ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/370155174

Applying ChatGPT and AI-powered tools to accelerate evidence reviews

Preprint · April 2023

DOI: 10.31219/osf.io/pcrqf

CITATIONS
0

READS
107

3 authors, including:

Kien Nguyen-Trung
Monash Sustainable Development Institute
38 PUBLICATIONS **82** CITATIONS
SEE PROFILE

Stefan Kaufman
Monash University (Australia)
12 PUBLICATIONS **123** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Understanding Rural Societies View project

Project Liên kết dân sự và phát triển nông thôn View project

All content following this page was uploaded by Kien Nguyen-Trung on 21 April 2023.

The user has requested enhancement of the downloaded file.

OSFPreprints DOI: https://doi.org/10.31219/osf.io/pcrqf License: CC-By Attribution 4.0 International

Applying ChatGPT and AI-powered tools to accelerate evidence reviews

Kien Nguyen-Trung*, Alexander K Saeri, Stefan Kaufman

BehaviourWorks Australia, Monash Sustainable Development Institute, Monash University

*Corresponding author. Dr. Kien Nguyen-Trung, Research Fellow, BehaviourWorks Australia, Monash Sustainable Development Institute, Monash University, Clayton Campus, VIC 3800, Australia. Kien.nguyen@monash.edu; ORCID: https://orcid.org/0000-0002-1782-7405

Dr. Alexander K Saeri, Research Fellow, BehaviourWorks Australia, Monash Sustainable Development Institute, Monash University, Clayton Campus, VIC 3800. alexander.saeri@monash.edu; ORCID: https://orcid.org/0000-0001-9254-0300

Dr. Stefan Kaufman, Senior Research Fellow, BehaviourWorks Australia, Monash Sustainable Development Institute, Monash University, Clayton Campus, VIC 3800. stefan.kaufman@monash.edu; ORCID: https://orcid.org/0000-0001-7973-3845

Please cite this work as:

Nguyen-Trung, K., Saeri, A. K., & Kaufman, S. (2023, April 20). Applying ChatGPT and AI-powered tools to accelerate evidence reviews. OSF Preprints, 20 April, 2023. https://doi.org/10.31219/osf.io/pcrqf

OSFPreprints DOI: https://doi.org/10.31219/osf.io/pcrqf License: CC-By Attribution 4.0 International

Ethical Statements and Declarations

Funding: The authors did not receive support from any organisation for the preparation of the manuscript or the research conducted herein.

Competing Interests: The authors have no relevant financial or non-financial interests to disclose.

Author Contributions: Author contributions are presented according to the International Council for Medical Journal Editors criteria for authorship.

Contribution to conception or design of the research: KNT, SK

Contribution to acquisition of the research: KNT, SK

Contributed to analysis or interpretation of data for the research: KNT, AKS.

Drafted the research or revisited the research critically for important intellectual content: KNT, AKS, SK.

Provided final approval of the research to be published: KNT, AKS, SK.

Agrees to be accountable for all aspects of the research in ensuring that questions related to the accuracy or integrity of any part of the research are appropriately investigated and resolved: KNT, AKS, SK.

OSFPreprints DOI: https://doi.org/10.31219/osf.io/pcrqf License: CC-By Attribution 4.0 International

Applying ChatGPT and AI-powered tools to accelerate evidence reviews

Abstract

Artificial Intelligence (AI) tools have been used to improve the productivity of evidence review and synthesis since at least 2016, with EPPI-Reviewer and Abstrackr being two prominent examples. However, since the release of ChatGPT by OpenAI in late 2022, a large language model with an intuitive chatbot interface, the use of AI-powered tools for research – especially those that deal with text-based data – has exploded. In this working paper, we describe how we used the AI-powered tools such as ChatGPT, ChatGPT for Sheets and Docs, Casper AI, and ChatPDF to assist several stages of an evidence review. Our goal is to demonstrate how AI-powered tools can boost research productivity, identify their current weaknesses, and provide recommendations for researchers looking to utilize them.

Keywords: Artificial intelligence, ChatGPT, evidence synthesis, large language model, research methods, literature review

OSFPreprints DOI: https://doi.org/10.31219/osf.io/pcrqf License: CC-By Attribution 4.0 International

Introduction

Early AI and machine learning tools, such as EPPI-Reviewer and Abstrackr, have assisted with the labor-intensive task of title and abstract screening in evidence review and synthesis (Thomas & Brunton, 2007; Tsou et al., 2020). However, the release of ChatGPT, an AI-based conversational large language model (LLM), as a public tool by OpenAI in November 2022 represented a major advancement in AI technologies (Sallam, 2023), particularly in the realm of text-based data analysis. ChatGPT, with the latest version being GPT-4, sets itself apart from earlier machine learning models by its ability to understand user requests and produce personalized, human-like responses, potentially impacting multiple industries (Lund & Wang, 2023). This large language model (LLM), coupled with its intuitive chatbot interface, has greatly expanded the potential applications of AI-powered tools in many areas such as climate change (Biswas, 2023a), public health (Biswas, 2023b), especially education (Firat, 2023; King & chatGPT, 2023). In the field of research and evidence synthesis, ChatGPT and AI-powered tools can assist in conducting literature review (Haman & Školník, 2023).

In this working paper, we explore the use of several AI-powered tools based on large language models (LLM), including ChatGPT (by OpenAI), ChatGPT for Sheets and Docs (by Talarian), Casper AI (by Casper AI), and ChatPDF (by ChatPDF), to assist in different stages of the evidence review process. The purpose of this working paper is twofold: to showcase the potential of AI-powered tools for evidence review and synthesis in research and academia, and to identify their current weaknesses and provide recommendations for researchers looking to utilize them.

Outline of a rapid systematic evidence review

In this working paper, based on the highly practical guidelines for evidence review outlined in Chapter 1 of BehaviourWorks Australia's 'The Method' Book (Bragge et al., 2020), we employed the eight basic steps of a rapid systematic evidence review as shown in Figure 1.

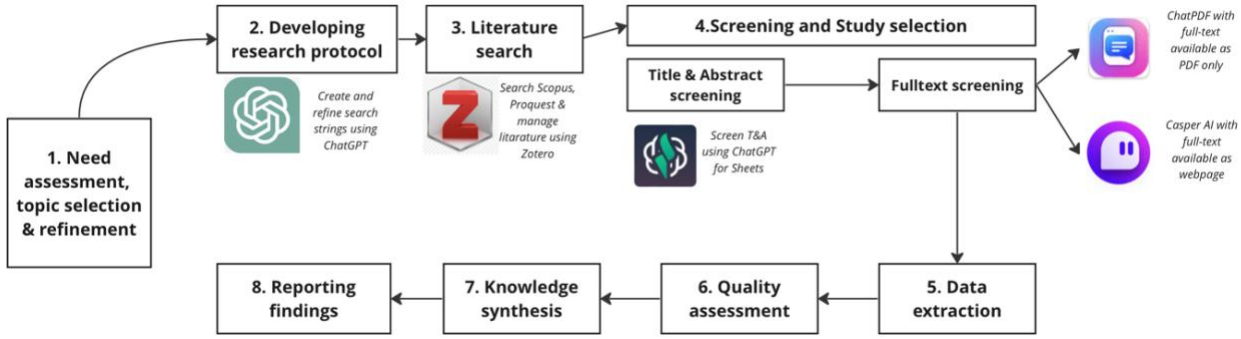

Figure 1. Evidence review process with the support of AI-powered tools

In this paper, our focus is on steps 2 and 4, as these are the steps where we have utilized AI-powered tools. Step 2 involves developing a research protocol that comprises the research question, database search string(s), inclusion & exclusion criteria, and method for searching, screening, extracting, and synthesizing findings. Our main focus in this step was using ChatGPT to support the development of search strings and adapt them to different databases. Moving to Step 4, we demonstrate how to use ChatGPT for Sheets to screen titles and abstracts, and ChatPDF, and Casper AI for full-text screening. For the remaining steps, we are of the view that large language models (LLM) and other AI-powered tools can significantly enhance the execution of each step in an evidence review process. We discuss this aspect in detail towards the end of the paper.

OSFPreprints DOI: https://doi.org/10.31219/osf.io/pcrqf License: CC-By Attribution 4.0 International

Rapid Systematic Evidence Review Procedure with AI-powered tools

Step 1: Needs assessment, topic selection, topic refinement

To experiment with the use of ChatGPT and other AI-powered tools in evidence review, we drew on a case study that aims to better understand the risks and opportunities associated with the participation of land managers in environmental markets in the Goulburn Broken catchment, Victoria, Australia (Kaufman & Nguyen-Trung, 2023). This study's research question and objectives are outlined as below:

Table 1: Research question and objectives of our case study

Primary research question	• Does land managers' participation in environmental markets build climate resilience?
Research objectives underpinning the research question	• List of environmental market participation outcomes that align with climate resilience outcomes (and those that potentially do not) • List of determinants (drivers and barriers) of participation in environmental markets building climate resilience • List of interventions / activities that appear to be relevant and effective in ensuring EM participation leads to climate resilience

Understanding this information determined this evidence review's need, topic selection and refinement, and the meaning and importance of keywords, which guided the development of our review protocol. A detailed description of the topic selection and context for the project is outside the scope of this brief working paper. We have not yet applied AI-powered tools to this stage of the evidence review.

Step 2: Developing a research protocol – using ChatGPT

In this section, we focus on the development of search strings. Other elements of the research protocol, such as inclusion and exclusion criteria and methods for searching, screening, extracting, and synthesising findings, were not developed through the use of AI.

OSFPreprints DOI: https://doi.org/10.31219/osf.io/pcrqf License: CC-By Attribution 4.0 International

We used the research objectives from Step 1 to develop a search string with key words, provide feedback on the responses from ChatGPT, and format it for the two databases we intended to search (ProQuest and Scopus). A detailed exchange with prompts and responses is provided in Appendix 1.

Summary of research question development:

Role	**Text**	**Commentary**
User	Please develop a search string that helps answer the following 3 objectives: List of environmental market participation outcomes that align with climate resilience outcomes (and those that potentially do not). [...]	Initial prompt to generate a search string
ChatGPT	Sure! Please find below search strings for each of the three research objectives: 1. List of environmental market participation outcomes that align with climate resilience outcomes (and those that potentially do not): ("environmental market*" OR "carbon market*" OR "carbon offset*" OR "carbon credit*" OR "cap and trade" OR "emissions trading") "climate resilience" OR [...]	
User	Could you please combine all three search strings into ONE only?	We sought to create a single string to address all objectives
ChatGPT	Sure, here is a combined search string for all three research objectives: ("environmental market*" OR "carbon market*" OR "carbon offset*" OR "carbon credit*" OR "cap and trade" OR "emissions trading") "climate resilience" OR [...]	
User	Did you forget to add "AND" to the search string?	The response did not include the correct AND operators between elements of the string

ChatGPT	[...] ("environmental market*" OR "carbon market*" OR "carbon offset*" OR "carbon credit*" OR "cap and trade" OR "emissions trading") AND ("climate resilience" OR [...]	
User	Very good. But in the first objective, there is a term "participation in environmental market". Where is it in the search string?	Although the string was now much better, we still found it missing the key words "participation".
ChatGPT	[...] ("participation in environmental market*" OR "carbon market*" OR "carbon offset*" OR "carbon credit*" OR "cap and trade" OR "emissions trading") AND ("climate [...]	
User	Adapt the following search string to be used in Proquest database: ("participation in environmental market*" OR "carbon market*" OR "carbon offset*" OR "carbon [...]	As the string was now good, we asked ChatGPT to adapt this string to Proquest. We also requested that ChatGPT add "NOFT" to every keyword. NOFT stands for "Anywhere except full text", namely, Proquest will only search for abstract, title and keywords.
ChatGPT	noft(((noft(determinants) OR noft(drivers) OR noft(barriers) OR noft(facilitators) OR noft(challenges) OR noft(obstacles) OR noft(opportunities) OR noft(factors) OR noft(enablers) OR noft(constraints) OR [...]	
User	Adapt this search string to be used in Scopus database: noft(((noft(determinants) OR noft(drivers) OR noft(barriers) OR noft(facilitators) OR noft(challenges) OR noft(obstacles) OR noft(opportunities) OR noft(factors) OR noft(enablers) OR [...]	We requested that the prompt be adapted for Scopus
ChatGPT	(TITLE-ABS-KEY (determinants OR drivers OR barriers OR facilitators OR challenges OR obstacles OR opportunities OR factors OR enablers OR constraints OR motivations OR	ChatGPT did a good job of adapting the string. Note that it recognised Scopus has the search

	`disincentives OR requirements OR conditions OR influences OR impacts)` `[...]`	function “Title-Abs-Key”, which is different from Proquest.

Below is a summary outlining the advantages and disadvantages of using ChatGPT at this particular stage.

Strengths:
- *ChatGPT facilitates the quick development of search strings from research objectives or questions.*
- *ChatGPT can assist in quickly adapting search strings from one database to another.*

Weaknesses:
- *ChatGPT does not understand the difference in significance of keywords in the search strings. Thus, researchers need to determine which keywords are crucial for their research objectives.*
- *ChatGPT occasionally makes errors, such as forgetting operators, which could compromise the search results. Researchers need to be able to verify and correct these mistakes.*

Step 3. Literature search of one or more databases and sources

Using these search strings, we conducted searches on the databases Proquest and Scopus. On Proquest, the string returned 56 records. We ended up removing one non-English study and having 55 studies for the next stage of review. On Scopus, the string returned 84 records. We limited the search to only include articles, reviews, book chapters, and conference papers. The final results were 81. Typically, initial searches in these databases result in a larger number of hits. However, for the purposes of this experiment, we accepted the small number of studies found.

We imported the records into Zotero, removed duplicates, and also added an extra three records based on expert consultation. After removing one non-English result, there were a total of 115 records for screening and study selection. We exported the Zotero database to spreadsheet format (as a CSV, comma-separated variable) so it could be manipulated in Google Sheets.

Step 4. Screening and study selection - using ChatGPT for Google Sheets

Title and abstract screening

Google Sheets allows for plugins to connect with other software. One recent plugin is ChatGPT for Sheets and Docs, which allows for the use of a new formula called *=gpt()*, which communicates with a web-based GPT system similar to ChatGPT. A spreadsheet needs to be activated using an API key from OpenAI before GPT tools can be used in that spreadsheet (Figure 2).

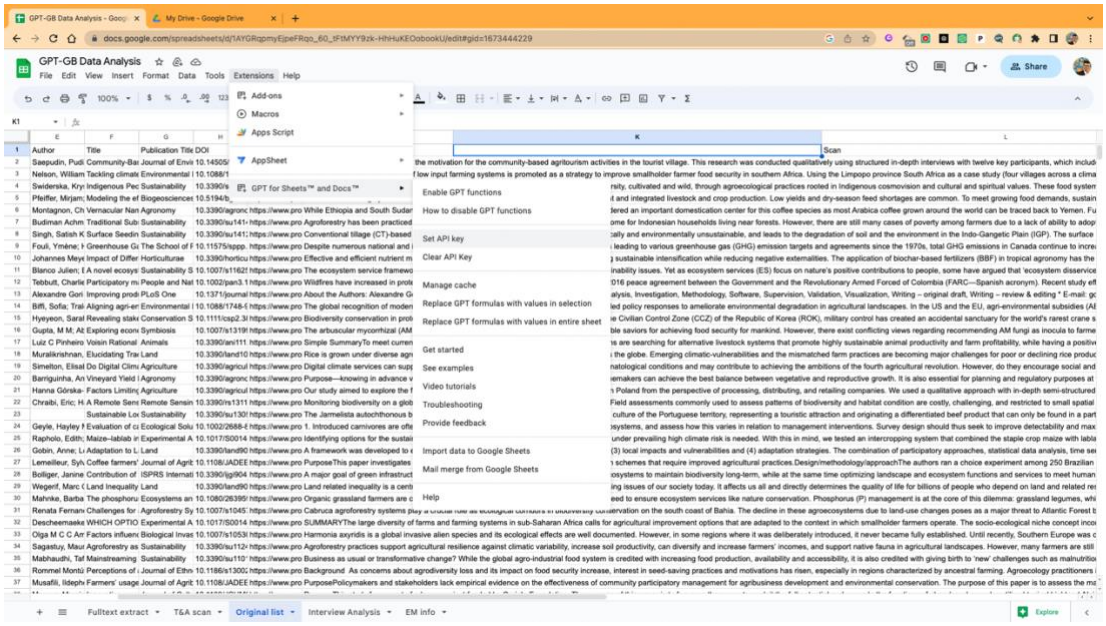

Figure 2. Enabling GPT functionality in a Google Sheet

We sought to use GPT to assist in determining the relevance of records to the research objectives of the project. We intended to exclude irrelevant records, and conduct full-text screening on the remaining records. Because the case study project focused on understanding the involvement of land managers and farmers in environmental markets and its impact on climate outcomes, relevant records were those that included:

1. Land managers and farmers

2. Certain types of environmental markets: Carbon market, ecosystem market, biodiversity market, and renewable energy market
3. Climate change or climate outcomes

To accelerate this screening process, we used GPT for Sheets to assess each abstract for each record and determine its relevance. It is worth noting that the records exported from Zotero already included a column containing study abstract.

In a column of the Google Sheet, we used the `=gpt()` formula (Figure 3):

```
=GPT("write yes if the abstract mentions farmer or land manager",
[Abstract cell reference])
```

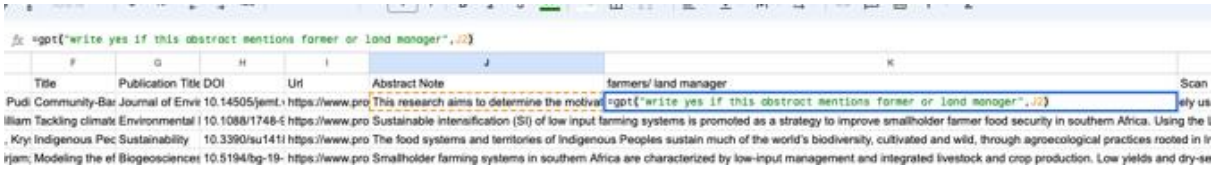

Figure 3. Using ChatGPT for Sheets to screen abstracts for relevance.

Researchers should conduct a thorough review of the results from the initial screening using ChatGPT for Sheets. It is important to cross-check the results with the abstract of the searched study and only proceed to the next step if the =gpt() formula returns accurate results.

As the result was good, we used this same formula for all 115 records to assess farmers / land managers (Figure 4):

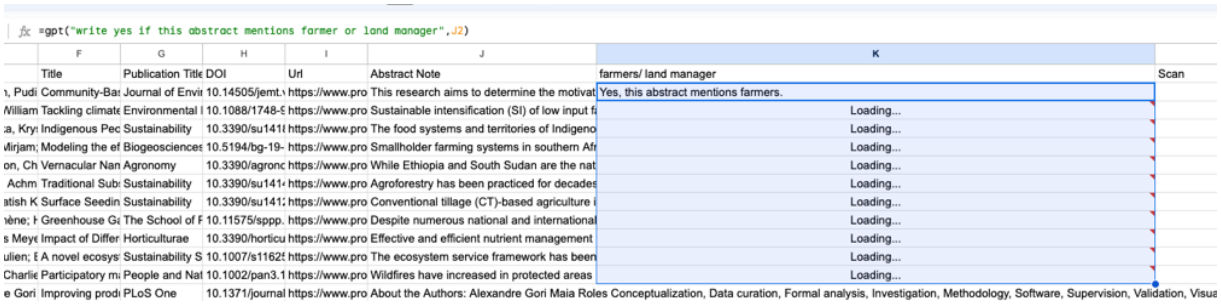

Figure 4. Copying the same ChatGPT for Sheets formula for each record

Using the same method, we searched for “carbon market”, “ecosystem market”, “biodiversity market”, “renewable energy market” and “climate change” in separate columns. It is worth noting that the researchers must be the ones who decide which keywords are important for the relevance of a study.

Once we had a result for each of these terms, we needed to combine these responses to decide on the relevance of the record. In a new column of the sheet, we used the =gpt() formula again:

```
=GPT(“write yes if any of these cells containing yes”, [farmer cell
reference : climate change cell reference])
```

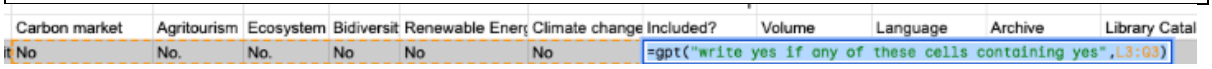

Figure 5. Using ChatGPT for Sheets formula to determine the relevance of the record

The results of the screening process were summarized in a new column titled "Included?". This column displayed the outcome of the keyword search for each document. Documents that did not contain the specific keywords we were looking for were marked as irrelevant and highlighted in grey. By sorting the entire dataset based on the "Included?" column, we could easily identify and remove 16 irrelevant studies from our analysis.

There are four cautions to keep in mind when developing a ChatGPT for Sheets formula: First, if the formula is too complicated, ChatGPT for Sheets may not perform accurately. For example, in our experiment, we wrote a formula:

```
=GPT("write yes if the abstract mentions either carbon market,
biodiversity market, natural capital market, agritourism market, water
quality market or ecosystem services", [Abstract cell reference])
```

Upon rechecking, we found that ChatGPT for Sheets sometimes failed to deliver the requested outcome. For example, it incorrectly returned a "no" for an abstract containing the

term "agritourism market." However, ChatGPT does possess some level of understanding of concepts, as demonstrated by our successful request:

```
=GPT("write yes if the abstract mentions any type of environmental
markets", [Abstract cell reference])
```

Although ChatGPT for Sheets returned relatively good results (e.g., identifying many types of environmental markets appearing in the abstracts), there were instances where it identified types of markets that were not environmental markets. For example, it classified 'agri-environmental subsidies' as an environmental market, which is inaccurate.

The third caution is that if we apply the same ChatGPT for Sheets formula to a large number of records at once, the likelihood of ChatGPT for Sheets making mistakes increases. As such, we recommend running it on a small number of records for better control.

The fourth caution is that after obtaining the results from your formulas, it is recommended to copy and paste them as values only. This will remove the ChatGPT for Sheets functions (i.e., =gpt()). Failure to do so will result in ChatGPT running the formulas again every time columns are inserted or moved around.

Below is a summary outlining the advantages and disadvantages of using ChatGPT for Sheets at this particular stage.

Strengths:

- *ChatGPT for Sheets can accelerate the screening of keywords in research.*
- *The most significant benefit is that ChatGPT for Sheets can quickly apply the screening results from the first instance to the remaining records.*
- *In some cases, ChatGPT for Sheets shows an ability to comprehend research concepts.*

Weaknesses:

- *ChatGPT for Sheets is currently better than keyword counting, but it has not yet reached the level of sufficiently understanding a research concept. Therefore, researchers need to experiment and check the formulas to find the best ones for their purposes.*
- *ChatGPT for Sheets can occasionally make mistakes, which requires researchers to cross-check the results.*
- *ChatGPT for Sheets automatically re-runs its formulas, which can result in the loss of obtained data or extra fees.*

OSFPreprints DOI: https://doi.org/10.31219/osf.io/pcrqf License: CC-By Attribution 4.0 International

Full-text screening

There are several options for full-text screening, depending on how the full-text can be accessed. At this stage, the usefulness of current AI-powered tools is somewhat limited due to their lack of a rich understanding of the research objectives, concepts, and context. However, it is still possible to take advantage of how these tools can assist in semantic search of full-text. Semantic search is a technique that seeks to understand the meaning and context behind search queries.

Instead of searching only for the keywords "carbon market" and "climate outcomes," a semantic search can identify sections where the text discusses the relationship between the two concepts. Using semantic search powered by large language models such as ChatGPT, researchers can more effectively identify relevant sections in the full-text documents and better understand the context in which the keywords and related concepts are discussed. This, in turn, can lead to a more comprehensive and accurate assessment of the available evidence during the screening process.

The tool to use will depend somewhat on the availability of full-text:

Full-text available online as web page	Use Casper AI to semantically search on web pages for inclusion criteria
Full-text available as PDF only	Use ChatPDF to semantically search PDFs for inclusion criteria

Full-text available online as web page: using Casper AI

Casper AI is a Chrome-based extension designed to help users summarize the content of web pages. It is particularly useful when documents are available as full-text web pages, which means Casper AI can access and process these texts directly.

We activated Casper AI and visited the website that contained the full-text of the record we wished to screen. Two useful ways to interact with Casper AI (see Figure 6) include:

- Using pre-built functions such as “Summarize this page”,
- Asking specific questions to understand the content of the full-text.

Below is an example of an exchange with Casper AI:

Role	**Text**	**Commentary**
User	`Does the paper mention any type of environmental market?`	Query to assess relevance of full-text
Casper	`The paper does not mention any specific type of environmental market, but it does emphasize the importance of environmental conservation as one of the dimensions of community-based agritourism. The study suggests that community-based agritourism can help to maintain local cultural and biological diversity through ecotourism management.`	

This information aided us in assessing the paper's relevance. Although Casper AI did not identify any specific type of environmental market, it was helpful to cross-check the paper. After a more detailed read, the research team decided that agritourism could be classified as a type of environmental market. This decision demonstrates that while AI-powered tools can serve as valuable assistants, the ultimate responsibility for making decisions regarding the inclusion and exclusion criteria rests with the researchers. Such decisions should reflect the research purpose, theoretical framework, and existing knowledge. Consequently, we advanced this paper to the next stage of data extraction.

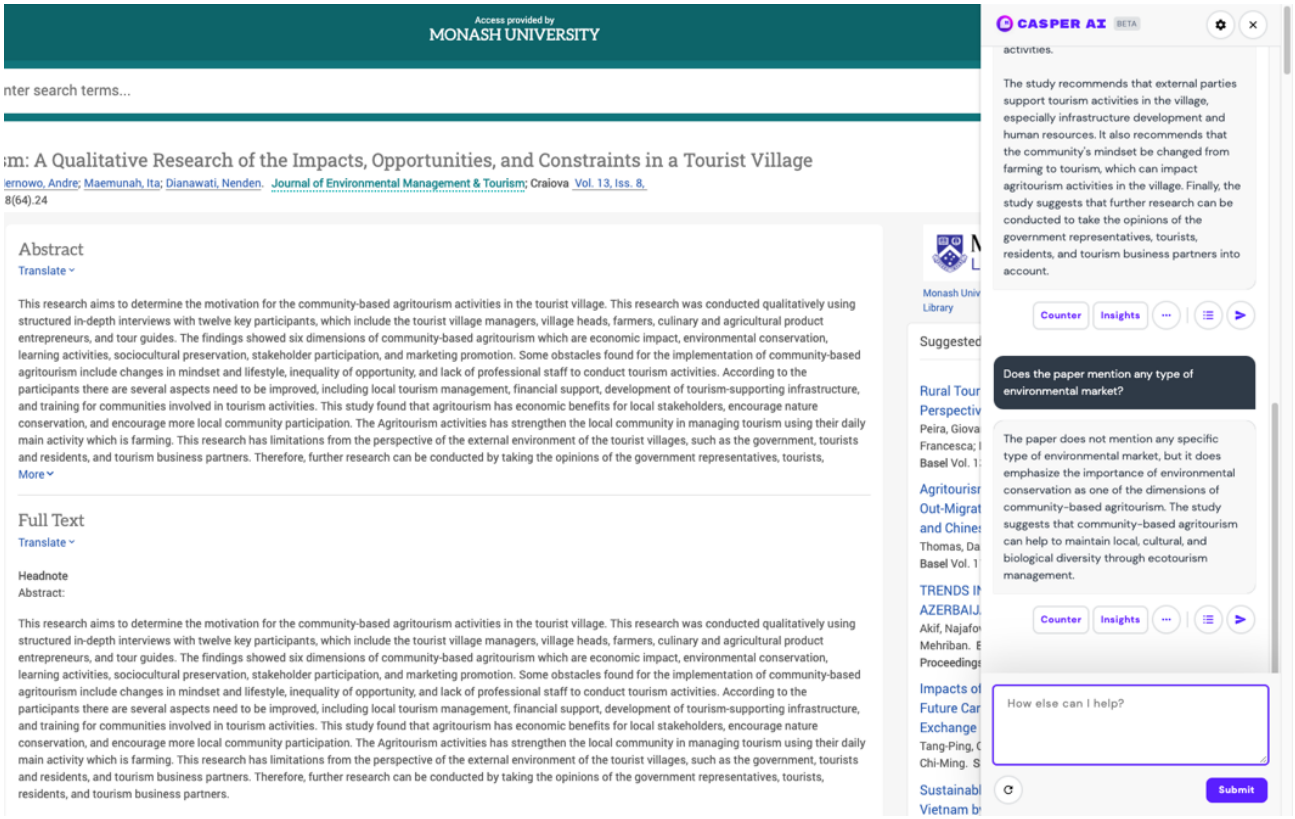

Figure 6. Using Casper AI to assess whether an online full-text record should be included in an evidence review

Full-text available as PDF only

If access to the full-text web page is not possible, an alternative is to obtain the full-text as a PDF and use one of two platforms to apply AI-powered tools to assess its relevance for an evidence review. ChatPDF allows the uploading of PDFs to a website, where you can then ask it questions about the content and conduct semantic search. Elicit is another AI tool that can search and summarise research papers. Elicit allows to ask specific questions of multiple PDFs but does not provide a chat function. We only discuss ChatPDF in this working paper for reasons of scope.

Using ChatPDF

OSFPreprints DOI: https://doi.org/10.31219/osf.io/pcrqf License: CC-By Attribution 4.0 International

Here we present two responses from using ChatPDF to assess a paper. In the first case, we did not assess a full-text paper as relevant from a cursory inspection but asked ChatPDF to determine whether we had missed something.

Role	**Text**	**Commentary**
User	Does the paper mention any type of environmental market?	Query to assess relevance of full-text
ChatPDF	I'm sorry, but I couldn't find any mention of an environmental market in the pages of the PDF that you provided. The paper mainly focuses on the development of a novel ecosystem service cascade model to navigate sustainability problems in a changing agricultural landscape in Brazil. It discusses the interactions between ecosystem services and ecosystem disservices and how they can be taken into account to support sustainability.	

To verify the results obtained from ChatPDF, we carefully reviewed the paper to ensure the accuracy of the AI-generated findings. It is crucial to cross-check the outcomes produced by AI-powered tools to maintain the quality of our research. Upon closer examination, we confirmed that the paper did not discuss the specific types of environmental markets as indicated by ChatPDF. Consequently, we decided to exclude the paper from our analysis.

In another case, ChatPDF helped identify that a paper did discuss carbon markets, which was helpful for us to assess the paper's relevance. We were able to follow-up on this determination with a request for specific quotations from the text so that we could read the most relevant sections of the paper.

Role	**Text**	**Commentary**
User	Does the paper mention any type of environmental market?	Query to assess relevance of full-text
ChatPDF	Yes, the paper mentions carbon markets in relation to agricultural producers in Alberta. A study conducted by the University of Alberta in 2012 found	

OSFPreprints DOI: https://doi.org/10.31219/osf.io/pcrqf License: CC-By Attribution 4.0 International

	that local markets and farm-specific factors affected participation in carbon markets more than global or provincial benefits (Page 9). However, the paper does not provide an extensive discussion of environmental markets beyond this example.	
User	So, please give some quotations on the impact of environmental market on GHG emissions	
ChatPDF	Certainly! Here are some quotations from the paper on the impact of environmental markets on GHG emissions: - "Mechanisms such as setting a carbon price, cap-and-trade systems and tax-based policies contribute to decreased GHG emissions." (Page 4) - "Agriculture has the potential to mitigate GHG emissions by applying BMPs that reduce emissions and increase carbon storage in soils." (Page 4) - "Participation in environmental markets can provide financial incentives for land managers, farmers, and landholders to adopt agricultural land management practices that reduce GHG emissions and increase carbon storage in soils." (Not a direct quotation from the paper, but a summary of its content based on information from Pages 4 and 13)	

Below is a summary outlining the advantages and disadvantages of using Casper AI/ChatPDF at this particular stage.

Strengths:
- *Casper AI/ChatPDF is helpful in summarizing the key findings of an article.*
- *Casper AI shows a better capacity to understand research concepts by identifying specific findings related to specific concepts or relationships between concepts.*
- *ChatPDF is more useful in identifying the pages that contain specific information.*

Weaknesses:
- *Casper AI/ChatPDF can sometimes misidentify information from an article, so it is better to narrow down to specific sections of the article to enhance its scanning capacity.*
- *Like ChatGPT for Sheets, both tools have not yet reached the level of understanding specific academic concepts. As a result, researchers need to spend time interacting with these tools through clear prompts to enhance their search capabilities.*

OSFPreprints DOI: https://doi.org/10.31219/osf.io/pcrqf License: CC-By Attribution 4.0 International

Other steps in the evidence review procedure

Step 5. Data extraction from included studies and Step 6. Quality assessment

Although we have not yet used AI-powered tools to assist in data extraction and quality assessment yet, we think that large language models such as ChatGPT can help researchers extract relevant data from included studies more quickly and accurately. Semantic search and identification of relevant information could save time and reduce the risk of human error during the data extraction process. These tools could apply quality assessment checklists or guidelines, such as the Cochrane Risk of Bias tool, to determine a score for each included paper. This could help researchers more accurately assess the credibility of the evidence and make informed decisions about the inclusion or exclusion of studies.

However, the tendency of these tools to confabulate data or become confused about what information is in the document (accurate, relevant) and what information is in its training dataset (inaccurate, misleading) is a cause for concern. Some recent work has suggested that careful questioning and prompting can improve the accuracy of responses from these models on documents.

Step 7. Knowledge synthesis

One potential area to apply large language models is in helping researchers identify patterns, trends, and relationships across the included papers. AI-powered tools can analyze large sets of information and generate summaries with guidance about theme, specific research questions, or other research objectives. While it is time consuming to re-analyze text data to assess the effectiveness of an intervention on another new outcome, it is likely that AI-powered tools will make replication and extension of existing review methods easier. Effect size extraction for meta-analysis or other aggregation such as vote counting could also be assisted.

OSFPreprints DOI: https://doi.org/10.31219/osf.io/pcrqf License: CC-By Attribution 4.0 International

Step 8. Reporting and disseminating findings

Beyond clearly written summaries of findings for inclusion in reports, papers, briefings, and other written materials, large language models could be applied to assist readers / stakeholders, who have a language barrier or no scientific background or tools, in interpreting the findings of research. For example, creating a PDF of the evidence review and uploading it to ChatPDF can help readers explore the findings of the work and apply it to their own context much more easily than the research team can. ChatGPT can help researchers adapt their writings into more user-friendly products with fewer jargons, less formal language, and more local language phrases. This is also a highly scalable approach to research dissemination, because it allows research users to engage with the findings on their own terms, empowering them to ask questions, seek clarifications, and apply the findings to their context. By streamlining and enriching the reporting and dissemination process, researchers can more effectively share their insights and contribute to the broader scientific community.

Conclusion and Implications

In conclusion, the integration of AI-powered tools, such as ChatGPT, ChatGPT for Sheets and Docs, and ChatPDF, has proven to be a valuable addition to the evidence review and synthesis process. Our working paper adds to the current exploration of utilizing machine learning specifically in screening titles and abstracts and more broadly in evidence review (Kebede et al., 2023). We argue that these tools offer various benefits for researchers, such as quickly generating search strings and accelerating keyword screening. These benefits include reducing the time, effort, and required number of researchers to conduct evidence reviews.

However, they each have weaknesses, such as inconsistent results and occasional errors, and the need for back-and-forth interaction and clear prompts to enhance search capabilities. Furthermore, although these AI-powered tools demonstrate some comprehension of research

OSFPreprints DOI: https://doi.org/10.31219/osf.io/pcrqf License: CC-By Attribution 4.0 International

concepts, they may misinterpret material or generate misleading descriptions or summaries of those concepts. This aligns with the efforts of other researchers who have highlighted the problems of inaccuracies, confabulation, plagiarism, and confusion that could negatively affect research quality, transparency, and human autonomy (Haman & Školník, 2023; van Dis et al., 2023). Hence, these tools should not be treated as independent research assistants capable of thinking and acting independently, and providing feedback to researchers in accordance with research goals and shared understanding. We advise researchers to exercise caution when using AI-powered technologies and to design a thorough research strategy to ensure effective monitoring and quality control. To improve their search capabilities and outcomes, each research team should invest time in connecting with them, testing appropriate prompts, and routinely fact-checking. Cross-checking among human researchers is critical to preventing individual researchers from being 'misled by false or biases information' produced by ChatGPT or AI-powered tools (van Dis et al., 2023).

In their examination of computer technologies/ intelligence over 20 years ago, Salomon and colleagues already developed the idea of "partnership with an intelligent tool", pointing out that this partnership transformed our perception of 'ability' from just about the capacity of a person's mind to the performance of a mind-machine collaboration (Salomon et al., 1991). Today, despite their limitations in grasping concept meaning and research tasks, ChatGPT and other AI-powered tools have a significant potential to transform the entire research process in general and evidence review and synthesis in particular. Large language models hold promise for enhancing data extraction, quality assessment, and knowledge synthesis in evidence reviews. They can also improve research communication and dissemination through quick summaries and interactive platforms, removing language barriers and empowering readers to engage with findings, as well as foster efficient knowledge sharing within the scientific and non-scientific communities. By embracing these technologies, researchers can

OSFPreprints DOI: https://doi.org/10.31219/osf.io/pcrqf License: CC-By Attribution 4.0 International

enhance their productivity, contribute to more robust and comprehensive evidence synthesis, and ultimately drive positive social change and impact on the world's most pressing problems at speed and scale.

In short, the COVID-19 pandemic has created an unprecedented demand for synthesised knowledge for policy-making (Bragge, 2022), the deployment of ChatGPT and other AI-powered tools could be a 'magic bullet' to assist address these demands. As the field of AI continues to advance, and especially given the recent news on AutoGPT, an open-source application that automates multi-step projects using OpenAI's GPT-4 language model (Mashable SEA, 2023), we anticipate that mind-AI partnership will become increasingly prevalent in research and academia. To keep up with this inconceivable pace, research institutions and academics around the world will have no choice but to fast incorporate AI-powered technologies into establishing their own organisational capabilities. The challenge is, how can we assure that the growth of this cooperation does not exacerbate the disparities between researchers in developed and developing countries? (van Dis et al., 2023).

References

Biswas, S. S. (2023a). Potential Use of Chat GPT in Global Warming. *Annals of Biomedical Engineering*. https://doi.org/10.1007/s10439-023-03171-8

Biswas, S. S. (2023b). Role of Chat GPT in Public Health. *Annals of Biomedical Engineering*. https://doi.org/10.1007/s10439-023-03172-7

Bragge, P. (2022). From Centuries to Hours: The Journey of Research into Practice. *Digital Government: Research and Practice*, *3*(2), 1–13. https://doi.org/10.1145/3529166

OSFPreprints DOI: https://doi.org/10.31219/osf.io/pcrqf License: CC-By Attribution 4.0 International

Bragge, P., Lennox, A., Curtis, J., Boulet, M., Slatter, P., Meis-Harris, J., & Wild, A. (2020). Chapter 1: Evidence Reviews. In *The Method Book* (pp. 1–17). BehaviourWorks Australia.

Firat, M. (2023). *How Chat GPT Can Transform Autodidactic Experiences and Open Education?* [Preprint]. Open Science Framework. https://doi.org/10.31219/osf.io/9ge8m

Haman, M., & Školník, M. (2023). Using ChatGPT to conduct a literature review. *Accountability in Research*, 1–3. https://doi.org/10.1080/08989621.2023.2185514

Kaufman, S., & Nguyen-Trung, K. (2023). *Environmental Markets and Climate Resilience for the Goulburn Broken region*. BehaviourWorks Australia, Monash Sustainable Development Institute, Monash University.

Kebede, M. M., Le Cornet, C., & Fortner, R. T. (2023). In-depth evaluation of machine learning methods for semi-automating article screening in a systematic review of mechanistic literature. *Research Synthesis Methods*, *14*(2), 156–172. https://doi.org/10.1002/jrsm.1589

King, M. R. & chatGPT. (2023). A Conversation on Artificial Intelligence, Chatbots, and Plagiarism in Higher Education. *Cellular and Molecular Bioengineering*, *16*(1), 1–2. https://doi.org/10.1007/s12195-022-00754-8

Lund, B. D., & Wang, T. (2023). Chatting about ChatGPT: How may AI and GPT impact academia and libraries? *Library Hi Tech News*. https://doi.org/10.1108/LHTN-01-2023-0009

Mashable SEA. (2023, April 14). What is Auto-GPT and why are hustle bros hype for it? Faster and easier ways for hustle bros to promote their Substack. *Mashable*. https://mashable.com/article/autogpt-chatgpt-hustle-bros-explainer

OSFPreprints DOI: https://doi.org/10.31219/osf.io/pcrqf License: CC-By Attribution 4.0 International

Sallam, M. (2023). *The Utility of ChatGPT as an Example of Large Language Models in Healthcare Education, Research and Practice: Systematic Review on the Future Perspectives and Potential Limitations*. https://doi.org/10.1101/2023.02.19.23286155

Salomon, G., Perkins, D. N., & Globerson, T. (1991). Partners in Cognition: Extending Human Intelligence with Intelligent Technologies. *Educational Researcher*, *20*(3), 2–9.

Thomas, J., & Brunton, J. (2007). *EPPI-Reviewer: Software for research synthesis*.

Tsou, A. Y., Treadwell, J. R., Erinoff, E., & Schoelles, K. (2020). Machine learning for screening prioritization in systematic reviews: Comparative performance of Abstrackr and EPPI-Reviewer. *Systematic Reviews*, *9*, 1–14.

van Dis, E. A. M., Bollen, J., Zuidema, W., van Rooij, R., & Bockting, C. L. (2023). ChatGPT: Five priorities for research. *Nature*, *614*(7947), 224–226. https://doi.org/10.1038/d41586-023-00288-7

OSFPreprints DOI: https://doi.org/10.31219/osf.io/pcrqf License: CC-By Attribution 4.0 International

Appendix 1. Full search string development

Role	Text	Commentary
User	Please develop a search string that helps answer the following 3 objectives: List of environmental market participation outcomes that align with climate resilience outcomes (and those that potentially do not). List of determinants (drivers and barriers) of participation in environmental markets building climate resilience List of interventions / activities that appear to be relevant and effective in ensuring EM participation leads to climate resilience	
ChatGPT	Sure! Please find below search strings for each of the three research objectives: 1. List of environmental market participation outcomes that align with climate resilience outcomes (and those that potentially do not): ("environmental market*" OR "carbon market*" OR "carbon offset*" OR "carbon credit*" OR "cap and trade" OR "emissions trading") "climate resilience" OR "resilience outcomes" OR "resilience benefits" OR "resilience co-benefits" OR "resilience advantages" OR "resilience drawbacks" OR "resilience limitations" OR "resilience trade-offs" OR "incompatible with resilience" OR "not aligned with resilience" 2. List of determinants (drivers and barriers) of participation in environmental markets building climate resilience: ("environmental market*" OR "carbon market*" OR "carbon offset*" OR "carbon credit*" OR "cap and trade" OR "emissions trading") "climate resilience" OR "resilience building" determinants OR drivers OR barriers OR challenges OR obstacles OR opportunities OR factors OR enablers OR constraints OR motivations OR disincentives OR requirements OR conditions OR influences OR impacts 3. List of interventions/activities that appear to be relevant and effective in ensuring EM participation leads to climate resilience:	

	("environmental market*" OR "carbon market*" OR "carbon offset*" OR "carbon credit*" OR "cap and trade" OR "emissions trading") "climate resilience" OR "resilience building" interventions OR activities OR strategies OR practices OR approaches OR measures OR initiatives OR policies OR programs OR schemes OR incentives OR support OR frameworks OR guidelines OR best practices OR success factors OR case studies	
User	Could you please combine all three search strings into ONE only?	It can be seen that ChatGPT did a good job of outlining keywords that align with the objectives. Since we only had limited resources (time, staff), we asked ChatGPT to combine these strings into one.
ChatGPT	Sure, here is a combined search string for all three research objectives: ("environmental market*" OR "carbon market*" OR "carbon offset*" OR "carbon credit*" OR "cap and trade" OR "emissions trading") "climate resilience" OR "resilience outcomes" OR "resilience benefits" OR "resilience co-benefits" OR "resilience advantages" OR "resilience drawbacks" OR "resilience limitations" OR "resilience trade-offs" OR "incompatible with resilience" OR "not aligned with resilience" determinants OR drivers OR barriers OR challenges OR obstacles OR opportunities OR factors OR enablers OR constraints OR motivations OR disincentives OR requirements OR conditions OR influences OR impacts interventions OR activities OR strategies OR practices OR approaches OR measures OR initiatives OR policies OR programs OR schemes OR incentives OR support OR frameworks OR guidelines OR best practices OR success factors OR case studies	
User	Did you forget to add "AND" to the search string?	It can be seen that ChatGPT forgot to use “AND” operator in its search string. As such, we asked it to add “AND” to ensure the equators are right.

OSFPreprints DOI: https://doi.org/10.31219/osf.io/pcrqf License: CC-By Attribution 4.0 International

ChatGPT	You're right, I apologize for the mistake. Here is the revised search string with the "AND" operator included: ("environmental market*" OR "carbon market*" OR "carbon offset*" OR "carbon credit*" OR "cap and trade" OR "emissions trading") AND ("climate resilience" OR "resilience outcomes" OR "resilience benefits" OR "resilience co-benefits" OR "resilience advantages" OR "resilience drawbacks" OR "resilience limitations" OR "resilience trade-offs" OR "incompatible with resilience" OR "not aligned with resilience") AND (determinants OR drivers OR barriers OR challenges OR obstacles OR opportunities OR factors OR enablers OR constraints OR motivations OR disincentives OR requirements OR conditions OR influences OR impacts) AND (interventions OR activities OR strategies OR practices OR approaches OR measures OR initiatives OR policies OR programs OR schemes OR incentives OR support OR frameworks OR guidelines OR best practices OR success factors OR case studies)	
User	Very good. But in the first objective, there is a term "participation in environmental market". Where is it in the search string?	Although the string was now much better, we still found it missing the key words "participation". As such, we asked:
ChatGPT	I apologize for the oversight. Here is the revised search string to include the missing term "participation in environmental market": ("participation in environmental market*" OR "carbon market*" OR "carbon offset*" OR "carbon credit*" OR "cap and trade" OR "emissions trading") AND ("climate resilience" OR "resilience outcomes" OR "resilience benefits" OR "resilience co-benefits" OR "resilience advantages" OR "resilience drawbacks" OR "resilience limitations" OR "resilience trade-offs" OR "incompatible with resilience" OR "not aligned with resilience") AND (determinants OR drivers OR barriers OR challenges OR obstacles OR opportunities OR factors OR enablers OR constraints OR motivations OR disincentives OR requirements OR conditions OR influences OR impacts) AND (interventions OR activities OR strategies OR practices OR approaches OR measures OR initiatives OR policies OR programs OR schemes OR incentives OR support OR frameworks OR guidelines OR best practices OR success factors OR case studies)	

OSFPreprints DOI: https://doi.org/10.31219/osf.io/pcrqf License: CC-By Attribution 4.0 International

	In this revised search string, the term "participation in environmental market" has been added at the beginning, followed by the other search terms, each separated by the "OR" operator, and then followed by the "AND" operator to ensure all terms are included in the search results.	
User	Adapt the following search string to be used in Proquest database: ("participation in environmental market*" OR "carbon market*" OR "carbon offset*" OR "carbon credit*" OR "cap and trade" OR "emissions trading" OR "biodiversity" OR "natural capital" or "environmental services") AND ("climate resilience" OR "resilience outcomes" OR "resilience benefits" OR "resilience co-benefits" OR "resilience advantages" OR "resilience drawbacks" OR "resilience limitations" OR "resilience trade-offs" OR "incompatible with resilience" OR "not aligned with resilience") AND (determinants OR drivers OR barriers OR facilitators OR challenges OR obstacles OR opportunities OR factors OR enablers OR constraints OR motivations OR disincentives OR requirements OR conditions OR influences OR impacts) AND (interventions OR activities OR strategies OR practices OR approaches OR measures OR initiatives OR policies OR programs OR schemes OR incentives OR support OR frameworks OR guidelines OR best practices OR success factors OR case studies) AND ("farmer*" OR "land manager*" OR "rancher*" OR "grower*" OR "agriculturist*" OR "agricultural producer*") AND ABSTRACT AND TITLE	As the string was now good, we asked ChatGPT to adapt this string to two target databases, Scopus and Proquest. We also requested that ChatGPT add "NOFT" to every keyword. NOFT stands for "Anywhere except full text", namely, Proquest will only search for abstract, title and keywords.
ChatGPT	noft(((noft(determinants) OR noft(drivers) OR noft(barriers) OR noft(facilitators) OR noft(challenges) OR noft(obstacles) OR noft(opportunities) OR noft(factors) OR noft(enablers) OR noft(constraints) OR noft(motivations) OR noft(disincentives) OR noft(requirements) OR noft(conditions) OR noft(influences) OR noft(impacts)) AND (noft("participation in environmental market*") OR noft("carbon market*") OR noft("carbon offset*") OR noft("carbon credit*") OR noft("cap and trade") OR noft("emissions trading") OR noft("biodiversity") OR noft("natural capital")) AND (noft("climate resilience") OR noft("resilience outcomes") OR noft("limitations")) AND (noft("farmer*") OR noft("land manager*") OR noft("rancher*") OR noft("grower*") OR noft("agriculturist*")) AND (noft(interventions OR activities OR strategies OR practices OR approaches OR measures OR initiatives OR policies OR programs OR schemes OR incentives OR	

	support OR frameworks OR guidelines OR best practices OR success factors OR case studies))))	
User	Adapt this search string to be used in Scopus database: noft(((noft(determinants) OR noft(drivers) OR noft(barriers) OR noft(facilitators) OR noft(challenges) OR noft(obstacles) OR noft(opportunities) OR noft(factors) OR noft(enablers) OR noft(constraints) OR noft(motivations) OR noft(disincentives) OR noft(requirements) OR noft(conditions) OR noft(influences) OR noft(impacts)) AND (noft("participation in environmental market*") OR noft("carbon market*") OR noft("carbon offset*") OR noft("carbon credit*") OR noft("cap and trade") OR noft("emissions trading") OR noft("biodiversity") OR noft("natural capital")) AND (noft("climate resilience") OR noft("resilience outcomes") OR noft("limitations")) AND (noft("farmer*") OR noft("land manager*") OR noft("rancher*") OR noft("grower*") OR noft("agriculturist*")) AND (noft(interventions OR activities OR strategies OR practices OR approaches OR measures OR initiatives OR policies OR programs OR schemes OR incentives OR support OR frameworks OR guidelines OR best practices OR success factors OR case studies))))	
ChatGPT	Here's the adapted search string for Scopus database: (TITLE-ABS-KEY (determinants OR drivers OR barriers OR facilitators OR challenges OR obstacles OR opportunities OR factors OR enablers OR constraints OR motivations OR disincentives OR requirements OR conditions OR influences OR impacts) AND TITLE-ABS-KEY ("participation in environmental market*" OR "carbon market*" OR "carbon offset*" OR "carbon credit*" OR "cap and trade" OR "emissions trading" OR "biodiversity" OR "natural capital") AND TITLE-ABS-KEY ("climate resilience" OR "resilience outcomes" OR "limitations") AND TITLE-ABS-KEY (farmer* OR "land manager*" OR rancher* OR grower* OR agriculturist*) AND TITLE-ABS-KEY (interventions OR activities OR strategies OR practices OR approaches OR measures OR initiatives OR policies OR programs OR schemes OR incentives OR support OR frameworks OR guidelines OR "best practices" OR "success factors" OR "case studies"))	ChatGPT did a good job of adapting the string. Note that it recognised Scopus has the search function “Title-Abs-Key”, which is different from Proquest.

OSFPreprints DOI: https://doi.org/10.31219/osf.io/pcrqf License: CC-By Attribution 4.0 International

Appendix 2. Some notes on pricing and how to manage costs

To be able to use ChatGPT with Google Sheets and Docs, you need to register API Key with OpenAI. This registration is not free. You need to add your credit card for billing. But as you will see, you only pay as you use. As you registered, you will be granted API key. You will need to store this key somewhere on your computer or personal notes. You need to use this key to activate ChatGPT for Google Sheet and Docs. So in creation of a new spreadsheet, you set API key, and then you can use ChatGPT as a new function.

One note is that, once you have the results run by chatGPT in specific cells/columns, you should copy the results to the same or other columns and choose *paste with value only*. Then you should delete the columns contains GPT functions. The reason is that if you do not do this step, every time you sort your data or move columns around, ChatGPT installed in Google Sheet will automatically run the functions again. This will cost you money.

Although we need to pay fee for ChatGPT for Google Sheet and Docs, the fee is quite cheap. With all the work we have done, it only costed us about US$2. You can also set up the hard and soft limit to prevent the running of ChatGPT costs more than what you expect.

View publication stats