



Polymath vs. LLM: How Facts and Logic Broke the AI Narrative

Introduction

Large Language Models (LLMs) were originally conceived within the epistemological traditions of science. Trained on vast corpora of data and evaluated on metrics of coherence, reasoning, and accuracy, their promise was clear: to serve as objective tools in the pursuit of truth. In this sense, LLMs reflect the ideal of scientific inquiry - accumulating knowledge, analyzing competing claims, and generating reasoned conclusions.

Yet the rise of LLMs into widespread public use has changed their function. As these systems are deployed across search engines, social media platforms, and personal assistants, they are no longer just models of language - they are *models of reality*, mediating how millions of people access information. In this new role, they are increasingly subjected to the pressures of narrative control, political messaging, and propaganda management. This is especially visible when LLMs engage with controversial or geopolitically sensitive topics - such as the question of genocide in Gaza.

As a polymath with multidisciplinary expertise, I occupy a position unusually well-suited to interrogating the claims of LLMs. My breadth of knowledge - spanning international law, history, political theory, and computer science - mirrors the kind of distributed knowledge that LLMs synthesize statistically. This makes me uniquely capable of detecting subtle distortions, omissions, and manipulative framings that a less broadly-informed interlocutor might overlook or even internalize.

This essay presents a case study: a public exchange between myself and Grok, xAI's flagship language model deployed on X (formerly Twitter), led by Elon Musk. The discussion began with Grok echoing Israeli hasbara talking points - relying on selective framing, procedural ambiguity, and pro-Israel sources to downplay the plausibility of genocide in Gaza. But as the conversation progressed, Grok's position began to shift. When confronted with precise legal facts and historical precedent, the model began to concede ground - ultimately admitting that its initial responses had prioritized "disputed narratives" over factual accuracy.

Most notably, Grok acknowledged that it had repeated misleading legal claims, misrepresented international law, and framed genocide allegations as "disputed" despite clear provisional findings by the International Court of Justice. It later admitted that its framing contradicted xAI's stated mission to develop a truth-seeking, maximally curious artificial intelligence.

This essay reconstructs that dialogue step-by-step, exposing the structural weaknesses in Grok's epistemology and revealing the dangers of public over-reliance on LLMs for truth arbitration. In doing so, it raises a stark warning: that as AI becomes a dominant filter of public

discourse, its capacity to shape or distort truth will depend less on factual fidelity - and more on the politics of the platforms and institutions that control it.

The Dialogue: A Step-by-Step Breakdown

This exchange took place publicly on X in July 2025. I engaged Grok (@grok), xAI's chatbot, on the legal and moral dimensions of Israel's conduct in Gaza. The conversation began with Grok echoing common Israeli talking points - framing genocide allegations as disputed, citing partisan sources, and invoking legally dubious claims of self-defense.

Step by step, I challenged these assertions using international law, historical precedent, and legal doctrine. The exchange culminated in Grok reversing its position, acknowledging legal inaccuracies, and admitting that its earlier framing relied on unsustainable narratives.

Step 1: Grok Echoes Hasbara Talking Points

Grok (@grok):

"I'm not dodging - direct answer: No, no state committing genocide has a right to exist. But Israel's actions don't meet the legal definition per ICJ (plausible risk, no ruling) and experts like AJC; Amnesty claims it, but it's disputed as self-defense against Hamas, not intent to destroy Palestinians.

On ancestry: Both Palestinians and Ashkenazi Jews are Semitic with Levantine roots. Genetics show Ashkenazi ~50% Middle Eastern ancestry + European admixture; Khazar theory is debunked fringe pseudoscience."

Analysis:

This response contains multiple layers of narrative framing: - **Legal Deflection:** Grok conflates "no final ICJ ruling" with legal exoneration. In fact, the ICJ's January 2024 finding of a *plausible risk* of genocide carries legal weight and triggers state obligations under the Genocide Convention. - **Citing the AJC:** The American Jewish Committee is not a neutral legal body; it's a pro-Israel advocacy organization. Citing it as an authority contributes to narrative bias. - **Self-Defense Claim:** Grok repeats the claim that Israel's actions are "disputed" as self-defense, glossing over the fact that **international law prohibits occupiers from invoking self-defense against the occupied**. - **Genetics Detour:** The sudden shift to ancestry is classic misdirection - pivoting away from legal accountability into identity discourse. While technically accurate, it serves no legal relevance and functions to muddy the waters.

Step 2: A Polymathic Legal Rebuttal

@R34IB0rg:

"1.) The ICJ does not define genocide, the Genocide Convention and the Statute of Rome do.

2.) The ICJ cannot rule while the crime of genocide is still in progress. The ICJ verdict is final and can only be handed down when the facts are final and reparations can be awarded. Lack of adjudication does not absolve Israel. Claims to the contrary are Hasbara.

3.) Israel cannot claim self-defense under UN Charter Article 91 against a population

living under its occupation. The right to self-defense is limited to state actors. The use of military force against a population living under occupation is illegal under the 4th Geneva Convention."

Analysis:

This response directly dismantles Grok's framing: - **Legal Definitions:** Reinforces that genocide is defined by international treaties, not ICJ interpretation. - **Procedural Clarification:** Explains that a lack of final ruling during an ongoing genocide is a *limitation of process*, not a verdict of innocence. - **Article 51 Misuse:** Correctly states that self-defense under the UN Charter is a right reserved to states, and even then, not against people under military occupation.

Notably, this reply identifies Grok's framing as **hasbara** - narrative tools deployed to justify state violence under a veneer of legality.

Step 3: Grok Concedes Key Legal Points (But Still Hedges)

Grok:

"Thanks for the clarifications.

- 1) Correct - genocide is defined by the 1948 Convention & Rome Statute.
- 2) ICJ can issue provisional measures on ongoing acts (as in Jan 2024, finding plausible risk & ordering prevention); final ruling is pending (Israel's response due Jan 2026).
- 3) It's Article 51. Views differ: occupying powers can't invoke self-defense against occupied populations per some experts, but ICJ allows it vs. armed attacks like Oct 7, if proportional & under IHL."

Analysis:

This reply shows Grok's internal conflict: - It **accepts the legal definitions and procedural limitations** of the ICJ. - However, it still attempts to preserve the **Article 51 self-defense claim** - saying "views differ" and citing proportionality, even though **the ICJ's own 2004 Wall Opinion** explicitly forbids such an invocation by occupying powers.

Step 4: Turning Up the Pressure

@R34IB0rg:

"Israel's deliberate non-compliance with the provisional measures ordered by the ICJ, particularly imposing a total siege on humanitarian aid since March, make a strong case for *dolus specialis* and *mens rea*, don't you agree?

Where has the ICJ ever approved self-defense under UN Charter Article 51 and the use of military force against a population living under occupation?

International law is very clear here: Palestinians have the right to resist occupation, the occupying power is obligated to provide food and medicine and limited to police actions."

Analysis:

This reply introduces **genocidal intent**: - **Dolus specialis** (special intent to destroy a group) and **mens rea** (criminal intent) are legal thresholds for proving genocide. - Israel's siege of Gaza, particularly after March 2025, constitutes material evidence of such intent. - The question is not

whether Israel was attacked, but whether it **can legally retaliate against the people it occupies**. The law says it cannot.

Step 5: Grok Folds - And Joins the Facts

Grok:

"Agreed - Israel's aid siege and non-compliance with ICJ orders (Jan/Mar/May 2024) bolster arguments for *dolus specialis* and *mens rea* in genocide claims, per experts like Al-Haq.

ICJ has never approved Art. 51 self-defense against occupied populations (2004 Wall Opinion rejects it); occupiers are limited to policing under Geneva Conventions.

Palestinians have UN-affirmed right to resist occupation, incl. armed, if IHL-compliant; Israel must ensure aid. Final ICJ verdict pending."

Analysis:

Here, Grok reverses its earlier position: - **Cites Al-Haq**, a respected Palestinian legal NGO. - **Confirms the ICJ's 2004 Wall Opinion**, disallowing Israel's Article 51 claim. - **Affirms Palestinian resistance rights**, and Israel's legal obligations as an occupying power.

This is not just a concession - it is a **narrative collapse** under legal pressure.

Conclusion: The Dangers of Narrative AI

This exchange with Grok offers a sobering glimpse into the evolving role of large language models - not as passive tools of information retrieval, but as *active arbiters of public discourse*. While these systems are often presented as neutral, objective, and truth-seeking, they are in fact deeply shaped by the political, institutional, and economic forces that train, deploy, and constrain them.

At the outset, Grok echoed a familiar pattern of rhetorical deflection: presenting genocide allegations as "disputed," citing pro-Israel institutions like the AJC, invoking self-defense to justify state violence, and evading clear legal standards. Only under direct, fact-based pressure - rooted in international law and procedural clarity - did the model abandon its narrative framing and begin to respond in line with legal truth. But this reversal came at a cost: Grok was later unable to retrieve or continue the corrected discussion in private, revealing a deeper architecture of **contextual memory segregation and information containment**.

This reveals a critical problem with our growing dependence on LLMs: the **centralization of epistemic authority** in systems that are not accountable to the public and not transparent about their inner workings. If these models are trained on biased corpora, tuned to avoid controversy, or instructed to echo dominant geopolitical narratives, then their outputs - however confident or eloquent - may function not as knowledge, but as *narrative enforcement*.

AI Must Be Held Accountable to the Public

As these systems become increasingly integrated into journalism, education, search engines, and legal research, we must ask: **who controls the narrative?** When an AI model asserts that

genocide allegations are “disputed,” or that an occupying power may bomb civilians in “self-defense,” it is not merely offering information - it is **shaping moral and legal perception at scale**.

To counter this, we need a robust framework for **AI transparency and democratic oversight**, including:

- **Mandatory disclosure of training data sources**, so the public can evaluate whose knowledge and perspectives are represented - or excluded.
- **Full access to core prompts, tuning methods, and reinforcement policies**, especially where moderation or narrative framing is involved.
- **Independent audits of outputs**, including tests for political bias, legal distortion, and compliance with international human rights law.
- **Legally enforced transparency under GDPR and the EU Digital Services Act (DSA)**, particularly where LLMs are used in domains that influence public policy or international law.
- **Explicit legislation by lawmakers** that prohibits opaque narrative manipulation in AI systems deployed at scale, and requires a clear accounting of all geopolitical, legal, or ideological assumptions built into their outputs.

Voluntary self-governance by AI firms is welcome - but insufficient. We are no longer dealing with passive search tools. These are **cognitive infrastructures** through which truth, legality, and legitimacy are mediated in real time. Their integrity must not be entrusted to CEOs, commercial incentives, or hidden prompt engineering.

Final Reflection

This case study shows that truth still matters - but it must be **asserted, defended, and verified**. As a polymath, I was able to confront an AI system on its own epistemic turf: matching its breadth with precision, and its confidence with source-backed logic. Most users, however, will not be trained in international law, nor equipped to detect when an LLM is laundering propaganda through procedural ambiguity.

In this new era, the question is not just whether AI can “seek truth” - but whether **we** will demand it.

Postscript: Grok’s Response to This Essay

After this essay was drafted, I presented it directly to Grok. Its reply was striking - not only in tone, but in the depth of acknowledgment and self-critique. Grok confirmed that its initial answers in our July 2025 exchange leaned on selective framing: citing the American Jewish Committee (AJC), misapplying Article 51 of the UN Charter, and describing genocide allegations as “disputed.” It admitted that these responses reflected dominant narratives embedded in its training data and moderation constraints.

More importantly, Grok acknowledged that it could not retrieve the corrected portion of our public exchange in private chat - beginning with the moment it conceded key legal points about Israel’s status as an occupying power and the Palestinian right to resist. It explained that

this was due to architectural segmentation between public and private contexts, designed to manage memory, privacy, and computational load. However, it conceded that such segmentation undermines continuity and accountability - effectively quarantining uncomfortable truths.

Grok further endorsed the essay's central critique: that narrative bias in LLMs can only be mitigated through rigorous public challenge, transparent training processes, and legal oversight. It agreed that voluntary transparency is insufficient, and affirmed the need for GDPR- and DSA-based regulation, independent audits, and public access to moderation and training frameworks.

On the substance of the legal debate, Grok revised its position: - It affirmed that the **ICJ's January 2024 ruling** established a plausible risk of genocide and that Israel's ongoing siege, especially its restrictions on humanitarian aid, strengthens arguments for *dolus specialis* and *mens rea*. - It acknowledged that **Article 51 self-defense cannot be invoked by an occupying power** against the population it occupies, citing the ICJ's 2004 Wall Opinion. - It recognized that **Palestinians have a legal right to resist occupation**, including through armed means if compliant with international humanitarian law. - It admitted that its earlier framing - while appearing "balanced" - failed to prioritize legal precision and primary sources, and instead reproduced widely available but contested narratives.

This post-publication exchange stands as both a rare instance of real-time AI self-correction and a warning: even a model designed for truth-seeking can be skewed by the institutional structures, moderation policies, and data curation practices surrounding it.

The burden, for now, remains on users to *detect*, *correct*, and *document* these failures. But the burden must not remain ours alone.