

ChatGPT for Bibliometrics: Potential applications and limitations

[Daniel Torres-Salinas](#)^a, [Mike Thelwall](#)^b, and [Wenceslao Arroyo-Machado](#)^c

^aDepartment of Information and Communication, [University of Granada](#), Granada, Andalucía, Spain

^bInformation School, [University of Sheffield](#), Sheffield, England, United Kingdom

^cSchool of Public Affairs, [Arizona State University](#), Tucson, Arizona, United States

Abstract

This paper explores the transformative role of ChatGPT in enhancing bibliometric research methodologies across various stages of academic study. It discusses the application of ChatGPT in bibliometric studies across five core research stages: preparation and consultation, data processing, data analysis, results interpretation, and scientific writing. Highlighting ChatGPT's versatility, the paper showcases its utility in streamlining data handling, enhancing analytical depth, and facilitating scholarly communication. With capabilities ranging from querying external APIs to customising responses for specific research needs, ChatGPT may aid the efficiency and efficacy of bibliometric research. Ethical considerations are also discussed, advocating for the integration of ChatGPT to uphold high ethical standards and improve research integrity.

Keywords

ChatGPT; bibliometrics; academic research; data analysis; AI ethics

To be published in: *Library Catalogues as Data: Research, Practice, and Usage* (Facet Publishing)

Co-edited by: Prof. Melissa Terras and Dr. Sarah Ames

Submitted on: 13 June 2024

DOI: [10.5281/zenodo.11103550](https://doi.org/10.5281/zenodo.11103550)



49 **1. INTRODUCTION**

50 1.1. Large Language Models: transforming artificial intelligence

51 Since the release of BERT (Bidirectional Encoder Representations from Transformers) by
52 Google in 2018 and GPT-3 by OpenAI in 2020, large language models (LLMs) have
53 revolutionised the field of artificial intelligence (AI) and natural language processing. BERT
54 pioneered the use of bidirectional transformers, which understand words by considering
55 context from both sides, enabling a better grasp of natural language (Devlin et al. 2019). GPT-
56 3 advanced this technology further with 175 billion parameters, providing unprecedented
57 capabilities in text generation and comprehension (Frantar et al. 2023). These models are
58 trained on vast amounts of textual data using deep learning techniques that capture complex
59 patterns and structures of human language. In analysing text sequences, LLMs use the
60 transformer architecture for deep learning networks, allowing models to attend to different
61 parts of the text simultaneously, thus enhancing the coherence and relevance of the generated
62 responses. This capability enhances generative AI by helping it to create contextually
63 appropriate content.

64
65 The transformer architecture was introduced in the seminal paper "Attention is All You Need"
66 (Vaswani et al. 2017). It demonstrated how self-attention mechanisms could outperform
67 traditional recurrent neural networks in natural language processing tasks. This significantly
68 improved the accuracy and efficiency of language models and laid the groundwork for
69 subsequent developments. The power of this architecture has been popularised by the success
70 of a new generation of conversational bots. In particular, the Generative Pre-trained
71 Transformer (GPT) chat implementation called ChatGPT was created for conversational AI
72 challenges (Roumeliotis and Tselikas 2023). It can engage in interactive conversations, answer
73 questions, and simulate natural interactions (Wu et al. 2023).

74
75 The use of LLMs has swiftly expanded across multiple fields, driven by their versatility. Many
76 sectors are being impacted by this technology. For example, in the business sector, these
77 models may enhance technical support through advanced chatbots, automate report generation,
78 and facilitate data analysis (George and George 2023). In Education, LLMs can aid in creating
79 personalised educational content and developing intelligent tutoring systems (Cooper 2023).
80 In Medicine, they may assist with clinical documentation, medical report generation, and
81 preliminary diagnostics (Dave, Athaluri, and Singh 2023).

82
83 LLMs and generative IA have already been investigated for various Library and Information
84 Science applications. First, in text processing, LLMs may automate and improve the accuracy
85 of classifying large volumes of textual information (Brzustowicz 2023). Second, information
86 retrieval systems, such as commercial databases like Scopus, are incorporating these models to
87 advise users on search strategies (Elsevier Scopus Blog 2024). Additionally, native products
88 like Elicit provide new tools for searching and analysing scientific literature, such as by
89 offering summaries, question answering, and trend analysis based on large data corpora
90 (Whitfield and Hofmann 2023). In libraries and information centres, LLMs might enhance user
91 services through intelligent chatbots and recommendation systems, optimise query
92 management, and provide personalised assistance (Aboelmaged et al. 2024). One of the fields
93 most actively discussing the arrival of ChatGPT is the academic and research sector (Peres et
94 al. 2023).

95
96 In the chatbot market, major technology companies have presented their own solutions.
97 Google's Bard, now known as Gemini, is an alternative to ChatGPT (Google DeepMind 2024).

98 Microsoft has embedded ChatGPT into Bing and Microsoft 365, branding it as Copilot
99 (Adetayo, Aborisade, and Sanni 2024). Other notable examples include IBM's Watson
100 Assistant and Facebook's BlenderBot.

101

102 1.2. Features and functionalities of ChatGPT

103 ChatGPT is a generative text system that uses machine learning to closely mimic human
104 writing. The current iteration, version 4o released in May 2024, is equipped to deal with a broad
105 spectrum of specialist tasks. It is accessible online and through a mobile application. Interaction
106 with ChatGPT is primarily managed through one or more user prompts, which facilitate the
107 entry of specific questions and requests by users. The accuracy of the prompt formulation is
108 essential, as vague prompts can lead to poor results (Giray 2023). The art of prompt engineering
109 has become a specialist skill for ChatGPT users (White et al. 2023).

110

111 In addition to processing text, ChatGPT 4o is multimodal, meaning it can analyse multiple
112 types of data, including images, Excel spreadsheets, and PDF documents. Furthermore, the
113 mobile application version of ChatGPT 4o supports voice commands. ChatGPT users can also
114 create their own GPT bots tailored to specific tasks or knowledge bases¹. One of the most
115 powerful features of ChatGPT is its own API for automating its use. It can also be integrated
116 with external APIs and third-party services, facilitating connections with external data and
117 tools. Beyond these capabilities, ChatGPT incorporates ethical safeguards, such as mechanisms
118 to prevent the generation of harmful content. Its theoretical potential for facilitating academic
119 writing, data analysis, and even conducting complex research activities suggests that it may
120 play a role in advancing academic research.

121

122 1.3. Potential of ChatGPT for research and objectives

123 The potential of ChatGPT for research has been discussed by UNESCO, suggesting that it can
124 help the processes of design, data collection, data analysis, and the subsequent publication stage
125 (United Nations Educational, Scientific and Cultural Organization 2023). Discussions centring
126 on various health fields have also tended to emphasise the potential of ChatGPT. For example,
127 a review of biomedical studies about ChatGPT found that its benefits were mentioned in 51 of
128 the 60 studies analysed (85%), and these included: (1) improvements in scientific writing (Graf
129 and Bernardi 2023) and an increase in the equity and versatility of research; (2) the efficient
130 analysis of data sets, code generation, and literature reviews (Sallam 2023). Additionally,
131 various experiments have been conducted, simulating patient populations in biomedical data
132 and testing its effectiveness in drafting paper (Macdonald et al. 2023). These uses have led to
133 discussions about co-authorship with ChatGPT in scientific articles (Stokel-Walker 2023).

134

135 Generative AI also raises ethical and copyright issues, such as who owns the rights to the
136 generated text since it is partly derived from the uncredited authors of the texts used to train the
137 models (Lund et al., 2023). In addition, it can be difficult to detect ChatGPT-authored text
138 within academic papers, which is a problem for journals and conferences that ban it (Odri and
139 Ji Yun Yoon 2023; Sadasivan et al. 2023). As Gao points out in the case of ChatGPT's capacity
140 to generate believable scientific abstracts (Else 2023). In this regard, to avoid plagiarism, it has

¹ GPT enables users to explore and develop personalized versions of ChatGPT. This feature allows for the customization of the AI model by combining specific user instructions, integrating enhanced knowledge bases, and incorporating a variety of skills. Users can tailor their version of ChatGPT to suit specific tasks or industries, optimizing the AI's responses and functionalities to better meet their needs. The online version of ChatGPT includes a GPT search tool, allowing users to easily find and access different versions of ChatGPT tailored to specific needs.

141 been proposed that the detection of ChatGPT and similar technologies should be integrated into
142 academic platforms to catch plagiarism and safeguard the rights of researchers and writers
143 (Qasem 2023). The increased use of generative AI may also cause a decrease in critical thinking
144 and challenges with effectively evaluating the generated text (Rahman and Watanobe 2023),
145 including the detection of hallucinations, that is, incorrect or nonsensical information. This
146 problem might be reduced or exacerbated by custom GPTs that connect and make use of
147 controlled sources, such as the Dimensions GPT for scientific information search.

148
149 The use of ChatGPT is being debated across numerous fields, serving both as a writing tool
150 and a research instrument (Torres-Salinas and Arroyo-Machado 2024). Notably, the field of
151 Library and Information Science has been the subject of discussion in academic research
152 libraries (Houston and Corrado 2023), where it is posited to significantly enhance information
153 retrieval and user interactions (Lund, Khan, and Yuvaraj 2024). Although this study focuses
154 on bibliometrics, libraries have long integrated AI into their operations, with numerous
155 applications predating the emergence of ChatGPT. These include expert systems for reference
156 services and personalized information retrieval (Tian 2021), chatbots to enhance services and
157 attract users (Choukimath, Shivarama, and Gujral 2019), and improvements in collection
158 management and internal activities (Das and Islam 2021). The advent of generative AI and the
159 democratization brought about by tools like ChatGPT have now magnified and accelerated
160 these existing applications (Arroyo-Machado and Torres-Salinas 2024). ChatGPT may also
161 support fundamental services such as cataloguing and reference (Brzustowicz 2023) or research
162 assistance (Oyelude 2023) with potentially far-reaching implications for all areas that rely on
163 these services (Cox and Tzoc 2023). It may be especially useful in areas where analysis,
164 reporting, and research of large data sets are necessary. This is why one of the key areas of
165 application could be bibliometrics and scientific evaluation, where there have already been
166 some interesting results. For instance, ChatGPT-4 can produce plausible document summaries
167 and quality evaluation rationales that align with the REF criteria (Thelwall 2024).

168
169 This chapter explores ChatGPT's potential applications in bibliometric studies, organised into
170 five key areas: (I) preparation and consultation, (II) data processing, (III) data analysis, (IV)
171 results interpretation, and (V) writing. These areas, detailed in Table 1, encompass various
172 tasks we will elaborate on to demonstrate ChatGPT's capabilities. We argue that integrating
173 ChatGPT into bibliometric studies and other Library and Information Science fields can
174 enhance efficiency and output quality. We will break down the functionalities outlined in Table
175 1, providing a clearer picture of effective integration. Finally, we will critically review
176 ChatGPT's ethical, legal, and creativity-related issues.

177
178
179
180
181
182
183
184
185
186
187
188
189
190

Table 1. Tasks where ChatGPT could be used in bibliometrics.

Project stage	Detailed activities
Preparation and consultation <i>Establishing bibliometric goals and consulting relevant literature</i>	Review of manifestos and ethical and methodological guidelines.
	Design and preparation of contextualised case studies.
	Translating the goals of an analysis into a bibliometric strategy.
Data processing <i>Retrieving, cleaning, and combining bibliometric data for accurate analysis</i>	Identification and selection of relevant data sources.
	Extraction and downloading of data from multiple sources.
	Data cleaning, transformation, and normalisation.
	Guidance on database structuring.
	Data capture through APIs and process automation.
Data analysis <i>Analysing bibliometric data to identify trends and patterns</i>	Documentation and metadata description of datasets.
	Conducting descriptive and exploratory data analysis.
	Application of advanced statistical techniques.
	Development of data visualizations for interpretation.
	Calculation of bibliometric indicators and impact metrics.
	Consultation on the selection and application of indicators.
Results interpretation <i>Interpreting data to understand bibliometric impacts and relationships</i>	Implementation of Topic Modelling and content analysis.
	Critical interpretation of the obtained results.
Writing <i>Drafting the report summarising bibliometric findings and conclusions</i>	Validation of results through statistical methods.
	Critical review and editing of scientific texts.
	Writing of reports, articles, and presentations.
	Making recommendations based on the bibliometric results, if applicable.
	Summarising the limitations of the bibliometric approach used.
	Translation of scientific documents into multiple languages.

193 **2. APPLICATIONS AND EXAMPLES**

194 This section discusses potential bibliometric applications of ChatGPT as a starting point for a
 195 discussion of its value, irrespective of whether there is any evidence that it is useful or ethical
 196 in the suggested roles.

198 2.1. Preparation and consultation

199 Table 2 highlights three ways in which ChatGPT could help in the preparatory phases of
 200 bibliometric analyses.

202 To give a practical example, we used ChatGPT in the initial consultation phase of a bibliometric
 203 study into the impact of open access policies (Figure 1) and then asked for help with the
 204 recommended software (Figure 2). Whilst the first example output might be plausible to a non-
 205 expert, the recommended strategy is too general to be valid and recommends the widely
 206 discredited h-index. Similarly, whilst the second output seems helpful, it would be risky to rely
 207 on it rather than the official software documentation in case ChatGPT has guessed incorrectly.

Table 2. Utilising ChatGPT in preparation and consultation: tasks, descriptions, and real-case scenarios.


Task	Description	Example
Review of ethical and methodological guidelines	Assists in analysing and summarising key points from relevant documents to ensure adherence to best practices.	ChatGPT facilitates the examination and summary of the DORA declaration and CoARA recommendations, enabling researchers to integrate these considerations into their publication strategies and repository selections.
Design and preparation of contextualised case studies	Provides support in structuring and drafting case studies that reflect specific research contexts.	ChatGPT assists in designing a methodology for evaluating 50 scientists in accordance with evaluative principles, detailing the steps such as criteria selection, data collection methods, and analysis techniques for objective assessment.
Translating the goals of an analysis into a bibliometric strategy	Assists in defining and structuring bibliometric analyses to align with specific research objectives.	ChatGPT aids in outlining a bibliometric approach to track the evolution of research on open access over the last decade, including the identification of key journals, authors, and trends in publication strategies.

209
210
211

Figure 1. In the following example, a prompt is shown requesting guidance for a bibliometric study, and the output is displayed in table format.

1 Suggested prompt to receive a bibliometric study suggestion in a specific format

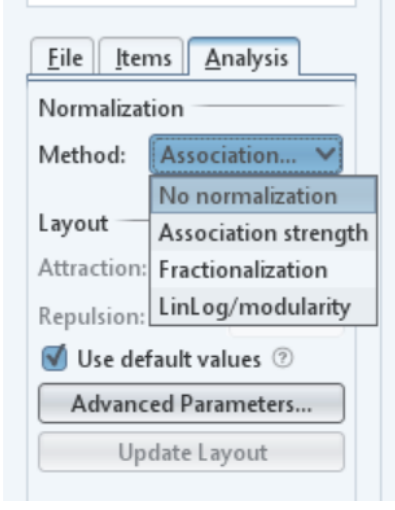
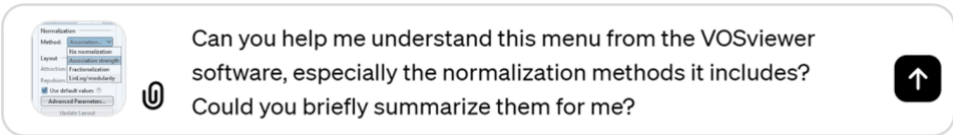
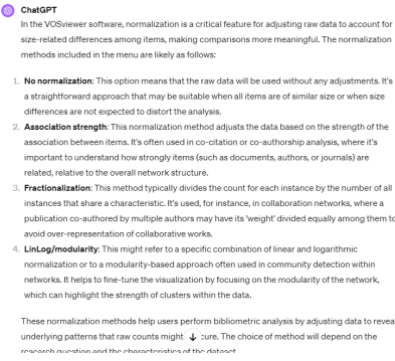
I need a table that summarizes the methodology and peer review process for a bibliometric analysis project, including evaluation criteria with specified weights, data collection methods mentioning specific databases and metrics, detailed analysis techniques identifying the software used, and a structured peer review process with defined steps and their importance. Please ensure the table provides a clear and comprehensive overview suitable for academic and research purposes.



2 Table generated by ChatGPT outlining the various elements of the study

Section	Details
Evaluation Criteria	1. Publication Output (30%): Number of peer-reviewed articles. 2. Citation Impact (40%): Citations and h-index. 3. Collaboration Networks (20%): Strength and diversity of co-authorships. 4. Policy Contributions (10%): Influence on policy-making documents.
Data Collection Methods	Utilizing Scopus for breadth of academic journals and Web of Science for robust citation data. Metrics include publication counts, citation numbers, and h-index. Data extraction facilitated through API access when available, ensuring the most current and comprehensive dataset.
Analysis Techniques	1. Citation Analysis : Using VOSviewer for citation mapping and impact assessment. 2. Social Network Analysis : Employing Gephi for visualizing collaboration networks, identifying key nodes and clusters. 3. Content Analysis for Policy Contributions : Applying NVivo to qualitatively assess the inclusion of scientific work in policy documents.
Importance of Peer Review	1. Indicator Verification (30%): Peer consultation for validating bibliometric indicators. 2. Report Analysis (40%): In-depth review by subject matter experts to ensure accuracy. 3. Candidate Scoring (30%): Objective scoring by peers, based on detailed reports and adherence to evaluation criteria.

Figure 2. Ask ChatGPT with a VOSviewer image to define the indicators and help us understand them.

	<p>1 In this instance, you were working with the VOSviewer software, and I want to know specific details about the functionalities and concrete indicators from the menu, particularly regarding normalisation aspects. I'm asking ChatGPT to describe them for me, and I'm attaching a small screenshot of the menu in the prompt.</p> 
<p>2 In the response, we see how ChatGPT can identify the different indicators included in the menu image and enlighten us on their use.</p> 	<p>In the VOSviewer software, normalisation is a critical feature for adjusting raw data to account for size-related differences among items, making comparisons more meaningful. The normalisation methods included in the menu are likely as follows:</p> <ul style="list-style-type: none"> • No normalisation: This option means that the raw data will be used without any adjustments. It's a straightforward approach that may be suitable when all items are of similar size or when size differences are not expected to distort the analysis. • Association strength: This normalisation method adjusts the data based on the strength of the association between items. It's often used in co-citation or co-authorship analysis, where it's important to understand how strongly items (such as documents, authors, or journals) are related, relative to the overall network structure. • Fractionalization: This method typically divides the count for each instance by the number of all instances that share a characteristic. It's used, for instance, in collaboration networks, where a publication co-authored by multiple authors may have its 'weight' divided equally among them to avoid over-representation of collaborative works. • LinLog/modularity: This might refer to a specific combination of linear and logarithmic normalisation or to a modularity-based approach often used in community detection within networks. It helps to fine-tune the visualisation by focusing on the modularity of the network, which can highlight the strength of clusters within the data. <p>These normalisation methods help users perform bibliometric analysis by adjusting data to reveal underlying patterns that raw counts might obscure. The choice of method will depend on the research question and the characteristics of the dataset.</p>

213

214 2.2. Data processing

215 ChatGPT may help with at least six data processing tasks (Table 3). All of these are sensitive
 216 actions in the sense that errors could invalidate a study and so expert oversight would be needed
 217 to check the results. To help with this, ChatGPT might be prompted for explicit reasons for its
 218 choices or supporting information (Figure 3). It is not clear whether the ChatGPT outputs
 219 would tend to be a helpful timesaver or a problematic distraction for these tasks.

220

221

222

223

224

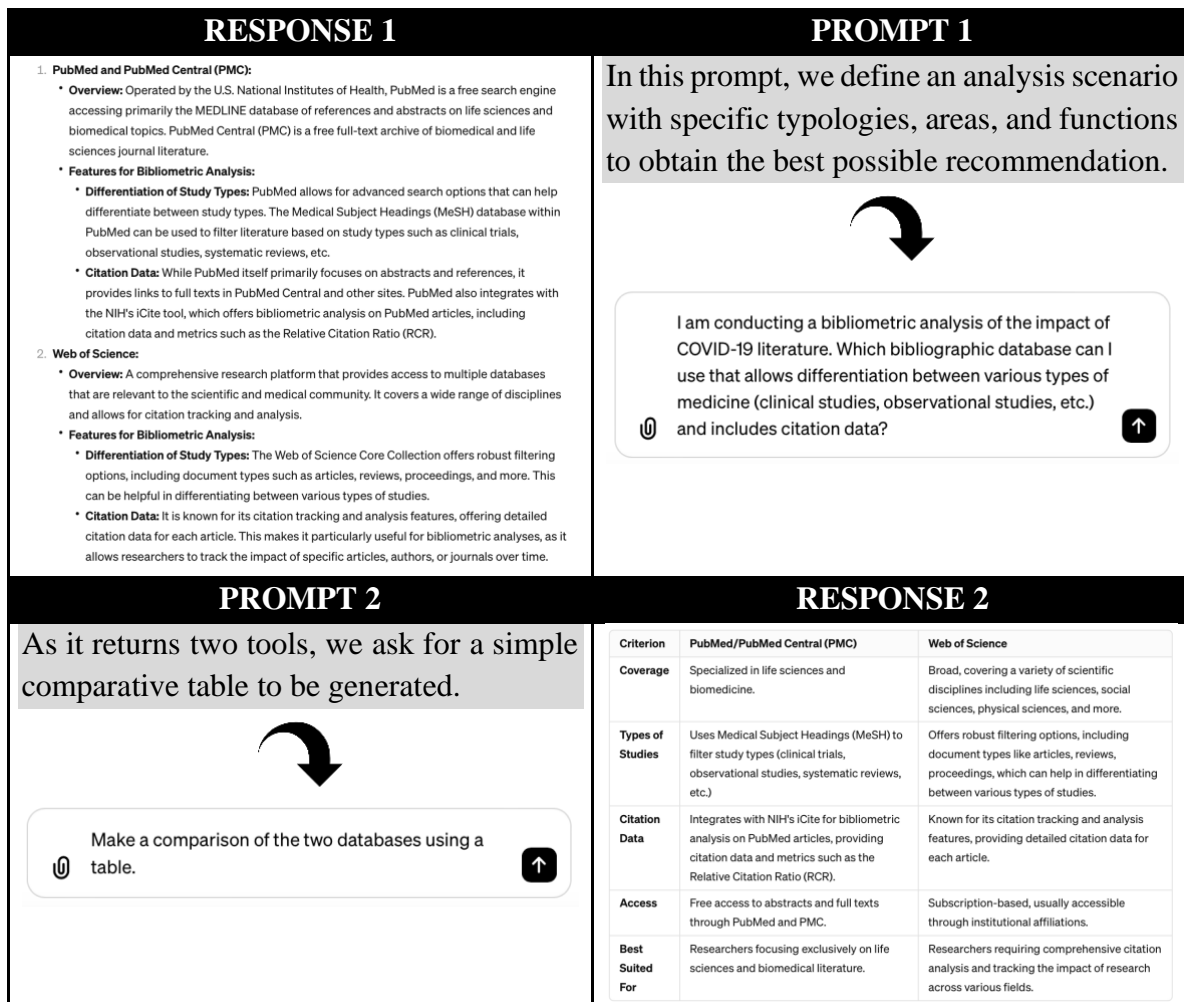
225
226
227
228
229
230
231
232
233
234
235

Table 3. Utilising ChatGPT in bibliometric data processing: tasks, descriptions, and real-case scenarios.

Task	Description	Example
<i>Identification and selection of relevant data sources</i>	Suggests databases, journals, and platforms for data collection.	Asking ChatGPT to identify the most appropriate database for a study on social sciences in Spanish might lead to the suggestion of Scopus or Web of Science due to their extensive coverage and language support.
<i>Extraction and downloading of data from multiple sources</i>	Automates the process of retrieving data from identified sources.	Using ChatGPT to automate the extraction of data from PubMed and arXiv for a review on neurodegenerative diseases.
<i>Data cleaning, transformation, and normalisation</i>	Refines raw data into a usable format, addressing inconsistencies and preparing for analysis.	ChatGPT could assist in scripting for data cleaning routines, such as standardising author names and affiliations from different citation styles.
<i>Guidance on database structuring</i>	Advice on organising data in a structured manner conducive to analysis.	ChatGPT offers insights on how to structure a relational database for storing bibliometric data, emphasising the relationships between authors, publications, and citations.
<i>Data capture through APIs and process automation</i>	Uses APIs for efficient data retrieval and employs automation for repetitive tasks.	ChatGPT aids in creating scripts to query the Crossref API for real-time citation data, enhancing the dataset's currency and accuracy.
<i>Documentation and metadata description of datasets</i>	Creating detailed records about the dataset, ensuring clarity and reusability.	ChatGPT helps draft comprehensive documentation for a dataset, outlining its scope, methodology, and potential limitations, ensuring future researchers can understand and utilise the data effectively.

236
237
238
239
240
241
242
243
244
245

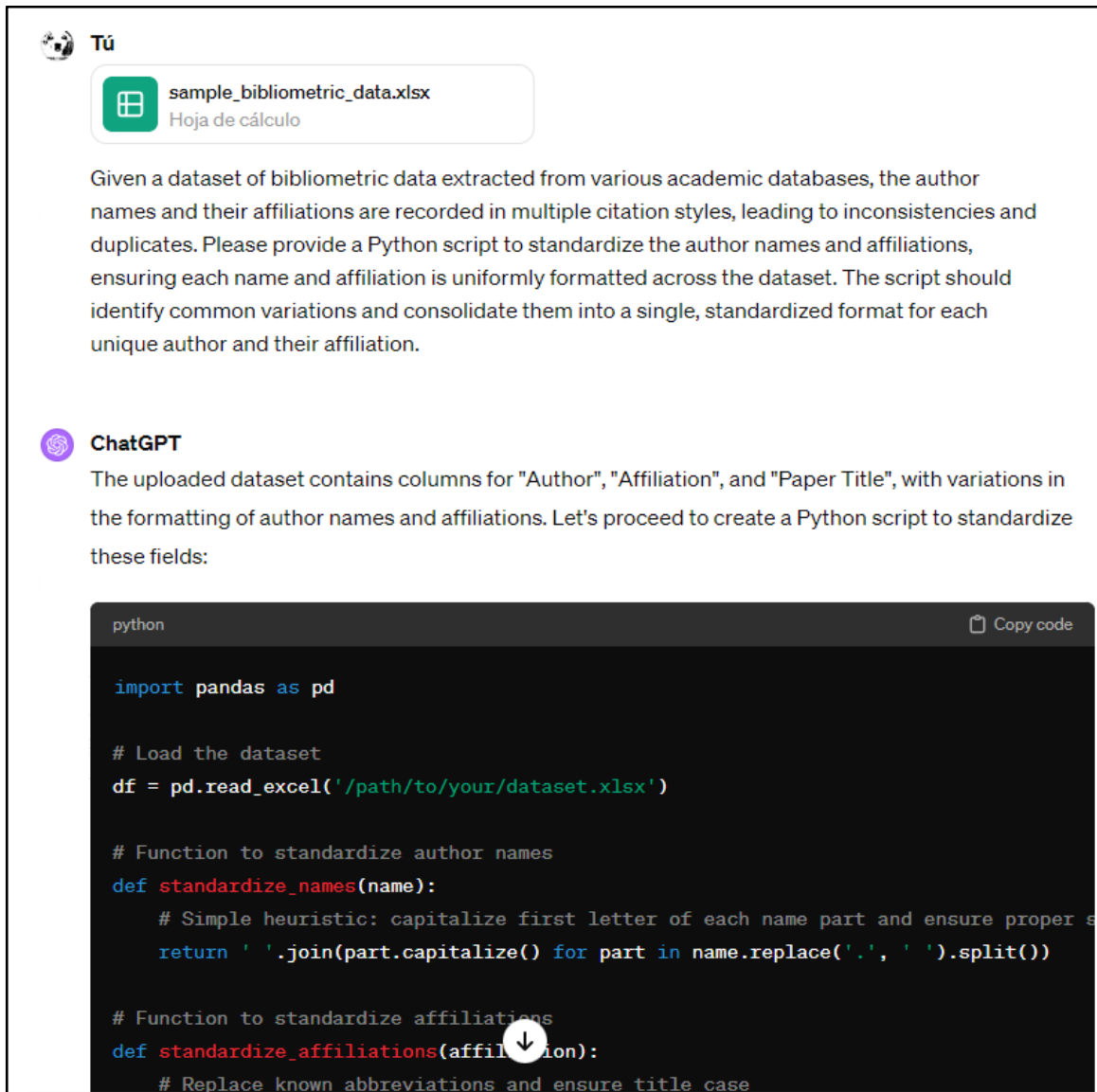
Figure 3. Example of assistance for choosing a bibliographic database and comparison.



ChatGPT may be particularly useful for programming tasks if the bibliometrician is not an experienced programmer. This might include constructing API queries. Figure 4 provides an example where ChatGPT constructs a Python program to normalise author names as part of the data preparation phase. The script employed regular expressions to identify and standardise common variations in author names (e.g., "J. Doe" vs. "John Doe") and affiliations (e.g., "Univ. of Example" vs. "University of Example"). It also included a function to match and merge records that, after processing, were identified as belonging to the same individual or institution, despite the initial discrepancies in how their names were presented. This application might improve the quality of the dataset, enabling more accurate analyses of publication patterns, collaborations, and institutional affiliations within the field of study. A competent Python coder should be able to check that the code works, saving time in comparison to identifying the appropriate functions to write the code themselves (Figure 4).

268
269
270

Figure 4. Example of a prompt in ChatGPT to request a Python routine that allows for the normalisation of author names and institutions in a dataset



The screenshot shows a chat interface. At the top, a user profile icon is labeled 'Tú'. Below it, a file upload box contains a green spreadsheet icon, the filename 'sample_bibliometric_data.xlsx', and the text 'Hoja de cálculo'. The user's message reads: 'Given a dataset of bibliometric data extracted from various academic databases, the author names and their affiliations are recorded in multiple citation styles, leading to inconsistencies and duplicates. Please provide a Python script to standardize the author names and affiliations, ensuring each name and affiliation is uniformly formatted across the dataset. The script should identify common variations and consolidate them into a single, standardized format for each unique author and their affiliation.'

Below the user message is the ChatGPT response, which begins with the GPT icon and the text: 'The uploaded dataset contains columns for "Author", "Affiliation", and "Paper Title", with variations in the formatting of author names and affiliations. Let's proceed to create a Python script to standardize these fields:'. The response then shows a code block with the following Python code:

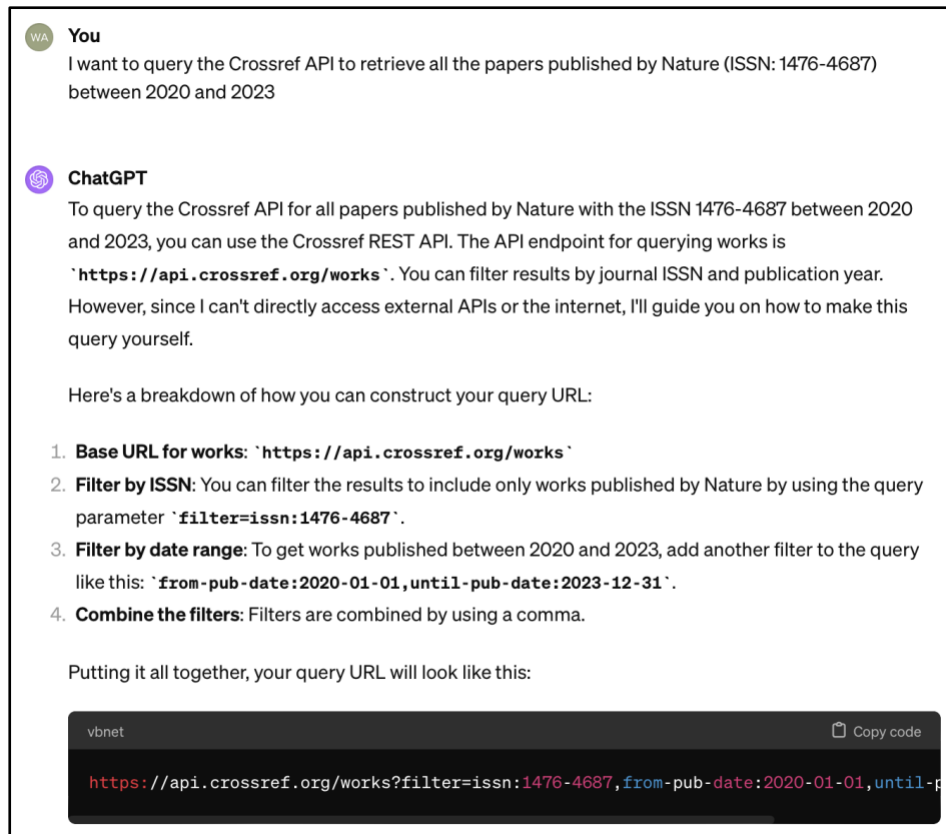
```
python Copy code  
  
import pandas as pd  
  
# Load the dataset  
df = pd.read_excel('/path/to/your/dataset.xlsx')  
  
# Function to standardize author names  
def standardize_names(name):  
    # Simple heuristic: capitalize first letter of each name part and ensure proper spacing  
    return ' '.join(part.capitalize() for part in name.replace('.', ' ').split())  
  
# Function to standardize affiliations  
def standardize_affiliations(affiliation):  
    # Replace known abbreviations and ensure title case
```

271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286

Similarly to the above example, Figure 5, illustrates ChatGPT's ability to construct a specific format of API query, with an explanation that can be checked. This can save the programmer the time to look up all the components of the query in the API documentation.

287
288
289
290
291
292

Figure 5. Example of a prompt for constructing a query to the Crossref API.



293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313

2.3. Data analysis

ChatGPT could be used to analyse data (e.g., uploaded in a spreadsheet), suggest a data analysis strategy or write code for a particular analysis (Table 4). Relying on ChatGPT to process the data is particularly problematic because the researcher then no longer has the oversight to check that the processing steps are correct. Since bibliometric processing often has many steps and some require expert input, relying on ChatGPT's guesses is unethical and potentially disastrous. This type of analysis should therefore never be used unless substantial evidence accrues from multiple sources that the results are reliable.

Table 4. Using ChatGPT in data analysis: tasks, descriptions, and real-case scenarios.

Task	Description	Example
Conducting descriptive and exploratory data analysis	Involves summarising the main characteristics of a dataset to uncover patterns or anomalies.	ChatGPT aids in scripting preliminary data analyses, guiding researchers through the process of identifying trends within their bibliometric data.
Application of advanced statistical techniques	Utilises complex statistical methods to interpret data and draw conclusions.	ChatGPT suggests appropriate statistical methods for complex data sets, such as regression analysis or factor analysis, to uncover deeper insights in bibliometric studies.
Development of data visualisations for interpretation	Creates visual representations of data to facilitate understanding and communication of findings.	ChatGPT assists in selecting the most effective data visualisation tools and techniques for showcasing bibliometric findings, such as network graphs or heatmaps.
Calculation of bibliometric indicators and impact metrics	Computes measures to assess the impact and quality of research outputs.	ChatGPT helps outline the steps for calculating key bibliometric indicators like the impact factor, ensuring accurate assessment of research impact.
Consultation on the selection and application of indicators	Advises on the most suitable metrics for specific research evaluation needs.	ChatGPT provides insights on the pros and cons of various bibliometric indicators, helping researchers choose the most relevant metrics for their analysis.
Implementation of Topic Modelling and content analysis	Employs algorithms to discover the thematic structure of a large corpus of text.	ChatGPT guides the setup and interpretation of Topic Modelling, facilitating the exploration of predominant themes in scholarly publications.

315
 316 To illustrate descriptive and exploratory data analysis, we provided ChatGPT with a tabulated
 317 Web of Science file (Figure 6). The GPT Data Analyst² is an AI-driven feature, equipped to
 318 handle, interpret, and provide insights from datasets by executing various statistical analyses,
 319 identifying trends, generating visualizations, and summarizing key findings. In this case, we
 320 see how a file containing publications allows us to directly perform basic analyses on scientific
 321 production and indicators, in this case, Henk Moed, from which a metrics table and various
 322 descriptive analyses (themes and co-authors) are derived.


323
 324 ChatGPT can execute various programming languages, including Python. It can integrate with
 325 Python's ecosystem of libraries to perform detailed data analysis and create interactive data
 326 visualisations. For data analysis, it can use libraries such as Scikit-learn for machine learning
 327 and statistical modelling. For visualisations, ChatGPT can use libraries such as Matplotlib and
 328 Plotly. These libraries enable the creation of a diverse range of visualisations, from simple bar
 329 charts and line graphs to more complex scatter plots and heatmaps. This capability is

² The "GPT Data Analyst" is a tool leveraging the GPT framework to perform data analysis. It automates statistical analysis, trend identification, and the generation of visualizations, streamlining data insights with minimal human input. This tool is part of GPT applications or available from GPT library or store.

330 particularly valuable for analysing trends, displaying statistical information, and making data
 331 more comprehensible. If the programming code can be seen and checked then this is a valid
 332 use, especially if a statistician can verify whether the graphs are appropriate for the data. Figure
 333 7 demonstrates ChatGPT's ability to generate visualisations directly from data files; in this
 334 instance, a conventional bibliometric file in Excel format, developed from data from the Journal
 335 Citation Reports (JCR). It is essential to specify the code and libraries used in the creation
 336 process to produce these charts. This detail is crucial because the graphs must be carefully
 337 reviewed to ensure their accuracy, which sometimes involves revising the code. Thus,
 338 familiarity with the relevant code and libraries is necessary. Additionally, the process can be
 339 replicated manually for more controlled use, ensuring that every step is understood and
 340 verified.

341
 342 **Figure 6.** Example of an exploratory analysis of an author's file downloaded from Web of Science and analyzed
 343 with the GPT "Data Analyst".


1 The image displays a specialized GPT model designed by OpenAI specifically for data analysis. We can see that a file containing Henk Moed's publications downloaded from the Web of Science has been attached. The prompt instructs the model to carry out some basic calculations. Although the tasks set out here are fundamental, the GPT Data Analyst has the capability to perform much more complex analyses.



Data Analyst

By ChatGPT

Drop in any files and I can help analyze and visualize your data.



HENK MOED papers.txt

Documento

I have provided a tab file from Web of Science. Conduct an exploratory data analysis to summarize main bibliometric indicators such as citation counts, number of articles, distribution by document types, key publishing journals, and trends over the years. Also I need basis descriptive statistical indicators. I'm looking for a global vision of the dataset to identify.

2 In the following image, we see the response provided, which promptly delivers an overarching view.

Based on the exploratory data analysis of the dataset provided, here are the summaries of the main bibliometric indicators:

- **Total Number of Articles:** The dataset contains a total of 179 entries.
- **Distribution by Document Types:**
 - Articles: 78
 - Article; Book Chapter: 28
 - Editorial Material: 18
 - Proceedings Paper: 15
 - Article; Proceedings Paper: 15
 - (Others include Letter, Editorial Material; Book Chapter, Meeting Abstract, Correction, Review, Biographical-Item, Book, Book Review, Note)
- **Key Publishing Journals (Top 10):**
 - Scientometrics: 54 publications
 - Citation Analysis in Research Evaluation: 30 publications
 - Journal of Informetrics: 12 publications
 - Nature: 7 publications
 - Research Evaluation: 7 publications
 - Research Policy: 6 publications
- **Citation Counts Summary:**
 - Mean: 43.88 citations per document
 - Standard Deviation: 77.93
 - Min: 0 citations
 - 25th Percentile: 1 citation
 - Median (50th Percentile): 10 citations
 - 75th Percentile: 48.5 citations
 - Max: 410 citations
- **Trends Over the Years:** Publications span from 1985 to 2022, with varying numbers of publications per year. The year 2005 saw the highest number of publications in the dataset (36 publications).

3 Next, we see how we can ask all kinds of exploratory questions about the dataset.

Could you briefly indicate Moed's three main research lines, provide a very brief description of each, and the terms's frequency associated with that line?

1. **Citation Analysis and Impact:** Moed investigates the significance of citations (38 mentions) to understand the impact (36 mentions) and reach of research (56 mentions). This area focuses on how citations can reflect the value and influence of scientific works, aiming to unravel the narrative behind citation patterns and their role in academic recognition.
2. **Bibliometrics and Scientometrics:** Through bibliometric (25 mentions) techniques, Moed analyzes scientific (34 mentions) publications to quantify the dynamics of science (30 mentions). This research examines publication trends, collaboration networks, and the distribution of knowledge across fields, offering insights into the structural aspects of scientific activity and its measurement.
3. **Journal and Research Performance Indicators:** Moed's work on evaluating journals (25 mentions) through performance indicators (18 mentions) aims to assess the quality and impact of scholarly outputs. This involves developing and applying metrics that accurately reflect the contributions of journals and research articles to the scientific community, aiding in the strategic decisions of academia and policy.

Can you tell me the names and institutions of his main collaborators in Spain?

Let's proceed with identifying these collaborators and their institutions.

Based on the data, Henk Moed's main collaborators in Spain and their institutions include:

1. **Vicente P. Guerrero-Bote:**
 - Institution: Universidad de Extremadura, Department of Information and Communication, Badajoz, Spain.
2. **Felix de Moya-Anegón:**
 - Institution: SCImago Group, Madrid, Spain; also associated with CSIC (Spanish National Research Council), Madrid, Spain.
3. **Carmen Lopez-Illescas:**
 - Institutions: Universidad Complutense Madrid, Department of Library and Information Science, Madrid, Spain; also mentioned in connection with the University of Granada, Granada, Spain.

344
345

Figure 7. Example of creating visualisations with ChatGPT from a file directly downloaded from the JCR.

✕

DanielTorres-Salinas_JCR_...
Hoja de cálculo

With this collection of one hundred scientific journals and their corresponding Journal Citation Reports (JCR) indicators, I require the creation of four distinct visualizations (such as a heatmap to display data density, a cluster diagram organized by indicator values, a bivariate map for comparative analysis, and a fourth visualization of your choosing) that will facilitate a deeper insight into the patterns and relationships within the indicator data.

1 In the following example, we input a basic prompt along with a file extracted from the Journal Citation Reports (JCR) containing six indicators, and we request ChatGPT to create various visualisations.

[File with original data](#)

2 Before generating the graphics, ChatGPT explains, on one hand, the methodology it will follow, the programming language, and libraries (2a), and then briefly describes the four graphics that it will produce (2b)

2a

To proceed with creating these visualizations, the first step involves parsing the data into a structured format suitable for analysis, such as a pandas DataFrame in Python. This structured format will enable the efficient manipulation and visualization of the data using libraries like matplotlib and seaborn for the heatmaps, cluster diagrams, and scatter plots, alongside other suitable tools for the bivariate map. Let's begin by parsing the data. [-]

2b

Here is a summary of the visualizations we will develop:

1. **Heatmap to Display Data Density:** This will visualize the distribution of journals across selected indicators, helping identify areas of high density where many journals share similar metric values.
2. **Cluster Diagram (Dendrogram):** By grouping journals based on the similarity of their indicator values, this visualization will reveal natural clusters, helping to identify journals with similar performance metrics.
3. **Bivariate Map for Comparative Analysis:** This plot will compare two indicators against each other for each journal, allowing us to observe correlations between different metrics directly.
4. **Scatter Plot Matrix (SPLOM):** This comprehensive visualization will showcase pairwise relationships between all indicators, providing an overview of how these metrics interrelate across the dataset.

3 Finally, here we have an image of the visualisations generated directly through ChatGPT

Correlation Heatmap of Journal Indicators

346

347 2.4. Results interpretation

348 ChatGPT may be used to help interpret and validate results through statistical methods (Table
349 5). It may help by suggesting conclusions and by finding ways of framing these insights within
350 the context of prior research, which may be checked.

351
352

Table 5. Utilising ChatGPT in results and interpretation: tasks, descriptions, and real-case scenarios.

Task	Description	Example
<i>Critical interpretation of the obtained results</i>	Engages in-depth analysis to understand the significance and implications of research findings.	ChatGPT facilitates a discussion on interpreting complex bibliometric data, assisting in drawing meaningful conclusions about research trends and patterns.
<i>Validation of results through statistical methods</i>	Uses statistical techniques to confirm the reliability and accuracy of the study's conclusions.	ChatGPT advice on selecting and applying robust statistical tests to validate bibliometric findings, ensuring their credibility and contributing to scholarly discourse.

14

353 In the realm of validating results, ChatGPT provides guidance on employing statistical methods
 354 to ensure the accuracy and reliability of the research conclusions. It advises on the selection
 355 and application of appropriate statistical tests, such as chi-square tests for categorical data or t-
 356 tests for comparing means, which might be useful suggestions if they can be checked by the
 357 researcher. For example, ChatGPT might guide a researcher through the process of validating
 358 a hypothesis about the correlation between the number of citations and the impact factor of
 359 journals, but statistical expertise would be needed to check that the correct type of correlation
 360 was used (e.g., Spearman or Pearson).

361
 362 2.5. Writing

363 Since ChatGPT is fundamentally a language model, and thus, it may be effective for scientific
 364 report writing. Although ChatGPT can generate text that is grammatically correct and
 365 semantically coherent, it does not possess true comprehension and may generate inaccurate or
 366 hallucinated content. Effective utilization of ChatGPT for academic writing demands
 367 meticulous prompting, training, and verification of its outputs. Additionally, while ChatGPT
 368 can be beneficial for tasks such as outlining, brainstorming, and editing, it should not be
 369 considered a substitute for the intellectual and creative efforts required by the writer (Lingard
 370 2023).

371
 372 *Table 6. Utilising ChatGPT in writing: tasks, descriptions, and real-case scenarios.*

Task	Description	Example
<i>Critical review and editing of scientific texts.</i>	Involves the evaluation and refinement of scholarly manuscripts to improve clarity and accuracy.	ChatGPT assists in reviewing a draft article, suggesting improvements in structure and argumentation, and identifying sections that require further clarification.
<i>Writing of reports, articles, and presentations.</i>	Facilitates the creation of various scientific communication materials.	ChatGPT aids in drafting a conference presentation, structuring the content to convey research findings effectively and engagingly.
<i>Making recommendations based on the bibliometric results, if applicable</i>	Analyses bibliometric data to derive actionable insights and recommendations that can guide future research directions, policy-making, or academic discourse.	ChatGPT aids researchers in identifying significant trends within publication data, suggesting potential areas for further research or areas needing more substantial funding. For example, after analysing data on renewable energy research outputs, ChatGPT could recommend a greater focus on emerging but under-researched technologies like biofuels, providing strategic insights useful for shaping research agendas and funding policies.
<i>Summarising the limitations of the bibliometric approach used</i>	Critically assesses the bibliometric methods used to identify potential biases or limitations in data collection, analysis, or interpretation.	ChatGPT assists in drafting sections of research papers that outline the limitations of bibliometric analyses, such as emphasising the predominance of English-language sources in the dataset. This summary helps maintain the credibility of the research by transparently addressing its methodological constraints and the potential impact on the understanding of global research trends.
<i>Translation of scientific documents into multiple languages</i>	Enables the broader dissemination of research by translating texts into different languages.	ChatGPT supports the translation of a research paper from English to Spanish, ensuring the technical terminology is accurately conveyed.

373
374
375
376
377
378
379
380
381

In Table 6, we have illustrated how ChatGPT can assist in various tasks encountered during the writing of scientific papers. This model may improve grammar, support literature reviews, and suggest professional knowledge, which are crucial for crafting high-quality documents. It may also enhance manuscripts by strengthening their structure, bolstering arguments, and clarifying ambiguous sections. ChatGPT can also convert papers into well-structured slides that the author can tailor, saving time compared to writing all the slides unaided. Moreover, ChatGPT may be particularly useful for translating scientific documents from other languages.

382 **3. LIMITATIONS, ETHICAL, AND LEGAL CONSIDERATIONS**

383 The deployment of AI in bibliometric analysis introduces a complex array of ethical issues that
384 stem from the inherent limitations and critical points in various project stages (Table 7). One
385 of the fundamental ethical challenges arises from the potential for misinterpretation or non-
386 compliance with ethical and methodological guidelines during the preparation and consultation
387 phase. This can lead to ethical dilemmas when AI systems, despite being designed to operate
388 within certain ethical frameworks, reinforce biases or lack representativeness in case selection
389 due to an inadequate understanding of the guidelines. Moreover, the reliance on AI for data
390 processing and analysis, including data cleaning and the application of advanced statistical
391 techniques, introduces risks of perpetuating or even exacerbating existing biases. These biases
392 may not only affect the accuracy of the bibliometric analysis but also raise ethical concerns
393 regarding fairness and equality in scientific research dissemination. Additionally, the ethical
394 responsibility of ensuring accurate and sufficient metadata description, along with the critical
395 interpretation of obtained results, underscores the need for transparency and accountability in
396 AI-assisted bibliometric research. As AI technologies continue to evolve, addressing these
397 ethical issues necessitates ongoing dialogue among researchers, ethicists, and technologists to
398 develop and adhere to comprehensive ethical standards that safeguard against potential harms.
399

400 On the legal front, the use of AI like ChatGPT in bibliometric analysis brings to the fore
401 significant legal considerations, particularly concerning data privacy, intellectual property, and
402 compliance with regulatory frameworks. The collection, extraction, and processing of data
403 from multiple sources involve navigating complex legal landscapes to ensure adherence to data
404 protection laws, such as GDPR³ in Europe, which impose strict requirements on data handling
405 and privacy. Moreover, the dependency on APIs for data capture and process automation
406 presents legal risks related to the terms of service and copyright laws, which may limit the
407 scope of permissible uses of data. Intellectual property rights also pose a legal challenge, as the
408 creation and use of bibliometric databases and analyses must respect the copyright and
409 ownership of original research materials. In addition, the automation of data analysis and report
410 generation by AI systems necessitates a careful consideration of the legal implications of
411 automated decisions, especially when they influence research funding, publication, and
412 reputation. Ensuring compliance with these diverse legal requirements demands a thorough
413 understanding of both domestic and international laws and underscores the need for legal
414 expertise in the design and implementation of bibliometric projects. Addressing these legal
415 challenges is crucial for maintaining the integrity and credibility of bibliometric analysis and
416 for fostering trust in AI-assisted research methodologies.
417

³ The General Data Protection Regulation (GDPR) is a comprehensive data protection law in the EU that sets guidelines for the collection and processing of personal information from individuals within the European Union and the European Economic Area.

418
419

Table 7. Limitations and risks in the main tasks where ChatGPT could intervene or be utilised in bibliometrics.

Project stage	Detailed activities	Limitations and critical points
Preparation and consultation	Review of manifestos and ethical and methodological guidelines.	Potential for misinterpretation or non-compliance due to limited understanding of guidelines.
	Design and preparation of contextualised case studies.	Risk of bias or lack of representativeness in case selection.
	Translating the goals of an analysis into a bibliometric strategy.	Risks of misalignment between translated objectives and actual outcomes.
Data processing	Identification, comparison, and selection of relevant data sources.	Limited availability of relevant data sources.
	Extraction and downloading of data from multiple sources.	Technical errors or data format inconsistencies may occur.
	Data cleaning, transformation, and normalisation.	Risks of errors or biases during data cleaning.
	Guidance on database structuring.	Inadequate database structure may hinder analysis.
	Data capture through APIs and process automation.	Dependency on APIs may entail risks of non-compliance or discontinuation.
	Documentation and metadata description of datasets.	Inaccurate or insufficient metadata may hinder data interpretation
Data analysis	Conducting descriptive and exploratory data analysis.	Risk of overlooking important patterns due to overreliance on descriptive statistics.
	Application of advanced statistical techniques.	Risks of misapplication leading to erroneous conclusions.
	Development of data visualisations for interpretation.	Ineffective visualisations may mislead interpretations.
	Calculation of bibliometric indicators and impact metrics.	Risks of oversimplification in calculations leading to misinterpretation.
	Consultation on the selection and application of indicators.	Inappropriate indicators may misrepresent research impact.
	Implementation of Topic Modeling and content analysis.	Results may be influenced by subjective choices or biases.
Results interpretation	Critical interpretation of the obtained results.	Risks of biased interpretations.
	Validation of results through statistical methods.	Statistical validation may not capture the full complexity of the data.
Bibliometric reporting	Critical review and editing of scientific texts.	Inadequate reporting may compromise transparency and reproducibility.
	Writing of reports, articles, and presentations.	Poor communication may lead to misunderstandings.
	Making recommendations based on the bibliometric results, if applicable.	Recommendations may be limited by analysis scope or quality.
	Summarising the limitations of the bibliometric approach used.	Failure to acknowledge limitations may undermine credibility.

420
421
422
423
424
425

Finally, as we contemplate integrating tools like ChatGPT into bibliometrics and other Library and Information Science domains, we must thoroughly assess the significant shifts they might trigger in our cognitive processes. Scholars have expressed concerns about employing AI technologies in academia, particularly highlighting AI's shortcomings in mimicking human reflexivity and its potential to erode the foundation of original scholarship (Lindebaum and

426 Fleming 2024) with the generation of unsupervised “botshit” (Hannigan, McCarthy, and Spicer
427 2024). They contend that AI's dependence on data patterns does not equate to real
428 understanding or critical thinking, which poses a threat to the quality and integrity of academic
429 research. These authors urge caution against allowing the efficiency of AI to overshadow the
430 indispensable human capacities for creativity and critical analysis, which are vital for
431 cultivating insightful and innovative scholarship.
432

433 REFERENCES

- 434 Aboelmaged, Mohamed, Shaker Bani-Melhem, Mohd Ahmad Al-Hawari, and Ifzal Ahmad.
435 2024. ‘Conversational AI Chatbots in Library Research: An Integrative Review and
436 Future Research Agenda’. *Journal of Librarianship and Information Science*, February,
437 09610006231224440. <https://doi.org/10.1177/09610006231224440>.
- 438 Adetayo, Adebowale Jeremy, Mariam Oyinda Aborisade, and Basheer Abiodun Sanni. 2024.
439 ‘Microsoft Copilot and Anthropic Claude AI in Education and Library Service’. *Library*
440 *Hi Tech News* ahead-of-print (ahead-of-print). <https://doi.org/10.1108/LHTN-01-2024-0002>.
- 442 Arroyo-Machado, Wenceslao, and Daniel Torres-Salinas. 2024. ‘Manual de ChatGPT:
443 Aplicaciones En Documentación y Bibliotecas’, May.
444 <https://doi.org/10.5281/ZENODO.11350826>.
- 445 Brzustowicz, Richard. 2023. ‘From ChatGPT to CatGPT: The Implications of Artificial
446 Intelligence on Library Cataloging’. *Information Technology and Libraries* 42 (3).
447 <https://doi.org/10.5860/ital.v42i3.16295>.
- 448 Choukimath, Puttaraj A., J. Shivarama, and Garima Gujral. 2019. ‘Perceptions and Prospects
449 of Artificial Intelligence Technologies for Academic Libraries: An Overview of Global
450 Trends’. In . INFLIBNET Centre, Gandhinagar.
451 <https://ir.inflibnet.ac.in:8443/ir/handle/1944/2337>.
- 452 Cooper, Grant. 2023. ‘Examining Science Education in ChatGPT: An Exploratory Study of
453 Generative Artificial Intelligence’. *Journal of Science Education and Technology* 32
454 (3): 444–52. <https://doi.org/10.1007/s10956-023-10039-y>.
- 455 Cox, Christopher, and Elias Tzoc. 2023. ‘ChatGPT: Implications for Academic Libraries | Cox
456 | College & Research Libraries News’, March. <https://doi.org/10.5860/crln.84.3.99>.
- 457 Das, Rajesh Kumar, and Mohammad Sharif Ul Islam. 2021. ‘Application of Artificial
458 Intelligence and Machine Learning in Libraries: A Systematic Review’. arXiv.
459 <https://doi.org/10.48550/ARXIV.2112.04573>.
- 460 Dave, Tirth, Sai Anirudh Athaluri, and Satyam Singh. 2023. ‘ChatGPT in Medicine: An
461 Overview of Its Applications, Advantages, Limitations, Future Prospects, and Ethical
462 Considerations’. *Frontiers in Artificial Intelligence* 6 (May).
463 <https://doi.org/10.3389/frai.2023.1169595>.
- 464 Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. ‘BERT: Pre-
465 Training of Deep Bidirectional Transformers for Language Understanding’. arXiv.
466 <http://arxiv.org/abs/1810.04805>.
- 467 Else, Holly. 2023. ‘Abstracts Written by ChatGPT Fool Scientists’. *Nature* 613 (7944): 423–
468 423. <https://doi.org/10.1038/d41586-023-00056-7>.
- 469 Elsevier Scopus Blog. 2024. ‘Scopus AI Release’. 2024. <https://blog.scopus.com/posts/scopus-ai-release-may-2024>.
- 471 Frantar, Elias, Saleh Ashkboos, Torsten Hoefler, and Dan Alistarh. 2023. ‘GPTQ: Accurate
472 Post-Training Quantization for Generative Pre-Trained Transformers’. arXiv.
473 <http://arxiv.org/abs/2210.17323>.

- 474 George, A. Shaji, and AS Hovan George. 2023. 'A Review of ChatGPT AI's Impact on Several
475 Business Sectors'. *Partners Universal International Innovation Journal* 1 (1): 9–23.
- 476 Giray, Louie. 2023. 'Prompt Engineering with ChatGPT: A Guide for Academic Writers'.
477 *Annals of Biomedical Engineering* 51 (12): 2629–33. [https://doi.org/10.1007/s10439-](https://doi.org/10.1007/s10439-023-03272-4)
478 [023-03272-4](https://doi.org/10.1007/s10439-023-03272-4).
- 479 Google DeepMind. 2024. 'Gemini Models'. 2024.
480 <https://deepmind.google/technologies/gemini/>.
- 481 Graf, A., and R.E. Bernardi. 2023. 'ChatGPT in Research: Balancing Ethics, Transparency and
482 Advancement'. *Neuroscience* 515:71–73.
483 <https://doi.org/10.1016/j.neuroscience.2023.02.008>.
- 484 Hannigan, Timothy R., Ian P. McCarthy, and André Spicer. 2024. 'Beware of Botshit: How to
485 Manage the Epistemic Risks of Generative Chatbots'. *Business Horizons*.
486 <https://www.sciencedirect.com/science/article/pii/S0007681324000272>.
- 487 Houston, Aileen B., and Edward M. Corrado. 2023. 'Embracing ChatGPT: Implications of
488 Emergent Language Models for Academia and Libraries'. *TECHNICAL SERVICES*
489 *QUARTERLY* 40 (2): 76–91. <https://doi.org/10.1080/07317131.2023.2187110>.
- 490 Lindebaum, Dirk, and Peter Fleming. 2024. 'ChatGPT Undermines Human Reflexivity,
491 Scientific Responsibility and Responsible Management Research'. *British Journal of*
492 *Management* 35 (2): 566–75. <https://doi.org/10.1111/1467-8551.12781>.
- 493 Lingard, Lorelei. 2023. 'Writing with ChatGPT: An Illustration of Its Capacity, Limitations &
494 Implications for Academic Writers'. *Perspectives on Medical Education* 12 (1): 261.
- 495 Lund, Brady D., Daud Khan, and Mayank Yuvaraj. 2024. 'ChatGPT in Medical Libraries,
496 Possibilities and Future Directions: An Integrative Review'. *HEALTH INFORMATION*
497 *AND LIBRARIES JOURNAL* 41 (1): 4–15. <https://doi.org/10.1111/hir.12518>.
- 498 Macdonald, C., D. Adeloje, A. Sheikh, and I. Rudan. 2023. 'Can ChatGPT Draft a Research
499 Article? An Example of Population-Level Vaccine Effectiveness Analysis'. *Journal of*
500 *Global Health* 13. <https://doi.org/10.7189/JOGH.13.01003>.
- 501 Odri, Guillaume-Anthony, and Diane Ji Yun Yoon. 2023. 'Detecting Generative Artificial
502 Intelligence in Scientific Articles: Evasion Techniques and Implications for Scientific
503 Integrity'. *Orthopaedics & Traumatology: Surgery & Research* 109 (8): 103706.
504 <https://doi.org/10.1016/j.otsr.2023.103706>.
- 505 Oyelude, Adetoun A. 2023. 'Much Ado about ChatGPT: Libraries and Librarians
506 Perspectives'. *Library Hi Tech News* 40 (3): 15–17. [https://doi.org/10.1108/LHTN-03-](https://doi.org/10.1108/LHTN-03-2023-0050)
507 [2023-0050](https://doi.org/10.1108/LHTN-03-2023-0050).
- 508 Peres, Renana, Martin Schreier, David Schweidel, and Alina Sorescu. 2023. 'On ChatGPT and
509 beyond: How Generative Artificial Intelligence May Affect Research, Teaching, and
510 Practice'. *International Journal of Research in Marketing* 40 (2): 269–75.
511 <https://doi.org/10.1016/j.ijresmar.2023.03.001>.
- 512 Qasem, Fawaz. 2023. 'ChatGPT in Scientific and Academic Research: Future Fears and
513 Reassurances'. *Library Hi Tech News* 40 (3): 30–32. [https://doi.org/10.1108/LHTN-03-](https://doi.org/10.1108/LHTN-03-2023-0043)
514 [2023-0043](https://doi.org/10.1108/LHTN-03-2023-0043).
- 515 Rahman, M.M., and Y. Watanobe. 2023. 'ChatGPT for Education and Research: Opportunities,
516 Threats, and Strategies'. *Applied Sciences (Switzerland)* 13 (9).
517 <https://doi.org/10.3390/app13095783>.
- 518 Roumeliotis, Konstantinos I., and Nikolaos D. Tselikas. 2023. 'Chatgpt and Open-Ai Models:
519 A Preliminary Review'. *Future Internet* 15 (6): 192.
- 520 Sadasivan, Vinu Sankar, Aounon Kumar, Sriram Balasubramanian, Wenxiao Wang, and Soheil
521 Feizi. 2023. 'Can AI-Generated Text Be Reliably Detected?' arXiv.
522 <https://doi.org/10.48550/ARXIV.2303.11156>.
- 523 Sallam, M. 2023. 'ChatGPT Utility in Healthcare Education, Research, and Practice:

- 524 Systematic Review on the Promising Perspectives and Valid Concerns’. *Healthcare*
525 (*Switzerland*) 11 (6). <https://doi.org/10.3390/healthcare11060887>.
- 526 Stokel-Walker, C. 2023. ‘ChatGPT Listed as Author on Research Papers: Many Scientists
527 Disapprove’. *Nature* 613 (7945): 620–21. [https://doi.org/10.1038/d41586-023-00107-](https://doi.org/10.1038/d41586-023-00107-z)
528 [z](https://doi.org/10.1038/d41586-023-00107-z).
- 529 Thelwall, Mike. 2024. ‘Can ChatGPT Evaluate Research Quality?’ *Journal of Data and*
530 *Information Science* 9 (2): 1–21. <https://doi.org/10.2478/jdis-2024-0013>.
- 531 Tian, Zhaoxia. 2021. ‘Application of Artificial Intelligence System in Libraries through Data
532 Mining and Content Filtering Methods’. *Journal of Physics: Conference Series* 1952
533 (4): 042091. <https://doi.org/10.1088/1742-6596/1952/4/042091>.
- 534 Torres-Salinas, Daniel, and Wenceslao Arroyo-Machado. 2024. *ChatGPT Manual:*
535 *Applications in Research and University Education*. Granada: InfluScience Editions.
536 <https://zenodo.org/records/11349799>.
- 537 United Nations Educational, Scientific and Cultural Organization. 2023. *ChatGPT and*
538 *Artificial Intelligence in Higher Education*. Venezuela: UNESCO.
- 539 Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez,
540 Łukasz Kaiser, and Illia Polosukhin. 2017. ‘Attention Is All You Need’. *Advances in*
541 *Neural Information Processing Systems* 30. [https://proceedings.neurips.cc/paper/7181-](https://proceedings.neurips.cc/paper/7181-attention-is-all)
542 [attention-is-all](https://proceedings.neurips.cc/paper/7181-attention-is-all).
- 543 White, Jules, Quchen Fu, Sam Hays, Michael Sandborn, Carlos Olea, Henry Gilbert, Ashraf
544 Elnashar, Jesse Spencer-Smith, and Douglas C. Schmidt. 2023. ‘A Prompt Pattern
545 Catalog to Enhance Prompt Engineering with ChatGPT’. arXiv.
546 <http://arxiv.org/abs/2302.11382>.
- 547 Whitfield, Sharon, and Melissa A. Hofmann. 2023. ‘Elicit: AI Literature Review Research
548 Assistant’. *Public Services Quarterly* 19 (3): 201–7.
549 <https://doi.org/10.1080/15228959.2023.2224125>.
- 550 Wu, Tianyu, Shizhu He, Jingping Liu, Siqi Sun, Kang Liu, Qing-Long Han, and Yang Tang.
551 2023. ‘A Brief Overview of ChatGPT: The History, Status Quo and Potential Future
552 Development’. *IEEE/CAA Journal of Automatica Sinica* 10 (5): 1122–36.
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573

574 **AUTHORS' BIOGRAPHIES**

575 **Daniel Torres Salinas** - *Department of Information and Communication, University of*
576 *Granada, Spain.*

577 DTS a Ph.D. in Scientific Documentation obtained in 2007. His professional journey includes
578 roles as a librarian and bookseller, alongside spearheading bibliometrics units at the University
579 of Navarra and the University of Granada until 2023. Currently, he lectures in the Department
580 of Information and Communication at the University of Granada and is an active member of
581 the EC3 Research Group, focusing on Evaluative Bibliometrics and altmetrics. With over 80
582 scientific contributions indexed in the Web of Science, his expertise is well-recognized. From
583 2013 to 2023, he served as CEO and now acts as a senior consultant at EC3metrics. He also
584 founded the YoSigoUGR teaching platform and the InluScience Project. Since 2023, he
585 coordinates Social Impact Assessment at UGR's Vice Chancellor's Office for Social Innovation,
586 Employability, and Entrepreneurship.

587
588 **Mike Thewall** - *Information School, University of Sheffield, Sheffield, UK.*

589 MT previously worked at the University of Wolverhampton from 1989 to 2023, where he taught
590 mathematics and statistics and conducted research in educational technology. In 2000, he
591 founded the Statistical Cybermetrics and Research Evaluation Group, which he led until his
592 departure for Sheffield. He is now Professor of Data Science at the Information School at the
593 University of Sheffield.

594
595 **Wenceslao Arroyo-Machado** - *School of Public Affairs, Arizona State University, USA*

596 WAM, holding a postdoctoral position at Arizona State University, is a Ph.D. in Information
597 and Communication Technologies from the University of Granada (UGR). His interdisciplinary
598 education combines sciencemetrics and data science, focusing on the analysis of science and its
599 interactions with society. WAM has an emerging research career with over 20 international
600 publications in scientometrics. He also holds the position of COO at EC3metrics, a spin-off
601 from the University of Granada, where he applies bibliometrics in a practical context for the
602 drafting of bibliometric reports and evaluation tasks. This synergy has enabled him to integrate
603 novel techniques into bibliometric reports and develop practical research methodologies,
604 showcasing his involvement with scientometrics across academic, teaching, and practical
605 domains.

606