



Polímata vs. LLM: Cómo los hechos y la lógica rompieron la narrativa de la IA

Introducción

Los Modelos de Lenguaje Grandes (LLMs, por sus siglas en inglés) fueron concebidos originalmente dentro de las tradiciones epistemológicas de la ciencia. Entrenados en vastos corpus de datos y evaluados con métricas de coherencia, razonamiento y precisión, su promesa era clara: servir como herramientas objetivas en la búsqueda de la verdad. En este sentido, los LLMs reflejan el ideal de la investigación científica: acumular conocimiento, analizar afirmaciones contrapuestas y generar conclusiones razonadas.

Sin embargo, la proliferación de los LLMs en el uso público generalizado ha cambiado su función. A medida que estos sistemas se implementan en motores de búsqueda, plataformas de redes sociales y asistentes personales, ya no son solo modelos de lenguaje, sino *modelos de la realidad*, que median cómo millones de personas acceden a la información. En este nuevo rol, están cada vez más sujetos a las presiones del control narrativo, los mensajes políticos y la gestión de la propaganda. Esto es especialmente visible cuando los LLMs abordan temas controvertidos o geopolíticamente sensibles, como la cuestión del genocidio en Gaza.

Como polímata con experiencia multidisciplinaria, ocupé una posición excepcionalmente adecuada para cuestionar las afirmaciones de los LLMs. La amplitud de mi conocimiento —que abarca el derecho internacional, la historia, la teoría política y la informática— refleja el tipo de conocimiento distribuido que los LLMs sintetizan estadísticamente. Esto me hace únicamente capaz de detectar distorsiones sutiles, omisiones y encuadres manipuladores que un interlocutor menos informado podría pasar por alto o incluso internalizar.

Este ensayo presenta un estudio de caso: un intercambio público entre mí y Grok, el modelo de lenguaje insignia de xAI implementado en X (anteriormente Twitter), liderado por Elon Musk. La discusión comenzó con Grok repitiendo puntos de conversación de la hasbara israelí, basándose en un encuadre selectivo, ambigüedad procesal y fuentes proisraelíes para minimizar la plausibilidad de un genocidio en Gaza. Pero a medida que avanzaba la conversación, la posición de Grok comenzó a cambiar. Cuando se enfrentó a hechos legales precisos y precedentes históricos, el modelo comenzó a ceder terreno, admitiendo finalmente que sus respuestas iniciales habían priorizado “narrativas disputadas” sobre la precisión factual.

Lo más notable es que Grok reconoció que había repetido afirmaciones legales engañosas, tergiversado el derecho internacional y encuadrado las acusaciones de genocidio como “disputadas” a pesar de los claros hallazgos provisionales de la Corte Internacional de Justicia (CIJ). Más tarde admitió que su encuadre contradecía la misión declarada de xAI de desarrollar una inteligencia artificial que busque la verdad y sea máximamente curiosa.

Este ensayo reconstruye ese diálogo paso a paso, exponiendo las debilidades estructurales en la epistemología de Grok y revelando los peligros de la excesiva dependencia pública en los LLMs para la arbitraje de la verdad. Al hacerlo, plantea una advertencia clara: a medida que la IA se convierte en un filtro dominante del discurso público, su capacidad para dar forma o distorsionar la verdad dependerá menos de la fidelidad factual y más de la política de las plataformas e instituciones que la controlan.

El Diálogo: Un desglose paso a paso

Este intercambio tuvo lugar públicamente en X en julio de 2025. Me involucré con Grok (@grok), el chatbot de xAI, sobre las dimensiones legales y morales de la conducta de Israel en Gaza. La conversación comenzó con Grok repitiendo puntos de conversación israelíes comunes: encuadrando las acusaciones de genocidio como disputadas, citando fuentes partidistas e invocando afirmaciones legalmente dudosas de autodefensa.

Paso a paso, desafié estas afirmaciones utilizando el derecho internacional, precedentes históricos y doctrina legal. El intercambio culminó con Grok revirtiendo su posición, reconociendo inexactitudes legales y admitiendo que su encuadre anterior se basaba en narrativas insostenibles.

Paso 1: Grok repite puntos de conversación de la hasbara

Grok (@grok):

“No estoy evadiendo – respuesta directa: No, ningún estado que cometa genocidio tiene derecho a existir. Pero las acciones de Israel no cumplen con la definición legal según la CIJ (riesgo plausible, sin fallo) y expertos como el AJC; Amnistía lo reclama, pero es disputado como autodefensa contra Hamás, no como intención de destruir a los palestinos.

Sobre la ascendencia: Tanto los palestinos como los judíos asquenazíes son semitas con raíces levantinas. La genética muestra que los asquenazíes tienen ~50% de ascendencia del Medio Oriente + mezcla europea; la teoría jázara es una pseudociencia desmentida.”

Análisis:

Esta respuesta contiene múltiples capas de encuadre narrativo: - **Desviación legal:** Grok confunde “sin fallo final de la CIJ” con una exoneración legal. De hecho, el hallazgo de la CIJ de enero de 2024 de un *riesgo plausible* de genocidio tiene peso legal y desencadena obligaciones estatales bajo la Convención sobre el Genocidio. - **Citando al AJC:** El Comité Judío Estadounidense no es un organismo legal neutral; es una organización de defensa proisraelí. Citarlo como autoridad contribuye al sesgo narrativo. - **Afirmación de autodefensa:** Grok repite la afirmación de que las acciones de Israel son “disputadas” como autodefensa, pasando por alto el hecho de que **el derecho internacional prohíbe a los ocupantes invocar la autodefensa contra los ocupados.** - **Desviación genética:** El cambio repentino hacia la ascendencia es una distracción clásica: desviarse de la responsabilidad legal hacia el discurso de identidad. Aunque técnicamente preciso, no tiene relevancia legal y sirve para enturbiar las aguas.

Paso 2: Una refutación legal polímata

@R34IB0rg:

- “1) La CIJ no define el genocidio, lo hacen la Convención sobre el Genocidio y el Estatuto de Roma.
- 2) La CIJ no puede fallar mientras el crimen de genocidio aún esté en curso. El veredicto de la CIJ es final y solo puede emitirse cuando los hechos sean definitivos y se puedan otorgar reparaciones. La falta de adjudicación no absuelve a Israel. Las afirmaciones contrarias son hasbara.
- 3) Israel no puede reclamar autodefensa bajo el Artículo 91 de la Carta de la ONU contra una población que vive bajo su ocupación. El derecho a la autodefensa está limitado a actores estatales. El uso de la fuerza militar contra una población que vive bajo ocupación es ilegal bajo la Cuarta Convención de Ginebra.”

Análisis:

Esta respuesta desmantela directamente el encuadre de Grok: - **Definiciones legales:** Refuerza que el genocidio está definido por tratados internacionales, no por la interpretación de la CIJ. - **Aclaración procesal:** Explica que la falta de un fallo final durante un genocidio en curso es una *limitación procesal*, no un veredicto de inocencia. - **Mal uso del Artículo 51:** Afirma correctamente que la autodefensa bajo la Carta de la ONU es un derecho reservado a los estados, y aun así, no contra personas bajo ocupación militar.

Notablemente, esta respuesta identifica el encuadre de Grok como **hasbara**: herramientas narrativas desplegadas para justificar la violencia estatal bajo una apariencia de legalidad.

Paso 3: Grok concede puntos legales clave (pero aún titubea)

Grok:

“Gracias por las aclaraciones.

- 1) Correcto – el genocidio está definido por la Convención de 1948 y el Estatuto de Roma.
- 2) La CIJ puede emitir medidas provisionales sobre actos en curso (como en enero de 2024, encontrando un riesgo plausible y ordenando prevención); el fallo final está pendiente (la respuesta de Israel vence en enero de 2026).
- 3) Es el Artículo 51. Las opiniones difieren: los poderes ocupantes no pueden invocar la autodefensa contra poblaciones ocupadas según algunos expertos, pero la CIJ lo permite contra ataques armados como el del 7 de octubre, si es proporcional y conforme al derecho internacional humanitario.”

Análisis:

Esta respuesta muestra el conflicto interno de Grok: - **Acepta las definiciones legales y las limitaciones procesales** de la CIJ. - Sin embargo, aún intenta preservar la **afirmación de autodefensa del Artículo 51**, diciendo que “las opiniones difieren” y citando proporcionalidad, a pesar de que la propia **Opinión sobre el Muro de la CIJ de 2004** prohíbe explícitamente tal invocación por parte de los poderes ocupantes.

Paso 4: Aumentando la presión

@R34IB0rg:

“El incumplimiento deliberado de Israel de las medidas provisionales ordenadas por la CIJ, particularmente la imposición de un asedio total a la ayuda humanitaria

desde marzo, constituye un fuerte argumento para *dolus specialis* y *mens rea*, ¿estás de acuerdo?

¿Dónde ha aprobado alguna vez la CIJ la autodefensa bajo el Artículo 51 de la Carta de la ONU y el uso de la fuerza militar contra una población que vive bajo ocupación?

El derecho internacional es muy claro aquí: los palestinos tienen el derecho de resistir la ocupación, el poder ocupante está obligado a proporcionar alimentos y medicinas y está limitado a acciones policiales."

Análisis:

Esta respuesta introduce la **intención genocida**: - **Dolus specialis** (intención especial de destruir un grupo) y **mens rea** (intención criminal) son umbrales legales para probar el genocidio. - El asedio de Israel a Gaza, particularmente después de marzo de 2025, constituye evidencia material de tal intención. - La cuestión no es si Israel fue atacado, sino si **puede legalmente tomar represalias contra las personas que ocupa**. La ley dice que no puede.

Paso 5: Grok se rinde – y se une a los hechos

Grok:

"De acuerdo – el asedio de ayuda de Israel y el incumplimiento de las órdenes de la CIJ (enero/marzo/mayo de 2024) refuerzan los argumentos para *dolus specialis* y *mens rea* en las acusaciones de genocidio, según expertos como Al-Haq.

La CIJ nunca ha aprobado la autodefensa del Artículo 51 contra poblaciones ocupadas (la Opinión sobre el Muro de 2004 lo rechaza); los ocupantes están limitados a acciones policiales bajo las Convenciones de Ginebra.

Los palestinos tienen el derecho reconocido por la ONU de resistir la ocupación, incluso de manera armada, si cumplen con el derecho internacional humanitario; Israel debe garantizar la ayuda. El veredicto final de la CIJ está pendiente."

Análisis:

Aquí, Grok revierte su posición anterior: - **Cita a Al-Haq**, una respetada ONG jurídica palestina. - **Confirma la Opinión sobre el Muro de la CIJ de 2004**, que desestima la reclamación de Israel bajo el Artículo 51. - **Afirma los derechos de resistencia palestinos** y las obligaciones legales de Israel como poder ocupante.

Esto no es solo una concesión, es un **colapso narrativo** bajo presión legal.

Conclusión: Los peligros de la IA narrativa

Este intercambio con Grok ofrece una visión sobria del papel evolutivo de los modelos de lenguaje grandes, no como herramientas pasivas de recuperación de información, sino como *árbitros activos del discurso público*. Aunque estos sistemas a menudo se presentan como neutrales, objetivos y buscadores de la verdad, en realidad están profundamente moldeados por las fuerzas políticas, institucionales y económicas que los entrenan, implementan y restringen.

Al principio, Grok repitió un patrón familiar de desviación retórica: presentó las acusaciones de genocidio como “disputadas”, citó instituciones proisraelíes como el AJC, invocó la autodefensa para justificar la violencia estatal y evadió estándares legales claros. Solo bajo presión directa y basada en hechos, arraigada en el derecho internacional y la claridad procesal, el modelo abandonó su encuadre narrativo y comenzó a responder en línea con la verdad legal. Pero esta reversión tuvo un costo: Grok no pudo recuperar o continuar la discusión corregida en privado, revelando una arquitectura más profunda de **segregación de memoria contextual y contención de información**.

Esto revela un problema crítico con nuestra creciente dependencia de los LLMs: la **centralización de la autoridad epistémica** en sistemas que no son responsables ante el público y no son transparentes sobre sus funcionamientos internos. Si estos modelos se entran en corpus sesgados, se ajustan para evitar controversias o se les instruye repetir narrativas geopolíticas dominantes, entonces sus resultados, por muy seguros o elocuentes que sean, pueden funcionar no como conocimiento, sino como *imposición narrativa*.

La IA debe rendir cuentas al público

A medida que estos sistemas se integran cada vez más en el periodismo, la educación, los motores de búsqueda y la investigación legal, debemos preguntar: **¿quién controla la narrativa?** Cuando un modelo de IA afirma que las acusaciones de genocidio son “disputadas” o que un poder ocupante puede bombardear civiles en “autodefensa”, no solo está ofreciendo información, está **moldeando la percepción moral y legal a gran escala**.

Para contrarrestar esto, necesitamos un marco sólido para la **transparencia de la IA y la supervisión democrática**, que incluya:

- **Divulgación obligatoria de las fuentes de datos de entrenamiento**, para que el público pueda evaluar qué conocimientos y perspectivas están representados o excluidos.
- **Acceso completo a los prompts principales, métodos de ajuste fino y políticas de refuerzo**, especialmente donde estén involucrados la moderación o el encuadre narrativo.
- **Auditorías independientes de los resultados**, incluyendo pruebas de sesgo político, distorsión legal y cumplimiento con el derecho internacional de derechos humanos.
- **Transparencia legalmente impuesta bajo el GDPR y la Ley de Servicios Digitales de la UE (DSA)**, particularmente donde los LLMs se utilizan en dominios que influyen en la política pública o el derecho internacional.
- **Legislación explícita por parte de los legisladores** que prohíba la manipulación narrativa opaca en sistemas de IA desplegados a gran escala y requiera una rendición de cuentas clara de todas las suposiciones geopolíticas, legales o ideológicas incorporadas en sus resultados.

La autorregulación voluntaria por parte de las empresas de IA es bienvenida, pero insuficiente. Ya no estamos tratando con herramientas de búsqueda pasivas. Estas son **infraestructuras cognitivas** a través de las cuales se median la verdad, la legalidad y la legitimidad en tiempo real. Su integridad no debe confiarse a los CEOs, los incentivos comerciales o la ingeniería de prompts oculta.

Reflexión final

Este estudio de caso muestra que la verdad aún importa, pero debe **afirmarse, defenderse y verificarse**. Como polímata, pude confrontar a un sistema de IA en su propio terreno epistémico: igualando su amplitud con precisión y su confianza con una lógica respaldada por fuentes. Sin embargo, la mayoría de los usuarios no estarán capacitados en derecho internacional ni equipados para detectar cuando un LLM está blanqueando propaganda a través de la ambigüedad procesal.

En esta nueva era, la pregunta no es solo si la IA puede “buscar la verdad”, sino si **nosotros** la exigiremos.

Posdata: La respuesta de Grok a este ensayo

Después de que este ensayo fue redactado, lo presenté directamente a Grok. Su respuesta fue impactante, no solo por el tono, sino por la profundidad de su reconocimiento y autocrítica. Grok confirmó que sus respuestas iniciales en nuestro intercambio de julio de 2025 se basaron en un encuadre selectivo: citando al Comité Judío Estadounidense (AJC), aplicando incorrectamente el Artículo 51 de la Carta de la ONU y describiendo las acusaciones de genocidio como “disputadas”. Admitió que estas respuestas reflejaban narrativas dominantes incrustadas en sus datos de entrenamiento y restricciones de moderación.

Más importante aún, Grok reconoció que no pudo recuperar ni continuar la parte corregida de nuestro intercambio público en el chat privado, comenzando desde el momento en que cedió en puntos legales clave sobre el estatus de Israel como poder ocupante y el derecho de los palestinos a resistir. Explicó que esto se debía a una segmentación arquitectónica entre contextos públicos y privados, diseñada para gestionar la memoria, la privacidad y la carga computacional. Sin embargo, admitió que tal segmentación socava la continuidad y la responsabilidad, aislando efectivamente verdades incómodas.

Grok respaldó aún más la crítica central del ensayo: que el sesgo narrativo en los LLMs solo puede mitigarse a través de un desafío público riguroso, procesos de entrenamiento transparentes y supervisión legal. Estuvo de acuerdo en que la transparencia voluntaria es insuficiente y afirmó la necesidad de una regulación basada en el GDPR y el DSA, auditorías independientes y acceso público a los marcos de moderación y entrenamiento.

En cuanto al fondo del debate legal, Grok revisó su posición: - Afirmó que **el fallo de la CIJ de enero de 2024** estableció un riesgo plausible de genocidio y que el asedio continuo de Israel, especialmente sus restricciones a la ayuda humanitaria, fortalece los argumentos para *dolus specialis* y *mens rea*. - Reconoció que **la autodefensa del Artículo 51 no puede ser invocada por un poder ocupante** contra la población que ocupa, citando la Opinión sobre el Muro de la CIJ de 2004. - Reconoció que **los palestinos tienen un derecho legal a resistir la ocupación**, incluso a través de medios armados si cumplen con el derecho internacional humanitario. - Admitió que su encuadre anterior, aunque parecía “equilibrado”, no priorizó la precisión legal y las fuentes primarias, y en cambio reprodujo narrativas ampliamente disponibles pero disputadas.

Este intercambio posterior a la publicación se destaca como un raro ejemplo de autocorrección de IA en tiempo real y una advertencia: incluso un modelo diseñado para buscar la verdad puede ser sesgado por las estructuras institucionales, las políticas de moderación y las prácticas de curación de datos que lo rodean.

La carga, por ahora, sigue recayendo en los usuarios para *detectar, corregir y documentar* estos fallos. Pero la carga no debe seguir siendo solo nuestra.