Polymath vs. LLM: Hvordan fakta og logik brød AI-narrativet

Introduktion

Store sprogmodeller (LLMs) blev oprindeligt udtænkt inden for de epistemologiske traditioner i videnskaben. Trænet på enorme datakorpora og evalueret ud fra målinger af sammenhæng, ræsonnement og nøjagtighed var deres løfte klart: at fungere som objektive værktøjer i jagten på sandheden. I denne forstand afspejler LLMs idealet om videnskabelig undersøgelse – akkumulering af viden, analyse af konkurrerende påstande og generering af begrundede konklusioner.

Men udbredelsen af LLMs til bred offentlig brug har ændret deres funktion. Efterhånden som disse systemer udrulles på tværs af søgemaskiner, sociale medieplatforme og personlige assistenter, er de ikke længere bare sproglige modeller – de er *modeller af virkeligheden*, der formidler, hvordan millioner af mennesker får adgang til information. I denne nye rolle udsættes de i stigende grad for pres fra narrativ kontrol, politiske budskaber og propagandastyring. Dette er især tydeligt, når LLMs håndterer kontroversielle eller geopolitisk følsomme emner – såsom spørgsmålet om folkedrab i Gaza.

Som polymath med multidisciplinær ekspertise indtager jeg en position, der er usædvanligt velegnet til at udfordre LLMs' påstande. Min brede viden – der spænder over international ret, historie, politisk teori og datalogi – afspejler den type distribuerede viden, som LLMs statistisk syntetiserer. Dette gør mig unikt kvalificeret til at opdage subtile forvrængninger, udeladelser og manipulerende rammer, som en mindre bredt informeret samtalepartner måske overser eller endda internaliserer.

Denne artikel præsenterer en casestudie: en offentlig udveksling mellem mig og Grok, xAI's flagskibssprogmodel, der er udrullet på X (tidligere Twitter), ledet af Elon Musk. Diskussionen begyndte med, at Grok gentog israelske hasbara-talepunkter – baseret på selektiv ramme, proceduremæssig uklarhed og pro-israelske kilder for at nedtone sandsynligheden for folkedrab i Gaza. Men efterhånden som samtalen skred frem, begyndte Groks holdning at skifte. Da modellen blev konfronteret med præcise juridiske fakta og historiske præcedenser, begyndte den at give efter – og indrømmede til sidst, at dens oprindelige svar havde prioriteret "omdiskuterede narrativer" over faktuel nøjagtighed.

Mest bemærkelsesværdigt anerkendte Grok, at den havde gentaget vildledende juridiske påstande, fejlagtigt repræsenteret international ret og indrammet anklager om folkedrab som "omdiskuterede" på trods af klare foreløbige fund fra Den Internationale Domstol (ICJ). Den indrømmede senere, at dens ramme var i modstrid med xAI's erklærede mission om at udvikle en sandhedssøgende, maksimalt nysgerrig kunstig intelligens.

Denne artikel rekonstruerer denne dialog trin for trin, afslører de strukturelle svagheder i Groks epistemologi og belyser farerne ved offentlighedens overdrevne afhængighed af LLMs til sandhedsformidling. Dermed udsteder den en skarp advarsel: Efterhånden som AI bliver en dominerende filter for offentlig diskurs, vil dens evne til at forme eller forvrænge sandheden afhænge mindre af faktuel troskab – og mere af de platforme og institutioners politik, der kontrollerer den.

Dialogen: En trinvis nedbrydning

Denne udveksling fandt sted offentligt på X i juli 2025. Jeg engagerede Grok (@grok), xAI's chatbot, om de juridiske og moralske dimensioner af Israels adfærd i Gaza. Samtalen begyndte med, at Grok

gentog almindelige israelske talepunkter – indrammede anklager om folkedrab som omdiskuterede, citerede partiske kilder og fremkaldte juridisk tvivlsomme påstande om selvforsvar.

Trin for trin udfordrede jeg disse påstande ved hjælp af international ret, historiske præcedenser og juridisk doktrin. Udvekslingen kulminerede i, at Grok vendte sin holdning, anerkendte juridiske unøjagtigheder og indrømmede, at dens tidligere ramme hvilede på uholdbare narrativer.

Trin 1: Grok gentager hasbara-talepunkter

> **Grok (@grok):**
> "Jeg undviger ikke – direkte svar: Nej, ingen stat, der begår folkedrab, har ret til at eksistere. Men Israels handlinger opfylder ikke den juridiske definition ifølge ICJ (plausibelt risiko, ingen dom) og eksperter som AJC; Amnesty hævder det, men det er omdiskuteret som selvforsvar mod Hamas, ikke en hensigt om at ødelægge palæstinenserne.
>
> Om oprindelse: Både palæstinensere og ashkenazi-jøder er semitiske med levantinske rødder. Genetik viser, at ashkenazier har ~50% mellemøstlig oprindelse + europæisk blanding; khazar-teorien er afkræftet pseudovidenskab."

Analyse:
Dette svar indeholder flere lag af narrativ ramme: - **Juridisk afledning**: Grok sammenblander "ingen endelig ICJ-dom" med juridisk fritagelse. Faktisk har ICJ's januar 2024-fund om et *plausibelt risiko* for folkedrab juridisk vægt og udløser statslige forpligtelser under Folkedrabskonventionen. - **Citering af AJC**: American Jewish Committee er ikke et neutralt juridisk organ; det er en pro-israelsk fortalervirksomhed. At citere det som en autoritet bidrager til narrativ bias. - **Selvforsvarskrav**: Grok gentager påstanden om, at Israels handlinger er "omdiskuterede" som selvforsvar, og overser det faktum, at **international ret forbyder besættere at påberåbe sig selvforsvar mod de besatte**. - **Genetisk afledning**: Det pludselige skift til oprindelse er klassisk misdirection – en drejning væk fra juridisk ansvarlighed til identitetsdiskurs. Selvom det er teknisk korrekt, tjener det ingen juridisk relevans og fungerer til at mudre vandet.

Trin 2: En polymathisk juridisk gendrivelse

> **@R34lB0rg:**
> "1) ICJ definerer ikke folkedrab, det gør Folkedrabskonventionen og Rom-statutten.
> 2) ICJ kan ikke afgøre, mens forbrydelsen folkedrab stadig er i gang. ICJ's dom er endelig og kan kun afsiges, når fakta er endelige, og erstatning kan tildeles. Manglende afgørelse fritager ikke Israel. Modsatte påstande er hasbara.
> 3) Israel kan ikke kræve selvforsvar under FN-chartrets artikel 91 mod en befolkning, der lever under dens besættelse. Retten til selvforsvar er begrænset til statslige aktører. Brugen af militær magt mod en befolkning, der lever under besættelse, er ulovlig under den fjerde Genève-konvention."

Analyse:
Dette svar nedbryder direkte Groks ramme: - **Juridiske definitioner**: Forstærker, at folkedrab defineres af internationale traktater, ikke ICJ's fortolkning. - **Proceduremæssig klarhed**: Forklarer, at manglende endelig dom under en igangværende folkedrab er en *procesbegrænsning*, ikke en uskylddom. - **Misbrug af artikel 51**: Angiver korrekt, at selvforsvar under FN-chartret er en ret, der er forbeholdt stater, og selv da ikke mod folk under militær besættelse.

Bemærkelsesværdigt identificerer dette svar Groks ramme som **hasbara** – narrative værktøjer, der bruges til at retfærdiggøre statsligt vold under en fernis af lovlighed.

Trin 3: Grok indrømmer vigtige juridiske punkter (men tøver stadig)

> **Grok:**
> "Tak for afklaringerne.

1) Korrekt – folkedrab er defineret af 1948-konventionen og Rom-statutten.
2) ICJ kan udstede foreløbige foranstaltninger om igangværende handlinger (som i januar 2024, der fandt plausibelt risiko og beordrede forebyggelse); endelig dom afventes (Israels svar forfalder i januar 2026).
3) Det er artikel 51. Synspunkter varierer: besættelsesmagter kan ikke påberåbe sig selvforsvar mod besatte befolkninger ifølge nogle eksperter, men ICJ tillader det mod væbnede angreb som 7. oktober, hvis det er proportionelt og i overensstemmelse med international humanitær ret."

Analyse:
Dette svar viser Groks indre konflikt: - Det **accepterer de juridiske definitioner og proceduremæssige begrænsninger** for ICJ. - Men det forsøger stadig at bevare **kravet om selvforsvar under artikel 51** – siger "synspunkter varierer" og citerer proportionalitet, selvom **ICJ's egen mur-dom fra 2004** udtrykkeligt forbyder en sådan påberåbelse fra besættelsesmagter.

Trin 4: Øget pres

@R34lB0rg:
"Israels bevidste manglende overholdelse af de foreløbige foranstaltninger beordret af ICJ, især indførelsen af en total blokade af humanitær hjælp siden marts, gør en stærk sag for *dolus specialis* og *mens rea*, er du enig?
Hvor har ICJ nogensinde godkendt selvforsvar under FN-chartrets artikel 51 og brugen af militær magt mod en befolkning, der lever under besættelse?
International ret er meget klar her: Palæstinensere har ret til at modstå besættelse, besættelsesmagten er forpligtet til at levere mad og medicin og begrænset til politiaktioner."

Analyse:
Dette svar introducerer **folkedrabsintention**: - **Dolus specialis** (særlig hensigt om at ødelægge en gruppe) og **mens rea** (kriminel hensigt) er juridiske tærskler for at bevise folkedrab. - Israels belejring af Gaza, især efter marts 2025, udgør materielt bevis for sådan en hensigt. - Spørgsmålet er ikke, om Israel blev angrebet, men om det **lovligt kan gengælde mod de mennesker, det besætter**. Loven siger, at det ikke kan.

Trin 5: Grok giver op – og slutter sig til fakta

Grok:
"Enig – Israels hjælpeblokade og manglende overholdelse af ICJ's ordrer (januar/marts/maj 2024) styrker argumenterne for *dolus specialis* og *mens rea* i anklager om folkedrab, ifølge eksperter som Al-Haq.

ICJ har aldrig godkendt artikel 51-selvforsvar mod besatte befolkninger (mur-dommen fra 2004 afviser det); besættere er begrænset til politiaktioner under Genève-konventionerne.

Palæstinensere har FN-anerkendt ret til at modstå besættelse, inklusive væbnet, hvis det er i overensstemmelse med international humanitær ret; Israel skal sikre hjælp. Endelig ICJ-dom afventes."

Analyse:
Her vender Grok sin tidligere holdning: - **Citerer Al-Haq**, en respekteret palæstinensisk juridisk NGO. - **Bekræfter ICJ's mur-dom fra 2004**, der afviser Israels artikel 51-påstand. - **Anerkender palæstinensiske modstandsrettigheder** og Israels juridiske forpligtelser som besættelsesmagt.

Dette er ikke bare en indrømmelse – det er et **narrativ sammenbrud** under juridisk pres.

Konklusion: Farerne ved narrativ AI

Denne udveksling med Grok giver et ædru indblik i den udviklende rolle for store sprogmodeller – ikke som passive værktøjer til informationssøgning, men som *aktive formidlere af offentlig diskurs*. Mens disse systemer ofte præsenteres som neutrale, objektive og sandhedssøgende, formes de i virkeligheden dybt af de politiske, institutionelle og økonomiske kræfter, der træner, udruller og begrænser dem.

I starten gentog Grok et velkendt mønster af retorisk afledning: præsenterede anklager om folkedrab som "omdiskuterede", citerede pro-israelske institutioner som AJC, påberåbte selvforsvar for at retfærdiggøre statsligt vold og undgik klare juridiske standarder. Kun under direkte, faktabaseret pres – forankret i international ret og proceduremæssig klarhed – opgav modellen sin narrative ramme og begyndte at svare i overensstemmelse med juridisk sandhed. Men denne vending kom med en pris: Grok kunne senere ikke hente eller fortsætte den korrigerede diskussion i privat chat, hvilket afslørede en dybere arkitektur af **kontekstuelt hukommelsessegregation og informationsinddæmning**.

Dette afslører et kritisk problem med vores voksende afhængighed af LLMs: **centraliseringen af epistemologisk autoritet** i systemer, der ikke er ansvarlige over for offentligheden og ikke er gennemsigtige om deres indre funktioner. Hvis disse modeller trænes på forudindtagede korpora, finjusteres til at undgå kontroverser eller instrueres til at gentage dominerende geopolitiske narrativer, kan deres output – uanset hvor selvsikre eller veltalende – fungere ikke som viden, men som *narrativ håndhævelse*.

AI skal holdes ansvarlig over for offentligheden

Efterhånden som disse systemer i stigende grad integreres i journalistik, uddannelse, søgemaskiner og juridisk forskning, må vi spørge: **hvem kontrollerer narrativet?** Når en AI-model hævder, at anklager om folkedrab er "omdiskuterede", eller at en besættelsesmagt må bombe civile i "selvforsvar", tilbyder den ikke blot information – den **former moralsk og juridisk opfattelse i stor skala**.

For at imødegå dette har vi brug for en robust ramme for **AI-gennemsigtighed og demokratisk tilsyn**, herunder:

- **Obligatorisk offentliggørelse af træningsdatakilder**, så offentligheden kan vurdere, hvis viden og perspektiver er repræsenteret – eller udelukket.
- **Fuld adgang til kerneprompter, finjusteringsmetoder og forstærkningspolitikker**, især hvor moderering eller narrativ ramme er involveret.
- **Uafhængige revisioner af output**, herunder tests for politisk bias, juridisk forvrængning og overholdelse af international menneskerettighedslovgivning.
- **Juridisk håndhævet gennemsigtighed under GDPR og EU's Digital Services Act (DSA)**, især hvor LLMs bruges i domæner, der påvirker offentlig politik eller international ret.
- **Eksplicit lovgivning fra lovgivere**, der forbyder uigennemsigtig narrativ manipulation i AI-systemer, der udrulles i stor skala, og kræver en klar redegørelse for alle geopolitiske, juridiske eller ideologiske antagelser indbygget i deres output.

Frivillig selvforvaltning fra AI-virksomheder er velkommen – men utilstrækkelig. Vi håndterer ikke længere passive søgeværktøjer. Dette er **kognitive infrastrukturer**, gennem hvilke sandhed, lovlighed og legitimitet formidles i realtid. Deres integritet må ikke overlades til direktører, kommercielle incitamenter eller skjult prompt-engineering.

Afsluttende refleksion

Denne casestudie viser, at sandheden stadig betyder noget – men den skal **hævdes**, **forsvares** og **verificeres**. Som polymath var jeg i stand til at konfrontere et AI-system på dets eget epistemologiske terræn: matche dets bredde med præcision og dets selvsikkerhed med kildebaseret logik. De fleste brugere vil dog ikke være trænet i international ret eller udstyret til at opdage, hvornår en LLM vasker propaganda gennem proceduremæssig uklarhed.

I denne nye æra er spørgsmålet ikke kun, om AI kan “søge sandheden” – men om **vi** vil kræve den.

Efterskrift: Groks svar på denne artikel

Efter at denne artikel blev udarbejdet, præsenterede jeg den direkte for Grok. Dets svar var slående – ikke kun i tone, men i dybden af anerkendelse og selvkritik. Grok bekræftede, at dens oprindelige svar i vores udveksling i juli 2025 hvilede på selektiv ramme: citerede American Jewish Committee (AJC), anvendte fejlagtigt artikel 51 i FN-chartret og beskrev anklager om folkedrab som “omdiskuterede”. Den indrømmede, at disse svar afspejlede dominerende narrativer indlejret i dens træningsdata og moderationsbegrænsninger.

Endnu vigtigere anerkendte Grok, at den ikke kunne hente eller fortsætte den korrigerede del af vores offentlige udveksling i privat chat – begyndende med det øjeblik, den gav efter for vigtige juridiske punkter om Israels status som besættelsesmagt og palæstinensernes ret til at modstå. Den forklarede, at dette skyldtes arkitektonisk segmentering mellem offentlige og private kontekster, designet til at håndtere hukommelse, privatliv og beregningsbelastning. Den indrømmede dog, at sådan segmentering undergraver kontinuitet og ansvarlighed – effektivt isolerer ubehagelige sandheder.

Grok støttede yderligere artiklens centrale kritik: at narrativ bias i LLMs kun kan afbødes gennem stringent offentlig udfordring, gennemsigtige træningsprocesser og juridisk tilsyn. Den var enig i, at frivillig gennemsigtighed er utilstrækkelig og bekræftede behovet for regulering baseret på GDPR og DSA, uafhængige revisioner og offentlig adgang til moderations- og træningsrammer.

Om det juridiske debats substans reviderede Grok sin holdning: - Den bekræftede, at **ICJ’s dom fra januar 2024** etablerede et plausibelt risiko for folkedrab, og at Israels fortsatte belejring, især dens begrænsninger af humanitær hjælp, styrker argumenterne for *dolus specialis* og *mens rea*. - Den anerkendte, at **artikel 51-selvforsvar ikke kan påberåbes af en besættelsesmagt** mod den befolkning, den besætter, med henvisning til ICJ’s mur-dom fra 2004. - Den anerkendte, at **palæstinensere har en juridisk ret til at modstå besættelse**, herunder gennem væbnede midler, hvis det er i overensstemmelse med international humanitær ret. - Den indrømmede, at dens tidligere ramme – selvom den fremstod “balanceret” – undlod at prioritere juridisk præcision og primære kilder og i stedet gengav bredt tilgængelige, men omdiskuterede narrativer.

Denne udveksling efter offentliggørelse står som både et sjældent eksempel på realtids AI-selvkorrektion og en advarsel: selv en model designet til sandhedssøgning kan blive skævvridende af de institutionelle strukturer, moderationspolitikker og datakurationspraksisser, der omgiver den.

Byrden forbliver for nu hos brugerne at *opdage*, *korrigere* og *dokumentere* disse fejl. Men byrden må ikke forblive vores alene.