

Pensamiento crítico e inteligencia artificial en la academia: procedimiento de análisis matricial cualitativo para evaluar sistemas de IA

Lluís Codina

Universitat Pompeu Fabra, España

<https://orcid.org/0000-0001-7020-1631>

Elisenda Aguilera-Cora

Universitat Pompeu Fabra, España

<https://orcid.org/0000-0003-0923-9192>

Carlos Lopezosa

Universitat de Barcelona, España

<https://orcid.org/0000-0001-8619-2194>

Pere Freixa

Universitat Pompeu Fabra, España

<https://orcid.org/0000-0002-9199-1270>

Codina, L., Aguilera-Cora, E., Lopezosa, C., & Freixa, P. (2025). Pensamiento crítico e inteligencia artificial en la academia: procedimiento de análisis matricial cualitativo para evaluar sistemas de IA. En J. Guallar, M. Váñez, & A. Ventura-Cisquella (Coords). *Comunicación digital. Tendencias y buenas prácticas* (pp. 171-183). Ediciones Profesionales de la Información. <https://doi.org/10.3145/cuviceom.12.esp>

Resumen

Este capítulo presenta el Procedimiento Matricial de Análisis de Sistemas de IA (MASIA), un método cualitativo basado en matrices, diseñado para evaluar el rendimiento y la calidad de los sistemas de inteligencia artificial generativa en entornos académicos. MASIA se centra en el análisis de tres componentes clave en las respuestas generadas por IA: síntesis narrativa, uso de fuentes y formulación de nuevos prompts. De esta forma, fomenta el pensamiento crítico en los usuarios de la IA y ofrece valiosas herramientas tanto para la docencia como para la investigación. El procedimiento define variables y parámetros analíticos que permiten la comparación de diferentes sistemas de IA, lo que facilita la toma de decisiones informada en entornos académicos y de investigación. Además, MASIA integra consideraciones éticas, como la trazabilidad, la atribución correcta y la prevención del plagio, lo que lo convierte en un instrumento flexible y adaptable a diversas necesidades y proyectos académicos. El capítulo concluye que MASIA es una herramienta sencilla pero potente para potenciar el pensamiento crítico, optimizar los procesos de enseñanza y aprendizaje y sentar las bases para la investigación comparativa sobre inteligencia artificial en el ámbito académico.

Palabras clave

Inteligencia artificial generativa; Evaluación cualitativa; Pensamiento crítico; Matrices de análisis; Ética académica; Sistemas de IA en el ámbito académico; Métodos de evaluación.

1. Introducción

El objetivo de este trabajo es presentar un procedimiento de análisis para evaluar el rendimiento y la calidad de los sistemas de inteligencia artificial generativa en entornos académicos.

El procedimiento, al que denominamos *Procedimiento Matricial de Análisis de Sistemas de IA* o MASIA, está concebido para evaluar sistemas de IA que, como parte de su respuesta no solamente proporcionan una síntesis narrativa, sino que también incluyen citas y las fuentes bibliográficas que han usado para generar el contenido.

El método de análisis favorece el pensamiento crítico de los usuarios de sistemas de IA, aporta elementos para procesos de enseñanza-aprendizaje y puede ser la base para desarrollar la toma de datos en procesos de investigación.

La utilización de fuentes como parte de la respuesta es una necesidad en el contexto académico porque permite verificar y ampliar la información proporcionada por la IA, así como mantener la cadena de atribuciones (High-Level Expert Group on Artificial Intelligence, 2019; Crompton y Burke, 2023; Kaebnick et al., 2023; Lund et al., 2023; Tilie et al., 2023; Gundersen et al., 2018; World Commission on the Ethics of Scientific Knowledge and Technology, 2019; Dwivedi et al., 2021; Bianchini et al., 2022). Esto último es doblemente conveniente, porque además de incrementar la calidad de la respuesta de las IA, previene el plagio o la inadecuada atribución, siendo ambos aspectos esenciales en trabajos académicos.

La base del procedimiento presentado aquí consiste en matrices de análisis basadas a su vez en esquemas formados por una serie de variables (Codina y Pedraza, 2016). La determinación de las variables procede de la actividad docente y de experiencias previas de los autores en el análisis y la utilización de sistemas de IA para trabajos académicos, muy especialmente, por la necesidad de aportar protocolos del uso crítico de sistemas de IA a estudiantes universitarios e investigadores predoctorales.

Desde la aparición de ChatGPT a finales de 2022, los autores han incorporado el uso de la inteligencia artificial en sus actividades docentes e investigadoras (Lopezosa y Codina, 2023; Lopezosa et al., 2023a; 2023b; Aguilera-Cora et al., 2024a; 2024b; Codina, 2025). Esta integración evidenció la necesidad de contar con una herramienta intelectual que permitiera formar a estudiantes e investigadores predoctorales en el uso adecuado de la IA (Codina y Garde, 2023). También, la necesidad de disponer de un instrumento que permitiera llevar a cabo estudios comparativos de la eficiencia de los sistemas de inteligencia artificial susceptibles de uso en la academia. Esto último, puede ser útil a efectos de investigación, o para que decisores económicos de las universidades, por ejemplo, puedan tener elementos de evaluación para adquirir sistemas de IA (Bhatia, 2023; Whitfield & Hofmann, 2024; Elsevier 2024).

Antes de presentar los componentes del método de evaluación debemos presentar unas precisiones terminológicas, cuya consideración forma parte del procedimiento en sí, así como deberemos considerar cuál es la composición de los resultados de una IA cuando responde a una instrucción del usuario.

2. Terminología

La terminología que se presenta a continuación se considera parte del procedimiento, por lo que resulta necesario establecer con precisión el uso de un conjunto de términos para su adecuada aplicación. Los presentamos en la tabla 1.

Tabla 1
Terminología del procedimiento de evaluación de sistemas de IA.

Término	Explicación
Fuentes bibliográficas	En un sistema de IA de tipo RAG (ver la definición más abajo) es la lista de documentos (artículos de revista, informes, páginas web, etc.) que justifican la respuesta. En el contexto del procedimiento de evaluación, este concepto de fuente es el que se tiene en cuenta si no se señala lo contrario.
Fuentes de información	En un sistema de IA de tipo RAG (ver la definición más abajo) las fuentes de información son los recursos que utiliza para localizar las fuentes en las que se fundamenta su respuesta. Fuentes de información características de las IA tipo RAG pueden ser bases de datos académicas o buscadores como Google.
Indicadores	En un procedimiento de evaluación los indicadores son las características que aportan información sobre aquello que quiere ser evaluado o comparado. P.e., en el análisis comparativo de la economía de las naciones, el índice de desempleo, la inflación o el PIB son indicadores. Por su propia naturaleza también son variables.
Matrices	Estructuras de información basadas en tablas que permiten extraer datos, presentarlos de forma visual y hacer análisis comparativos. En el procedimiento presentado aquí el uso de matrices se considera normativo. La palabra tabla es equivalente. El uso del término matriz introduce la idea de que se trata de una tabla que cumple ciertas condiciones, las más importantes son que se trata de tablas homogéneas y que se organizan de forma que las filas son entidades y las columnas son las propiedades de las entidades.

Término	Explicación
Modelo de IA	También llamado <i>Large Language Model</i> (LLM). Es la denominación técnica de las inteligencias artificiales de tipo generativo por tratarse de su base tecnológica. Un modelo de IA o una LLM no es utilizable por un usuario final, ya que su uso requiere programación y APIs, por lo que los usuarios finales en general trabajan con IA a través de sistemas de IA.
Página de resultados	La respuesta de un sistema de IA se presenta al usuario en una página de resultados que suele tener al menos tres componentes: la síntesis narrativa, una lista de fuentes y un conjunto de nuevos <i>prompts</i> sugeridos.
Parámetros	En un procedimiento de evaluación los parámetros agrupan indicadores o variables. La idea es que un grupo de variables sirve para caracterizar un aspecto significativo de cierta complejidad de aquello que se pretende evaluar. Por ejemplo, un parámetro de las naciones en análisis comparativos es su economía, otro su demografía o su sistema político, etc. Pero para caracterizar a cada parámetro es necesario utilizar indicadores desagregados como el índice de desempleo en el caso de la economía, junto con otros como el PIB, etc.
Prompt	Instrucción en lenguaje natural que utiliza el usuario para obtener la respuesta de una IA.
Prompts sugeridos	Lista de nuevos <i>prompts</i> que algunos sistemas de IA proporcionan como parte de la respuesta a un <i>prompt</i> .
Rapid review	Una modalidad de revisión en la que se omiten algunos de los controles habituales de las <i>systematic reviews</i> para obtener resultados inmediatos con valor preliminar. Las síntesis narrativas de una IA pueden equipararse a una forma de <i>rapid review</i> .
Retrieval Augmented Generation (RAG)	La generación aumentada de recuperación (RAG) consiste en mejorar las respuestas generadas por sistemas de IA al combinar su conocimiento de su base de entrenamiento con información recuperada en tiempo real de fuentes externas tales como bases de datos académicas o especializadas o de buscadores generalistas como Google o Bing. La mayor parte de los sistemas de IA especializados en uso académico son de tipo RAG. También algunas IA generalistas como Perplexity. Y se supone que debería serlo Google cuando integre de forma efectiva su buscador con su IA.
Revisión de la literatura	La revisión de la literatura es un proceso sistemático de búsqueda, selección, análisis y síntesis de la información existente sobre un tema específico. Implica evaluar críticamente los estudios previos, identificar patrones, debates y vacíos en el conocimiento actual, y presentar una visión integral y organizada del estado del arte en ese campo. En sistemas de IA aplicados a la academia, la producción de revisiones de la literatura es una de sus funciones principales.
Scratchpad / Bloc de notas	El <i>scratchpad</i> o bloc de notas es la sección que precede a la síntesis narrativa donde la IA presenta la cadena de razonamientos que ha seguido para llevar a cabo sus tareas, en nuestro caso, la tarea consiste en resolver las cuatro fases que conducen a una síntesis narrativa. La posibilidad de examinar el <i>scratchpad</i> implica la posibilidad, entre otras cosas, de comprobar si la IA ha comprendido los objetivos de la tarea y ha acertado en su enfoque.
Síntesis narrativa	Una síntesis narrativa es el resultado de analizar un conjunto de fuentes utilizando un esquema bien definido y de reunir en un resumen textual coherente los conocimientos principales derivados del análisis de estos documentos. La síntesis narrativa es tanto un producto derivado de una revisión de la literatura como parte de la respuesta de una IA generativa.
Sistema de IA	Se compone de uno o más modelos de IA (también llamados LLM por <i>Large Language Model</i>), de una o varias capas de software, p.e., para consultar bases de datos (p.e., bases de datos académicas), para gestionar referencias, etc., así como de una interfaz de usuario para facilitar el uso de todo el sistema.
Utilidades	Consisten en funciones complementarias que suelen ofrecer los programas de software además de sus funciones principales. Por ejemplo, en un sistema de IA enfocado a la academia, una utilidad puede consistir en recursos para gestionar referencias.
Variables	Una variable es una propiedad de alguna entidad que puede adoptar valores diferentes para cada entidad, o a lo largo del tiempo en la misma entidad. Registrar estos valores permite analizar, caracterizar y comparar entidades. También se pueden denominar indicadores si se desea enfatizar su carácter heurístico.

Fuente: elaboración propia

3. Composición de la página de resultados de un sistema de IA

En una página de resultados de una IA del tipo que estamos considerando analizar podemos determinar la existencia de tres componentes principales:

1. Síntesis narrativa.
2. Fuentes consultadas.
3. Nuevos *prompts*.

En lo que sigue, presentaremos cada uno de estos componentes a efectos de evaluación. Pero antes debemos señalar un cuarto elemento que, aunque no forma parte de la página de resultados, podemos encontrar formando parte de la misma interfaz:

4. Utilidades y funciones adicionales y específicas de cada sistema.

3.1. Síntesis narrativa

- **Trazables mediante un *scratchpad* visible.** El *scratchpad* o bloc de notas es una sección que precede a la síntesis narrativa en la cual la IA presenta de modo transparente y trazable la cadena de razonamientos que ha seguido para resolver la tarea. La posibilidad de examinar esta cadena de razonamientos permite detectar posibles sesgo u otra clase de errores, así como permite verificar si la IA ha entendido bien la tarea, y en todo caso facilita la trazabilidad del proceso seguido.
- **Articuladas.** Significa que la síntesis se presenta mediante alguna clase de estructura. Por ejemplo, en apartados separados y posiblemente organizados por títulos y siguiendo alguna clase de lógica en la distribución de secciones.
- **Presentan coherencia y cohesión.** La coherencia consiste en la interrelación de las oraciones que forman el texto mediante su adecuada vinculación con el tema principal del texto. Se manifiesta por la unidad temática, la ausencia (relativa) de redundancia y la progresión lógica de las ideas. La cohesión, por su parte, se manifiesta en la interrelación de las oraciones por la gramática. Se determina fundamentalmente por los conectores.
- **Presentan conectividad.** El final de cada párrafo anticipa el siguiente, y el inicio de los párrafos siguientes conectan con los anteriores esta se propiedad se hace evidente mediante el uso de conectores. La conectividad se incrementa si hay una sección que reunifica las ideas principales, o algún apartado con una función equivalente.
- **Son (relativamente) extensas.** A igualdad de los demás criterios son preferibles síntesis extensas. Ya que hablamos de un rango que puede ir de las 200 palabras a las 3.000 palabras, son preferibles las que, en caso necesario, se pueden acercar más a este último límite.
- **Son multimodales.** Además de texto incluyen algún formato adicional, p.e., tablas, fichas, mapas conceptuales, mapas mentales o diagramas.

3.2. Fuentes

En las AI de tipo académico, las fuentes suelen ser documentos, informes y artículos de revistas científicas. En definitiva, las fuentes son las que permiten atribuir las ideas y contenidos componentes de la síntesis a sus creadores originales. En el contexto académico tenemos el imperativo categórico de usar sistemas de IA que junto a la síntesis narrativa generada aporten las fuentes en las que se han basado. Este es el motivo de preferir sistemas de IA de tipo RAG.

El motivo de esta preferencia es que la ausencia de fuentes en la respuesta promovería la ruptura de la cadena de atribuciones y en consecuencias predispondría al plagio. Nótese que estamos separando la realización (o no) de plagio del hecho de que algunos sistemas de IA faciliten o promuevan, de facto, el plagio al ofrecer respuestas sin fuentes. Lógicamente debemos preferir sistemas de IA que, al menos, no promuevan el plagio.

A efectos de nuestra propuesta de análisis, son preferibles las IA que en relación con las fuentes:

- Presentan **capilaridad**. Son preferibles las AI que asignan las fuentes a nivel de frases, y en su defecto a nivel de párrafos, mejor que como una lista final que afecta a la totalidad indiferenciada de la síntesis narrativa. La capilaridad implica también conectividad, ya que cuando en un párrafo con (p.e.) tres ideas relacionadas cada fuente está vinculada a cada una de las ideas, en lugar de situar las tres fuentes al final del párrafo o al final de la totalidad de la síntesis.
- Proporcionan **formatos** de citación bien formados. Esto es, con datos referenciales completos y en su caso con enlaces viables.

Además, el usuario de una IA está obligado a verificar y revisar las fuentes, no solo para evaluar los argumentos, sino también para poder atribuir ideas y contenidos de terceros, a sus verdaderos autores mediante el sistema convencional de citación.

3.3. Nuevos *prompts*

Algunas IA ofrecen, como un tercer componente destacado de sus respuestas, una lista de nuevos *prompts* o de nuevas preguntas. Estas propuestas pueden tener escaso interés o pueden ser muy incisivas. En este último caso tienen un evidente valor heurístico. Por tanto, son preferibles las IA que generan, como parte de la página de resultados, *prompts* adicionales, de los que evaluaremos:

- **Oportunidad**. Esto es, ¿los nuevos *prompts* sugeridos son adecuados a los objetivos de la búsqueda?
- **Variedad de enfoques**. Con esta variable nos referimos a la pregunta: ¿ofrecen nuevas facetas o enfoques no considerados en el *prompt* original?

3.4. Utilidades e idiofunciones

Algunos sistemas de IA presentan una o más características que son específicas y únicas del sistema considerado. Frente a las funciones comunes podemos hablar de *idiofunciones* esto es, funciones únicas de cada sistema particular y por tanto únicamente presentes en el sistema considerado. Por ejemplo, una IA puede presentar una funcionalidad consistente en extraer conceptos u otra consistente en poder diseñar matrices de análisis de las referencias.

Se trata de *idiofunciones* porque son propias y únicas de cada IA. Aunque es posible que con el tiempo estas funciones se vayan estandarizando (y el concepto pierda sentido), actualmente son significativas estas diferencias y resulta útil considerarlas.

3.5. Matrices de análisis

Con ayuda de los conceptos anteriores, ya podemos presentar los elementos de análisis, que articulamos en parámetros y variables según muestra la tabla 2.

Tabla 2
Variables de análisis.

Parámetro	Código	Variables /Pregunta de chequeo
1. Síntesis narrativa	1.1	Scrathpad ¿Presenta un <i>scratchpad</i> con la cadena de razonamientos seguidos por la IA? ¿Este documento, se puede seguir consultando una vez completada la tarea?
	1.2	Articulación ¿La síntesis narrativa se presenta organizada o articulada en diversos apartados o se presenta como un continuo sin estructura definida?
	1.3	Coherencia y conexión ¿Se conserva la una unidad temática global en el conjunto de la síntesis narrativa y en el seno de cada párrafo? ¿Hay conexión entre los apartados, secciones o párrafos de la síntesis narrativa?
	1.4	Extensión ¿Es una síntesis de extensión adecuada a los objetivos? ¿Cuántas palabras forman la síntesis narrativa? ¿Ofrece versiones alternativas en relación con la extensión?
	1.5	Multimodalidad ¿La página de resultados incluye solo texto o incluye otras morfologías de la información, como p.e., diagramas? ¿En caso de no ofrecerlas al inicio, las propone como alternativa?
2. Fuentes	2.1	Número ¿Cuántas fuentes cita en total?
	2.2	Diversidad ¿Las fuentes exhiben diversidad adecuada a los objetivos del <i>prompt</i> ? Nota: una misma base de datos no es una limitación <i>a priori</i> de la diversidad. ¿El sistema permite diferenciar si las fuentes pertenecen a textos académicos, de prensa, literatura gris u otras fuentes no regladas?
	2.3	Capilaridad ¿Las fuentes están conectadas al menos al nivel de los párrafos o secciones?
	2.4	Bien formadas ¿Las fuentes están presentadas en un formato que resulta fácil para exportar, gestionar y citar las fuentes?
3. Prompts sugeridos	3.1	Oportunidad ¿Los nuevos <i>prompts</i> sugeridos parecen adecuados u oportunos dada la necesidad de información?
	3.2	Variedad ¿Los <i>prompts</i> son variados y ayudan a ampliar el foco del tema?
4. Idiofunciones	4.1	Funciones propias y exclusivas de cada sistema considerado Además de las funciones comunes examinadas, ¿el sistema presenta otras funciones propias?

Tabla 3

Puntuaciones teóricas.

Parámetro	Código	Variables	Puntuación teórica
Síntesis narrativa	1.1	Scratchpad	0-3
	1.2	Articulación	
	1.3	Conexión	
	1.4	Extensión	
	1.5	Multimodalidad	
Fuentes	2.1	Número	
	2.2	Diversidad	
	2.3	Capilaridad	
	2.4	Bien formadas	
Prompts sugeridos	3.1	Oportunidad	
	3.2	Variedad	
Idiofunciones	4.1	Funciones propias y exclusivas de cada sistema considerado	0-3

La escala de puntuación se ofrece a modo de ejemplo. Para cada uso, los responsables pueden determinar (de forma justificada) otras escalas. En este caso, hemos usado una escala característica de evaluaciones heurísticas en el ámbito de la usabilidad de sistemas de información, y corresponde a la siguiente estimación:

Puntuación	Interpretación
0	Ausencia de función o variable considerada
1	La función o variable aparece en una expresión mínima
2	La función o variable está correctamente implementada pero admite mejoras
3	La función o variable está plenamente implementada

En esta clase de procedimientos de puntuación, las primeras puntuaciones se asignan de modo intuitivo, y se van ajustando a medida que el examen de nuevos casos permite hacer comparaciones. Las puntuaciones se terminan de ajustar cuando se han examinado todos los casos. También es habitual que haya dos analistas encargados de asignar las puntuaciones de modo independiente, luego se comparan las puntuaciones y las discrepancias se resuelven por consenso. No obstante, la escala y el procedimiento concreto para asignar los valores se pueden establecer en cada proyecto concreto.

Tabla 4
Tabla de extracción de datos

Código	Variable	Puntuación
	Síntesis narrativa	
1.1	Scratchpad	
1.2	Articulación	
1.3	Conexión	
1.4	Extensión	
1.5	Multimodalidad	
	Fuentes	
2.1	Número	
2.2	Diversidad	
2.3	Capilaridad	
2.4	Bien formadas	
	Prompts adicionales	
3.1	Oportunidad	
3.2	Variedad	
	Idiofunciones	
4.1	Funciones propias y exclusivas de cada sistema considerado	
TOTAL		

Tabla de síntesis comparativa

Sistema	Síntesis	Fuentes	Prompts	Idiofunción	TOTAL

Las tablas anteriores son ejemplos habituales de sistemas de análisis basados en el uso de matrices. Para cada proyecto concreto, los responsables pueden modificar los aspectos que consideren convenientes.

3.6. Otros modos de evaluación

Es evidente que pueden desarrollarse diferentes modos de evaluación. Por citar algunas alternativas que nos parecen significativas, podemos citar la evaluación en base a tareas, como en Font-Julián et al. (2024) en la que se desarrolla un modelo de evaluación que combina procedimientos de análisis cualitativos y cuantitativos bajo determinadas tareas y con un grupo de dos o más usuarios como jueces que consensúan sus evaluaciones.

Además, hay métodos de evaluación mediante *benchmarking* automatizados, como los que pueden verse en el portal *Artificial Analysis* (<https://artificialanalysis.ai/>) donde se comparan periódicamente decenas de modelos de lenguaje en base a una batería de tests.

3.7. Aportación diferencial de este modo de evaluación

Cualquier modo de evaluación puede ser de utilidad, todo depende del contexto y de los objetivos de cada caso. El modo de evaluación cualitativo que proponemos aquí tiene una triple función:

- Reforzar el pensamiento crítico de los usuarios respecto a la IA.
- Aportar un método de enseñanza/aprendizaje y de adquisición de habilidades en el uso de sistemas de IA.
- Disponer de un procedimiento de evaluación y de análisis comparativo de sistemas de IA basado en el análisis matricial cualitativo de parámetros e indicadores.

3.8. Procedimiento de geometría variable

El procedimiento MASIA presentado aquí aporta esquemas de análisis que se pueden aplicar tal como se presentan, cosa para la que están concebidos. Pero, por un lado, no es imprescindible utilizar todas las variables, y por otro lado se pueden añadir nuevas variables o incluso considerar otros parámetros. La idea de este método de evaluación en su esencia es la que sigue:

- Los diseñadores del análisis, con cualquiera de los tres objetivos enunciados antes, pueden considerar la conveniencia de añadir nuevas variables o de retirar alguna de las mismas.
- A partir de una concepción de la estructura de respuestas y de utilidades de un sistema de IA se propone una serie de variables de análisis, para cuya evaluación se dispone de una guía sucinta y mínimamente viable.
- Los análisis se llevan a cabo mediante matrices como las mostradas en las tablas precedentes, lo que facilita tanto la toma de datos como las operaciones de comparación.

Si el análisis se lleva a cabo con objetivos de investigación, el equipo que desee aplicar el procedimiento MASIA debe añadir los protocolos necesarios de validez, fiabilidad, transparencia y trazabilidad que son habituales en las investigaciones de calidad.

3.9. Fases del procedimiento

Concebimos el procedimiento MASIA como parte de un flujo de trabajo donde se distinguen tres fases. Una fase propia del procedimiento que se presenta en este capítulo y dos fases más que deberían tenerse en cuenta siempre que se utilizan sistemas de IA en contextos académicos, independientemente de las especificidades que determinen su uso.

3.10. Preparación

Se trata de la fase inicial, donde el investigador determina si es conveniente usar sistemas de IA para resolver una tarea, así como seleccionar la más adecuada. En esta primera fase el investigador debe tener en cuenta el objetivo de la tarea a resolver, así como los distintos sistemas de IA que mejor pueden contribuir a su consecución, si es el caso, ya sea mediante herramientas gratuitas o de suscripción, propia o institucional. También se deben valorar aspectos relacionados con la privacidad y la protección de datos. Por último, es conveniente hacer una

prueba o pretest antes de usar formalmente un sistema de IA, especialmente cuando se usa uno por primera vez, para facilitar una evaluación inicial de la eficacia de la herramienta en un área que el investigador conozca bien. Se recomienda aplicar esta fase a cualquier modelo que integre el uso de sistemas de IA en contextos académicos.

3.11. Análisis

Esta fase comprende el procedimiento MASIA para evaluar sistemas de IA que presentan, en su página de resultados, cómo mínimo los siguientes tres elementos: síntesis narrativa, fuentes consultadas y nuevos *prompts*. El investigador evalúa distintos sistemas de IA mediante el uso de matrices y variables, a las que asigna una puntuación en base al cumplimiento de una serie de preguntas de chequeo, que le permiten identificar qué sistema de IA es el más adecuado.

3.12. Ética investigadora

Esta fase debe guiar todo el proceso, desde la concepción de la tarea o proyecto de investigación para el cual se usarán sistemas de IA, hasta su culminación. Como en el caso de la fase de preparación, la ética investigadora debe formar parte de cualquier proyecto de investigación que use sistemas de IA, independientemente del enfoque o modelo adoptado. La fase de ética investigadora contempla la documentación del proceso, mediante el registro de *prompts* utilizados y fecha y contexto de interacción, o capturas de pantalla para fomentar la transparencia y la reproducibilidad. También se deben validar los resultados y verificar las fuentes obtenidas en los sistemas utilizados. Otro aspecto importante es evitar el plagio, tanto en respuestas basadas en trabajos previos como en contenido generado por los propios sistemas de IA. Además, en contextos académicos, es esencial mantener la trazabilidad de los trabajos originales y citar adecuadamente el uso de sistemas de IA.

4. Conclusiones

Hemos presentado un procedimiento relativamente simple para analizar y comparar sistemas de IA aplicados a trabajos y entornos académicos. El procedimiento cubre diferentes funciones, la primera y tal vez más destacada es que apoya el pensamiento crítico de los usuarios de las IA, al hacerles evaluar aspectos significativos de las IA.

Otra función es que, por la misma razón enunciada más arriba, pero no reducible a ella, se puede utilizar en procesos de enseñanza aprendizaje en la formación de estudiantes universitarios e investigadores predoctorales.

Por último, es un procedimiento que puede utilizarse con fines de investigación. Para ello, los miembros del equipo disponen de esquemas sistemáticos con los que analizar y comparar diversas IA a partir de un mismo *prompt* o de un pequeño conjunto de *prompts*.

En general, entendemos que este procedimiento, como señalamos más arriba, es fácil de interpretar y de aplicar, sin perjuicio de su capacidad heurística, que puede ser notable gracias al esfuerzo que los analistas apliquen en el análisis.

5. Financiación

Este trabajo forma parte del proyecto “Parámetros y estrategias para incrementar la relevancia de los medios y la comunicación digital en la sociedad: curación, visualización y visibilidad (CUVICOM)”. Ayuda PID2021-123579OB-I00 financiada por MICIU/AEI/10.13039/501100011033 y por FEDER, UE.

6. Referencias

Aguilera-Cora, E., Lopezosa, C., & Codina, L. (2024a). *Scopus AI beta: Functional analysis and cases* (DigiDoc Editorial Series. DigiDoc Reports). Pompeu Fabra University, Department of Communication. <https://repositori.upf.edu/handle/10230/58658>

Aguilera-Cora, E., Lopezosa, C., Fernández-Cavia, J., & Codina, L. (2024b). Accelerating research processes with Scopus AI: A place branding case study. *Pan-American Journal of Communication*, 6(1). <https://doi.org/10.21555/rpc.v6i1.3088>

Bianchini, S., Müller, M., & Pelletier, P. (2022). Artificial intelligence in science: An emerging general method of invention. *Research Policy*, 51(10), 104604. <https://doi.org/10.1016/j.respol.2022.104604>

Bhatia, P. (2023). ChatGPT for academic writing: A game changer or a disruptive tool? *Journal of Anaesthesiology Clinical Pharmacology*, 39(1), 1. https://doi.org/10.4103/joacp.joacp_84_23

Codina, L. (2025). *Sistemas de inteligencia artificial para la síntesis del conocimiento: marcos teóricos y revisiones de la literatura*. [Presentación]. Facultad de Comunicación, Universidad Pompeu Fabra. <http://hdl.handle.net/10230/70242>

Codina, L., & Garde, C. (2023). Uso de ChatGPT en la docencia universitaria: fundamentos y propuestas[Informe]. *Facultad de Comunicación, Universidad Pompeu Fabra*. <http://hdl.handle.net/10230/570157>

Codina, L., & Pedraza-Jiménez, R. (2016). Características y componentes de un sistema de análisis de medios digitales: el SAAMD. R. Pedraza-Jiménez, L. Codina, & J. Guallar (Eds.), *Calidad en sitios web: método de análisis general, e-commerce, imágenes, hemeroteca y turismo* (pp. 15–40). UOC Publicaciones.

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, 20(1). <https://doi.org/10.1186/s41239-023-00392-8>

Dwivedi, Y.K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J.S., Eirug, A., Galanos, V., Ilavarasan, P.V., Janssen, M., Jones, P., Kar, A.K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., ... Williams, M.D. (2021). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, 57, 101994. <https://doi.org/10.1016/j.ijinfomgt.2019.08.002>

Elsevier. (2024, January 16). *Launch of Scopus AI to help researchers navigate the world of research*. <https://www.elsevier.com/about/press-releases/launch-of-scopus-ai-to-help-researchers-navigate-the-world-of-research>

- Font-Julián, C. I., Orduña-Malea, E., & Codina, L. (2024). ChatGPT Search as a tool for scholarly tasks: Evolution or return? *Infonomy*, 2(5). <https://doi.org/10.3145/infonomy.24.059>
- Gundersen, O. E., Gil, Y., & Aha, D.W. (2018). On reproducible AI: Towards reproducible research, open science, and digital scholarship in AI publications. *AI Magazine*, 39(3), 56–68. <https://doi.org/10.1609/aimag.v39i3.2816>
- High- Level Expert Group on Artificial Intelligence. (2019, April 8). *Ethics guidelines for trustworthy AI*. European Commission. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- Kaebnick, G. E., Magnus, D., Kao, A., Hosseini, M., Resnik, D. B., Dubljević, V., Rentmeester, C. A., Gordijn, B., & Cherry, M. J. (2023). Editor's statement on the responsible use of generative AI technologies in scholarly journal publishing. *Medicine, Health Care and Philosophy*, 26(4), 499–503. <https://doi.org/10.1007/s11019-023-10176-6>
- Lopezosa, C., & Codina, L. (2023). Testing Bard: Probando Bard: así funciona la Inteligencia Artificial Generativa de Google. *Anuario ThinkEPI*, 17. <https://doi.org/10.3145/thinkepi.2023.e17a25>
- Lopezosa, C., Codina, L., & Ferran-Ferrer, N. (2023a). *ChatGPT como apoyo a las systematic scoping reviews: integrando la inteligencia artificial con el framework SALSA*. Col·lecció del CRICC. Barcelona: Universitat de Barcelona.
- Lopezosa, C., Codina, L., Boté-Vericad, J. J. (2023b). *Testeando ATLAS.ti con OpenAI: hacia un nuevo paradigma para el análisis cualitativo de entrevistas con inteligencia artificial*. Barcelona: Universitat Pompeu Fabra. Departament de Comunicació. [30] p. (Serie Editorial DigiDoc. DigiDoc Reports).
- Lund, B., Wang, T., Mannuru, N. R., Nie, B., Shimray, S. R., & Wang, Z. (2023). ChatGPT and a new academic reality: Artificial intelligence-written research papers and the ethics of large language models in scholarly publishing. *Journal of the Association for Information Science and Technology*, 74(5), 570–581. <https://doi.org/10.1002/asi.24750>
- Tlili, A., Huang, R., Mustafa, M. Y., Zhao, J., Bozkurt, A., Xu, L., ... Wang, H. (2023). Speaking of transparency: Are all artificial intelligence (AI) literature reviews in education transparent? *The Journal of Applied Learning & Teaching (JALT)*, 6(2). <https://doi.org/10.37074/jalt.2023.6.2.15>
- UNESCO. (2023, September 8). *Guidance for generative AI in education and research*. <https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research>
- Whitfield, S., & Hofmann, M. A. (2023). Elicit: AI literature review research assistant. *Public Services Quarterly*, 19(3), 201–207. <https://doi.org/10.1080/15228959.2023.2224125>
- World Commission on the Ethics of Scientific Knowledge and Technology. (2019) *Preliminary study on the ethics of artificial intelligence*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000367823>