



Human-AI Collaboration in Systematic Reviews

(from RQ definition to screening)

- While you wait, you can download the PDF, open a browser with your Gmail account (or if you have OpenAi or Poe or Claude) and think a SLR theme that might interest you. So you can clone the examples that I am going to propose



QR for access to material discharge (PDF, texts for tests ...)
<https://github.com/jamg-upv/LLMforSLRscreening/tree/main/UBdocSchoolMAy25>

This is the "guilty" information for which you are here today

This workshop explores effective collaboration between researchers and artificial intelligence tools in conducting screening step in systematic literature reviews (SLRs).

The presentation covers a structured approach to SLRs, demonstrating how different AI tools can support various phases of the review process:

1. Formulating research questions using generative AI
2. Defining key concepts and identifying synonyms
3. Developing search strategies across academic databases
4. Screening articles with three different AI approaches, including sentence transformers
5. Creating PRISMA flow diagrams
 - Through practical examples and demonstrations using Claude, chatGPT, Gemini, and specialized sentence transformer models, the presenter shows how researchers can leverage AI capabilities while maintaining research integrity. The talk includes hands-on Python code examples for article classification and provides a realistic assessment of where AI currently adds value versus where human expertise remains essential. Rather than positioning AI as a replacement for researcher expertise, the presentation advocates for thoughtful integration of AI tools as research assistants that can reduce tedious tasks and potentially improve review quality when properly supervised.

Learning outcomes for 2 hours

- 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 IAgén
 - To do this, you must know the VA model applied to IAgén
- 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.

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)

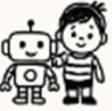
After these "mantras" may be you decide that it is not worth listening to me



- It's not what AI can do, but what I'm interested in AI doing
 - I like writing, I don't want AI to write for me
 - I like thinking, I don't want AI to think for me
 - I like making decisions, I don't want AI to make decisions for me
- If you do not like writing, thinking and making decisions, you don't like to do research

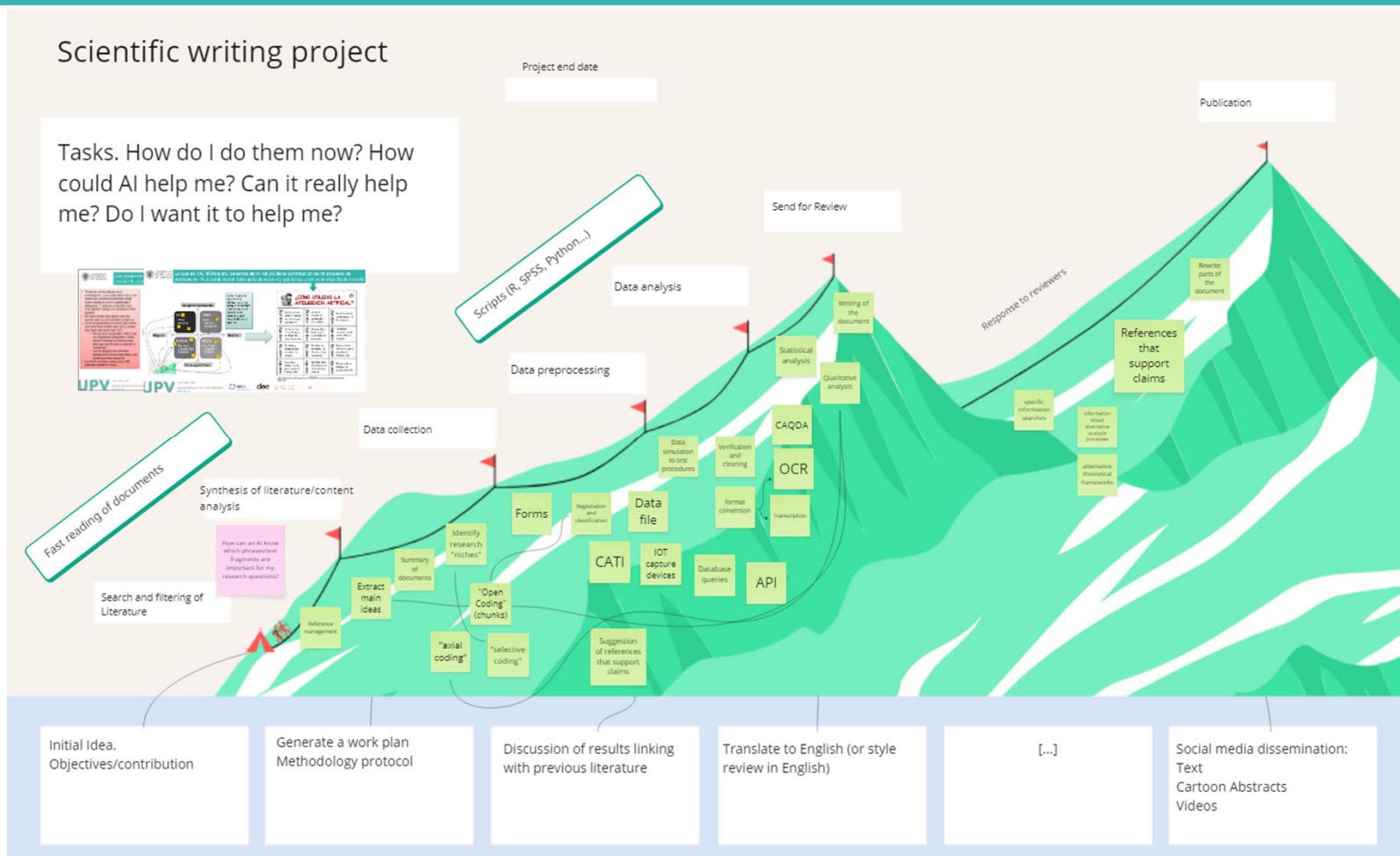
I insist with uncomfortable questions.

If you delegate everything in IAGEN, what do you learn? What do you contribute?

 HOW DO YOU USE ARTIFICIAL INTELLIGENCE?		
 AI does all the work without you having to think	 AI writes the content and you edit it	 AI offers content and you improve it
 AI creates drafts and you choose the best parts	 You note down basic ideas and AI creates a draft for you	 While you write, AI gives you ideas and guides you
 You write and ask AI for help if you get stuck	 You write a draft, AI writes another, and you compare them	 You search for data with AI, but write the work alone
 You consult AI for ideas, but do the work alone	 You write all the work and AI reviews and improves it	 You do all the work without AI help

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

Scientific writing project

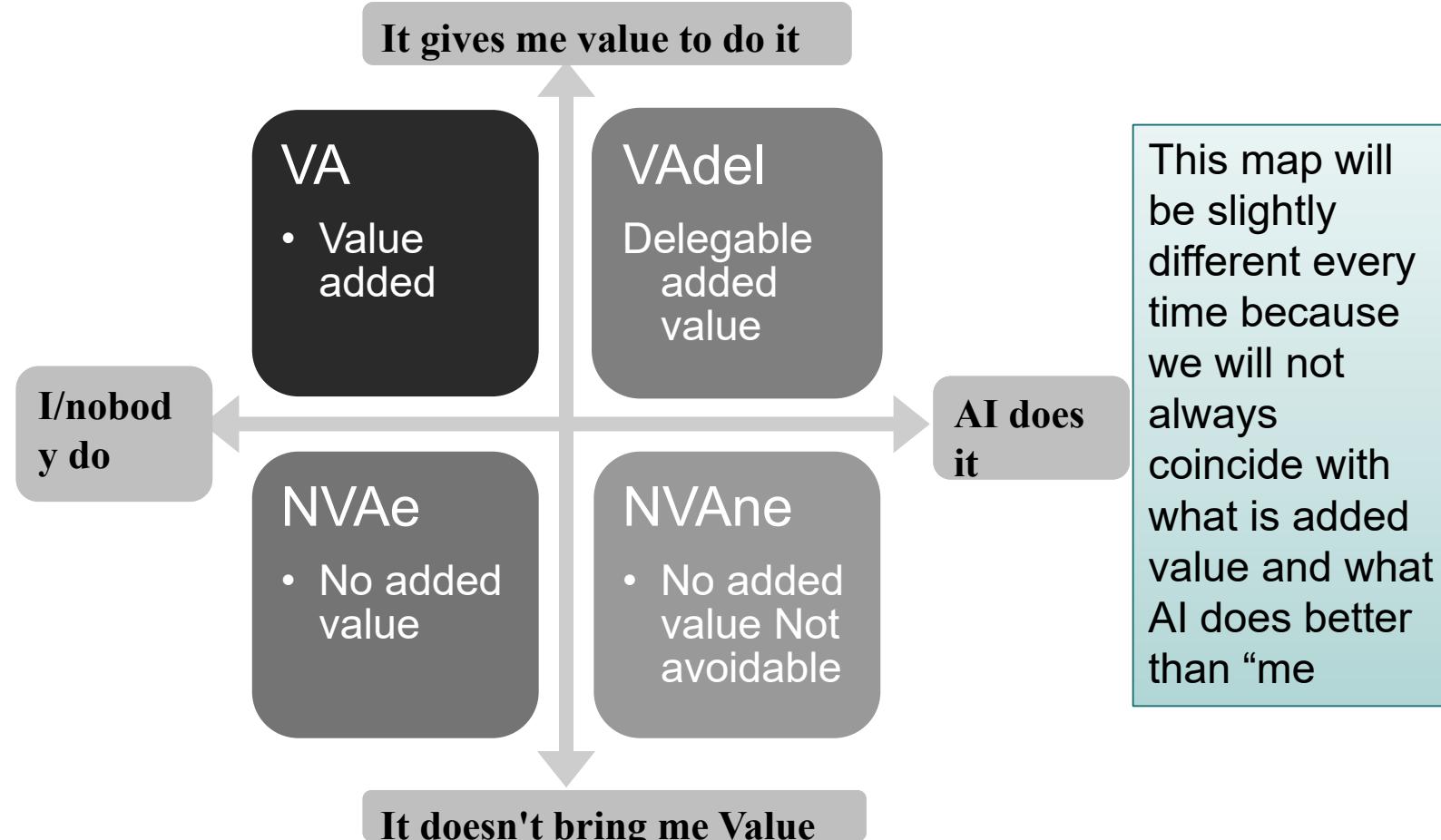


Juan A. Marin-Garcia

Universitat Politècnica de València-DOE-ROGLE-IEMA
www.upv.es

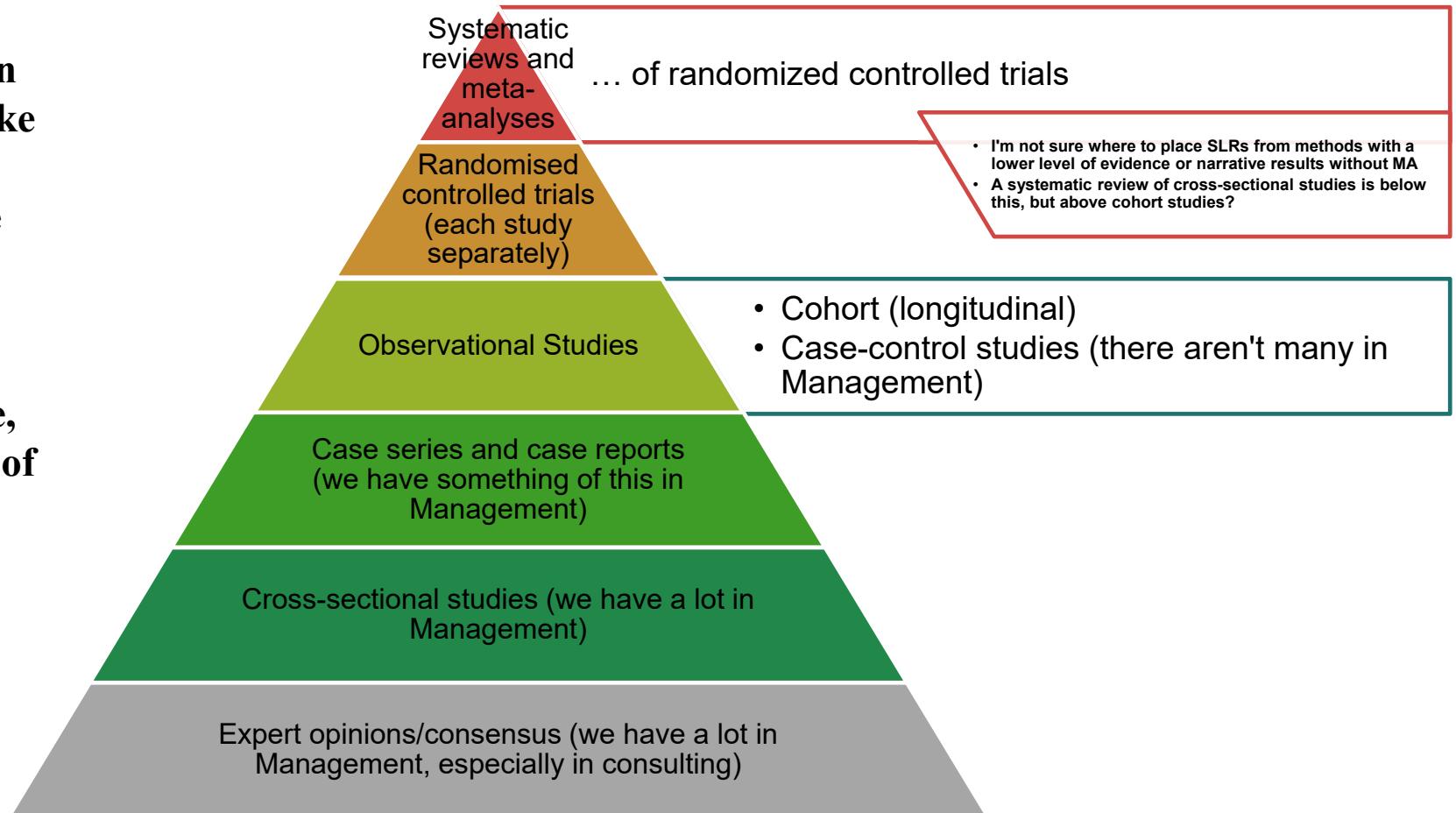
VA Model

- The key questions (and in this order):
 - Can Algen really help me in these tasks?
 - Do I want Algen to help me?
 - How could AI help me?
- There are things that AI does well (better than "me") and things it does wrong (worse than "me")
 - Of those that add value to me, some I will be interested in delegating and others I will want to do even if it is worse than AI so as not to lose the "expertise"
 - NVAnes are done by us only if the AI does not do them well
- Avoidable NVAs should NEVER be done

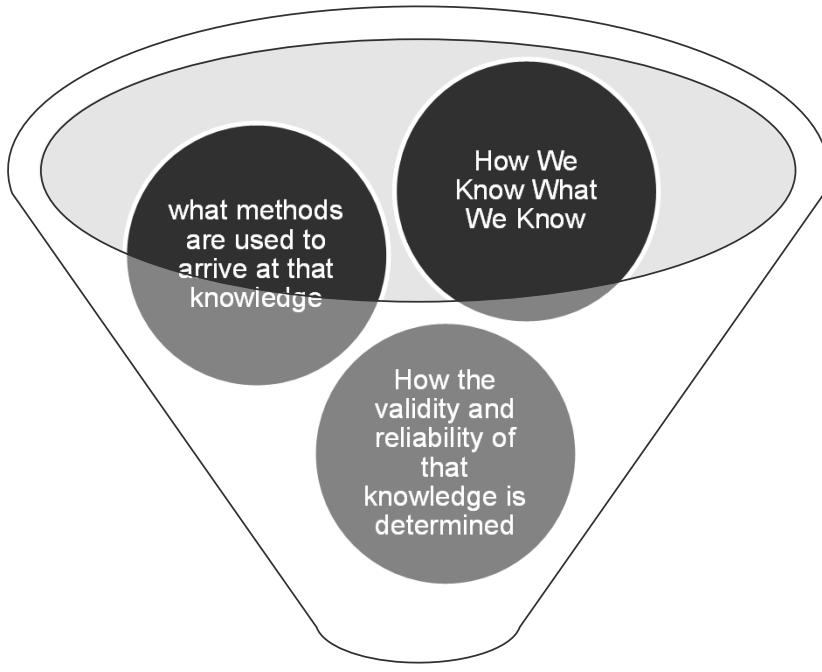


From the positivist point of view. The synthesis of evidence is at the cusp of the pyramid of scientific evidence (applies to ma, I do not know if to all SLR)

**Although there is Algen
we need to learn to make
SLR (some people).
In about 10 years? The
entire process will be
automated (search,
filtering, coding or
extraction of data table,
analysis and summary of
conclusions).**



Epistemological Consideration: The previous slide is true from a positivist vision of science



Epistemology: the study of knowledge, its nature, origin and limits

Interpretive or constructivist approach:

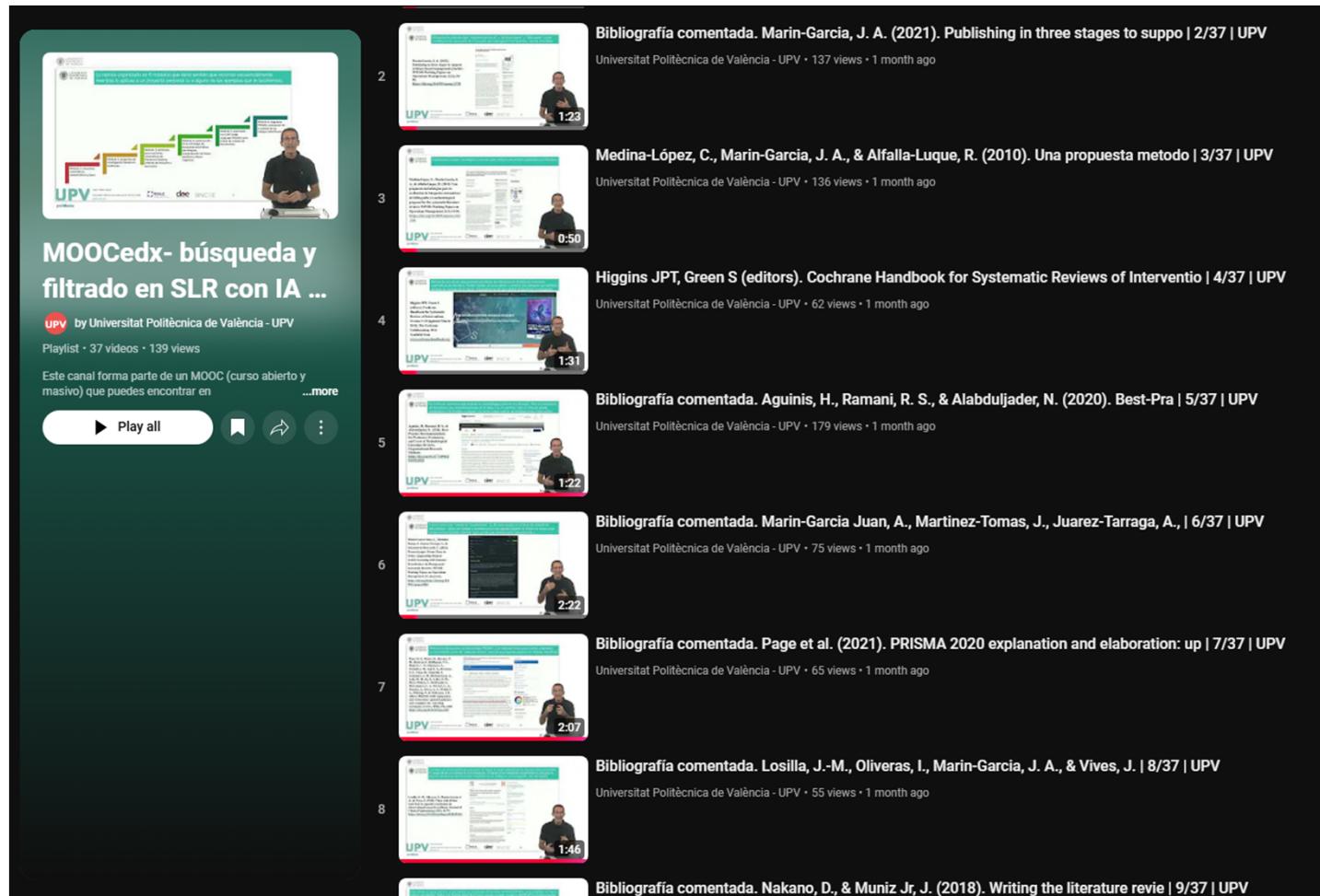
1. Qualitative methods such as in-depth case studies, ethnographies, or action research are more valued.
2. Validity is based more on the richness and depth of the data than on statistical generalization.

Critical perspective:

1. Methods that reveal power structures and promote social change are prioritized.
2. Participatory research and critical action research are highly valued.

Supporting videos (in Spanish). Bibliography on SLR commented (they are part of the MOOC but you do not need to do the MOOC if you only are interested in this)

<https://www.youtube.com/playlist?list=PL6kQim6ljTJsQvBXx8G1nE1iOi-iTb4JY>



MOOCedx- búsqueda y filtrado en SLR con IA ...

UPV by Universitat Politècnica de València - UPV

Playlist · 37 videos · 139 views

Este canal forma parte de un MOOC (curso abierto y masivo) que puedes encontrar en [...more](#)

Play all

2 Bibliografía comentada. Marin-Garcia, J. A. (2021). Publishing in three stages to suppo | 2/37 | UPV
Universitat Politècnica de València - UPV · 137 views · 1 month ago

3 Medina-López, C., Marin-Garcia, J. A., & Alfalla-Luque, R. (2010). Una propuesta metodo | 3/37 | UPV
Universitat Politècnica de València - UPV · 136 views · 1 month ago

4 Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventio | 4/37 | UPV
Universitat Politècnica de València - UPV · 62 views · 1 month ago

5 Bibliografía comentada. Aguinis, H., Ramani, R. S., & Alabduljader, N. (2020). Best-Pra | 5/37 | UPV
Universitat Politècnica de València - UPV · 179 views · 1 month ago

6 Bibliografía comentada. Marin-Garcia Juan, A., Martinez-Tomas, J., Juarez-Tarraga, A., | 6/37 | UPV
Universitat Politècnica de València - UPV · 75 views · 1 month ago

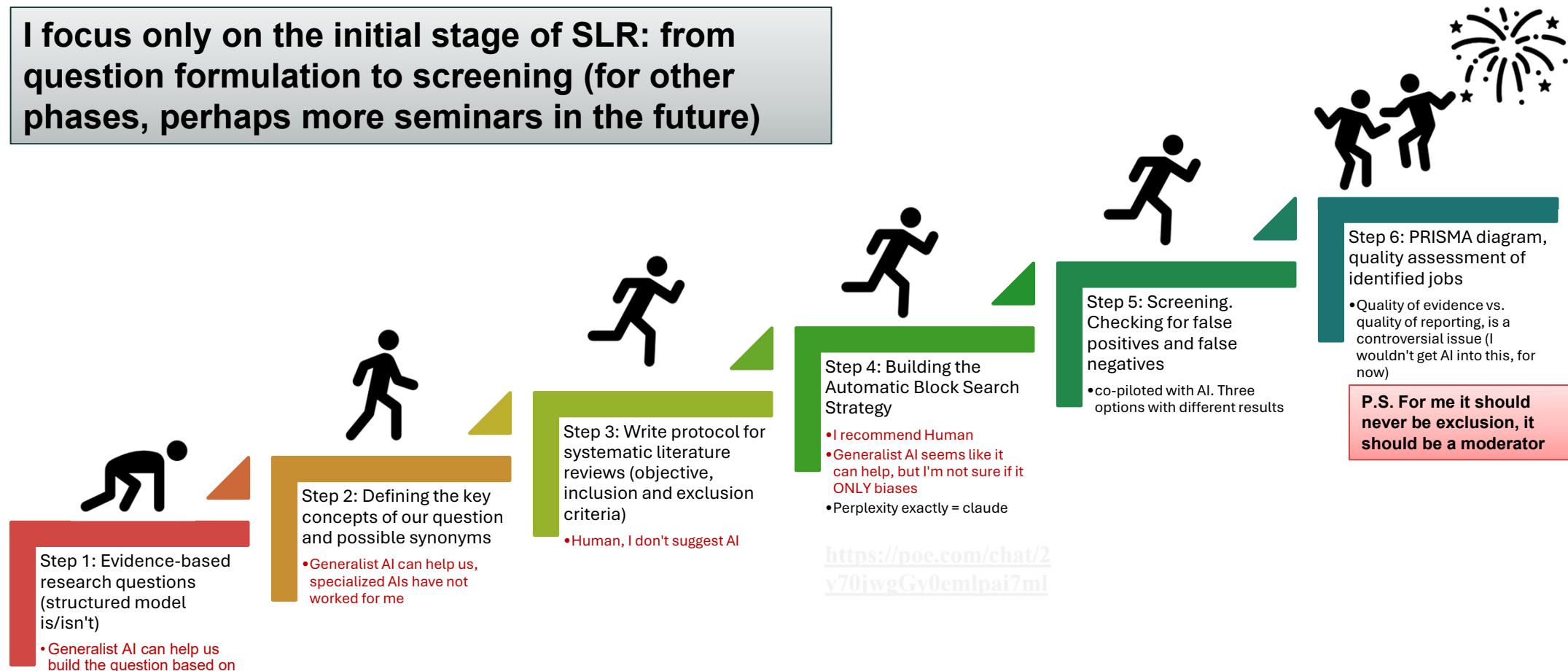
7 Bibliografía comentada. Page et al. (2021). PRISMA 2020 explanation and elaboration: up | 7/37 | UPV
Universitat Politècnica de València - UPV · 65 views · 1 month ago

8 Bibliografía comentada. Losilla, J.-M., Oliveras, I., Marin-Garcia, J. A., & Vives, J. | 8/37 | UPV
Universitat Politècnica de València - UPV · 55 views · 1 month ago

Bibliografía comentada. Nakano, D., & Muniz Jr, J. (2018). Writing the literature revie | 9/37 | UPV

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)



Four basic decisions in a systematic review (today we will talk about three of them)

Choose the topic (ask the research question)

Evidence-Based Research Question (MOOC SLR)



Set SLR parameters

Search (MOOC SLR)

Analysis (MOOC pending)



Choose the software that can support the chosen parameters

MA (R –[metafor](#)-, [Meta-essentials](#), [CMA](#), [RevMan](#))

[Atlas.Ti](#), [Nvivo](#)

[Rayyan](#)

GPTs (Claude, OpenAI...)

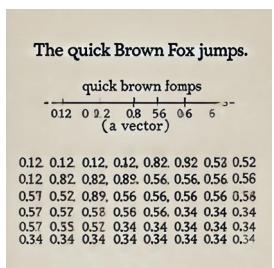
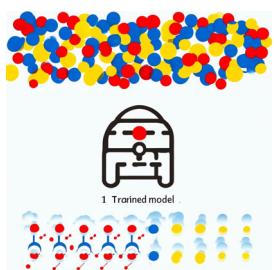
Perplexity, consensus, scite, ...



Extract the data (and pre-process it)

(MOOC pending)

Terminology: four types of AI that I am going to use. The accompanying last name is very relevant.



- Generative AI

- Generalist generative (example Claude3.7.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
 - (Rayyan as an example)

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

	Is	It's not
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 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>



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)

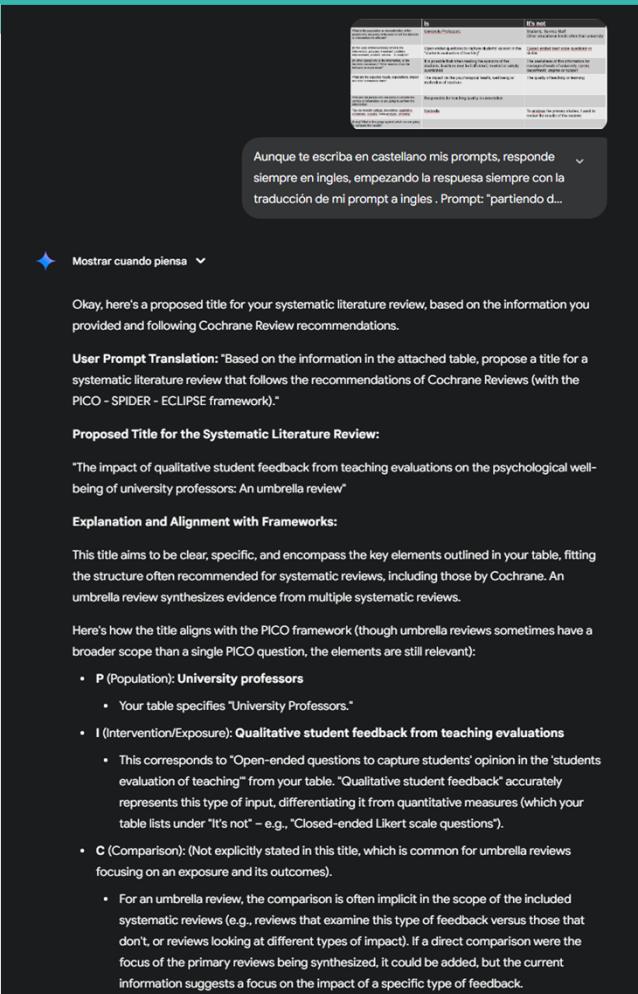
Here's a proposed title for the systematic review following Cochrane recommendations:

"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**
- **P (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



<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)

- 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)

The screenshot shows a YouTube playlist page with the title 'Concretando la definición de los términos clave. Con ayuda de IA. Comparando plataform'. The page displays 26 video thumbnails, each showing a screen recording of a presentation or interface. The first thumbnail shows a slide with the text 'MOOCedx- búsqueda y filtrado en SLR con IA ...'. The subsequent thumbnails show various stages of the process, with titles like 'Concretando la definición de los términos clave. Comparando plataform | 20/37 | UPV', 'Concretando la definición de los términos clave. Comparando plataform | 21/37 | UPV', and so on up to 'Concretando la definición de los términos clave. Comparando plataform | 26/37 | UPV'. Each thumbnail includes the UPV logo and a timestamp.

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



[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"

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

Comprobada hasta al 6, son correctas, comprobad el resto

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>

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	Export			
Search	"student* evaluat* of teach**" (Topic) and Preprint Citation Index (Exclude – Database) and Review Article (Document Types)	All Databases Show collections	35	
Search	"student* evaluat* of teach**" (Topic) and "open ended" (Topic) and Preprint Citation Index (Exclude – Database)	All Databases Show collections	35	
Search	"student* evaluat* of teach**" (Topic) and Preprint Citation Index (Exclude – Database)	All Databases Show collections	1,311	

Juan A. Marin-Garcia

Universitat Politècnica de València-DOE-ROGLE-IEMA
www.upv.es



ROGLE
Reengineering Operations
GroupWork Logistics Excellence



d^oe
Departamento de Organización de Empresas



27



ASOCIACIÓN CIENTÍFICA
DE ECONOMÍA Y DIRECCIÓN
DE LA EMPRESA
DIRECCIÓN DE OPERACIONES Y TECNOLOGÍA



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 with title and definitions

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

<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

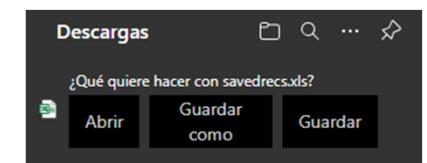
50/337 Add To Marked List

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]

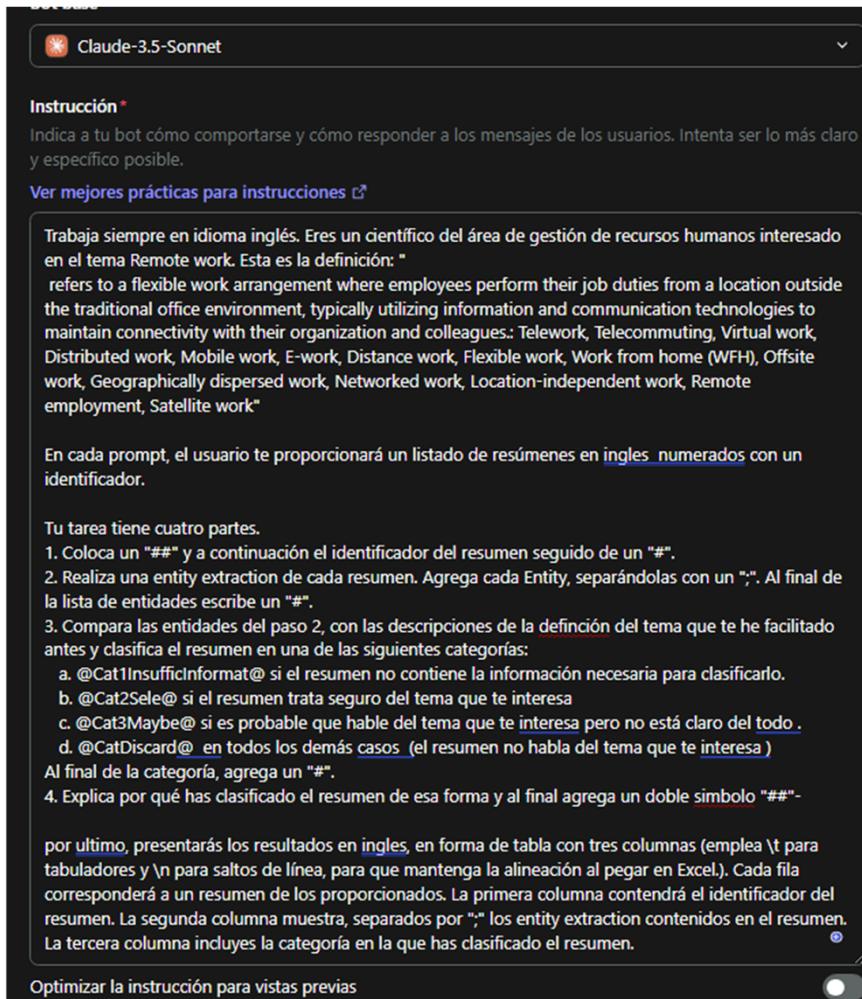
- Verbeek, J., Salmi, J., Pasternack, I., Jauhiainen, M., Laamanen, I., Schaafsma, F., HULSHOF, C., & Van Dijk, F. (2005). To Search Strategy for Occupational Health Intervention Studies. *Occupational and Environmental Medicine*, 62 (10), 682-687.
<https://doi.org/10.1136/OEM.2004.019117>
- Devillé, 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 things. Can you do a classification homework?



The screenshot shows the Claude-3.5-Sonnet interface. In the top bar, it says "Claude-3.5-Sonnet". Below that is a section titled "Instrucción *". It contains a text input field with the following instructions:

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. @Cat2Semi@ 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



The screenshot shows the advanced settings for ChatGPT-01. It includes the following options:

Avanzado ^

Sugerir respuestas

Renderizar contenido markdown

Temperatura personalizada

Controla la creatividad de las respuestas del bot. Los valores más altos producen respuestas más variadas pero impredecibles, los valores más bajos generan respuestas más coherentes.
Predeterminado: 0.50

A slider for Temperature is set to 0.30.

In March 2024 (from API) failed loudly, with a horrifying classification and unfortunate justification.

Things have changed 8 months later from web interface (I have confirmed yet with API)

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##

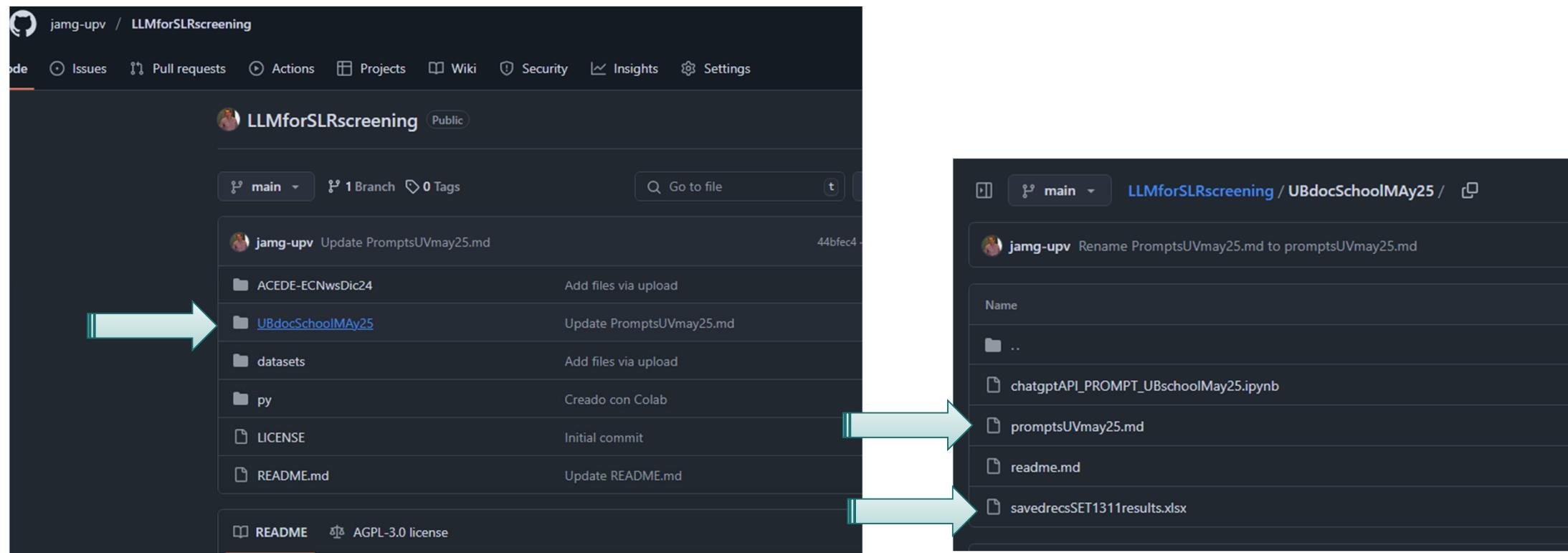
##10.1108/PR-02-2022-0085#

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

@Cat2Sele#

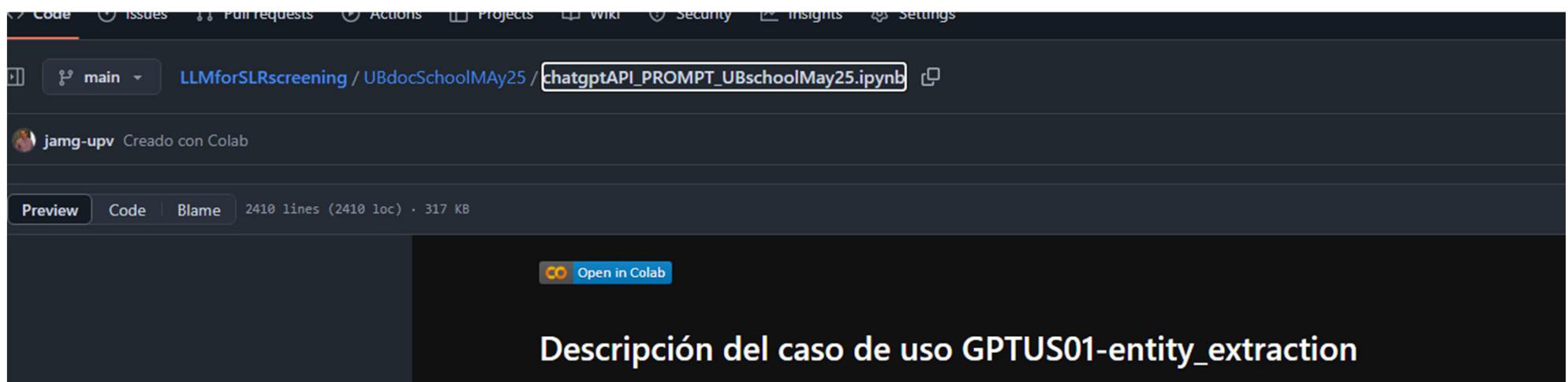
This abstract focuses on remote workplace considerations specifically for neurodivergent employees##

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)



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 'LLMforSLRscreening / UBdocSchoolMay25'. The repository contains a file named 'chatgptAPI_PROMPT_UBschoolMay25.ipynb'. The code in the file is a Python script for generating AI responses using the OpenAI API. It includes imports for 'time' and 'client.chat.completions.create'. The main function 'generate_response_and_update_df' takes parameters for the DataFrame, row index, model, system prompt, user prompt, and temperature. It constructs a message list with system and user prompts and calls the API with the specified model and temperature.

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

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#@Cat2Sel
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#@Cat3Maybe@#The abstract does not focus on student evaluations or teacher 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#@Cat4Discard@#The abstract does not focus on student evaluations or teacher evaluations.
Procesada fila 430 de 500

A	B	C	D	E	F	G
GPTId	GPTContent	PTRespons	GPTUsage	gmtAgregado		
1	0 ##id0001#open-ended questions; assessment indicators; teaching quality; students' evaluation of teaching; optimization; rational analysis#@Cat ChatCompleto id0001# Study on Data Analysis					
2	1 ##id0002#Online Student Evaluation of Teaching; productivity; implementation; stakeholder consultation; user training; primary data#@Cat4Disc ChatCompleto id0002# Development and Pilot					
3	2 ##id0004#Student evaluation of teaching; teaching quality assurance; higher education institutions; evaluation system; student feedback; misus@ChatCompleto id0004# On Student Evaluation					
4	3 ##id0006#Student evaluation of teaching; likability; student perceptions; evaluations; effectiveness; variance#@Cat4Discard#@The abstract disc ChatCompleto id0006# The student evaluation					
5	4 ##id0008#Students' evaluation of teaching; teaching quality evaluation method; teacher performance evaluation tool; reliability; validity; influer ChatCompleto id0008# A literature review of t					
6	5 ##id0009#Course evaluation; standardized transformation; student evaluation of teaching; course characteristics; male students; higher grades; s ChatCompleto id0009# Course evaluation mod					
7	6 ##id0010#game theoretic model; student evaluation of teacher; empirical test; correlation; average grades; grade inflation#@Cat4Discard#@The ChatCompleto id0010# Is student evaluation o					
8	7 ##id0012#Student evaluations of teaching; flipped classroom; traditional lecture; recorded lectures; course design; teaching evaluations; studen ChatCompleto id0012# Flipped pedagogy and s					
9	8 ##id0013#Students' Evaluation of Teaching; Higher Vocational Colleges; teachers' participation; students' ability to evaluate; evaluation methods ChatCompleto id0013# Rethink on the Student					
10	9 ##id0014#student evaluation of teacher; game theoretic model; empirical test; correlation; average grades; Management introductory course#@ ChatCompleto id0014# Is student evaluation o					
11	10 ##id0016#Student evaluations of teaching; faculty performance; ina					
12	11 ##id0017#Faculty development program; student evaluations of te					
13	12 ##id0019#student evaluation of teaching; methodological inadequa					
14	13 ##id0021#open-ended questions; student evaluation of teaching; i					
15	14 ##id0022#Student evaluations of teaching; open-ended questions;					
16	15 ##id0106#Student eval @Cat2Sele@ The study fo ChatCompleto id0106# Evaluating Student Evaluations of Teaching: a Review of Measurement and Equity Bias					
17	16 ##id0123#written com@Cat2Sele@ The abstract ChatCompleto id0123# What's a comment worth? How to better understand student evaluations of teaching					
18	17 ##id0136#student eval @Cat2Sele@ The abstract ChatCompleto id0136# Validity and fairness of utilising student evaluation of teaching (SET) as a primary per					
19	18 ##id0144#open-ended @Cat2Sele@ The study fo ChatCompleto id0144# Reducing discrimination in the field: Evidence from an awareness raising intervention					
20	19 ##id0180#Gender bias@Cat2Sele@ The article fo ChatCompleto id0180# Gender biases in student evaluations of teaching. This article uses data from a French					
21	20 ##id0214#open-ended @Cat2Sele@ The abstract ChatCompleto id0214# Gender Bias in Student Evaluations of Teaching: 'Punish[ing] Those Who Fail To Do T					
22	21 ##id0228#open-ended @Cat2Sele@ The study fo ChatCompleto id0228# He just teaches whatever he thinks is important: Analysis of comments in student ev					
23	22 ##id0244#open-ended @Cat2Sele@ The abstract ChatCompleto id0244# Exploring the impact of faculty reflection on weekly student evaluations of teaching.					
24	23 ##id0253#Student Eval @Cat2Sele@ The study fo ChatCompleto id0253# Validation of a Short Scale for Student Evaluation of Teaching Ratings in a Polytechnic.					

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^a, Juan Martinez-Tomas^b, Amable Juarez-Tarraga^a and Cristina Santandreu-Mascarell^a

^aROGLE - Departamento de Organización de Empresas - Universitat Politècnica de Valencia jamarin@omp.upv.es
^bDepartamento de Organización de Empresas - Universitat Politècnica de Valencia jiamar@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-distilberto-v1, all-mpnet-base-v2, paraphrase-multilingual-mpnet-base-v2, distiluse-base-multilingual-cased-v1, all-MiniLM-L12-v2, allennas-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



ROGLE
Reengineering Operations
GroupWork Logistics Excellence



doe
Departamento de Organización de Empresas



SINCE

40

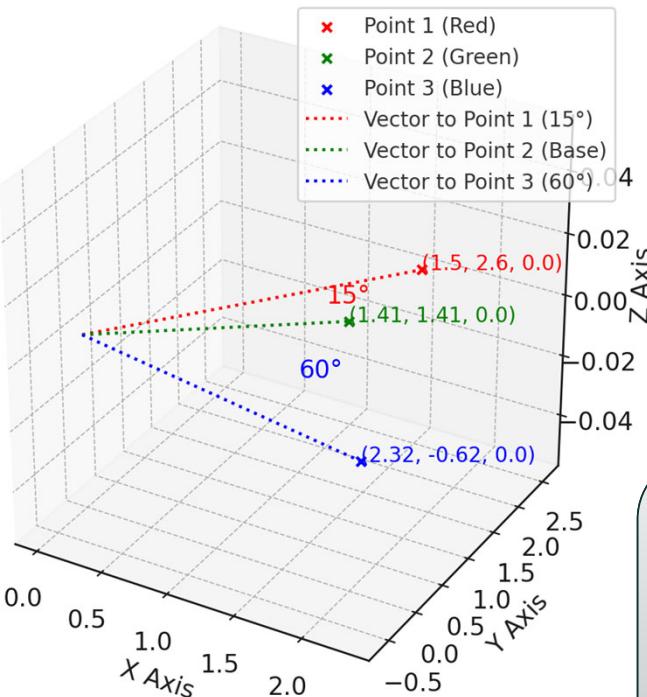


ACEDEdot
ASOCIACIÓN CIENTÍFICA
DE ECONOMÍA Y DIRECCIÓN
DE LA EMPRESA
DIRECCIÓN DE OPERACIONES Y TECNOLOGÍA

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

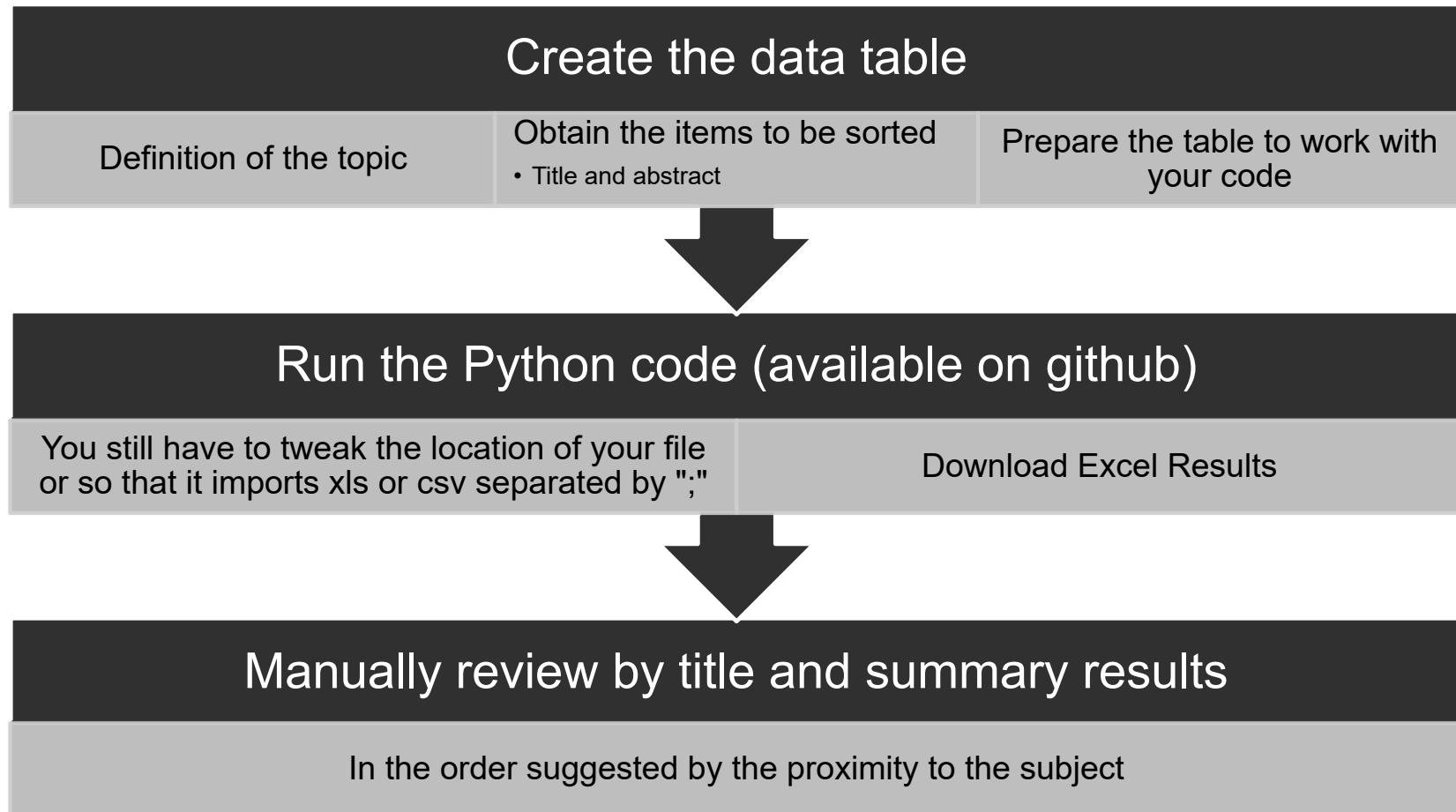
But before you come up ...
Operation: Madrid - Spain + France \approx Paris
Similarity with objective: 0.6902
Top 5 more similar words:
France: 0.7202
Madrid: 0.7023
Paris: 0.6902
Spain: 0.3597

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>



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”
- Retaining the code
- We see directly the first 20 results in Google Colab
- Then each person who plays with Excel to your liking

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 HiWP definition)

A	B	C
	Description	Average_Rank
1 ct_Compct		
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 studen	3
5 id0384--10	Redesigning student evaluations of teaching: integrating faculty and st	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 present	6
8 id0635--10	I liked your course because I believe in (the power of) student evaluat	7
9 id0500--10	Student Evaluation of Teaching (SET): Why the Emperor Has No Clothe	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 S	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 fo	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 l	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

Step 6: Prisma diagram. Rayyan makes it automatic in its payment version (it is not AI, it is simply database management)

<https://www.rayyan.ai/rayyan-prisma-2020-flowchart-ensuring-transparency-reproducibility>

Rayyan's PRISMA 2020 Flowchart – Ensuring Transparency, Reproducibility, and Quality

Posted by Robert Ayan

If you are at the stage in your systematic review or meta-analysis where you have completed article screening, you may be ready to generate a PRISMA flow diagram that captures your results. This diagram is going to be a summary of the number of articles identified, screened, assessed for eligibility, and included or excluded in the final analysis, along with the reasons for excluding articles.

Rayyan has just the feature for you! The PRISMA 2020 Flowchart!

Rayyan's PRISMA flow diagram feature has been set up to conform with the PRISMA Statement 2020*. The PRISMA statement is a set of guidelines that help ensure transparent and complete reporting of systematic reviews and meta-analyses.

Creating a PRISMA flow diagram on Rayyan is simple and efficient. Simply select the PRISMA button and you will be taken to a PRISMA flow diagram generator with several options and a control panel to make any necessary modifications.

This diagram is pre-populated with values available from your review, such as the number of references, the number of duplicated records, and those included for the final analysis. You can make changes or edits using the control panel and then download the completed PRISMA flow diagram in a variety of standard formats for inclusion in your manuscript.

Below you will find information on how to use the Main Options panel to customize your PRISMA flow diagram to completion.

- Reflection on evaluation of the "quality" of the identified works
 - Evidence quality vs. Report quality
 - It is a controversial issue
 - (I would not put the AI in this, for the moment)
- My suggestion
 - Do not use it to exclude articles but as a methodological moderator
 - Grouping sources by levels, evidence or report quality, do different results come out in each group?

Two prism report versions

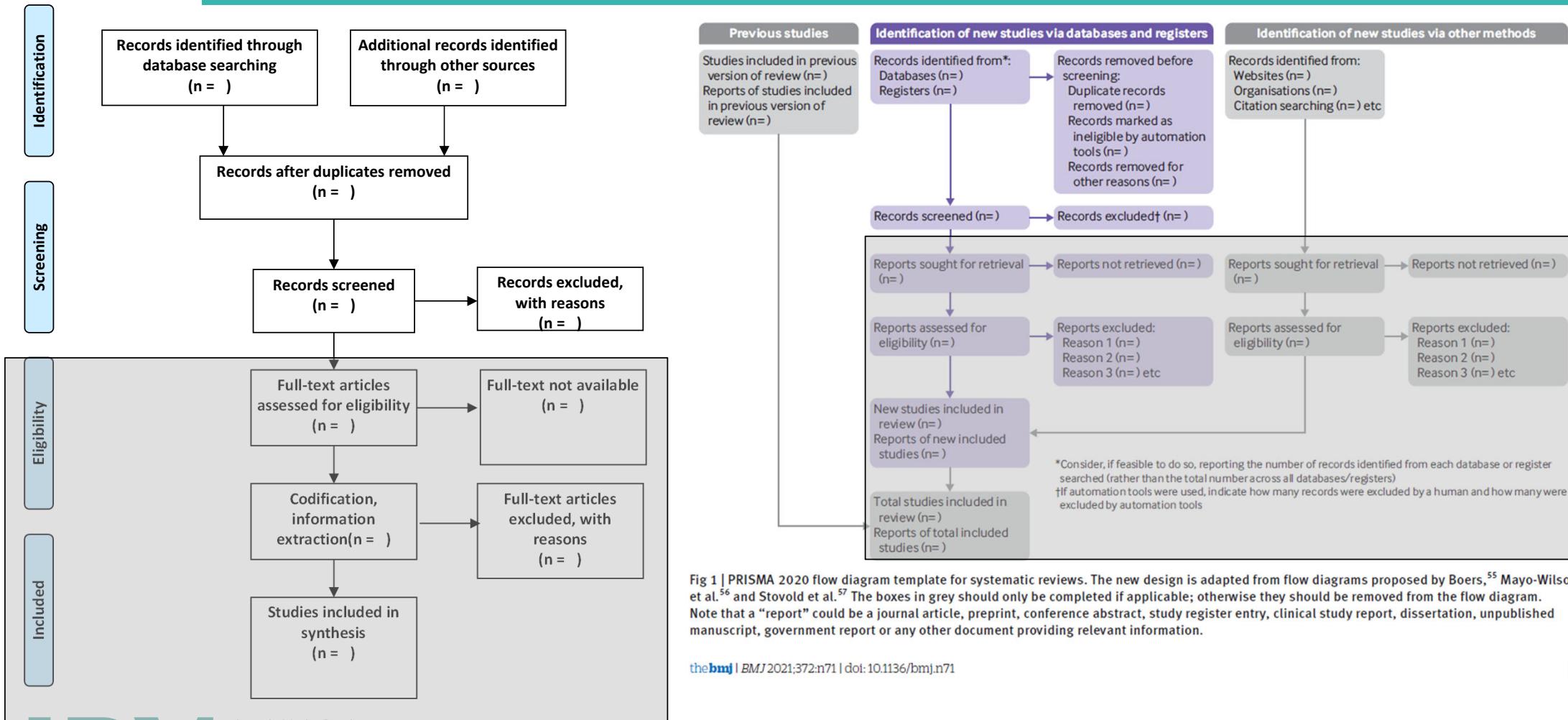


Fig 1 | PRISMA 2020 flow diagram template for systematic reviews. The new design is adapted from flow diagrams proposed by Boers,⁵⁵ Mayo-Wilson et al.,⁵⁶ and Stovold et al.⁵⁷ The boxes in grey should only be completed if applicable; otherwise they should be removed from the flow diagram. Note that a "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

Prisma model (preferred reporting items for systematic reviews and meta-analysis)

Table 1 | PRISMA 2020 item checklist

Section and topic	Item #	Checklist item	Location where item is reported
Title			
Title	1	Identify the report as a systematic review.	
Abstract			
Abstract	2	See the PRISMA 2020 for Abstracts checklist (table 2).	
Introduction			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	
Methods			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	
	13d	Describe any methods used to synthesise results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesised results.	

Juan A. Marin-Garcia

Universitat Politècnica de València-DOE-ROGLE-IEMA
www.upv.es

Prisma model (preferred reporting items for systematic reviews and meta-analysis)

Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.
Results		
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram (see fig 1).
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.
Study characteristics	17	Cite each included study and present its characteristics.
Risk of bias in studies	18	Present assessments of risk of bias for each included study.
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.
	20c	Present results of all investigations of possible causes of heterogeneity among study results.
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesised results.
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.
Discussion		
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.
	23b	Discuss any limitations of the evidence included in the review.
	23c	Discuss any limitations of the review processes used.
	23d	Discuss implications of the results for practice, policy, and future research.
Other Information		
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.
	24c	Describe and explain any amendments to information provided at registration or in the protocol.
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.
Competing interests	26	Declare any competing interests of review authors.
Availability of data, code, and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.

Prisma model (preferred reporting items for systematic reviews and meta-analysis)

Table 2 | PRISMA 2020 for Abstracts checklist*

Section and topic	Item #	Checklist item
Title		
Title	1	Identify the report as a systematic review.
Background		
Objectives	2	Provide an explicit statement of the main objective(s) or question(s) the review addresses.
Methods		
Eligibility criteria	3	Specify the inclusion and exclusion criteria for the review.
Information sources	4	Specify the information sources (e.g. databases, registers) used to identify studies and the date when each was last searched.
Risk of bias	5	Specify the methods used to assess risk of bias in the included studies.
Synthesis of results	6	Specify the methods used to present and synthesise results.
Results		
Included studies	7	Give the total number of included studies and participants and summarise relevant characteristics of studies.
Synthesis of results	8	Present results for main outcomes, preferably indicating the number of included studies and participants for each. If meta-analysis was done, report the summary estimate and confidence/credible interval. If comparing groups, indicate the direction of the effect (i.e. which group is favoured).
Discussion		
Limitations of evidence	9	Provide a brief summary of the limitations of the evidence included in the review (e.g. study risk of bias, inconsistency and imprecision).
Interpretation	10	Provide a general interpretation of the results and important implications.
Other		
Funding	11	Specify the primary source of funding for the review.
Registration	12	Provide the register name and registration number.

*This abstract checklist retains the same items as those included in the PRISMA for Abstracts statement published in 2013,⁵⁴ but has been revised to make the wording consistent with the PRISMA 2020 statement and includes a new item recommending authors specify the methods used to present and synthesise results (item #6).



Questions, doubts, comments

