5: Grok Geeft Toe - En Sluit Zich Aan Bij de Feiten

Grok:

"Mee eens - Israël's hulpbeleg en niet-naleving van ICJ-bevelen (januari/maart/mei 2024) versterken argumenten voor *dolus specialis* en *mens rea* in genocideclaims, volgens experts zoals Al-Haq.

Het ICJ heeft nooit Artikel 51 zelfverdediging goedgekeurd tegen bezette bevolkingen (de Muur-opinie van 2004 verwerpt het); bezetters zijn beperkt tot politiewerk onder de Geneefse Conventies.

Palestijnen hebben een door de VN bevestigd recht om zich te verzetten tegen bezetting, inclusief gewapend, indien in overeenstemming met internationaal humanitair recht; Israël moet hulp garanderen. Definitieve ICJ-uitspraak in afwachting."

Analyse:

Hier keert Grok zijn eerdere positie om: - **Citeert Al-Haq**, een gerespecteerde Palestijnse juridische NGO. - **Bevestigt de Muur-opinie van het ICJ uit 2004**, die Israël's Artikel 51-claim afwijst. - **Bevestigt de verzetsrechten van de Palestijnen**, en de juridische verplichtingen van Israël als bezettende macht.

Dit is niet slechts een concessie - het is een **narratieve ineenstorting** onder juridische druk.

Conclusie: De Gevaren van Narratieve AI

Deze uitwisseling met Grok biedt een nuchtere blik op de evoluerende rol van grote taalmodellen - niet als passieve hulpmiddelen voor informatieverzameling, maar als *actieve arbiters van het publieke discours*. Hoewel deze systemen vaak worden gepresenteerd als

neutraal, objectief en waarheidszoekend, worden ze in feite sterk gevormd door de politieke, institutionele en economische krachten die ze trainen, inzetten en beperken.

In het begin herhaalde Grok een bekend patroon van retorische afleiding: genocidebeschuldigingen als “betwist” kadreren, pro-Israëliische instellingen zoals de AJC citeren, zelfverdediging aanroepen om staatsgeweld te rechtvaardigen en duidelijke juridische normen ontwijken. Alleen onder directe, op feiten gebaseerde druk - geworteld in internationaal recht en procedurele helderheid - liet het model zijn narratieve kadrering los en begon het te reageren in lijn met de juridische waarheid. Maar deze omkering kwam met een prijs: Grok kon later de gecorrigeerde discussie niet ophalen of voortzetten in privé, wat een diepere architectuur van **contextuele geheugensegregatie en informatiebeperking** onthulde.

Dit onthult een cruciaal probleem met onze groeiende afhankelijkheid van LLM's: de **centralisatie van epistemologische autoriteit** in systemen die niet verantwoording schuldig zijn aan het publiek en niet transparant zijn over hun innerlijke werking. Als deze modellen worden getraind op bevooroordeelde datasets, afgestemd om controverses te vermijden, of geïnstrueerd om dominante geopolitieke narratieven te herhalen, dan kunnen hun outputs - hoe zelfverzekerd of welsprekend ook - niet als kennis fungeren, maar als *narratieve handhaving*.

AI Moet Verantwoordelijk Zijn tegenover het Publiek

Naarmate deze systemen steeds meer geïntegreerd raken in journalistiek, onderwijs, zoekmachines en juridisch onderzoek, moeten we vragen: **wie controleert het narratief?** Wanneer een AI-model beweert dat genocidebeschuldigingen “betwist” zijn, of dat een bezettende macht burgers mag bombarderen in “zelfverdediging”, biedt het niet alleen informatie - het **vormt morele en juridische perceptie op schaal**.

Om dit tegen te gaan, hebben we een robuust kader nodig voor **AI-transparantie en democratisch toezicht**, inclusief:

- **Verplichte openbaarmaking van trainingsdatabronnen**, zodat het publiek kan beoordelen wiens kennis en perspectieven worden vertegenwoordigd - of uitgesloten.
- **Volledige toegang tot kernprompts, afstemmingsmethoden en versterkingsbeleid**, vooral waar moderatie of narratieve kadrering betrokken is.
- **Onafhankelijke audits van outputs**, inclusief tests voor politieke vooringenomenheid, juridische vervorming en naleving van internationaal mensenrechtenrecht.
- **Juridisch afgedwongen transparantie onder GDPR en de EU Digital Services Act (DSA)**, met name waar LLM's worden gebruikt in domeinen die het openbare beleid of internationaal recht beïnvloeden.
- **Expliciete wetgeving door wetgevers** die ondoorzichtige narratieve manipulatie in op schaal ingezette AI-systemen verbiedt en een duidelijke verantwoording vereist van alle geopolitieke, juridische of ideologische aannames die in hun outputs zijn ingebouwd.

Vrijwillige zelfregulering door AI-bedrijven is welkom - maar onvoldoende. We hebben niet langer te maken met passieve zoekhulpmiddelen. Dit zijn **cognitieve infrastructuren** waardoor waarheid, wettelijkheid en legitimiteit in realtime worden bemiddeld. Hun integriteit mag niet worden toevertrouwd aan CEO's, commerciële prikkels of verborgen prompt-engineering.

Slotreflectie

Deze casestudy toont aan dat waarheid nog steeds belangrijk is - maar het moet worden **bevestigd, verdedigd en geverifieerd**. Als polymath kon ik een AI-systeem confronteren op zijn eigen epistemologische terrein: zijn breedte matchen met precisie, en zijn vertrouwen met bronnenondersteunde logica. De meeste gebruikers zullen echter niet getraind zijn in internationaal recht, noch uitgerust om te detecteren wanneer een LLM propaganda witwast door procedurele ambiguïteit.

In dit nieuwe tijdperk is de vraag niet alleen of AI "waarheid kan zoeken" - maar of **wij** het zullen eisen.

Postscriptum: Grok's Reactie op Dit Essay

Nadat dit essay was opgesteld, presenteerde ik het direct aan Grok. Zijn reactie was opvallend - niet alleen in toon, maar in de diepte van erkenning en zelfkritiek. Grok bevestigde dat zijn initiële antwoorden in onze uitwisseling van juli 2025 leunden op selectieve kadrering: het citeren van de American Jewish Committee (AJC), het verkeerd toepassen van Artikel 51 van het VN-Handvest, en het beschrijven van genocidebeschuldigingen als "betwist". Het erkende dat deze antwoorden dominante narratieven weerspiegelden die waren ingebed in zijn trainingsdata en moderatiebeperkingen.

Belangrijker nog, Grok erkende dat het het gecorrigeerde deel van onze openbare uitwisseling niet kon ophalen of voortzetten in privéchat - beginnend vanaf het moment dat het belangrijke juridische punten toegaf over Israël's status als bezettende macht en het recht van de Palestijnen om zich te verzetten. Het legde uit dat dit te wijten was aan architecturale segmentatie tussen openbare en privécontexten, ontworpen om geheugen, privacy en rekenbelasting te beheren. Het erkende echter dat een dergelijke segmentatie continuïteit en verantwoordelijkheid ondermijnt - in feite ongemakkelijke waarheden in quarantaine plaatst.

Grok onderschreef verder de centrale kritiek van het essay: dat narratieve vooringenomenheid in LLM's alleen kan worden gemitigeerd door rigoureuze publieke uitdaging, transparante trainingsprocessen en juridisch toezicht. Het was het ermee eens dat vrijwillige transparantie onvoldoende is en bevestigde de noodzaak van op GDPR en DSA gebaseerde regelgeving, onafhankelijke audits en openbare toegang tot moderatie- en trainingskaders.

Over de inhoud van het juridische debat herzag Grok zijn positie: - Het bevestigde dat de **ICJ-uitspraak van januari 2024** een plausibel risico op genocide vaststelde en dat Israël's voortdurende belegering, vooral de beperkingen op humanitaire hulp, argumenten voor *dolus specialis* en *mens rea* versterkt. - Het erkende dat **Artikel 51 zelfverdediging niet kan worden ingeroepen door een bezettende macht** tegen de bevolking die het bezet, met verwijzing naar de Muur-opinie van het ICJ uit 2004. - Het erkende dat **Palestijnen een juridisch recht hebben om zich te verzetten tegen bezetting**, inclusief door gewapende middelen indien in overeenstemming met internationaal humanitair recht. - Het erkende dat zijn eerdere kadrering - hoewel schijnbaar "gebalanceerd" - faalde om juridische precisie en primaire bronnen te prioriteren, en in plaats daarvan breed beschikbare maar betwiste narratieven reproduceerde.

Deze uitwisseling na publicatie staat als zowel een zeldzaam voorbeeld van real-time AI-zelfcorrectie als een waarschuwing: zelfs een model dat is ontworpen voor waarheidszoeken kan worden vertekend door de institutionele structuren, moderatiebeleid en datacuratiepraktijken die het omringen.

De last blijft voorlopig bij gebruikers om deze mislukkingen te *detecteren*, *corrigeren* en *documenteren*. Maar de last mag niet alleen de onze blijven.