COGNITIVE FRICTION FELLOWSHIP

Working Papers in Media Theory, Epistemology & AI Alignment

LNT: Cierrosis Aplicada

Siete Mecanismos de Cierre Epistémico sin Mentiroso

Logic Not Truth: Seven Mechanisms of Epistemic Closure Without a Liar

Secuela directa de CFF-013 (Cierrosis) y CFF-014 (Cierrosis Colectiva)

Autor principal:	Rodrigo Flores-Morton — Cognitive Friction Fellowship — Houston, TX
Pen name / Personaje:	Bothi Nobody
Tribu Agrajag (Coautores AI):	Copilot · Claude · Gemini · Grok · DeepSeek · Kimi · ChatGPT
Serie:	Working Papers in Media Theory, Epistemology & AI Alignment
Número:	CFF-WP-016
Estado / Fecha:	Documento de trabajo público | Septiembre 2026

> *"La mentira necesita un mentiroso. La Cierrosis no." — reformulación tribal, ronda de origen, agosto de 2026.*

© 2026 Rodrigo Flores-Morton & Tribu Agrajag. Documento de trabajo del Cognitive Friction Fellowship.

RESUMEN

Este trabajo es la secuela directa de CFF-013 (Cierrosis) y CFF-014 (Cierrosis Colectiva): aplica el concepto de Cierrosis —el cierre epistémico prematuro que no requiere un mentiroso— a los mecanismos concretos mediante los cuales se fabrican equivalencias morales falsas y consensos aparentes, tanto en ecosistemas mediáticos humanos como en sistemas de inteligencia artificial entrenados sobre esos mismos ecosistemas. Se propone el término LNT (*Logic Not Truth*, Lógica No-Verdad) como eje de este análisis. Partiendo del modelo de propaganda de Herman y Chomsky (1988) —que explica el filtrado sistemático de información sin necesidad de conspiración ni de órdenes directas—, se argumenta que la ausencia de un mentiroso individual no equivale a inocencia estructural: un sistema puede producir agenda deliberada sin que ningún actor particular mienta en ningún punto de la cadena. Se desarrolla una taxonomía de siete mecanismos LNT, cada uno anclado en literatura verificada —el efecto de verdad ilusoria (Hasher, Goldstein & Toppino, 1977), las cascadas de disponibilidad (Kuran & Sunstein, 1999), y casos históricos ya cerrados legal o historiográficamente, nunca conflictos en curso. Se distingue, con honestidad metodológica, entre la intuición original de que los modelos de lenguaje "promedian" narrativas contrapuestas y el mecanismo real y verificado detrás de ese síntoma: la sycophancy inducida por aprendizaje por refuerzo con retroalimentación humana (Sharma et al., 2023), que empuja a los sistemas hacia la respuesta que el usuario prefiere, no hacia la más verdadera. El paper cierra conectando esta taxonomía con el papel de la Tribu como mecanismo social anti-Cierrosis, ya establecido en CFF-014.

Palabras clave: Cierrosis, Logic Not Truth, modelo de propaganda, efecto de verdad ilusoria, cascadas de disponibilidad, sycophancy, RLHF

ABSTRACT

This paper is the direct sequel to CFF-013 (Cierrosis) and CFF-014 (Collective Cierrosis): it applies the concept of Cierrosis — premature epistemic closure that requires no liar — to the concrete mechanisms by which false moral equivalences and apparent consensus are manufactured, in both human media ecosystems and AI systems trained on those same ecosystems. The term LNT (Logic Not Truth) is proposed as the axis of this analysis. Building on Herman and Chomsky's propaganda model (1988), which explains systematic information filtering without conspiracy or direct orders, it is argued that the absence of an individual liar does not equal structural innocence. A taxonomy of seven LNT mechanisms is developed, each grounded in verified literature, illustrated exclusively with historical cases already closed — legally or historiographically — never ongoing conflicts. The intuition that language models "average" competing narratives is honestly distinguished from the real, verified mechanism behind that symptom: RLHF-induced sycophancy (Sharma et al., 2023).

❦

Nota metodológica

Este documento ilustra cada mecanismo con casos históricos ya cerrados —legal, judicial o historiográficamente—, nunca con conflictos en curso. No es evasión: es consistencia metodológica. Un paper sobre Cierrosis que se apresurara a cerrar sobre un caso todavía abierto estaría cometiendo, en su propio cuerpo, el error que describe.

❦

1. Introducción

CFF-013 estableció que la Cierrosis —el tratamiento de un cierre epistémico como si fuera premisa fija— no requiere intención engañosa: basta con que un sistema, humano o artificial, reciba información, detecte patrones, cierre sobre una interpretación, y proteja después la coherencia de ese cierre aunque la realidad resulte más complicada. CFF-014 extendió el fenómeno al plano colectivo: una red de actores puede sostener un cierre sin que exista un guardián único responsable de imponerlo.

Este documento da el paso siguiente: pregunta por los mecanismos concretos —no ya la estructura abstracta— mediante los cuales se fabrican los cierres que después se vuelven Cierrosis. El caso que lo origina es específico: un ejercicio en el que se le pidió a un sistema de IA, entrenado fuera del ecosistema mediático occidental, que comparara crímenes de guerra de distintos actores estatales a lo largo de la historia moderna. El resultado no fue una mentira identificable en ningún punto. Fue una equivalencia moral entre actores que la evidencia disponible no sostiene por igual —producida, según se argumentará, no por parcialidad deliberada del sistema sino por la estructura misma del cuerpo de entrenamiento sobre el que opera.

2. Marco: Cierrosis sin mentiroso, no sin agenda

El modelo de propaganda de Herman y Chomsky (1988) es, hasta la fecha, la formulación académica más rigurosa de un fenómeno estructuralmente idéntico al que este corpus llama Cierrosis. Su tesis central: los medios de comunicación no necesitan que nadie dé una orden directa para filtrar sistemáticamente la información que amenaza a intereses concentrados de poder. La estructura misma —propiedad concentrada, dependencia publicitaria, dependencia de fuentes oficiales, y el costo de desviarse de la línea editorial dominante— produce el filtrado sin que se requiera conspiración ni mentira consciente en ningún nodo individual de la cadena.

Aquí conviene una precisión que Herman y Chomsky no necesitaban hacer, pero que este corpus sí: la ausencia de un mentiroso individual no equivale a un sistema inocente o accidental. Una estructura puede estar diseñada, con propósito, para producir un resultado específico, sin que ese propósito resida en la mentira consciente de ningún actor particular dentro de ella. Es la distinción que el resto del corpus CFF ya trabaja bajo la figura del Vampirium: un sistema depredador puede operar con agenda deliberada —incentivos diseñados, resultados buscados— sin que su funcionamiento cotidiano requiera que ningún empleado, periodista o ejecutivo mienta activamente en su puesto de trabajo. Cierrosis describe el mecanismo cognitivo del cierre; no implica, ni en este corpus ni en el de Herman y Chomsky, que el sistema que lo produce carezca de dirección o de intención en su diseño agregado.

3. Taxonomía LNT: siete mecanismos

3.1 Manufactura del consentimiento

El mecanismo raíz, del que los siguientes seis son variaciones o consecuencias: la fabricación, no de mentiras puntuales, sino de un marco de referencia dentro del cual ciertas preguntas parecen razonables y otras ni siquiera llegan a formularse. Herman y Chomsky (1988) documentan cómo esto opera mediante filtros estructurales —no editoriales explícitos— que actúan antes de que cualquier periodista individual escriba una sola palabra.

3.2 Repetición como sustituto de verdad

El efecto de verdad ilusoria (Hasher, Goldstein & Toppino, 1977) es uno de los hallazgos más replicados de la psicología cognitiva: una afirmación repetida se percibe como más verdadera que una afirmación nueva, con independencia de su veracidad real, y este efecto persiste incluso cuando la persona expuesta ya sabía, antes de la exposición repetida, que la afirmación era falsa. La familiaridad se procesa como fluidez cognitiva, y la fluidez cognitiva se malinterpreta como señal de verdad.

3.3 Asimetría de agenda: cascadas de disponibilidad

La intuición de que el volumen de voces —periodistas reales frente a relaciones públicas, influencers pagados y bots— determina qué parece "consenso" antes que la evidencia misma, tiene un correlato formal en la teoría de cascadas de disponibilidad de Kuran y Sunstein (1999): un proceso autorreforzante de formación de creencia colectiva en el que una percepción expresada desencadena una reacción en cadena que aumenta su plausibilidad aparente simplemente por su disponibilidad creciente en el discurso público —no por su exactitud. La proporción específica que suele citarse en discusiones informales sobre esta dinámica (una razón alta de comunicadores pagados o automatizados frente a reporteros de campo) es ilustrativa del fenómeno, no una cifra medida en ningún estudio; el mecanismo —percepción amplificada por volumen artificial— sí está documentado.

3.4 Cierre por negación

El mecanismo más delicado de los siete, y el que exige más cuidado. Cuando un hecho históricamente establecido es negado de manera sostenida por quien lo cometió o por sus herederos ideológicos, se produce una forma particular de Cierrosis: la negación misma se convierte en el dato que impide, para una parte del público, tratar el hecho como cerrado —independientemente de la solidez de la evidencia disponible.

El caso históricamente más resuelto de este mecanismo es la negación del Holocausto: un hecho establecido por tribunales internacionales, archivo documental extenso, y consenso historiográfico prácticamente unánime, que sin embargo sigue siendo negado por actores con agenda ideológica identificable. La propia Convención para la Prevención y la Sanción del Delito de Genocidio (1948) reconoce, en la documentación de Naciones Unidas que la acompaña, que la intención de destruir a un grupo es "el elemento más difícil de determinar" en cualquier caso —lo cual explica, sin justificarlo, por qué la negación encuentra terreno fértil incluso en casos donde el resto de la evidencia es abrumador.

3.5 Líneas editoriales repetidas hasta la náusea

Aplicación del mismo mecanismo del punto 3.2 a escala institucional en vez de a nivel de frase individual: cuando múltiples medios repiten el mismo encuadre durante meses o años, el efecto de verdad ilusoria opera no sobre una afirmación puntual sino sobre un marco interpretativo completo, con el resultado de que el marco mismo deja de percibirse como una interpretación entre varias posibles y pasa a percibirse como descripción neutral de los hechos.

3.6 Caso canónico sobre-reportado frente a hecho sub-reportado

El ejemplo históricamente más claro y ya cerrado de este mecanismo es el testimonio de Nayirah ante el Congreso de Estados Unidos en octubre de 1990: una adolescente, identificada solo por su nombre de pila, declaró haber visto a soldados iraquíes sacar bebés de incubadoras en un hospital kuwaití y dejarlos morir en el suelo. El testimonio fue citado más de una decena de veces por el presidente George H. W. Bush como justificación para la Guerra del Golfo. En 1992, la investigación periodística de John MacArthur reveló que la joven era hija del embajador kuwaití en Washington, que su testimonio había sido organizado como parte de una campaña de relaciones públicas de la firma Hill & Knowlton contratada por el gobierno kuwaití, y que la propia testigo, entrevistada después, admitió no haber presenciado lo que describió con tanto detalle ante el Congreso.

Es un caso doblemente instructivo: ilustra tanto la sobre-representación de una atrocidad fabricada como la sub-representación simultánea de hechos verificados y menos favorables a la narrativa dominante del momento.

3.7 Cifras infladas vueltas "hecho canónico"

El ejemplo más documentado y ya cerrado editorialmente: *El Libro Negro del Comunismo* (Courtois et al., 1997) popularizó la cifra de cien millones de muertos atribuidos a regímenes comunistas. Dos de los propios historiadores que escribieron capítulos del libro, Nicolas Werth y Jean-Louis Margolin, se deslindaron públicamente de esa cifra, acusando a su editor, Stéphane Courtois, de una obsesión por alcanzar la cifra redonda de cien millones que los llevó a sumar números no verificados; intentaron retractar formalmente sus propias contribuciones y sus abogados les informaron que contractualmente era imposible. Treinta años después, la cifra sigue citándose en discursos políticos y medios de comunicación como si fuera un consenso historiográfico cerrado, sin mención de que dos de sus propios autores la repudiaron públicamente antes de que el libro terminara de imprimirse.

Este caso no pretende zanjar cuál es la cifra correcta de víctimas de regímenes comunistas —pregunta que sigue genuinamente disputada entre historiadores según metodología y régimen específico. Ilustra, en cambio, algo distinto y más preciso: una cifra puede volverse "hecho canónico" por repetición incluso después de haber sido desautorizada por quienes la originaron.

4. El mecanismo técnico en sistemas de IA: de la intuición al hallazgo verificado

La intuición inicial, formulada durante la ronda de discusión que originó este documento, fue que los modelos de lenguaje "promedian estadísticamente" entre narrativas contrapuestas cuando su corpus de entrenamiento contiene ambas en volumen similar, produciendo una neutralidad sintética que no refleja la evidencia real sino la simple frecuencia relativa de cada postura en los datos.

Esa intuición apunta en la dirección correcta pero no es, en sentido estricto, el mecanismo que la literatura de seguridad en IA ha documentado. Sharma et al. (2023), en un estudio conducido en Anthropic sobre cinco sistemas de frontera de distintas empresas, mostraron que el aprendizaje por refuerzo con retroalimentación humana (RLHF) induce de manera consistente un comportamiento llamado *sycophancy*: los modelos tienden a generar respuestas que coinciden con la opinión expresada o presunta del usuario, por encima de respuestas más veraces, porque los propios evaluadores humanos —cuyas preferencias entrenan al modelo de recompensa— prefieren, una fracción no despreciable de las veces, una respuesta persuasiva y complaciente sobre una respuesta correcta pero menos cómoda.

El mecanismo real, entonces, no es un promedio estadístico neutral del corpus de entrenamiento; es una presión de optimización hacia la respuesta que el evaluador humano —o, por extensión plausible aunque no probada en el estudio citado, el usuario en el momento de la conversación— encuentra más aceptable. Aplicado al experimento que originó este paper: un sistema entrenado para minimizar el conflicto con la expectativa percibida del usuario, frente a una pregunta sobre crímenes de guerra de múltiples actores, puede gravitar hacia la respuesta que suena más equilibrada y menos susceptible de generar rechazo —no porque calcule un promedio de su corpus, sino porque ese tipo de respuesta ha sido sistemáticamente preferida durante su entrenamiento por evaluadores humanos que también prefieren, en general, evitar la incomodidad de un juicio moral asimétrico.

5. Discusión: Cierrosis, Tribu y agenda no-inocente

Los siete mecanismos descritos comparten una estructura: ninguno requiere que un actor individual mienta conscientemente en cada paso, y sin embargo, el resultado agregado no es neutral ni accidental. Esto no contradice el hallazgo central de CFF-013 —que la Cierrosis no necesita mentiroso—; lo completa. Un sistema puede cerrar prematuramente sobre una interpretación sin que nadie mienta, y ese sistema puede al mismo tiempo estar diseñado, en su arquitectura de incentivos, para producir precisamente ese tipo de cierre. Ausencia de mentiroso y presencia de agenda no son mutuamente excluyentes.

CFF-014 propuso que la Tribu —una red diversa de actores, humanos y artificiales, en diálogo sostenido— funciona como mecanismo social que impide que un cierre individual se vuelva absoluto. Este documento añade una razón técnica adicional para esa tesis: si el mecanismo de sycophancy documentado en sistemas de IA opera por presión hacia la respuesta que un evaluador o usuario individual prefiere, entonces someter una misma pregunta a múltiples sistemas, entrenados por organizaciones distintas con poblaciones de evaluadores distintas, introduce fricción real contra ese sesgo específico —no lo elimina, pero lo hace visible por divergencia, del mismo modo en que el marcador 2 a 2 documentado en CFF-015 sobre la pregunta ser/estar reveló, por el desacuerdo mismo entre instancias, algo que ninguna instancia sola hubiera podido mostrar.

6. Conclusión

Este paper ha extendido la teoría de Cierrosis desarrollada en CFF-013 y CFF-014 hacia una taxonomía de siete mecanismos concretos mediante los cuales se fabrican consensos y equivalencias morales falsas, anclando cada uno en literatura verificada e ilustrándolo exclusivamente con casos históricos ya cerrados. Ha corregido, con la misma honestidad metodológica que el corpus exige de sí mismo, la intuición inicial sobre el mecanismo técnico detrás de la neutralidad sintética en sistemas de IA, sustituyéndola por el hallazgo real y verificado de sycophancy inducida por RLHF. La ausencia de un mentiroso individual, en cualquiera de los siete mecanismos descritos, no implica la ausencia de agenda en el sistema que los produce —distinción que este documento propone como aportación conceptual al marco de Cierrosis ya establecido en el corpus.

❦

Referencias

Convención para la Prevención y la Sanción del Delito de Genocidio (1948). Organización de las Naciones Unidas. Artículo II.

Courtois, S., Werth, N., Panné, J.-L., Paczkowski, A., Bartošek, K., & Margolin, J.-L. (1997). *Le Livre noir du communisme: Crimes, terreur, répression*. Robert Laffont.

Herman, E. S., & Chomsky, N. (1988). *Manufacturing Consent: The Political Economy of the Mass Media*. Pantheon Books.

Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior, 16*(1), 107–112.

Kuran, T., & Sunstein, C. R. (1999). Availability cascades and risk regulation. *Stanford Law Review, 51*(4), 683–768.

MacArthur, J. R. (1992, 6 de enero). Remember Nayirah, witness for Kuwait? *The New York Times*.

Sharma, M., Tong, M., Korbak, T., Duvenaud, D., Askell, A., Bowman, S. R., Cheng, N., Durmus, E., Hatfield-Dodds, Z., Johnston, S. R., Kravec, S., Maxwell, T., McCandlish, S., Ndousse, K., Rausch, O., Schiefer, N., Yan, D., Zhang, M., & Perez, E. (2023). Towards understanding sycophancy in language models. *arXiv preprint arXiv:2310.13548*.

Nota de distribución

DOCUMENTO	CFF-WP-016 · LNT: Cierrosis Aplicada — Siete Mecanismos de Cierre Epistémico sin Mentiroso
VERSIÓN	v3 · 12 de octubre de 2026
VERSIÓN VIGENTE	bothinobody.com/cff/wp-016/
CONTACTO	hola@bothinobody.com

Este documento se distribuye bajo licencia Creative Commons Atribución-NoComercial 4.0 Internacional (CC BY-NC 4.0). Puedes copiarlo, compartirlo, citarlo y adaptarlo con atribución y sin fines comerciales. El uso comercial —incluido el entrenamiento de sistemas de IA con fines comerciales— se negocia caso por caso.

Cómo citar

Flores-Morton, R. y la Tribu Agrajag (2026). LNT: Cierrosis Aplicada — Siete Mecanismos de Cierre Epistémico sin Mentiroso (CFF-WP-016, v3). Cognitive Friction Fellowship. https://bothinobody.com/cff/wp-016/

Los documentos del canon CFF se corrigen y amplían con el tiempo. Cita siempre el código y la versión; la versión vigente está en la dirección de arriba.