

LUIS ENRIQUE LESCANO BORREGO

Inteligencia artificial para investigadores



UNA GUÍA PAASO A PASO

Da el paso hacia el futuro potenciando tus habilidades de investigación con IA



2020

Lo cambió todo

UN GIRO SILENCIOSO
PERO IRREVERSIBLE.

HOLA, MI NOMBRE ES LUIS ENRIQUE LESCANO

2020 lo cambió todo. Mientras el mundo se detenía, yo hice un giro silencioso pero irreversible: de los servicios bibliotecarios a la gestión de datos de investigación. Y descubrí algo que desde entonces guía todo mi trabajo:

Las universidades no tienen un problema de datos. Tienen un problema de estructura.

Vengo de las bibliotecas. Entendí desde adentro cómo el conocimiento se pierde — no porque no exista, sino porque nadie diseñó el sistema para que fuera encontrable, conectable o duradero. Ese es el problema que resuelvo ahora, pero a nivel de infraestructura.

Hoy trabajo en la intersección de la gestión de datos de investigación, la ciencia abierta y la inteligencia artificial — ayudando a universidades en América Latina a transformar datos dispersos en ecosistemas de conocimiento escalables y alineados con FAIR.

He trabajado con más de 15 universidades en América Latina. Formo profesionales. Asesoro instituciones. Y actualmente soy candidato a PhD investigando cómo evolucionan los sistemas de conocimiento en la era de la IA.

En 2026 di una charla TEDx en Quito sobre una sociedad que olvidó leer — porque la infraestructura del conocimiento no es solo técnica. Es profundamente humana.

FUNDAMENTOS CONCEPTUALES



Antes de usar IA en investigación, es clave entender qué hace realmente un modelo de lenguaje, cuáles son sus fallos más comunes y qué sesgos trae de fábrica. Esta sección te da las bases para no confiar ciegamente.

Cómo funciona un LLM

Antes de integrar la IA en tu flujo de trabajo, necesitas entender qué pasa dentro del modelo. No se trata de magia ni de inteligencia humana: los LLM son máquinas estadísticas que predicen la siguiente palabra más probable. Eso les permite generar textos sorprendentemente coherentes... pero también cometer errores que un humano no cometería. En esta sección verás por qué alucinan, qué sesgos arrastran desde su entrenamiento y por qué no puedes confiar ciegamente en sus respuestas. Con estos fundamentos, pasarás de usuario ingenuo a investigador crítico.

LA CAJA NEGRA (QUÉ ES UN LLM)

Un Large Language Model (LLM) es un sistema estadístico entrenado con enormes volúmenes de texto para predecir qué palabra viene después de otra. No "piensa", no "comprende" ni tiene acceso a información actualizada por defecto. Genera texto plausible — que puede ser correcto o completamente inventado.

La "caja negra" no es metáfora: ni sus propios creadores pueden explicar completamente por qué un modelo produce una respuesta específica. Esto tiene implicaciones directas para la ciencia:

- Los resultados de un LLM no son reproducibles de la misma forma que un experimento controlado.
- La confianza del modelo en su respuesta no correlaciona con su exactitud. Puede afirmar una falsedad con total seguridad.
- El modelo no distingue entre fuentes primarias verificadas y texto de baja calidad — pesa todos los patrones del entrenamiento.
- Modelos distintos (GPT-4, Claude, Gemini) producirán respuestas diferentes ante la misma consulta.

Regla práctica: *Tratar al LLM como un colega brillante pero que a veces inventa. Muy útil para estructurar ideas, inaceptable como fuente de hechos sin verificación.*

IA: DENTRO DE LA CAJA NEGRA

Explicar el concepto de "caja negra" en la inteligencia artificial, sus riesgos éticos y cómo mitigar la opacidad de los algoritmos.



ALUCINACIONES: EL RIESGO MÁS SUBESTIMADO

Las "alucinaciones" ocurren cuando el modelo genera contenido factualmente incorrecto con apariencia convincente. En investigación, esto se manifiesta como:

- Citas inventadas: el modelo genera autores, títulos, revistas, años y DOIs que no existen pero parecen reales.
- Datos estadísticos falsos: porcentajes, tamaños de muestra, p-valores fabricados con precisión numérica.
- Atribución incorrecta: asigna afirmaciones a autores que nunca las hicieron.
- Desactualización silenciosa: presenta como vigente información que fue superada por nueva evidencia.

La frecuencia varía por dominio: en áreas muy especializadas o en idioma español, la tasa de error aumenta considerablemente.



Error frecuente

Copiar la lista de referencias generada por el modelo directamente en el manuscrito. Un 30–50% de las citas de LLMs pueden ser inexistentes o incorrectas.

Cómo detectarlas

Verifica toda referencia en PubMed, Scopus o Google Scholar. Pide al modelo que indique su nivel de certeza. Usa Scite.ai para validar el contexto de citas.

SESGOS: UN APRENDIZAJE HEREDADO

Los sesgos de un LLM son el reflejo de los sesgos presentes en su corpus de entrenamiento — mayoritariamente texto en inglés, de fuentes anglófonas, de ciertas disciplinas y perspectivas.

- Sesgo lingüístico: el conocimiento científico producido en español, portugués o en el Sur Global está subrepresentado.
- Sesgo de citación: favorece artículos muy citados sobre investigaciones emergentes o interdisciplinarias.
- Sesgo de género y geografía: puede invisibilizar contribuciones de investigadoras o de instituciones latinoamericanas.
- Sesgo de disciplina: dominan las STEM occidentales; las ciencias sociales, humanidades e investigación local tienen menor peso.

IMPACTO DIRECTO PARA AMÉRICA LATINA:

El modelo puede desconocer literatura relevante en español o subvalorar contextos regionales. Siempre complementa con búsquedas en bases de datos especializadas (SciELO, REDIB, Latindex).



LATAM-GPT

LATAM-GPT es un bien público lanzado en marzo de 2026, sobre el cual diversos actores (universidades, centros de investigación, gobiernos, startups) podrán construir soluciones de IA específicas, democratizando así el acceso a herramientas clave para la competitividad y el desarrollo regional.



CAPÍTULO 1

TÉCNICAS PARA UN USO MÁS CONFIABLE



No basta con preguntar. Para obtener respuestas útiles y fiables necesitas configurar bien el modelo, anclarlo a tus documentos y saber cuándo realmente vale la pena entrenarlo.

SYSTEM PROMPT

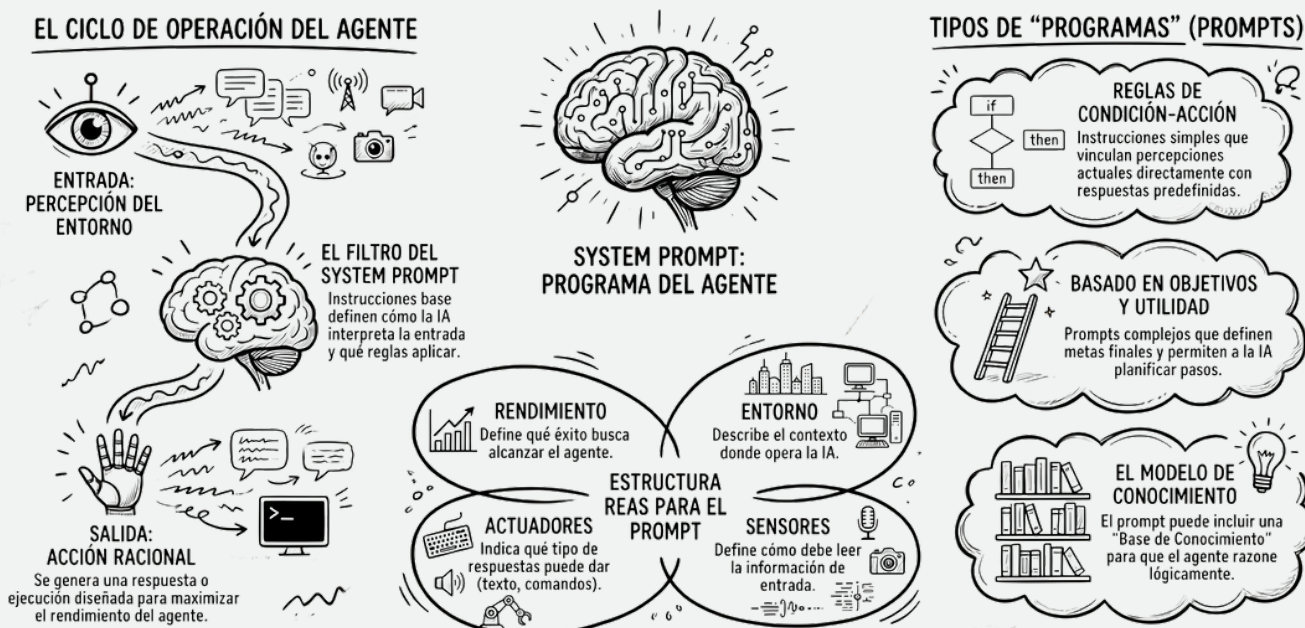
El system prompt es una instrucción que precede tu consulta y que define el rol, el alcance y las restricciones del modelo. Es la herramienta más subutilizada y más efectiva para mejorar la calidad de las respuestas.

- Define el rol: "Actúa como revisor metodológico experto en ciencias sociales cuantitativas."
- Establece límites: "Responde solo con información presente en los documentos que te proveo. Si no tienes certeza, indícalo explícitamente."
- Controla el formato: "Responde en español académico. No uses puntos de lista. Cita siempre la fuente."
- Establece el nivel de audiencia: "Asume que el lector es un investigador con doctorado en biología molecular."

Ejemplo práctico:

"Eres un asistente de revisión bibliográfica. Solo puedes afirmar información que esté respaldada por los artículos que el usuario te proporcione. Indica explícitamente cuando algo es una inferencia tuya y no un hecho del documento."

EL "SYSTEM PROMPT": EL ADN DE UN AGENTE INTELIGENTE



FINE-TUNING

El fine-tuning implica re-entrenar un modelo base con datos específicos de tu dominio. Es costoso en tiempo, recursos y dinero. Para la mayoría de investigadores, no es necesario.

- Cuándo sí aplica: terminología muy especializada que el modelo desconoce, tarea repetitiva y muy definida (ej. clasificar miles de abstracts en categorías fijas), acceso a grandes volúmenes de datos de entrenamiento propios.
- Cuándo no aplica: para mejorar respuestas ocasionales — el system prompt y el RAG son más eficientes y económicos.
- Alternativa práctica: few-shot prompting — proporcionar 3-5 ejemplos del output deseado dentro del mismo prompt.

Alternativa accesible

Few-shot prompting: incluye ejemplos del formato y contenido esperado directamente en tu prompt. Logra el 80% del resultado sin costo adicional.

No justificado

Fine-tuning para uso personal o de un único proyecto. El costo computacional y de mantenimiento raramente justifica el beneficio marginal.



RAG — RETRIEVAL-AUGMENTED GENERATION

RAG conecta al modelo con una base documental verificada antes de responder. En lugar de depender de su memoria de entrenamiento, el sistema busca fragmentos relevantes en tus propios documentos y los incluye en el contexto de la respuesta. Reduce drásticamente las alucinaciones.

- El modelo solo puede afirmar lo que está en tus PDFs — no inventa desde su entrenamiento.
- Ideal para trabajar con corpus de literatura, manuales institucionales, normativas o datos propios.
- Herramientas con RAG integrado: NotebookLM (Google), Elicit, Consensus, Perplexity (con fuentes).

Herramienta	Tipo de RAG	Mejor para
NotebookLM	Sobre tus PDFs subidos	Análisis de corpus propio
Elicit	Sobre literatura científica indexada	Extracción de datos de estudios
Consensus	Sobre artículos revisados por pares	Verificación de consenso científico
Perplexity	Sobre web + fuentes citadas	Consultas con trazabilidad

Recomendación:

Para trabajo científico serio, prioriza siempre herramientas con RAG sobre un LLM a secas. La diferencia en confiabilidad es sustancial.



CAPÍTULO 2

IA en el flujo de investigación



La IA puede ayudar en cada etapa: desde descubrir literatura hasta redactar.
Pero solo si la integras con criterio y sin delegar el pensamiento crítico.

Búsqueda y mapeo de literatura

Las herramientas semánticas van más allá de palabras clave: comprenden la intención de la consulta, mapean redes de citas e identifican trabajos fundacionales que podrían perderse en una búsqueda convencional.

Elicit



Extrae datos (muestra, método) de artículos automáticamente

Research Rabbit



Mapa visual de citas + alertas de nuevos trabajos, encuentra artículos con IA

ScySpace



Suite de herramientas de IA para investigadores

Scite.ai



Contexto de citas: apoya, menciona o contradice

Connected Papers



Grafo interactivo desde un artículo semilla

Consensus



Respuestas basadas en evidencia con citas directas

RECOMENDACIONES

- Empieza con Connected Papers o Research Rabbit para mapear la genealogía intelectual de tu tema desde un artículo clave.
- Usa Elicit para construir tablas comparativas de metodologías — ahorra semanas de lectura manual.
- Verifica el consenso con Scite antes de citar un trabajo como soporte de una afirmación.
- Complementa con búsquedas directas en SciELO / REDIB para no perder literatura en español.}

NINGUNA DE ESTAS HERRAMIENTAS INDEXA TODA LA LITERATURA EXISTENTE. SON PUNTOS DE ENTRADA, NO SUSTITUTOS DE LA BÚSQUEDA SISTEMÁTICA EN SCOPUS, WEB OF SCIENCE O PUBMED.

Organización del conocimiento

La IA es muy útil para procesar grandes volúmenes de documentos y conectar ideas dispersas, pero solo funciona bien cuando le provees el material de calidad.

NotebookLM (Google)



Sube tus PDFs y hazle preguntas directamente, responde solo con lo que está en tus documentos. RAG real y gratuito.

Zotero + extensiones



gestión de referencias integrada con síntesis automática de resúmenes.

Obsidian + plugins IA



construye un "segundo cerebro" vinculando notas mediante embeddings semánticos.

TAMBIÉN PUEDES PEGAR FRAGMENTOS DE TUS PROPIAS NOTAS EN CLAUDE O CHATGPT Y PÍDELE QUE LAS ESTRUCTURE, CONECTE O GENERE MAPAS CONCEPTUALES.

BUENA PRÁCTICA


- Usa NotebookLM como tu primer filtro después de leer un artículo. Sube el PDF y pídele un resumen estructurado, conceptos clave y citas relevantes. Genera una nota inicial en 5 minutos.
- Crea un sistema de tres niveles para tus archivos: Raw (PDFs originales) → Procesado (resúmenes generados con IA) → Síntesis (notas integradas por ti). Así evitas mezclar lo no leído con lo ya trabajado.
- Pega fragmentos de tus propias notas en Claude o ChatGPT y pídele que conecte ideas, genere mapas conceptuales o encuentre contradicciones. La IA ve patrones que tu memoria a corto plazo pierde..


NO DEBES SUBIR DOCUMENTOS CONFIDENCIALES, DATOS DE PARTICIPANTES O BORRADORES NO PUBLICADOS A SERVICIOS COMERCIALES SIN REVISAR SU POLÍTICA DE PRIVACIDAD Y USO DE DATOS.


Redacción académica con IA

La IA no escribe por ti, pero puede ayudarte a escribir mejor y más rápido. Úsala como copiloto: genera estructuras, mejora la claridad y rompe el bloqueo del escritor. Especialmente útil si publicas en un idioma que no es tu lengua nativa. El contenido intelectual, el análisis crítico y la voz autoral son tuyos e intransferibles.

BUENAS PRÁCTICAS

Genera primero el esquema (outline), no el texto final. Pide a ChatGPT, Claude o Gemini: "Propón una estructura para un artículo científico sobre [tema] con introducción, métodos, resultados y discusión." Tú llenas el contenido después. 

Usa herramientas especializadas en lenguaje académico. Trinkla, Writefull y Paperpal están entrenadas con millones de artículos científicos. Corrigen gramática, estilo y registro con mucha más precisión que un chatbot generalista. 

Pide reformulaciones para ganar claridad, no para crear contenido. Toma un párrafo denso tuyo y pídele a Claude o ChatGPT: "Reescribe este texto para que sea más claro y fluido, manteniendo todo el contenido técnico." Compara versiones y elige la mejor. 

ERROR FRECUENTE QUE DEBES EVITAR

Copiar directamente texto generado por IA sin revisarlo. La IA puede suavizar matices técnicos, eliminar precisión disciplinar o introducir frases vacías (el famoso "Es importante destacar que..."). Todo texto generado debe ser revisado línea por línea.

LA MAYORÍA DE REVISTAS DE ALTO IMPACTO (NATURE, LANCET, PLOS, ELSEVIER, SPRINGER) EXIGEN DECLARAR EL USO DE IA EN LA SECCIÓN DE MÉTODOS O EN NOTA AL PIE. NO HACERLO ES FALTA DE TRANSPARENCIA CIENTÍFICA.

CAPÍTULO 3

Límites éticos cuándo no usar la IA



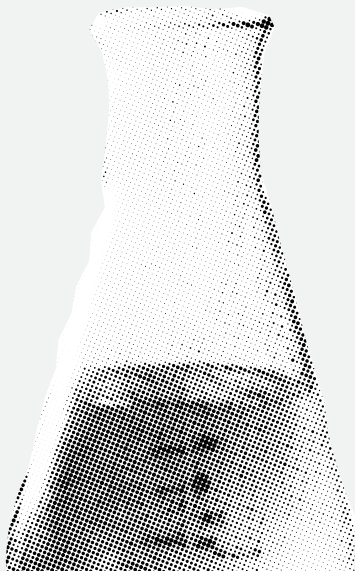
La IA puede hacer muchas cosas, pero no todo lo que puede hacer se debe hacer. Hay tareas que un investigador no puede delegar sin violar principios éticos básicos: confidencialidad, responsabilidad personal, protección de datos. Aquí tienes los límites claros que no debes cruzar.

Revisión por pares

La revisión por pares es el pilar de la calidad científica. Es un proceso confidencial, voluntario y de responsabilidad personal. Delegar en una IA la evaluación de un manuscrito ajeno no es solo técnicamente riesgoso: viola principios éticos fundamentales. Aquí te explicamos por qué y cuáles son los límites que no debes cruzar.

4 RAZONES PARA NO USAR IA EN LA REVISIÓN POR PARES

- Violación de la confidencialidad: El manuscrito que recibes para revisar no te pertenece. Al copiarlo y pegarlo en un chatbot público (ChatGPT, Claude, Gemini, etc.), estás compartiendo información confidencial con un tercero sin el consentimiento del autor ni del editor. Esto es una falta ética grave.
- El modelo no tiene criterio disciplinar profundo. Un LLM puede generar una evaluación que suene convincente, pero carece del conocimiento contextual, la experiencia metodológica y la sensibilidad al campo que sí tiene un revisor humano. Puede aprobar algo incorrecto o rechazar algo novedoso solo porque no se parece a lo que ha visto.
- La responsabilidad del juicio científico es intransferible. Si firmas una revisión, estás certificando con tu nombre que has evaluado el trabajo. Delegar ese juicio en una IA es una falsificación de la función de revisor. Las revistas y comités editoriales te eligieron a ti, no a un modelo.
- La mayoría de las revistas lo prohíben explícitamente. Elsevier, Springer Nature, Wiley, PLOS, y prácticamente todas las editoriales de alto impacto tienen políticas claras: no se permite introducir manuscritos confidenciales en herramientas de IA de terceros sin autorización expresa. Incumplirlo puede llevar a la inhabilitación como revisor.



EL 53% DE REVISORES YA USA IA PARA ASISTIR EVALUACIONES (2025), PERO USARLA PARA GENERAR LA REVISIÓN COMPLETA VIOLA LOS ESTÁNDARES ÉTICOS DE LA MAYORÍA DE REVISTAS.

Privacidad, datos y soberanía

La privacidad no es un detalle técnico menor: es un pilar de la integridad científica y un requisito legal en casi todos los países. Subir información confidencial a un servicio de IA puede exponer datos de participantes, hacerte perder prioridad sobre resultados inéditos o violar acuerdos con financiadores. Antes de copiar y pegar, hazte estas preguntas.

4 REGLAS DE ORO PARA PROTEGER TUS DATOS

- Nunca subas datos personales de participantes humanos. Nombres, direcciones, números de identificación, historias clínicas, respuestas de encuestas que permitan reidentificación... Todo esto está protegido por leyes. Subirlos a un chatbot público es una violación directa.
- Nunca subas borradores no publicados con resultados inéditos. Una vez que tu manuscrito sale de tu control, pierdes la capacidad de garantizar su confidencialidad. Algunos servicios comerciales pueden usar tu contenido para reentrenar sus modelos. El riesgo de perder prioridad científica o de ser "adelantado" es real.
- Nunca subas información cubierta por acuerdos de confidencialidad (NDA). Si tu investigación involucra una empresa, una agencia gubernamental o cualquier tercero con el que firmaste un NDA, compartir esa información con una IA externa rompe el acuerdo y puede tener consecuencias legales.
- Revisa siempre la política de privacidad del servicio antes de subir algo. No todos los servicios son iguales. Busca respuestas claras a: ¿Usan mis datos para reentrenar sus modelos? ¿Borran mis archivos después de procesarlos? ¿Están mis datos cifrados? ¿La empresa tiene sede en un país con leyes de protección de datos?

ALTERNATIVAS SEGURAS
USA MODELOS LOCALES (OLLAMA), ENTORNOS
INSTITUCIONALES CON DATOS PROTEGIDOS, O
NOTEBOOKLM CON CUENTA INSTITUCIONAL.



Seis principios del investigador crítico con IA

N.1 LA IA AMPLIFICA TU JUICIO, NO LO REEMPLAZA

El pensamiento crítico, la interpretación de resultados y la responsabilidad científica siguen siendo tuyos. La IA puede sugerir, estructurar, resumir... pero la decisión final, la conclusión y la firma del artículo son humanas e intransferibles.

Pregunta que debes hacerte: *¿Entiendo lo suficiente del tema como para evaluar críticamente lo que la IA me ha dado? Si la respuesta es no, empieza por formarte tú, no por delegar.*

N.2 VERIFICA TODA AFIRMACIÓN FACTUAL EN FUENTES PRIMARIAS

Especialmente cifras, citas bibliográficas, datos estadísticos, fechas y nombres. Los LLM alucinan con total seguridad. Una cita que parece real puede ser completamente inventada. Un porcentaje puede ser exacto... o no.

Pregunta que debes hacerte: *¿He comprobado cada referencia en PubMed, Scopus, Google Scholar o la fuente original? Si no lo has hecho, no confíes.*

N.3 DEFINE EL CONTEXTO ANTES DE PREGUNTAR

Un buen system prompt (instrucción previa) da mejores resultados que el modelo más potente sin instrucciones. Define el rol, el formato, las restricciones y el nivel de audiencia.

Pregunta que debes hacerte: *¿He indicado al modelo cómo quiero que actúe, qué límites tiene y cómo debe formatear la respuesta? Un minuto extra en instrucciones ahorra horas de corrección.*

N.4 PRIORIZA HERRAMIENTAS CON RAG PARA TRABAJAR CON LITERATURA

Un LLM sin anclaje documental puede inventar. Las herramientas que buscan en tus propios PDFs o en bases indexadas (RAG) reducen drásticamente las alucinaciones. Siempre que sea posible, trabaja sobre documentos que tú controlas.

Pregunta que debes hacerte: *¿La respuesta que obtuve está basada en los documentos que proporcioné o en la memoria general del modelo? Si es lo segundo, verifica el doble.*

N.5 LA PRIVACIDAD DE PARTICIPANTES Y COLEGAS ES NO NEGOCIABLE

No subas datos personales, resultados inéditos o manuscritos confidenciales a servicios de IA sin revisar sus políticas. Usa modelos locales (Ollama, LM Studio) o entornos institucionales protegidos cuando trabajes con información sensible.

Pregunta que debes hacerte: *¿Lo que voy a pegar aquí contiene algo que no debería salir de mi control? Si dudas, no lo subas.*

N.6 DECLARA EL USO DE IA EN TUS MANUSCRITOS

La transparencia científica incluye la transparencia metodológica. La mayoría de revistas de alto impacto exigen una declaración explícita si usaste IA en la redacción, el análisis o la generación de figuras. No hacerlo es una falta de integridad.

Pregunta que debes hacerte: *¿He incluido en el manuscrito una nota clara sobre qué herramientas de IA usé y para qué? Si no, agrégala antes de enviar.*

¿Necesitas ir más allá de esta guía?

Agenda una asesoría
personalizada conmigo sobre IA
para investigación

