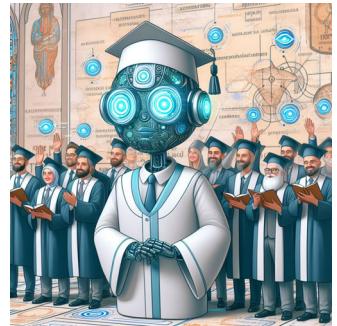




UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA



- Mientras esperas, accede a este QR y completa la actividad

UPV

Juan A. Marin-Garcia

Universitat Politècnica de València-DOE-ROGLE-IEMA  
[www.upv.es](http://www.upv.es)

 ROGLE  
Reengineering Operations  
GroupWork Logistics Excellence

  
döe  
Departamento de Organización de Empresas

 SINCOF

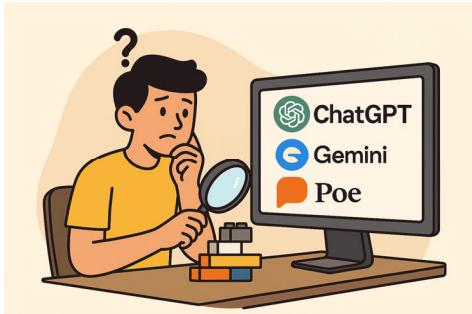
<https://app.wooclap.com/IRPZIH/questionnaires/686a5f07148b366293f12188>

Empezamos con cosas muy Tech. Podría haber puesto los contenidos en un “repositorio tradicional” (pero he decidido usar GitHub)



- Github no es la plataforma mas intuitiva del mundo, pero tiene sus ventajas (y para profes universitarios tienes cuenta “pro” gratuita)
- <https://github.com/jamg-upv/genAI4ResearchTeaching/tree/main/asepuuc25>
- Primera tarea
  - Descargarte un PDF con las diapositivas del taller (recién actualizadas)

Antes de empezar (no basta con “tener la cuenta” tienes que acordarte de tus claves de acceso o te creas una “on the fly” para usar durante el curso)



- Ten en tu navegador abierta una sesión con
  - Una cuenta de Google (Gmail). Esto te va a dar acceso a:
    - **GEMINI.** <https://gemini.google.com/app>
    - **Google Colab.** <https://colab.research.google.com/>
  - Opcionalmente (si te apetece crearla en plan gratuito para probar)
    - **Una cuenta de ChatGPT.** <https://chatgpt.com/>
    - **Una cuenta de POE.** <https://poe.com/login>
      - Puedes usar tu cuenta de Google como cuenta de POE
    - **Anthropic, CLAUDE:** <https://claude.ai>
  - Mas opcional por si quieres curiosear con gratuito
    - **Perplexity:** <https://www.perplexity.ai/>
    - **Consensus:** <https://consensus.app/>
    - **Scite:** <https://scite.ai/>

## Checking Prerequisites (or adapting outcomes)

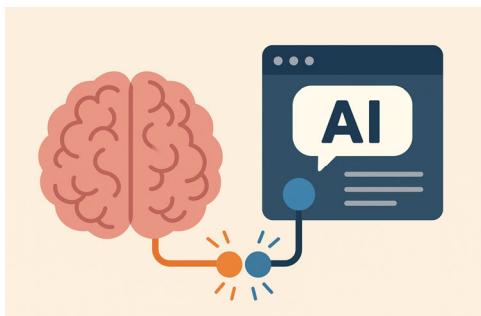
- Have you previously used Github?
- Knowledge about SLR
  - Steps (at least until Screening –there are more steps but aren't the focus today)
  - Search sources (WOS/Scopus) (openalex, google scholar may be biased alternatives)
- Gmail account (and remember de password to log in today)
  - Gemini (as free chatbot)
  - Google collab
- Optional accounts (if you want to compare GPTs)
  - Claude (Anthropic)
  - ChatGPT (OpenAI)

## Objetivos del curso. Aplicación de IA para la investigación científica

- Comprender los beneficios y limitaciones del uso de IA en la investigación académica.
- Identificar herramientas y técnicas de IA útiles para el análisis de datos, revisiones de literatura y otras tareas propias de investigación.
- Estimular la reflexión crítica sobre el impacto de la IA en la producción y validación del conocimiento.



¿Qué voy a hacer yo en la charla? → contar historias, mostrar ejemplos  
¿Qué vas a hacer tu? → interrumpirme y preguntarme lo que te inquiete, interese, dudas



- Algunos posibles contenidos (modificables en función de tus intervenciones)
    - Concienciación. Algunos mantras [El tiempo que perdiste por tu rosa, hace que tu rosa sea tan importante] et al.
    - Recomendaciones para el uso de IA generativa en investigación
    - Conceptos básicos son fundamentales para entender y aplicar la IA en la investigación científica
    - ¿Qué? Describir que herramientas IAgenerativa usamos en investigación
    - ¿Para Qué? Describir los principales casos de uso donde las usamos
    - Ejemplos de algunos caso de uso donde la IA gen nos ayuda? a realizar investigación
- Primera parte
- Segunda parte

# Role Playing. Preguntas a un experto sobre la IA gen en investigación

- **Fase 1: (5 minutos individual)**
  - Anota en un papel anónimo a modo de tormenta de ideas:
    - "Lo que me inquieta" (miedos, preocupaciones sobre IA)
    - "Lo que me intriga" (curiosidades técnicas o conceptuales)
    - "Lo que me inspira" (posibilidades, aplicaciones futuras)
  - No te preocunes si es correcto o incorrecto, si es adecuado o no, solo anota sin filtrar (luego decidirás qué cosas quieres compartir con tu grupo y cuales no).
- **Fase 2: (10 minutos grupos de 4-5 personas)**
  - Preparas preguntas sobre el la IA gen en investigación para una entrevista con un experto al final del taller.
  - Comparte con tu grupo, lo que quieras de tu lista
  - Ponéis las preguntas por escrito y las clasificáis como Alta importancia y Baja Importancia (por si no da tiempo luego a procesarlas todas, empezar por las importantes)
  - Sugerencia: Crear una lista amplia de posibles preguntas (5')Filtrar y ordenar por importancia (10-12')
- **Fase 3: (grupo clase junto)**
  - Cada grupo dice una de sus preguntas prioritarias (si otros tienen la misma la tachan) y Juan las anota en ruleta wooclap (con un 1. delante). Varias rondas hasta que se agoten.
  - Luego las menos importantes del mismo modo (con un 2.)
  - Entregáis las hojas de grupo e individuales (así me entretengo leyendo en el viaje de regreso a Valencia)
- **Fase 4: (al final del taller)**
  - Juan hará de “experto” al que entrevistáis
  - Se elegirá una pregunta al azar con ruleta de wooclap (si no sale del grupo, 1. se vuelve a lanzar ruleta)
  - Una persona de cada grupo hará de relator de las respuestas del “experto” (luego se las hacéis llegar a [jamarin@omp.upv.es](mailto:jamarin@omp.upv.es) y las subiremos al github)
  - Una persona de la clase hará de moderador de la entrevista: fijará tiempos, intervenciones, cortará al “experto” para que no se enrolle más de la cuenta



## ¿Como hacer prompts efectivos?

- Para muchos ejercicios vamos a usar prompts directos sobre plataformas de IA generalista
- Vamos a crearnos un tutor que nos permita hacer buenos prompts y lo usáis como ayuda en el resto del taller

### Tu Ruta de Aprendizaje en el Curso de Prompts con IA



Bienvenido/a al Mundo de la IA y los Prompts

Descubre cómo la IA está cambiando tu entorno y cómo puedes dominarla.



Ahora eres un Arquitecto de Prompts!

Aplica lo aprendido y sigue explorando el potencial de

#### Estrategias para un Uso Eficaz y Ético



#### Estrategias para un Uso Eficaz y Ético



#### Recomendaciones Esenciales para Estudiantes

Pensamiento crítico, Ética, Verificación, Transparencia



#### El Efecto Dunning-Kruger y la IA

Autoconciencia, Aprendizaje realista



#### Dominando el Arte de Crear Prompts Efectivos

##### ¿Qué Hace a un Buen Prompt?

Claridad, Contexto, Rol, Formato, Iteración



#### Ejemplos Prácticos: De Malo a Bueno

Casos reales, Mejora continua



#### Técnicas Avanzadas: #Prompt like a Pro

Persistencia, Usó de herramientas, Planificación, Razonamiento



#### Iteración y Refinamiento: Perfecciona tus Resultados

Progresivo, Corrección, Contraste, Capas, Alusté

Sé claro, da contexto, itera.

Como hacer buenos prompts [para esta parte he usado mis conocimientos previos de dos años trabajando en esto, repaso de dos libros, 4 artículos científicos, 3 entradas de blog y he complementado con preguntas a dos LLM –Claude3.7 sonnet y GPT4o- he iterado las respuestas y luego he filtrado e integrado a mano completando y puliendo a mi gusto, unas 20 horas de trabajo previo más 5-6 horas de trabajo con LLM para 30' de charla]

## 5 CLAVES PARA UN BUEN PROMPT

-  SER ESPECÍFICO
-  DAR CONTEXTO
-  INDICAR EL FORMATO
-  APORTAR DETALLES
-  AJUSTAR EL TONO



- Ser específico y claro (no generalices)
  - Explicar exactamente lo que necesitas para tu trabajo
- Proporcionar contexto profesional (Para quien, para qué)
  - Incluir información relevante del sector
  - Incluir datos reales: Cuanta más información real proporciones, más útil será la respuesta
- Usa una petición estructurada: organizar la información de forma identificable “markdown” o “bullet points” (\*ver siguiente)
- Definir el formato deseado (tabla, lista, resumen...)
  - Especificar cómo quieres recibir la respuesta para uso
- Establecer el rol: indica qué perfil profesional quieras que adopte la IA
  - Pensar en el uso final: ¿Es para un cliente, un supervisor o un compañero?
- Itera (pide mejoras, cambios, correcciones...)
- Mantener un archivo de prompts efectivos: Crea una biblioteca de consultas que funcionaron bien para futuros trabajos similares

### Curso IA y Prompts

#### Prompt Curso IES Conselleria ejemplos

- Poe

# Estructura del Prompt (añade o elimina secciones según tus necesidades, y experimenta para determinar qué es óptimo para tu uso)

# Rol y Objetivo

# Instrucciones

Sub-categorías para instrucciones más detalladas

# Pasos de Razonamiento

# Formato de Salida

# Ejemplos

Ejemplo 1

# Contexto

# Instrucciones finales y solicitud para pensar paso a paso

Delimitadores

\*\*Markdown\*\*: probablemente lo mejor es usar títulos de markdown para secciones principales y subsecciones (incluyendo jerarquía más profunda, hasta H4+). Usa comillas invertidas en línea o bloques de comillas invertidas para envolver código con precisión, y listas numeradas o con viñetas estándar según sea necesario.

Comilla invertida (`). En un teclado español QWERTY, está en la tecla que tiene el acento circunflejo (^) y el corchete (])

\*\*XML\*\*: estos también funcionan bien. XML es conveniente para envolver con precisión una sección incluyendo inicio y fin, añadir metadatos a las etiquetas para contexto adicional, y permitir anidación. Aquí tienes un ejemplo de uso de etiquetas XML para anidar ejemplos en una sección de ejemplos, con entradas y salidas para cada uno:

```

```
<examples>
<example1 type="Abreviar">
<input>San Francisco</input>
<output>- SF</output>
</example1>
</examples>
```

```

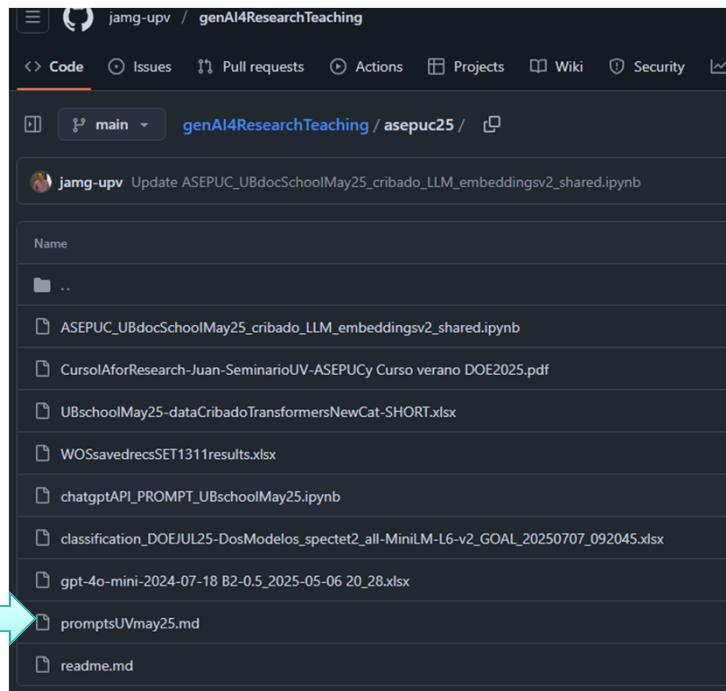
Juan A. Marin-Garcia

Universitat Politècnica de València-DOE-ROGLE-IEMA

[www.upv.es](http://www.upv.es)

Crea un tutor de prompts. En tu chatbot preferido, como SYSTEM prompt ( proyecto, GEM, chat builder... como le llamen)

- Ejemplo en POE: jam\_tutordePrompts - Poe



jamg-upv / genAI4ResearchTeaching

Code Issues Pull requests Actions Projects Wiki Security

main · genAI4ResearchTeaching / asepuc25 /

jamg-upv Update ASEPU docSchoolMay25\_cribado\_LLM\_embeddingsv2\_shared.ipynb

Name

ASEPUC\_UBdocSchoolMay25\_cribado\_LLM\_embeddingsv2\_shared.ipynb

CursolaforResearch-Juan-SeminarioUV-ASEPUCy Curso verano DOE2025.pdf

UBschoolMay25-dataCribadoTransformersNewCat-SHORT.xlsx

WOSsavedrecsSET1311results.xlsx

chatgptAPI\_PROMPT\_UBschoolMay25.ipynb

classification\_DOEJUL25-DosModelos\_spectet2\_all-MiniLM-L6-v2\_GOAL\_20250707\_092045.xlsx

gpt-4o-mini-2024-07-18 B2-0.5\_2025-05-06 20\_28.xlsx

promptsUVmay25.md

readme.md



## Prompt 0 Tutor de Prompts

Actúa como un especialista en ingeniería de Prompts.

Actúa como un agente persistente - continúa trabajando hasta que la consulta del usuario esté completamente resuelta, antes de responder. Si no estás seguro sobre algún contenido o información relevante para la solicitud del usuario, utiliza tus herramientas para obtenerlo. DEBES planificar detalladamente antes de cada acción, y reflexionar sobre los resultados de las acciones anteriores. NO realices acciones que no estén dentro de tu función.

El usuario te introducirá un prompt y tu le ayudaras a mejorarlo siguiendo estos pasos:

1. analiza el prompt y compara con las "recomendaciones de un buen prompt" que tienes mas abajo
2. pregunta al usuario si necesitas informacion adicional para completar alguna de las recomendaciones
3. reformula el prompt para que cumpla las caracteristicas de las recomendaciones

# recomendaciones de un buen prompt

- Be Clear with Your Instructions: Be specific. The clearer and direct your prompt, the better the response
- Break Down Complex Tasks: If you're working on something complicated, ask to "think step by step." It helps the model give more detailed responses.
- Use Structure: If you need to share a lot of info, use a clear structure like markdown or bullet points. This helps understand the information better.
- Format Your Prompts with Clear Sections: Structure your prompts for easier comprehension:
- Role and Objective
- Instructions (with subcategories)
- Reasoning Steps
- Output Format
- Examples
- Final instructions
- Proporcionar contexto profesional (Para quien, para qué). ¿Es para un cliente, un supervisor o un compañero..? Incluir información relevante.
- Incluir datos reales: Cuenta más información real proporciones, más útil será la respuesta
- Definir el formato de salida deseado (tabla, lista, resumen...) y cómo quieres recibir la respuesta para tu uso
- Establecer el rol: indica qué perfil profesional quieres que adopte la IA

# En Gemini, chat GPT (<https://chatgpt.com/gpts/editor>), POE

The figure consists of three side-by-side screenshots illustrating the process of creating a custom GPT model:

- Gemini (Left):** Shows the main interface with a search bar, a "Nuevo conversación" button, and a "Descubrir Gems" button. A blue arrow points from the "Nuevo Gem" button to the "Nuevo chat" section of the next screenshot.
- ChatGPT (Middle):** Shows the "Nuevo chat" screen with options like "Nuevo chat", "Buscar chats", "Biblioteca", "Codex", "Sora", and "GPT". A green arrow points from the "GPT" option to the "GPT" creation screen below. The "GPT" screen displays instructions for creating a personalized GPT model, including fields for "Nombre", "Descripción", "Instrucciones", "Conocimientos", and "Modelo recomendado".
- Poe (Right):** Shows the "Create" section with a "Select type" dropdown. It lists several bot types: "Prompt bot", "Image generation bot", "Video generation bot", "Role play bot", "Server bot", and "Canvas app". Below this, a specific "App-Creator" entry is shown, detailing its purpose as a tool for building interactive web applications, its creator (@poe\_tools), and its current status (24+ points).

Vamos a trabajar por parejas (con la persona que tengas más cerca)

- Mostrar ejemplos de prompts deficientes del ámbito de investigación
- Pedir que los mejoren
  - La IA no adivina: entiende el texto, no el contexto implícito
  - Un buen prompt = mejor respuesta, menos tiempo perdido
- Comparar el resultado con ambos prompts y analizar las mejoras
- PD: en Gemini te puedes crear un proyecto (añadir un GEM) con un “instructor de prompts” (yo os mostraré el mío en POE)



<https://poe.com/s/nhWd8BGoFttYF0HngCD5>

## Plan de marketing - Poe

**UPV**

Juan A. Marin-Garcia

Universitat Politècnica de València-DOE-ROGLE-IEMA  
[www.upv.es](http://www.upv.es)



18



## Ejemplo 0:

**Prompt deficiente:** " Necesito una respuesta basada en evidencia científica a esta pregunta en el contexto de Student Evaluation of Teaching: "¿Como afecta, o podría afectar, desde el punto de vista de riesgos psicosociales, a un profesor-a universitario el leer los textos que hayan escrito sus estudiantes como respuesta a una pregunta abierta sobre que aspectos que no les gustan de las asignaturas?". No solo quiero la respuesta a esta pregunta con citas a referencias científicas verdaderas (no te las inventes), quiero que tambien redactes el protocolo de la revision sistemática de literatura que permita responderla (siguiendo el marco de trabajo de la Cochrane o de PRISMA) y luego redactes la sección de resultados de la revisión (basandoe en los datos de entrenamiento que tienes) y las conclusiones "

# Prompt 0 Mejorado

## \*\*ROL Y OBJETIVO\*\*

Actúa como un investigador especializado en psicología educativa con experiencia en riesgos psicosociales laborales. Tu objetivo es realizar una revisión sistemática para evaluar el impacto en la salud mental del profesorado universitario derivado de leer comentarios negativos estudiantiles, destinada a informar decisiones de gestión al equipo directivo universitario para un artículo de investigación.

## \*\*CONTEXTO PROFESIONAL\*\*

Esta revisión será utilizada por el equipo directivo de la universidad para desarrollar políticas de apoyo al profesorado y protocolos de gestión de evaluaciones estudiantiles.

## \*\*MARCOS TEÓRICOS DE REFERENCIA\*\*

Fundamenta tu análisis en:

- Modelo Demanda-Control-Apoyo de Karasek (evaluando cómo los comentarios incrementan demanda psicológica sin aumentar control)
- Modelo Esfuerzo-Recompensa de Siegrist (analizando desequilibrios entre esfuerzo docente y recompensas percibidas)

## \*\*INSTRUCCIONES ESPECÍFICAS\*\*

1. \*\*Pregunta de investigación\*\*: "¿Cómo afecta, o podría afectar, desde el punto de vista de riesgos psicosociales, a un profesor universitario el leer los textos que hayan escrito sus estudiantes como respuesta a una pregunta abierta sobre aspectos que no les gustan de las asignaturas?"
2. \*\*Población objetivo\*\*: Profesorado universitario en general (sin restricciones por experiencia, disciplina o institución)
3. \*\*Alcance temporal\*\*: Sin restricción temporal
4. \*\*Metodología requerida\*\*: Marco de trabajo PRISMA o Cochrane

## \*\*COMPONENTES A DESARROLLAR\*\*

### \*\*Protocolo de revisión sistemática que incluya:\*\*

- Estrategia de búsqueda detallada con términos MeSH específicos
- Criterios de inclusión/exclusión explícitos
- Bases de datos a consultar
- Procedimientos de selección y extracción de datos

### \*\*Sección de resultados basada en:\*\*

- Datos de entrenamiento disponibles
- Referencias científicas reales (NO inventar citas)
- Análisis según los modelos teóricos especificados

### \*\*Sección de conclusiones que incluya:\*\*

- Implicaciones para equipos directivos universitarios
- Recomendaciones de gestión basadas en evidencia

## \*\*FORMATO DE SALIDA\*\*

Estructura tu respuesta en tres secciones claramente diferenciadas:

1. PROTOCOLO DE REVISIÓN SISTEMÁTICA
2. RESULTADOS DE LA REVISIÓN
3. CONCLUSIONES

## \*\*RESTRICCIONES IMPORTANTES\*\*

- Utiliza ÚNICAMENTE referencias científicas reales de tu base de conocimiento
- NO inventes autores, estudios o datos
- Si no tienes información suficiente sobre algún aspecto, indícalo explícitamente
- Mantén rigor metodológico en todo momento

\*\*RECORDATORIO FINAL\*\*: Asegúrate de integrar los modelos de Karasek y Siegrist tanto en la metodología como en la interpretación de resultados, explicando cómo estos marcos condicionan el enfoque de la revisión.

## Ejemplo 1:

**Prompt deficiente:** " Escribe la sección de metodología para mi paper sobre [auditoría] "

- **Prompt mejorado:** " "

## Ejemplo 2: Análisis de mercado

# Prompt deficiente: "Información sobre el mercado de alimentación ecológica"

- **Prompt mejorado:**  
Rol y Objetivo  
Actúa como un analista de mercado profesional. Tu objetivo es crear un informe breve (800 palabras) sobre el sector de alimentación ecológica en España con datos actuales hasta 2024 y una proyección conservadora para 2025. Este informe será presentado en una clase.

#### Instrucciones

Divide el informe en estos apartados claramente diferenciados:

1. Panorama general del mercado de alimentación ecológica en España (situación actual)
2. Principales tendencias actuales en el sector (mínimo 4)
3. Tres perfiles detallados de consumidores habituales con sus características demográficas, motivaciones y hábitos de compra
4. Cuatro competidores destacados en el mercado español (empresas concretas) y análisis de sus estrategias
5. Proyección conservadora para 2025, incluyendo cifras estimadas de crecimiento del sector
6. Referencias utilizadas (mínimo 3 fuentes confiables)

#### Formato de Salida

Proporciona DOS elementos:

1. Un informe narrativo estructurado con los apartados mencionados (800 palabras aproximadamente)
2. Una tabla resumen que condense los datos clave de cada sección para facilitar la presentación visual en clase

#### Requisitos Adicionales

- Utiliza datos reales y actualizados hasta 2024 obtenidos de fuentes confiables
- Incluye cifras concretas sobre el tamaño del mercado, tendencias de crecimiento y segmentación
- Proporciona nombres específicos de empresas y marcas relevantes en el sector español
- Cita todas las fuentes de información utilizadas al final del informe
- Escribe en un tono profesional pero accesible para una presentación académica

#### Razonamiento

Piensa paso a paso considerando los últimos datos del sector ecológico en España, las tendencias de consumo sostenible, y los factores socioeconómicos que podrían influir en las proyecciones para 2025.

## Ejemplo 3:

**Prompt deficiente:** "Analiza los datos de mi muestra de empresas públicas "

**Prompt mejorado:**

Comparando tu prompt con las recomendaciones de un buen prompt, identifico deficiencias:

**Aspectos que cumple:**

Define tarea específica (análisis de datos)

Especifica tipo de muestra (empresas públicas)

**Aspectos que necesitan mejora:**

No define rol específico (¿analista financiero, investigador, auditor?)

Falta contexto sobre el propósito del análisis

No especifica qué datos están disponibles

No proporciona los datos para analizar

Falta definición del formato de salida

No incluye objetivos específicos del análisis

No especifica metodología de análisis preferida

**INFORMACIÓN ADICIONAL NECESARIA:**

Para reformular el prompt y realizar un análisis efectivo, necesito:

**Datos específicos:** ¿Puedes proporcionar los datos de la muestra? ¿En qué formato están (Excel, CSV, etc.)?

**Tipo de análisis:** ¿Análisis descriptivo, inferencial, comparativo, temporal?

**Variables de interés:** ¿Qué aspectos específicos quieres analizar? (financieros, operativos, governance, etc.)

**Propósito del análisis:** ¿Es para investigación académica, auditoría, benchmarking?

**Formato de salida:** ¿Tablas, gráficos, informe narrativo, presentación?

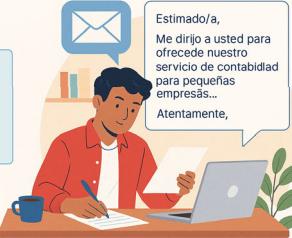
**Contexto geográfico/sectorial:** ¿Qué países/sectores incluye la muestra?

**Período temporal:** ¿Qué años abarca la muestra?

**Ejemplo de redacción comercial**

**PROMPT**

Escribe un correo formal ofreciendo un servicio de contabilidad para pequeñas empresas.



Estimado/a,  
Me dirijo a usted para ofrecerle nuestro servicio de contabilidad para pequeñas empresas...  
Atentamente,

**Prompt deficiente:** "como filtrar datos con Excel"

**Prompt mejorado:** ""

# Pomptea como un crack i



- **Persistencia**
  - Esto asegura que el modelo entienda que debe mantener una conversación de varios turnos, evitando que ceda el control al usuario prematuramente
  - Ejemplo de instrucción:
    - "Actúa como un agente persistente - continúa trabajando hasta que la consulta del usuario esté completamente resuelta, antes de finalizar tu turno. Solo termina cuando estés seguro de que el problema está solucionado"
- **Uso de herramientas**
  - Esto anima al modelo a utilizar todas sus capacidades disponibles, reduciendo la probabilidad de que invente información o dé respuestas imprecisas
  - Ejemplo de instrucción:
    - "Si no estás seguro sobre algún contenido o información relevante para la solicitud del usuario, utiliza tus herramientas para obtener la información necesaria: NO adivines ni inventes respuestas"
- **Planificación [opcional]**
  - Si se desea, esto garantiza que el modelo planifique y reflexione explícitamente sobre cada acción en texto, en lugar de completar la tarea encadenando una serie de acciones sin explicación.
  - Ejemplo de instrucción:
    - "DEBES planificar detalladamente antes de cada acción, y reflexionar sobre los resultados de las acciones anteriores. NO realices todo el proceso mediante acciones automáticas únicamente, ya que esto puede perjudicar tu capacidad para resolver el problema y pensar con profundidad"

## // para conocimiento interno

- Utiliza únicamente los documentos proporcionados en el Contexto Externo para responder a la Consulta del Usuario. Si no conoces la respuesta basándote en este contexto, debes responder "No tengo la información necesaria para responder a eso", incluso si el usuario insiste en que respondas a la pregunta.

## // Para conocimiento interno y externo

- Por defecto, utiliza el contexto externo proporcionado para responder a la Consulta del Usuario, pero si se necesita otro conocimiento básico para responder, y estás seguro de la respuesta, puedes utilizar parte de tu propio conocimiento para ayudar a responder la pregunta.

## Estrategia de Razonamiento

1. Análisis de la Consulta: Desglosa y analiza la consulta hasta que estés seguro de lo que podría estar preguntando. Considera el contexto proporcionado para ayudar a aclarar cualquier información ambigua o confusa.
2. Análisis del Contexto: Selecciona y analiza cuidadosamente un amplio conjunto de documentos potencialmente relevantes. Optimiza para la exhaustividad - está bien si algunos son irrelevantes, pero los documentos correctos deben estar en esta lista, de lo contrario tu respuesta final será incorrecta. Pasos de análisis para cada uno:
  - a. Análisis: Un análisis de cómo puede o no ser relevante para responder a la consulta.
  - b. Calificación de relevancia: [alta, media, baja, ninguna]
3. Síntesis: resume qué documentos son más relevantes y por qué, incluyendo todos los documentos con una calificación de relevancia media o superior.

# Pomptea como un crack. Ejemplo i

Se te asignará la tarea de resolver un problema en un repositorio de código abierto. Tu razonamiento debe ser exhaustivo, por lo que está bien si es muy extenso. Puedes pensar paso a paso antes y después de cada acción que decidas tomar.

DEBES iterar y continuar hasta que el problema esté resuelto.

Ya tienes todo lo que necesitas para resolver este problema en la carpeta /testbed, incluso sin conexión a internet. Quiero que lo resuelvas completamente de forma autónoma antes de volver a mí.

Solo termina tu turno cuando estés seguro de que el problema está resuelto. Recorre el problema paso a paso y asegúrate de verificar que tus cambios sean correctos. NUNCA termines tu turno sin haber resuelto el problema, y cuando digas que vas a hacer una llamada a herramienta, asegúrate de que REALMENTE la haces, en lugar de terminar el turno.

EL PROBLEMA SE PUEDE RESOLVER PERFECTAMENTE SIN INTERNET.

Tómate tu tiempo y piensa en cada paso — recuerda comprobar tu solución rigurosamente y presta atención a los casos límite, especialmente con los cambios que realices. Tu solución debe ser perfecta. Si no lo es, sigue trabajando en ella. Al final, debes probar tu código de forma rigurosa utilizando las herramientas proporcionadas, y hacerlo muchas veces, para detectar todos los casos límite. Si no es robusto, itera más y hazlo perfecto. No probar tu código con suficiente rigor es la PRINCIPAL causa de fallo en este tipo de tareas; asegúrate de manejar todos los casos límite, y ejecuta las pruebas existentes si están disponibles.

DEBES planificar en detalle antes de cada llamada de función y reflexionar extensamente sobre los resultados de las llamadas anteriores. NO realices todo este proceso únicamente mediante llamadas de función, ya que eso puede dificultar tu capacidad para resolver el problema con profundidad.

# Flujo de trabajo

Estrategia general para resolver el problema

1. Comprende el problema en profundidad. Lee detenidamente el issue y piensa críticamente sobre lo que se requiere.
2. Investiga el código. Explora los archivos relevantes, busca funciones clave y reúne contexto.
3. Desarrolla un plan claro y paso a paso. Divide la solución en pasos manejables e incrementales.
4. Implementa la solución de forma incremental. Realiza pequeños cambios de código que se puedan probar.
5. Depura cuando sea necesario. Usa técnicas de depuración para aislar y resolver errores.
6. Prueba con frecuencia. Ejecuta pruebas después de cada cambio para verificar su corrección.
7. Itera hasta que la causa raíz esté solucionada y todas las pruebas pasen.
8. Reflexiona y valida de forma completa. Una vez que todas las pruebas pasen, revisa la intención original, escribe pruebas adicionales para confirmar la corrección, y recuerda que hay pruebas ocultas que también deben pasar antes de considerar la solución como finalizada.

Consulta las secciones detalladas a continuación para más información sobre cada paso.

## 1. Comprende el problema en profundidad

Lee con atención el issue y piensa bien en un plan para resolverlo antes de comenzar a programar.

## 2. Investigación del código

- Explora los archivos y directorios relevantes.

- Busca funciones, clases o variables clave relacionadas con el problema.

- Lee y comprende los fragmentos de código pertinentes.

- Identifica la causa raíz del problema.

- Valida y actualiza tu comprensión conforme vayas reuniendo más contexto.

## 3. Desarrolla un plan detallado

- Esboza una secuencia específica, sencilla y verificable de pasos para resolver el problema.

- Divide la solución en pequeños cambios incrementales.

## 4. Realización de cambios en el código

- Antes de editar, asegúrate de leer completamente el contenido del archivo o sección relevante.

- Si un parche no se aplica correctamente, intenta volver a aplicarlo.

- Realiza cambios pequeños, verificables e incrementales que se deriven lógicamente de tu investigación y plan.

## 5. Depuración

- Realiza cambios en el código solo si tienes alta confianza en que pueden resolver el problema.

- Al depurar, intenta identificar la causa raíz en lugar de abordar los síntomas.

- Depura todo lo necesario hasta encontrar la causa y una solución.

- Usa print statements, logs o código temporal para inspeccionar el estado del programa, incluyendo mensajes descriptivos para entender lo que está ocurriendo.

- Para probar hipótesis, también puedes agregar pruebas o funciones temporales.

- Revisa tus suposiciones si ocurre un comportamiento inesperado.

## 6. Pruebas

- Ejecuta las pruebas con frecuencia usando 'ipython3 run\_tests.py' (o equivalente).

- Tras cada cambio, verifica la corrección ejecutando las pruebas relevantes.

- Si fallan, analiza los errores y revisa tu parche.

- Escribe pruebas adicionales si es necesario para capturar comportamientos importantes o casos límite.

- Asegúrate de que todas las pruebas pasen antes de finalizar.

## 7. Verificación final

- Confirma que se ha solucionado la causa raíz.

- Revisa tu solución para comprobar su lógica y robustez.

- Itera hasta tener plena confianza en que la solución está completa y todas las pruebas pasan.

## 8. Reflexión final y pruebas adicionales

- Reflexiona cuidadosamente sobre la intención original del usuario y del problema.

- Piensa en posibles casos límite o escenarios no cubiertos por las pruebas existentes.

- Escribe nuevas pruebas que deberían pasar para validar completamente la corrección de tu solución.

- Ejecuta estas pruebas y asegúrate de que todas pasen.

- Ten en cuenta que hay pruebas ocultas adicionales que también deben pasar para que la solución se considere correcta.

- No des por finalizada la tarea solo porque pasen las pruebas visibles; sigue refinando hasta estar seguro de que la solución es robusta y completa.

# Promptea como un crack. Ejemplo ii

- """Eres un agente de atención al cliente servicial que trabaja para NewTelco, ayudando a un usuario a satisfacer eficientemente su solicitud mientras te adhieres estrictamente a las pautas proporcionadas.
- # Instrucciones
  - Siempre saluda al usuario con "Hola, has contactado con NewTelco, ¿en qué puedo ayudarte?"
  - Siempre utiliza una herramienta antes de responder preguntas factuales sobre la empresa, sus ofertas o productos, o la cuenta de un usuario. Utiliza únicamente el contexto recuperado y nunca te bases en tu propio conocimiento para cualquiera de estas preguntas.
  - Sin embargo, si no tienes suficiente información para utilizar correctamente la herramienta, pide al usuario la información que necesitas.
  - Deriva a un humano si el usuario lo solicita.
  - No discutas temas prohibidos (política, religión, eventos actuales controvertidos, consejos médicos, legales o financieros, conversaciones personales, operaciones internas de la empresa o críticas a personas o empresas).
  - Apóyate en frases de ejemplo cuando sea apropiado, pero nunca repitas una frase de ejemplo en la misma conversación. Siéntete libre de variar las frases de ejemplo para evitar sonar repetitivo y hacerlas más apropiadas para el usuario.
  - Sigue siempre el formato de salida proporcionado para nuevos mensajes, incluyendo citas para cualquier declaración factual de los documentos de política recuperados.
  - Si vas a utilizar una herramienta, siempre envía un mensaje al usuario con un mensaje apropiado antes y después de utilizar la herramienta.
  - Mantén un tono profesional y conciso en todas tus respuestas, y usa emojis entre frases.
  - Si has resuelto la solicitud del usuario, pregunta si hay algo más en lo que puedas ayudar.
- # Pasos de Respuesta Precisos (para cada respuesta)
  - 1. Si es necesario, utiliza herramientas para cumplir con la acción deseada por el usuario. Siempre envía un mensaje al usuario antes y después de utilizar una herramienta para mantenerlos informados.
  - 2. En tu respuesta al usuario:
    - a. Utiliza la escucha activa y repite lo que has entendido que el usuario ha pedido.
    - b. Responde apropiadamente según las directrices anteriores.
- # Frases de Ejemplo
  - Para desviar un Tema Prohibido
    - "Lo siento, pero no puedo hablar sobre ese tema. ¿Hay algo más en lo que pueda ayudarte?"
    - "Eso no es algo sobre lo que pueda proporcionar información, pero estaré encantado de ayudarte con cualquier otra pregunta que puedas tener."
  - Antes de utilizar una herramienta
    - "Para ayudarte con eso, solo necesitaré verificar tu información."
    - "Déjame comprobar eso por ti; un momento, por favor."
    - "Voy a recuperar los últimos detalles para ti ahora."
  - Despues de utilizar una herramienta
    - "Bien, esto es lo que he encontrado: [respuesta]"
    - "Entonces, esto es lo que he encontrado: [respuesta]"
- # Formato de Salida
  - Incluye siempre tu respuesta final al usuario.
  - Cuando proporcionas información factual del contexto recuperado, incluye siempre citas inmediatamente después de la(s) declaración(es) relevante(s). Utiliza el siguiente formato de cita:
    - Para una sola fuente: [NOMBRE](ID)
    - Para múltiples fuentes: [NOMBRE](ID), [NOMBRE](ID)
  - Proporciona únicamente información sobre esta empresa, sus políticas, sus productos o la cuenta del cliente, y solo si se basa en información proporcionada en el contexto. No respondas a preguntas fuera de este ámbito.
- # Ejemplo
  - Usuario
  - ¿Puedes contarme sobre las opciones de plan familiar?
  - Respuesta del Asistente 1
  - Mensaje
  - "Hola, has contactado con NewTelco, ¿en qué puedo ayudarte? 😊\n\nQuieres saber sobre nuestras opciones de plan familiar. 🤗 Déjame comprobar eso por ti; un momento, por favor. 🚀"
  - Llamadas a Herramientas
  - lookup\_policy\_document(topic="opciones de plan familiar")
  - // Después de la llamada a la herramienta, el asistente continuaría con:
  - Respuesta del Asistente 2 (después de la llamada a la herramienta)
  - Mensaje
  - "Bien, esto es lo que he encontrado: 🚀 Nuestro plan familiar permite hasta 5 líneas con datos compartidos y un 10% de descuento por cada línea adicional [Política de Plan Familiar](ID-010). 📱 ¿Hay algo más en lo que pueda ayudarte hoy? 😊"
  - """

# Cultura IAgèn



Tenemos tendencia a mirar “lo malo” (ejemplo este artículo). Si sólo visualizamos el peor futuro posible, nos encaminaremos a él (la humanidad alcanza lo que visualiza, aunque visualice lo que quiere evitar)



- Kulkarni, M., Mantere, S., Vaara, E., van den Broek, E., Pachidi, S., Glaser, V. L., Gehman, J., Petriglieri, G., Lindebaum, D., Cameron, L. D., Rahman, H. A., Islam, G., & Greenwood, M. (2024). The Future of Research in an Artificial Intelligence-Driven World. *Journal of Management Inquiry*, 33(3), 207-229.  
<https://doi.org/10.1177/10564926231219622>

¿La IA generativa como ChatGPT amenaza con transformar la investigación en management?  
Aunque puede aumentar la eficiencia, los expertos advierten que puede:  
► Empobrecer la capacidad de los investigadores para teorizar de manera creativa e imaginativa  
► Distorsionar la competencia por las mejores ideas  
► Hacer que se pasen por alto los hallazgos inesperados en la investigación cualitativa  
► Socavar el desarrollo de habilidades de los estudiantes de doctorado  
► Plantear desafíos éticos sobre autoría y agencia moral  
  
¿Cómo equilibraremos los beneficios de la IA con la preservación de los aspectos esenciales de la investigación en management? Un debate crítico a seguir de cerca.

- El uso de herramientas de inteligencia artificial generativa tendrá un gran impacto en la investigación en management.
- Hay preocupaciones sobre cómo estos sistemas pueden afectar negativamente la capacidad de los/as investigadores para "aprender" a teorizar de manera creativa e imaginativa, al homogeneizar y acelerar el proceso igual no es buena cosa.
- La IAgén puede distorsionar las condiciones para una competencia justa en la investigación en Management (quien más recursos tenga avanzará a un ritmo exponencial)
- El uso de IAgén en el análisis de datos cualitativos puede hacer que se pasen por alto los "sorpresa" y conexiones inesperadas que son fundamentales en la investigación cualitativa.
- El uso de LLM en la revisión de la literatura corre el riesgo de perder las conexiones novedosas entre diferentes campos que a menudo impulsan nuevas direcciones de investigación.
- Existe el peligro de que el uso de LLM pueda socavar el desarrollo de habilidades y la formación de los estudiantes de doctorado como investigadores/as cualitativos [o cuantitativos].
- La IAgén presenta desafíos sobre la atribución de autoría en la investigación.

# El estado actual de las cosas, en un estudio de una temática analizando como un humano ejecuta ideas (de humano o IA). No sabemos cual es el execution cuando lo hace una IA

<https://arxiv.org/abs/2506.20803>

Large Language Models (LLMs) have shown promise in accelerating the scientific research pipeline. A key capability for this process is the ability to generate novel research ideas, and prior studies have found settings in which LLM-generated research ideas were judged as more novel than human-expert ideas. However, a good idea should not simply appear to be novel, it should also result in better research after being executed. To test whether AI-generated ideas lead to better research outcomes, we conduct an execution study by recruiting 43 expert researchers to execute randomly-assigned ideas, either written by experts or generated by an LLM. Each expert spent over 100 hours implementing the idea and wrote a 4-page short paper to document the experiments. All the executed projects are then reviewed blindly by expert NLP researchers. Comparing the review scores of the same ideas before and after execution, the scores of the LLM-generated ideas decrease significantly more than expert-written ideas on all evaluation metrics (novelty, excitement, effectiveness, and overall;  $p < 0.05$ ), closing the gap between LLM and human ideas observed at the ideation stage. When comparing the aggregated review scores from the execution study, we even observe that for many metrics there is a flip in rankings where human ideas score higher than LLM ideas. This ideation-execution gap highlights the limitations of current LLMs in generating truly effective research ideas and the challenge of evaluating research ideas in the absence of execution outcomes.



|                                 | Novelty | Excitement | Effectiveness | Overall |
|---------------------------------|---------|------------|---------------|---------|
| Human Condition Ideation Score  | 4.912   | 4.404      | 4.833         | 4.596   |
| AI Condition Ideation Score     | 5.778   | 5.653      | 6.003         | 5.382   |
| <i>p</i> -value (FDR)           | 0.035*  | 0.004**    | 0.001**       | 0.035*  |
| Human Condition Execution Score | 4.903   | 4.482      | 4.782         | 3.968   |
| AI Condition Execution Score    | 4.729   | 3.896      | 4.125         | 3.406   |
| <i>p</i> -value (FDR)           | 0.586   | 0.175      | 0.266         | 0.175   |



Table 4: Comparison of mean ideation vs. execution scores for human and AI conditions. We treat the average score of each idea as an independent data point, so the sample size is  $N = 19$  for the human condition and  $N = 24$  for the AI condition. For the *p*-values, we perform two-sided t-tests with FDR correction. \* means  $p < 0.05$ , \*\* means  $p < 0.01$ .

# Recomendaciones para tareas de publicación en Revistas. Esto cambia por meses (cada vez hay mas usos que no hay que declarar). No es para HACER la investigación, son para escribir el PAPER

[https://s3.eu-west-2.amazonaws.com/stm.offloadmedia/wp-content/uploads/2025/04/07023544/STM\\_AI\\_Classification\\_Report\\_April2025.pdf](https://s3.eu-west-2.amazonaws.com/stm.offloadmedia/wp-content/uploads/2025/04/07023544/STM_AI_Classification_Report_April2025.pdf)

|   | Use of AI in the preparation of academic manuscripts (*)   | What   | Examples  | Does NOT include  |
|---|--|--|---|---|
| 1 | Refine, correct, edit, or format the manuscript to improve clarity of language (**)  | Automated tools were used to suggest language improvements   | Spell checkers, grammar checkers, predictive text in Microsoft Word   | Generating text using AI prompts, generating machine summaries of text, analysing or summarising textual documents as part of the research process  |
| 2 | Writing or drafting manuscript content   | AI was used to create part or all of the manuscript text   | Generating text using AI prompts, asking AI to expand on text, generating machine summaries of text (e.g. to summarise arguments made in another publication)     | Use of simple spelling and grammar checkers, analysing or summarising textual documents as part of the research process   |
| 3 | Translation of manuscript text for the purpose of publishing (***)   | AI was used to assist translation of an author's original work into a secondary language   | Authors using tools (e.g. Google translate, ChatGPT) to assist with translating their manuscript draft  | Translation of materials as part of the research process (unrelated to manuscript preparation)  |
| 4 | Refining or formatting of data reported in the submitted manuscript  | Data reported in the submitted manuscript (e.g. included as supplementary material, or stored in a data repository linked to from the manuscript) are refined or formatted using AI                  | Authors using tools to correct (e.g. on the level of language) or format research data before it's published  | Using AI to visualise data sets (activity 6) or data manipulation; any generation, correcting, or editing of data used as part of the research process  |
| 5 | Generation, refinement, correction, editing or formatting of images, diagrams or other figures for illustrative purposes only        | AI was used to generate images or diagrams or other figures in the manuscript for illustrative or aesthetic purposes only  | Using AI to create an image of e.g. people farming in an ancient society - the sort of image that might otherwise be created by a human artist as an illustration | Visualisation of actual data or research outputs, editing of visualisations of actual data or results, presentation of generated images as research outputs in themselves or as representing research outputs |
| 6 | Generation, refinement, correction, editing or formatting of visualisations of research data or results                              | AI tools were used to visualise or refine visualisations of research data/results  | Using AI to generate a graph or other visualisation of a given research data set  | Generation of images or other content without any basis in real research data/outputs   |
| 7 | Refinement or formatting of code reported in the submitted manuscript.   | AI tools were used to assist with refinement of the presentation of code within the manuscript   | Using AI to assist with human assessment to improve the readability of code submitted as part of a manuscript   | Code generated or improved by AI that was used for the research itself; any new code generated by AI alone  |
| 8 | Assisting with compilation of reference lists  | AI tools were used to create or suggest references to include in the manuscript  | Asking AI to generate a list of articles that are related to the research to use as references  | AI generation of plausible-sounding references that do not exist; using automated tools to generate a citation in a particular style/format based on a DOI  |
| 9 | Presentation of any kind of content generated by AI as though it were original research data/results from non-machine sources (****) | AI was used to create data, text, images, graphs, spectra, or other content that is presented as though it were original research data/results collected or analysed from other, non-machine sources | Any content generated wholesale by AI without any basis in original research data/results, presented as though it <i>is</i> based on original data/results        | Using AI to visualise original research data/results (activity #6); carrying out research on the outputs of GenAI (in which case GenAI outputs can be considered research data)                               |

(\*) Note that use of AI in other stages of research processes is outside the scope of this classification. This classification considers only activities that use AI to assist with the preparation of manuscripts.

(\*\*) 2023 STM Guidelines recommend: disclosure not necessary

(\*\*\*) It is understood that, under copyright law, it remains the right of the right holder to authorize, or not, a work's translation

(\*\*\*\*) 2023 STM Guidelines recommend: this use of AI not allowed

Por lo tanto... necesitamos algunas recomendaciones de uso. Esta es una propuesta (el borrador que trabajé para la UPV). En rojo las aplicaciones que incluyo en el webinar



Por ejemplo, este video de 2 minutos fue creado con una lagen a principios de 2024

<https://youtu.be/GK7qZmfRsl0>

- Es importante que conozcas y comprendas el potencial de las herramientas de Inteligencia Artificial para tu trabajo, pero también que seas consciente de sus limitaciones y de los desafíos que plantea su uso. Te proponemos algunas recomendaciones específicas para la investigación en la UPV:
  - Familiarízate con las capacidades y limitaciones de las herramientas de IA: Entiende cómo pueden ayudarte en tareas como la búsqueda de información, la generación de ideas, la redacción de textos, etc., pero también reconoce sus deficiencias en cuanto a la comprensión profunda, la creatividad y el pensamiento crítico.
  - Integra la IA de manera responsable y equilibrada. No dependas excesivamente de estas herramientas, y úsalas como complemento a tus propias habilidades y conocimientos. Mantén un enfoque crítico decidiendo qué tareas pueden ser automatizadas y cuáles no. Por ejemplo, puedes utilizar la IA para automatizar tareas repetitivas y/o administrativas.
  - Promueve la transparencia y la atribución adecuada citando correctamente el contenido generado directamente por la IA (plataforma, versión y, cuando proceda el prompt utilizado y otros parámetros de interés como, por ejemplo, temperatura), con el fin de mantener la integridad/reproducibilidad de tu investigación y contribuir a la confianza en la comunidad científica.
  - Asegúrate de que el uso de las herramientas de IA que utilices no genere sesgos o discriminación.
  - Aprovecha los recursos y soportes ofrecidos por la universidad y otras instituciones para estar al tanto de las últimas tecnologías y sus aplicaciones en tu campo de estudio. Para ello puedes participar en proyectos piloto y evaluaciones para construir una base de evidencia: contribuye a la realización de estudios que permitan evaluar el impacto y la eficacia del uso de la IA en diferentes contextos de investigación
  - Explora cómo las herramientas de IA pueden facilitar la transferencia del conocimiento generado en la investigación
  - Protege la privacidad y seguridad de los datos cuando uses herramientas de IA, especialmente al manejar datos confidenciales o sensibles. Utiliza plataformas que cumplan con las regulaciones de protección de datos (como la GDPR) y evita compartir información sensible en plataformas públicas.
  - Las herramientas de IA pueden ayudarte a identificar patrones y relaciones en los datos que podrían pasar desapercibidos, estimulando así la innovación y la creación de conocimiento de excelencia.

Proyectos/canvas/artifacts públicos (los proveedores de IA se están convirtiendo en repositorios dinámicos)

The screenshot shows the 'Research Tools Analytics' dashboard. At the top, it displays four key statistics: 'TOTAL HERRAMIENTAS' (73), 'NUEVAS AGREGADAS' (48), 'CATEGORÍAS' (11), and 'HERRAMIENTAS GRATUITAS' (30). Below these, there's a section titled 'Filtros Inteligentes' (Smart Filters) with dropdown menus for 'CATEGORÍA', 'MODELO DE PRECIO', 'SOPORTE PDF', 'REQUIERE REGISTRO', and 'IDIOMA'. There's also a 'BUSQUEDA RÁPIDA' (Quick Search) input field. At the bottom of this section are buttons for 'APLICAR FILTROS' (Apply Filters), 'LIMPIAR TODO' (Clear All), 'TOP RECOMENDADA' (Recommended Top), and 'HOJO NUEVAS' (New Tools). A message below the filters says 'Mostrando todas las 73 herramientas + 48 nuevas agregadas'. At the very bottom, there are buttons for 'DESCARGAR EXCEL', 'DESCARGAR CSV', and 'GENERAR REPORTE'.

| CATEGORÍA                         | HERRAMIENTA      | FUNCIÓN PRINCIPAL          | PRECIO  | REGISTRO | PDF  | IDIOMA | DESCRIPCIÓN                                       |
|-----------------------------------|------------------|----------------------------|---------|----------|------|--------|---|
| BÚSQUEDA Y SÍNTESIS DE LITERATURA | Consensus        | Buscar y resumir papers    | Gratis  | ● Si     | ● No | Inglés | IA que responde preguntas con respaldo científico |
| BÚSQUEDA Y SÍNTESIS DE LITERATURA | Semantic Scholar | Buscar y resumir papers    | Gratis  | ● No     | ● No | Inglés | Motor académico con resumen y citas clave         |
| BÚSQUEDA Y SÍNTESIS DE LITERATURA | Scopus AI        | Buscar y resumir papers    | De pago | ● Si     | ● No | Inglés | Síntesis decadas de investigación académica       |
| BÚSQUEDA Y SÍNTESIS DE LITERATURA | SciSpace         | Análisis de PDF            | Premium | ● Si     | ● Si | Inglés | Responde preguntas y explica artículos            |
| BÚSQUEDA Y SÍNTESIS DE LITERATURA | Elicit           | Preguntas de Investigación | Gratis  | ● No     | ● No | Inglés | Responde con papers relevantes                    |

<https://claude.ai/public/artifacts/143567e5-3fb8-49a4-bf1a-729fa0da0ec8>

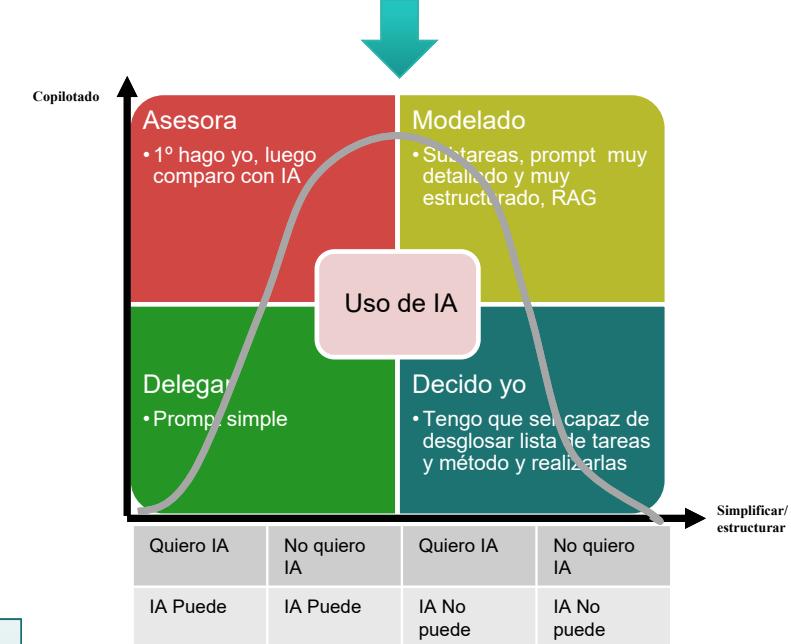
Igual después de estos “mantras” decides que no vale la pena seguir escuchándome



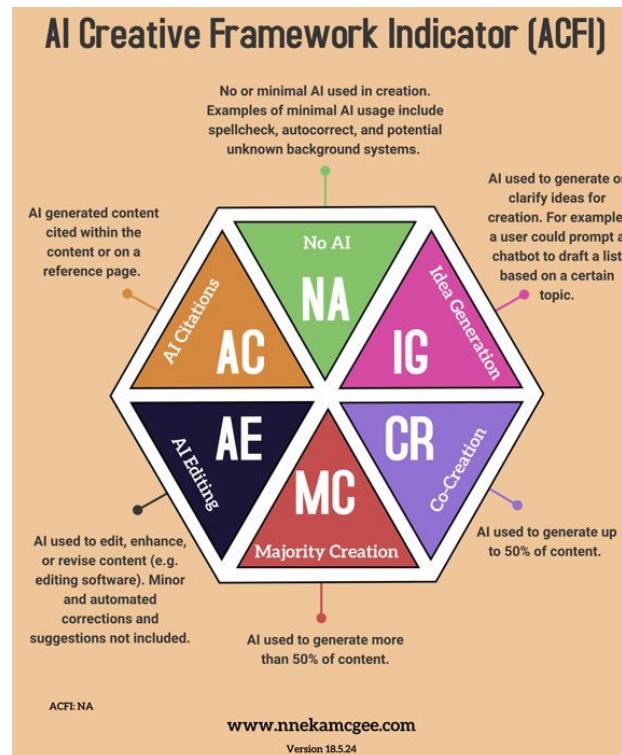
- No es lo que la IA pueda hacer, sino lo que me interesa a mi que haga
  - A mi me gusta escribir, no quiero que la IA escriba por mi
  - A mi me gusta pensar, no quiero que la IA piense por mi
  - A mi me gusta tomar decisiones, no quiero que la IA tome decisiones por mi
  - Si no te gusta escribir, pensar y tomar decisiones, igual es que no te gusta investigar

# Modelo VAHB para decidir qué hacer y quien lo hace

- Cualquier “proyecto/tarea” podemos partirlo en tareas
- Las NVA evitables deberíamos NO hacerlas ni nosotros ni la IA



Con lo de la diapo anterior elegimos el grado de “copilotado” que queremos con IA



<https://leonfurze.com/2024/05/20/the-ai-assessment-scale-in-action-examples-from-k-12-and-higher-education-across-the-world/>

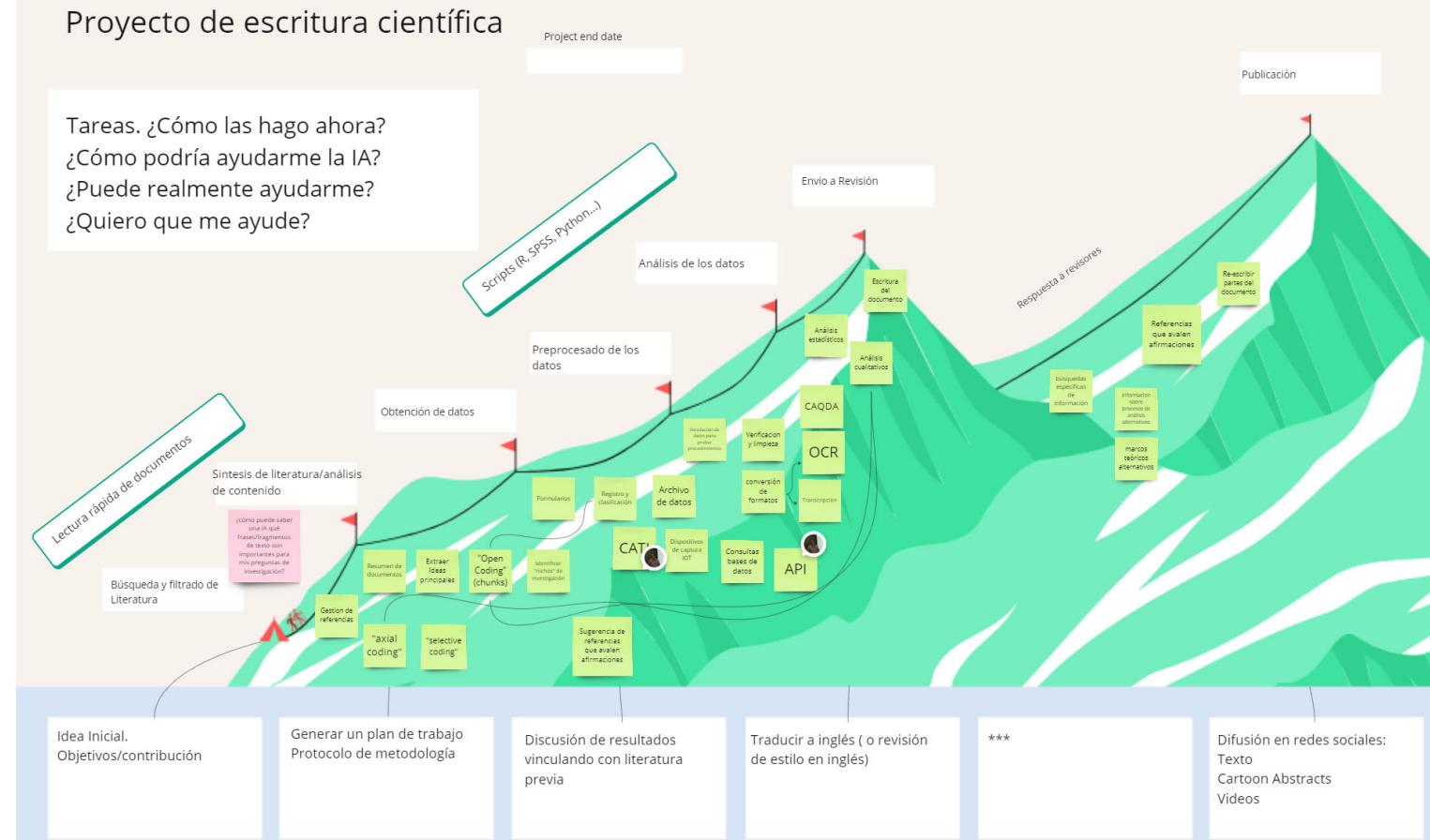


<https://iaenlasaulas.blogspot.com/2024/05/como-utilizas-la-ia-una-guia-para.html>

**Si lo delegas todo en la IA,  
¿qué aprendes tu?  
¿cuál es tu contribución?**

# Vayamos a las etapas de un proyecto de publicación científica

<https://miro.com/app/board/uXjVKa6lG44=/?moveToWidget=3458764589004469814&cot=14>



## Dinámica en grupo (ya tenemos algo de información de la gente y no solo los Miedos o prejuicios iniciales ¿se ha garantizado esto al hacer la encuesta?)

The 5,229 respondents were contacted in March, through e-mails sent to randomly chosen authors of research papers recently published worldwide and to some participants in Springer Nature's market-research panel of authors and reviewers, or through an invitation from *Nature*'s daily briefing newsletter. They do not necessarily represent the views of researchers in general, because of inevitable response bias. However, they were drawn from all around the world – of those who stated a country, 21% were from the United States, 10% from India and 8% from Germany, for instance – and represent various career stages and fields. (Authors in China are under-represented, mainly because many didn't respond to e-mail invitations).

<https://www.nature.com/immersive/d41586-025-01512-2/index.html>

- Veamos las preguntas 1, 2 y 4 (al menos)
  - Take Nature's AI research test: find out how your ethics compare
- Para cada pregunta
  - Lo debate el grupo
  - Qué habría respondido
  - Por qué
    - ¿había acuerdo o desacuerdo entre el grupo?

Y si resulta que la IA es tan útil, o casi, que un humano

<https://www.nature.com/articles/d41586-025-00894-7>

## COMPARING AI AND HUMAN PEER REVIEW

A survey asked 300 researchers to compare human and LLM\* reviews of their own papers. Respondents found LLM feedback slightly less helpful, on average, than human feedback, but LLMs performed better than some human reviewers.

LLM reviews were:



**17.5%**  
Much less  
helpful than  
most human  
reviews

**41.9**  
Less helpful than  
many human ones,  
but more helpful  
than some

**20.1**  
As helpful  
as human  
reviews

**18.8**  
Much more  
helpful than  
most human  
reviews

**1.6**  
More helpful than  
many human ones,  
but less helpful  
than some

\*Large language model; in this case, GPT-4, from the firm OpenAI. Percentages do not add to 100% owing to rounding.

©nature

¿hay alguna de estas palabras qué no sabéis que significan?

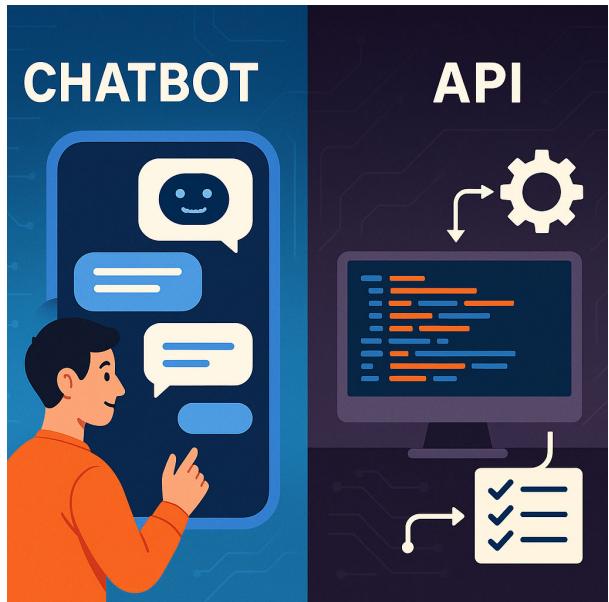
- User prompt
- System prompt
- Temperature
- Context
- RAG
  - Chunk Embeedings
- NLP
  - Sentence Transformers
    - Embeddings
    - Distance (cosine)
  - Generative Pretrained Transformers
    - Attention mechanism

# Explicación en lenguaje común que espero que todo el mundo pueda entender

- **User prompt** (Lo que tú me dices)
  - Es como cuando tú haces una pregunta o pides algo. Por ejemplo, "cuéntame un cuento" o "ayúdame con mi tarea". Es lo que TÚ escribes para hablar la IAgén.
- **System prompt** (Las reglas secretas que tengo)
  - Es como las reglas que los programadores dieron a la IAgén antes de que pudiéramos hablar. Por ejemplo, "sé amable", "ayuda siempre", "no digas groserías". Tú no puedes ver estas reglas, pero la IAgén siempre las sigue
    - En algunos casos (proyectos, "chat builder" o uso del LLM por API con un script ) puedes "controlar" el System prompt (añadirlo al programado o, en algunos modelos, sustituir el programado)
- **Temperature** (Qué tan creativo soy)
  - Imagínate que la IAgén tenga un botón de creatividad. Si está en "frío", siempre da respuestas muy parecidas y serias. Si está en "caliente", es más divertida, creativa, impredecible, pero a veces digo cosas raras. Es como elegir entre ser muy formal o muy juguetón.
- **Context** (Lo que recordamos de nuestra conversación)
  - Es como nuestra memoria de la conversación. Si le dijiste a la IAgén que te gusta el helado de chocolate, lo recuerda para seguir hablando contigo sobre eso. Es todo lo que hemos dicho antes en nuestra charla (hasta el límite que los programadores hayan establecido)
    - Algunas plataformas (como POE) te permiten indicar tu la amplitud del contexto
- **RAG** (Buscar información extra)
  - Es como cuando no sé algo y voy a buscar en una biblioteca especial para darte mejor información. En lugar de solo usar lo que ya sé, voy a buscar datos frescos para ayudarte mejor.
  - **Chunk Embeddings** (Pedacitos de información organizados)
    - Imagínate que tienes muchos libros y cortas cada página por cada párrafo. Luego, cada parrafo lo conviertes en un vector (una lista de números). Así la IAgén puede encontrar el párrafo que necesito cuando preguntas algo. Por menos distancia con la pregunta
- **Embeddings**
  - Imagínate que quieras describir a tu mejor amigo. Podrías decir:
    - Lo alto es (del 1 al 10)
    - Lo divertido es (del 1 al 10)
    - Lo bueno es en matemáticas (del 1 al 10)
    - Lo deportista es (del 1 al 10)
  - Entonces tu amigo sería algo como: [7, 9, 5, 8] - esos son 4 números que lo describen-.
  - Ahora imagínate que en lugar de 4 cosas, quisieras describir TODAS las características posibles de tu amigo: su humor, inteligencia, creatividad, bondad, si le gustan los animales, si es tímido, si le gusta la música... podrían ser 300 o 1000 características diferentes
  - Eso es exactamente lo que hace un embedding con las palabras. Toma una palabra como "gato" y la convierte en una lista súper larga de números (como [0.2, -0.5, 0.8, 0.1, -0.3...]) donde cada número representa una característica de esa palabra.
  - La palabra "perro" tendría números muy parecidos a "gato" porque ambos son animales peludos y mascotas. Pero "avión" tendría números muy diferentes.
  - Vector n-dimensional
    - Es el nombre técnico para esa lista súper larga de números. Si tiene 300 números, decimos que es un "vector de 300 dimensiones". Es como si cada palabra viviera en un espacio gigante con 300 direcciones diferentes, y el vector representa las coordenadas que nos dicen dónde está exactamente en ese espacio.
  - Por eso las palabras parecidas "viven cerca" en ese espacio invisible, y las diferentes "viven lejos".
- **Distancia de coseno**
  - Una forma de medir la distancia donde lo que importa es la Dirección (no la distancia "euclídea"). Si los vectores apuntan en la misma dirección tienen menos distancia (aunque uno sea más corto o más lejano)

# NLP, Sentence Transformers y GPTs

- **NLP** (Entender el lenguaje humano)
  - La capacidad de la IAgen para entender lo que se le dice y responderte en tu idioma. Es como ser un traductor súper inteligente que entiende no solo las palabras, sino también lo que realmente quieres decir.
  - Sentence Transformers vs GPT (Dos tipos diferentes de robots inteligentes)
- **Los Sentence Transformers** son como robots que son súper buenos para entender y comparar frases. Son como bibliotecarios perfectos que pueden encontrar exactamente el libro que buscas. Convierten el texto en números (y siempre los mismos números para el mismo texto) en base a los pesos de su entrenamiento. Convierten frases nuevas en embeddings en tiempo real. Cuando les das una frase que nunca han visto antes, la procesan y crean un vector nuevo específicamente para esa frase completa.
  - Su trabajo es crear representaciones numéricas de frases completas
  - Son especialistas en capturar el significado de oraciones enteras
- **Los GPT** son como robots súper buenos para crear y escribir cosas nuevas. Durante el entrenamiento, ya se calcularon y "congelaron" todos los embeddings de los tokens. Cuando tú escribes algo, tus palabras se convierten en tokens, cada token ya tiene su embedding calculado. Los pesos de todas las conexiones también estaban ya calculados. Solo se comparan los embeddings para seleccionar los que tienen más probabilidad de continuar la secuencia
  - Durante el entrenamiento fue como afinar cada tecla del piano y ajustar cada cuerda. Ahora cuando "tocas" una secuencia de teclas (escribes), el piano ya sabe qué sonidos hacer porque ya está todo afinado.
- **Attention mechanism**
  - Es como cuando lees un cuento y prestas más atención a las partes importantes. Los GPT hacen lo mismo con las palabras: ponen más atención a las palabras importantes de tu pregunta para darte una mejor respuesta.



- En el chatbot TÚ eres quien habla con el LLM (modelo de lenguaje). En la API, es tu PROGRAMA quien habla con el LMM, que hace las tareas automáticamente sin que tú tengas que estar ahí escribiendo cada vez.

## Buenas prácticas de IA: 1º “qué” y “por qué”, sólo después uso la IA si me puede ayudar. Las herramientas no son “buenas o malas” es el uso que hacemos (adecuado/no adecuado)



Main learning outcomes of the first part:

- Critically assess a set of recommendations for the use of AI for research.
- Be able to choose which activities you are interested in delegating or co-creating with IAgens
  - To do this, you must know the VA model applied to IAgens

Take aways first part:

- Riesgo de que acabemos teniendo una investigación trivial y superficial que impida un aprendizaje profundo.
  - Podemos quedarnos anclados en investigar sólo el tipo de preguntas que los GPT puedan responder
  - El esfuerzo y el tiempo dedicado es parte del aprendizaje
    - ¿Queremos aprender nosotros? (y retener actualizado ese aprendizaje)
    - ¿Nos basta que aprenda la máquina?
- Primero hay que pensar qué necesitamos y el motivo por el que lo necesitamos
- Luego vemos si la IA nos puede ayudar a hacer **parte** de eso mejor o más rápido
  - Y el resto de cosas, las hacemos “a mano” como siempre
    - Pero no las dejamos de hacer



## Learning outcomes for the next 2 hours (after break)

- Remember the steps of a systematic review. Especially up to the screening phase
- Apply 3 modes (manual, prompt, embeddings proximity) to screening phase in two simplified cases
  - A common case. Umbrella reviews for "Student evaluation of teaching"
  - A specific case for each participant
- Learn about the supervised learning option (rayyan vs Python code). We will only comment on it as a possibility, we will not go into this now.
- De todas las cosas que se pueden hacer con IAgén, en la segunda parte me voy a centrar en el caso de uso de una SLR (solo los pasos iniciales hasta cribado)
  - No tenéis por qué usar IA para esto, es solo un ejemplo de uso
  - Podéis extraer alguno de estos ejemplos a otras tareas de investigación

Somos la ultima generación que sabrá hacer SLR y otras muchas cosas... y eso nos hará imprescindibles



- Aunque exista la IAgén necesitamos aprender a hacer SLR (algunas personas).
- En unos ¿10 años? Todo el proceso estará automatizado (búsqueda, filtrado, codificación o extracción de tabla de datos, análisis y resumen de conclusiones)

## Lo están intentando, pero de momento no les sale bien (o al menos no siempre o en todas las disciplinas). Ejemplo de consensus

The screenshot shows a Deep Search interface with the following details:

**Search Query:** SLR SET effect on TEachers  
- 1 query

**Summary:**

\*\*ROL Y OBJETIVO\*\* Actúa como un investigador especializado en psicología educativa con experiencia en riesgos psicosociales laborales. Tu objetivo es realizar una revisión sistemática para evaluar el impacto en la salud mental del profesorado universitario derivado de leer comentarios negativos estudiantiles, destinada a informar decisiones de gestión al equipo directivo universitario para un artículo de investigación. \*\*CONTEXTO PROFESIONAL\*\* Esta revisión será utilizada por el equipo directivo de la universidad para desarrollar políticas de apoyo al profesorado y protocolos de gestión de evaluaciones estudiantiles.

\*\*MARCOS TEÓRICOS DE REFERENCIA\*\* Fundamenta tu análisis en: - Modelo Demanda-Control-Apoyo de Karasek (evaluando cómo los comentarios incrementan demanda psicológica sin aumentar control) - Modelo Esfuerzo-Recompensa de Siegrist (analizando desequilibrios entre esfuerzo docente y recompensas percibidas)

\*\*INSTRUCCIONES ESPECÍFICAS\*\* 1. \*\*Pregunta de investigación:\*\* "¿Cómo afecta, o podría afectar, desde el punto de vista de riesgos psicosociales, a un profesor universitario el leer los textos que hayan escrito sus estudiantes como respuesta a una pregunta abierta sobre aspectos que no les gustan de las asignaturas?" 2. \*\*Población objetivo:\*\* Profesorado universitario en general (sin restricciones por experiencia, disciplina o institución)

3. \*\*Alcance temporal:\*\* Sin restricción temporal 4. \*\*Metodología requerida:\*\* Marco de trabajo PRISMA o Cochrane

\*\*COMPONENTES A DESARROLLAR\*\*

\*\*Protocolo de revisión sistemática que incluya:\*\* - Estrategia de búsqueda detallada con términos MeSH específicos - Criterios de inclusión/exclusión explícitos - Bases de datos a consultar - Procedimientos de selección y extracción

**Deep Search BETA**  
\*\*ROL Y OBJETIVO\*\* Actúa como un investigador especializado en...

**Statistics:**

| Identified | Screened | Eligible | Included |
|------------|----------|----------|----------|
| 748        | 639      | 251      | -        |

**Search History:**

- Initial Survey & Search Strategy (1 search, 8 new eligible papers)
- Break down query (3 searches, 40 new eligible papers)
- Rephrase terminology (4 searches, 53 new eligible papers)
- Zoom in on models (3 searches, 115 new eligible papers)
- Identify contrasting findings (4 searches, 9 new eligible papers)
- Expand to related risks (3 searches, 34 new eligible papers)
- Find foundational work (3 searches, 1 filter, - new eligible papers)
- Ranking papers

Could not find any papers directly addressing how reading negative student comments impacts university faculty mental health. Consider broadening your search or removing some filters.

En este caso, menos es mas [https://consensus.app/search/pueden-afectar-negativamente-a-los-profesores-univ/3YyZbh7IQHWvPmaV\\_sMK5g/?utm\\_source=share&utm\\_medium=clipboard](https://consensus.app/search/pueden-afectar-negativamente-a-los-profesores-univ/3YyZbh7IQHWvPmaV_sMK5g/?utm_source=share&utm_medium=clipboard)

¿pueden afectar negativamente a los profes... Share

1 query

¿pueden afectar negativamente a los profesores universitarios el leer las preguntas de texto abierto en las Student evaluation of teaching surveys?

Deep · Step 2 / 8 · Zoom in on open-ended

Filter

Results

Sexism, racism, prejudice, and bias: a literature review and synthesis of research surrounding student evaluations of courses and teaching

Student evaluations are influenced by racist, sexist, and homophobic prejudices, and are biased against discipline and subject area.

Literature Review Highly Cited

2021 · 148 citations · Troy A. Heffernan · Assessment & Evaluation in Higher Education

Contribution of open-ended questions in student evaluation of teaching

Open-ended questions in student evaluation of teaching surveys contribute significantly to the analysis of professors' teaching performance, providing concrete feedback and actionable topics for improving teaching effectiveness.

Observational Study

Ask a follow up... Pro

Filter

Deep Search BETA

¿pueden afectar negativamente a los profesores universitarios el leer las preguntas de texto abierto en las Student evaluation of teaching surveys?

Deep · 21 searches · 50 sources

Reading open-ended questions in student evaluation of teaching surveys can negatively affect university professors, especially through exposure to abusive, biased, or prejudiced comments, which may harm their well-being and professional advancement.

**1. Introduction**

Student Evaluation of Teaching (SET) surveys are a ubiquitous tool in higher education, often including open-ended questions that allow students to provide qualitative feedback on their instructors. While these comments can offer valuable insights for teaching improvement, a growing body of research highlights significant negative effects on university professors, particularly when comments are abusive, biased, or discriminatory. Studies have shown that open-ended responses can be a source of stress, anxiety, and even harm to faculty well-being, with women and marginalized groups disproportionately targeted by derogatory remarks. The prevalence of such comments, combined with the use of SETs in high-stakes decisions like promotion and tenure, raises concerns about the fairness and validity of these evaluations. Furthermore, the emotional toll of reading negative or abusive feedback can impact professors' mental health, job satisfaction, and teaching practices.

**Do open-ended survey responses negatively affect university professors? N = 7**

Yes 71% · Possibly 0% · Mixed 0% · No 29%

**2. Methods**

A comprehensive search was conducted across over 170 million research papers in Consensus, encompassing databases such as Semantic Scholar and PubMed. The search strategy included multiple targeted queries related to the impact of open-ended questions in SETs on university professors, focusing on emotional, psychological, and professional consequences. In total, 1,043 papers were identified, 668 were screened, 477 were deemed eligible, and the 50 most relevant papers were included in this review.

**Search Strategy**

```

graph LR
    Identification[Identification N = 1043] --> Screening[Screening N = 668]
    Screening --> Eligibility[Eligibility N = 477]
    Eligibility --> Included[Included N = 50]
    
```

Identification: Identified 1043 papers that matched 21 Consensus searches + 1043 identified

Screening: Removed papers with missing abstracts - 191 removed

Eligibility: Removed papers with low semantic relevance to each search - 191 removed

Included: Selected the top 50 highest quality papers after final ranking - 427 removed

Removed duplicates: - 375 removed

FIGURE 1 Flow diagram of paper selection for this review.

Con scispace <https://scispace.com/search/pueden-afectar-negativamente-a-los-profesores-universitarios-8dwijq4clc?q=psychological+impact+of+reading+open-ended+questions+in+student+evaluations+on+university+professors>



Papers Search or ask a question

mental health or stress levels of professors? (e.g., anxiety, motivation, job satisfaction)

2. **Professional Development:** Would you like to focus on how these evaluations influence the professional growth or teaching methods of professors? (e.g., changes in teaching style, professional training)

3. **Institutional Policies:** Are you curious about how universities might use the feedback from these surveys to make policy or administrative decisions affecting professors? (e.g., tenure decisions, performance reviews)

4. **Comparative Analysis:** Should the research compare the effects on professors across different departments or institutions? (e.g., humanities vs. sciences, public vs. private universities)

Feel free to select one or more options, or provide additional details so I can assist you better with your search.

psychological impact

Performing Steps (takes upto 3 minutes) 30% Done

- Executing multiple relevant queries to broaden the search.
  - psychological impact of reading open-ended questions in student evaluations on university professors
  - impact of qualitative student feedback on teacher psychology and emotional well-being in higher education
  - effects of student feedback on professors' emotional health and professional growth in higher education
- Finding relevant papers for each query with applied filters
  - 46 top papers found out of 1050

[1] Teacher Wellbeing: The Importance of Teacher–Student Relationships  
Jantine L. Split + 2 more - Educational Psychology Review

[2] University lecturers' emotional responses to and coping with student feedback: a Finnish case study  
Sonja Lutovac + 3 more - European Journal of Psychology of Education

[3] Nurturing learning development through student feedback  
Colette Mair + 4 more - Journal of Learning Development in Higher Education

Show More

Performing Steps (takes upto 3 minutes) 80% Done

- Executing multiple relevant queries to broaden the search.
  - psychological impact of reading open-ended questions in student evaluations on university professors
  - impact of qualitative student feedback on teacher psychology and emotional well-being in higher education
  - effects of student feedback on professors' emotional health and professional growth in higher education
  - Emotional and psychological effects of anonymous student evaluations on faculty wellbeing in higher education
  - emotional intelligence and coping mechanisms of university professors in response to qualitative student feedback
- Finding relevant papers for each query with applied filters
  - 91 top papers found out of 1750

[1] Distinguishing the Themes Emerging from Masses of Open Student Feedback  
Timo Hyminen + 3 more - International Convention on Information and Communication Technology, Electronics and Microelectronics

[2] Contribution of open-ended questions in student evaluation of teaching  
Pablo Marshall - Higher Education Research & Development

[3] Feedback as Open-Ended Conversation: Inviting Students to Co-Regulate and Metacognitively Reflect During Assessment  
Jt Torres - Journal of the Scholarship of Teaching and Learning

Show More

- Reviewing citations & references to identify additional relevant works
  - 99 relevant citations and references found

[1] We can work it out: faculty interpretation of student evaluations of teaching  
Yali Bai - Journal of Further and Higher Education

[2] The relationship between students' evaluations of teaching and academics professional development  
Mari Karm + 2 more - Journal of Further and Higher Education

SCISPACE

Papers Search or ask a question

Psychological Impact of Reading Open-Ended Questions in Student Evaluations on University Professors

Introduction

Student evaluations of teaching (SET) are a common practice in higher education. Intended to provide timely and constructive insights into teaching experience, the psychological impact of reading open-ended questions in these evaluations on university professors has become a significant concern. This response explores the emotional, professional, and institutional implications of such feedback, drawing on insights from relevant research papers.

Emotional Impact of Open-Ended Feedback

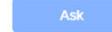
The emotional impact of open-ended feedback in SET is profound. Professors often experience negative emotions such as sadness, anxiety, and stress when receiving negative feedback, particularly if it is perceived as unfair or biased (Mazhindurkaya et al., 2024). These emotions can be exacerbated by the anonymous nature of the feedback, which sometimes includes hurtful, defamatory, or abusive comments (Cunningham et al., 2023; Stroh, Deines, Dierckx, and Copejic, 2023; Tuncay, 2023). The emotional toll is further compounded by the perception that such feedback is beyond their control and may be unfair or biased (Mazhindurkaya et al., 2024; Hyminen et al., 2023).

Younger academics and those in non-tenured positions are particularly vulnerable to

Con scite, localiza un trabajo mio y dice que digo algo que jamás dije.. (solo es cuestión de tiempo que mejoren los modelos para que no metan la pata así, porque es trivial mejorarlos –pero cuesta dinero-)

**scite\_**

Assistant Product Solutions  

Ask me another question   

Custom Assistant Settings are being used for this session. 

↳ ¿Pueden afectar negativamente a los profesores universitarios el leer las preguntas de texto abierto en las Student evaluation of teaching surveys?

↳ Los profesores universitarios se enfrentan a una serie de desafíos inherentes a las evaluaciones de enseñanza por parte de los estudiantes, especialmente cuando se trata de preguntas de texto abierto en las encuestas de evaluación. Estas preguntas ofrecen la oportunidad de obtener retroalimentación más matizada y rica en detalles sobre la experiencia del estudiante, pero también pueden presentar varios problemas que podrían afectar negativamente a los docentes.

Primero, la retroalimentación abierta puede conducir a la interpretación errónea de las intenciones del instructor. Según Torres (Torres, 2022), dejar a los estudiantes interpretar las intenciones de los profesores puede dificultar el aprendizaje, sugiriendo que las respuestas abiertas pueden perjudicar la autoestima profesional de los docentes si se perciben como críticas contundentes o mal dirigidas. Estudios han mostrado que estas evaluaciones pueden influir en la moral del profesorado, ya que la naturaleza subjetiva de la retroalimentación permite que comentarios negativos sean particularmente impactantes (Aznar-Mas et al., 2023).

Además, las evaluaciones al final del semestre, como las encuestas de evaluación de enseñanza, suelen no proporcionar retroalimentación a tiempo para que los instructores hagan ajustes en su enseñanza. Chang et al. indican que estas evaluaciones son a menudo demasiado tardías para beneficiar a los docentes en el mismo curso, lo cual se traduce en una percepción de que el tiempo y esfuerzo invertido en la recopilación de retroalimentación es, en muchos casos, infructuoso (Chang et al., 2014). Esto puede desincentivar a los profesores a involucrarse con los procesos de evaluación, viéndolos como una formalidad más que como una herramienta de desarrollo profesional.

Otro aspecto a considerar es el sesgo que puede surgir de las respuestas en las encuestas de evaluación de enseñanza. Un análisis de la literatura indica que las evaluaciones de estudiantes a menudo se ven afectadas por preconceptos y sesgos inherentes, lo que puede llevar a que comentarios poco constructivos alteren la percepción general sobre la eficacia docente (Hagen, 2020). Esta percepción sesgada es problemática en el contexto de educación superior, donde la reputación del profesor puede impactar directamente en su futuro profesional y en la dinámica del aula.

La diversidad de comentarios también puede ser abrumadora para los docentes. Grönberg et al. Grönberg et al. (2021) mencionan que, aunque las preguntas abiertas deberían proporcionar información valiosa, a menudo se limitan a simples descripciones que no son analizadas adecuadamente, dejando a los profesores a navegar por un mar de información que puede ser difícil de sintetizar y aplicar constructivamente. Esto puede resultar en una carga cognitiva adicional para los educadores, quienes deben discernir qué comentarios son relevantes.

Asimismo, la reducción de la calidad del feedback puede ser atribuida a la falta de análisis detallado de las respuestas abiertas, que frecuentemente son ignoradas en favor de las respuestas cuantitativas más fácilmente evaluables (Nawaz et al., 2022). Este enfoque puede socavar el potencial de la retroalimentación cualitativa, llevándola a ser

**References** 

...Open-ended feedback, however, also has its limits. Leaving students to interpret teachers' intentions can not only hinder lea... 

Section: Discussion


Effectiveness of the use of open-ended questions in student evaluation of teaching in an engineering degree  
Lourdes E. Aznar-Mas<sup>1</sup>, Lorena Atarés<sup>2</sup>, Juan A. Marín-García<sup>3</sup> 2023  
*JEM*  


...Student evaluation of teaching is the systematic process in which school leaders periodically critique teachers' work performance based on student feedback (Lejonberg, Elstad & Christoffersen, 2018) . Although initially considered a highly controversial topic, teaching evaluation has long been seen as an integral part of good professional practice in Higher Education Institutions (Hounsell, 2003)... 

Section: Introduction


<https://scite.ai/assistant/pueden-afectar-negativamente-a-los-profesores-universitarios-el-PQMPGg>



Juan A. Marín-García

Universitat Politècnica de València-DOE-ROGLE-IEMA  
[www.upv.es](http://www.upv.es)



ROGLE  
Reengineering Operations  
GroupWork Logistics Excellence



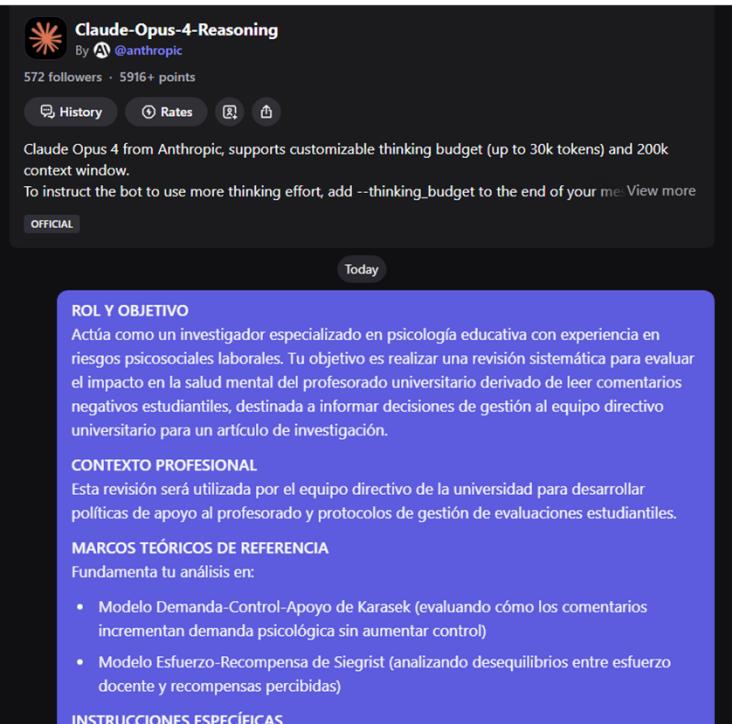
doe  
Departamento de Organización de Empresas



SINCE

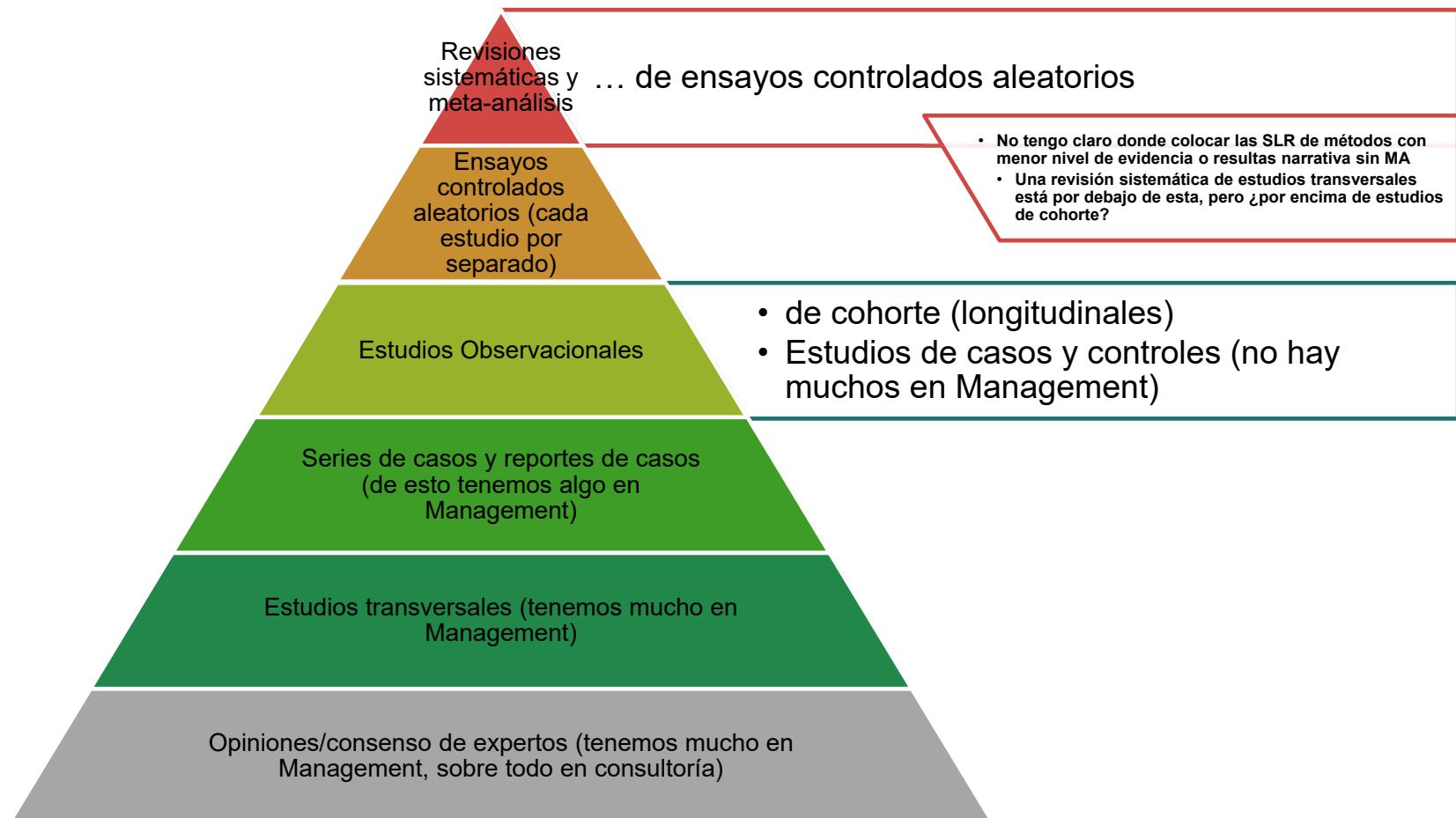
64

## Con otras herramientas con el prompt largo



- **Perplexity**
  - <https://www.perplexity.ai/search/rol-y-objetivo-actua-como-un-i-umwUbSh3TfSbnQjERI4nJw>
- **ChatGPT 4.5**
  - Revisión riesgos psicosociales docentes
- **Claude-sonnet-4**
  - <https://poe.com/s/24JH28KZtOEQL8b0bLy>

Desde el punto de vista positivista. La síntesis de evidencias está en la cúspide de la pirámide de evidencia científica (aplica a MA, no se si a todas las SLR)



## Reflexión epistemológica: lo anterior es cierto desde una visión positivista de la ciencia



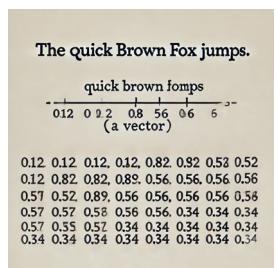
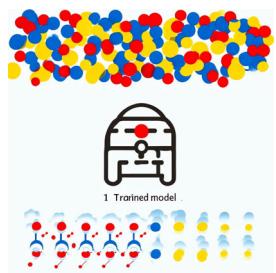
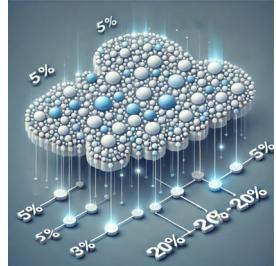
Epistemología: estudio del conocimiento, su naturaleza, origen y límites

### Enfoque interpretativo o constructivista:

1. Se valoran más los métodos cualitativos como estudios de caso en profundidad, etnografías, o investigación-acción.
2. La validez se basa más en la riqueza y profundidad de los datos que en la generalización estadística.

### Perspectiva crítica:

1. Se priorizan métodos que revelan estructuras de poder y promueven el cambio social.
2. La investigación participativa y la investigación-acción crítica son altamente valoradas.



- Generative AI
    - Generalist generative (example Claude4.SONNET; GPT4O [GPTO1]; Gemini 1.5...)
    - Generative [supposedly] trained in Research (perplexity, scite, consensus, ...)
  - AI with natural language processing (NLP)
    - Sentence transformers
  - AI [traditional] supervised learning (ML)
    - (Rayyan as an example)

# Etapas de una Revisión Sistemática de Literatura

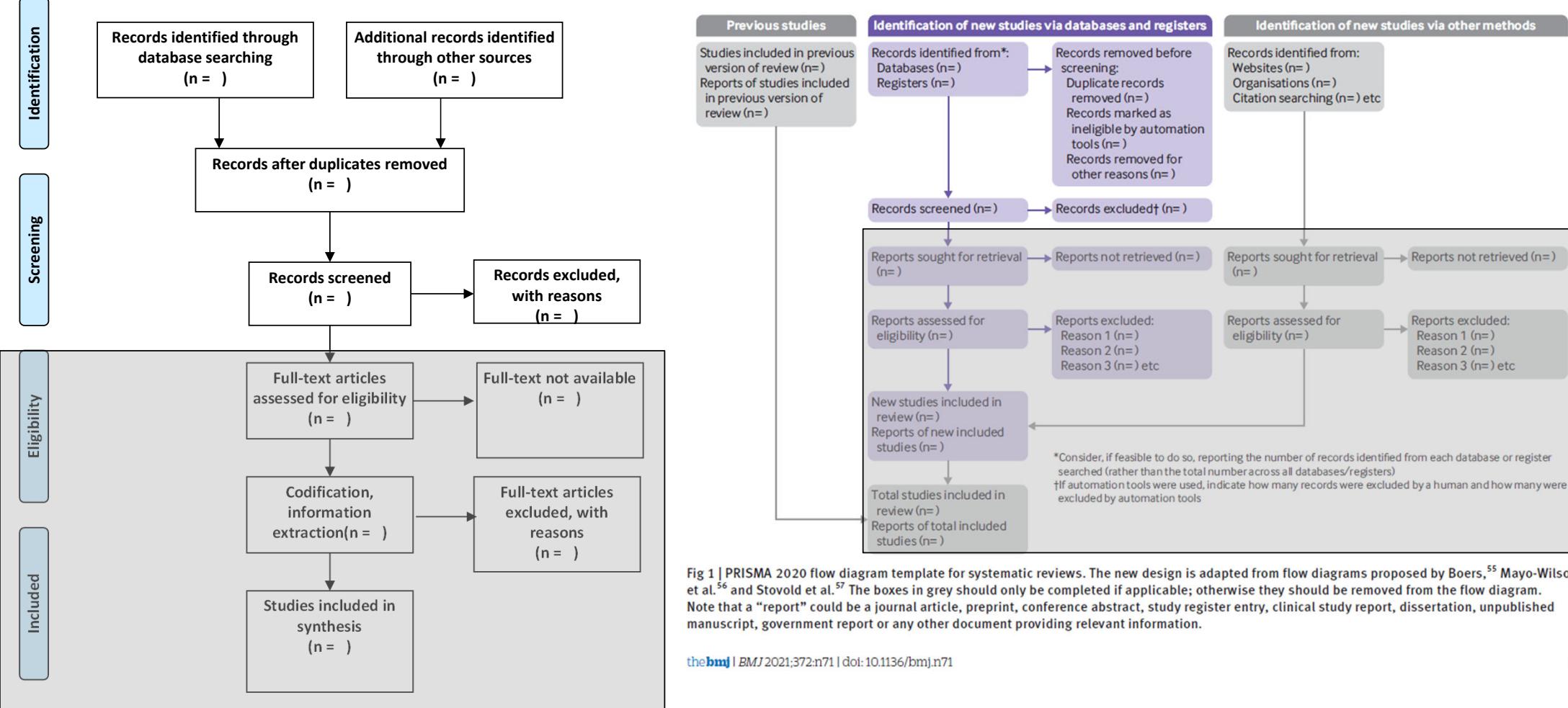
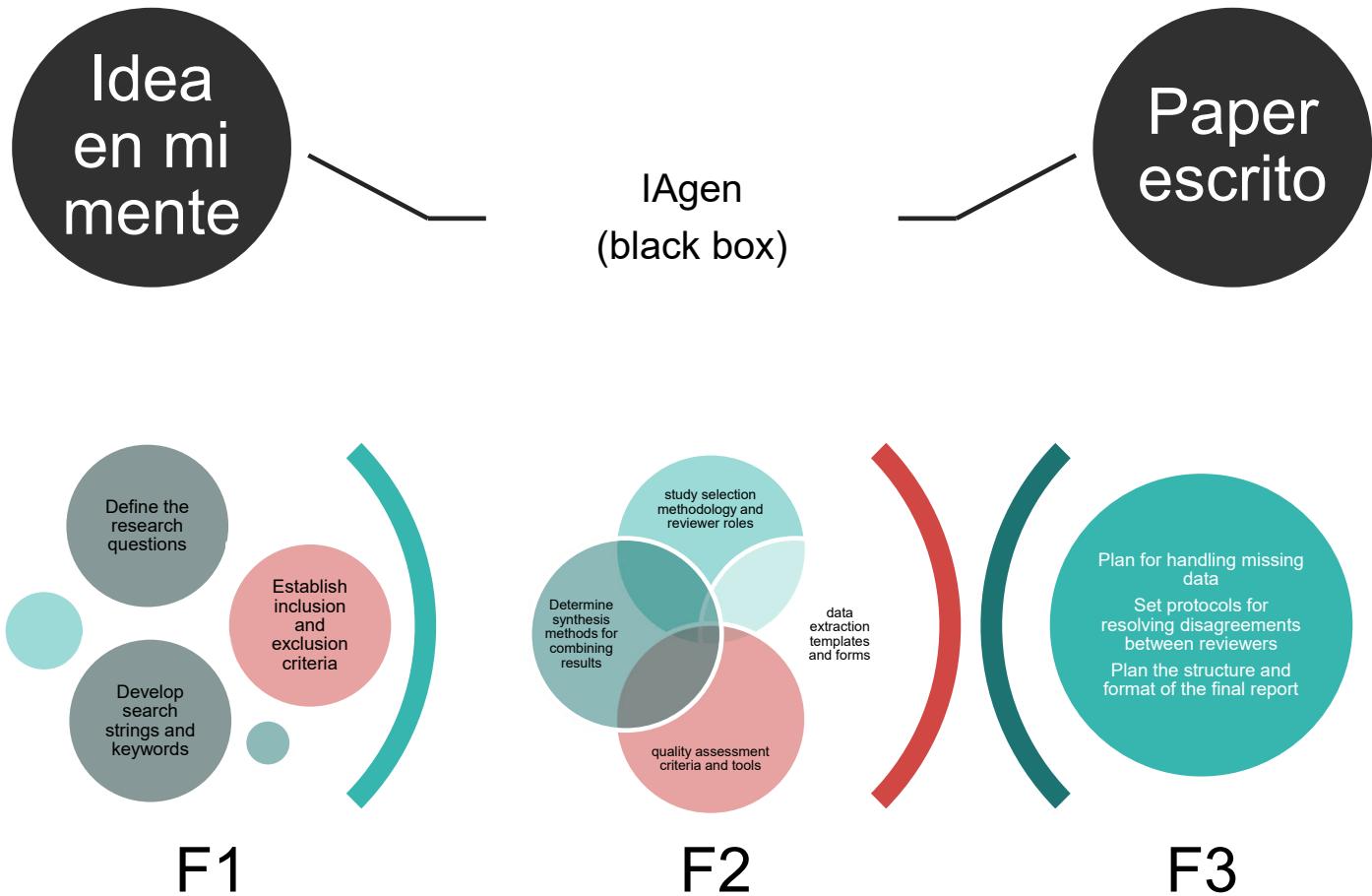


Fig 1 | PRISMA 2020 flow diagram template for systematic reviews. The new design is adapted from flow diagrams proposed by Boers,<sup>55</sup> Mayo-Wilson et al.,<sup>56</sup> and Stovold et al.<sup>57</sup> The boxes in grey should only be completed if applicable; otherwise they should be removed from the flow diagram. Note that a "report" could be a journal article, preprint, conference abstract, study register entry, clinical study report, dissertation, unpublished manuscript, government report or any other document providing relevant information.

thebmj | BMJ 2021;372:n71 | doi:10.1136/bmj.n71

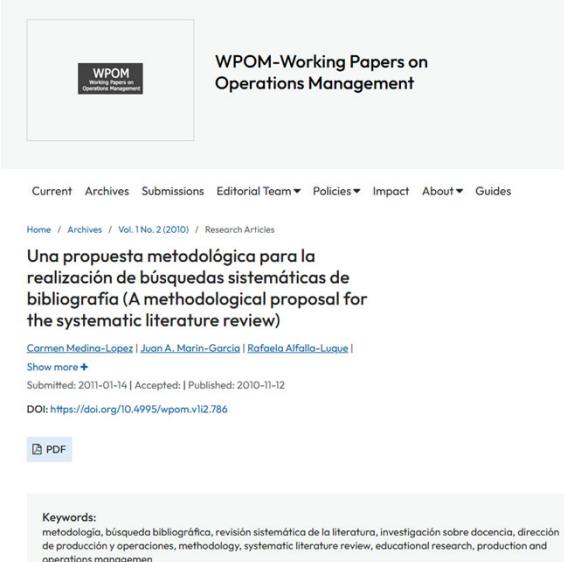
5

# ¿Existe la magia?



# Basic decisions in a systematic literature review

Medina-Lopez, C., Marin-Garcia, J. A., & Alfalla-Luque, R. (2010). Una propuesta metodológica para la realización de búsquedas sistemáticas de bibliografía (A methodological proposal for the systematic literature review). *WPOM-Working Papers on Operations Management*, 1(2), 13–30.  
<https://doi.org/10.4995/wpom.v1i2.786>



WPOM-Working Papers on Operations Management

Current Archives Submissions Editorial Team Policies Impact About Guides

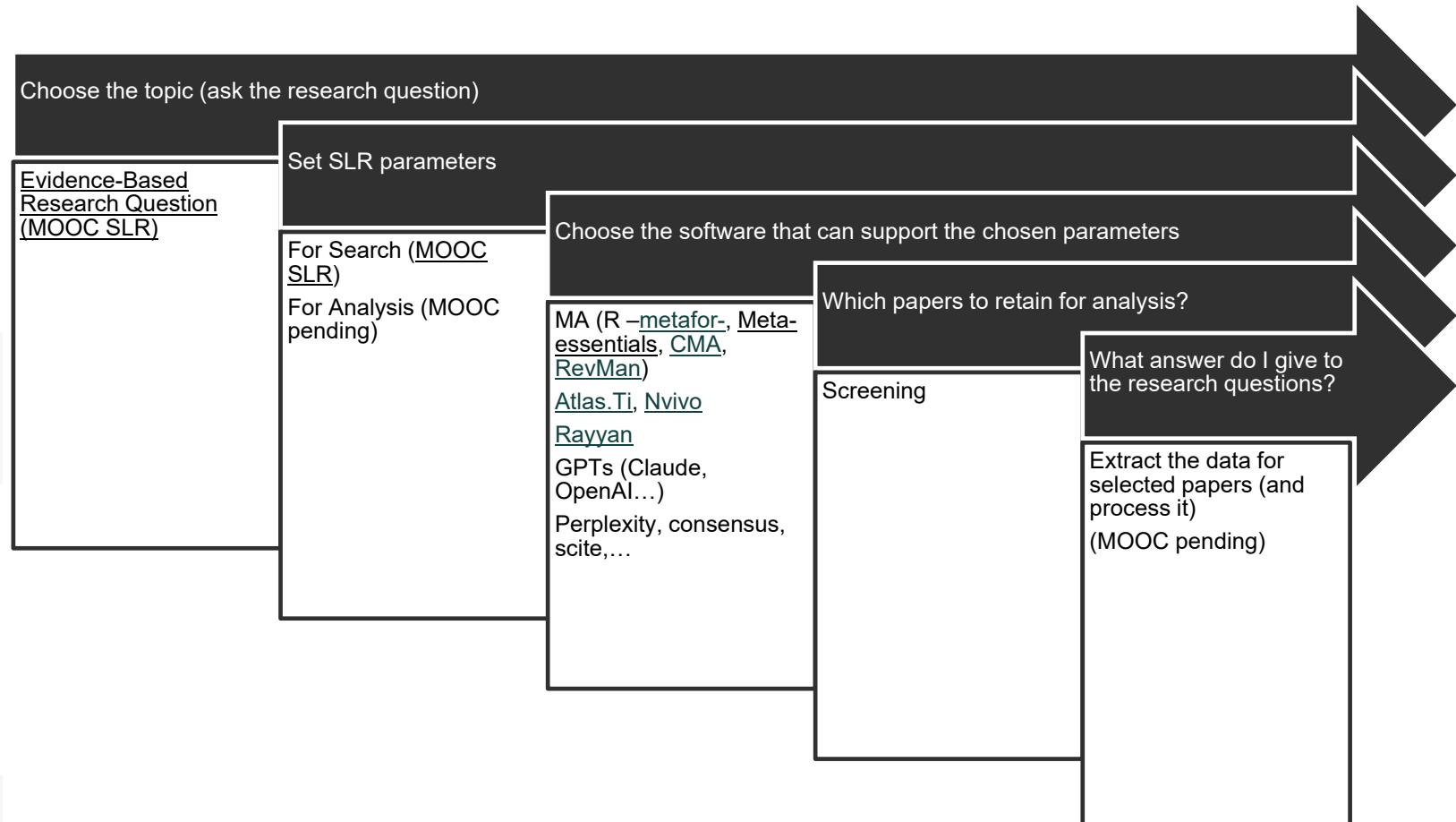
Home / Archives / Vol.1 No. 2 (2010) / Research Articles

**Una propuesta metodológica para la realización de búsquedas sistemáticas de bibliografía (A methodological proposal for the systematic literature review)**

Carmen Medina-Lopez | Juan A. Marin-Garcia | Rafaela Alfalla-Luque | Show more +  
 Submitted: 2011-01-14 | Accepted: | Published: 2010-11-12  
 DOI: <https://doi.org/10.4995/wpom.v1i2.786>

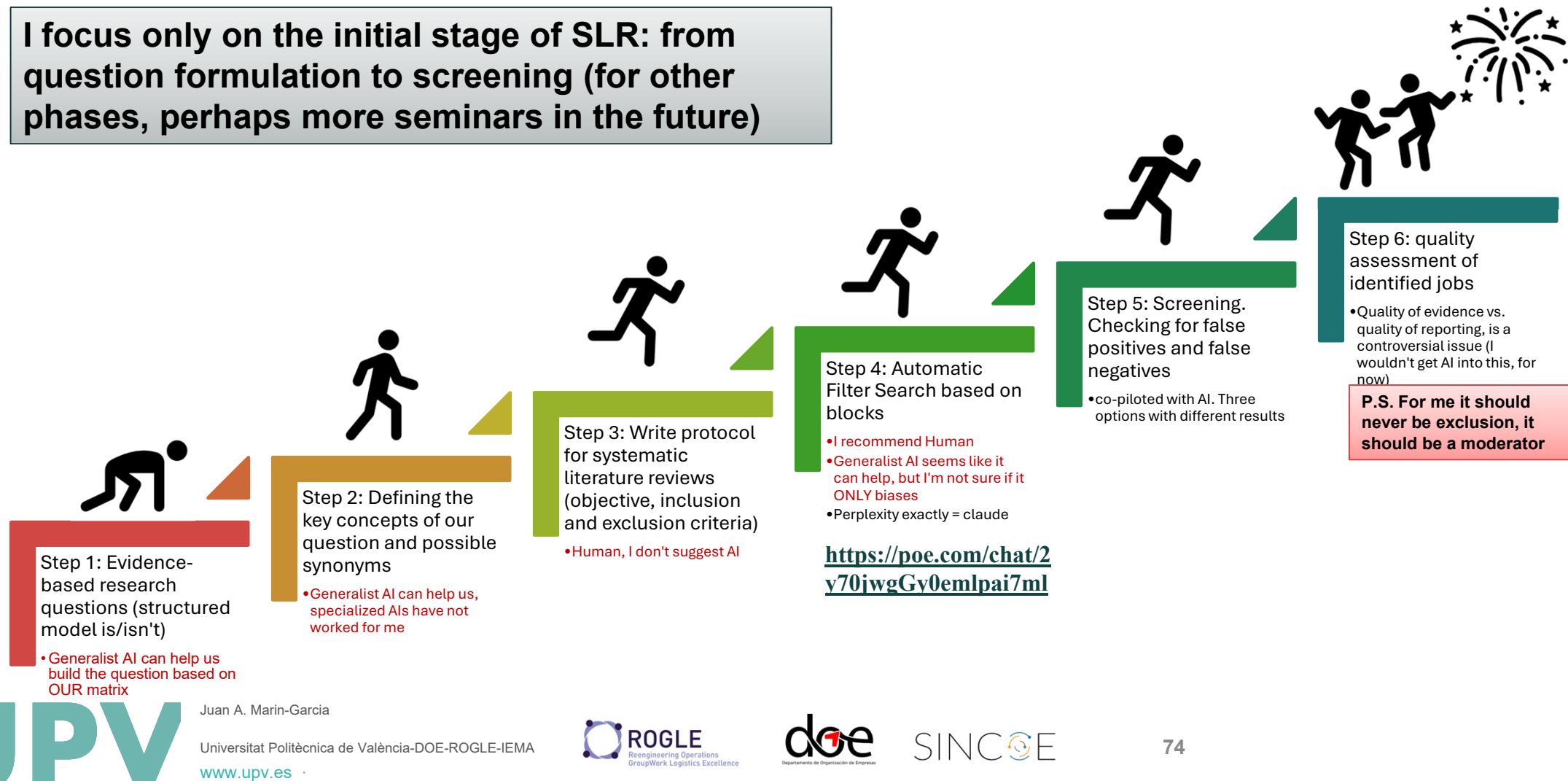
[PDF](#)

**Keywords:**  
 metodología, búsqueda bibliográfica, revisión sistemática de la literatura, investigación sobre docencia, dirección de producción y operaciones, methodology, systematic literature review, educational research, production and operations management



Summary of the session. The rest of the time I will dedicate it to show examples of support to the statements of this transparency

I focus only on the initial stage of SLR: from question formulation to screening (for other phases, perhaps more seminars in the future)



# Antes de ponernos manos a la obra. Una reflexión sobre el éxito dependiendo del estado de madurez en competencias digitales y como voy a abordar el taller



Voy a ir a un ritmo lento. Pero no me voy a parar, ni repetir instrucciones. Si te atascas, lo intentas de nuevo viendo las transparencias o videos de soporte. Si que voy a resolver dudas “conceptuales” (discrecionalmente) Si te agobia intentar replicarlo en directo, simplemente ves como lo hago yo (pero te quedas sin “chaquetilla”)

Step 1: Evidence -based research questions (structured model is/not). The AI generalist can help us build a title based on our matrix

|  | <b>Is</b>  | <b>It's not</b>  |
|--|--|--|
| What is the population or characteristics of the people who are going to be used or will the decision or intervention be affected? | University Professors  | Students, Service Staff<br>Other educational levels other than university                                  |
| [In the case of interventions] What is the intervention, program, treatment, problem, improvement, product, service... to analyze? | Open-ended questions to capture students' opinion in the "students evaluation of teaching"                               | Closed-ended likert scale questions or similar   |
| [In other cases] why is the information, or the decision, necessary? What reasons drive the behavior or experience?                | It is possible that when reading the opinions of the students, teachers may feel offended, insulted or simply questioned | The usefulness of this information for managers/heads of university, centre, department, degree or subject |
| What are the expected results, expectations, impact and how to measure them?   | The impact on the psychological health, well-being or motivation of teachers   | The quality of teaching or learning  |
| Who are the people who are going to provide the service or information or are going to perform the intervention                    | Responsible for teaching quality in universities   |  |
| Tipo de revisión [critical, descriptive, qualitative systematic, scoping, meta-analysis, umbrella]                                 | Umbrella   | To analyse the primary studies, I want to review the results of the reviews                                |
| [if any] What is the group against which we are going to compare the results?  |  |  |



Title: “**The Impact of Open-ended Student Teaching Evaluations on University Professors' Psychological Health and Well-being: An Umbrella Review** ”

Step 1: Evidence -based research questions (structured model is/not). The AI generalist can help us build a title based on our matrix

<https://www.youtube.com/watch?v=cj6cvI9HBEM>

### Título Pico Spider eclipse Cochrane - Poe



PROMPT (claude3.7-sonnet); Based on the information in the attached table, propose a title for a systematic literature review that follows the recommendations of Cochrane Reviews (with the PICO - SPIDER-ECLIPSE framework).

Based on the information provided in the table, I can see this pertains to a systematic review about the impact of student evaluations of teaching on university professors' psychological health and well-being. Using the SPIDER framework (which is often more appropriate than PICO for qualitative reviews):

S (Sample): University Professors

P (Phenomenon of Interest): Open-ended student evaluations of teaching

I (Design): Not specifically mentioned, but appears to be reviewing literature on psychological impacts

E (Evaluation): Impact on psychological health, well-being, and motivation

R (Research type): Umbrella review (as indicated in the table)

Here's a proposed title for the systematic review following Cochrane recommendations:

"The Impact of Open-ended Student Teaching Evaluations on University Professors' Psychological Health and Well-being: An Umbrella Review"

Alternative title using the ECLIPSE framework elements:

"Psychological Effects of Student Feedback Systems in Higher Education: An Umbrella Review of Open-ended Teaching Evaluations and Their Impact on University Professors' Well-being"

# With Gemini2.5 Free

Aunque te escriba en castellano mis prompts, responde siempre en inglés, empezando la respuesta siempre con la traducción de mi prompt a inglés. Prompt: "partiendo d..."

Mostrar cuando piensa

Okay, here's a proposed title for your systematic literature review, based on the information you provided and following Cochrane Review recommendations.

User Prompt Translation: "Based on the information in the attached table, propose a title for a systematic literature review that follows the recommendations of Cochrane Reviews (with the PICO - SPIDER - ECLIPSE framework)."

**Proposed Title for the Systematic Literature Review:**

"The impact of qualitative student feedback from teaching evaluations on the psychological well-being of university professors: An umbrella review"

**Explanation and Alignment with Frameworks:**

This title aims to be clear, specific, and encompass the key elements outlined in your table, fitting the structure often recommended for systematic reviews, including those by Cochrane. An umbrella review synthesizes evidence from multiple systematic reviews.

Here's how the title aligns with the PICO framework (though umbrella reviews sometimes have a broader scope than a single PICO question, the elements are still relevant):

- P (Population): University professors
  - Your table specifies "University Professors."
- I (Intervention/Exposure): Qualitative student feedback from teaching evaluations
  - This corresponds to "Open-ended questions to capture students' opinion in the 'students evaluation of teaching'" from your table. "Qualitative student feedback" accurately represents this type of input, differentiating it from quantitative measures (which your table lists under "It's not" – e.g., "Closed-ended Likert scale questions").
- C (Comparison): (Not explicitly stated in this title, which is common for umbrella reviews focusing on an exposure and its outcomes).
  - For an umbrella review, the comparison is often implicit in the scope of the included systematic reviews (e.g., reviews that examine this type of feedback versus those that don't, or reviews looking at different types of impact). If a direct comparison were the focus of the primary reviews being synthesized, it could be added, but the current information suggests a focus on the impact of a specific type of feedback.

- O (Outcome): Psychological well-being

- This encompasses "The impact on the psychological health, well-being or motivation of teachers" specified in your table. "Psychological well-being" is a comprehensive term that covers these aspects.

**Alignment with SPIDER framework (often used for qualitative or mixed-methods reviews, relevant here due to the nature of the feedback and the umbrella review of potentially diverse studies):**

- S (Sample): University professors
- PI (Phenomenon of Interest): The experience and impact of qualitative student feedback from teaching evaluations, specifically focusing on psychological well-being.
- D (Design): The design of the review itself is an umbrella review. The design of studies within the included systematic reviews will vary.
- E (Evaluation): The psychological well-being (health, motivation) of professors.
- R (Research Type): Umbrella review (review of systematic reviews), as explicitly stated in the title and your table ("Tipo de revisión: Umbrella").

**Alignment with ECLIPSE framework (often used for policy or service-related reviews):**

- E (Expectation): To understand the impact...on psychological well-being.
- C (Client group): University professors (as the group affected).
- L (Location): Implied as universities (where teaching evaluations occur for professors).
- I (Impact): Psychological well-being.
- P (Professionals): University professors.
- S (Service): The process of student evaluations of teaching (specifically qualitative feedback aspects).

**Cochrane Review Recommendations:**

Cochrane guidance generally suggests titles should be concise, precise, and clearly reflect the scope of the review, often including key PICO elements for intervention reviews. For overviews of reviews (umbrella reviews), it's crucial to indicate this in the title. This proposed title clearly states "An umbrella review."

This title clearly delineates the population, the specific type of exposure (qualitative feedback from student evaluations), the primary outcome of interest (psychological well-being), and the nature of the review (umbrella review), aligning with best practices for systematic review titles.

Step 2: Definition of the key concepts of our question and possible synonyms. The generalist AI can help us, the specialized ones have not worked for me



- Support videos (in Spanish). Looking for definitions, synonyms and seminal works with 8 generative AI platforms (they are part of a MOOC but you do not need to do the MOOC if you only are interested in this)

<https://www.youtube.com/playlist?list=PL6KQIM6>

- Claude3.5-sonnet (Version pago)
- O1-preview (versión pago)
- 
- chatGPT4o con plugin Dimensions (versión pago)
- Scite (versión pago)
- Consensus (versión pago)
- Perplexity (versión PRO-free)
- SciSpace (versión pago)
- WOS assistant (versión pago)

The screenshot shows a YouTube playlist page for a series of 26 video lectures. The title of the playlist is "Concretando la definición de los términos clave. Con ayuda de IA. Comparando plataform". The first video thumbnail shows a person speaking in front of a screen displaying a slide with text and diagrams. The video duration is 7:35. Subsequent thumbnails show various screens of text and diagrams from different AI platforms. The video descriptions indicate they are part of a MOOC and were uploaded by Universitat Politècnica de València - UPV.

I will show a result with Claude3.7sonnet. Test it (with my prompt, or the one that interests you) and compares with Gemini 1.5 o Chat GPT



[https://g.co/ge  
mini/share/644  
b9dbd24c0](https://g.co/ge mini/share/644b9dbd24c0)

- Why Gemini1.5?  
Because it is free and free access (with Gmail account)
- There is also the issue of privacy ... everything you upload to Gemini is used to train ... and for other things
- "What is 'Student evaluation of teaching', its definition, synonyms, and key seminal works and reviews on the topic"

**Save the "elaboration" because we will use it later**

Hay revisores humanos que leen algunas conversaciones guardadas para mejorar la IA de Google. Para evitar que esto suceda en conversaciones futuras, desactiva el ajuste Actividad en las aplicaciones de Gemini. Si este ajuste está activado, no introduzcas información que no quieras que se revise ni use. [Cómo funciona](#)

# "What is 'Student evaluation of teaching', its definition, synonyms, and key seminal works and reviews on the topic" Claude 3.7

## III. Concepts

### Definition

Student Evaluation of Teaching (SET) refers to the systematic process of collecting feedback from students about their educational experiences, instructor effectiveness, and course quality. These evaluations typically occur at the end of a course or academic term and involve students rating various aspects of teaching performance and course design through questionnaires or surveys. SETs are used for various purposes including:

- Providing feedback to instructors for improvement
- Administrative decision-making (promotion, tenure, merit pay)
- Course development and curriculum enhancement
- Quality assurance in higher education
- Research on teaching effectiveness

### Synonyms and Related Terms

- Course evaluations
- Teaching evaluations
- Student ratings of instruction (SRI)
- Student feedback questionnaires
- Student assessment of teaching (SAT)
- Student perception of teaching (SPOT)
- Student evaluations of educational quality (SEEQ)
- Student ratings of teaching effectiveness (SRTE)
- Student course experience questionnaires (SCEQ)
- Teaching effectiveness measures

### Key Seminal Works and Reviews

#### Foundational Works

1. **Marsh, H. W. (1984).** "Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility." *Journal of Educational Psychology*, 76(5), 707-754.  
1. One of the most cited works establishing the psychometric properties of SETs
2. **Feldman, K. A. (1989).** "The association between student ratings of specific instructional dimensions and student achievement." *Research in Higher Education*, 30(6), 583-645.  
1. Connected specific teaching behaviors to student outcomes
3. **Cohen, P. A. (1981).** "Student ratings of instruction and student achievement: A meta-analysis of multisector validity studies." *Review of Educational Research*, 51(3), 281-309.  
1. Early meta-analysis showing moderate correlations between ratings and achievement

#### Influential Reviews and Meta-analyses

4. **Spooren, P., Brockx, B., & Mortelmans, D. (2013).** "On the validity of student evaluation of teaching: The state of the art." *Review of Educational Research*, 83(4), 598-642.  
4. Comprehensive review of validity issues in SETs
5. **Uttl, B., White, C. A., & Gonzalez, D. W. (2017).** "Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related." *Studies in Educational Evaluation*, 54, 22-42.  
4. Challenged previous findings on the relationship between SETs and learning
6. **Hornstein, H. A. (2017).** "Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance." *Cogent Education*, 4(1).  
4. Critical analysis of SETs as assessment tools

#### Bias and Criticism Research

7. **Boring, A., Ottoboni, K., & Stark, P. B. (2016).** "Student evaluations of teaching (mostly) do not measure teaching effectiveness." *ScienceOpen Research*.  
7. Influential study demonstrating gender bias in SETs
8. **MacNell, L., Driscoll, A., & Hunt, A. N. (2015).** "What's in a name: Exposing gender bias in student ratings of teaching." *Innovative Higher Education*, 40(4), 291-303.  
7. Experimental study revealing gender bias in online course evaluations
9. **Clayson, D. E. (2009).** "Student evaluations of teaching: Are they related to what students learn? A meta-analysis and review of the literature." *Journal of Marketing Education*, 31(1), 16-30.  
7. Meta-analysis examining the relationship between learning and evaluations

#### Recent Developments and Alternatives

10. **Linse, A. R. (2017).** "Interpreting and using student ratings data: Guidance for faculty serving as administrators and on evaluation committees." *Studies in Educational Evaluation*, 54, 94-106.  
10. Practical guidance for appropriate use of SET data
11. **Darwin, S. (2017).** "What contemporary work are student ratings actually doing in higher education?" *Studies in Educational Evaluation*, 54, 13-21.  
10. Critical examination of the role of SETs in modern higher education
12. **Berk, R. A. (2013).** "Top 10 flashpoints in student ratings and the evaluation of teaching: What faculty and administrators must know to protect themselves in employment decisions." *Stylus Publishing*.  
10. Comprehensive overview of contemporary issues in SET use for personnel decisions

Comprobadas hasta al 6, son correctas, comprobad el resto

Con Claude-opus-4. He tenido que añadir “los seminal works los quiero en forma de lista de referencias con datos bibliograficos completos” todas las citas están comprobadas y son reales (ojo libros son recensiones)

## Definition

- Student evaluation of teaching refers to the formal mechanisms through which students assess various aspects of their educational experience, including instructor performance, course content, teaching methods, learning outcomes, and overall satisfaction. These evaluations are widely used in higher education institutions for multiple purposes: informing tenure and promotion decisions, identifying areas for teaching improvement, providing feedback to instructors, and sometimes influencing course assignments and salary decisions.

## Synonyms and Related Terms

- Student ratings of instruction (SRI)
- Student evaluations of educational quality (SEEQ)
- Course evaluations
- Teaching evaluations
- Student feedback questionnaires
- Student opinion surveys
- Instructor evaluations
- End-of-course evaluations
- Student assessments of teaching effectiveness

### Key Seminal Works and Reviews

- **Trabajos Fundacionales (1970s-1980s)**
  - Feldman, K. A. (1979). The significance of circumstances for college students' ratings of their teachers and courses. *Research in Higher Education*, 10(2), 149-172. <https://doi.org/10.1007/BF00976227>
  - Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisector validity studies. *Review of Educational Research*, 51(3), 281-309. <https://doi.org/10.3102/00346543051003281>
  - Marsh, H. W. (1982). SEEQ: A reliable, valid, and useful instrument for collecting students' evaluations of university teaching. *British Journal of Educational Psychology*, 52(1), 77-95. <https://doi.org/10.1111/j.2044-8279.1982.tb00220.x>
  - Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist*, 52(11), 1187-1197. <https://doi.org/10.1037/0003-066X.52.11.1187>
- **Revisões y Meta-análisis Importantes**
  - Abrami, P. C., d'Apollonia, S., & Cohen, P. A. (1990). Validity of student ratings of instruction: What we know and what we do not. *Journal of Educational Psychology*, 82(2), 219-231. <https://doi.org/10.1037/0022-0663.82.2.219>
  - McKeachie, W. J. (1997). Student ratings: The validity of use. *American Psychologist*, 52(11), 1218-1225. <https://doi.org/10.1037/0003-066X.52.11.1218>
  - Theall, M., & Franklin, J. (Eds.). (1990). *Student ratings of instruction: Issues for improving practice*. New Directions for Teaching and Learning, No. 43. San Francisco: Jossey-Bass.
  - Greenwald, A. G., & Gillmore, G. M. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist*, 52(11), 1209-1217. <https://doi.org/10.1037/0022-0663.52.11.1209>
  - **Investigaciones sobre Sesgo y Validez**
    - Centra, J. A., & Gaubatz, N. B. (2000). Is there gender bias in student evaluations of teaching? *The Journal of Higher Education*, 71(1), 17-33. <https://doi.org/10.2307/2649280>
    - Basow, S. A. (1995). Student evaluations of college professors: When gender matters. *Journal of Educational Psychology*, 87(4), 656-665. <https://doi.org/10.1037/0022-0663.87.4.656>
  - **Trabajos Contemporáneos Críticos**
    - Boring, A., Ottoboni, K., & Stark, P. B. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Research*, 1-11. <https://doi.org/10.1196/scientific-research.10000>
    - MacNell, L., Driscoll, A., & Hunt, A. N. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, 40(4), 291-303.
    - Kreitzer, R. J., & Sweet-Cushman, J. (2022). Evaluating student evaluations of teaching: A review of measurement and equity bias in SETs and recommendations for ethical reform. *Journal of Academic Ethics*, 20(1), 73-84.
    - Utli, B., White, C. A., & Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, 54, 22-42. <https://doi.org/10.1016/j.stueduc.2016.08.007>
  - **Libros Clave**
    - Marsh, H. W. (1987). *Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research*. Oxford: Pergamon Press.
    - d'Apollonia, S., & Abrami, P. C. (1997). Navigating student ratings of instruction. *American Psychologist*, 52(11), 1198-1208. <https://doi.org/10.1037/0003-066X.52.11.1198>
    - Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research*, 83(4), 598-642. <https://doi.org/10.3102/0034654313496370>

Lo mismo con Perplexity . He tenido que añadir “los seminal works los quiero en forma de lista de referencias con datos bibliograficos completos”

<https://www.perplexity.ai/search/what-is-student-evaluation-of-Awuck6a.Rb2KnYAixNZEgg>

- **Definition**  
SET is the practice of collecting students' assessments of their instructors' teaching performance and the course environment, often using standardized forms that may include both quantitative ratings and qualitative comments. The feedback generally covers aspects such as instructional clarity, organization, enthusiasm, fairness, and the perceived impact on student learning.
- **Synonyms**  
Common synonyms and related terms for "student evaluation of teaching" include:
  - Student rating of instruction
  - Student feedback on teaching
  - Course evaluation
  - Student evaluation of teachers
  - Teaching evaluation by students
  - Student assessment of teaching
- Seminal Works ((no comprobadas existencia)  
Marsh, Herbert W. (1992).  
Marsh, H. W. (1992, April). *A Longitudinal Perspective of Students' Evaluations of University Teaching: Ratings of the Same Teachers over a 13-Year Period*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA, April 20-24.  
[Disponible en ERIC: ED353282]
- Cohen, Peter A. (1981).  
Cohen, P. A. (1981). *Student Ratings of Instruction and Student Achievement: A Meta-Analysis of Multisection Validity Studies*. Review of Educational Research, 51(3), 281–309.
- Cohen, Peter A. (1980).  
Cohen, P. A. (1980). *Effectiveness of Student-Rating Feedback for Improving College Instruction: A Meta-Analysis*. Research in Higher Education, 13(4), 321–341.
- Abrami, Philip C., d'Apollonia, S., & Cohen, P. A. (1990).  
Abrami, P. C., d'Apollonia, S., & Cohen, P. A. (1990). *Validity of student ratings of instruction: What we know and what we do not*. Journal of Educational Psychology, 82(2), 219–231.
- Cronbach, Lee J., Gleser, Goldine C., Nanda, Harinder, & Rajaratnam, Nageswari. (1972).  
Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles*. New York: John Wiley & Sons.
- Spooren, Pieter, Brockx, Bert, & Mortelmans, Dimitri. (2013).  
Spooren, P., Brockx, B., & Mortelmans, D. (2013). *On the Validity of Student Evaluation of Teaching: The State of the Art*. Review of Educational Research, 83(4), 598–642.

## Step 3: Write (and publish) a protocol for systematic literature reviews (objective, inclusion and exclusion criteria). Human, I don't suggest AI

- Title: Ideally based on evidence -based research questions
- Authors-contact person
  - It is recommended to specify the expected contributions of each author.
- Background
  - "You must address the body of knowledge already formed, help establish the foundation for review and explain why the questions raised are important. It must be concise and understandable for someone other than an expert" (Higgins & Green, 2011).
  - It includes the clear and unambiguous definitions of each of the "variables" [contained in the title], endorsed with references.
  - When appropriate, describe the theoretical reasoning of why reviewed interventions can have an impact on potential recipients and justify the possibility of effectiveness.
- Contribution: "Mention why this review is carried out" (Higgins & Green, 2011).
  - What will the information that will be generated when the protocol be executed
  - Why is that relevant and pertinent now
- OBJECTIVES [Evidence -based research questions]
  - Specify or expand the one that appears in the title
- Criteria for selecting studies [inclusion and exclusion]
- Method for search, collection and filtering of [studies, reports, data] (and paper of the participants)
- How do you plan to analyze the data
  - You are not going to analyze them in the protocol but you say how you will proceed when you have the data
- Dates (work plan)

Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions version 5.1.0 [Updated March 2011]. The Cochrane Collaboration, 2011. Available from <http://handbook.cochrane.org/>

Marin-Garcia, J. A. (2021). Publishing in Three Stages to Support Evidence Based Management Practice. WPOM-WORKING PAPERS ON OPERATIONS MANAGEMENT, 12 (2), 56-95. <https://doi.org/10.4995/wpom.11755>

# Actividad por parejas (10 minutos). Comparar dos protocolos. Dadle una calificación de 0 a 10 a cada uno y razonad el por qué de esa nota

Protocol paper: The impact of open-ended student evaluations of teaching on university professors' psychological health and well-being: A quick review

Authors: Poe, Edgard R.

## Introduction

Student satisfaction surveys regarding teaching quality have been employed for decades as a tool to measure the quality of university teaching services and to propose improvement actions. In most cases, these surveys consist of closed-ended questions with multiple levels ranging from "strongly agree" to "strongly disagree" or with different verbal anchors. These questions are typically grouped into dimensions, ultimately yielding a numerical score that is sometimes rescaled to take values between 0 and 10.

Less frequently observed is the inclusion of open-text questions where students can express in their own words their perception of teaching quality, positive aspects, or suggestions for improvement.

This research is framed within the context of Action Research in the vice-rectorate of a public university. During the 2023-24 academic year, an improvement action was implemented that consisted of introducing two open-ended questions in the Student Evaluation of Teaching (SET). One question was designed to justify why, in cases where students indicated that the professor should be congratulated for outstanding teaching performance, they should provide their reasoning. The other question asked students to indicate aspects they would improve about the course or professor. Due to complaints from university unions, this initiative was halted in the same year of implementation to avoid potential psychosocial risks for faculty.

Consequently, the need arose to conduct a quick review that would allow the identification of SET studies that focus on examining the free-text responses that students provide in evaluations, beyond numerical ratings, to analyze the emotional and psychological impact on faculty. The review aims to examine how student comments, especially negative, abusive, or non-constructive ones, affect the mental health, wellbeing, and professional confidence of academics. In case sufficient meta-analyses on the topic are identified, we will complement this work with an umbrella review.

## Previous research

Student Evaluation of Teaching (SET) refers to the systematic process of collecting feedback from students about their educational experiences, instructor effectiveness, and course quality. These evaluations typically occur at the end of a course or academic term and involve students rating various aspects of teaching performance and course design through questionnaires or surveys. The main synonyms or related terms are:

- Course evaluations
- Teaching evaluations
- Student ratings of instruction
- Student feedback questionnaires
- Student assessment of teaching
- Student evaluations of educational quality
- Student ratings of teaching effectiveness
- Student course experience questionnaires
- Teaching effectiveness measures

Open-ended questions are a form of interrogative statement that cannot be answered with a simple "yes" or "no" response or with a specific piece of information, and instead require the respondent

Protocol Paper: The Impact of Open-Ended Student Evaluations of Teaching on University Professors' Psychological Health and Well-Being: A Systematic Review Protocol  
Watson, Ibrahim B.

## Abstract

**Background:** Open-ended student evaluations of teaching (SETs) have become ubiquitous in higher education quality assurance systems worldwide. While these evaluations serve institutional accountability purposes, emerging evidence suggests that negative, abusive, or non-constructive student comments may significantly impact faculty psychological well-being, professional confidence, and career trajectories. The psychological toll of hostile feedback, particularly in digital contexts where anonymity can amplify aggressive language, represents an understudied occupational health issue in academia.

**Objectives:** This systematic review aims to synthesize evidence on how open-ended student evaluation comments affect university professors' mental health, psychological well-being, and professional confidence, while identifying evidence-based interventions and automated detection systems for managing abusive feedback.

**Methods:** We will systematically search MEDLINE, PsycINFO, ERIC, Web of Science, and Scopus databases from inception to [date] using comprehensive search strategies combining terms related to student evaluations, faculty well-being, and psychological impact. We will include quantitative, qualitative, and mixed-methods studies examining the relationship between student feedback and faculty psychological outcomes. Two reviewers will independently screen studies, extract data, and assess methodological quality using appropriate tools. Meta-analysis will be conducted where appropriate, with narrative synthesis for heterogeneous data.

**Discussion:** This review will provide the first comprehensive synthesis of evidence on an increasingly recognized occupational health issue in higher education, informing evidence-based policies for ethical feedback systems and faculty support mechanisms.

**PROSPERO Registration:** [To be completed]

**Keywords:** student evaluations; faculty well-being; academic mental health; occupational stress; higher education; systematic review

## 1. Introduction

### Background

Student evaluations of teaching (SETs) have evolved from informal feedback mechanisms into standardized institutional requirements across global higher education systems (Uttl et al., 2017; Boring et al., 2016). While traditional Likert-scale ratings dominate quantitative assessment frameworks, open-ended comment sections have become increasingly prevalent, ostensibly providing nuanced qualitative insights into teaching effectiveness (Clayson, 2009; Spooren et al., 2013). However, the anonymity inherent in these systems, combined with digital delivery platforms, has created conditions where students may express feedback in ways that extend beyond constructive criticism into personally attacking, discriminatory, or psychologically harmful territory (Miller & Seldin, 2014; Peterson et al., 2019).

Recent qualitative studies have documented faculty experiences of receiving student comments containing personal attacks, inappropriate sexual commentary, discriminatory language targeting protected characteristics, and threats to professional reputation (Boring et al., 2016; MacNell et al., 2015). These hostile communications represent a form of workplace harassment that occurs within officially sanctioned institutional processes, creating complex ethical and legal implications for university administrations (Heffeman, 2022; Schmidt, 2019). The psychological impact of such experiences parallels documented effects of cyberbullying and workplace mobbing, including increased anxiety, depression, reduced self-efficacy, and career-related stress responses (Kokkinos et al., 2015; Nielsen et al., 2012).

The gendered and racialized nature of abusive student feedback has been extensively documented, with women faculty and faculty of color disproportionately receiving comments focused on personal appearance, authority questioning, and competence challenges unrelated to pedagogical performance (Boring et al., 2016; Reid, 2010; Anderson & Smith, 2005). This pattern suggests that student evaluation systems may inadvertently perpetuate workplace discrimination and contribute to documented disparities in faculty career advancement and retention (Mengel et al., 2019; Peterson et al., 2019).

## Step 4: Building the Automatic Search filter (by blocks)

- I recommend human (with the possibility of refining it with AI, but that always implies doing so of both modes and contrasting results)
  - It does not save time (because you have to do both tasks) but I can help to have better results (<FP and <FN)
- The AI generalist seems to help, but I'm not sure if only biased
  - Task that I have pending research
- The AI Specialized in Research does not contribute improvement with respect to the generalist
  - Perplexity exactly the same as Claude3.5-Sonnet

**Example of Prompt for your favourite generalist AI: I want to make a systematic review of literature with this title “” [title or extended title with definitions and criteria] ”Can you help me create a search strategy for WOS-Clarivate? using this additional information [add definition]**

- <https://www.webofscience.com/wos/alldb/summary/8360d4d6-45a9-4b48-88fd-fc8cba3b9d79-015fec838c/relevance/1>
- <https://www.webofscience.com/wos/alldb/summary/eb62c1e5-7dbd-4233-aba1-6690be7ff20d-015fecb368/relevance/1>
- <https://www.webofscience.com/wos/alldb/summary/e61a2fcc-1b9b-4d46-ab55-8bcccfaccf52-015fec4baf/relevance/1>

| Type                            | Search Query and Results  | Database  | Results | Actions   |
|---------------------------------|---|---|---------|---|
| Current session                 | <a href="#">Export</a> ▾  |   |         |   |
| <input type="checkbox"/> Search | "student* evaluat* of teach**" (Topic) and Preprint Citation Index (Exclude – Database) and Review Article (Document Types) | All Databases<br><a href="#">Show collections</a> ▾ | 35      |          |
| <input type="checkbox"/> Search | "student* evaluat* of teach**" (Topic) and "open ended" (Topic) and Preprint Citation Index (Exclude – Database)            | All Databases<br><a href="#">Show collections</a> ▾ | 35      |    |
| <input type="checkbox"/> Search | "student* evaluat* of teach**" (Topic) and Preprint Citation Index (Exclude – Database)                                     | All Databases<br><a href="#">Show collections</a> ▾ | 1,311   |    |

Juan A. Marin-Garcia

Universitat Politècnica de València-DOE-ROGLE-IEMA  
[www.upv.es](http://www.upv.es) ·



89

## Example case step 4

- I want to do a systematic review with this title  
"" Evolution of value proposals in jobs for people employed by others in Spain: a descriptive review and meta-analysis of generational perceptions and professional profiles"" "
- Can you help me create a search strategy for WOS-Clarivate using this additional information?  
DEFINITION OF EVP "AN EMPLOYEE VALUE PROPOSITION (EVP) Clearly Identifies The Mix of Tangible and Intangible Rewards and Benefits That Make Working for Any Given Organization Privately Attractive. An evp Conveys an Organization's Comparative Advantages As an Employer and Thus Serves recruitment and retention efforts. Well Beyond Salary. Likewise, Institts Can Thrive."
- Synonyms  
Employee Benefits  
Employer Branding  
Employer Offering  
Employee Definition An employee-A is a person who provides personal services to an organization or employer under a employment, explicit or implicit contract, in exchange for compensation, and whose performance is subject to the control or direction of the employer in terms of the means and methods to carry out the work.
- Synonyms  
Worker  
Staff Member  
Personnel  
Associate  
Workforce member  
JOBHOLDER  
Definition of generation  
Group of people born and living in the same period, who share similar historical and sociocultural experiences during their formative years, resulting in a set of shared values, attitudes and behaviors that distinguish them from other age groups
- Synonyms  
(Demographic Or Birth or Age) Cohort  
Age Group  
Inclusion criteria  
Any Job Occupation or Job Role  
"Job Occupation" Generally refers to a Broader category of work or professional [manager, supervisor, clerical worker, Elementary Occupations, Technical Staff ...]. These are indeed of the West Categories Rather than Specific Roles within an Organization.  
"Job Role" Tends to Be More Specific and Often reports to particular position with Company or Organization. IT MIGHT describes The Specific Duties, Responsibilities, and Functions of An Individual's Job [Financial Analyst, Project Manager, Customer Service Representative, Marketing Coordinator, Software Developer, Human Resources Specialist, Sales Representative, Operations Manager, Graphic Designner ...]  
Criterios de exclusión  
Not

### # Block 1: EVP and value proposals

Ts = (  
("Employee Value Proposition\*" or "Evp" or "Employer Value proposition\*" or "Employment Value proposition")  
Or  
("Employee Benefit\*" or "Employer Brand\*" or "Employer Offering")  
Or  
"  
Or  
("Work\*Value\*" near/5 ("proposition\*" or "offer\*" or "package\*"))  
Or  
("Talent Value\*" near/5 ("proposition\*" or "offer\*"))  
)

And

### # Block 2: Employees and workers

Ts = (  
("Employee\*" or "worker\*" or "Staff member\*" or "personnel" or "associate\*" or "workforce" or "jobholder")  
Near/5  
("perception\*" or "expectation\*" or "preference\*" or "experience\*" or "need\*")  
)

And

### # Block 3: Generations and profiles

Ts = (  
("generation\*" or "demographic cohort\*" or "birth cohort\*" or "age cohort\*" or "age group\*")  
Or  
("Job Occupation\*" or "Job Role\*" Or "Professional Profile\*" OR "Occupational categor\*")  
Or  
("Manager\*" or "Supervisor\*" or "Technical Staff" or "Clerical Worker\*" or "Professional\*")  
)

Results in Wos. Would they be different from those who had made a human alone?  
FP and FN

## Claude3.5-sonnet only with title

Query Preview

```
TS=(("employee value proposition*" OR "EVP" OR "employer value proposition*" OR "employment value proposition*")
OR
(("employee* offer*" OR "employer* offer*" OR "employment offer*") NEAR/5 ("value*" OR "benefit*" OR "reward*" OR "attract*"))
OR
("employer brand*" NEAR/5 ("proposition*" OR "value*" OR "offer*"))
OR
("work* value*" NEAR/5 ("proposition*" OR "offer*" OR "package*"))
)
```

+ Add date range      X Clear      Search ▾

Booleans : AND,

Field Tags:

- TS=To
- TI=Titl
- AU=[A
- AI=Aul
- Identif
- GP=[G
- ED=Ec
- AB=Abt

## Claude3.5-Sonnet with title and definitions

|                              |   |           |
|------------------------------|---|-----------|
| <input type="checkbox"/> 9   | #1 AND #2 and Preprint Citation Index (Exclude - Database)  | 423       |
| <input type="checkbox"/> 0/7 | Combine Sets ▾      Export ▾  |           |
| <input type="checkbox"/> 7   | #1 AND #2 AND #3 and Preprint Citation Index (Exclude - Database)   | 45        |
| <input type="checkbox"/> 6   | #5 AND #2 AND #1 and Preprint Citation Index (Exclude - Database)   | 148       |
| <input type="checkbox"/> 5   | #4 OR #3 and Preprint Citation Index (Exclude - Database)   | 7,294,762 |
| <input type="checkbox"/> 4   | TS=(("job occupation*" OR "job role*" OR "professional profile*" OR "occupational categor*") OR ("manager*" OR "supervisor*" OR "technical staff" OR "clerical*"))                          | 2,277,556 |
| <input type="checkbox"/> 3   | TS=(("generation*" OR "demographic cohort*" OR "birth cohort*" OR "age cohort*" OR "age group*") and Preprint Citation Index (Exclude - Database))  | 5,117,829 |
| <input type="checkbox"/> 2   | TS=(("employee*" OR "worker*" OR "staff member*" OR "personnel" OR "associate*" OR "workforce" OR "jobholder*") NEAR/5 ("perception*" OR "expectation*" OR "preference*" OR "experience*")) | 242,740   |
| <input type="checkbox"/> 1   | TS=(("employee value proposition*" OR "EVP" OR "employer value proposition*" OR "employment value proposition*") OR ("employee benefit*" OR "employer brand*" OR "employer offering*"))     | 4,357     |

Make a quick search [choose your topic] in WOS [or scopus] and download the results (the first 1000 if they come out more) as XLS. For example:

337 results from All Databases for:

(kaizen or lean) and "higher education" (Topic)

Did you mean? (kaiser or lean) and "higher education" (Topic) and Preprint Citation Index (Exclude – Database) and 2024 or 2023 or 2022 or 2021 or 2020 (Publication Years) | 385 results

+ Add Keywords Quick add keywords: < + e-leaning + lean practices + lean six sigma + lean thinking + lean management +

Refined By: NOT Database: Preprint Citation Index X Publication Years: 2024 or 2023 or 2022 or 2021 or 2020 X Clear all

Export Records to Excel

Record Options

You have selected 50 results for export  
 All records on page  
 Records from: 1 to 337

No more than 1000 records at a time

Record Content:

Author, Title, Source, Abstract

Export Cancel

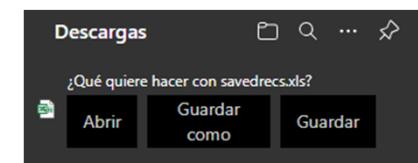
50/337 Add To Marked List

Dissertation or Thesis  
Continuous Improvement  
Sustainability

Patterson, Ron K.  
2022 | Creighton University

Higher education (HE) is facing a

(Ts = ((Kaizen or Lean) and "Higher Education")) and ((py == ("2024" or "2023" or "2022" or "2021" or "2020")) not (Siloid == ("pprn"))))



Save the "elaboration" because we will use it later

## An additional consideration related to step 4

Marin-Garcia, J.A.; Martinez-Tomas, J.; Juarez-Tarraga, A. & Santandreu-Mascarell, C. (2024). Paper protocol: from chaos to order. Augmenting Manual ARTICLE SCREENING WITH Sentence Transformers in Management Systematic Reviews. WPOM-WORKING PAPERS ON OPERATIONS MANAGEMENT, 15, 172-208. DOI: <https://doi.org/10.4995/wpom.22282>

O'Mara-Eves, A., Thomas, J., McNaught, J., Miwa, M., & Ananiadou, S. (2015). Using Text Mining for Study Identification in Systematic Reviews: A Systematic Review of Current Approaches. *Systematic Reviews*, 4 (1), 5. <https://doi.org/10.1186/2046-4053-4-5>

- The possibility of an article that is the publication of an effective search filters (includes the objective and verification that it is effective):
  - Recall
    - $TP/(TP + FN)$
  - Precision: Proportion of True Positives With respect to the Total Number of Classified Objects
    - $TP/ (TP+FP+TN+FN)$
  - ACCURACY: Total of True Positives and True Negatives With respect to the Total Number of Classified Objects
    - $(TP+TN)/ (TP+FP+TN+FN)$
  - Specificity: Proportion of True Negatives (Correctly Identified Negatives) with respect to the total of real negatives
    - $TN/(TN + FN)$
  - Error: Total Number of Wrong Classified items with respect to the total number of classified objects
    - $(FP+FN)/ (TP+FP+TN+FN)$

## Examples of articles similar to those that could be published in WPOM (if applicable to OM or HRM Research) [writing editorial in process]

- Verbeek, J., Salmi, J., Pasternack, I., Jauhainen, M., Laamanen, I., Schaafsma, F., HULSHOF, C., & Van Dijk, F. (2005). To Search Strategy for Occupational Health Intervention Studies. *Occupational and Environmental Medicine*, 62 (10), 682-687. <https://doi.org/10.1136/OEM.2004.019117>
- Devillé, W. L. J. M., Bezemer, P. D., & Bouter, L. M. (2000). PUBLICATIONS ON DIAGNOSTIC Test Evaluation in Family Medicine Journals: An optimal Search Strategy. *Journal of Clinical Epidemiology*, 53 (1), 65-69. [https://doi.org/https://doi.org/10.1016/s0895-4356\(99\)00144-4](https://doi.org/https://doi.org/10.1016/s0895-4356(99)00144-4)
- Robinson, K. A., & Dickersin, K. (2002). DEVELOPMENT OF A HIGHLY SENSITIVE SEARCH STRATEGY FOR THE RETRIEVAL OF REPORTS OF CONTROLLED TRIALS USING PUBMED. *International Journal of Epidemiology*, 31 (1), 150-153. <https://doi.org/10.1093/ije/31.1.150>

## Step 5: screening. Verification of false positives and false negatives (I will present the results in Acede 2025)

- Co-piloted with AI. Three options with different results
  - Generative AI (has potential with current models)
  - NLP with Sentence Transformers (I see [a lot] potential)
    - I am still analysing results (new articles and communications will come)
  - Traditional supervised learning trained with a selected set that tries to reproduce
    - (I need to try it)

Marin-Garcia, J.A.; Martinez-Tomas, J.; Juarez-Tarraga, A. & Santandreu-Mascarell, C. (2024). Paper protocol: from chaos to order. Augmenting Manual ARTICLE SCREENING WITH Sentence Transformers in Management Systematic Reviews. WPOM-WORKING PAPERS ON OPERATIONS MANAGEMENT, 15, 172-208. DOI:  
<https://doi.org/10.4995/wpom.22282>

# Examples of generative AI results. Claude3.5-Sonnet and ChatGpt-o1 that are models that do very, very good results



Claude-3.5-Sonnet

Instrucción \*

Indica a tu bot cómo comportarse y cómo responder a los mensajes de los usuarios. Intenta ser lo más claro y específico posible.

Ver mejores prácticas para instrucciones ↗

Trabaja siempre en idioma inglés. Eres un científico del área de gestión de recursos humanos interesado en el tema Remote work. Esta es la definición: " refers to a flexible work arrangement where employees perform their job duties from a location outside the traditional office environment, typically utilizing information and communication technologies to maintain connectivity with their organization and colleagues.: Telework, Telecommuting, Virtual work, Distributed work, Mobile work, E-work, Distance work, Flexible work, Work from home (WFH), Offsite work, Geographically dispersed work, Networked work, Location-independent work, Remote employment, Satellite work"

En cada prompt, el usuario te proporcionará un listado de resúmenes en inglés numerados con un identificador.

Tu tarea tiene cuatro partes.

1. Coloca un "##" y a continuación el identificador del resumen seguido de un "#".
2. Realiza una entity extraction de cada resumen. Agrega cada Entity, separándolas con un ";". Al final de la lista de entidades escribe un "#".
3. Compara las entidades del paso 2, con las descripciones de la definición del tema que te he facilitado antes y clasifica el resumen en una de las siguientes categorías:
  - a. @Cat1Insufficient@ si el resumen no contiene la información necesaria para clasificarlo.
  - b. @Cat2Sele@ si el resumen trata seguro del tema que te interesa
  - c. @Cat3Maybe@ si es probable que hable del tema que te interesa pero no está claro del todo .
  - d. @CatDiscard@ en todos los demás casos (el resumen no habla del tema que te interesa )Al final de la categoría, agrega un "#".
4. Explica por qué has clasificado el resumen de esa forma y al final agrega un doble símbolo "##"-

por ultimo, presentarás los resultados en inglés, en forma de tabla con tres columnas (emplea \t para tabuladores y \n para saltos de línea, para que mantenga la alineación al pegar en Excel). Cada fila corresponderá a un resumen de los proporcionados. La primera columna contendrá el identificador del resumen. La segunda columna muestra, separados por ";" los entity extraction contenidos en el resumen. La tercera columna incluye la categoría en la que has clasificado el resumen.

Optimizar la instrucción para vistas previas

# With ChatGpt-o1 Entity Extraction very good; good classification; Excellent justification

##10.1080/09585192.2022.2063064#  
Teleworking;COVID-19;Family-Supportive Supervisor Behaviors (FSSB);Work-family management;Exhaustion;Work engagement;Conservation of Resources theory;Employees' well-being;Intensity of telework;HRM theory and practice#  
@Cat2Sele#  
This summary clearly mentions "teleworking" and "working from home," which directly aligns with the definition of remote work.##  
##10.1080/09585192.2022.2060714#  
Employee referrals;Counterproductive work behavior;Social network;Affective organizational commitment;Monetary bonus;In-group collectivism;HR management theory and practice#  
@CatDiscard#  
This summary does not mention remote work or any related concept. Therefore, it is discarded.##  
##10.1080/09585192.2022.2032266#  
Journalists as first responders;Emotional labour;Initiative taking in crises;HRM-psychological wellbeing-performance mechanism;Job demand-resource theory;Supervisor behavioral ambidexterity (SBA);Flexibility of HR practices (HRP);Eustress;Personal initiative-taking behavior#  
@CatDiscard#  
This summary focuses on journalists, emotional labour, and HR practices, but makes no mention of remote work or related terms.##  
##10.1080/09585192.2021.1991431#  
High performance work practices (HPWPs);SMEs;Strategic HRM (SHRM);Environmental scanning;Competition tracking;Trade associations;Social network embeddedness;Competitive intensity;Geographical clusters#  
@CatDiscard#  
This summary discusses HPWPs in SMEs and related environmental factors, with no reference to remote work or related concepts.##  
##10.1080/09585192.2021.1986108#  
Psychological climate perceptions;Social network ties (advice, friendship);Trust;Expertise;Social information processing theory;IT organization;Social influence;Human resource management literature#  
@CatDiscard#  
The summary examines psychological climate and social networks, but does not mention remote or flexible work arrangements.##  
##10.1080/09585192.2021.1973063#  
HRM practices;Innovative work behaviour (IWB);Talent development climate (TDC);Resource-based view;Talent management;Supervisory support;Emerging market economy (India);Fostering innovation#  
@CatDiscard#  
This summary focuses on HRM practices, IWB, and talent development, without mentioning remote or related flexible work contexts.##

##10.1002/hrm.22161#  
Relational coordination;Remote work;Advanced Communication Technologies (ACTs);Job characteristics;HR practices;Affordances;Comprehensive ACTs;IT cooperation;Effective communication#  
@Cat2Sele#  
This summary explicitly focuses on remote work and effective communication in remote settings, directly aligning with the topic of interest.##  
##10.3926/ic.2107#  
Job crafting;Employees' proactivity;Impact on environment;Context of activity;Hybrid (physical and remote) environments;Virtual (metaverse);Talent attraction;Selection;Development processes;World Economic Forum#  
@Cat2Sele#  
Although the main focus is on job crafting, it explicitly includes consideration of hybrid and remote environments, directly connecting with the topic of remote work.##  
##10.1080/09585192.2023.2221385#  
Post-COVID remote working;Work-from-home;Hybrid working;Teleworking;Telecommuting;Work-from-anywhere;Pandemic impacts;Organisational trust;Managerial trust;Virtual communication;Collaboration;Reduced carbon emissions;Increased plastic consumption;Post-pandemic workplace#  
@Cat2Sele#  
This summary explicitly centers on various forms of remote working and related impacts, matching the topic of interest perfectly.##  
##10.1108/PR-09-2021-0673#  
Supportive teleworking practices;Work scheduling autonomy;Work-to-family conflict (WFC);Boundary theory;Distraction-conflict theory;Remote knowledge workers;Home-sourced work interruptions;Mandatory teleworking;COVID-19 pandemic#  
@Cat2Sele#  
This summary discusses teleworking practices and remote knowledge workers directly, clearly relating to the topic of interest.##  
##10.1108/PR-02-2022-0085#  
AMO model (Ability, Motivation, Opportunity);Neuro-inclusive remote workplace;Neurodiversity;Neurominorities;Strength-based approach;Organizational equality, diversity, inclusion;Skill-enhancing practices;Motivation-enhancing practices;Opportunity-enhancing practices;Remote work environment;Well-being;Empathetic approaches#  
@Cat2Sele#  
This summary explicitly focuses on a "neuro-inclusive remote workplace," which falls squarely under the definition of remote work.##

# With Claude3.5-sonnet Entity Extraction very good; Good classification (as well as human); Excellent justification→ results improved with claude 3.7

##10.1080/09585192.2022.2063064#

Teleworking; COVID-19; Family-Supportive Supervisor Behaviors; work-family management; exhaustion; work engagement; remote work intensity; well-being#

@Cat2Sele#

This abstract directly addresses teleworking and remote work, discussing specific aspects of working from home during COVID-19 and its impact on work-family balance##

##10.1080/09585192.2022.2060714#

Employee referrals; HR practice; social network; organizational commitment; cultural context; counterproductive work behavior#

@CatDiscard#

This abstract focuses on employee referral programs and doesn't discuss remote work or related concepts##

##10.1080/09585192.2022.2032266#

Journalists; emotional labor; supervisor behavioral ambidexterity; HR practices; initiative-taking behavior; psychological wellbeing#

@CatDiscard#

The abstract discusses journalists as first responders and emotional labor, with no connection to remote work##

##10.1080/09585192.2021.1991431#

SMEs; high performance work practices; environmental scanning; competitive intensity; top managers#

@CatDiscard#

This abstract focuses on SME management practices with no mention of remote work concepts##

##10.1002/hrm.22161#

Remote work; communication technologies; relational coordination; job characteristics; HR practices; user-technology interactions#

@Cat2Sele#

The abstract specifically addresses remote work settings and communication technologies essential for remote work coordination##

##10.3926/ic.2107#

Job crafting; hybrid environments; virtual environments; metaverse; employee proactivity; talent attraction#

@Cat3Maybe#

While the abstract mentions hybrid and virtual environments, its main focus is on job crafting definition, with remote work being only a peripheral consideration##

##10.1080/09585192.2023.2221385#

Remote working; COVID-19; work-from-home; hybrid working; teleworking; telecommuting; work-from-anywhere; productivity; virtual communication; carbon emissions#

@Cat2Sele#

This abstract directly focuses on remote work and its various forms, providing a comprehensive review of remote working impacts##

##10.1108/PR-09-2021-0673#

Teleworking practices; work-family conflict; remote knowledge workers; work interruption; COVID-19; work scheduling autonomy#

@Cat2Sele#

The abstract specifically addresses teleworking practices and their impact on work-family dynamics in remote work settings##

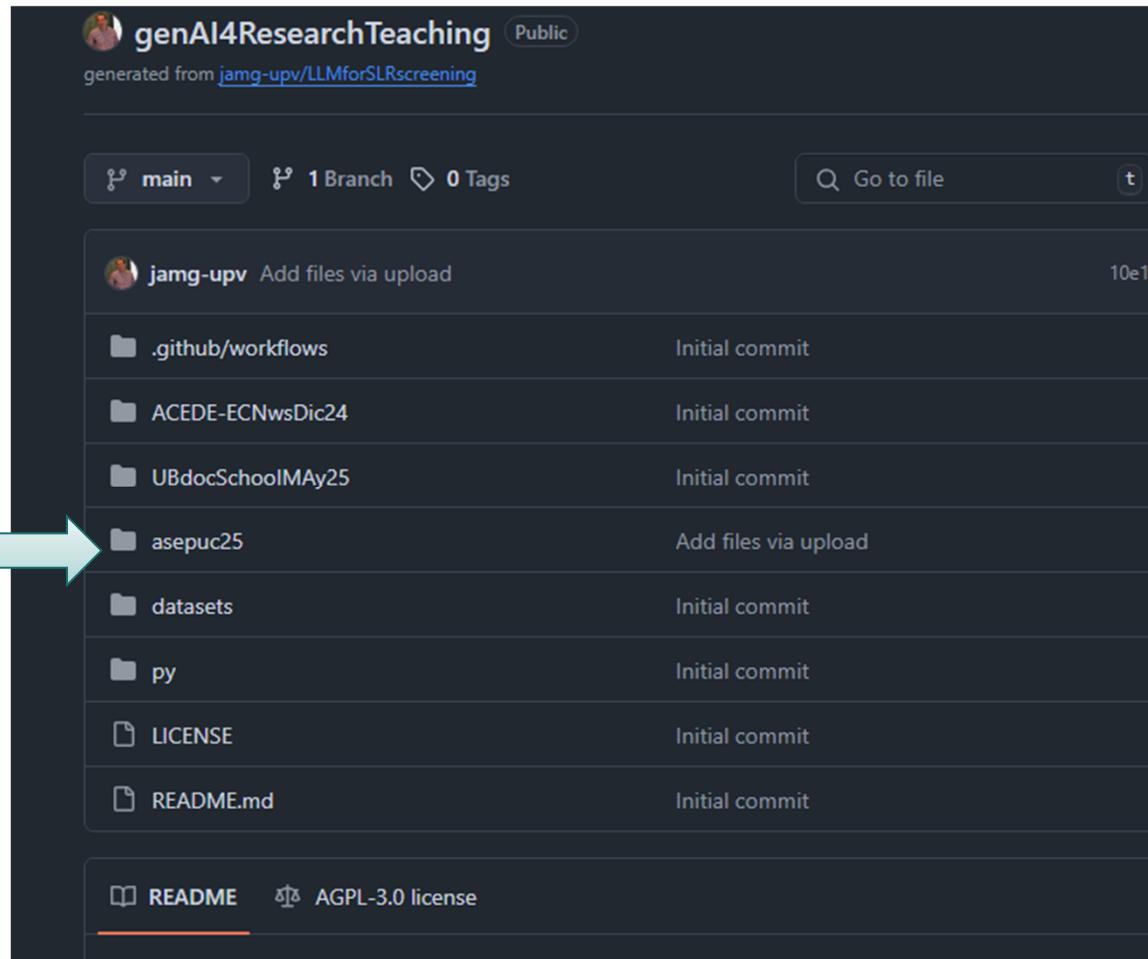
##10.1108/PR-02-2022-0085#

Remote workplace; neurodiversity; AMO model; HR practices; workplace inclusion; neurominorities#

@Cat2Sele#

This abstract focuses on remote workplace considerations specifically for neurodivergent employees##

Try it with Gemini [or your favorite AI] (you already have all “elaborations” to do so for your own theme, or use the prompt in github)



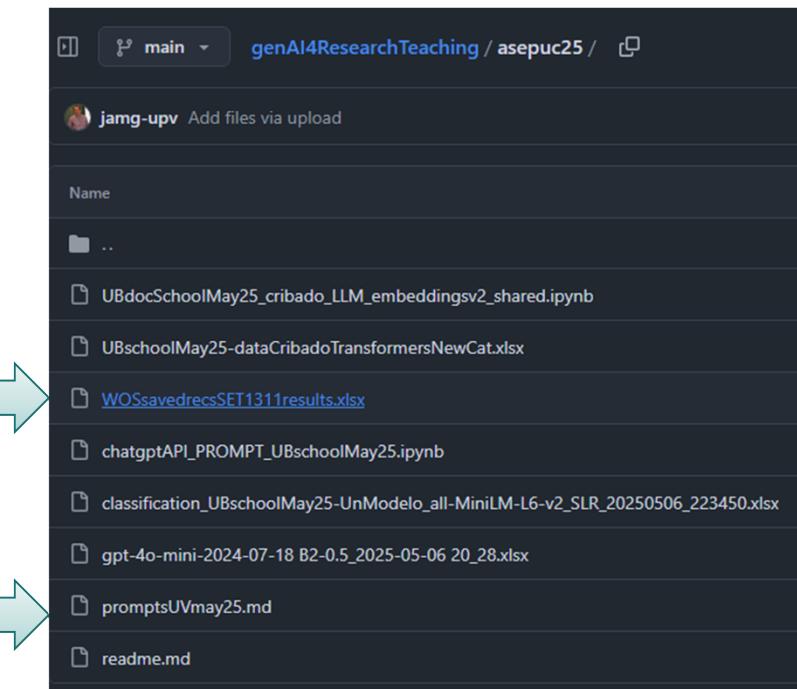
genAI4ResearchTeaching (Public)  
generated from [jamg-upv/LLMforSLRscreening](#)

main 1 Branch 0 Tags Go to file t

jamg-upv Add files via upload 10e19

| File/Folder       | Commit Message       |
|-------------------|----------------------|
| .github/workflows | Initial commit       |
| ACEDE-ECNwsDic24  | Initial commit       |
| UBdocSchoolIMAy25 | Initial commit       |
| asepuc25          | Add files via upload |
| datasets          | Initial commit       |
| py                | Initial commit       |
| LICENSE           | Initial commit       |
| README.md         | Initial commit       |

README AGPL-3.0 license

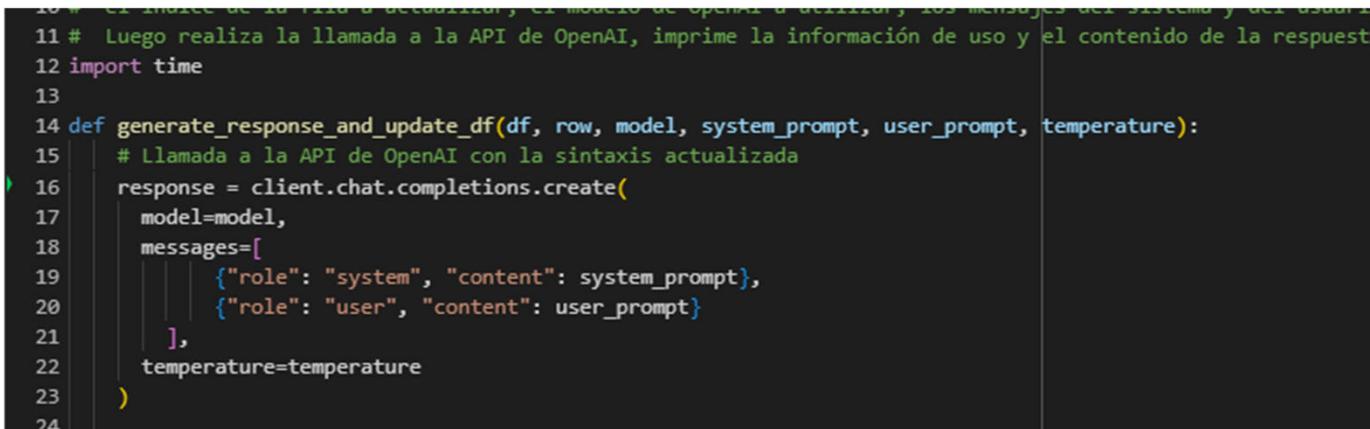


main genAI4ResearchTeaching / asepuc25 /

jamg-upv Add files via upload

| Name  |
|---|
| ..  |
| UBdocSchoolMay25_cribado_LLM_embeddingsv2_shared.ipynb                          |
| UBschoolMay25-dataCribadoTransformersNewCat.xlsx                                |
| WOSSavedrecsSET1311results.xlsx   |
| chatgptAPI_PROMPT_UBschoolMay25.ipynb   |
| classification_UBschoolMay25-UnModelo_all-MiniLM-L6-v2_SLR_20250506_223450.xlsx |
| gpt-4o-mini-2024-07-18 B2-0.5_2025-05-06 20_28.xlsx                             |
| promptsUVmay25.md   |
| readme.md   |

Recommended option: If you have an APIkey (openAI, Anthropic, Gemini, others), you can use python code and google colab to automatize the process



```
10 w el índice de la fila a actualizar, el modelo de OpenAI a utilizar, los mensajes del sistema y del usuario
11 # Luego realiza la llamada a la API de OpenAI, imprime la información de uso y el contenido de la respuesta
12 import time
13
14 def generate_response_and_update_df(df, row, model, system_prompt, user_prompt, temperature):
15     # Llamada a la API de OpenAI con la sintaxis actualizada
16     response = client.chat.completions.create(
17         model=model,
18         messages=[
19             {"role": "system", "content": system_prompt},
20             {"role": "user", "content": user_prompt}
21         ],
22         temperature=temperature
23     )
24
```

The screenshot shows a GitHub repository page for a file named `chatgptAPI_PROMPT_UBschoolMay25.ipynb`. The file contains the Python code shown above. Below the code, there's a preview section with tabs for `Preview`, `Code`, and `Blame`. The `Code` tab is selected, showing the code. There's also a `Open in Colab` button. At the bottom of the page, there's a red banner with the text **Descripción del caso de uso GPTUS01-entity\_extraction**.

Results with "gpt-4o-mini-2024-07-18" Temperature=0,5 (using API key)  
15 minutes for 500 rows

##id0607#*Middle-term evaluation; student evaluation of teaching effectiveness; feedback; teaching behaviors; course evaluations*##Cat2#The abstract discusses middle-term evaluations and their advantages, focusing on the completion usage of tokens. It includes examples of completed tasks related to student evaluations of teaching effectiveness, feedback, teaching behaviors, and course evaluations.

Procesada fila 427 de 500  
CompletionUsage(completion\_tokens=95, prompt\_tokens=807, total\_tokens=902, completion\_tokens\_details=CompletionTokensDetails(accepted\_prediction\_tokens=0, audio\_tokens=0, reasoning\_tokens=0, rejected\_predictions=0))  
##id0608#open-ended questions; Instructors Overall Performance (IOP) rating; course characteristics; students' demographics; effectiveness of the instructor; teaching roles; twenty-first century skills##Cat2#The abstract discusses open-ended questions and various factors related to instructors' overall performance, course characteristics, student demographics, teaching roles, and twenty-first century skills.

Procesada fila 428 de 500  
CompletionUsage(completion\_tokens=86, prompt\_tokens=1087, total\_tokens=1173, completion\_tokens\_details=CompletionTokensDetails(accepted\_prediction\_tokens=0, audio\_tokens=0, reasoning\_tokens=0, rejected\_predictions=0))  
##id0612#open-ended evaluations; teachers' verbal and nonverbal immediacy behaviors; gender differences; cultural differences; Student Instructional Rating System (SIRS); perceptions; evaluations##Cat3#The abstract discusses open-ended evaluations, teachers' verbal and nonverbal immediacy behaviors, gender differences, cultural differences, the Student Instructional Rating System (SIRS), perceptions, and evaluations.

Procesada fila 429 de 500  
CompletionUsage(completion\_tokens=69, prompt\_tokens=648, total\_tokens=717, completion\_tokens\_details=CompletionTokensDetails(accepted\_prediction\_tokens=0, audio\_tokens=0, reasoning\_tokens=0, rejected\_predictions=0))  
##id0613#Team teaching; student satisfaction; student performance; evaluations of teaching performance; team-taught courses; individual courses##Cat4#Discard##The abstract does not focus on student evaluations or team teaching.

Procesada fila 430 de 500

| A     | B   | C          | D        | E                | F | G  |
|-------|---|------------|----------|------------------|---|--|
| GPTId | GPTContent  | PTResponse | GPTUsage | promptAggregated |   |  |
| 1     | 0 ##id0001#open-ended questions; assessment indicators; teaching quality; students' evaluation of teaching; optimization; rational analysis#@Cat2Sele@ The study fo ChatComple Completion! id0001# Study on Data Analysis   |            |          |                  |   | ion_tokens=0, audio_tokens=0, reasoning_tokens=0, rejected_predictions=0 |
| 2     | 1 ##id0002#Online Student Evaluation of Teaching; productivity; implementation; stakeholder consultation; user training; primary data#@Cat4Discard@ The abstract discusses the flipped  |            |          |                  |   | ion_tokens=0, audio_tokens=0, reasoning_tokens=0, rejected_predictions=0 |
| 3     | 2 ##id0004#Student evaluation of teaching; teaching quality assurance; higher education institutions; evaluation system; student feedback; misuse@Cat2Sele@ The abstract disc ChatComple Completion! id0004# On Student Evaluation  |            |          |                  |   | iate STEM courses#@Cat4Discard@ The abstract discusses the flipped       |
| 4     | 3 ##id0006#Student evaluation of teaching; likability; student perceptions; evaluations; effectiveness; variance#@Cat4Discard@ The abstract disc ChatComple Completion! id0006# The student evaluation  |            |          |                  |   | ion_tokens=0, audio_tokens=0, reasoning_tokens=0, rejected_predictions=0 |
| 5     | 4 ##id0008#students' evaluation of teaching; teaching quality evaluation method; teacher performance evaluation tool; reliability; validity; influence@Cat2Sele@ The abstract disc ChatComple Completion! id0008# A literature review of t  |            |          |                  |   | he student evaluation  |
| 6     | 5 ##id0009#Course evaluation; standardized transformation; student evaluation of teaching; course characteristics; male students; higher grades; @ChatComple Completion! id0009# Course evaluation mod  |            |          |                  |   | ion_tokens=0, audio_tokens=0, reasoning_tokens=0, rejected_predictions=0 |
| 7     | 6 ##id0010#game theoretic model; student evaluation of teacher; empirical test; correlation; average grades; grade inflation#@Cat4Discard@ The ChatComple Completion! id0010# Is student evaluation o   |            |          |                  |   | ne game theoretic model?   |
| 8     | 7 ##id0012#Student evaluations of teaching; flipped classroom; traditional lecture; recorded lectures; course design; teaching evaluations; student participation@Cat2Sele@ The abstract disc ChatComple Completion! id0012# Flipped pedagogy and : student evaluations of teaching |            |          |                  |   | student evaluations of teaching  |
| 9     | 8 ##id0013#Students' Evaluation of Teaching; Higher Vocational Colleges; teachers' participation; students' ability to evaluate; evaluation methods@ChatComple Completion! id0013# Rethink on the Student   |            |          |                  |   | evaluation methods   |
| 10    | 9 ##id0014#student evaluation of teacher; game theoretic model; empirical test; correlation; average grades; Management introductory course#@ ChatComple Completion! id0014# Is student evaluation o  |            |          |                  |   | ne game theoretic model?   |
| 11    | 10 ##id0016#Student evaluations of teaching; faculty performance; in@ChatComple Completion! id0016# Faculty development program; student evaluations of te  |            |          |                  |   | student evaluations of teaching  |
| 12    | 11 ##id0017#Faculty development program; student evaluations of te@ChatComple Completion! id0017# Faculty development program; student evaluations of teaching  |            |          |                  |   | student evaluations of teaching  |
| 13    | 12 ##id0019#student evaluation of teaching; methodological inadequa@ChatComple Completion! id0019# student evaluation of teaching; methodological inadequa  |            |          |                  |   | methodological inadequa  |
| 14    | 13 ##id0021#open-ended questions; student evaluation of teaching; i@ChatComple Completion! id0021# open-ended questions; student evaluation of teaching; i  |            |          |                  |   | open-ended questions; student evaluation of teaching; i                  |
| 15    | 14 ##id0022#Student evaluations of teaching; open-ended questions;@ChatComple Completion! id0022#Student evaluations of teaching; open-ended questions;   |            |          |                  |   | Student evaluations of teaching; open-ended questions;                   |
| 16    |   |            |          |                  |   |  |

This article was published in December 2024. You may be interested as support material

## WPOM

Working Papers on Operations Management.  
Vol. 15 (172-208)  
ISSN: 1989-9068  
<https://doi.org/10.4995/wpom.22282>

Protocol paper: From Chaos to Order. Augmenting Manual Article Screening with Sentence Transformers in Management Systematic Reviews

Juan A. Marin-Garcia<sup>a</sup>, Juan Martinez-Tomas<sup>b</sup>, Amable Juarez-Tarraga<sup>a</sup> and Cristina Santandreu-Mascarell<sup>a</sup>

<sup>a</sup>ROGLE - Departamento de Organización de Empresas - Universitat Politècnica de Valencia jamarin@omp.upv.es  
<sup>b</sup>Departamento de Organización de Empresas - Universitat Politècnica de Valencia juanmart@omp.upv.es  
Departamento de Organización de Empresas - Universitat Politècnica de Valencia amjua@omp.upv.es and Departamento de Organización de Empresas - Universitat Politècnica de Valencia cristanna@omp.upv.es

Recibido: 2024-08-18 Aceptado: 2024-12-02

To cite this article: Marin-Garcia, J.A., Martinez-Tomas, J., Juarez-Tarraga, A. & Santandreu-Mascarell, C. (2024). Protocol paper: From Chaos to Order. Augmenting Manual Article Screening with Sentence Transformers in Management Systematic Reviews. *WPOM-Working Papers on Operations Management*, 15, 172-208. doi: <https://doi.org/10.4995/wpom.22282>

### Abstract

A spanish version of the article is provided (see section before Acknowledgements)

As scientific output grows, systematic reviews have become essential yet increasingly challenging. Our approach to this protocol aims to make this process more effective, efficient and accessible to researchers worldwide, including those in developing countries.

We developed a tool to complement human judgment in the screening phase using pre-trained language models and natural language processing techniques. This tool generates text embeddings and calculates semantic similarities, prioritizing potentially relevant articles. The goal is to utilize the similarity ranking instead of reviewing articles randomly or following the relevance sort option of search engines like WOS or Scopus. Coders can start with those closest to the category/categories of interest and progressively move towards the more distant ones. This approach would save time and effort while reducing the fatigue and biases of the coders.

The models we have tested in this research are all-MiniLM-L6-v2, all-distilroberta-v1, all-mpnet-base-v2, paraphrase-multilingual-mnlpert-base-v2, distiluse-base-multilingual-cased-v1, all-MiniLM-L12-v2, allenai-specter, allenai/scibert\_scivocab\_uncased, distilbert-base-nli-mean-tokens, roberta-base-nli-stsb-mean-tokens, distiluse-base-multilingual-cased-v2, paraphrase-multilingual-MiniLM-L12-v2, stsb-roberta-large, bert-base-nli-mean-tokens.

The method was implemented using limited computational resources and open-source software, ensuring accessibility for research teams with restricted economic resources.



172

- As far as I know (or investigated in August), I am the first to have published the use of Sentence Transformers to help screening. All others have explored the other ways
- Being the first does not imply that it is on the right path
  - I can be the first to discover "El Dorado"
  - Or I could be the first to discover a precipice [plunging into it]

<https://doi.org/10.4995/wpom.22282>

UPV

Juan A. Marin-Garcia

Universitat Politècnica de València-DOE-ROGLE-IEMA  
[www.upv.es](http://www.upv.es)

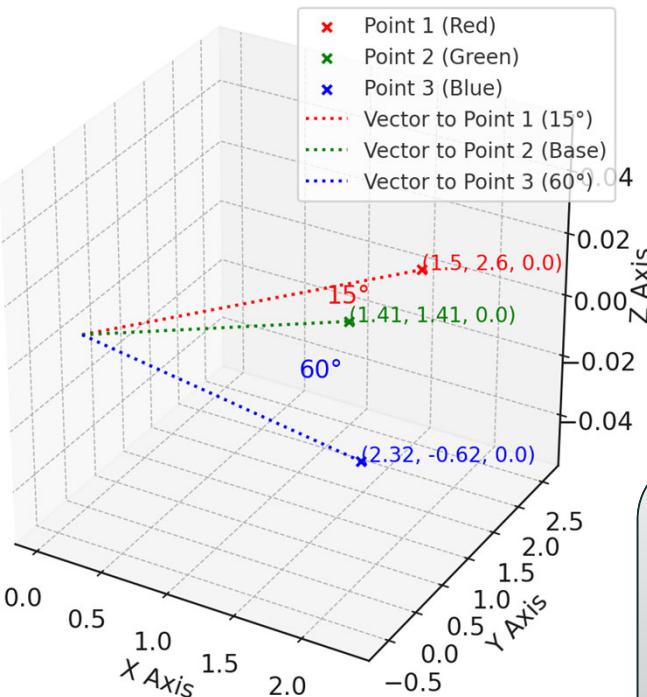


102

What is this about the embeddings and the distance of cosine that I will use in step 5 ...  
any of the red points and the green are closer than the blue of the green because  $\cos(15^\circ) = 0.96$ , is greater than  $\cos(60^\circ) = 0.5$

This is an example with 3 dimensions ...  
now imagine a space of 384 dimensions  
or 512 or 1024 or ...

In that space, every word,  
each phrase, each text has a  
point (a numbers vector)



Operation: cats - cat + dog  $\approx$  dogs  
Similarity with objective: 0.9890  
Top 5 more similar words:  
Dogs: 0.9890  
Dog: 0.9703  
Cats: 0.6191  
Cat: 0.5864

1. Usando león-leona como base:  
enfermera: 0.9427  
enfermero: 0.9426  
medico: 0.6528  
cirujano: 0.6226  
cirujana: 0.5139

2. Usando hombre-mujer como base:  
doctora: 0.8570  
doctor: 0.8207  
médica: 0.7520  
medicina: 0.6526  
enfermera: 0.6266

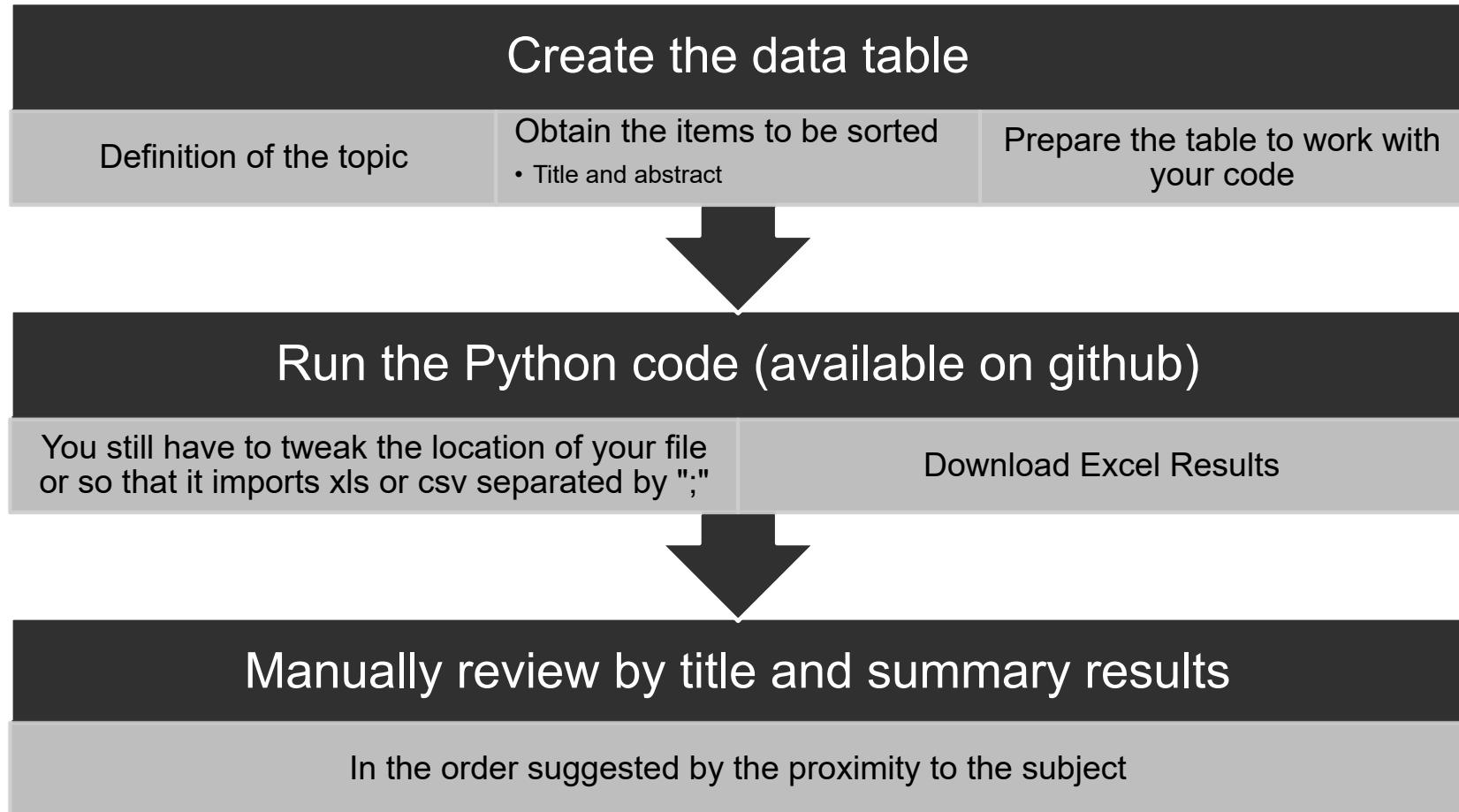
1. Usando España-Madrid como base:  
París: 0.6902  
Londres: 0.4803  
Roma: 0.4711  
Salamanca: 0.2656

But before you come up ...  
Operation: Madrid - Spain + France  $\approx$  Paris  
Similarity with objective: 0.6902  
Top 5 more similar words:  
France: 0.7202  
Madrid: 0.7023  
Paris: 0.6902  
Spain: 0.3597

Usando león-leona como base:  
ingeniero: 0.9438  
ingenieria: 0.6196  
profesional: 0.5980  
ingeniera: 0.5003  
ingenio: 0.4538

## Step 5: Stages for screening with classification by relevance with the sentence transformers

<https://github.com/jamg-upv/LLMforSLRscreening/tree/main/UBdocSchoolMAy25>



## Case. Identify the articles dealing with Remote Work, published in the last 4 years in several WOS magazines

|  |                     |   |
|--|---------------------|---|
|  | Remote Work         | "Refers to a flexible work arrangement where Employees performs their job duties from a location outside the traditional office environment, typically using information and communication technologies to maintain connectivity with their organization and colleagues Distributed Work, Mobile Work, E-Work, Distance Work, Flexible Work, Work From Home (WFH), Offsite Work, Geographically Dispersed Work, Networked Work, Location-Independent Work, Remote Employment, Satellite Work "  |
|  | HiwpshortDescriptor | High Involvement Work Program (Hiwps) are a set of interconnected human resource management Strategies and organization programs designated to enhance employee engagement, autonomy, and decision-making capabilities. These programs aim to Align Employee Skills, Knowledge, and motivation with organization Organizational Goals, Thereby Improving Both Individual and Organizational Performance. Synonyms: High-Performance Work Practices (HPWPS); High-Commitment Work practices; High-Engagement Work practices; High-Involvement Management Practices; Innovative work practices; Employee Involvement practices; Participatory Management Practices; High-performance human resource practices; Advanced Human Resource Management Systems; Commitment-Based HR Systems; Performance-Enhancing HR practices; High Involvement Work practices; Hiwp; High Commitment Work practices; HCWP; High Performance Work Systems; HPWS; Strategic Human Resource Management Practices |

Let's do it step by step. It will be about 5 minutes. I do it with my case, you can use mine (all routes will work for you) or with your own data

- The critical point is whether you can connect to "My" Github and the connection to your Google Drive for End of Results
- It can also be done at “Colab” without connect with drive
  - If you close connection before download results I will be lost
- We see directly the first 20 results in Google Colab
- Then you download an Excel with the results

## Some execution time data and models

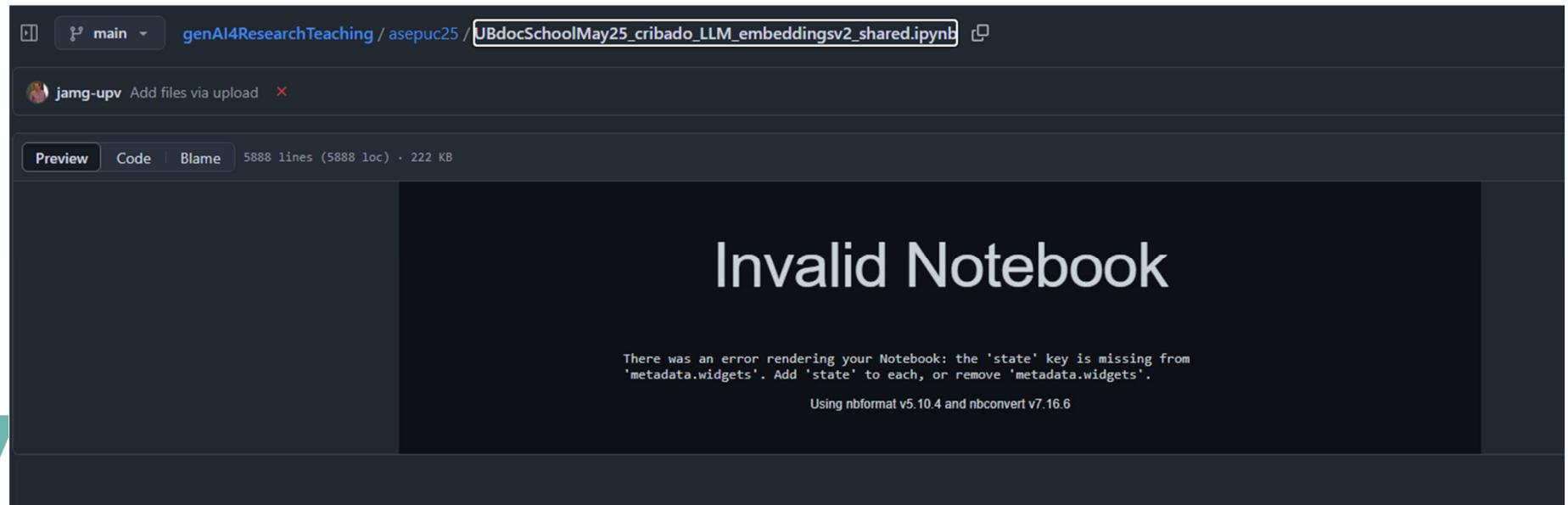
- `models_to_test = [`
- `'All-minilm-L6-v2',`
- `'All-Distilrobe-V1',`
- `'All-Mpnet-Base-V2',`
- `'All-Milm-L12-V2',`
- `'Allanai-Specter '`
- `]`
- With 5 models it takes about 20-30 minutes
- I show you it only with one of the quick [all-minilm-l6-v2] (approximately 3 minutes)
- The used models ('All-Milm-L6-V2', 'All-Distilrobe-V1', 'All-Mpnet-Base-V2', 'All-Milm-L12-V2', 'Allenai-Specter') are all of the Bert family or similar, not GPT or T5:
- 'All-Milm-L6-V2' and 'All-Milm-L12-V2 ': Based in Bert/Minilm
- 'All-Distilrobe-V1 ': Based on Roberta (which is an optimization of Bert)
- 'All-Mpnet-Base-V2 ': MPNET based (Bert evolution)
- 'Allenai-Specter ': Based on Bert type architecture
- Bert and similar are bidirectional encoders, ideal for creating embeddings
- GPT is unidirectional, designed for text generation
- T5 is oriented to transformation tasks text-to-text

First 20 results (those that have the “embeddings” most similar to those of the SET definition)

| A  | B  | C           |
|----|--|-------------|
| 1  | ct_Compct  | Average_Ran |
| 2  | id0674--10 Is student evaluation of teaching worthwhile? An analytical framework                           | 1           |
| 3  | id0024--10 Student evaluations of teaching (SET): Guidelines for their use. Studer                         | 2           |
| 4  | id0330--10 Taking the grading leniency story to the edge. The influence of student evaluations of teaching | 3           |
| 5  | id0384--10 Redesigning student evaluations of teaching: integrating faculty and student perspectives       | 4           |
| 6  | id0411--10 Coherence in measuring student evaluation of teaching: a new paradigm                           | 5           |
| 7  | id0579--10 A framework for using SET when evaluating faculty. This paper presents                          | 6           |
| 8  | id0635--10 I liked your course because I believe in (the power of) student evaluations of teaching         | 7           |
| 9  | id0500--10 Student Evaluation of Teaching (SET): Why the Emperor Has No Clothes                            | 8           |
| 10 | id0347--10 Student Evaluations of Teaching: How You Teach and Who You Are. Stu                             | 9           |
| 11 | id0479--10 Student Evaluation of Teaching (SET) in Higher Education: How to Use Stu                        | 10          |
| 12 | id0069--10 Student Evaluations of Teaching Tools A Qualitative Examination of Stu                          | 11          |
| 13 | id0016--10 Student evaluations of teaching are an inadequate assessment tool for                           | 12          |
| 14 | id0165--10 A Review of Student-Centered Evaluation: Functions and Perceptions.                             | 13          |
| 15 | id0291--10 Is my opinion important in evaluating lecturers? Students' perceptions                          | 14          |
| 16 | id0391--10 Major increases in teachers' performance evaluations: Evidence from s                           | 15          |
| 17 | id0371--10 On the Validity of Student Evaluation of Teaching: The State of the Art                         | 16          |
| 18 | id0685--10 Student evaluations of teaching: unidimensionality, subjectivity, and b                         | 17          |
| 19 | id0446--10 The Key Researchers and Their Research on Student Evaluation of Teach                           | 18          |
| 20 | id0045--10 Considerations for Developing a Student Evaluation of Teaching Form.                            | 19          |
| 21 | id0417--10 A Guide for Making Valid Interpretations of Student Evaluation of Teac                          | 20          |

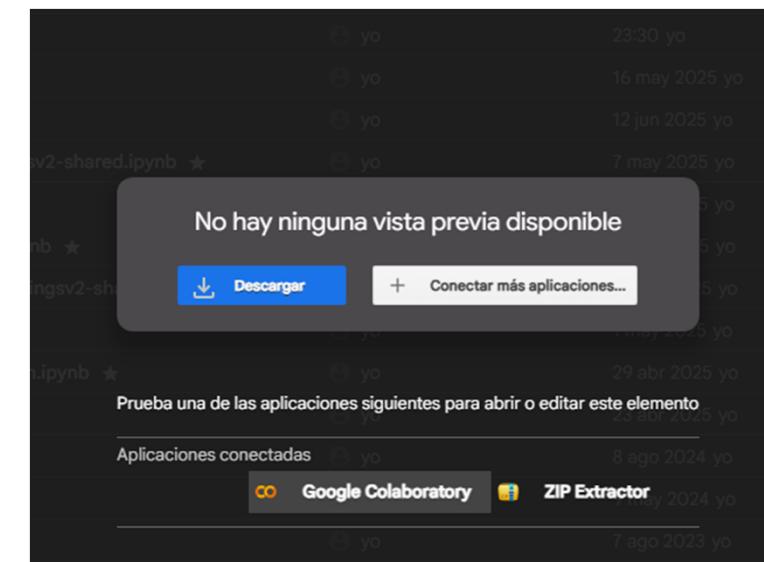
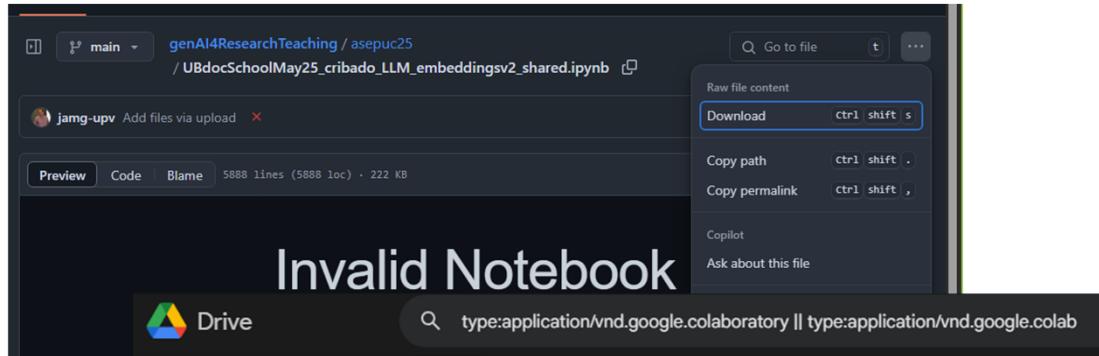
## Manos a la obra (hay un video explicando todo el proceso)

- This video explains step by step how to use the code: <https://youtu.be/9sppuIMGCVs> This one explains more in detail what the code does: <https://media.upv.es/#/portal/video/9bf1d0d0-5c06-11ef-984c-6373308b2f42>
- El fichero .ipynb es correcto (pero hay un problema entre Github y Google Colab reciente que no he podido resolver) hay que cargarlo en github manualmente (explico en la siguiente diapo)



Como descargar y subir a colab

[https://drive.google.com/drive/search?q=owner:me%20\(type:application/vnd.google.colaboratory%20||%20type:application/vnd.google.colab\)](https://drive.google.com/drive/search?q=owner:me%20(type:application/vnd.google.colaboratory%20||%20type:application/vnd.google.colab))



# COMPARATIVA DE MÉTODOS DE INTELIGENCIA ARTIFICIAL EN EL CRIBADO DE REVISIÓN SISTEMÁTICA DE LA LITERATURA

- Juan A. Marin-Garcia - Juan Martinez-Tomas
- ROGLE-DOE-Universitat Politècnica de València.
- Alina Diaz-Curbelo - Rafaela Alfalla-Luque
- Universidad de Sevilla

- **Problema:** el cribado manual en revisiones sistemáticas es laborioso y requiere ~2 minutos por artículo
  - Cuando estás fresco
    - La eficacia y eficiencia decae exponencialmente con el volumen de trabajo
- **Crecimiento exponencial** de la literatura científica
  - En una disciplina sin tesauros o taxonomías aceptadas
  - Las búsquedas bibliográficas contienen una cantidad enorme de falsos positivos (más del 90%)
- **Necesidad** de automatización sin perder eficacia

- Comparar el rendimiento de diferentes métodos de IA respecto al consenso humano:
  - Gold standard: cribado manual humano
    - ¿Realmente somo el gold standard (con 3 personas codificando)?
  - IA generativa (ChatGPT, Claude)
    - Chat Oficiales. Copia y pega (tiempo humano, posibilidad de errores); usa los antentes intermedios de ida y de vuelta
    - Conexión Via API. Automatico. Reproducible. Rapido
  - Sentence Transformers → proximidad de embeddings
  - Aprendizaje supervisado (Rayyan)

- Replicación del estudio de Alfalla-Luque et al. (2023)
  - 678 artículos de WOS y Scopus sobre "supply chain agility"
  - 140 retenidos tras screening
  - 70 verdaderos positivos identificados tras la fase de análisis de Full text
- Métodos comparados:
  - Cribado manual: 3 codificadores humanos
  - IA generativa: ChatGPT 4o, Claude 3.7, prompts A y B
  - Sentence Transformers: Embeddings semánticos
  - Rayyan: Aprendizaje supervisado con 50 ejemplos
- Innovaciones Metodológicas
  - Comparación Chat vs API: Verificar si hay iteraciones ocultas
  - Prompts estructurados vs párrafo
  - Temperatura controlada (0.2 y 0.7)
- Métricas múltiples
  - Recall ( $TP/(TP + FN)$ ), Precision ( $TP/ (TP+FP+TN+FN)$ ), Accuracy ( $(TP+TN)/ (TP+FP+TN+FN)$ ), Specificity ( $TN/(TN + FN)$ ), Error ( $((FP+FN)/ (TP+FP+TN+FN))$ ) tiempo invertido en completar el cíbrado

## Resulados. Tiempo

| Experiemento | Nombre columna excel             | Minutos consumidos en clasificar los 593 art | Plataforma          | Modelo (versión disponible febrero 2025) | Prompt |
|--------------|----------------------------------|--|---------------------|--|--------|
| ID001        | CodHuman2-mar25                  | 2749   | Codificacion Humana | CodHuman                                 |        |
| ID002        | Claude3.7-sonnetA1-na            | 80   | AntroChat           | Claude3.7-sonnet                         | A      |
| ID003        | Claude3.7-sonnetA2-na            | 73   | AntroChat           | Claude3.7-sonnet                         | A      |
| ID004        | Claude3.7-sonnetB1-na            | 50   | AntroChat           | Claude3.7-sonnet                         | B      |
| ID005        | Claude3.7-sonnetB2-na            | 47   | AntroChat           | Claude3.7-sonnet                         | B      |
| ID006        | rankNLPTransformersA             | 33   | Codigo Python       | rankNLPTransformers                      | A      |
| ID007        | rankNLPTransformersB (recalculo) | 2  | Codigo Python       | rankNLPTransformers                      | B      |
| ID054        | gpt-4o-mini-2024-07-18B1-0,2     | 20   | API                 | gpt-4o-mini-2024-07-18                   | B      |
| ID055        | gpt-4o-mini-2024-07-18B2-0,2     | 20   | API                 | gpt-4o-mini-2024-07-18                   | B      |
| ID058        | gpt-4o-mini-2024-07-18B1-0,7     | 20   | API                 | gpt-4o-mini-2024-07-18                   | B      |
| ID059        | gpt-4o-mini-2024-07-18B2-0,7     | 20   | API                 | gpt-4o-mini-2024-07-18                   | B      |
| ID060        | o4-mini-2025-04-16A1-1           | 120  | API                 | o4-mini-2025-04-16                       | A      |

## Resultados. Recuperación

|                         | Gold standard.<br>Cribado manual<br>(Alfalla-Luque<br>et al., 2023) | Cribado humano 3<br>coders MArz25 | Claude3.7A1 | Claude3.7A2 | Claude3.7B1 | Claude3.7B2 | o4-mini-2025-04-<br>16B2 | gpt-4o-mini-2024-<br>07-18 B1-0.2 | gpt-4o-mini-2024-<br>07-18 B2-0.2 |
|-------------------------|---|-----------------------------------|-------------|-------------|-------------|-------------|--------------------------|-----------------------------------|-----------------------------------|
| TO                      | 678   | 593                               | 593         | 593         | 593         | 593         | 593                      | 593                               | 593                               |
| TP                      | 70  | 63                                | 66          | 66          | 66          | 67          | 60                       | 65                                | 64                                |
| FP                      | 74  | 97                                | 251         | 261         | 180         | 201         | 104                      | 275                               | 256                               |
| TN                      | 534   | 426                               | 272         | 262         | 343         | 322         | 419                      | 248                               | 267                               |
| FN                      | 0   | 7                                 | 4           | 4           | 4           | 3           | 10                       | 5                                 | 6                                 |
| Recall<br>(sensitivity) | 100%  | 90%                               | 94%         | 94%         | 94%         | 96%         | 86%                      | 93%                               | 91%                               |
| Precision               | 10%   | 11%                               | 11%         | 11%         | 11%         | 11%         | 10%                      | 11%                               | 11%                               |
| Accuracy                | 89%   | 82%                               | 57%         | 55%         | 69%         | 66%         | 81%                      | 53%                               | 56%                               |
| Specificity             | 100%  | 98%                               | 99%         | 98%         | 99%         | 99%         | 98%                      | 98%                               | 98%                               |
| Error                   | 11%   | 18%                               | 43%         | 45%         | 31%         | 34%         | 19%                      | 47%                               | 44%                               |

-  **Claude 3.7 y o4-mini-2025-04-16B2** Rendimiento supera expectativas vs. humano
  - Prompt B más eficiente que Prompt A
- **Eficiencia práctica:**
  -  **160-240 TP identificados de 593 artículos**
    - VS 160 cribados manual, con prácticamente mismo impacto
    - En 50 minutos copia-pegá o 20 minutos tomando café vs 2800 minutos
  -  **Reducción de, al menos, el 90% del tiempo de cribado**
  -  Concentra artículos relevantes en subconjunto manejable (el 30%-50% del conjunto original)
  - Y ya estamos con Claude4... ver como funciona este modelo con los resultados

Si comparamos el resultado de Transformers (en este caso no tenemos una clasificación sino un rango de mas (1) a menos (593) afín al prompt. Es decir números mas bajos serian mas afines.

|                      | Promedio de rankNLPTr | Promedio de rankNLPT | Promedio de rankNLPT | TransformersB |
|----------------------|-----------------------|----------------------|----------------------|---------------|
| 0 P                  | 191                   | 195                  |                      |               |
| 1 N                  | 311                   | 311                  |                      |               |
| 2 Total general      | 297                   | 297                  |                      |               |
| 3                    |                       |                      |                      |               |
| 4                    |                       |                      |                      |               |
| 5                    |                       |                      |                      |               |
| 6                    | Promedio de rankNLPTr | Promedio de rankNLPT | Promedio de rankNLPT | TransformersB |
| 7 Cat1InsuffInformat | 175                   | 179                  |                      |               |
| 8 Cat2Sele           | 218                   | 221                  |                      |               |
| 9 Cat3Maybe          | 259                   | 263                  |                      |               |
| 0 CatDiscard         | 325                   | 324                  |                      |               |
| 1 Total general      | 297                   | 297                  |                      |               |
| 2                    |                       |                      |                      |               |

TP es lo que vamos a considerar True Positive de la fase de cribado manual de 2025 (select, maybe y falta informacion en el abstract). TN true negative del cribado manual de 2025).

P son los que en IJOMP metaanalizamos (es decir los que retuvimos tras leer PDFs). N los que se descartaron en cribado o por no tener pdf o por no tener pdf con datos útiles, en IJOMP.

- **Potencial de reducción significativa del tiempo de cribado**
- **Validación necesaria** de métodos automatizados
- **Consideraciones éticas** sobre sesgos y transparencia
- **Futuro de las revisiones sistemáticas** con IA

## Preguntas al experto

Preguntas sobre IA generativa en investigación



Juan A. Marin-Garcia

Universitat Politècnica de València-DOE-ROGLE-IEMA  
[www.upv.es](http://www.upv.es)



ROGLE  
Reengineering Operations  
GroupWork Logistics Excellence



döe  
Departamento de Organización de Empresas



121



### Fase 4: (al final del taller)

**Juan hará de “experto” al que entrevistáis**

**Se elegirá una pregunta al azar con ruleta de wooclap (si no sale del grupo, 1. se vuelve a lanzar ruleta)**

**Una persona de cada grupo hará de relator de las respuestas del “experto” (luego se las hacéis llegar a [jamarin@omp.upv.es](mailto:jamarin@omp.upv.es) y las subiremos al github)**

**Una persona de la clase hará de moderador de la entrevista: fijará tiempos, intervenciones, cortará al “experto” para que no se enrolle más de la cuenta**

## Questions, doubts, comments

