

Chapter 19 / Capítulo 19

Critical reading of scientific articles on
health: A practical guide (Spanish Edition)
ISBN: 978-9915-9928-3-9
DOI: 10.62486/978-9915-9928-3-9.Ch19
Pages: 333-343



© 2026 The authors. This is an open access
book chapter distributed under the terms of
the Creative Commons Attribution (CC BY)
4.0 License.

Use of NotebookLM in the Comparative Analysis of Scientific Evidence

Uso de NotebookLM en el análisis comparativo de evidencia científica

Mauricio Alejandro Paz del Rio  

Fidel Aguilar Medrano  

Leonel Rivero Castedo  

1 Clínica Los Andes. La Paz, Bolivia.

2 Hospital Municipal de Morón Ostaciana B. De Lavignolle. Buenos Aires, Argentina.

3 Hospital General de Agudos José María Ramos Mejía. Buenos Aires, Argentina.

4 Universidad Privada Franz Tamayo (UNIFRANZ). Facultad de Ciencias de la Salud. Santa Cruz, Bolivia.

ABSTRACT

Introduction: Comparing multiple studies is an essential competency in Evidence-Based Medicine (EBM), but it is a complex process due to the heterogeneity of study designs, populations, interventions, and outcomes. In this context, NotebookLM emerges as an artificial intelligence tool designed to facilitate the organization, synthesis, and exploration of scientific documents from user-provided sources, promoting a structured approach to comparative evidence analysis. **Development:** NotebookLM allows users to summarize articles, identify similarities and discrepancies, formulate cross-sectional comparisons, and reduce initial cognitive load by providing responses supported by citations from the original documents. Its greatest utility lies in the preliminary stages of document analysis, supporting the organization of the corpus and the generation of comparative hypotheses. However, the interpretation of methodological quality, risk of bias, effect size, and clinical applicability continues to depend on the researcher's critical judgment. **Conclusions:** The responsible integration of NotebookLM into EBM demonstrates that artificial intelligence should be understood as a complementary resource and not as a substitute for scientific reasoning. Its true value lies in enhancing the efficiency of document analysis, while the validity of the conclusions continues to rely on critical evaluation, verification of original sources, and methodological interpretation by trained professionals.

Keywords: Evidence-Based Medicine; Artificial Intelligence; Literature Review; Biomedical Research; Clinical Decision Making

RESUMEN

Introducción: La comparación de múltiples estudios constituye una competencia esencial en la Medicina Basada en la Evidencia (MBE), pero representa un proceso complejo debido a la heterogeneidad de diseños, poblaciones, intervenciones y desenlaces. En este contexto, NotebookLM emerge como una herramienta de inteligencia artificial orientada a facilitar la organización, síntesis y exploración de documentos científicos a partir de fuentes proporcionadas por el usuario, favoreciendo una aproximación estructurada al análisis comparativo de la evidencia. **Desarrollo:** NotebookLM permite resumir artículos, identificar similitudes y discrepancias, formular comparaciones transversales y reducir la carga cognitiva inicial mediante respuestas sustentadas en citas de los documentos originales. Su mayor utilidad se concentra en las etapas preliminares del análisis documental, apoyando la organización del corpus y la generación de hipótesis comparativas. No obstante, la interpretación de la calidad metodológica, el riesgo de sesgo, la magnitud del efecto y la aplicabilidad clínica continúa dependiendo del juicio crítico del investigador. **Conclusiones:** La integración responsable de NotebookLM en la MBE evidencia que la inteligencia artificial debe entenderse como un recurso complementario y no como un sustituto del razonamiento científico. Su verdadero valor radica en potenciar la eficiencia del análisis documental, mientras que la validez de las conclusiones sigue sustentándose en la evaluación crítica, la verificación de las fuentes originales y la interpretación metodológica realizada por profesionales capacitados.

Palabras clave: Medicina Basada en la Evidencia; Inteligencia Artificial; Revisión de la Literatura; Investigación Biomédica; Toma de Decisiones Clínicas

1. CASO CLÍNICO

Un residente de segundo año de Medicina Interna recibe a una paciente de 68 años con fibrilación auricular no valvular, enfermedad renal crónica estadio 3 y antecedente de sangrado digestivo alto hace dos años. Durante la discusión del caso surge una pregunta clínica frecuente pero compleja: ¿cuál es la mejor estrategia anticoagulante para reducir eventos tromboembólicos sin aumentar de forma desproporcionada el riesgo de sangrado en este perfil de paciente?

El residente revisa un ensayo clínico, luego una cohorte retrospectiva, después una revisión sistemática y finalmente una guía clínica reciente. Muy pronto nota que el problema ya no es solo “leer un artículo”, sino comparar varios documentos con diseños distintos, poblaciones no idénticas, desenlaces diferentes y conclusiones que no siempre apuntan en la misma dirección. Un ensayo informa superioridad en eficacia, una cohorte sugiere mejor seguridad, una revisión sistemática reporta heterogeneidad importante y la guía resume recomendaciones con distintos niveles de evidencia.

En ese momento surge una dificultad que es cada vez más común en la práctica académica y clínica: el lector dispone de mucha información, poco tiempo y le cuesta ordenar la información, contrastarla y sintetizarla. No basta con leer por separado; se necesita construir una visión comparativa. En este contexto, herramientas como NotebookLM pueden ser útiles como apoyo inicial, porque permiten cargar varios documentos, interrogarlos de manera conjunta, solicitar resúmenes comparativos y ubicar, con apoyo de citas al texto fuente, los puntos de coincidencia o discrepancia entre los estudios. Sin embargo, su utilidad depende de que el lector mantenga el control metodológico y no confunda una síntesis automática con una evaluación crítica formal.

Este capítulo parte de esa necesidad real: aprender a usar NotebookLM como una herramienta de apoyo para comparar evidencia científica, sin perder de vista que el verdadero análisis crítico sigue dependiendo del razonamiento del clínico o investigador.

2. OBJETIVOS

NotebookLM tiene como objetivo facilitar la organización, comprensión y exploración de múltiples documentos cargados por el usuario. Busca resumir información, responder preguntas con base en las fuentes proporcionadas, organizar contenidos y generar productos de apoyo al aprendizaje, como esquemas, autoevaluaciones, flashcards, resúmenes en audio y presentaciones.

Además, su finalidad es ayudar a comparar estudios, identificando similitudes y diferencias entre diseños, poblaciones, intervenciones, desenlaces y resultados. Sin embargo, su objetivo no es sustituir la lectura crítica ni la valoración metodológica formal, sino servir como apoyo para una síntesis preliminar y estructurada de la evidencia científica.

3. INTRODUCCIÓN

Leer un solo artículo ya exige habilidades complejas: identificar la pregunta de investigación, comprender el diseño, valorar la validez interna, interpretar los resultados y decidir si el estudio es aplicable al paciente o al contexto concreto. Sin embargo, en la práctica clínica real y en la formación académica avanzada, rara vez basta con un único artículo. La mayoría de las decisiones relevantes obligan a comparar varios estudios, a veces de distinto tipo, distinta calidad metodológica y diferente pertinencia clínica.

La comparación de varios artículos añade un nivel adicional de complejidad. Ya no se trata únicamente de entender qué concluye cada estudio, sino de responder preguntas como estas: ¿los pacientes eran comparables?, ¿los desenlaces fueron equivalentes?, ¿las diferencias observadas se explican por la intervención o por el diseño?, ¿la heterogeneidad es clínica, metodológica o estadística?, ¿existe coherencia entre los hallazgos o solo una apariencia superficial de consenso?

Este trabajo comparativo suele ser lento y cognitivamente demandante. Exige revisar muchos textos, subrayar secciones clave, tomar notas, construir tablas de extracción y volver varias veces a los documentos originales. En ese escenario, las herramientas de inteligencia artificial ancladas en fuentes específicas pueden servir como apoyo para organizar información, resumirla y facilitar una primera visión panorámica. NotebookLM fue diseñado precisamente como un asistente de investigación apoyado en las fuentes que el propio usuario carga, y sus respuestas incluyen citas en línea que permiten regresar al texto fuente y verificar la afirmación. Esa característica lo diferencia de un chatbot generalista y lo hace particularmente atractivo para tareas de organización y comparación documental.

No obstante, la utilidad práctica de una herramienta no elimina sus limitaciones epistemológicas. Una síntesis generada por IA puede ahorrar tiempo, pero también puede simplificar en exceso, omitir matices relevantes o inducir sesgos de interpretación si el lector delega demasiado. Las revisiones recientes sobre IA y síntesis de evidencia muestran un crecimiento acelerado del campo, pero también señalan que la supervisión humana sigue siendo indispensable y que la evidencia actual no justifica reemplazar el juicio experto por automatización generativa sin control metodológico.

Por ello, este capítulo no presenta a NotebookLM como un sustituto de la lectura crítica, sino como un instrumento de apoyo dentro de un proceso más amplio de análisis comparativo de evidencia científica.

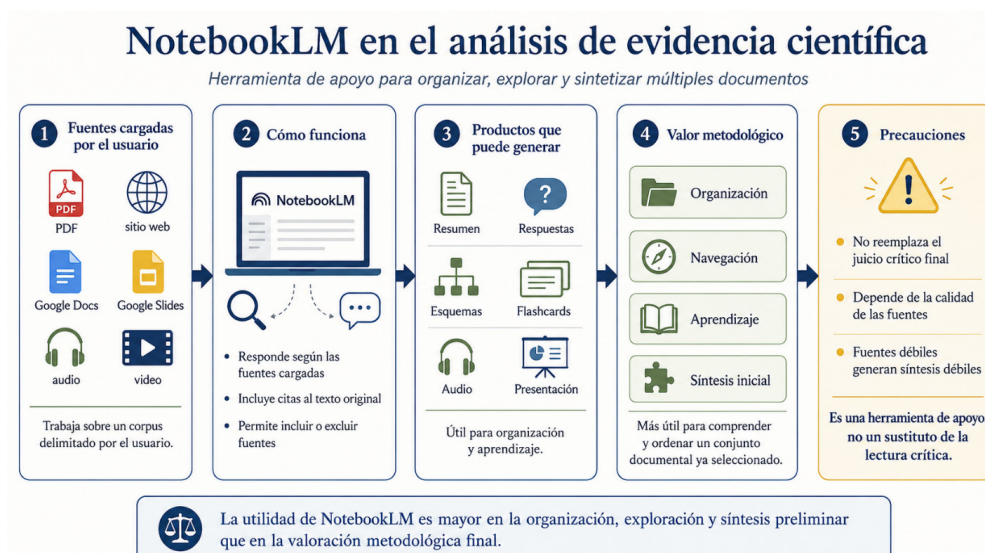
4. FUNDAMENTOS DE NOTEBOOKLM EN EL ANÁLISIS DE EVIDENCIA CIENTÍFICA

NotebookLM es un asistente de investigación impulsado por inteligencia artificial que trabaja sobre un conjunto de fuentes delimitadas por el propio usuario. De acuerdo con la documentación oficial de Google, permite cargar documentos como PDF, sitios web, archivos de Google Docs, Google Slides, audio, videos y otras fuentes compatibles, y luego interactuar con todo lo cargado para obtener resúmenes, respuestas y distintos productos organizativos. Su rasgo más importante para fines académicos es que las respuestas se generan a partir de las fuentes cargadas, con citas en línea que permiten revisar el fragmento o la ubicación del contenido en el documento original.

Desde el punto de vista metodológico, esto implica que NotebookLM no opera, idealmente, como una conversación abierta sobre “cualquier cosa”, sino como un entorno de trabajo documental. Cuando el usuario formula una pregunta, el sistema recupera la información más relevante dentro de las fuentes disponibles que se cargaron y construye una respuesta apoyada en ese material. Además, el usuario puede incluir o excluir determinadas fuentes para refinar la respuesta, lo cual resulta muy valioso cuando se desea contrastar subgrupos de documentos, por ejemplo: ensayos por un lado y revisiones por otro, o estudios recientes frente a estudios clásicos.

Esta arquitectura lo vuelve especialmente útil para el análisis de evidencia científica cuando el problema no es crear contenido original, sino comprender mejor un conjunto documental ya seleccionado. En otras palabras, NotebookLM puede ser más valioso en la etapa de organización, navegación, aprendizaje y síntesis inicial del material que en la etapa de juicio metodológico final.

Figura 1: NotebookLM como herramienta de apoyo en el análisis de evidencia científica.



Fuente: Elaboración propia apoyada con inteligencia artificial.

Es importante subrayar que el valor de NotebookLM depende de la calidad de documentos que se le carga. Si el usuario utiliza fuentes débiles, no pertinentes o desactualizadas, la herramienta no corregirá ese error de origen; solo organizará y resumirá un conjunto de documentos ya sesgado. En este sentido, la herramienta hereda tanto las fortalezas como las debilidades del

material que recibe.

5. APORTES DE NOTEBOOKLM EN LA COMPARACIÓN DE ARTÍCULOS CIENTÍFICOS

Cuando se trabaja con varios artículos sobre una misma pregunta, NotebookLM puede aportar valor en cinco planos principales.

Primero, puede resumir múltiples documentos y ofrecer una visión general del conjunto. Esto es útil cuando el lector ya seleccionó artículos relevantes, pero necesita una entrada rápida al panorama global antes de iniciar una lectura más minuciosa. En vez de revisar de manera lineal cada texto desde cero, puede obtener una síntesis preliminar que le permita ubicar los grandes temas, las intervenciones estudiadas, los desenlaces frecuentes y las líneas de conclusión dominantes.

Segundo, puede organizar hallazgos. Por ejemplo, en nuestro caso el residente puede pedir que agrupe todos los estudios que encontró según diseño, población, intervención, comparador o desenlace. Esta función es especialmente útil cuando el corpus contiene ensayos, cohortes, revisiones narrativas y guías, y se requiere ordenar la información antes de construir una tabla comparativa formal.

Tercero, puede detectar patrones repetidos. Si varios artículos coinciden en que cierto beneficio aparece solo en una subpoblación, o si varios estudios identifican el mismo desenlace adverso, NotebookLM puede ayudar a visualizar esa recurrencia y a ponerla sobre la mesa para análisis posterior. Del mismo modo, puede señalar discrepancias entre estudios, lo cual es muy valioso porque la divergencia suele ser el punto de partida de una lectura crítica más profunda.

Cuarto, puede facilitar preguntas cruzadas entre documentos. Por ejemplo: “¿Qué diferencias de edad media existen entre los estudios que encontraron beneficio y los que no?”, “¿cuáles usaron mortalidad como desenlace primario y cuáles usaron desenlaces compuestos?”, o “¿qué estudios excluyeron pacientes con insuficiencia renal avanzada?”. Este tipo de interrogación transversal es precisamente donde una herramienta documental puede ahorrar tiempo.

Quinto, puede reducir la carga cognitiva inicial. La literatura sobre automatización de síntesis de evidencia muestra que la IA está siendo cada vez más usada para acelerar pasos del proceso, pero esa aceleración es útil sobre todo cuando se mantiene supervisión humana y verificación metodológica. En términos prácticos, NotebookLM puede ayudar a pasar más rápido del “caos documental” a un mapa preliminar del problema.

Sin embargo, ninguno de estos aportes equivale por sí mismo a una conclusión científica robusta. El hecho de que la herramienta identifique coincidencias o discrepancias no significa que ya haya ponderado adecuadamente el riesgo de sesgo, la magnitud del efecto, la heterogeneidad clínica o la aplicabilidad externa. Esa parte sigue siendo humana.

6. USO PRÁCTICO DE NOTEBOOKLM PARA ANALIZAR MÚLTIPLES ARTÍCULOS

El uso más provechoso de NotebookLM comienza antes de abrir la herramienta. El primer paso no es “preguntar a la IA”, sino seleccionar bien el corpus. El usuario debe definir con claridad la pregunta clínica o académica, idealmente en formato estructurado, y luego cargar artículos que respondan realmente a esa pregunta. En esta fase conviene evitar mezclar documentos demasiado heterogéneos si no existe una razón metodológica para hacerlo.

Una vez cargados los documentos, puede solicitarse una primera síntesis general con preguntas como: “resume los hallazgos principales de estos estudios”, “ordena los artículos por diseño” o “identifica en qué puntos coinciden y en qué puntos discrepan”. Esa primera interacción no reemplaza la lectura, pero permite construir una orientación inicial del terreno.

El siguiente paso consiste en formular preguntas comparativas específicas. Esta es probablemente la fase más útil desde el punto de vista metodológico. En lugar de pedir resúmenes vagos, conviene hacer preguntas dirigidas: “compara las características basales de las poblaciones”, “menciona diferencias en los criterios de inclusión”, “separa los desenlaces primarios y secundarios”, “indica cuáles estudios reportan intervalos de confianza y cuáles no”, “resume los eventos adversos reportados” o “muestra qué estudios tienen mayor seguimiento”.

Después, el lector puede pedir una síntesis preliminar conjunta, por ejemplo: “integra los hallazgos sobre eficacia” y luego, por separado, “integra los hallazgos sobre seguridad”. Esta separación es útil porque evita que la herramienta mezcle planos conceptuales distintos.

Finalmente, la fase más importante consiste en volver al artículo original. Toda afirmación

relevante hallada en NotebookLM debe ser corroborada en el texto fuente, especialmente si afecta la interpretación clínica o académica. Las citas en línea son una ventaja justamente porque facilitan esa verificación. Desde la lógica de la lectura crítica, NotebookLM debe funcionar como una mesa de trabajo que ayuda a localizar, ordenar y resumir, pero no como una autoridad final.

Figura 2. Interfaz de carga de fuentes en NotebookLM.



Fuente: Elaboración propia a partir de captura de pantalla de NotebookLM(Google LLC, 2026).

En la práctica, un flujo razonable sería el siguiente: selección de estudios → carga en NotebookLM → resumen inicial → preguntas comparativas específicas → extracción preliminar de similitudes y diferencias → verificación en textos originales → tabla comparativa propia del lector → interpretación crítica final. Esa última tabla, idealmente construida por el investigador o clínico, sigue siendo el núcleo del análisis serio.

7. ANÁLISIS COMPARATIVO DE EVIDENCIA CIENTÍFICA ASISTIDO POR INTELIGENCIA ARTIFICIAL

Comparar evidencia científica no significa solo reunir artículos; significa interpretar por qué coinciden o por qué discrepan. Allí es donde la inteligencia artificial puede ayudar, pero solo dentro de un marco metodológico claro.

El lector debe aprender a contrastar al menos cinco dimensiones. La primera es el **diseño del estudio**. No tiene el mismo peso interpretativo un ensayo clínico aleatorizado que una cohorte retrospectiva o una serie de casos. Si NotebookLM resume conclusiones “similares”, el lector debe preguntarse si esas conclusiones proceden de evidencias metodológicamente equivalentes o de diseños con grados de confianza muy diferentes.

La segunda **dimensión es la población estudiada**. A veces dos artículos parecen contradictorios, pero en realidad estudian pacientes distintos: unos incluyen ancianos frágiles, otros adultos jóvenes; unos se realizan en cuidados intensivos, otros en consulta externa; unos excluyen comorbilidades mayores, otros las incorporan. La IA puede ayudar a identificar esas diferencias descriptivas, pero corresponde al lector valorar su impacto clínico.

La tercera es el **desenlace**. Un estudio puede mostrar beneficio en un desenlace intermedio y otro no mostrar efecto en mortalidad. Si esto no se distingue con claridad, se corre el riesgo de comparar resultados que en realidad no son equivalentes.

La cuarta dimensión es la **medida del efecto y su precisión**. No basta con que un estudio “favorezca” una intervención; hay que revisar cómo se expresó ese hallazgo, si hubo intervalos de confianza estrechos o amplios, si el resultado fue clínicamente importante, y si la magnitud del efecto podría explicarse por sesgo o azar.

La quinta dimensión es la **calidad metodológica**. Una discrepancia entre estudios puede nacer de diferencias reales en la intervención, pero también de problemas de sesgo de selección, sesgo de información, pérdidas al seguimiento, análisis incompletos o sobreajuste estadístico. Aquí la IA puede señalar que existen diferencias; pero la interpretación del valor de esas diferencias

es una tarea esencialmente crítica y humana.

Las revisiones recientes sobre IA aplicada a síntesis de evidencia y uso de LLM en salud coinciden en algo fundamental: estas tecnologías **pueden acelerar tareas, organizar información y asistir al usuario**, pero no están listas para sustituir el control humano, especialmente en dominios de alto impacto como la medicina. La colaboración humano-IA parece ser el modelo más razonable: la herramienta ayuda a navegar el volumen informativo, mientras el clínico o investigador conserva la responsabilidad de juzgar validez, pertinencia y aplicabilidad.

8. APLICACIONES CLÍNICAS, ACADÉMICAS Y DE INVESTIGACIÓN

En el **ámbito clínico**, NotebookLM puede ser útil para preparar discusiones de casos en las que existen varios artículos relevantes y se necesita una comparación rápida antes de una sesión académica o una decisión compartida. Puede ayudar, por ejemplo, a ordenar la evidencia preliminar sobre un tratamiento, una prueba diagnóstica o un factor pronóstico.

En el **ámbito docente**, resulta especialmente valioso en clubs de revistas, seminarios de residentes y clases de lectura crítica. Un grupo puede cargar varios artículos y usar la herramienta para generar una visión panorámica que luego se discuta críticamente. Esto no sustituye la lectura individual, pero sí puede enriquecer la discusión grupal al hacer más visibles las discrepancias entre estudios.

En el ámbito de investigación, puede servir para apoyar la construcción de marcos teóricos, revisiones narrativas, introducciones de tesis o presentaciones académicas. También puede ser útil cuando el investigador desea identificar rápidamente qué líneas temáticas se repiten en varios artículos y qué vacíos persisten.

En todos estos escenarios, el beneficio principal no es “pensar por el usuario”, sino facilitar el acceso estructurado al contenido de múltiples documentos. Esa es una diferencia central entre usar una IA como asistente de organización documental y usarla como reemplazo del análisis científico.

9. RIESGOS, LIMITACIONES Y PRECAUCIONES EN EL USO DE NOTEBOOKLM

La primera limitación de NotebookLM es su dependencia absoluta de los documentos cargados. Si los documentos seleccionados son incompletos, sesgados, irrelevantes o de baja calidad, la herramienta organizará bien un problema mal planteado, pero no lo corregirá. La calidad de la síntesis no puede superar la calidad del material fuente.

La segunda limitación es que, aunque sus respuestas estén ancladas en fuentes, eso no elimina el **riesgo de simplificación excesiva**. Las herramientas generativas tienden a condensar, y al condensar pueden perder matices metodológicos importantes. En comparación de estudios, un matiz omitido puede cambiar toda la interpretación: diferencias en dosis, tiempos de seguimiento, definición de desenlaces o criterios de exclusión pueden desaparecer en un resumen demasiado limpio. La literatura reciente ha insistido en que los LLM pueden sobre-generalizar, omitir contexto o inducir confianza excesiva si el usuario no verifica.

La tercera limitación es el **riesgo de sesgo de automatización**. Cuando una herramienta presenta una síntesis clara y coherente, el usuario puede asumir que esa síntesis ya representa el “consenso” de la evidencia. Pero una síntesis automática no es equivalente a una revisión sistemática rigurosa, no reemplaza la evaluación del riesgo de sesgo ni resuelve por sí sola la heterogeneidad entre estudios. Las revisiones recientes en evidencia sintética asistida por IA señalan precisamente que el uso de GenAI sin supervisión humana no está respaldado por la evidencia actual.

La cuarta limitación es **metodológica**: NotebookLM puede ayudar a comparar textos, pero no garantiza que el usuario esté comparando entidades verdaderamente comparables. Si se mezclan documentos con preguntas de investigación distintas, poblaciones no equivalentes o fines divergentes, la comparación resultante puede ser superficial o engañosa.

La quinta limitación es **ética y profesional**. En salud, cualquier uso de IA exige cautela porque los errores no son solo conceptuales: pueden traducirse en decisiones clínicas equivocadas, docencia deficiente o conclusiones de investigación mal fundamentadas. La literatura ética sobre LLM en medicina insiste en la necesidad de supervisión humana, transparencia y responsabilidad profesional.

10. RECOMENDACIONES PARA UN USO RESPONSABLE EN EL ANÁLISIS COMPARATIVO

NotebookLM debe usarse después de una selección inicial razonada de artículos, no como sustituto de la búsqueda bibliográfica ni del cribado metodológico. Conviene trabajar con artículos de investigación definidos por una pregunta clara y evitar mezclar documentos heterogéneos sin un propósito explícito.

Es recomendable formular preguntas comparativas concretas y no solo pedir “resúmenes”. Cuanto más específica sea la pregunta, más probable será que la herramienta recupere la información pertinente dentro del conjunto documental. La propia ayuda oficial de NotebookLM indica que, cuando existen muchas fuentes, una pregunta precisa mejora la localización del contenido relevante.

Toda afirmación importante debe verificarse en el artículo original. Esto es especialmente obligatorio cuando la respuesta afecta interpretación metodológica, medidas de efecto, resultados numéricos o implicaciones clínicas. La herramienta puede servir para encontrar rápidamente el lugar donde está una afirmación, pero no debe eximir de leer el fragmento o el estudio completo.

Tampoco debe utilizarse como reemplazo de una revisión sistemática formal. Una revisión sistemática exige protocolo, estrategia de búsqueda reproducible, criterios de elegibilidad, evaluación de riesgo de sesgo, métodos explícitos de síntesis y, cuando corresponde, metaanálisis. NotebookLM no suple ese proceso; a lo sumo puede asistir algunas etapas cognitivas o documentales.

Figura 3. Flujo práctico para el análisis comparativo de múltiples artículos con NotebookLM.



Fuente: Elaboración propia con apoyo de inteligencia artificial, 2026.

Por último, el usuario debe mantener siempre una postura epistemológica prudente: una salida bien redactada no es sinónimo de una salida metodológicamente correcta. La claridad expresiva de la IA no debe confundirse con validez científica.

11. UTILIDAD DENTRO DE LA MEDICINA BASADA EN LA EVIDENCIA

La Medicina Basada en la Evidencia no consiste en acumular artículos, sino en integrar tres componentes: la mejor evidencia disponible, la experiencia clínica y los valores y contexto del paciente. En este marco, NotebookLM puede ser útil porque ayuda a ordenar la evidencia disponible cuando esta se encuentra dispersa en varios documentos. Puede facilitar la transición entre la búsqueda bibliográfica y la lectura crítica comparativa, y reducir parte de la carga inicial de organización.

Sin embargo, la esencia de la MBE permanece intacta. La validez interna de un estudio, la fuerza del diseño, la precisión de los resultados, la relevancia clínica de los desenlaces y la aplicabilidad a un paciente concreto no pueden delegarse por completo a una herramienta automatizada. Incluso si NotebookLM ayuda a integrar y resumir, la decisión final sigue dependiendo de un juicio clínico y metodológico humano. Dicho de otro modo: NotebookLM

puede ser una herramienta útil dentro de la MBE, pero no redefine la MBE ni la reemplaza. Su papel es instrumental, no rector.

12. PREGUNTAS DE AUTOEVALUACIÓN

PREGUNTAS DE SELECCIÓN MÚLTIPLE

1. ¿Cuál es la característica más relevante de NotebookLM para el análisis comparativo de artículos científicos?
 - a) Genera respuestas sin depender de documentos previos.
 - b) Trabaja sobre fuentes cargadas por el usuario y ofrece respuestas ancladas en ellas.
 - c) Sustituye la necesidad de leer artículos originales.
 - d) Realiza automáticamente metaanálisis completos.
2. ¿Cuál de las siguientes tareas sí puede beneficiarse claramente de NotebookLM?
 - a) Reemplazar la evaluación del riesgo de sesgo.
 - b) Decidir por sí solo la aplicabilidad clínica final.
 - c) Organizar y resumir varios artículos para una comparación preliminar.
 - d) Garantizar que todos los estudios incluidos tengan alta calidad metodológica.
3. ¿Qué riesgo importante existe al usar NotebookLM para comparar estudios?
 - a) Que no pueda trabajar con más de un documento.
 - b) Que convierta automáticamente cohortes en ensayos clínicos.
 - c) Que el usuario confunda una síntesis automática con un consenso científico formal.
 - d) Que siempre responda con información ajena a las fuentes.
4. ¿Cuál es la mejor forma de usar NotebookLM en lectura crítica?
 - a) Como sustituto del juicio clínico.
 - b) Como apoyo para organizar y explorar el corpus antes de la interpretación crítica final.
 - c) Como mecanismo para evitar revisar los textos originales.
 - d) Como reemplazo de una revisión sistemática.
5. ¿Qué diferencia práctica importante existe entre NotebookLM y un LLM generalista?
 - a) NotebookLM no usa lenguaje natural.
 - b) NotebookLM solo sirve para crear presentaciones.
 - c) NotebookLM se apoya en un corpus documental delimitado por el usuario y ofrece citas al contenido fuente.
 - d) NotebookLM elimina completamente el riesgo de error interpretativo.

PREGUNTAS ABIERTAS

6. Explique por qué comparar varios artículos es metodológicamente más complejo que leer uno solo.
7. Describa dos aportes concretos y dos limitaciones importantes de NotebookLM en el análisis de evidencia científica.
8. ¿Por qué una síntesis generada por NotebookLM no debe confundirse con una revisión sistemática?
9. Mencione cómo integraría NotebookLM en un club de revistas o en una sesión bibliográfica.
10. Explique por qué el juicio crítico humano sigue siendo indispensable aunque la herramienta entregue respuestas con citas.

RESPUESTAS

1. b. NotebookLM destaca porque trabaja sobre fuentes aportadas por el usuario y genera respuestas apoyadas en ese corpus. Este rasgo es central en el capítulo, ya que permite una comparación más controlada de artículos que en un modelo generalista.
2. c. NotebookLM es útil para resumir, organizar y contrastar varios documentos. No puede reemplazar por sí mismo la evaluación metodológica final ni garantizar calidad científica del corpus.
3. c. Uno de los riesgos más importantes es interpretar una síntesis clara como si fuera una conclusión definitiva o un consenso formal. Esta confusión puede producir sesgo de automatización y llevar a errores en lectura crítica.

4. b. La herramienta debe utilizarse como apoyo inicial y organizativo. La interpretación crítica final siempre requiere retorno al texto original, valoración metodológica y juicio clínico.

5. c. La diferencia más útil para este capítulo es que NotebookLM trabaja con un conjunto documental concreto y muestra citas al material fuente, lo cual facilita verificación y rastreo del contenido.

6. Porque el lector no solo debe entender cada estudio por separado, sino también contrastar diseño, población, intervención, comparadores, desenlaces, medidas de efecto, calidad metodológica y aplicabilidad. La complejidad surge de integrar y jerarquizar diferencias, no solo de resumir resultados.

7. Entre los aportes: organización del corpus, resúmenes conjuntos, identificación de patrones y discrepancias, apoyo para preguntas transversales. Entre las limitaciones: dependencia del material cargado, simplificación excesiva, riesgo de automatización acrítica y ausencia de evaluación metodológica autónoma.

8. Porque una revisión sistemática requiere métodos explícitos y reproducibles: estrategia de búsqueda, criterios de inclusión, evaluación del riesgo de sesgo y síntesis formal. NotebookLM puede ayudar a organizar documentos, pero no reemplaza ese proceso.

9. Podría cargarse un grupo de artículos previamente seleccionados, pedir un resumen comparativo inicial, identificar coincidencias y discrepancias, y luego discutir críticamente cada punto con el grupo revisando los textos originales.

10. Porque incluso una respuesta con citas puede simplificar, omitir matices o resaltar información sin ponderar adecuadamente su importancia metodológica. La cita ayuda a verificar, pero no reemplaza el análisis del lector.

REFERENCIAS

- Google. Learn about NotebookLM. NotebookLM Help. 2026. Disponible en la ayuda oficial de Google NotebookLM.
- Google. Use chat in NotebookLM. NotebookLM Help. 2026. Disponible en la ayuda oficial de Google NotebookLM.
- Google. Frequently asked questions – NotebookLM Help. NotebookLM Help. 2026. Disponible en la ayuda oficial de Google NotebookLM.
- Clark J, Barton B, Albarqouni L, Byambasuren O, et al. Generative artificial intelligence use in evidence synthesis: A systematic review. *Res Synth Methods*. 2025. Available at: <https://doi.org/10.1017/rsm.2025.16>.
- Lieberum JL, Toews M, Metzendorf MI, Heilmeyer F, Siemens W, Haverkamp C, et al. Large language models for conducting systematic reviews: On the rise, but not yet ready for use—a scoping review. *J Clin Epidemiol*. 2025;181:111746. Available at: <https://doi.org/10.1016/j.jclinepi.2025.111746>.
- Harasgama S, Tunnicliffe DJ, Newman T, Campbell J, Haby MM, et al. Artificial intelligence tools for automating evidence synthesis: Scoping review. *J Med Internet Res*. 2026;28:e81597.
- Sousa MSA, Peiris S, Figueiró MF, Haby MM, Baraldi AC, Reveiz L, Souza JP. The landscape of artificial intelligence tools and platforms for evidence synthesis: A scoping review. *Syst Rev*. 2026;15(1). Available at: <https://doi.org/10.1186/s13643-025-02842-y>.
- Wang Z, Cao L, Danek B, Jin Q, Lu Z, Sun J. Accelerating clinical evidence synthesis with large language models. *NPJ Digit Med*. 2025;8(1):509. Available at: <https://doi.org/10.1038/s41746-025-01840-7>.
- Pratte M, Thirukumar S, Zhang C, Slessarev M, Basmaji J, Prager R. Can large language models approximate the results of meta-analyses in critical care? A meta-research study. *J Crit Care*. 2025;92:155358. Available at: <https://doi.org/10.1016/j.jcrc.2025.155358>.
- Haltaufderheide J, Ranisch R. The ethics of ChatGPT in medicine and healthcare: A systematic review on large language models (LLMs). *NPJ Digit Med*. 2024;7:183. Available at: <https://doi.org/10.1038/s41746-024-01157-x>.
- Tam TYC, Sivarajkumar S, Kapoor S, Stolyar AV, Polanska K, McCarthy KR, et al. A framework for human evaluation of large language models in healthcare derived from literature review. *NPJ Digit Med*. 2024;7(1):258. Available at: <https://doi.org/10.1038/s41746-024-01258-7>.
- Li HY, Fu JF, Python A, et al. Implementing large language models in health care: Clinician-focused review with interactive guideline. *J Med Internet Res*. 2025;27:e71916. Available at: <https://doi.org/10.2196/71916>.
- Ng KKY, Matsuba I, Zhang PC. RAG in health care: A novel framework for improving communication and decision-making by addressing LLM limitations. *NEJM AI*. 2025;2(1):Alra2400380.