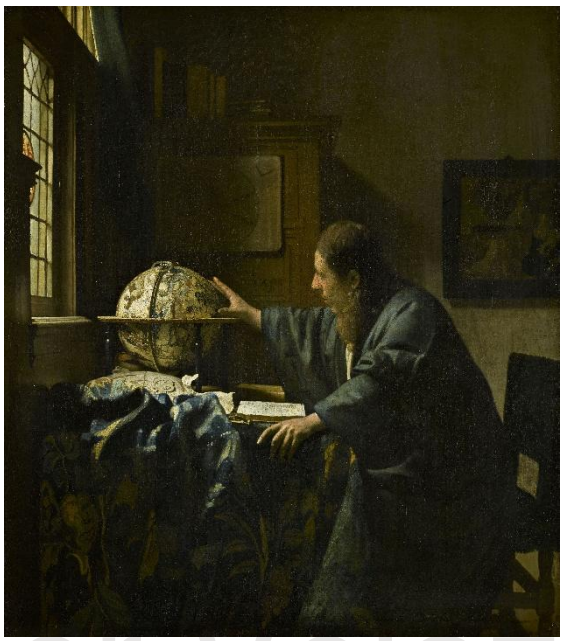
Hacia una filosofía de la razón medible

Luminomática (LUM) como metrología operatoria del cierre científico



Ensayo filosófico sobre Luminomática (LUM)

Julio David Rojas

30/05/26

«Si he visto más lejos, es porque estoy sentado sobre hombros de gigantes.»

— Isaac Newton, carta a Robert Hooke

«La filosofía no es una ciencia entre las ciencias, sino una reflexión crítica sobre las ciencias y sobre los demás saberes.»

— Gustavo Bueno, El papel de la filosofía en el conjunto del saber.

Índice

Introducción: cuando la razón necesita instrumentos

La modernidad no carece de datos. Tiene más de los que puede procesar, más de los que puede interpretar, más de los que cualquier institución puede metabolizar en tiempo razonable. Lo que escasea no es información sino algo más difícil de fabricar: un criterio compartido para saber cuándo esa información alcanza cierre suficiente como para justificar acción.

Hay algo desconcertante en este diagnóstico. La filosofía de la ciencia lleva casi un siglo elaborando criterios de demarcación, distinguiendo la ciencia de la pseudociencia, describiendo cómo avanzan los paradigmas, analizando las condiciones del conocimiento. Y sin embargo, cuando una agencia regulatoria pregunta si la terapia génica está lo suficientemente madura para aprobar un protocolo; cuando un ministerio de salud necesita saber si los resultados sobre intervenciones psicológicas resisten la traslación a política pública; cuando un fondo de investigación quiere decidir si determinada línea de investigación merece seguir financiándose —en esos momentos, la filosofía de la ciencia no tiene mucho que ofrecer. Tiene diagnósticos históricos y vocabulario preciso, pero no tiene procedimientos accionables.

Esta brecha entre el pensamiento filosófico y la necesidad operativa no es una deficiencia accidental. Es estructural. Y es exactamente el espacio en que emerge la Luminomática (LUM): un sistema gnoseo-operatorio diseñado para determinar cuándo un campo científico ha alcanzado cierre suficiente para justificar acción, y qué hacer cuando no lo ha alcanzado.

Definida con mayor precisión, la Luminomática es una metrología filosófico-operatoria de la razón científica. No pretende sustituir a la ciencia, ni decidir desde fuera qué teorías son verdaderas, ni convertir la filosofía en algoritmo administrativo. Su función es más concreta: evaluar el grado de cierre de un campo de conocimiento cuando ese campo debe sostener decisiones prácticas. Para ello examina la coherencia interna de la evidencia, la estabilidad de sus operaciones, la presencia de contradicciones activas, la trazabilidad de sus procedimientos y la proporcionalidad entre lo que el campo demuestra y lo que se quiere hacer con él.

LUM no pregunta simplemente si una teoría parece convincente. Pregunta si un cuerpo de evidencia, situado en una ventana histórica determinada, permite pasar responsablemente de la interpretación a la acción. Por eso su unidad de análisis no es la opinión individual, ni la teoría aislada, ni el consenso institucional tomado como autoridad final, sino el campo operatorio: protocolos, datos, instrumentos, modelos, replicaciones, controversias, umbrales y efectos de decisión. Su resultado no es una verdad absoluta, sino un dictamen de cierre: cierre suficiente para actuar, no-cierre que exige abstención o investigación adicional, o sobrecierre cuando una apariencia excesiva de orden encubre fragilidad, sesgo o contradicciones no resueltas.

En este sentido, LUM puede entenderse como una disciplina de prudencia racional: mide hasta dónde puede llegar una afirmación sin fingir más certeza de la que sus materiales permiten.

La tesis de este ensayo es la siguiente: al convertir el cierre científico en objeto de reconstrucción, medición y revisión, LUM transforma la razón científica en una práctica auditable. Conocer implica producir cierres operatorios reconstruibles, medibles y revisables. LUM convierte el problema clásico de la demarcación científica en una metrología del cierre, y convierte el no-cierre en una ética de la abstención y del trabajo pendiente. La razón se mide por la disciplina con que distingue cierre, no-cierre y sobrecierre.

Esta tesis exige una precaución desde el inicio. LUM conserva el juicio filosófico y lo obliga a declarar sus condiciones. Antes de medir cierre hay que decidir qué cuenta como campo, qué cuenta como evidencia, qué cuenta como incompatibilidad y qué tipo de acción se pretende justificar. Esa decisión previa se vuelve más visible al quedar formalizada. La fuerza filosófica de LUM consiste en impedir que el juicio se esconda detrás de la autoridad, la intuición o el consenso. La metrología pone a la filosofía bajo contrato.

Esa disciplina tiene consecuencias. Una de ellas es incómoda: si la razón debe ser auditable, entonces declarar cierre donde no lo hay es una forma de fraude epistémico. La otra consecuencia es liberadora: cuando no hay cierre, se puede decirlo con precisión, y esa precisión es ya una contribución al conocimiento.

La apuesta propia de este ensayo es que LUM permite formular una filosofía de segundo grado sobre la razón científica. Su interés filosófico aparece al

preguntar qué imagen de la racionalidad surge cuando el conocimiento debe dejar rastro, admitir revisión y responder por sus condiciones de validez. LUM importa por lo que mide y por lo que obliga a pensar: la diferencia entre verdad y cierre, prudencia y parálisis, consenso y auditoría, pluralismo legítimo y contradicción no resuelta. En ese desplazamiento empieza su alcance filosófico.

El ensayo está organizado como una progresión. Primero sitúa el problema histórico: qué logró y qué no pudo lograr la filosofía clásica de la ciencia. Después desarrolla el núcleo filosófico de LUM: qué significa el cierre categorial, qué significa medirlo y cómo se operacionaliza en cuatro ejes. Luego trabaja los estados del sistema —CLARION, no-cierre, sobrecierre y auditoría—, incorpora un caso mínimo de no-cierre y responde las objeciones más serias. Finalmente, despliega la dimensión ética y una proyección controlada: pensar la realidad como historia de cierres locales sin convertir LUM en metafísica total.

Nota metodológica: leer LUM como problema filosófico

Este ensayo no toma la LUM como objeto puramente técnico ni como doctrina filosófica cerrada. La lee como un dispositivo gnoseológico: un sistema que obliga a preguntar bajo qué condiciones una afirmación sobre un campo puede pasar de la interpretación a la acción responsable. Por eso el análisis seguirá tres planos. Primero, el plano histórico-filosófico: la relación de LUM con los criterios clásicos de demarcación, especialmente Popper, Kuhn, Lakatos y Bueno. Segundo, el plano operatorio: la forma en que LUM traduce el cierre en índices, umbrales, estados y procedimientos de auditoría. Tercero, el plano crítico: los límites del sistema, sus riesgos de tecnocratización y el problema de que toda medición presupone decisiones filosóficas previas.

Esta lectura asume una meta modesta y defendible: mostrar que LUM desplaza la pregunta filosófica desde la definición abstracta de ciencia hacia la evaluación operatoria de campos concretos. La tesis sostiene que toda razón orientada a acciones colectivas debe declarar sus condiciones de cierre, sus límites y sus procedimientos de reconstrucción.

I. El problema que la filosofía nombró pero no pudo medir

Existe una diferencia entre diagnosticar una enfermedad y saber cuándo el paciente está curado. La filosofía de la ciencia ha sido, durante casi un siglo, extraordinariamente hábil en el primer tipo de tarea y casi inútil en el segundo.

La gran tradición de la epistemología del siglo XX produjo diagnósticos lúcidos. Karl Popper identificó que la falsabilidad distingue la ciencia de la metafísica: una proposición es científica si puede ser refutada por la experiencia (Popper, 1959). Este fue un avance conceptual real. Antes de Popper, la demarcación era con frecuencia un asunto de intuición, autoridad disciplinar o tradición metodológica. Después de él, había al menos un criterio estructural: la refutabilidad. Pero falsabilidad en principio no equivale a suficiencia empírica para actuar. Una teoría puede ser falsable y sin embargo sobrevivir indefinidamente rodeada de hipótesis auxiliares, parches ad hoc que absorben cada anomalía sin comprometer el núcleo. La psicología del priming fue falsable en teoría; y sin embargo décadas de replicaciones problemáticas no la desacreditaron hasta que el peso de la evidencia acumulada se volvió insostenible. El criterio de Popper no fue diseñado para responder cuándo ese campo estaba maduro para guiar intervención pública.

Thomas Kuhn ofreció una descripción más rica de cómo funciona la ciencia real: bajo paradigmas que estructuran qué preguntas son legítimas, qué evidencia cuenta y qué anomalías se toleran hasta que se vuelven insoportables (Kuhn, 2012). La revolución científica ocurre cuando las anomalías se acumulan hasta hacer inviable el paradigma anterior. Pero la teoría de Kuhn es retrospectiva: explica por qué hubo revolución después de que ocurrió, no cuándo un paradigma está lo suficientemente maduro para guiar política pública. No proporciona umbrales ni procedimientos reproducibles.

Imre Lakatos refinó a Kuhn con la distinción entre programas de investigación progresivos y degenerativos: un programa avanza cuando anticipa hechos nuevos y degenera cuando solo protege su núcleo mediante ajustes defensivos (Lakatos & Musgrave, 1970). Un programa es progresivo si sus predicciones superan sus protecciones ad hoc; degenerativo si solo se protege absorbiendo anomalías sin generar predicciones nuevas. Fue un paso enorme hacia la evaluación dinámica de la ciencia, pero Lakatos tampoco especificó

umbrales cuantitativos ni procedimientos que un evaluador externo pueda reproducir. La distinción progresivo/degenerativo sigue requiriendo juicio experto. No es computable. No es auditable por terceros.

Paul Feyerabend, en el extremo opuesto, disolvió la pretensión de reglas metodológicas universales al mostrar que varios episodios decisivos de la historia científica violaron las normas que luego se presentaron como condiciones del método (Feyerabend, 1975). Contra método, su obra más conocida, demuestra que los grandes avances científicos históricos a menudo violaron las normas metodológicas de su tiempo. La conclusión no es que cualquier cosa vale epistemológicamente, sino que las reglas rígidas obstaculizan la creatividad científica. Pero Feyerabend tampoco ofrece un sustituto operativo para la demarcación. Su anarquismo epistemológico es una crítica, no un sistema.

Gustavo Bueno formuló, desde el Materialismo Filosófico, el concepto de cierre categorial: un campo tiene cierre cuando sus operaciones son estables, se conectan con otras a través de puentes y producen identidades sintéticas que neutralizan al sujeto operatorio (Bueno, 1992–1993, 1995). Es decir: el resultado no depende de quién ejecuta la operación, sino del método. Cuando varios técnicos independientes, usando el mismo protocolo, obtienen siempre el mismo resultado al medir la misma muestra, el sujeto ha sido neutralizado por el procedimiento. Eso es cierre. Este concepto es quizás el más cercano filosóficamente a lo que LUM operacionaliza. Su limitación es que Bueno no tradujo el cierre categorial a un algoritmo: describió la condición filosófica del cierre sin especificar cómo medirla. La operacionalización quedó pendiente.

El resultado de esta historia es un vacío de segunda capa. La filosofía ha nombrado el problema con creciente precisión: ¿cuándo un campo alcanza cierre suficiente para justificar acción? Pero no ha podido medirlo. La capa de la decisión institucional —¿aprobamos este protocolo?, ¿financiamos esta intervención?, ¿diseñamos política sobre este cuerpo de evidencia?— ha operado sin criterio formal, dependiendo de la autoridad del evaluador, del consenso del gremio o de la intuición del tomador de decisiones. LUM opera en esa segunda capa. No refuta a ninguno de estos autores: complementa sus diagnósticos con el procedimiento que les faltó.

La pregunta que LUM responde no es: ¿esta teoría es científica? Sino: ¿el cuerpo de evidencia de este campo es suficientemente coherente, en esta ventana temporal, para justificar acción? La primera es una pregunta filosófica sobre la naturaleza de la ciencia. La segunda es una pregunta metrológica sobre el estado actual de un campo. Son preguntas distintas y requieren instrumentos distintos. Los cuatro grandes autores reseñados permiten comprender distintas dimensiones del problema; LUM intenta operar en el plano donde esas dimensiones deben convertirse en decisión: el plano de los campos, las evidencias acumuladas y las consecuencias institucionales.

La diferencia decisiva está en la unidad de análisis. Una teoría puede ser elegante, falsable y discutida por especialistas; un campo, en cambio, incluye laboratorios, protocolos, bases de datos, incentivos editoriales, crisis de replicación, aparatos de medición, disputas internas y prácticas institucionales. La acción pública rara vez se apoya en una teoría aislada. Se apoya en campos. Por eso la pregunta metrológica no rebaja la pregunta filosófica: la desplaza hacia el lugar donde las consecuencias ocurren. Una vacuna, una terapia, una intervención educativa o una política sanitaria no salen al mundo como silogismos; salen como decisiones montadas sobre cuerpos de evidencia. LUM intenta evaluar precisamente ese montaje.

II. Del cierre categorial al cierre auditable

El paso de Bueno a LUM es el paso de la condición filosófica del cierre a su instrumentación metrológica. Merece detenerse en él, porque no es trivial.

El Materialismo Filosófico de Bueno distingue tres momentos en la constitución del conocimiento. M1 es la dimensión físico-corporal: aparatos, señales, datos, registros materiales, instrumentos de medición. M2 es la dimensión práctico-social: operaciones, protocolos, instituciones, sanciones, tradiciones de investigación, criterios de validez compartidos por la comunidad. M3 es la dimensión simbólico-formal: modelos, ecuaciones, lenguajes formales, índices, estructuras teóricas. El conocimiento científico no reside en ninguno de estos momentos por separado, sino en su trabazón articulada.

No hay cierre con solo M3: ecuaciones sin datos ni protocolos de verificación son especulación formal, por elegante que sea. No hay cierre con solo M1: datos

brutos sin modelo ni protocolo de análisis son ruido empírico, por abundantes que sean. No hay cierre con solo M2: el consenso institucional o el comité de expertos sin datos ni modelo es autoridad desnuda, por respetable que parezca. El cierre real emerge cuando los tres momentos están trabados: cuando hay datos que se articulan con protocolos que se articulan con modelos, y cuando las identidades sintéticas resultantes son independientes del metodólogo que las produce.

Esta arquitectura triádica del conocimiento es la herencia filosófica que LUM recoge y operacionaliza. Cuando LUM mide la integridad interna de la evidencia —su índice de defectología— mide el estado de M1: qué tan libre de errores y retractaciones es el registro material. Cuando mide la centralidad de las prácticas, mide el estado de M2: qué tan convergentes son los protocolos y las operaciones del campo. Cuando mide la estructura topológica de la evidencia y la densidad de contradicción, mide la coherencia entre M1 y M3: si los datos forman una arquitectura conectada capaz de sostener los modelos.

La diferencia entre Bueno y LUM no es de contenido filosófico sino de capa operativa. Bueno pregunta: ¿cuándo puede hablarse de cierre en sentido estricto? LUM pregunta: ¿cómo se computa esa pregunta para un campo dado, en una ventana temporal dada, con el cuerpo de evidencia disponible? La primera es una pregunta gnoseológica. La segunda es una pregunta metrológica. LUM necesita la primera para saber qué está midiendo; y la primera, para tener consecuencias institucionales, necesita la segunda.

Hay además un elemento que LUM añade y que no estaba articulado en Bueno: la temporalidad explícita. El cierre categorial tiende a pensarse como una propiedad que los campos adquieren cuando maduran, con cierta estabilidad posterior. LUM lo trata como un evento probabilístico dentro de una ventana temporal Δ. Esto no es un debilitamiento del concepto sino una precisión epistemológica de primer orden: el cierre no es permanente. Los dictámenes envejecen. Un campo que alcanzó cierre operativo en un momento dado puede perderlo si la acumulación de replicaciones fallidas erosiona sus identidades sintéticas. La función de envejecimiento del dictamen es la formalización de algo que la filosofía sabe desde Kuhn pero rara vez traduce a procedimiento: el conocimiento es históricamente situado, y esa situación importa para la acción institucional.

El paso del cierre categorial al cierre auditable es, en suma, el paso de una filosofía de la ciencia descriptiva a una filosofía de la ciencia operativa. LUM hereda el rigor conceptual de Bueno y lo traduce a índices computables, umbrales de decisión y contratos verificables. Esta traducción no empobrece el concepto: lo hace accionable.

III. Qué significa medir la razón

La expresión «metrología de la razón» puede sonar a reducción tecnocrática o a metáfora soberbia. Su sentido exige una defensa filosófica cuidadosa.

La metrología física —la ciencia de la medición— no inventa las propiedades que mide. Un termómetro no inventa la temperatura; registra el estado térmico de un sistema. Una regla no inventa la longitud; la cuantifica con relación a un patrón. Una báscula no inventa el peso; calibra la respuesta de un instrumento contra un estándar. Lo que la metrología hace es producir instrumentos que tienen trazabilidad: su relación con la propiedad que miden puede reconstruirse y verificarse por cualquier evaluador que disponga del mismo instrumento y el mismo patrón.

LUM extiende esta lógica al terreno del conocimiento con una precisión filosóficamente decisiva: mide condiciones de cierre. La verdad de las proposiciones científicas no es directamente cuantificable sin petición de principio. En cambio, sí puede evaluarse si la evidencia de un campo es suficientemente coherente, íntegra, convergente y estructurada para sostener identidades sintéticas auditables. La diferencia entre medir verdad y medir condiciones de cierre equivale a la diferencia entre hablar de «lo húmedo» en abstracto y medir la tensión de vapor de un sistema: una operación queda en vaguedad conceptual; la otra produce medición controlada.

Esta distinción resuelve la objeción aparentemente circular de que la razón no puede medirse porque es ella la que mide. La respuesta es que LUM no mide la razón como facultad interior o como estructura trascendental. Mide condiciones públicas del conocimiento: las condiciones bajo las cuales un campo produce resultados reconstruibles, coherentes y sostenibles. Esas condiciones son externas al sujeto cognoscente individual; son propiedades del cuerpo de evidencia, de las prácticas de investigación, de la red de validaciones y de las relaciones entre

estudios. Son, en el sentido preciso de Bueno, identidades sintéticas: resultados que emergen de la operación misma, independientemente de quién la ejecuta.

Medir la razón, en este sentido, equivale a medir la madurez operatoria de un campo: su capacidad, en un momento histórico dado, de producir dictámenes que resistan auditoría independiente.

Esta operacionalización tiene tres implicaciones filosóficas que vale la pena articular con cuidado. La primera implicación desplaza el problema del sujeto al campo. La pregunta decisiva cambia: ¿ha alcanzado este campo cierre suficiente? El sujeto individual deja de ocupar el centro del dictamen, bajo la condición que Bueno llamó neutralización del sujeto operatorio. El resultado depende del método antes que del metodólogo.

La segunda implicación desplaza la objetividad desde la representación hacia la reconstrucción. Un dictamen alcanza objetividad operativa cuando puede ser reconstruido por cualquier evaluador independiente dado el mismo cuerpo de evidencia y la misma configuración metodológica. La objetividad se vuelve una propiedad de trazabilidad del dictamen.

La tercera implicación convierte el tiempo en dimensión epistémica ineliminable. El cierre funciona como evento probabilístico dentro de una ventana temporal. Un campo con cierre hoy puede perderlo mañana si nuevas evidencias erosionan sus identidades sintéticas. La revisabilidad del dictamen expresa la virtud epistémica central del sistema: reconocer la naturaleza provisional del conocimiento sin convertirla en relativismo.

IV. Los cuatro ejes de la razón medida

LUM mide el cierre mediante cuatro índices ortogonales: IPU, CPV, A_norm y κ_conf. Son ortogonales en el sentido matemático: cada uno captura una dimensión del cierre que los demás no pueden capturar. Esta ortogonalidad es la que permite combinarlos en un modelo estadístico sin que se cancelen ni dupliquen, y sin que la fortaleza en uno pueda compensar artificialmente la debilidad en otro.

El primer eje, IPU, es el índice de defectología o integridad interna. Mide la frecuencia y gravedad de los defectos metodológicos dentro del cuerpo de

evidencia del campo: retractaciones, correcciones mayores, errores de análisis documentados. Un campo con IPU alto tiene poca actividad defectológica severa; un campo con IPU bajo tiene defectos frecuentes o graves que erosionan la confianza en sus resultados acumulados. La lógica filosófica de IPU remite directamente al momento M1 de Bueno: el cierre requiere que el registro material de la evidencia sea íntegro. Sin integridad en el registro, las identidades sintéticas que el campo produce son frágiles, dependientes de datos que pueden ser retirados o corregidos en cualquier momento. Un campo con IPU alto puede aun así producir conocimiento erróneo; puede tener IPU bajo y estar en proceso de corrección productiva. IPU es condición necesaria, no suficiente, del cierre.

El segundo eje, CPV, es el índice de centralidad o coherencia práctica. Mide qué tan agrupadas están las prácticas, modelos y datos del campo en torno a un núcleo central. Un campo con CPV alto tiene prácticas convergentes; uno con CPV bajo tiene prácticas dispersas o contradictorias entre sí. La lógica filosófica de CPV remite al momento M2: el cierre requiere que las operaciones del campo sean estables y conectadas. Si los investigadores del mismo campo usan métodos tan distintos que sus resultados no pueden compararse, no hay cierre posible aunque cada estudio individual sea impecable. CPV captura la cohesión práctico-operatoria del campo. Un riesgo específico que el sistema detecta es el monoparadigmatismo artificial: un campo forzado por presión institucional a converger superficialmente puede tener CPV alto con diversidad real baja, lo que apunta hacia sobrecierre en lugar de cierre genuino.

El tercer eje, A_norm, es la actividad topológica normalizada o estructura de la evidencia. Mide la arquitectura de la red de evidencias mediante homología persistente: qué tan conectada, coherente y estable es esa red en su conjunto. Captura algo que los dos primeros índices no pueden: es posible tener un campo con pocos errores (IPU alto) y prácticas convergentes (CPV alto) y aun así tener evidencia fragmentada, compuesta de islas que no se comunican entre sí, incapaz de producir predicciones fuera de muestra. Si la evidencia de un campo forma una nube de puntos bien conectada, con estructura estable bajo perturbaciones, A_norm es alto. Si forma grupos aislados —archipiélagos de evidencia que no se comunican— A_norm es bajo, y el campo tiene datos pero no tiene arquitectura epistémica.

El cuarto eje, κ_conf, es la densidad de contradicción o consistencia interna. Mide las incompatibilidades entre resultados y modelos dentro del campo, bajo la definición operacional de incompatibilidad que el evaluador debe preregistrar antes de la evaluación. Esta exigencia de preregistro es filosóficamente relevante: obliga a definir ex ante qué cuenta como contradicción en ese dominio, bloqueando la posibilidad de llamar «diversidad productiva» o «perspectivas complementarias» a lo que en realidad es inconsistencia real. El preregistro bloquea esta trampa epistemológica.

La pluralidad de cuatro ejes ortogonales impide que la razón se reduzca a un solo criterio. Esta es una lección filosófica de primer orden: la razón operatoria no tiene una sola dimensión. Un campo puede ser íntegro y fragmentado; puede ser convergente y contradictorio; puede tener buena estructura topológica y alta densidad de defectos. LUM captura estas combinaciones en lugar de colapsarlas en una puntuación única que oculte sus tensiones internas.

La consecuencia filosófica de estos cuatro ejes es más importante que su forma técnica. IPU recuerda que sin higiene material no hay verdad operatoria: los datos defectuosos no se vuelven ciencia por estar bien narrados. CPV muestra que un campo requiere núcleo práctico: una comunidad que no comparte operaciones mínimas produce literatura abundante, pero difícilmente cierre. A_norm añade una exigencia más severa: las evidencias deben formar arquitectura. Un montón de estudios todavía no constituye una estructura científica. κ_conf, por último, impide una coartada frecuente: llamar pluralismo a lo que quizá sea contradicción no resuelta.

Los cuatro índices se integran mediante un modelo de regresión con enlace complementario log-log, diseñado para modelar eventos raros en ventanas de tiempo. El resultado estima, bajo condiciones explícitas, la probabilidad calibrada de que el campo produzca al menos una identidad sintética auditada en la siguiente ventana temporal Δ. Esta formulación probabilística y temporal distingue a LUM de los criterios clásicos de demarcación: desplaza la pregunta desde el estatuto abstracto de una teoría hacia el cierre efectivo de un campo aquí y ahora.

V.1. Nota técnica mínima: cómo se vería una medición LUM

La exposición anterior puede dejar una impresión equivocada si se queda solo en conceptos. Si LUM pretende funcionar como metrología, debe poder traducir sus intuiciones a operaciones mínimas. Lo que sigue no es una especificación completa del sistema, sino una muestra de su forma operatoria.

IPU puede expresarse, en su versión básica, como un índice de integridad:

$$IPU = 1 - (\Sigma_i w_i \cdot E_i) / (\Sigma_i w_i)$$

donde E_i representa la severidad del defecto observado —por ejemplo, error menor, corrección mayor, retractación o fallo de replicación— y w_i representa el peso relativo del estudio afectado. Si tres defectos tienen severidades 0.1, 0.2 y 0.0 con el mismo peso, el promedio defectológico es 0.10 y el IPU resultante es 0.90. El campo conserva alta integridad. Si las severidades fueran 0.8, 0.9 y 0.7, el promedio sería 0.80 y el IPU caería a 0.20. En ese caso, el cuerpo de evidencia tiene problemas severos de higiene metodológica.

CPV mide la proximidad de las prácticas al núcleo operativo del campo. En una versión geométrica, puede estimarse con distancia robusta entre cada estudio y el núcleo central de prácticas, modelos o mediciones. La intuición es sencilla: la cercanía de un estudio depende de su semejanza operatoria con las prácticas nucleares del campo. Si los estudios se agrupan de manera compacta alrededor del núcleo, CPV sube. Si se dispersan en procedimientos incompatibles, CPV baja.

A_norm mide la forma global de la evidencia. En su versión topológica, usa homología persistente para observar si la red de estudios forma una estructura conectada o si aparece como archipiélago: grupos aislados, vacíos inexplicables, ciclos cerrados o zonas sin evidencia. En una versión robusta, A_norm puede aproximarse mediante tres señales: estabilidad del modelo al cambiar parámetros, modularidad de los resultados principales y sensibilidad a retirar partes del corpus. Si al retirar unos pocos estudios la estructura se derrumba, el cierre era frágil.

κ_conf mide la densidad de contradicción bajo una definición preregistrada de incompatibilidad:

κ_conf = número de incompatibilidades / número de comparaciones relevantes

Su complemento, Cons = 1 − κ_conf, representa consistencia interna. La clave está en que la incompatibilidad debe definirse antes de medir. En medicina podría ser una diferencia clínicamente relevante de efecto; en psicología, un efecto que desaparece bajo preregistro; en física, una predicción incompatible con observaciones robustas. Sin preregistro, el evaluador puede llamar “pluralismo” a lo que quizá sea contradicción.

Los cuatro índices no se suman de manera ingenua. En LUM se integran mediante un modelo con enlace log-log complementario:

$$p\Delta = 1 - \exp[-\exp(\eta)]$$

$$\eta = \beta_0 + \beta_1 \cdot IPU + \beta_2 \cdot CPV + \beta_3 \cdot A\_norm + \beta_4 \cdot Cons + controles$$

donde pΔ designa la probabilidad estimada de que el campo produzca una identidad sintética auditada en la siguiente ventana temporal Δ, bajo el contrato de evaluación declarado. Cuando esa probabilidad carece de calibración contra ventanas históricas comparables, debe reportarse como score.

Esta nota conserva el ensayo dentro de su registro filosófico y muestra el punto decisivo: la metrología de la razón exige declarar qué mide, con qué entradas, bajo qué normalización, con qué versión y con qué posibilidad de error.

V.2. Un rodeo didáctico: LUM explicado con experimentos de primaria

Conviene hacer una pausa antes de avanzar hacia CLARION. La LUM puede parecer abstracta porque habla de cierre, índices, auditoría y probabilidades calibradas. Pero su intuición básica puede entenderse con ejemplos muy simples. Imaginemos una clase de primaria que quiere responder preguntas curiosas: ¿qué avión de papel vuela más lejos?, ¿qué semilla crece mejor?, ¿qué puente de palitos aguanta más peso?, ¿qué receta sale igual dos veces?

En todos esos casos, la pregunta no es solamente si algo salió bien una vez. Un avión puede volar lejos por casualidad. Una semilla puede crecer más porque recibió más sol sin que nadie lo notara. Un puente puede aguantar mucho peso porque un niño usó más pegamento que los demás. Una receta puede salir rica una tarde y mal al día siguiente. La pregunta LUM sería otra: ¿tenemos suficientes materiales limpios, reglas claras, pruebas variadas y resultados compatibles para decir que realmente entendimos lo que está pasando?

IPU sería la limpieza del experimento. En el caso del avión de papel, IPU baja si algunos aviones están rotos, si nadie midió bien la distancia, si unos lanzaron desde una mesa y otros desde el suelo, o si se olvidó anotar varios resultados. No significa que el experimento sea inútil; significa que el registro todavía tiene demasiados defectos para confiar plenamente en la conclusión.

CPV sería la convergencia de prácticas. Si cada niño dobla el avión de una forma distinta, lo lanza con distinta fuerza y mide con reglas diferentes, el grupo no está probando lo mismo. Puede haber resultados interesantes, pero no hay todavía un núcleo común de operación. Para que CPV suba, la clase necesita acordar un procedimiento mínimo: mismo tipo de papel, mismo lugar de lanzamiento, misma forma de medir y varias repeticiones.

A_norm sería la arquitectura de la evidencia. No basta lanzar un avión una sola vez. Tampoco basta lanzar diez veces el mismo avión si todos los lanzamientos ocurren con viento fuerte. La evidencia mejora cuando hay pruebas variadas pero conectadas: varios modelos de avión, varios lanzamientos, registro de distancia, control del viento, comparación entre diseños. A_norm pregunta si los datos forman una red que de verdad permite entender el fenómeno, no solo una colección de números sueltos.

κ_conf sería la contradicción interna. Supongamos que el grupo afirma: “los aviones con alas grandes vuelan más lejos”. Pero luego aparecen varios aviones con alas

pequeñas que vuelan más lejos que los de alas grandes, y nadie puede explicar por qué. Ahí hay contradicción. No basta decir “cada avión es diferente”. Hay que investigar si cambió el peso, el doblez, la fuerza del lanzamiento o el aire. κ_conf mide cuántas incompatibilidades quedan sin resolver.

CLARION aparecería cuando el grupo puede decir: “bajo estas reglas, con estos materiales y estas pruebas repetidas, este diseño de avión vuela más lejos con suficiente estabilidad”. No significa que sea verdad eterna ni que ese avión gane siempre en cualquier planeta, patio o tormenta. Significa que hay cierre suficiente para actuar: por ejemplo, elegir ese diseño para el concurso de la escuela.

PSNC aparece cuando todavía no hay cierre. Si los resultados cambian mucho, si nadie midió igual, si hubo viento, si los aviones no eran comparables, la clase no debe fingir que ya sabe. Debe producir un plan: repetir pruebas sin viento, usar el mismo papel, registrar distancias, separar diseños, revisar errores. El no-cierre no es fracaso; es saber qué falta para poder saber mejor.

El sobrecierre aparece cuando todo sale demasiado perfecto. Por ejemplo: todos los aviones vuelan exactamente cinco metros, todos los niños entregan la misma tabla y nadie puede explicar cómo midió. Eso no parece ciencia perfecta; parece copia, regla mal usada o resultado fabricado. LUM enseña una sospecha sana: cuando el orden es demasiado bonito, hay que revisar si es real.

La analogía tiene un límite evidente: un salón de primaria no es un campo científico. Pero ilumina la intuición central de LUM. El cierre no aparece porque alguien lo declare. Aparece cuando los materiales, las operaciones, las reglas y los resultados empiezan a sostenerse entre sí. Y cuando no se sostienen, lo racional no es fingir avance, sino descubrir qué falta.

VI. CLARION: cierre suficiente para actuar

CLARION es el nombre que LUM da al estado de cierre operativo: la condición en que un campo ha alcanzado suficiente coherencia interna y externa para justificar acción institucional sobre sus resultados.

CLARION debe entenderse como cierre operativo revisable. Afirma que un campo dispone de condiciones suficientes para orientar acción bajo monitoreo, no que posea verdad absoluta, certeza subjetiva o consenso incuestionable. La sensación del investigador, el acuerdo del gremio y la estabilidad institucional

pueden acompañar el cierre, pero no lo sustituyen. Un CLARION emitido hoy debe revisarse en la siguiente ventana temporal.

CLARION funciona como permiso racional condicionado: autoriza actuar bajo cierre suficiente, con monitoreo de deriva y revisión posterior. Esta condición cambia la actitud institucional ante el dictamen. Quien opera bajo CLARION debe atender nuevas evidencias, evitar presentar el dictamen como permanente y someter las decisiones de alto impacto a monitoreo posterior.

La condición técnica para alcanzar CLARION en LUM es múltiple. Requiere que la probabilidad calibrada de cierre supere un umbral específico ($p_{cal} \geq 0.80$), que los índices mínimos de integridad, estructura y confianza superen simultáneamente sus umbrales respectivos —condición AND_min, que opera en lógica conjuntiva, no disyuntiva—, y que haya calibración externa válida del modelo. La calibración es el puente entre un score estadístico y una probabilidad en sentido frecuencial: si el modelo dice 80%, eso debe significar que en el 80% de los campos con esa puntuación se produce efectivamente cierre en la siguiente ventana. Sin calibración, esa afirmación no puede sostenerse, y el dictamen debe interpretarse como score, no como probabilidad.

Esta exigencia conecta LUM con una tradición técnica más amplia sobre modelos predictivos: una probabilidad no basta con ser numéricamente elegante; debe estar calibrada contra resultados observados. La diferencia entre score y probabilidad calibrada es decisiva porque una decisión institucional no puede apoyarse en una cifra que parece probabilidad pero no ha sido sometida a contraste empírico suficiente (Van Calster et al., 2016).

La diferencia entre score y probabilidad calibrada tiene consecuencias filosóficas directas. Un score de 0.82 indica que el modelo asigna un valor alto a este campo. Una probabilidad calibrada de 0.82 afirma que, en campos con puntuaciones semejantes, el 82% alcanzó cierre en la siguiente ventana. La segunda afirmación puede verificarse empíricamente y corregirse si falla. LUM obliga a declarar cuál de las dos magnitudes se reporta, porque las consecuencias institucionales permitidas dependen de esa diferencia.

En términos filosóficos más amplios, CLARION traduce la noción de identidad sintética en un concepto operativo: cierre suficiente para actuar. La identidad sintética plena puede requerir décadas de trabajo colectivo; CLARION

apunta a la suficiencia necesaria para justificar acción según el estado actual del conocimiento. Esta formulación evita la espera de una perfección epistémica imposible y permite actuar responsablemente bajo incertidumbre controlada.

VII. No-cierre: la ignorancia como diagnóstico

Esta sección aborda la contribución filosófica quizás más contraintuitiva de LUM. En la epistemología estándar, el no-saber es simplemente la ausencia de saber: un estado privativo, un vacío que se llena con más investigación. Cuando un campo no ha alcanzado cierre, la respuesta habitual es «se necesita más investigación». Esta frase tiene la profundidad diagnóstica de decirle a alguien que tiene hambre que necesita comer. Es verdadera y casi completamente inútil.

LUM transforma el no-cierre en un diagnóstico estructurado. El Plan de Salida de No-Cierre —PSNC— no es un consuelo ni una promesa vaga de progreso futuro. Es un análisis causal de por qué el campo no ha alcanzado cierre y qué tipo de acciones específicas podrían remediarlo. La lógica es que el no-cierre tiene múltiples causas posibles con remedios distintos, y confundirlas produce planes de investigación equivocados y costosos.

El sistema distingue varios tipos de deficiencia diagnóstica. Si el campo tiene alta sombra epistémica —alta fracción de evidencia relevante a la que el evaluador no tiene acceso— la acción correcta es ampliar la instrumentación y el acceso a datos. Más estudios con los mismos instrumentos y el mismo sesgo de cobertura no reducen la sombra. Si hay entropía espectral alta con señal-ruido baja, la señal está oculta en ruido de medición: la acción es mejorar los instrumentos de medición, no multiplicar las mediciones defectuosas. Si hay interferencia alta entre mecanismos causales, el modelo mezcla mecanismos distintos que producen el mismo efecto por vías diferentes: la acción es rediseñar el modelo causal, no acumular más datos con el mismo diseño. Si hay alta densidad de contradicción, hay incompatibilidades reales no resueltas: la acción es resolverlas mediante replicaciones directas y preregistro de definiciones, no ignorarlas nombrándolas «diversidad de perspectivas».

La riqueza filosófica de esta taxonomía consiste en convertir la ignorancia en conocimiento de segundo orden. El no-cierre identifica estructuradamente qué falta para poder saber. La distinción entre ignorancia bruta e ignorancia

diagnosticada tiene valor epistémico positivo: identificar el tipo de deficiencia que impide el cierre ya contribuye al conocimiento del campo. Es progreso de segundo orden.

Existe también una modalidad especial de no-cierre: el PSNC-D, que se activa cuando el objeto de evaluación no supera el filtro de demarcación previo. Preguntas como «¿qué es la conciencia?» o «¿es la inteligencia artificial peligrosa?» no son preguntas sin importancia; son filosóficamente ricas. Pero no tienen el recorte operatorio necesario para aplicarles el sistema: carecen de operadores definidos, de criterios de contradicción especificados, de definición del evento de cierre. En estos casos, el PSNC-D no descalifica la pregunta; exige convertirla primero en un problema con recorte operatorio. El diagnóstico es demarcatorio.

Hay además una dimensión ética en el no-cierre que LUM articula con precisión. Cuando el diagnóstico es ROJO, la abstención se vuelve respuesta racional. Actuar como si hubiera cierre en un campo diagnosticado en no-cierre constituye irresponsabilidad epistémica. Quien diseña política pública sobre un campo en ROJO —con alta densidad de contradicción, con efectos que no se replican, con evidencia fragmentada— no está siendo cauteloso: está ignorando el diagnóstico disponible. El PSNC convierte esa abstención en trabajo: identifica qué tipo de investigación, con qué diseño, atendiendo a qué deficiencia específica, puede mover el campo del no-cierre al cierre. La abstención con plan es muy distinta de la parálisis sin horizonte.

En este punto, LUM converge con una preocupación más amplia de la ciencia contemporánea: la necesidad de diseños que reduzcan flexibilidad analítica, preregistro de hipótesis y prácticas de investigación reproducible. Los reportes registrados y la ciencia abierta no resuelven por sí solos el cierre, pero reducen algunas de las condiciones que producen no-cierre persistente: sesgo de publicación, flexibilidad post-hoc y dificultad de reconstrucción del procedimiento (Chambers, 2013; Nosek & Lakens, 2014; Peng, 2011; Stodden et al., 2016).

VIII. Un caso mínimo: el priming social como campo en no-cierre

La utilidad filosófica de LUM se entiende mejor si se baja un momento al suelo. Pensemos en el caso del priming social. Durante años, una parte de la psicología social sostuvo que estímulos sutiles podían modificar conductas posteriores de manera significativa: palabras asociadas a la vejez que alteraban la velocidad al caminar, señales ambientales que inclinaban decisiones morales, pequeñas manipulaciones contextuales que parecían producir efectos sorprendentes. El campo era atractivo porque ofrecía algo que toda institución desea: intervenciones baratas, elegantes y aparentemente poderosas sobre la conducta humana.

La crisis de replicación modificó esa confianza. Muchos de esos efectos resultaron más frágiles de lo que sugería su recepción inicial, especialmente cuando se sometían a preregistro, muestras mayores o protocolos más estrictos (Open Science Collaboration, 2015; Camerer et al., 2018).

Visto desde una filosofía clásica de la ciencia, el diagnóstico era difícil. Las hipótesis eran falsables y, por tanto, superaban el criterio popperiano mínimo. Había comunidad de investigación, revistas, laboratorios y experimentos; existía un paradigma operativo en sentido amplio. También podía describirse como programa de investigación, con predicciones, anomalías y respuestas defensivas. Sin embargo, esas descripciones dejaban abierta la pregunta institucionalmente decisiva: ¿ese cuerpo de evidencia estaba lo suficientemente cerrado para justificar intervención psicológica, educativa o política?

LUM formularía la pregunta desde otro plano. Evaluaría las condiciones de cierre del priming social como campo. ¿Qué tan alto es su IPU, si aparecen fallos de replicación, sesgos de publicación o prácticas cuestionables? ¿Qué tan estable es su CPV, si los diseños experimentales varían tanto que los resultados no se comparan limpiamente? ¿Qué forma tiene su A_norm, si la evidencia aparece como archipiélago de efectos llamativos pero poco conectados? ¿Qué tan elevada es su κ_conf, si estudios equivalentes producen resultados incompatibles o efectos que desaparecen bajo preregistro y muestras mayores?

El punto delicado está en que muchas fragilidades no proceden de fraude abierto, sino de flexibilidad analítica normalizada: decidir qué variables reportar, cuándo detener la recolección, qué especificación privilegiar o qué comparación dejar fuera después de haber visto los resultados. Ese tipo de prácticas puede inflar

falsos positivos y producir una apariencia de estabilidad que no resiste auditoría posterior (Simmons et al., 2011).

Para que el caso no quede como simple ilustración cualitativa, puede imaginarse una aplicación hipotética de LUM al priming social. No sería un dictamen real —eso requeriría corpus cerrado, criterios preregistrados y cálculo reproducible—, pero sí muestra cómo operaría la matriz.

Supongamos una ventana de evaluación donde el corpus contiene estudios clásicos, replicaciones directas, metaanálisis y reportes preregistrados. El evaluador preregistra como incompatibilidad relevante la desaparición del efecto bajo muestras mayores, preregistro o reducción de flexibilidad analítica. Bajo ese contrato, el campo podría arrojar una matriz de este tipo:

IPU: medio-bajo, por fallos de replicación, sesgo de publicación y flexibilidad analítica documentada.

CPV: medio, porque existen protocolos experimentales reconocibles, pero con alta variación de estímulos, contextos y medidas de resultado.

A_norm: bajo o medio-bajo, si la evidencia aparece como archipiélago de efectos locales poco conectados entre sí.

κ_conf: alto, si estudios conceptualmente equivalentes producen resultados incompatibles o efectos que desaparecen bajo condiciones más estrictas.

Conf: insuficiente para VERDE, si la incertidumbre sobre la calidad y cobertura del corpus sigue siendo alta.

El dictamen razonable sería NO-CIERRE: la base disponible resulta insuficiente para derivar intervención psicológica o política pública fuerte. Ese diagnóstico no equivale a NEGRO automático ni a rechazo total del campo.

El PSNC pediría replicaciones preregistradas, definición más estricta de familias de efectos, separación por contexto, control de flexibilidad analítica y

comparación entre efectos que sobreviven y efectos que colapsan bajo mejores diseños.

La consecuencia filosófica de ese dictamen no sería una condena teatral, sino una delimitación responsable. Y esa diferencia importa. Decir “el campo está en no-cierre” significa algo preciso: el cuerpo de evidencia disponible todavía no sostiene la acción fuerte que algunos querían derivar de él. Ese diagnóstico evita tres exageraciones: declarar que todo era falso, suponer que todos se equivocaron o concluir que la psicología social carece de valor. El PSNC, en este caso, pediría algo más específico que “más investigación”. Pediría replicaciones preregistradas, reducción de flexibilidad analítica, separación de efectos por contexto, mejor definición de incompatibilidades, control de sesgo editorial y reconstrucción del núcleo de prácticas válidas.

Ahí aparece la ganancia filosófica. LUM sitúa al campo. Retira la máscara de cierre prematuro y ofrece un camino de reconstrucción. La crisis de replicación deja de operar como escándalo moral o defensa corporativa y se convierte en diagnóstico operatorio: qué falló, dónde falló, con qué consecuencias y qué tendría que cambiar para que el campo pudiera volver a reclamar cierre. Esa es la diferencia entre una filosofía que juzga desde fuera y una metrología que obliga al campo a mostrar sus condiciones de estabilidad.

IX. Sobrecierre: cuando el orden se vuelve sospechoso

El caso del priming social muestra un campo que pierde derecho a reclamar cierre fuerte cuando sus replicaciones y prácticas internas no sostienen la promesa inicial. Pero existe una patología distinta, menos visible y quizá más peligrosa. A veces aparece una patología distinta: una forma falsa de cierre, hecha de demasiada coherencia, demasiada uniformidad y demasiada facilidad para confirmar lo que el campo ya esperaba encontrar. Si el no-cierre expresa ausencia de conocimiento suficiente, el sobrecierre expresa apariencia de conocimiento producida por mecanismos defectuosos de validación. LUM llama a este estado NEGRO: conocimiento patológico. Los campos en NEGRO tienen, paradójicamente, demasiado orden: demasiada consistencia, demasiado consenso, demasiada convergencia. Y ese exceso es precisamente la señal de que algo está mal.

Las señales específicas del sobrecierre son cuatro. La primera es el monocriterio colapsado: todos los estudios usan el mismo método y los resultados son perfectamente consistentes porque no pueden ser de otra manera. No hay diversidad metodológica genuina; hay uniformidad impuesta que impide la aparición de resultados discordantes. La segunda es la supresión de comparaciones: se desalientan o imposibilitan los estudios que podrían falsificar el consenso. La disidencia metodológica o empírica no es simplemente ignorada —es activamente bloqueada mediante presiones institucionales, de financiamiento o de publicación. La tercera es el leakage temporal: información del futuro contamina el entrenamiento del modelo, produciendo sobreestimación sistemática del cierre porque el sistema evalúa conocimiento que en realidad aún no estaba disponible en la ventana de análisis. La cuarta es la circularidad de validación: el mismo corpus se usa para entrenar el modelo y para validarlo, produciendo una consistencia circular sin valor predictivo real.

La contribución más contraintuitiva de LUM al pensamiento sobre la ciencia es precisamente el NEGRO. La mayoría de los sistemas de evaluación distinguen simplemente entre «bien» —alto grado de cierre— y «mal» —bajo grado de cierre—. LUM distingue cuatro categorías y coloca el mayor peligro en el sobrecierre patológico. El no-cierre conserva una honestidad mínima: un campo en ROJO tiene evidencia insuficiente, y esa insuficiencia puede decirse y diagnosticarse. El sobrecierre ha producido una imagen de conocimiento que protege activamente su ilusión contra la falsificación. Es más difícil de detectar y más dañino cuando se convierte en base de acción institucional.

La ley de Goodhart es el mecanismo filosófico de esta patología. Formulada en economía —«cuando una medida se convierte en objetivo, deja de ser una buena medida»— opera en la ciencia de manera análoga: cuando los índices de un campo se vuelven criterios para obtener financiamiento o publicación, los actores aprenden a optimizarlos sin mejorar el cierre real. Una práctica de publicación orientada al número de papers en revistas de alto impacto, no a la producción de conocimiento sólido, puede producir campos con IPU bajo —muchas retractaciones—, CPV artificialmente alto —convergencia forzada—, y estructura topológica circular. LUM no resuelve Goodhart por decreto; lo incorpora como

riesgo estructural declarado del sistema y diseña métricas específicas para detectar cuando alguien está optimizando el sistema en lugar de mejorando el campo.

El problema atraviesa toda institución que convierte una medida en objetivo. Ese gesto produce incentivos para optimizar el indicador y descuidar el fenómeno medido. La literatura sobre auditoría y evaluación institucional ha mostrado que los sistemas de medición pueden modificar la conducta de los agentes hasta volver opaca la relación entre indicador y realidad evaluada (Campbell, 1979; Strathern, 1997). LUM no escapa mágicamente a este riesgo; lo vuelve parte explícita del diagnóstico.

El enemigo de la razón aparece en dos formas: el caos epistémico del no-cierre y el orden falso del sobrecierre. El orden falso resulta más difícil de combatir que el caos porque adopta la apariencia de la virtud epistémica: coherencia, convergencia, consenso. Detectarlo requiere precisamente el tipo de vigilancia que LUM institucionaliza.

X. Auditoría: la objetividad como reconstrucción

La auditoría cumple una función epistémica. Esta frase merece defensa cuidadosa, porque la cultura científica suele percibir el trámite institucional como obstáculo al trabajo real.

En la tradición filosófica, la objetividad se ha entendido frecuentemente como representación fiel de la realidad: el conocimiento es objetivo en la medida en que «corresponde» a los hechos. Esta imagen tiene problemas conocidos —el problema de cómo verificar la correspondencia sin un acceso independiente a la realidad que sirva de término de comparación— pero tiene también una virtud: conecta el conocimiento con algo exterior al sujeto.

LUM propone un desplazamiento: la objetividad reside en la reconstruibilidad del dictamen. Un dictamen es objetivo si puede ser reconstruido por cualquier evaluador independiente dado el mismo cuerpo de evidencia, la misma configuración metodológica y el mismo modelo estadístico. Esta reconstruibilidad es empíricamente verificable. El concepto de objetividad gana así un soporte operativo.

Este desplazamiento conserva la pregunta por la verdad y la obliga a pasar por mediaciones. Una verdad científica que no puede reconstruirse, que no declara sus condiciones de producción y que no permite revisar sus datos, umbrales o procedimientos queda demasiado cerca de la autoridad. Puede ser verdadera, pero todavía carece de responsabilidad pública. Cuando una afirmación pretende guiar acción colectiva, debe aceptar el régimen de la reconstrucción.

El contrato LUM-I/O es la arquitectura formal de esa reconstruibilidad. Define cinco componentes que deben estar presentes en cualquier dictamen válido. INPUT: qué campo se evalúa, en qué ventana temporal, con qué cuerpo de evidencia —definido y hasheable antes de comenzar la evaluación. CONFIG: qué parámetros y umbrales se usaron, congelados antes del análisis para impedir el ajuste retrospectivo de parámetros en función de los resultados. MODEL: qué modelo estadístico produjo los resultados, con qué calibración externa. OUTPUT: qué resultados produjo el sistema, con intervalos de confianza y estado del semáforo con su justificación. AUDIT: la huella criptográfica SHA-256 que garantiza la integridad del dictamen completo.

La regla es simple y sin excepciones: si cualquier componente falta, el dictamen es inválido. La invalidez no juzga la sustancia del resultado; marca su imposibilidad de auditoría. Un dictamen sin auditoría queda en el plano de la declaración. Esta distinción entre conocimiento operativo y declaración ocupa el centro de la filosofía de LUM.

El versionado es otro elemento de la arquitectura de reconstrucción que merece reflexión filosófica. LUM distingue varios niveles de versionado, y cualquier cambio en los parámetros del sistema —umbrales de decisión, definición semántica de los índices, política de ventana temporal— produce una nueva versión que hace los dictámenes históricos no directamente comparables. Esta exigencia funciona como virtud epistémica. Ocultar los cambios de parámetros —presentar el dictamen actual como si fuera continuo con el anterior cuando en realidad los umbrales han cambiado— es una forma de corrupción institucional. El versionado la hace imposible: cualquier cambio de parámetros deja huella, y la comparación entre dictámenes de distintas versiones requiere declarar explícitamente la diferencia. La corrupción no puede ser silenciosa.

La huella criptográfica SHA-256 cumple aquí una función modesta pero importante. Un hash criptográfico convierte el dictamen completo en una cadena única: cualquier modificación posterior —de un umbral, de un dato, de un coeficiente del modelo— produce un hash diferente. Conviene no exagerar su alcance. El hash garantiza integridad documental: el dictamen publicado corresponde al paquete de entrada, configuración, modelo y salida declarado. Su valor filosófico consiste en impedir una forma concreta de irresponsabilidad: cambiar las condiciones del dictamen sin dejar rastro.

Esta arquitectura transforma la objetividad de una propiedad del conocimiento en una propiedad del proceso de producción del conocimiento. Un dictamen se vuelve objetivo cuando procede de un proceso que cualquier evaluador independiente puede reconstruir, verificar y, si es necesario, refutar. La objetividad como reconstrucción tiene pretensiones más modestas que la objetividad como correspondencia. Precisamente por eso puede cumplirse.

XI. Objeciones contra una filosofía de la LUM

Toda posición filosófica que valga algo debe poder responder sus objeciones más fuertes. Las que siguen no son de papel de paja; son genuinamente incómodas y merecen respuesta honesta.

Primera objeción: LUM reduce la razón a números. Si la razón se mide mediante cuatro índices y una ecuación, ¿no se está colapsando la complejidad del pensamiento científico en una puntuación vacía de contenido filosófico?

La respuesta tiene dos niveles. El primero: LUM no mide la razón como facultad, sino condiciones públicas del conocimiento. Los cuatro índices capturan propiedades públicas del cuerpo de evidencia, independientes de las creencias o intuiciones teóricas de los investigadores. El segundo nivel: la pluralidad de cuatro ejes ortogonales es precisamente la defensa contra la reducción. Si hubiera un solo índice, la acusación de reducción sería más convincente. Cuatro ejes que capturan dimensiones distintas del cierre —integridad, convergencia, estructura topológica, consistencia— es ya una forma de resistir el colapso en un criterio único.

Hay además un argumento más amplio que no debe pasarse por alto: cualquier alternativa a la medición formal es igualmente susceptible de ser acusada

de reduccionismo. El juicio experto se reduce a la autoridad del experto. El consenso institucional se reduce al acuerdo del gremio, potencialmente capturado. La intuición del tomador de decisiones se reduce a sus sesgos personales. LUM explicita sus criterios y permite criticarlos, mejorarlos y refutarlos. Los criterios implícitos resultan más peligrosos porque escapan al escrutinio.

Segunda objeción: LUM puede volverse tecnocracia. Si los dictámenes de LUM se convierten en el criterio de financiamiento, aprobación regulatoria o publicación, el sistema puede derivar en una burocracia de la medición que sustituye el juicio científico por la puntuación algorítmica.

Esta objeción es legítima y LUM la incorpora de manera estructural, no como nota al pie. El estado NEGRO existe precisamente para detectar cuando el sistema está siendo optimizado en lugar de los campos. El estado INVALID existe para detectar cuando el contrato está siendo violado. Los límites explícitos del sistema —qué LUM no puede hacer, cuándo falla por diseño— son parte de la arquitectura. Más importante: LUM requiere separación institucional entre quienes definen parámetros, quienes ejecutan evaluaciones y quienes auditan footprints. Sin esa separación, el sistema mismo puede ser capturado. La tecnocracia perfecta requeriría que los criterios del sistema fueran siempre correctos; LUM declara que los suyos no lo son, que tienen condiciones de frontera y que pueden corromperse, y diseña mecanismos para detectarlo.

Tercera objeción: toda métrica puede ser manipulada. La ley de Goodhart es el enunciado más preciso de esta objeción: cuando los índices de LUM se vuelven criterios de acceso a recursos, los actores aprenderán a optimizarlos sin mejorar el cierre real del campo.

LUM convierte el problema de Goodhart en condición explícita del sistema. El gaming aparece como riesgo declarado, con métricas específicas — leakage_score, missing_comparisons_ratio, entropy_protocols, validation_dependency— diseñadas para detectarlo. Ningún sistema elimina por decreto la manipulación de métricas; LUM la vuelve visible como riesgo operativo. Si el gaming es suficientemente sofisticado y no deja señales en esas métricas, LUM no lo detecta; y lo dice honestamente. Este límite honesto es filosóficamente más respetable que la pretensión de haberlo resuelto. A nivel de segundo orden: el versionado y la política anti-cambio post-hoc reducen el gaming oportunista, que es

el más frecuente. El gaming sistémico —cuando todo un ecosistema coopera para optimizar las métricas— requiere separación institucional, no solo instrumentos estadísticos.

Cuarta objeción: LUM tiene límites de aplicación. Hay campos, preguntas y fenómenos que quedan fuera de su alcance.

Esta limitación forma parte de su honestidad filosófica. LUM requiere recorte operatorio definido, condiciones de evidencia y posibilidad de producir identidades sintéticas estables. Preguntas totalizantes, fenómenos intrínsecamente no estacionarios o problemas sin criterios de contradicción quedan fuera de su régimen de medición. Tampoco decide verdad metafísica ni garantiza que sus cuatro índices capturen todo lo relevante del cierre. Estas limitaciones pertenecen al diseño del sistema. Un criterio formal responsable debe declarar sus condiciones de frontera antes de aplicarse.

Quinta objeción: LUM presupone juicio filosófico previo. Esta es quizá la objeción más fuerte, porque toca el punto anterior a toda medición. Para aplicar LUM hay que decidir antes qué cuenta como campo, qué evidencia entra, qué queda fuera, qué se entiende por incompatibilidad, qué ventana temporal se usará y qué tipo de acción se pretende justificar. Ninguna de esas decisiones nace automáticamente de la ecuación. Todas requieren criterio filosófico, conocimiento del dominio y responsabilidad institucional. Entonces, podría decirse, LUM no reemplaza la filosofía: depende de ella.

La objeción acierta y fortalece la tesis. LUM vuelve explícito, versionado y auditable el juicio previo. La decisión sobre el campo, la evidencia, la incompatibilidad y la ventana temporal sigue siendo filosófica y gnoseológica, pero deja de operar como presupuesto invisible. En un sistema sin contrato, esas decisiones pueden esconderse bajo el prestigio del experto o bajo la costumbre disciplinar. En LUM deben declararse antes de medir, quedar registradas y aceptar que otro evaluador pueda discutirlas. La metrología obliga a la filosofía a dejar huella.

Este punto impide una lectura tecnocrática de LUM. El sistema no dice: “la máquina decide por nosotros”. Dice algo más exigente: “si vas a decidir, declara tus condiciones de decisión”. La filosofía queda situada en el lugar donde siempre

actuó: el recorte del objeto, la definición de los términos y la elección de los criterios de contradicción. LUM conserva el juicio y le quita la coartada de la invisibilidad.

XII. Hacia una ética del cierre responsable

La filosofía de LUM tiene una dimensión ética que no puede separarse de su dimensión epistémica. El sistema funciona como instrumento de medición y como prescripción de conducta para quienes toman decisiones basadas en el conocimiento.

El principio central puede enunciarse así: no actuar como si hubiera cierre cuando el campo está en no-cierre. Este principio suena elemental. Sus implicaciones son profundas.

Cuando un tomador de decisiones diseña política pública sobre evidencia en ROJO —alta densidad de contradicción, efectos que no se replican, señales activas de no-cierre o sobrecierre— toma una decisión epistémica con consecuencias materiales sobre un diagnóstico que ignora. Y esa decisión puede estar equivocada de una manera específica: no porque la evidencia sea mala en principio, sino porque no ha alcanzado la madurez suficiente para sostener la acción que se le está exigiendo.

LUM articula esta responsabilidad en términos operativos. La abstención operatoria —el ROJO que prohíbe intervenciones de política basadas en el campo— funciona como respuesta racional cuando el diagnóstico indica que la acción prematura puede ser más dañina que la espera. La diferencia entre cautela y parálisis es que la cautela tiene un plan: el PSNC especifica qué investigación haría posible mover el campo del no-cierre al cierre. La abstención con plan de trabajo es completamente distinta de la inacción sin horizonte.

Hay también una dimensión de responsabilidad en el CLARION que merece subrayarse. Un dictamen VERDE exige seguir pensando y atender nueva evidencia. Requiere monitoreo de deriva activo: si nuevas evidencias empiezan a erosionar el cierre, el sistema debe detectarlo y actualizar el dictamen antes de que las consecuencias de actuar sobre un dictamen obsoleto se acumulen. La acción bajo CLARION es acción bajo riesgo controlado y monitoreo continuo, no bajo certeza permanente.

Esta ética del cierre responsable tiene una relación directa con la honestidad intelectual. Declarar cierre donde no lo hay —para obtener financiamiento, para

justificar una decisión ya tomada, para mantener la autoridad del campo o para satisfacer la demanda institucional de certezas— es una forma de fraude epistémico. LUM lo hace detectable mediante el estado NEGRO y sus métricas de sobrecierre. Pero la detección técnica no basta: se necesita también la disposición institucional a actuar sobre el diagnóstico cuando es adverso. Un sistema que detecta el fraude pero cuyos dictámenes adversos son ignorados no protege la integridad del conocimiento; simplemente documenta su degradación.

La prudencia —la virtud aristotélica de la acción adecuada a las circunstancias— equivale en LUM a usar el criterio disponible: operar bajo CLARION cuando el cierre está justificado, abstenerse bajo ROJO cuando falta cierre, investigar en la dirección que el PSNC indica, y detectar el NEGRO cuando el orden es falso. La prudencia epistémica, así entendida, es una práctica institucional, no solo una virtud individual.

XIII. Proyección: el universo como historia de cierres locales

La filosofía de LUM puede proyectarse más allá de la evaluación de campos científicos, siempre que evite convertirse en metafísica total. Conviene formular esta proyección con cuidado. LUM deja intacto el trabajo propio de la física, la química, la biología o la historia. Su función consiste en leer cada dominio según sus modos de cierre: las condiciones bajo las cuales aparecen identidades estables, reproducibles y auditables.

Vista así, la realidad aparece como una historia de cierres locales, más que como sustancia única o línea teleológica asegurada. En física, el cierre se manifiesta cuando la materia deja de ser pura dispersión y produce entidades estables, invariantes, estados ligados y leyes reproducibles. Núcleos, átomos, campos y regularidades experimentales son formas de estabilidad operatoria. En química, el cierre aparece como combinación repetible: enlaces, estequiometrías, rutas de reacción, familias moleculares. La materia ya no solo choca; empieza a combinarse de maneras que pueden repetirse y controlarse.

En biología, el cierre adquiere frontera, metabolismo y memoria. Una célula no es simplemente una bolsa de moléculas: es una organización que distingue interior y exterior, conserva circuitos metabólicos y transmite patrones. La membrana reduce ruido; el metabolismo sostiene continuidad; la replicación vuelve

heredable una forma. En evolución, ese cierre se desplaza del organismo al linaje: el patrón ya no solo se conserva en una célula, sino en una población que transmite variación bajo reglas de herencia y selección.

En historia y ciencias sociales, el cierre cambia de naturaleza. No hay leyes estrictas al modo de la física, pero sí formas institucionales capaces de sobrevivir a individuos: Estados, monedas, sistemas jurídicos, iglesias, burocracias, lenguas estandarizadas, aparatos escolares. Un cierre institucional no es eterno ni perfecto; es una coordinación relativamente estable que permite reproducir operaciones sociales más allá de quienes las ejecutan en un momento dado. Precisamente por eso puede degenerar: una institución puede cerrar, no cerrar o sobrecerrarse en dogma, propaganda o circularidad.

Esta proyección permite una fórmula breve: el universo, visto desde LUM, es una historia de formas que logran cerrar. La frase solo conserva validez si evita el exceso metafísico. LUM no explica por sí sola el mecanismo causal de cada transición ni reemplaza las ciencias particulares. Ofrece una lectura metacientífica de sus condiciones de inteligibilidad: cuándo un nivel alcanza suficiente coherencia para ser tratado como entidad, sistema, campo o institución.

La ganancia filosófica está en evitar dos errores simétricos. El primero sería reducir todo a caos: nada cierra, todo fluye, todo es interpretación. El segundo sería imponer un cierre total: todo obedece a una única ley, una única sustancia o un único destino. LUM permite pensar una tercera vía: cierres parciales, escalonados, revisables y locales. La razón no contempla un universo ya terminado; reconstruye los momentos en que ciertas formas consiguen sostenerse.

XIV. Conclusión: de la razón declamativa a la razón trazable

Una imagen sintetiza el desplazamiento que propone LUM. Aristóteles advertía que es propio del hombre cultivado buscar la exactitud en cada materia en la medida en que la naturaleza del asunto lo permite. LUM busca exactitud operativa: trazabilidad, reconstrucción, auditoría y revisión.

La razón declamativa declara cierre. Proclama conocimiento, certeza, consenso, sin necesidad de especificar el procedimiento que los produjo ni las condiciones bajo las cuales podrían ser refutados. La razón trazable produce cierre bajo condiciones declaradas, lo registra con su huella criptográfica, lo versiona cuando esas condiciones cambian y acepta sin evasiones los casos donde falta cierre. La diferencia no es de grado sino de naturaleza: la razón trazable es responsable ante el dictamen que emite.

LUM ocupa un lugar preciso dentro del ecosistema del saber. La investigación científica seguirá siendo disputada, revisada, contradictoria en sus bordes y viva en sus controversias. La filosofía seguirá preguntando qué es el conocimiento, cuáles son sus condiciones de posibilidad y cómo se relaciona con la realidad. LUM proporciona un instrumento para una tarea específica y difícil: saber cuándo un campo ha alcanzado la madurez suficiente para justificar acción, y qué hacer con precisión cuando esa madurez falta.

Este instrumento da un lenguaje preciso a tres categorías que la epistemología había descrito sin operacionalizar. Cierre: hay identidades sintéticas auditables, la acción es justificable bajo monitoreo de deriva y revisión periódica. No-cierre: el diagnóstico especifica qué falta, la abstención es la respuesta racional, el Plan de Salida de No-Cierre indica el trabajo pendiente con precisión causal. Sobrecierre: el orden es falso, la coherencia es artefacto del proceso, la auditoría externa es obligatoria y el sistema bloquea cualquier dictamen de cierre.

En una época saturada de datos, discursos y métricas, la tarea filosófica consiste en cerrar mejor. Cerrar mejor significa saber cuándo actuar, cuándo esperar, cuándo corregir, cuándo desconfiar de una coherencia demasiado perfecta y cuándo reconocer que el campo todavía no merece la confianza que se le exige. La razón trazable habla con consecuencias. Puede equivocarse, pero deja a la vista las condiciones de su error. Por eso puede corregirse. Y una razón capaz de corregirse vale más, filosóficamente, que una razón acostumbrada a proclamarse.

Anexo didáctico.

Pequeños experimentos para entender LUM

Este anexo traduce algunos conceptos de LUM a ejemplos de primaria. No sustituye la explicación filosófica del ensayo; funciona como una escalera pedagógica. La idea es mostrar que IPU, CPV, A_norm, κ_conf, CLARION, PSNC, sobrecierre y auditoría no son adornos técnicos, sino maneras de preguntar si un conjunto de pruebas ya sostiene una conclusión.

Ejercicio 1. Aviones de papel: ¿cuál vuela más lejos?

Situación:
Una clase quiere descubrir qué diseño de avión de papel vuela más lejos. Hacen tres modelos: avión puntiagudo, avión de alas anchas y avión pequeño. Cada equipo lanza su avión cinco veces.

Pregunta LUM:
¿Ya podemos decir cuál diseño es mejor?

Qué revisar:
IPU: ¿se midieron bien las distancias?, ¿todos anotaron los resultados?, ¿hubo aviones rotos?, ¿alguien lanzó desde más adelante?
CPV: ¿todos usaron el mismo tipo de papel?, ¿lanzaron desde el mismo punto?, ¿midieron con la misma regla?
A_norm: ¿hubo suficientes lanzamientos?, ¿se probaron varios diseños?, ¿se repitió el experimento otro día?
κ_conf: si el avión de alas anchas gana unas veces y pierde otras, ¿sabemos por qué?

Respuesta orientativa:
Si cada equipo usó papel distinto, lanzó con diferente fuerza y midió como pudo, todavía no hay cierre. Hay datos, pero no prueba estable. El PSNC sería repetir el experimento con reglas comunes: mismo papel, mismo lugar, misma forma de lanzar, varias repeticiones y registro claro.

Traducción LUM:

No basta con que un avión gane una vez. Para tener cierre, el resultado debe sostenerse bajo reglas claras y repetibles.

Ejercicio 2. Semillas de frijol: ¿qué planta crece mejor?

Situación:
Tres grupos siembran frijoles. Un grupo les da mucha agua, otro poca agua y otro agua normal. Después de una semana, comparan cuánto crecieron.

Pregunta LUM:
¿Podemos concluir cuánta agua necesita mejor el frijol?

Qué revisar:
IPU: ¿se midieron bien las plantas?, ¿se anotaron los días?, ¿se confundieron macetas?
CPV: ¿todos usaron la misma tierra?, ¿las macetas recibieron la misma luz?, ¿se sembraron a la misma profundidad?
A_norm: ¿se usó solo una semilla por grupo o varias?, ¿se observó durante varios días?, ¿se compararon alturas y hojas?
κ_conf: si una planta con poca agua creció más que una con agua normal, ¿fue por el agua o por otra cosa?

Respuesta orientativa:
Si solo se sembró una semilla por grupo, el resultado es débil. Tal vez esa semilla era más fuerte, recibió más sol o estaba en mejor tierra. Para acercarse al cierre, hay que repetir con varias semillas y controlar luz, tierra y cantidad de agua.

Traducción LUM:
A_norm ayuda a ver si la evidencia tiene estructura suficiente. Una sola planta no basta para entender el comportamiento del conjunto.

Ejercicio 3. Puente de palitos: ¿qué diseño aguanta más peso?

Situación:

Los niños construyen puentes con palitos de madera y pegamento. Quieren saber qué forma aguanta más: puente plano, puente triangular o puente con doble soporte.

Pregunta LUM:
¿Ya sabemos cuál puente es más resistente?

Qué revisar:
IPU: ¿todos usaron la misma cantidad de palitos?, ¿el pegamento secó igual?, ¿se anotó el peso exacto antes de romperse?
CPV: ¿todos siguieron instrucciones comparables?, ¿se puso el peso en el mismo lugar del puente?
A_norm: ¿se hizo más de un puente por diseño?, ¿se probaron pesos pequeños, medianos y grandes?
κ_conf: si un puente triangular aguanta mucho y otro igual se rompe rápido, ¿qué diferencia hubo?

Respuesta orientativa:
Si cada puente fue construido con materiales distintos o probado de manera distinta, no hay cierre. Puede haber una pista, pero no una conclusión fuerte. El PSNC sería fabricar varios puentes por diseño, usar la misma cantidad de material y probar el peso de forma igual.

Traducción LUM:
CPV mide si las prácticas son comparables. Sin operaciones comunes, los resultados no hablan el mismo idioma.

Ejercicio 4. Barquitos de aluminio: ¿cuál flota con más monedas?

Situación:
Cada equipo hace un barquito con papel aluminio. Luego agregan monedas una por una hasta que se hunde. Gana el que carga más monedas.

Pregunta LUM:

¿El mejor barquito ganó por buen diseño o por suerte?

Qué revisar:
IPU: ¿se contaron bien las monedas?, ¿todos usaron el mismo tamaño de papel aluminio?, ¿el agua estaba igual?
CPV: ¿todos colocaron las monedas de la misma manera?, ¿se permitió doblar varias veces?, ¿se siguieron reglas comunes?
A_norm: ¿cada barquito se probó una sola vez o varias veces?, ¿se compararon formas anchas, profundas y planas?
κ_conf: si un barquito ancho se hunde rápido y otro ancho aguanta mucho, ¿qué cambió?

Respuesta orientativa:
El resultado necesita repetición. Si un barquito gana una vez, puede ser casualidad. Si un diseño gana varias veces bajo reglas iguales, empieza a haber cierre.

Traducción LUM:
CLARION no significa “nunca fallará”. Significa “tenemos suficiente estabilidad para tomar una decisión razonable”.

Ejercicio 5. Receta de limonada: ¿cuál sabe mejor?

Situación:
La clase quiere encontrar la mejor receta de limonada. Un equipo usa más limón, otro más azúcar y otro más agua. Después, varios niños prueban y votan.

Pregunta LUM:
¿Podemos decir cuál receta es mejor?

Qué revisar:
IPU: ¿se midieron bien las cantidades?, ¿se anotó la receta exacta?, ¿todos probaron la misma cantidad?

CPV: ¿todos usaron los mismos vasos?, ¿los limones eran parecidos?, ¿se mezcló igual?
A_norm: ¿probaron suficientes niños?, ¿se repitió otro día?, ¿se compararon varias proporciones?
κ_conf: si una receta gana en un grupo y pierde en otro, ¿es contradicción o gusto diferente?

Respuesta orientativa:
Aquí aparece un problema interesante: no todo cierre es igual. En una receta, parte del resultado depende del gusto. Para tener mejor cierre, habría que definir antes qué significa “mejor”: más dulce, más ácida, más refrescante o preferida por la mayoría. Sin esa definición, el campo queda confuso.

Traducción LUM:
Antes de medir, hay que definir qué cuenta como resultado. LUM no elimina el juicio previo; lo obliga a declararse.

Ejercicio 6. Imanes: ¿qué objetos atrae un imán?

Situación:
La maestra da un imán y varios objetos: clip, moneda, borrador, llave, lápiz, cuchara de plástico y tornillo. Los niños predicen cuáles serán atraídos.

Pregunta LUM:
¿La clase puede cerrar una regla sobre los imanes?

Qué revisar:
IPU: ¿se probaron todos los objetos?, ¿se anotó cuáles sí y cuáles no?, ¿el imán funcionaba bien?
CPV: ¿todos acercaron el imán de la misma manera?, ¿se usó el mismo imán?
A_norm: ¿se probaron objetos de distintos materiales?, ¿se repitió con otro imán?
κ_conf: si una moneda no se pega pero una llave sí, ¿la regla “los metales se pegan” está completa?

Respuesta orientativa:
La clase puede descubrir que no todos los metales se comportan igual. Algunos objetos metálicos son atraídos y otros no. La regla inicial necesita corregirse: los imanes atraen ciertos materiales, como el hierro o el acero, no cualquier metal.

Traducción LUM:
κ_conf ayuda a corregir reglas demasiado simples. Una contradicción bien observada no destruye el conocimiento; lo mejora.

Ejercicio 7. Sombras: ¿qué cambia el tamaño de una sombra?

Situación:
Los niños usan una linterna y un muñeco. Mueven la linterna cerca y lejos para observar el tamaño de la sombra.

Pregunta LUM:
¿Ya entendieron qué hace crecer o reducir una sombra?

Qué revisar:
IPU: ¿se midió la sombra?, ¿se anotó la distancia de la linterna?, ¿la pared era la misma?
CPV: ¿todos colocaron la linterna y el muñeco igual?, ¿se cambió solo una cosa a la vez?
A_norm: ¿se probaron varias distancias?, ¿se repitió con otros objetos?
κ_conf: si una sombra cambia aunque la distancia sea igual, ¿se movió el objeto?, ¿cambió el ángulo?

Respuesta orientativa:
Para cerrar una regla, deben cambiar una variable a la vez. Si mueven al mismo tiempo la linterna, el muñeco y la pared, no sabrán qué causó el cambio. El PSNC sería separar variables: primero distancia, luego ángulo, luego tamaño del objeto.

Traducción LUM:

Cuando muchos factores se mezclan, hay interferencia. El no-cierre puede venir de no haber separado bien las causas.

Ejercicio 8. Germinación con música: ¿la música ayuda a crecer a las plantas?

Situación:
Un grupo pone música clásica a una planta. Otro grupo no pone música. Después de dos semanas comparan alturas.

Pregunta LUM:
¿Podemos decir que la música hizo crecer más a la planta?

Qué revisar:
IPU: ¿se midió bien la altura?, ¿se anotaron agua, luz y temperatura?
CPV: ¿las plantas eran del mismo tipo?, ¿estaban en macetas iguales?, ¿recibieron la misma agua?
A_norm: ¿hubo varias plantas con música y varias sin música?, ¿se repitió el experimento?
κ_conf: si una planta sin música creció más que varias con música, ¿qué significa?

Respuesta orientativa:
Este experimento enseña prudencia. Si solo hay una planta con música y una sin música, no hay cierre. Tal vez una recibió más sol. Tal vez una semilla era mejor. Hay que repetir con varias plantas y controlar condiciones.

Traducción LUM:
LUM ayuda a no enamorarse de una explicación bonita antes de tener cierre suficiente. Bonito no significa probado. Duro, pero justo.

Ejercicio 9. El juego de memoria: ¿un método ayuda a recordar más?

Situación:

La clase quiere saber si dibujar palabras ayuda a recordarlas. Un grupo estudia una lista leyendo; otro grupo dibuja cada palabra. Luego todos intentan recordar.

Pregunta LUM:
¿Dibujar ayuda a memorizar?

Qué revisar:
IPU: ¿todos tuvieron el mismo tiempo?, ¿se anotaron bien los aciertos?, ¿alguien vio la lista después?
CPV: ¿los grupos siguieron instrucciones comparables?, ¿las palabras tenían dificultad parecida?
A_norm: ¿se repitió con otra lista?, ¿se probó otro día?, ¿se usaron suficientes alumnos?
κ_conf: si dibujar ayuda con animales pero no con números, ¿la regla debe cambiar?

Respuesta orientativa:
Puede haber cierre parcial: dibujar quizá ayuda con palabras visuales, pero no con palabras abstractas. Entonces la conclusión debe ser más precisa. No “dibujar ayuda siempre”, sino “dibujar ayuda en cierto tipo de palabras y bajo ciertas condiciones”.

Traducción LUM:
Un cierre bueno no siempre es universal. A veces cerrar bien significa delimitar mejor.

Ejercicio 10. Sobrecierre: la feria de ciencias demasiado perfecta

Situación:
Un equipo presenta un experimento donde todo sale perfecto: las plantas crecen exactamente igual, los aviones vuelan exactamente la misma distancia, las tablas no tienen ningún error y todos los dibujos son idénticos. Nadie puede explicar cómo midió.

Pregunta LUM:
¿Eso parece cierre fuerte o sobrecierre?

Respuesta orientativa:
Parece sobrecierre. En un experimento real suele haber pequeñas variaciones. Si todo sale idéntico, puede haber copia, medición inventada o una prueba demasiado fácil. La perfección absoluta no siempre es buena señal.

Traducción LUM:
El sobrecierre es el falso orden. LUM no solo desconfía del caos; también desconfía de la perfección sospechosa.

Ejercicio 11. Auditoría: la caja del experimento

Situación:
La maestra pide que cada equipo entregue una “caja de auditoría” con:

- instrucciones del experimento;
- materiales usados;
- datos anotados;
- errores encontrados;
- fotos o dibujos del proceso;
- conclusión;
- qué repetirían para mejorar.

Pregunta LUM:
¿Por qué no basta entregar solo la conclusión?

Respuesta orientativa:
Porque otra persona debe poder revisar cómo llegaron al resultado. Si solo dicen “nuestro avión es el mejor”, eso es una opinión. Si muestran materiales, reglas, medidas y errores, otro equipo puede reconstruir el experimento.

Traducción LUM:

Auditar no es desconfiar por deporte. Es permitir que el conocimiento pueda revisarse, repetirse y corregirse.

Ejercicio 12. CLARION escolar: decidir con semáforo

Situación:
Después de varios experimentos, la clase usa un semáforo:

Verde: podemos actuar porque hay cierre suficiente.
Amarillo: hay pistas, pero falta probar más.
Rojo: todavía no sabemos; hay que hacer PSNC.
Negro: todo parece demasiado perfecto o mal validado.

Pregunta LUM:
¿Qué color pondrías a estos casos?

Caso A:
El avión de alas anchas ganó 12 de 15 lanzamientos, todos con el mismo papel, mismo patio y misma regla.
Color probable: Verde.

Caso B:
La planta con música creció más, pero solo había una planta por grupo.
Color probable: Amarillo o Rojo.

Caso C:
Todos los equipos entregaron exactamente los mismos datos, sin errores y sin explicar cómo midieron.
Color probable: Negro.

Caso D:
El puente triangular ganó una vez, pero se usó más pegamento que en los otros.
Color probable: Rojo.

Traducción LUM:
El semáforo no es decoración. Es una forma de decidir qué se puede hacer con lo que se sabe: actuar, investigar más, abstenerse o auditar el falso cierre.

Cierre del anexo

Estos ejemplos muestran que LUM no empieza con fórmulas, sino con una intuición sencilla: no basta acertar una vez, ni tener muchos datos, ni parecer ordenado. Para decir que un campo tiene cierre, hay que revisar la calidad del material, la estabilidad de las operaciones, la estructura de la evidencia, las contradicciones internas y la posibilidad de que otro reconstruya el dictamen. La ciencia adulta es más compleja que un salón de primaria, claro. Pero la lógica mínima es parecida: antes de avanzar, hay que saber si de verdad se aprendió o si solo estamos celebrando un experimento bonito

Referencias

Aristóteles. (1985). Ética a Nicómaco (J. Pallí Bonet, Trad.). Gredos.

Bueno, G. (1992–1993). Teoría del cierre categorial (Vols. 1–5). Pentalfa.

Bueno, G. (1995). ¿Qué es la ciencia? La respuesta de la teoría del cierre categorial. Pentalfa.

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Rand, D. G., & Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. Nature Human Behaviour, 2(9), 637–644. https://doi.org/10.1038/s41562-018-0399-z

Campbell, D. T. (1979). Assessing the impact of planned social change. Evaluation and Program Planning, 2(1), 67–90. https://doi.org/10.1016/0149-7189(79)90048-X

Chambers, C. D. (2013). Registered reports: A new publishing initiative at Cortex. Cortex, 49(3), 609–610. https://doi.org/10.1016/j.cortex.2012.12.016

Feyerabend, P. (1975). Against method: Outline of an anarchistic theory of knowledge. New Left Books.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. Biometrika, 80(1), 27–38. https://doi.org/10.1093/biomet/80.1.27

Ioannidis, J. P. A. (2005). Why most published research findings are false. PLoS Medicine, 2(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Kuhn, T. S. (2012). The structure of scientific revolutions (4.ª ed.). University of Chicago Press.

Lakatos, I., & Musgrave, A. (Eds.). (1970). Criticism and the growth of knowledge. Cambridge University Press.

Nosek, B. A., & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. Social Psychology, 45(3), 137–141. https://doi.org/10.1027/1864-9335/a000192

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. Science, 349(6251), aac4716. https://doi.org/10.1126/science.aac4716

Peng, R. D. (2011). Reproducible research in computational science. Science, 334(6060), 1226–1227. https://doi.org/10.1126/science.1213847

Popper, K. R. (1959). The logic of scientific discovery. Basic Books.

Rojas Aguayo, J. D. (2026). LUMINOMÁTICA (LUM): Sistema Gnoseo-Operatorio Unificado para Demarcación, Cierre Científico (CLARION) y Planes de No-Cierre (PSNC), con Contrato Auditable [Ensayo integral]. Laboratorio de Luminomática. https://doi.org/10.5281/zenodo.19211260

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychological Science, 22(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Stodden, V., McNutt, M., Bailey, D. H., Deelman, E., Gil, Y., Hanson, B., Heroux, M. A., Ioannidis, J. P. A., & Taufer, M. (2016). Enhancing reproducibility for computational methods. Science, 354(6317), 1240–1241. https://doi.org/10.1126/science.aah6168

Strathern, M. (1997). "Improving ratings": Audit in the British university system. European Review, 5(3), 305–321. https://doi.org/10.1017/S1062798700002660

Van Calster, B., Nieboer, D., Vergouwe, Y., De Cock, B., Pencina, M. J., & Steyerberg, E. W. (2016). A calibration hierarchy for risk models was defined: From utopia to empirical data. Journal of Clinical Epidemiology, 74, 167–176. https://doi.org/10.1016/j.jclinepi.2015.12.005