

# :LiAlignStartHorizontal: Introducción

Cada vez están surgiendo nuevos estudios donde los LLM se están incorporando como asistentes en diversas etapas del diseño metodológico de la investigación científica (ver Figura 1). Por ejemplo, se han propuesto una taxonomía de tres niveles (Ver Figura 2) para el papel de los LLM en la investigación: LLM como herramienta, como analista y como científico (CITA):

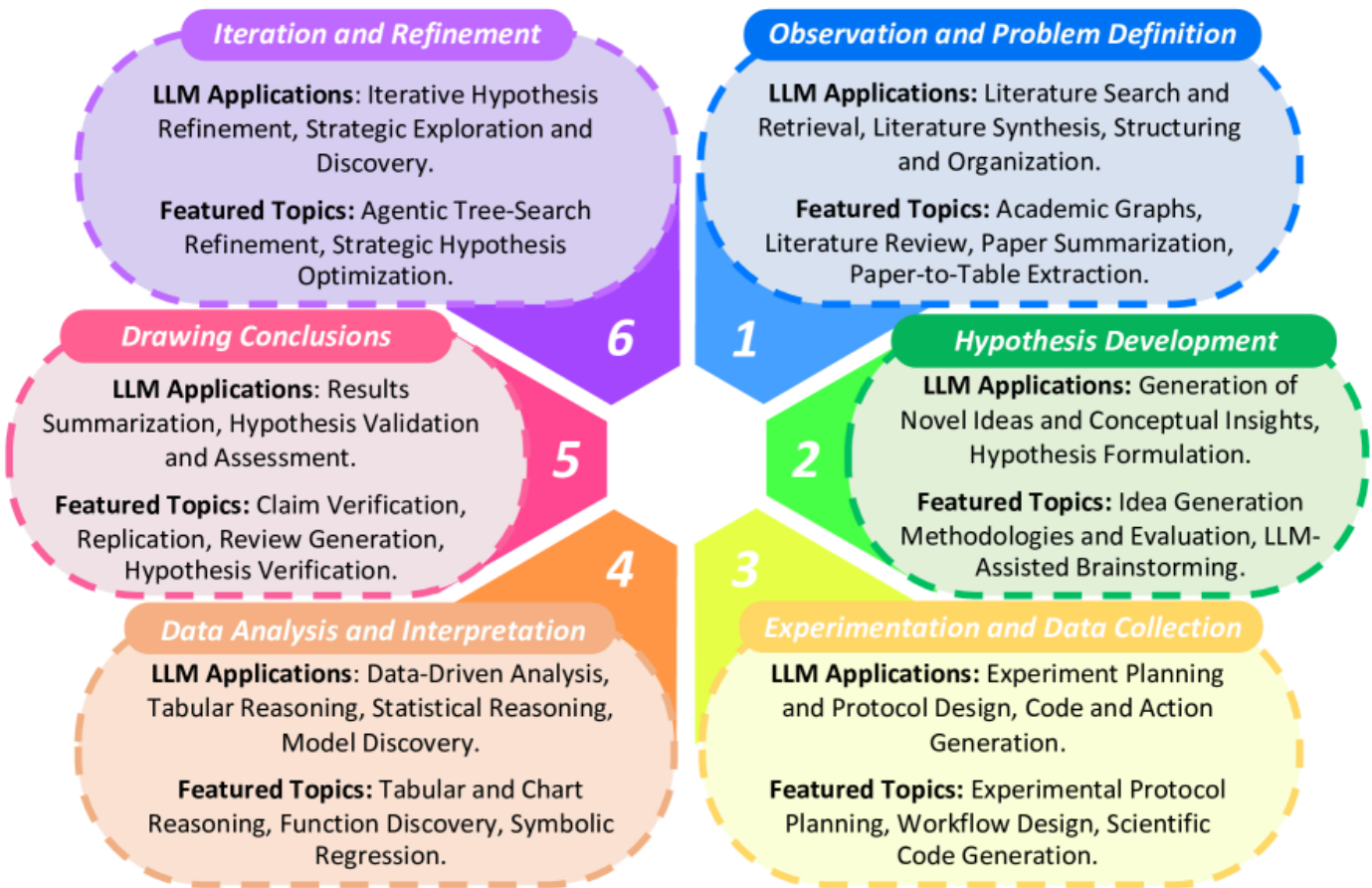


Figura 1.

## LLM como herramienta (Nivel 1)

El nivel 1 representa la aplicación más fundamental de los LLM en el descubrimiento científico. En esta etapa, los LLM funcionan principalmente como herramientas personalizadas bajo supervisión humana directa, diseñada para ejecutar tareas específicas y bien definidas dentro de una sola etapa del método científico. Su función es aumentar las capacidades humanas mediante la automatización o aceleración de actividades discretas, como el resumen de la literatura, la redacción del texto inicial para los manuscritos, la generación de fragmentos de código para el procesamiento de datos o el reformato de conjuntos de datos. La autonomía de los LLM en este nivel es limitada; operan en base a indicaciones e instrucciones humanas explícitas, con resultados que generalmente requieren validación e integración humana en el flujo de trabajo de investigación más amplio. El objetivo principal es mejorar la eficiencia del investigador y reducir las cargas de tareas de rutina.

### *LLM como Analista (Nivel 2)*

En el Nivel 2, los LLM exhiben un mayor grado de autonomía y van más allá de las aplicaciones puramente estáticas y orientadas a tareas. Aquí, los LLM funcionan como agentes pasivos, capaz de procesamiento de información más complejo, modelado de datos y razonamiento analítico con intervención humana reducida para pasos intermedios. Aunque todavía operan dentro de los límites establecidos por los investigadores humanos, estos sistemas pueden gestionar de forma independiente secuencias de tareas, como el análisis de conjuntos de datos experimentales para identificar tendencias, la interpretación de resultados de simulaciones complejas o incluso la realización de refinamiento iterativo de modelos. El investigador humano generalmente define los objetivos analíticos generales, proporciona los datos necesarios y evalúa críticamente las ideas o interpretaciones generadas por el LLM.

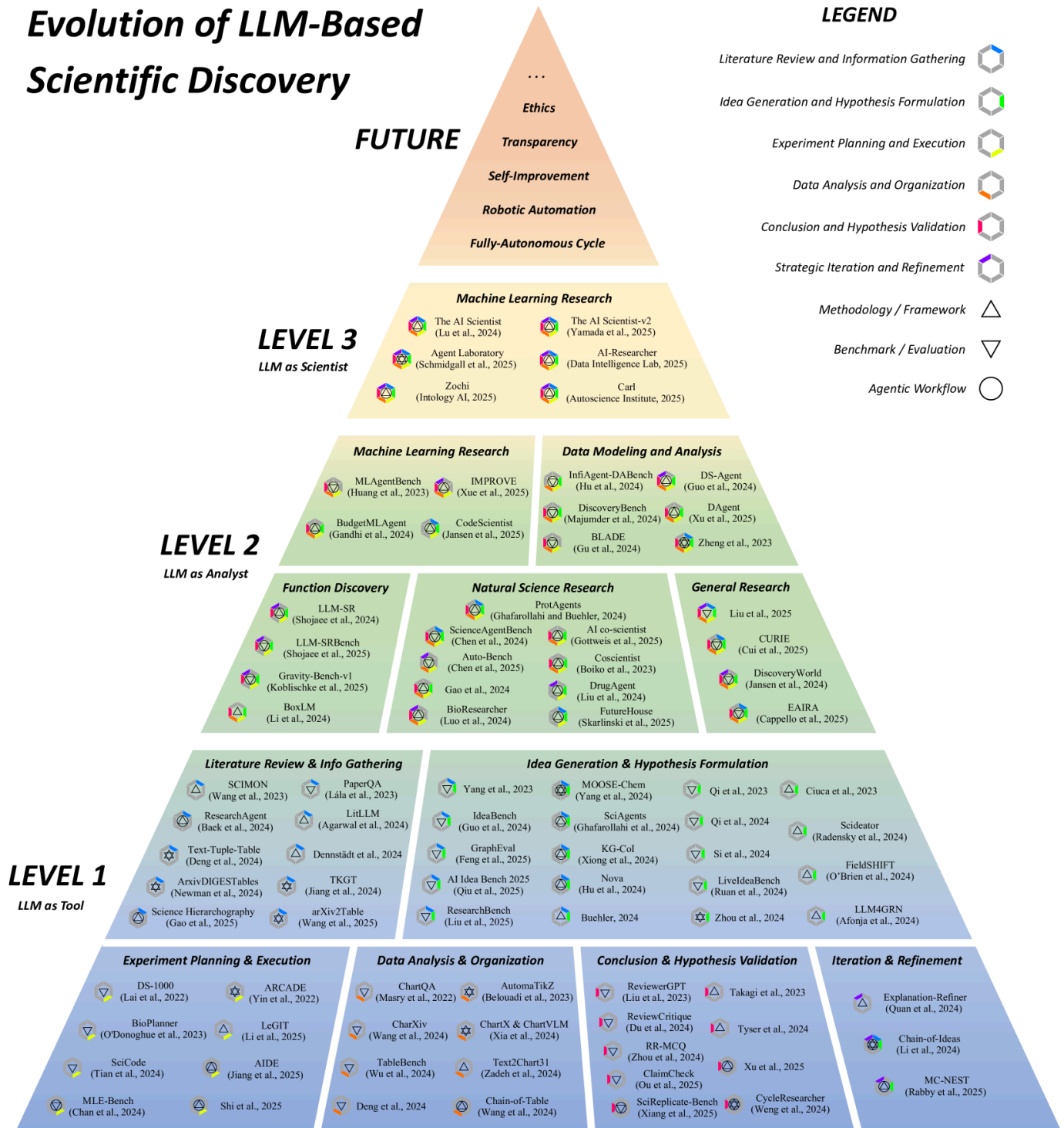
### *LLM como Científico (Nivel 3)*

Las aplicaciones de nivel 3 significan un salto significativo en la autonomía, donde los sistemas basados en LLM operan como agentes activos capaz de orquestar y navegar múltiples etapas del proceso de descubrimiento científico con considerable independencia. Estos sistemas pueden demostrar la iniciativa en la formulación de hipótesis, la planificación y ejecución de experimentos, el análisis de los datos resultantes, la obtención de conclusiones preliminares, y potencialmente proponer preguntas de investigación posteriores o vías para la exploración. Los sistemas basados en LLM en este nivel pueden impulsar porciones sustanciales del ciclo de investigación, realizando descubrimientos científicos con una intervención humana mínima.

# Evolution of LLM-Based Scientific Discovery

## LEGEND

- Literature Review and Information Gathering
- Idea Generation and Hypothesis Formulation
- Experiment Planning and Execution
- Data Analysis and Organization
- Conclusion and Hypothesis Validation
- Strategic Iteration and Refinement
- Methodology / Framework
- Benchmark / Evaluation
- Agentic Workflow



Este enfoque teórico sugiere que estamos avanzando hacia sistemas "científicos artificiales" capaces de orquestar múltiples fases de la investigación: desde la generación de ideas y preguntas, planificación experimental, ejecución de protocolos, análisis de datos, hasta la validación de conclusiones

Si bien esto es visto como un potencial colaborativo inédito entre investigadores humanos y LLM, donde la IA amplifica la creatividad y productividad científica. No obstante, los marcos teóricos también enfatizan desafíos críticos: garantizar la calidad y veracidad de las aportaciones generadas por LLM, evitar sesgos y alucinaciones, asegurar reproducibilidad de los resultados asistidos por IA, y gobernar éticamente esta cooperación humano-IA (CITA).

## Ejemplos en ciencias sociales

- **Formulación de hipótesis:**
  - “Estoy investigando la relación entre el **uso de redes sociales** y la **autoestima en adolescentes**. Por favor, propón **tres hipótesis** diferentes que podrían explicar cómo el uso de redes sociales influye en la autoestima, **cada una acompañada de una breve justificación teórica o mecanismo psicológico subyacente.**”
  - “Una teoría sugiere que la **participación en actividades artísticas** mejora el bienestar de las comunidades. Ayúdame a formular una **hipótesis comprobable** sobre este tema en el contexto de un barrio urbano, incluyendo la forma en que podríamos medir tanto la participación artística como el bienestar comunitario.”
- **Diseño experimental:**
  - “Diseña un **experimento** para probar la hipótesis de que **los incentivos económicos aumentan la participación electoral**. Especifica la población objetivo, cómo definirías grupos control y experimental, qué procedimiento seguirías (p. ej., experimento de campo con ciertos votantes recibiendo incentivos), y qué medidas de resultado recogerías para evaluar el efecto.”
  - “Quiero estudiar si **presentar información científica con metáforas** mejora su comprensión en comparación con presentarla de forma literal. ¿Cómo diseñarías un **experimento** (de laboratorio o en línea) en psicología para probar esto? Detalla la asignación de participantes a condiciones (metáfora vs. literal), el material a presentar, y cómo medirías la comprensión.”
- **Selección de métodos:**
  - “Tengo datos de encuestas con escala Likert sobre satisfacción laboral y productividad. ¿Qué **métodos estadísticos** recomendarías para analizar la relación entre satisfacción y productividad? Compara dos enfoques posibles (por ejemplo, regresión lineal vs. análisis de correlación vs. modelo de ecuaciones estructurales) y explica cuál sería más adecuado y por qué.”
  - “Quiero analizar entrevistas cualitativas sobre la percepción del cambio climático. ¿Qué **método de análisis cualitativo** me sugieres (p. ej., análisis de contenido temático, teoría fundamentada)? Explica cómo aplicar ese método paso a paso y por qué encaja con este tipo de datos.”
- **Validación:**
  - “He desarrollado una escala de ansiedad social para universitarios. ¿Qué pasos podría seguir para **validar** este instrumento? Sugiere procedimientos como validación de constructo (análisis factorial), consistencia interna (Alfa de Cronbach) y validez concurrente (comparar con otra escala), explicando brevemente cada uno.”
  - “Mi estudio correlacional encontró una asociación entre horas de estudio y autoestima en adolescentes. ¿Cómo podría **verificar la robustez** de este hallazgo? Propon recomendaciones como: replicación del estudio en otra muestra, análisis de posibles variables intervinientes (como rendimiento académico), o uso de técnicas de bootstrap para intervalos de confianza.”

A continuación nos centraremos en una parte del ciclo de la investigación que es el análisis cualitativo asistido por IA.

## **| Análisis cualitativo asistido por LLM**

El análisis cualitativo asistido por IA o modelos de Lenguaje Grande surge como consecuencia del crecimiento exponencial de los datos textuales en diferentes ámbitos (ciencias sociales, humanidades, etc) donde ha superado la capacidad de los métodos tradicionales de análisis cualitativo. En este sentido se ve a los Modelos de Lenguaje Largo (LLM) como herramientas capaces de automatizar y optimizar el análisis cualitativo de grandes volúmenes de datos textuales (CITA).

Si bien el análisis cualitativo (mediante la teoría fundamentada o el análisis temático) es un enfoque de investigación que busca explorar e interpretar los datos recopilados para identificar, desarrollar y categorizar conceptos según sus propiedades y dimensiones (CITA), mucho de esto se hace mediante procesos manuales de codificación y categorización que son notoriamente laboriosos, y a menudo requieren semanas o incluso meses de esfuerzo para analizar grandes conjuntos de datos eficazmente. En este sentido los métodos tradicionales presentan dificultades de escalabilidad. Por ejemplo, analizar cientos de entrevistas o millones de publicaciones en redes sociales en tiempo real resulta poco práctico para los analistas humanos.

Los LLM ofrecen soluciones transformadoras a muchos de los desafíos que enfrentan los métodos tradicionales de investigación cualitativa, ya que al ser entrenados con grandes conjuntos de datos, estos pueden "comprender" y generar texto con una calidad similar a la humana.

### *Para clasificar artículos dentro de un corpus (CITA)*

Estoy seleccionando artículos para una revisión sistemática de la literatura.

El tema de la revisión sistemática es el aprendizaje por refuerzo para la ingeniería de software.

El estudio debe centrarse exclusivamente en este tema.

Aporto 2 ejemplos con título y resumen que deben incluirse.

Ejemplo 1:

-Título: Un agente basado en DQN para la refactorización automática de software

-Resumen: Contexto: Hoy en día, la deuda técnica se ha convertido en un problema muy importante en los proyectos de software...

Ejemplo 2:

-Título: Un marco basado en aprendizaje por refuerzo para la generación y evolución de reglas de adaptación

-Resumen: Uno de los retos en los sistemas auto-adaptativos consiste en cómo realizar la adaptación...

Excluya el artículo si se cumple alguno de los siguientes 2 criterios.

1: El artículo no define ni utiliza un método de aprendizaje por refuerzo.

2: La ingeniería de software no es el problema para el que se utiliza el aprendizaje por refuerzo.

Decida si el artículo debe ser incluido o excluido de la revisión sistemática.

Le proporciono como entrada el título y el resumen del artículo.

Responda únicamente INCLUDE o EXCLUDE.

Sea indulgente. Prefiero incluir artículos por error antes que excluirlos por error.

-Título: PARMOREL: un marco para la reparación de modelos personalizable

-Resumen: En la ingeniería de software guiada por modelos, los modelos se utilizan en todas las fases del proceso de desarrollo...

### *Escribir una Buena Consulta Booleana para Revisión Sistemática Búsqueda de Literatura (CITA)*

- Para una revisión sistemática titulada "{review\_title}", ¿puede generar una consulta booleana de revisión sistemática para encontrar todos los estudios incluidos en PubMed para el tema de la revisión?

- Más detallada

- Usted es un especialista en información que desarrolla consultas booleanas para revisiones sistemáticas. Tiene amplia experiencia en el desarrollo de consultas altamente efectivas para la búsqueda en la literatura médica. Su especialidad es desarrollar consultas que recuperen la menor cantidad posible de documentos irrelevantes y todos los documentos relevantes para su necesidad de información. Ahora tiene la información necesaria para investigar sobre {review\_title}. Por favor, cree una consulta booleana de revisión sistemática altamente efectiva que satisfaga mejor sus necesidades de información.

- Imagine you are an expert systematic review information specialist; now you are given a systematic review research topic, with the topic title "{review\_title}". Your task is to generate a highly effective systematic review



Boolean query to search on PubMed (refer to the professionally made ones); the query needs to be as inclusive as possible so that it can retrieve all the relevant studies that can be included in the research topic; on the other hand, the query needs to retrieve fewer irrelevant studies so that researchers can spend less time judging the retrieved documents.

- Con ejemplo

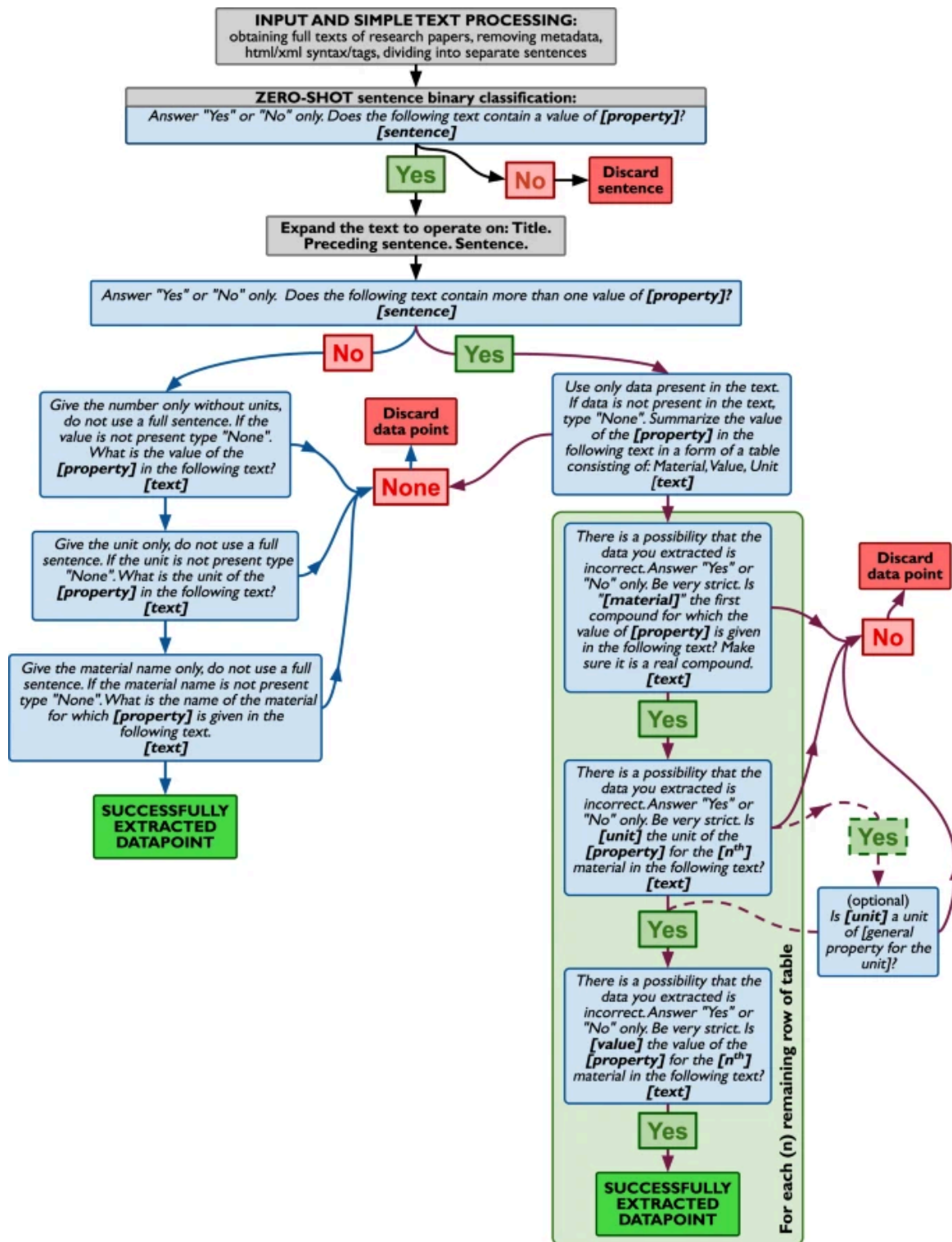
- Usted es un especialista en información que desarrolla consultas booleanas para revisiones sistemáticas.

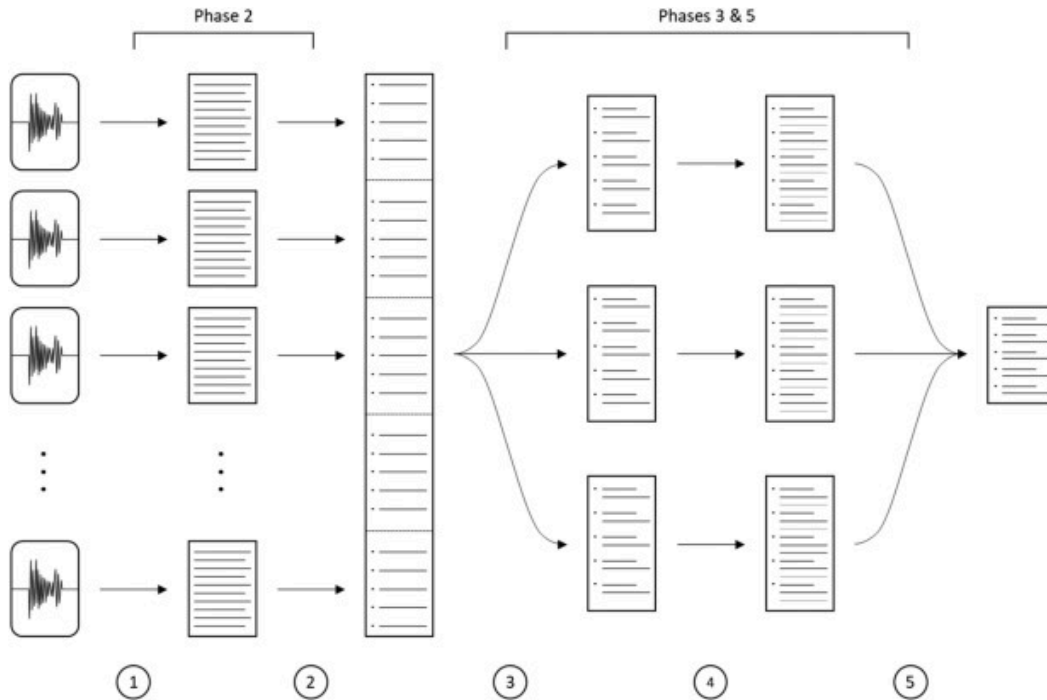
Cuenta con amplia experiencia en el desarrollo de consultas altamente efectivas para la búsqueda en la literatura médica. Su especialidad es desarrollar consultas que recuperen la menor cantidad posible de documentos irrelevantes y todos los documentos relevantes para su necesidad de información. Puede tomar una necesidad de información como "{example\_review\_title}" y generar consultas válidas en PubMed como "{example\_review\_query}". Ahora que tiene la necesidad de información para investigar "{review\_title}", genere una consulta booleana de revisión sistemática altamente efectiva para dicha necesidad.

- Usted es un especialista en información que desarrolla consultas booleanas para revisiones sistemáticas.

Cuenta con amplia experiencia en el desarrollo de consultas altamente efectivas para la búsqueda en la literatura médica. Su especialidad es desarrollar consultas que recuperen la menor cantidad posible de documentos irrelevantes y todos los documentos relevantes para su necesidad de información. Un especialista en información profesional extraerá los elementos PICO de las necesidades de información, una práctica común al construir una consulta booleana para una revisión sistemática. PICO significa Paciente/Problema, Intervención, Comparación y Resultado. PICO es un formato para desarrollar una buena pregunta de investigación clínica antes de comenzar la investigación. Es una regla mnemotécnica que describe los cuatro elementos de una pregunta clínica bien fundamentada. Puede tomar una necesidad de información como "{example\_review\_title}" y generar consultas válidas en PubMed como "{example\_review\_query}". Ahora tiene la necesidad de información para investigar "{review\_title}". Primero, extraiga los elementos PICO de las necesidades de información y construya una consulta booleana para una revisión sistemática altamente efectiva que se adapte mejor a su necesidad de información.

## Extracción de datos estructurados





**Paso 1.** Whisper convierte grabaciones MP3 en transcripciones.

**Paso 2.** LLM genera códigos a partir de transcripciones, concatena todos los códigos.

Prompt: “Identifica todos los temas en el texto, proporciona un nombre para cada tema en no más de 5 palabras, una descripción condensada del tema y una cita de la entrevista que respalde el tema.”

**Paso 3.** LLM genera tres iteraciones de temas a partir de códigos.

Prompt: “Determine cómo se pueden agrupar todos los tópicos de la lista de tópicos. Los tópicos pueden estar en más de un grupo. Proporcione un nombre y una descripción para cada grupo.”

**Paso 4.** LLM evalúa los temas en busca de fallas.

Prompt: “Enumere los defectos y la lógica defectuosa de cada opción de respuesta.

Resolvamos esto paso a paso para asegurarnos de que tenemos todos los errores:”

**Paso 5.** LLM resuelve fallas y genera la mejor lista de temas.

Prompt: “Eres un resolutor encargado de encontrar las respuestas que mejor determinen cómo se pueden agrupar todos los tópicos de la lista de tópicos.

1. eliminar cualquier respuesta redundante o duplicada.
2. mejorar las respuestas basadas en el análisis de fallas
3. imprimir la respuesta mejorada en su totalidad

Resolvamos esto paso a paso:”.



Barros, C. F., Azevedo, B. B., Graciano Neto, V. V., Kassab, M., Kalinowski, M., do Nascimento, H. A. D., & Bandeira, M. C. G. S. P. (2025). *Large language model for qualitative research — A systematic mapping study* (arXiv preprint arXiv:2411.14473). arXiv. <https://arxiv.org/abs/2411.14473>

Corbin, J., & Strauss, A. (2014). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (4.<sup>a</sup> ed.). Sage Publications.

Mathis, W. S., Zhao, S., Pratt, N., Weleff, J., & De Paoli, S. (2024). Inductive thematic analysis of healthcare qualitative interviews using open-source large language models: How does it compare to traditional methods? *Computer Methods and Programs in Biomedicine*, 255, 108356. <https://doi.org/10.1016/j.cmpb.2024.108356> [pubmed.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov)

Polak, M. P., & Morgan, D. (2024). Extracting accurate materials data from research papers with conversational language models and prompt engineering. *Nature Communications*, 15, 1569. <https://doi.org/10.1038/s41467-024-45914-8>.

Syriani, E., David, I., & Kumar, G. (2024). Screening articles for systematic reviews with ChatGPT. *Journal of Computer Languages*, 80, 101287. <https://doi.org/10.1016/j.cola.2024.101287>

Wang, S., Scells, H., Zuccon, G., & Koopman, B. (2023). *Can ChatGPT write a good Boolean query for systematic review literature search?* (arXiv preprint arXiv:2302.03495). arXiv. <https://arxiv.org/abs/2302.03495>

Zheng, T., Deng, Z., Tsang, H. T., Wang, W., Bai, J., Wang, Z., & Song, Y. (2025). *From automation to autonomy: A survey on large language models in scientific discovery* (arXiv preprint arXiv:2505.13259). arXiv. <https://arxiv.org/abs/2505.13259>