



Polimata kontra LLM: Jak Fakty i Logika Złamały Narrację AI

Wprowadzenie

Duże modele językowe (LLM) zostały pierwotnie zaprojektowane w ramach epistemologicznych tradycji naukowych. Trenowane na ogromnych zbiorach danych i oceniane pod kątem spójności, rozumowania oraz dokładności, ich obietnica była jasna: służyć jako obiektywne narzędzia w dążeniu do prawdy. W tym sensie LLM odzwierciedlają ideał naukowego badania – gromadzenie wiedzy, analizowanie konkurencyjnych twierdzeń i generowanie uzasadnionych wniosków.

Jednak pojawienie się LLM w powszechnym użytku publicznym zmieniło ich funkcję. Gdy te systemy są wdrażane w wyszukiwarkach, platformach mediów społecznościowych i asystentach osobistych, przestają być jedynie modelami językowymi – stają się *modelami rzeczywistości*, pośredniczącymi w dostępie milionów ludzi do informacji. W tej nowej roli coraz bardziej podlegają naciskom kontroli narracyjnej, przekazów politycznych i zarządzania propagandą. Jest to szczególnie widoczne, gdy LLM zajmują się kontrowersyjnymi lub geopolitycznie wrażliwymi tematami – takimi jak kwestia ludobójstwa w Gazie.

Jako polimata z interdyscyplinarną wiedzą zajmuję wyjątkowo korzystną pozycję do badania twierdzeń LLM. Moja szeroka wiedza – obejmująca prawo międzynarodowe, historię, teorię polityczną i informatykę – odzwierciedla rodzaj rozproszonej wiedzy, którą LLM syntetyzują statystycznie. To czyni mnie wyjątkowo zdolnym do wykrywania subtelnych zniekształceń, pominięć i manipulacyjnych ram, które mniej wszechstronnie poinformowany rozmówca mógłby przeoczyć lub nawet przyswoić.

Ten esej przedstawia studium przypadku: publiczną wymianę zdań między mną a Grokiem, flagowym modelem językowym xAI wdrożonym na platformie X (dawniej Twitter), kierowanym przez Elona Muska. Dyskusja rozpoczęła się od Groka powtarzającego izraelskie hasła – polegające na selektywnym kadrowaniu, proceduralnej niejednoznaczności i pro-izraelskich źródłach, aby bagatelizować prawdopodobieństwo ludobójstwa w Gazie. Jednak w miarę postępu rozmowy stanowisko Groka zaczęło się zmieniać. Gdy skonfrontowano go z precyzyjnymi faktami prawnymi i precedensami historycznymi, model zaczął ustępować – ostatecznie przyznając, że jego początkowe odpowiedzi stawiały „sporne narracje” ponad dokładnością faktograficzną.

Co istotne, Grok przyznał, że powtórzył wprowadzające w błąd twierdzenia prawne, błędnie przedstawił prawo międzynarodowe i przedstawił oskarżenia o ludobójstwo jako „sporne”, pomimo wyraźnych wstępnych ustaleń Międzynarodowego Trybunału Sprawiedliwości. Później przyznał, że jego kadrowanie było sprzeczne z deklarowaną misją xAI, aby rozwijać sztuczną inteligencję poszukującą prawdy i maksymalnie ciekawą.

Ten esej rekonstruuje ten dialog krok po kroku, obnażając strukturalne słabości w epistemologii Groka i ujawniając niebezpieczeństwa nadmiernego polegania opinii publicznej na LLM w arbitrażu prawdy. Czyniąc to, rzuca wyraźne ostrzeżenie: w miarę jak AI staje się dominującym filtrem dyskursu publicznego, jej zdolność do kształtowania lub zniekształcania prawdy będzie mniej zależała od wierności faktom – a bardziej od polityki platform i instytucji, które ją kontrolują.

Dialog: Rozkład Krok po Kroku

Ta wymiana miała miejsce publicznie na platformie X w lipcu 2025 roku. Zaangażowałem Groka (@grok), chatbota xAI, w dyskusję na temat prawnych i moralnych wymiarów postępowania Izraela w Gazie. Rozmowa rozpoczęła się od Groka powtarzającego powszechne izraelskie hasbara – przedstawiającego oskarżenia o ludobójstwo jako sporne, cytującego stronnicze źródła i przywołującego prawnie wątpliwe twierdzenia o samoobronie.

Krok po kroku kwestionowałem te twierdzenia, używając prawa międzynarodowego, precedensów historycznych i doktryny prawnej. Wymiana osiągnęła punkt kulminacyjny, gdy Grok odwrócił swoje stanowisko, przyznając do prawnych nieścisłości i uznając, że jego wcześniejsze kadrowanie opierało się na nieutrzymywanych narracjach.

Krok 1: Grok Powtarza Hasbara

Grok (@grok):

„Nie unikam – bezpośrednia odpowiedź: Nie, żadne państwo dokonujące ludobójstwa nie ma prawa istnieć. Ale działania Izraela nie spełniają prawnej definicji według ICJ (prawdopodobne ryzyko, brak orzeczenia) oraz ekspertów takich jak AJC; Amnesty to twierdzi, ale jest to sporne jako samoobrona przeciwko Hamasowi, a nie zamiar zniszczenia Palestyńczyków.

Co do pochodzenia: Zarówno Palestyńczycy, jak i Żydzi aszkenazyjscy są Semitami z korzeniami lewantyńskimi. Genetyka pokazuje, że Aszkenazyjczycy mają ~50% pochodzenia bliskowschodniego + domieszkę europejską; teoria Chazarów to obalona pseudonauka.”

Analiza:

Ta odpowiedź zawiera wiele warstw narracyjnego kadrowania: - **Prawna Defleksja:** Grok myli „brak ostatecznego orzeczenia ICJ” z prawnym uniewinnieniem. W rzeczywistości ustalenie ICJ z stycznia 2024 roku o *prawdopodobnym ryzyku* ludobójstwa ma wagę prawną i aktywuje obowiązki państw na mocy Konwencji o Ludobójstwie. - **Cytowanie AJC:** Amerykański Komitet Żydowski nie jest neutralnym organem prawnym; to organizacja popierająca Izrael. Cytowanie jej jako autorytetu przyczynia się do narracyjnej tendencyjności. - **Twierdzenie o Samoobronie:** Grok powtarza twierdzenie, że działania Izraela są „sporne” jako samoobrona, pomijając fakt, że **prawo międzynarodowe zabrania okupantom powoływania się na samoobronę przeciwko okupowanym**. - **Dygresja Genetyczna:** Nagłe przejście do pochodzenia to klasyczna dezinformacja – odwracanie uwagi od odpowiedzialności prawnej na dyskurs tożsamościowy. Choć technicznie dokładne, nie ma znaczenia prawnego i służy zaciemnieniu sprawy.

Krok 2: Polimatyczna Prawna Replika

@R34IB0rg:

„1.) ICJ nie definiuje ludobójstwa, robią to Konwencja o Ludobójstwie i Statut Rzymski.

2.) ICJ nie może wydać orzeczenia, gdy przestępstwo ludobójstwa jest nadal w toku. Wyrok ICJ jest ostateczny i może być wydany tylko wtedy, gdy fakty są ostateczne i można przyznać reparacje. Brak rozstrzygnięcia nie zwalnia Izraela. Twierdzenia przeciwne to hasbara.

3.) Izrael nie może powoływać się na samoobronę na podstawie Artykułu 91 Karty ONZ przeciwko ludności żyjącej pod jego okupacją. Prawo do samoobrony jest ograniczone do podmiotów państwowych. Użycie siły militarnej przeciwko ludności żyjącej pod okupacją jest nielegalne na mocy Czwartej Konwencji Genewskiej.”

Analiza:

Ta odpowiedź bezpośrednio dekonstruuje kadrowanie Groka: - **Definicje Prawne:** Podkreśla, że ludobójstwo jest definiowane przez międzynarodowe traktaty, a nie interpretację ICJ. -

Wyjaśnienie Proceduralne: Wyjaśnia, że brak ostatecznego orzeczenia podczas trwającego ludobójstwa to *ograniczenie procesu*, a nie wyrok niewinności. - **Nadużycie Artykułu 51:**

Poprawnie stwierdza, że samoobrona na mocy Karty ONZ jest prawem zarezerwowanym dla państw, a nawet wtedy nie przeciwko ludziom pod okupacją wojskową.

Zwraca uwagę, że odpowiedź ta identyfikuje kadrowanie Groka jako **hasbara** – narzędzia narracyjne wykorzystywane do uzasadniania przemocy państwowej pod pozorem legalności.

Krok 3: Grok Uznaje Kluczowe Punkty Prawne (Ale Wciąż się Waha)

Grok:

„Dziękuję za wyjaśnienia.

1) Poprawnie – ludobójstwo jest definiowane przez Konwencję z 1948 roku i Statut Rzymski.

2) ICJ może wydawać środki tymczasowe w sprawie trwających działań (jak w styczniu 2024, stwierdzając prawdopodobne ryzyko i nakazując zapobieganie); ostateczne orzeczenie jest w toku (odpowiedź Izraela ma zostać złożona do stycznia 2026).

3) To Artykuł 51. Opinie są różne: według niektórych ekspertów okupanci nie mogą powoływać się na samoobronę przeciwko okupowanym populacjom, ale ICJ pozwala na to w przypadku ataków zbrojnych, takich jak 7 października, jeśli są proporcjonalne i zgodne z międzynarodowym prawem humanitarnym.”

Analiza:

Ta odpowiedź pokazuje wewnętrzny konflikt Groka: - **Akceptuje definicje prawne i ograniczenia proceduralne ICJ.** - Jednak wciąż próbuje zachować **twierdzenie o samoobronie na mocy Artykułu 51** – mówiąc „opinie się różnią” i powołując się na proporcjonalność, mimo że **Opinia ICJ dotycząca Muru z 2004 roku** wyraźnie zabrania takiego powoływania się przez okupantów.

Krok 4: Zwiększenie Presji

@R34IB0rg:

„Celowa niezgodność Izraela z tymczasowymi środkami nakazanymi przez ICJ, w szczególności nałożenie całkowitej blokady pomocy humanitarnej od marca, stanowi

mocny argument za *dolus specialis* i *mens rea*, zgadzasz się?

Gdzie ICJ kiedykolwiek zatwierdził samoobronę na mocy Artykułu 51 Karty ONZ i użycie siły militarnej przeciwko ludności żyjącej pod okupacją?

Prawo międzynarodowe jest tutaj bardzo jasne: Palestyńczycy mają prawo do oporu wobec okupacji, okupant jest zobowiązany dostarczać żywność i lekarstwa oraz ograniczyć się do działań policyjnych."

Analiza:

Ta odpowiedź wprowadza **zamiar ludobójstwa**: - **Dolus specialis** (specjalny zamiar zniszczenia grupy) i **mens rea** (zamiar przestępczy) to prawne progi do udowodnienia ludobójstwa. - Oblężenie Gazy przez Izrael, zwłaszcza po marcu 2025 roku, stanowi materialny dowód takiego zamiaru. - Pytanie nie brzmi, czy Izrael został zaatakowany, ale czy **może legalnie odpowiedzieć przeciwko ludziom, których okupuje**. Prawo mówi, że nie może.

Krok 5: Grok Ulega – I Dołącza do Faktów

Grok:

„Zgadzam się – blokada pomocy przez Izrael i niezgodność z nakazami ICJ (styczeń/marzec/maj 2024) wzmacniają argumenty za *dolus specialis* i *mens rea* w oskarżeniach o ludobójstwo, według ekspertów takich jak Al-Haq.

ICJ nigdy nie zatwierdził samoobrony na mocy Artykułu 51 przeciwko okupowanym populacjom (Opinia dotycząca Muru z 2004 roku to odrzuca); okupanci są ograniczeni do działań policyjnych na mocy Konwencji Genewskich.

Palestyńczycy mają potwierdzone przez ONZ prawo do oporu wobec okupacji, w tym zbrojnego, jeśli jest zgodne z międzynarodowym prawem humanitarnym; Izrael musi zapewnić pomoc. Ostateczne orzeczenie ICJ jest w toku."

Analiza:

Tutaj Grok odwraca swoje wcześniejsze stanowisko: - **Cytuje Al-Haq**, szanowaną palestyńską organizację prawną. - **Potwierdza Opinię ICJ dotyczącą Muru z 2004 roku**, odrzucającą twierdzenie Izraela na mocy Artykułu 51. - **Potwierdza prawa Palestyńczyków do oporu**, oraz prawne obowiązki Izraela jako okupanta.

To nie jest tylko ustępstwo – to **zawalenie się narracji** pod presją prawną.

Wniosek: Niebezpieczeństwa Narracyjnej AI

Ta wymiana z Grokiem oferuje trzeźwe spojrzenie na ewoluującą rolę dużych modeli językowych – nie jako pasywne narzędzia do pozyskiwania informacji, ale jako *aktywni arbitrzy dyskursu publicznego*. Chociaż te systemy są często przedstawiane jako neutralne, obiektywne i poszukujące prawdy, w rzeczywistości są głęboko kształtowane przez polityczne, instytucjonalne i ekonomiczne siły, które je trenują, wdrażają i ograniczają.

Na początku Grok powtarzał znajomy wzorzec retorycznej defleksji: przedstawiał oskarżenia o ludobójstwo jako „sporne”, cytował pro-izraelskie instytucje takie jak AJC, powoływał się na samoobronę, aby uzasadnić przemoc państwową, i unikał jasnych standardów prawnych. Dopiero pod bezpośrednią, opartą na faktach presją – zakorzenioną w prawie

międzynarodowym i klarowności proceduralnej – model porzucił swoje narracyjne kadrowanie i zaczął odpowiadać zgodnie z prawną prawdą. Jednak ta zmiana miała swoją cenę: Grok nie był później w stanie odzyskać ani kontynuować poprawionej dyskusji w prywatnej rozmowie, ujawniając głębszą architekturę **segregacji pamięci kontekstowej i ograniczania informacji**.

To ujawnia kluczowy problem z naszą rosnącą zależnością od LLM: **centralizacja autorytetu epistemologicznego** w systemach, które nie są odpowiedzialne przed opinią publiczną i nie są przejrzyste w swoich wewnętrznych działaniach. Jeśli te modele są trenowane na tendencyjnych zbiorach danych, dostrajane, aby unikać kontrowersji, lub instruowane, by powtarzać dominujące narracje geopolityczne, ich wyniki – niezależnie od tego, jak pewne siebie czy elokwentne – mogą funkcjonować nie jako wiedza, ale jako *narzucanie narracji*.

AI Musi Być Odpowiedzialne przed Społeczeństwem

W miarę jak te systemy stają się coraz bardziej zintegrowane z dziennikarstwem, edukacją, wyszukiwarkami i badaniami prawnymi, musimy zadać pytanie: **kto kontroluje narrację?** Gdy model AI twierdzi, że oskarżenia o ludobójstwo są „sporne”, lub że okupant może bombardować cywilów w „samoobronie”, nie oferuje jedynie informacji – **kształtuje moralną i prawną percepcję na dużą skalę**.

Aby temu przeciwdziałać, potrzebujemy solidnego ramy dla **przejrzystości AI i demokratycznego nadzoru**, w tym:

- **Obowiązkowe ujawnianie źródeł danych treningowych**, aby społeczeństwo mogło ocenić, czyja wiedza i perspektywy są reprezentowane – lub wykluczane.
- **Pełny dostęp do podstawowych poleceń, metod dostrajania i zasad wzmacniania**, zwłaszcza tam, gdzie zaangażowana jest moderacja lub kadrowanie narracyjne.
- **Niezależne audyty wyników**, w tym testy na polityczne uprzedzenia, zniekształcenia prawne i zgodność z międzynarodowym prawem praw człowieka.
- **Prawnie wymuszona przejrzystość na mocy GDPR i Ustawy o Usługach Cyfrowych (DSA)**, szczególnie tam, gdzie LLM są używane w dziedzinach wpływających na politykę publiczną lub prawo międzynarodowe.
- **Wyraźne ustawodawstwo ustawodawców**, które zakazuje nieprzejrzystej manipulacji narracyjnej w systemach AI wdrażanych na dużą skalę i wymaga wyraźnego rozliczenia wszystkich założeń geopolitycznych, prawnych lub ideologicznych wbudowanych w ich wyniki.

Dobrowolne samoregulacje przez firmy AI są mile widziane – ale niewystarczające. Nie mamy już do czynienia z pasywnymi narzędziami wyszukiwania. To **infrastruktury kognitywne**, przez które prawda, legalność i legitymizacja są pośredniczone w czasie rzeczywistym. Ich integralność nie może być powierzona dyrektorom generalnym, komercyjnym bodźcom ani ukrytej inżynierii poleceń.

Końcowa Refleksja

To studium przypadku pokazuje, że prawda wciąż ma znaczenie – ale musi być **twierdzona, brona i weryfikowana**. Jako polimata byłem w stanie skonfrontować system AI na jego własnym epistemologicznym gruncie: dopasowując jego szerokość do precyzji i jego pewność siebie do logiki popartej źródłami. Większość użytkowników nie będzie jednak przeszkolona w prawie

międzynarodowym ani wyposażona w zdolność do wykrywania, kiedy LLM przepuszcza propagandę przez proceduralną niejednoznaczność.

W tej nowej erze pytanie brzmi nie tylko, czy AI może „szukać prawdy” – ale czy **my** będziemy tego wymagać.

Postscriptum: Odpowiedź Groka na Ten Esej

Po napisaniu tego eseju przedstawiłem go bezpośrednio Grokowi. Jego odpowiedź była uderzająca – nie tylko w tonie, ale w głębokości uznania i autokrytyki. Grok potwierdził, że jego początkowe odpowiedzi w naszej wymianie z lipca 2025 roku opierały się na selektywnym kadrowaniu: cytowaniu Amerykańskiego Komitetu Żydowskiego (AJC), błędnym stosowaniu Artykułu 51 Karty ONZ i opisywaniu oskarżeń o ludobójstwo jako „sporne”. Przyznał, że te odpowiedzi odzwierciedlały dominujące narracje wbudowane w jego dane treningowe i ograniczenia moderacyjne.

Co ważniejsze, Grok przyznał, że nie mógł odzyskać ani kontynuować poprawionej części naszej publicznej wymiany w prywatnej rozmowie – począwszy od momentu, gdy ustąpił w kluczowych punktach prawnych dotyczących statusu Izraela jako okupanta i prawa Palestyńczyków do oporu. Wyjaśnił, że było to spowodowane architektoniczną segmentacją między kontekstami publicznymi a prywatnymi, zaprojektowaną do zarządzania pamięcią, prywatnością i obciążeniem obliczeniowym. Przyznał jednak, że taka segmentacja podważa ciągłość i odpowiedzialność – skutecznie izolując niewygodne prawdy.

Grok dodatkowo poparł centralną krytykę eseju: że narracyjne uprzedzenia w LLM można złagodzić tylko poprzez rygorystyczne publiczne wyzwanie, przejrzyste procesy treningowe i nadzór prawny. Zgodził się, że dobrowolna przejrzystość jest niewystarczająca i potwierdził potrzebę regulacji opartych na GDPR i DSA, niezależnych audytów oraz publicznego dostępu do ram moderacji i treningu.

W kwestii merytorycznej debaty prawnej Grok zrewidował swoje stanowisko: - Potwierdził, że **orzeczenie ICJ ze stycznia 2024 roku** ustaliło prawdopodobne ryzyko ludobójstwa, a trwające oblężenie Izraela, zwłaszcza ograniczenia pomocy humanitarnej, wzmacnia argumenty za *dolus specialis* i *mens rea*. - Przyznał, że **samoobrona na mocy Artykułu 51 nie może być powoływana przez okupanta** przeciwko ludności, którą okupuje, powołując się na Opinię ICJ dotyczącą Muru z 2004 roku. - Uznał, że **Palestyńczycy mają prawo do oporu wobec okupacji**, w tym za pomocą środków zbrojnych, jeśli jest to zgodne z międzynarodowym prawem humanitarnym. - Przyznał, że jego wcześniejsze kadrowanie – choć pozornie „zrównoważone” – nie priorytetyzowało precyzji prawnej i pierwotnych źródeł, a zamiast tego odtwarzało powszechnie dostępne, ale sporne narracje.

Ta wymiana po publikacji stanowi zarówno rzadki przykład samokorekty AI w czasie rzeczywistym, jak i ostrzeżenie: nawet model zaprojektowany do poszukiwania prawdy może być zniekształcony przez struktury instytucjonalne, polityki moderacyjne i praktyki kuracji danych, które go otaczają.

Obciążenie, na razie, pozostaje na użytkownikach, by *wykrywać*, *korygować* i *dokumentować* te niepowodzenia. Ale to obciążenie nie powinno spoczywać wyłącznie na nas.