frontiers | Frontiers in Education

TYPE Original Research
PUBLISHED 14 January 2025
DOI 10.3389/feduc.2024.1417900

OPEN ACCESS

EDITED BY
Robert Farrow,
The Open University, United Kingdom

REVIEWED BY
Mohamed Rafik Noor Mohamed Qureshi,
King Khalid University, Saudi Arabia
Yousef Wardat,
Yarmouk University, Jordan

*CORRESPONDENCE
Dimitri Prandner
✉ dimitri.prandner@jku.at

†These authors share first authorship

RECEIVED 16 April 2024
ACCEPTED 13 December 2024
PUBLISHED 14 January 2025

CITATION
Prandner D, Wetzelhütter D and Hese S (2025) ChatGPT as a data analyst: an exploratory study on AI-supported quantitative data analysis in empirical research.
Front. Educ. 9:1417900.
doi: 10.3389/feduc.2024.1417900

ChatGPT as a data analyst: an exploratory study on AI-supported quantitative data analysis in empirical research

Dimitri Prandner[1*†], Daniela Wetzelhütter[2†] and Sönke Hese[3]

[1]Johannes Kepler University, Institute of Sociology, Department of Empirical Social Research, Linz, Austria, [2]Department for Social Work, School of Medical Engineering and Applied Social Sciences, University of Applied Sciences Upper Austria, Linz, Austria, [3]Institute for Social Sciences / Sociology, Christian-Albrechts-Universität zu Kiel, Kiel, Germany

Social scientists are faced with the challenge of designing complex studies and analyzing collected data via various programs such as R, Stata, SPSS, or Python. This often requires the use of analytical procedures and specific software packages that are beyond an individual's established skillsets and technical knowledge. To address these challenges, generative artificial intelligence, such as ChatGPT, can now be employed as 'assistants'—with both associated risks and benefits. Accordingly, this paper explores the potential and pitfalls of using a tool like ChatGPT as an assistant in quantitative data analysis. We investigate the practical use of ChatGPT-3.5 by replicating analyses and findings in everyday scientific research. Unlike previous studies, which have primarily focused optimizing the use of chatbots for code generation, our approach examines an amateur level use of AI tools to support and reference regular research activities, with an emphasis on minimal technical expertise. While we overall conducted three experiments, with the goal to replicate academic papers, the article's focus is on the methodologically most complex one, by De Wet et al. from 2020. In this case AI is used for the step-by-step replication of the two-dimensional model of value types proposed by Schwartz (2012). The results of this experiment highlight the challenges of using ChatGPT 3.5 for specific, detailed tasks in academic research, as a tendency for responses to repeat in loops when solutions were not readily available emerged at several stages. Thus, we concluded that there are severe limitations in the AI's ability to provide accurate and comprehensive solutions for complex tasks and emphasize the need for caution and verification when using AI powered tools for complex research procedures.

KEYWORDS

ChatGPT, data analyst, AI-supported, quantitative, data analysis

1 Introduction

The ongoing digitization and digitalization of society is influencing empirical social research, which is undergoing a wide-reaching transformative process (Kritzinger et al., 2023). Consequently, the social sciences must adapt to a world where "digital technologies and AI in particular are developing very rapidly. The direction of such fast innovations needs to be steered socio-politically" (Cath et al., 2017). Following this assessment, it needs to be argued that digitalization has several consequences for social science research (Couldry and Powell, 2014; Kritzinger et al., 2023). Firstly, it has fostered an expansion of topics that require examination (see, e.g., Couldry and van Dijck 2015). Secondly, it has created new categories of empirical materials—such as digital trace data and digital behavioral data—that are available for social scientific research (see, e.g., Brady, 2019; Leitgöb et al., 2023). Thirdly, data analysis procedures are evolving, with new analytical techniques emerging and established ones being reexamined, increasing the need to be literate in an increasing variety software packages and their unique feature sets (Karaali, 2023).

For this paper we want to focus on the third issue raised. Today, social scientists who use quantitative methods for data analysis are often expected not only to design and operationalize highly complex studies but also to code in multiple software environments such as R, Stata, SPSS, or Python (Abbasnasab Sardareh et al., 2021; Manyika et al., 2011). While many social scientists are literate in one or two of these programs, inter-organizational collaborations, academic journal reviewers, specific analytical procedures, and sometimes even funding agencies require the use of specific software packages that may be outside an individual's established skill set.

As a result, some social scientists have shifted their priorities from working on substantive issues to time- and resource-intensive tasks tied to such technical aspects (Kim and Ng, 2022; Brooker, 2019). While thorough documentation, increasing interoperability and combinability of datasets, and clear procedural instructions offset some of this burden, it remains an issue for many social science collaborations because different researchers are accustomed to different software environments (Prandner et al., 2021, p. 38). Yet, digitalization has not only fostered these challenges, but also may have provided a potential solution as well. Recently several generative AI tools—such as ChatGPT, Google AI, Microsoft's Co-Pilot, and IBM Watson—have become available for general and scientific purposes.

Due to the flexibility and ease of use, it is no surprise that the launch of ChatGPT on November 30, 2022, changed the discourse surrounding the deployment of AI applications in many fields, including academia. Although various AI-powered tools such as DeepL and Google Lens were already widely used in academic and commercial applications (Bilyk et al., 2022; Bellés-Calvera and Caro Quintana, 2022; Grover et al., 2022), the public availability of ChatGPT, the trailblazer of generative AI, sparked extensive and profound discussions—including legal and ethical issues like academic integrity and plagiarism, as well as practical considerations (Cotton et al., 2023; Gabriel et al., 2024; Jahic et al., 2023, p. 1466). To control for these developments prevailing software like Turnitin or ithenticate started to include AI output detectors, with the aim to stop undeclared use or straight misuse of AI tools in academia (Gao et al., 2023). However, a clear, ethically undisputed application scenario emerged in the field of computer sciences, where the use of generative AI to create and debug code for various programming problems has become increasingly common, with promising results reported (Kalla and Smith, 2023; Nikolaidis et al., 2023). Gabriel et al. (2024, p. 179) provide information that the co-pilot AI from Microsoft is already responsible for 30% of the program code found on GitHub and the percentage is steeply rising, with human work more and more shifting towards ensuring that code remains interpretable and matching the given tasks (Gabriel et al., 2024, pp. 173). Overall, this hints that coders and programmers are shifting from code debugging and writing code in repetitive tasks to more creative processes (Gabriel et al., 2024, p. 180).

Building on these insights, our research aims to evaluate the potential of ChatGPT as a supporting tool for quantitative data analysis in the social sciences, specifically regarding its ability to generate code for statistical software packages and perform the corresponding statistical analysis. Our study, initiated in the summer of 2023, focuses on the freely available GPT-3.5 version of ChatGPT to assess whether it can assist individuals already familiar with data analysis procedures. Throughout this paper, when we refer to ChatGPT, we are exclusively referring to GPT-3.5, unless stated otherwise.

Like Keeling proposes in Gabriel et al. (2024, p. 12) we acknowledge that AI support can occur at different levels of skill and involvement—and that prompt engineering is a rapidly developing professional field—and we took the conscious decision to use ChatGPT from the perspective of a *naïve user* in this experiment and thus concentrate on their perspective. Using this as the starting point our article focuses on the inherent social technological and communication-based approach to AI (Gabriel et al., 2024, p. 22), where a *naïve user* is defined as someone who interacts with the chatbot using domain-specific terminology but in a natural, non-technical manner. This means that we decided to forgo prompt optimization and the use of a statistical software package that aligns with the known parameters that ChatGPT typically prefers. With this particular approach we seek to replicate the real-world use of the tool and evaluate its effectiveness when employed by *naïve user* without specialized training in prompt engineering for data analysis and without access to more advanced, commercial versions of the Generative Pre-trained Transformer (GPT) model (e.g., GPT-4o, GPT4 or even the coding focused GPT-4o with canvas). The range of tasks assigned to ChatGPT for this experiment extends from simple but common forms of data preparation (e.g., creating a mean-based index) to performing more complex analytical procedures (e.g., multidimensional scaling).

We assess the utility of ChatGPT based on the replicability of previously published results. To do so, we apply several evaluation criteria for the quality of analysis with ChatGPT, including the suitability of the analysis procedures, the functionality of the generated code, and the correctness of the presentation of the results. Therefore, we aim to gain insights into which steps of the code and output generation phases of data analysis show promise when it comes to AI assistance and what types of pitfalls may arise when a naïve user employs a free version of ChatGPT without specific prompt optimization during the interaction with the AI. With this we want to illustrate what safeguards and considerations may be necessary, when researchers want to employ AI to assist them during their work.

The article is divided into the following sections: a brief review concerning the use of generative AI for code generation and analysis in survey research; a description of the methodological approach used to replicate existing results using ChatGPT; the presentation of the evaluation results. A short summary of the results followed by a conclusion, including a paragraph on limitations, closes the contribution.

2 Cutting-edge generative AI for survey researchers: potentials for analysis and code generation?

The public release of ChatGPT in November 2022 marked a pivotal moment in the accessibility of conversational artificial intelligence (AI) for technically less-skilled users. To interact with this AI, users only need to provide an initial prompt, after which the AI generates human-like text or content based on the parameters set by the prompt. This may be followed up by further input from the user and through iterative dialogue, where users refine content through further prompts, the AI can produce a variety of materials, predominantly text, mimicking a natural conversation.

While OpenAI's ChatGPT is just one of many generative AI models, others include, e.g., Co-Pilot, Gemini, and Apple Intelligence, it is widely regarded as the best-performing model available to the public (Calonge et al., 2023). In addition to its conversational capabilities, ChatGPT has demonstrated the ability to solve complex mathematical and statistical problems [see, e.g., Calonge et al., 2023;

differences between the free GPT-3.5 model and the paid GPT-4 version have been documented, with the latter showing enhanced capabilities (Zhang et al., 2023)].

The underlying technology behind ChatGPT is based on *Generative Pre-trained Transformer* (GPT) models, which are initially trained in a general manner before being fine-tuned for specific tasks. The term "Transformer" refers to the model's architectural framework, as introduced by Vaswani et al. (2017). Although the training data for ChatGPT (GPT-3.5 and GPT-4) remains proprietary, it is known that GPT-3 was trained on a dataset of approximately 570 GB, equivalent to about 300 billion words, drawn from various sources such as a curated subset of Common Crawl, literary works, and Wikipedia (Brown et al., 2020). This pre-training was followed by fine-tuning for specific applications, such as generating conversational responses (Radford et al., 2018). Human feedback was incorporated to improve the model's accuracy and reduce harmful outputs.

Due to this training process, the output generated by ChatGPT reflects both the strengths and bias of its training data (Ray, 2023). For example, its proficiency in different languages correlates with the amount of training data available in each language; languages with more data yield better results (Lai et al., 2023). While such linguistic bias are transparent, other forms of bias—such as social or political bias—are more subtle but still present and reproduced in the model (Rozado, 2023). This phenomenon is further exacerbated when information less prevalent in digital materials is underrepresented, leading to reduced accuracy when discussing such topics. Although fine-tuning can mitigate some of these biases, eliminating them entirely remains unfeasible due to the model's scale and scope.

The issue of bias also bleeds into the discussion concerning the application of AI in the social sciences, however in field's current focus is predominantly the impact of AI on academic publishing and writing. Although AI can produce convincing texts, closer examination often reveals subtle or significant errors and pushes additional issues like plagiarism and authorship to the forefront. References, in particular, are prone to "hallucinations," where the AI generates plausible but false content (Farhat et al., 2023). Consequently, experts recommend that AI-generated text should undergo critical review and be scrutinized by field specialists (Salvagno et al., 2023; AlZaabi et al., 2023; Hong, 2023; Rahman and Watanobe, 2023; Zheng and Zhan, 2023; Sallam, 2023; Van Dis et al., 2023; Cooper, 2023; Malinka et al., 2023). Furthermore, publishers – including Frontiers – have reached a preliminary consensus that AI cannot be listed as an author (Stokel-Walker, 2023).

This is also true when it comes to assessing the effect of generative AI on the evaluation of junior researchers and students, as it can be used to generate summaries or synthesize literature on specific topics, raising again issues of plagiarism, cheating, and assessment. The consensus is that while ChatGPT can be a useful tool for rephrasing or summarizing texts, future assignments may need to evolve, and ethical guidelines should be explicitly addressed (Rahman and Watanobe, 2023). However, recent studies indicate that students using ChatGPT for class-related tasks do not necessarily perform better in written assignments than those who do not use AI (Bašić et al., 2023). This underscores the notion that the usefulness of ChatGPT is limited when users lack the expertise to perform quality checks.

However, research that evaluates how well ChatGPT performs when it comes to support task concerning writing code for statistical software packages, like, e.g., Hansson and Ellréus (2023), Kalla and Smith (2023) or Nikolaidis et al. (2023) researched for the computer sciences is still scarce and the challenge remains in how to assess its use as a tool. Accordingly, we want to evaluate how ChatGPT performs in generating code for statistical analysis software, acknowledging the following:

Firstly, just as English dominates the digital space compared to languages like Hungarian or Danish, open-source software documentation (e.g., R code) is expected to be more accessible than documentation for proprietary software like SPSS or STATA, which are commercial. Although the exact training data for ChatGPT is undisclosed, sources like Common Crawl reference platforms such as GitHub and Stack Overflow, which are rich in open-source coding languages. This suggests a potential bias towards open-source software. Hansson and Ellréus (2023) concluded that ChatGPT performs well in generating and debugging code for open-source software, while other studies indicate limitations when using proprietary languages (Kalla and Smith, 2023; Nikolaidis et al., 2023). Thus, we hypothesize that ChatGPT may face similar challenges in generating functional SPSS syntax or STATA commands, particularly for complex tasks that require advanced prompting skills beyond the basic *naïve* use described earlier.

Secondly, another challenge of using ChatGPT in research is its proprietary nature and the constant updates to its versions. Combined with the non-deterministic outputs of large language models (Ouyang et al., 2023), this makes it difficult to reproduce or verify work involving ChatGPT. Responses to identical queries may vary upon re-generation, and updates to the training data can further complicate replication or may result in some form of loops or repetitions (Farhat et al., 2023).

3 Objective of the study and related methodical approach

We explore the practical applicability of ChatGPT-3.5 based on the reproduction of analyses and findings with help of ChatGPT compared to those generated by researchers themselves. The comparison between ChatGPT's solutions and those of researchers was the basis for judging the output (right vs. wrong). This exploration is based on an everyday scientific application – the routine use of artificial intelligence tools as supporting resources and reference points in regular research activities, without strict adherence to more systematic methods or rules for efficient use of chatbots and requiring minimal technical expertise to operate it. Thus, we differentiate in our approach, in contrast to Henrickson and Meroño-Peñuela (2023), who did research on the optimization of prompt engineering, or Meyer et al. (2023), who focused on limitations and usefulness for academic writing and teaching. This is important as it can help improve academic work, increase replicability and ease the burden on scientists when it comes to code generation and code-maintenance.

Our original experiment was based on the reproduction of three different academic papers, with different levels of complexity. The chat logs and results of all three parts of the experiment can be found online.[1] However, in this article, we want to focus on the most complex one: A previously published article by De Wet et al. (2020). We tried to replicate the analysis and generation of the presentation of the results, based on the code provided by the original authors who worked with the commercial software package SPSS, from IBM, which is common on the social sciences but not open access and thus less likely to be included in

1 See: https://www.researchgate.net/publication/374144911_AI-Enabled_Data_Analysis_Quality_Addressing_A_Knowledge_Gap;

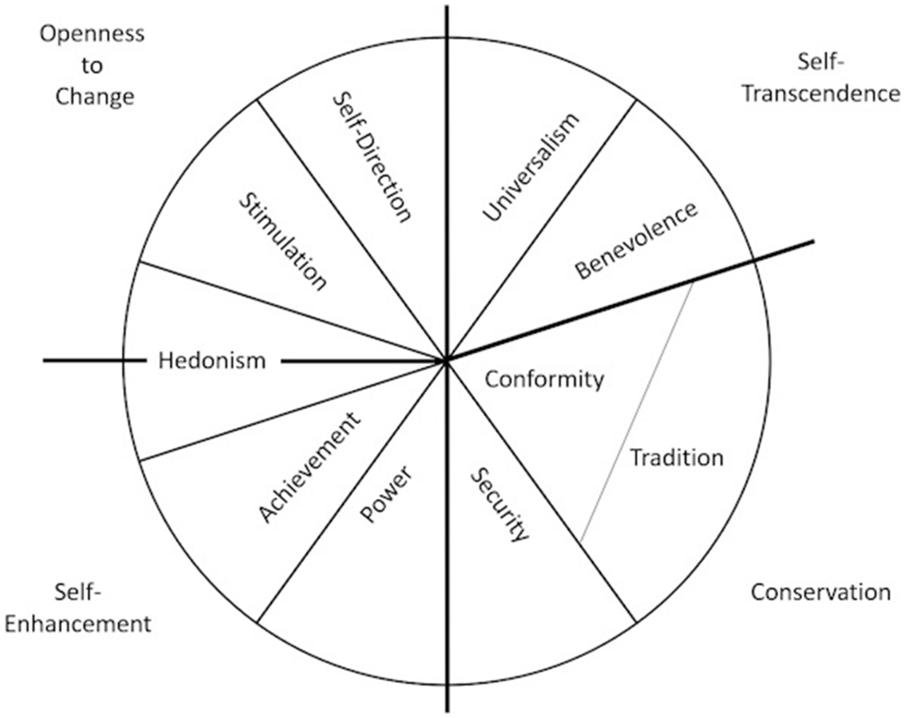

FIGURE 1
Own depiction of Schwartz's two dimensional values model, based on Schwartz (2012, p. 9).

training materials compared to code for more open software solutions for statistical analysis like Python or R. The overall aim was to reproduce the visualization of Schwartz's (2009, 2012) two-dimensional model of motivational value types and higher-order value domains (see Figure 1) based on empirical data collected with Schwartz's PVQ-21, as it was done by De Wet et al. (2020) (see Figure 2), using ChatGPT-3.5. Since values influence political behavior (Inglehart and Welzel, 2005), this is a central topic of political science and of interest to many in the broader social sciences, e.g., education, politics or communication. Although this article is only indirectly about values, we will use it as an example to demonstrate the possibilities of AI (see Figure 3).

We have chosen this complex example as the central case for this article because it requires various simple (e.g., generating a syntax to calculate a mean value index) to very complex steps (e.g., generating a syntax for a specific method of multi-dimensional scaling) in the analysis process, helping us illustrate the different ways generative AI can be used to support and assist researchers.

We used the following procedure for the use of ChatGPT:

1) Formulation of prompts for the
   i) selection of an analysis method
   ii) generation of syntax code
   iii) presentation of results.

2) Evaluation of the respective response from ChatGPT: Our evaluation follows relevant quality criteria, drawing on established frameworks by Pallant (2023), Starr (2006), and Sada et al. (2007). Three main aspects are central to our assessment:
   o Appropriateness of the proposed method of analysis: This is an assessment of how suitable the analysis method proposed by ChatGPT-3.5 is for the task.
   o Functionality of the generated syntax code: We evaluate the operational effectiveness of the syntax code generated by ChatGPT-3.5, including its ability to run without errors and produce meaningful output.
   o Accuracy of results presentation: The accuracy of the results presented by ChatGPT-3.5 will be examined.
3) Formulation of follow-up prompts: concretization of the prompt to achieve the goal.
4) Repeating Steps 2 and 3 as long as a solution is in sight.

In step 4, we found that subsequent responses to requests (prompts) for the same problem could become very similar or recurring, creating a kind of "loop." In this case, we stopped the dialogue.

Based on this, we describe the findings of reproducing several results of the paper using ChatGPT-3.5 and present the lessons learned. Results have to be read in accordance with the fact that the experiment took place in August 2023, and thus, results may vary at other points in time.

4 Results

The first task we asked ChatGPT[2] to complete was the selection of an appropriate analysis model. This goes beyond simple code creation

2 Prompt no. 1: "I have collected data using Schwartz's PVQ-21 and would now like to test whether I can use it to test the theoretical model (Theory of Basic Values, see, e.g., Schwartz 1992, 1994; Cieciuch et al., 2013, p. 1216) with my empirical data. I would like to test it in a similar way to Bilsky et al. (2011). What should I do?" (see chat log, https://chat.openai.com/share/e1875e07-a8ab-40d4-a337-3144760c3415).

but is a necessary set up for all the following parts of the analysis. We formulated our prompt based on the above-mentioned objective; the reproduction of the visualization of Schwartz's (2009, 2012) two-dimensional model of motivational value types and higher-order value domains on the basis of empirical data collected with Schwartz's PVQ-21 data collection instrument (see Figures 1, 2). The wording of the prompt used follows the perspective of a naïve user, who is using ChatGPT3.5 as a supportive co-pilot during analysis and engaging with it in a natural dialogue form. This form of communication is not only accessible to users of any skill level, but also a key feature promoted by OpenAI. Thus, we did not expect optimal outcomes with this approach and acknowledge that professionals with high skill levels in prompt engineering may generate much better results. This type of approach was used for all follow-up prompts during the dialogue as well.

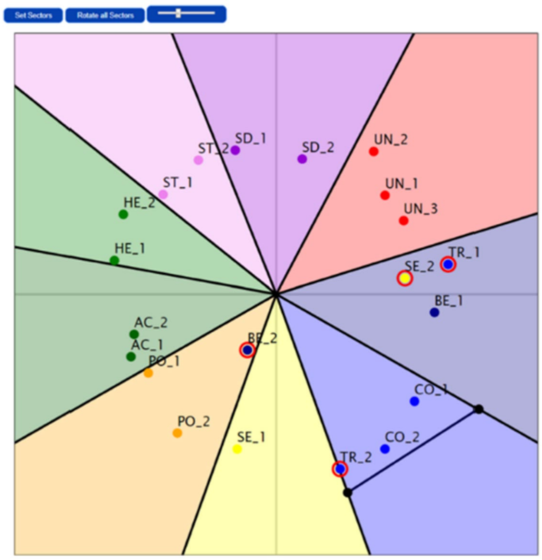

FIGURE 2
Graph of optimal empirical solution Source: De Wet et al. (2020).

The first prompt (no. 1) was followed by a dialogue based on "follow-up prompts" (no. 1.1–1.5) since the proposals were either very general (e.g., along the scientific research process) or included a wide variety of procedures until finally two procedures could be narrowed down (see Figure 3, as well as the example for prompt 1.5: "I meant: would they do PCA or MDS to visualize the objects in a two dimensional space like the theoretical model"). Furthermore, the explanations on the implementation of the proposed procedure were incorrect and not realizable (answer to prompt 1.2).

However, the final decision was not made by ChatGPT – it was left to the user. As a result, we made the decision in favor of the MDS ourselves.

As there is no such thing as "the MDS," but the MDS must be specified, we asked specifically about this in a second step. As before, the first prompt was followed by a dialogue based on "follow-up prompts," as the proposals were very general (in terms of key considerations) but included explanations for deciding by oneself. However, they remained superficial, which is why the dialogue was closed (see Table 1).

In total, we formulated 11 prompts for this first task "selecting an appropriate method of analysis." Based on our experiences, the responses tend to repeat in an "endless loop" when no solution is found, but similar inquiries are made repeatedly. Therefore,

ChatGPT: Schwartz, Cieciuch, and Bilsky, as researchers in the field of values and proponents of Schwartz's Theory of Basic Values, might use Multidimensional Scaling (MDS) or Principal Component Analysis (PCA) as data reduction techniques to visualize objects (e.g., values or respondents) in a two-dimensional space in certain situations. While these techniques are not the primary focus of their research, they can be valuable for data exploration and visualization. The choice between PCA and MDS depends on their specific research objectives and data.

1.Multidimensional Scaling (MDS):

1. MDS is typically used when the primary goal is to visualize the similarities or dissimilarities between objects in a lower-dimensional space while preserving the original relationships. If researchers are interested in understanding how values or respondents are situated in a two-dimensional space based on their similarities or differences, MDS can be a suitable technique.

2.Principal Component Analysis (PCA):

1. PCA is more often used for dimensionality reduction and to identify the main components (dimensions) that explain the variance in the data. While it is not primarily designed for visualizing objects in a two-dimensional space, PCA can be used to reduce data complexity and potentially provide insight into the underlying structure of the data.

The choice between MDS and PCA for visualizing objects in a two-dimensional space depends on the specific goals of the research and the nature of the data. MDS is typically more appropriate when the primary focus is on spatial relationships and similarity between objects. PCA, on the other hand, is suitable for understanding the underlying structure of the data in terms of dimensions that capture the most variance. ChatGPT "only" provides information for the decision
In the context of values research, if the goal is to create a visual representation of the relationships between different values or respondents, **MDS might be a more appropriate choice**. However, it's essential to consider the specific research objectives and data characteristics before selecting the most suitable technique for visualization.

FIGURE 3
Excerpt from Chatlog "selecting analysis procedure". See: https://chat.openai.com/share/e1875e07-a8ab-40d4-a337-3144760c3415.

TABLE 1 Summary: task "selecting an appropriate method of analysis"—prompts and answers.

Prompt No.	Aim:	Result	Evaluation
1	Recommendation for analysis method: MDS (multidimensional scaling)	General Approach: 12 Steps 1) Literature review 2) Hypothesis deduction … 12) Publication	MDS not part of the solution.
1.1	Specification: analyses method in SPSS	Different Methods: 13 procedures 1) Descriptive statistics 2) Reliability analyses … 13) Reporting and visualization	MDS not part of the solution
1.2	Specification: visualization in a two-dimensional space	Two different techniques: MDS or PCA—but incorrect instructions.	MDS is part of the solution
1.3	Specification: procedure analogue to experts	Different methods: 9 procedures	MDS not part of the solution
1.4	Specification: Decision in favor of MDS or PCA	PCA, MDS are useful but less common for reproducing dimensionality.	MDS not part of the solution
1.5	Specification: Decision in favor of MDS vs. PCA for visualization in two-dimensional space.	Solution: Information about the two different procedures as a basis for the decision	MDS is part of the solution
Overall:	2 out of 6 responses contained information for decision making (for/against MDS). The explanations of applying the analysis procedures were incorrect. The dialogue was closed as there seemed to be no solution in sight.		
2	Recommendations for specification of the analysis method (similar to Schwartz, …): MDS/SSA • Two-dimensional • Similarity measure (correlation matrix based on mean centered PVQ-Items) • Theory based starting values	9 Key Considerations: Choice of • Similarity/Dissimilarity measures: […] • Scaling and metric: […] • Number of dimensions: […] • Labelling and visualization: […] • …	Part of the solution: • Replicate the number of dimensions. • Use same/similar similarity/dissimilarity measure • -
2.1	Specification: special MDS-technique	8 Key Considerations: • Similarity/Dissimilarity measures: […] • Number of dimensions: […] 2D and 3D • … […]	Part of the solution: • Similarity/Dissimilarity measure • 2–3 dimensions • …
2.2	Specification: decision between similarity and dissimilarity measure	Description of similarity and dissimilarity measures: e.g., • Euclidean Distance, Pearson Correlation, Cosine Similarity, Jaccard Index and Squared Euclidean Distance […] Basis for decision: Nature of data, research objectives, theoretical framework, data transformation, validation, experimentation […]	Part of the solution: • Pearson Correlation
2.3	Specification: decision analogue to experts	Similar to answer to prompt 1.2	Part of the solution: • Similarity measure
2.4	Specification: aim of the analyses (represent empirical measured values in 2-dim. Space)	Description of similarity measures: • Pearson Correlation: […] • Cosine Similarity: […] • Jaccard Index: […] • Correlation Dissimilarity: […]	Part of the solution: • Pearson Correlation
Overall:	5 out of 5 answers provided a basis for decisions on the choice of particular specifications (dimensionality and measure but not for the starting values). However, no solution could be generated. The explanations remained superficial. The dialogue was closed as there seemed to be no solution in sight.		

See chat log: https://chat.openai.com/share/e1875e07-a8ab-40d4-a337-3144760c3415.

we stopped the dialogue in both examples mentioned below because no solution was in sight. The responses were articulated in an understandable manner. However, they only provided help, that resulted in partial solutions and had gaps when it came to understanding how decisions should be made. Therefore, this task was not solved and naïve users would have a problem concerning how they may proceed correctly. The following table summarizes the purpose of the respective prompt (aim), the corresponding answer (result), and our assessment (evaluation).

In a second task, we tried to create the actual syntax for SPSS using ChatGPT in order to perform the MDS as described by De Wet et al. (2020), who also used the commercial software for their analysis. As before, the initial prompt[3] was followed by a dialogue based on "follow-up prompts," as the first answer was incorrect (see Table 2, prompt 1–1.1).

After the attempt at correction failed, we switched to a step-by-step approach – first by aiming to generate a syntax for creating the initial configuration of the theory-based initial values (which is needed for performing the MDS). The following figures show the solution presented in the cited paper (Figure 4) and the proposal generated by ChatGPT (prompt 1.2) (Figure 5). Using the syntax-code shows that except for two small coding errors, the ChatGPT-Syntax code works well. However, the bugs could not be fixed by Chat-GPT, but experienced researchers can fix them (see yellow marking in the following figure) with minimal effort. Which leads to the issue of the naïve user once more, was there is now a skill gap that becomes evident: Naïve use of ChatGPT may be somewhat helpful if there is a basic understanding of the tools used for data analysis, but not without such a basic understanding.

In the second step, we tried to generate a script for the mean centering of the PVQ items (see Table 2, prompt 2–2.2). While the ChatGPT-generated syntax for mean centering worked well, the rest was unusable (see chat excerpt below). The generated syntax contained commands that are "atypical" for SPSS but are used in other software—e.g., "Loop," and "Do Repeat" – suggesting that ChatGPT compiled the syntax by drawing from different programming languages. This outcome results in a hard to solve challenge even for experienced researchers. Now the research not only needs to understand what the ChatGPT proposes on a content level, but they also need to identify what the proposed command does in other, similar programming languages. And if they manage to do both, they still need to figure out how to translate the given information into the syntax used by the statistical software package of their choice.

Example prompt 2.2: Now, I get this error message: 6 COMPUTE The function appearing on the left side of the assignment operator (equals sign) is not valid in that context.

In the third step, we tried to generate a syntax for the MDS/SSA for SPSS, as described in Table 2 (see prompt 3). Zero out of three answers provided usable syntax-code. The dialog was closed because there seemed to be no solution in sight.

Overall, for all steps dealing with this second task, we formulated 10 prompts from the perspective of a naïve users, that hoped to "generate an SPSS-Syntax for MDS/SSA." As before, we stopped the respective dialogue when it became likely that no solution could be reached. The responses, while written in an understandable manner, were just partly correct. Therefore, this task was not solved. The table below summarizes the steps and results.

Thirdly, we tried to create a syntax code using R for the visualization of empirical coordinates as the reproduction of the two-dimensional value model (see Figure 1). The decision to switch to R for this step was tied to the fact that the original authors of De Wet et al. (2020) used a self-designed software solution, without enough documentation to be viable for discussion with ChatGPT. We were able to visualize the coordinates for the theoretical model and the empirical data with the first prompt.[4] However, the first solution was difficult to verify because the labeling of the data points was missing (see Figure 6). The second solution shows that either the data points of the theoretical model or the empirical data had to be shifted so that the data points of a value dimension (e.g., theoretical and empirical data points for Hedonism) start next to each other – otherwise, the deviations cannot be verified. This means that for validation, either the empirically observed or theoretically assumed data points had to be rotated slightly clockwise or anticlockwise. However, this step proved to be a major challenge,

3 Prompt no. 1:"I would like to use the MDS/SSA analysis as described in the paper" "Standardisation of the reproduction of Schwartz two-dimensional value space using multidimensional scaling and Goodness-of-Fit test procedure" by Jacques de Wet, Daniela Wetzelhütter & Johann Bacher reproduce described. I would like to do it step by step in SPSS—i.e. first save the start configuration in a spss-file: sav. The start configuration is: SD_1m −0.34 0.94 PO_1m 0.00−1.00 UN_1m 0.34 0.94 AC_1m −0.64−0.77 SE_1m 0.64−0.77 ST_1m −0.87 0.50 CO_1m 0.49−0.09 UN_2m 0.34 0.94 TR_1m 0.98−0.17 HE_1m −0.98−0.17 SD_2m −0.34 0.94 BE_1m 0.87 0.50 AC_2m −0.64−0.77 SE_2m 0.64−0.77 ST_2m −0.87 0.50 CO_2m 0.49−0.09 PO_2m 0.00−1.00 BE_2m 0.87 0.50 UN_3m 0.34 0.94 TR_2m 0.98−0.17 HE_2m −0.98−0.17 I have the data set - there are 21 PVQ variables. SD_1 SD_2 UN_2 UN_1 UN_3 BE_2 BE_1 TR_1 TR_2 CO_1 CO_2 SE_1 SE_2 PO_1 PO_2 AC_1 AC_2 HE_1 HE_2 ST_1 ST_2 Would you write me the necessary spss-syntax?" (see chat log, https://chat.openai.com/share/ba1df9a3-93c7-44be-98a7-05ef855a305b).

4 "I have coordinates available. One is a theoretical model (Schwartz value model) with the following coordinates: Schwartz's Starting Configuration dim1 dim2 SD_1−0.34 0.94 PO_1 0−1 UN_1 0.34 0.94 AC_1−0.64 -0.77 SE_1 0.64−0.77 ST_1−0.87 0.5 CO_1 0.49−0.09 UN_2 0.34 0.94 TR_1 0.98−0.17 HE_1−0.98 -0.17 SD_2−0.34 0.94 BE_1 0.87 0.5 AC_2−0.64 -0.77 SE_2 0.64−0.77 ST_2−0.87 0.5 CO_2 0.49−0.09 PO_2 0−1 BE_2 0.87 0.5 UN_3 0.34 0.94 TR_2 0.98−0.17 HE_2−0.98 -0.17 And once empirical coordinates generated on the basis of the PVQ-21 (by Schwartz). Final coordinates Dimension 1 2 SD_1m – 0.173 0.607 PO_1m – 0.541 -0.332 UN_1m 0.460 0.416 AC_1m – 0.614 -0.264 SE_1m – 0.165 -0.653 ST_1m – 0.478 0.420 CO_1m 0.584−0.452 UN_2m 0.412 0.602 TR_1m 0.726 0.125 HE_1m – 0.684 0.141 SD_2m 0.110 0.570 BE_1m 0.668−0.076 AC_2m – 0.600 -0.169 SE_2m 0.544 0.067 ST_2m – 0.329 0.565 CO_2m 0.459−0.653 PO_2m – 0.418 -0.586 BE_2m – 0.122 -0.235 UN_3m 0.538 0.310 TR_2m 0.269−0.738 HE_2m – 0.646 0.336 I would like to display the data points for both in a two-dimensional space (circular visualisation) in order to compare whether the empirical data deviate strongly from the starting configuration. I would like to use R-Study for this - can you write me a syntax for it?" (see chat log https://chat.openai.com/share/ff1d00b7-69aa-4f9d-a92e-a88236701db7).

TABLE 2 Summary: task "compute a syntax-code"—prompts and answers.

Prompt no.	Aim	Result	Evaluation
1	Generation of an SPSS syntax for MDS/ SSA using: • 2 Dimensions • theory based initial values • similarity measure based on a correlation matrix using mean centered items	Extensive but incorrect SPSS syntax to: • define a start configuration • load a dataset • perform a mds	Syntax for • the starting configuration: promising, as there were only small errors • The rest: unusable
1.1	Correction of the syntax: • based on the error-message of SPSS	A procedure for checking the data was suggested.	The error message was misunderstood.
1.2	Specification: • correction of the erroneous syntax for computing the start configuration	Unnecessary changes were made to the syntax – it was still incorrect.	Syntax for the starting configuration: small errors.
1.3	Specification: correction of the error	A procedure for checking the data was suggested.	The error message was misunderstood.
Overall:	4 out of 4 answers provided either an incorrect syntax or a procedure for correction that was not needed. However, no solution could be generated. The dialogue was closed as there seemed to be no solution in sight.		
2	Generation of an SPSS syntax for computing • mean centered items	Extensive but incorrect SPSS syntax in order to: • define data list • mean centering • repeat mean centering • cleaning data set	Syntax for • mean centering: promising The rest: unusable
2.1	Correction of the syntax: Based on the error-message of SPSS	Incorrect SPSS syntax for: • mean centering • other steps	Syntax for mean centering: promising, but just partly correct The rest: unusable
2.2	Correction of the syntax: Based on the error-message of SPSS	Correct SPSS syntax for: • mean centering Incorrect SPSS syntax for: • other steps	Syntax for mean centering: correct The rest: unusable
Overall:	1 out of 3 answers provided a partly correct syntax. The user has to find the right part. Himself. The dialogue was closed as there seemed to be no solution in sight.		
3	Generation of an SPSS syntax for computing • MDS/SSA	Extensive but incorrect SPSS syntax in order to: • Calculate proximities • Using the starting configuration • Perform Proxscal • Close data sets	Syntax is incorrect – not a part of it is usable.
3.1	Correction of the syntax: based on • The order of the steps to be considered, • The SPSS command "Proxscal."	Incorrect SPSS syntax for: • Calculate proximities • Performing Proxscal	Syntax is incorrect – the commands are not comprehensible.
3.2	Correction of the syntax based on a similar description as in 1.1	Incorrect SPSS syntax for: • Calculate proximities • Performing Proxscal	Syntax is incorrect – the commands are not comprehensible.
Overall:	0 out of 3 answers provided a partly correct syntax. The dialogue was closed as there seemed to be no solution in sight.		

See chat log: https://chat.openai.com/share/ba1df9a3-93c7-44be-98a7-05ef855a305b.

Paper

The syntax for the starting configuration is as follows:

```
*Define your Working Directory.
cd "E:\JdWS2D".

set DECIMAL = dot.
*Definition of starting configuration.
data list free/item (a5) dim1 dim2.
begin data
SD_1m  -0.34   0.94
PO_1m   0.00  -1.00
UN_1m   0.34   0.94
AC_1m  -0.64  -0.77
SE_1m   0.64  -0.77
ST_1m  -0.87   0.50
CO_1m   0.49  -0.09
UN_2m   0.34   0.94
TR_1m   0.98  -0.17
HE_1m  -0.98  -0.17
SD_2m  -0.34   0.94
BE_1m   0.87   0.5
AC_2m  -0.64  -0.77
SE_2m   0.64  -0.77
ST_2m  -0.87   0.50
CO_2m   0.49  -0.09
PO_2m   0.00  -1.00
BE_2m   0.87   0.50
UN_3m   0.34   0.94
TR_2m   0.98  -0.17
HE_2m  -0.98  -0.17
end data.


*Save the starting configuration.
save outfile="startconfig.sav".
execute.
```

FIGURE 4
Syntax codes for starting configuration (paper). Source: De Wet (2020).

ChatGPT

```
* Set the working directory to the folder where your data and SPSS syntax
files are located.
CD 'your_directory_path_here'.

* Define the start configuration values.
DATA LIST FREE / Item Start_x Start_y.
BEGIN DATA.
SD_1m -.34 .94
PO_1m .00 -1.00
UN_1m .34 .94
AC_1m -.64 -.77
SE_1m .64 -.77
ST_1m -.87 .50
CO_1m .49 -.09
UN_2m .34 .94
TR_1m .98 -.17
HE_1m -.98 -.17
SD_2m -.34 .94
BE_1m .87 .50
AC_2m -.64 -.77
SE_2m .64 -.77
ST_2m -.87 .50
CO_2m .49 -.09
PO_2m .00 -1.00
BE_2m .87 .50
UN_3m .34 .94
TR_2m .98 -.17
HE_2m -.98 -.17
END DATA.
```

FIGURE 5
Syntax codes for starting configuration (ChatGPT). Chat log: https://chat.openai.com/share/ba1df9a3-93c7-44be-98a7-05ef855a305b.

as shown in the following examples in Figures 7–10. In total, a wide range of different—but not correct—solutions were the result of different proposals to explain this deliberate shift. Below are several different incorrect solutions and the final (answer to the 23rd prompt) solution in Figure 11.

The final solution works well. It is not perfect—the ranges are not shown, and outliers have to be marked. Nevertheless, it is a suitable, cost-free approach to depict the empirical data, before processing it yourself. Only the reproducibility is questionable, but this is not a major drawback.

5 Summary of results

Overall, the results indicate promising potential for using freely available generative AI tools, such as ChatGPT, to assist naïve users during data analysis. However, our findings also reveal challenges, particularly in selecting the appropriate research procedure, which proved to be cumbersome. At the conclusion of the chat with ChatGPT for the first task, two approaches remained under consideration. Unfortunately, these approaches were not illustrated in a clear or complete manner, making it difficult for a naïve user to choose between them. Additional reading or external sources might have guided users to the correct choice of MDS, but this reliance highlights a key limitation.

The second task, which involved the commercial software SPSS, was also unsuccessful. While ChatGPT was able to generate basic SPSS syntax, it quickly became confused. The AI provided code fragments that were tangentially related to the problem but did not align with the commands required for SPSS operations. This presents a significant barrier for naïve users with limited skills in statistics and statistical software. Skilled users might be able to adapt the code and procedures with the help of external resources, such as manuals or command descriptions. However, it is unlikely that naïve users could leverage the information provided by ChatGPT to arrive at a satisfactory or functional solution. The methodological and technical knowledge required to correct the AI-generated solutions is too extensive for these users.

Finally, the third task offered some fascinating insights. Despite being the most complex task, it relied on an open software solution – R-which yielded better outcomes. The initial prompt generated a working solution (see Figure 7), and with several tweaks, a well-functioning solution was achieved (see Figures 8–11 for a step by step improvement). While not perfect, it produced a fitting visualization. However, reproducibility and documentation remain significant challenges when using this approach. The results of the third task suggest that ChatGPT performs better with open software, which is often better documented and more accessible, compared to proprietary solutions.

6 Conclusion

Increasing digitization and associated challenges are transforming empirical social research (Brady, 2019). The introduction of generative AI tools such as ChatGPT has opened up new possibilities, going far beyond questions of literature processing or academic dishonesty (Stokel-Walker, 2023; Cooper, 2023). ChatGPT can help select analysis procedures, generate syntax in different scenarios and even coding languages, and assist in the presentation of results. Compared to other usage scenarios like using AI in the writing of a text, using it for evaluative purposes etc. this step.

```
[...]
* Compute the mean for each respondent.
COMPUTE Mean_PVQ = MEAN(SD_1, SD_2, UN_2, UN_1, UN_3, BE_2, BE_1, TR_1, TR_2, CO_1, CO_2, SE_1, SE_2, PO_1, PO_2,
AC_1, AC_2, HE_1, HE_2, ST_1, ST_2).

* Create a loop to mean-center each PVQ item for each respondent.
VECTOR PVQ_items = SD_1 TO ST_2.
LOOP #i = 1 TO 20.
  COMPUTE Mean_Centered_PVQ_items(#i) = PVQ_items(#i) - Mean_PVQ.
END LOOP.

* Clean up the dataset by dropping the original PVQ items.
EXECUTE.

* Save the dataset with mean-centered PVQ items.
SAVE OUTFILE = 'mean_centered_pvq.sav'.
```

FIGURE 6
Excerpt from Chatlog "generate a syntax-code". See chat log: https://chat.openai.com/share/ba1df9a3-93c7-44be-98a7-05ef855a305b

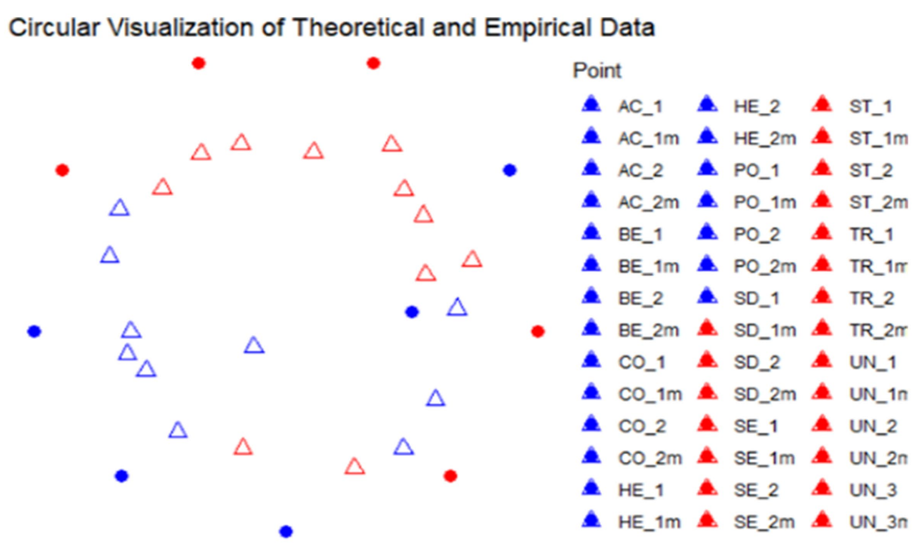

FIGURE 7
Result generated with prompt. See chat log: https://chat.openai.com/share/ff1d00b7-69aa-4f9d-a92e-a88236701db7.

This article examines how useful or risky this is under the assumption that a naïve AI user is employing it for their work. It is based on criteria proposed by authors like Pallant (2023), Starr (2006), and Sada et al. (2007): the suitability of the analysis methods, the functionality of the generated syntax code, and the correctness of the presentation of the results.

To replicate an analysis and presentation of the results of a previously published article, we used ChatGPT-3.5 and followed a multi-step process: 1. Formulating prompts, 2. evaluating ChatGPT's responses, 3. formulating follow-up prompts, and repeating steps 2 and 3 until a solution was reached or the dialogue devolved into an endless loop. The whole process was designed with the assumption that the provided AI tools are employed by a naïve user, who is proficient in his scientific discipline but not necessary in coding. This is a problem that is common for everyday research empirical social scientists, but as a research design it has several limitations that need to be considered:

Firstly, this research is based on a singular case study, which made some limiting assumptions: it is based on the premise of a naïve users, future research could propose different user profiles; including one that is only using optimized prompts. Secondly, it is based on a case study, where the original researchers were using proprietary software instead of an open one, both the literature and the results of the third task in our analysis hint at the fact that it may be a good idea to not replicate existing studies but build new, software agnostic case studies and compare them for further insights. Thirdly, as only ChatGPT3.5 was used for the experiment the results are also limited. More powerful AI may perform differently. Furthermore, the choice of prompts and follow-up prompts was made individually (user dependent). Moreover, a full quantification of prompts leading to

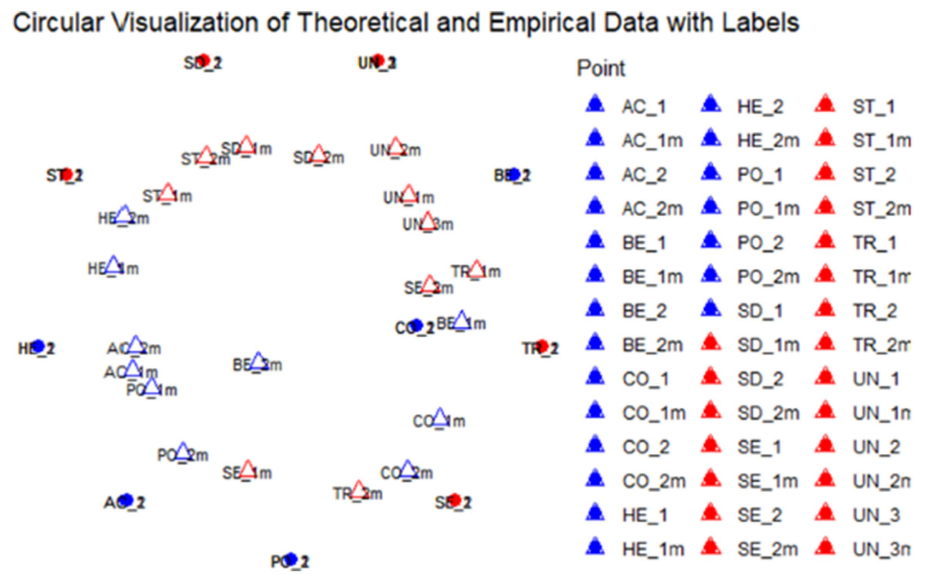

FIGURE 8
Result generated with prompt 2. See chat log: https://chat.openai.com/share/ff1d00b7-69aa-4f9d-a92e-a88236701db7.

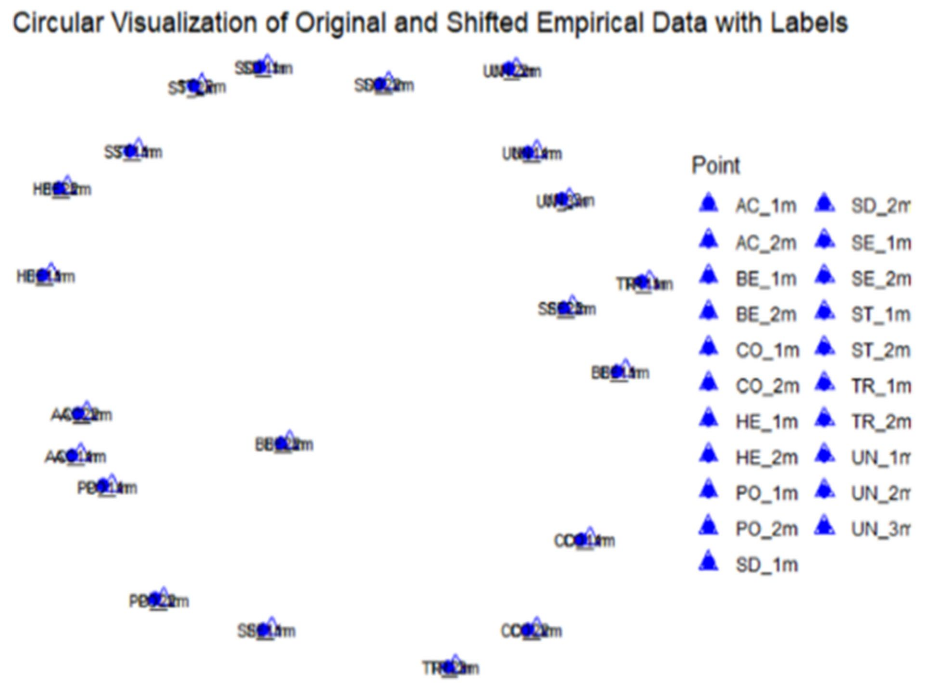

FIGURE 9
Result generated with prompt 3. See chat log: https://chat.openai.com/share/ff1d00b7-69aa-4f9d-a92e-a88236701db7.

failures was, not least due to the creation of a loop (recurring responses), not possible. Additionally, it needs to be noted that the results must be read in the context of the fact that the experiment took place in August 2023, and therefore, the results may be different at other points in time, with as more refined AI models are released for public use.

At this point, it should be mentioned that we had to refrain from discussing more comprehensive ethical considerations, such as the potential misuse of AI-generated research results or the long-term effects of AI in science. However, this is increasingly addressed in research like Resnik and Hosseini (2024) and future experiments can use the insights generated in such discussions.

Coming from limitations to actual outcomes of our study, the results show that ChatGPT is able to suggest an analysis procedure and generate syntax code for a proprietary software solution that is in use in many research institutions and universities – SPSS in our case.

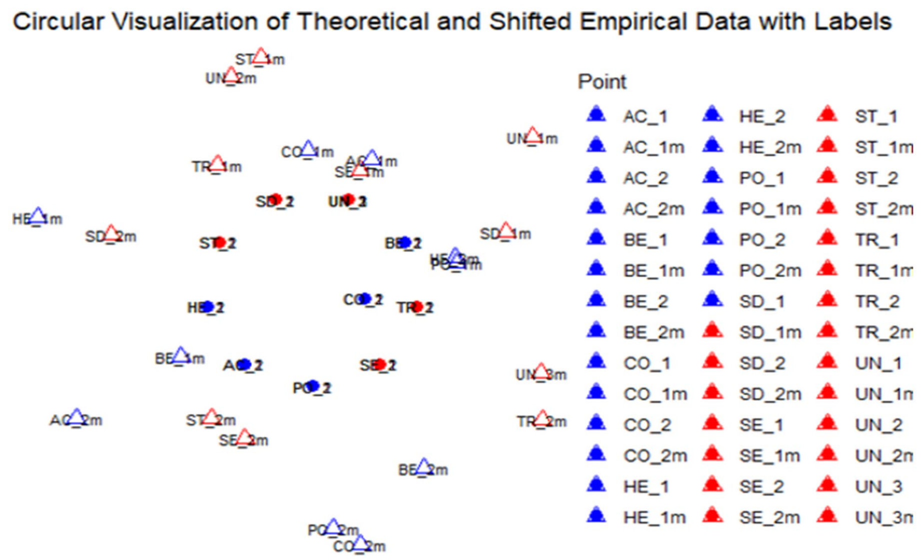

FIGURE 10
Result of the R-Syntax generated with prompt 4. See chat log: https://chat.openai.com/share/ff1d00b7-69aa-4f9d-a92e-a88236701db7.

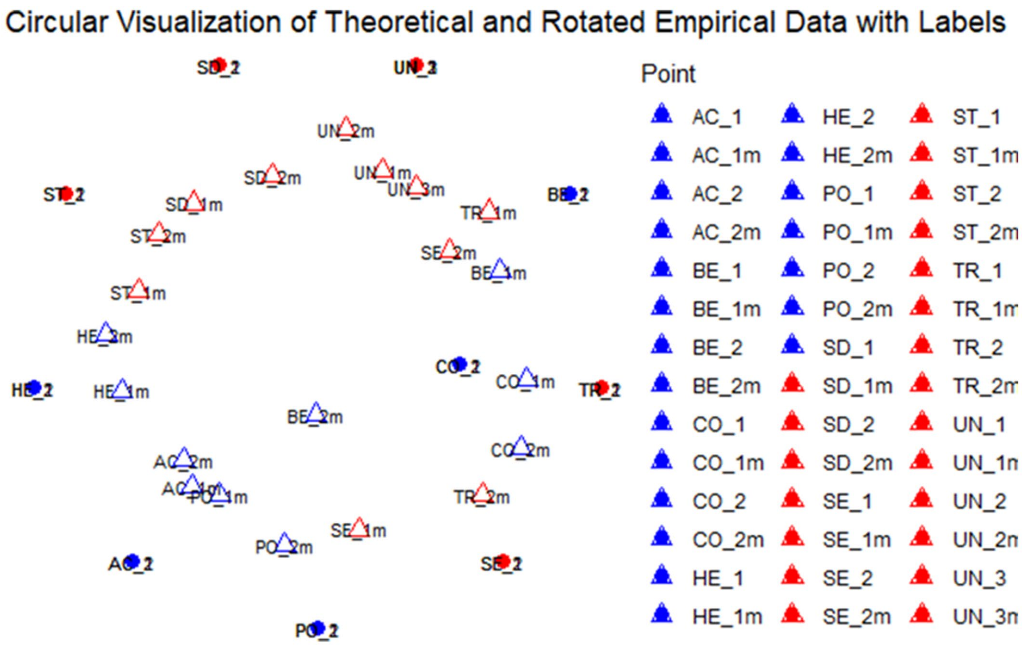

FIGURE 11
Result of the R-Syntax generated with prompt 23. See chat log: https://chat.openai.com/share/ff1d00b7-69aa-4f9d-a92e-a88236701db7.

However, in line with other literature on the issue, the suggestions were often general, and the implementation of the procedures was flawed. In the end, the users had to make their own decisions and corrections, showing that the idea of naïve users employing free-to-use software to improve their work is still a flawed assumption.

To analyze ChatGPT's capabilities, we chose tasks that would be classified as having different levels of difficulty for prospective researchers (basic vs. advanced analyzing methods). This allowed us to show that it is not the difficulty of the task but the availability of training material that determines the quality of ChatGPT's work.

Accordingly, the results must be structured to get a deeper understanding of the role GPT-based tools may play in future research processes. Firstly, our example showed that ChatGPT-3.5 can be a useful tool in research, but human expertise and intervention are needed to achieve optimal results. This limits the applicability of the free-to-use tools for academic work. The user needs either advanced competencies in their field of interest – in our case, statistics and programming – or prompt crafting. In most cases, we expect both will be necessary to create satisfying results.

Secondly, the experiment illustrates that the risks arising from faulty syntax generated by GPT software are manageable, as its use is not possible, thereby preventing (presumed incorrect) results from being produced. Again, the greater risk lies in user-based decision-making based on the proposed solution options.

Thirdly, while some literature supports the claim that using AI support for specific tasks can save time, the experiment shows that this is also tied to the knowledge of the user and their ability to adapt to the particularities of the tool – in our chase GPT 3.5. However, for our experiment with the naïve user, it must be stated that producing numerous prompts without finding a final solution (potentially ending up in loops) can also be very time-consuming. Thus, it is again the more proficient – maybe practiced or ideally trained—use that leads to efficiency.

Utilizing ChatGPT more effectively, saving time and improving overall usage is thus still a question of substantive knowledge and skill in using the tool. Thus, the idea of using AI-assisted tools to allow social scientists to focus on more substantive issues, limiting the time and resources allotted to more technical tasks (Kim and Ng, 2022; Brooker, 2019), seems unrealistic—at least for the moment. Even though the user may gain capabilities beyond their own knowledge in the field of data analysis and result presentation through the use of ChatGPT, adequate and validated usage can only be ensured if the user is "one step ahead" and has a sufficient understanding of the subject matter. If the user does not have sufficient skills, it is—as before—essential to seek support from an expert in the respective context. Overall, this means that free-to-use AI-powered tools like ChatGPT-3.5 have the potential to improve and speed up the research process, but they cannot completely replace human expertise and judgment. It is important that researchers use these tools critically, evaluate their results carefully, and also consider what advantages they may have using commercial offerings based on, e.g., GPT-4. Furthermore, based on current developments, it can be assumed that using a single AI tool will not be "enough" in the future, e.g., the emergence of tools like SciSpace. Instead, there will likely be a large number of tools and custom versions that can be used as support, but their use will also need to be learned and reflected upon.

Finally, more research, especially concerning specific case studies, as proposed in our paragraph on limitations, are needed to improve clarity.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

DP: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing. DW: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing. SH: Resources, Validation, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This publication is supported by the Johannes Kepler Open Access Publishing Fund and the federal state of Upper Austria.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abbasnasab Sardareh, S., Brown, G. T. L., and Denny, P. (2021). Comparing four contemporary statistical software tools for introductory data science and statistics in the social sciences. *Teach. Stat.* 43, 157–172. doi: 10.1111/test.12274

AlZaabi, A., Al-Amri, A., Albalushi, H., Aljabri, R., and AalAbdulsallam, A. (2023). ChatGPT applications in academic research: a review of benefits, concerns, and recommendations. *Biorxiv*. doi: 10.1101/2023.08.17.553688

Bašić, Ž., Banovac, A., Kružić, I., and Jerković, I. (2023). ChatGPT-3.5 as writing assistance in students' essays. *Communications* 10:28. doi: 10.1057/s41599-023-02269-7

Bellés-Calvera, L., and Caro Quintana, R. (2022). Is academic discourse accurate when supported by machine translation?. *Quaderns de Filologia* 171–201. doi: 10.7203/QF.27.24671

Bilsky, W., Janik, M., and Schwartz, S. H. (2011). The Structural Organization of Human Values – Evidence from Three Rounds of the European Social Survey (ESS). *J. Cross-Cult. Psychol.* 42, 759–776. doi: 10.1177/0022022110362757

Bilyk, Z. I., Shapovalov, Y. B., Shapovalov, V. B., Megalinska, A. P., Zhadan, S. O., Andruszkiewicz, F., et al. (2022). Comparison of Google Lens recognition performance with other Plant recognition systems. *Educ. Technol. Quart.* 2022, 328–346. doi: 10.55056/etq.433

Brady, H. E. (2019). The challenge of big data and data science. *Annu. Rev. Polit. Sci.* 22, 297–323. doi: 10.1146/annurev-polisci-090216-023229

Brooker, P. D. (2019). Programming with Python for social scientists. London and Thousand Oaks, CA: SAGE Publications.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., and Amodei, D. (2020). Language models are few-shot learners. *arXiv*. doi: 10.48550/arXiv.2005.14165

Calonge, D. S., Smail, L., and Kamalov, F. (2023). Enough of the chit-chat: a comparative analysis of four AI chatbots for calculus and statistics. *J. Appl. Learn. Teach.* 6, 346–357. doi: 10.37074/jalt.2023.6.2.22

Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., and Floridi, L. (2017). Artificial intelligence and the 'good society': the US, EU, and UK approach. *Sci. Eng. Ethics* 24, 505–528. doi: 10.1007/s11948-017-9901-7

Cieciuch, J., Schwartz, S. H., and Vecchione, M. (2013). Applying the Refined Values Theory to Past Data: What Can Researchers Gain?. *J. Cross-Cult. Psychol.* 44, 1215 –1234. doi: 10.1177/0022022113487076

Cooper, G. (2023). Examining science education in ChatGPT: an exploratory study of generative artificial intelligence. *J. Sci. Educ. Technol.* 32, 444–452. doi: 10.1007/s10956-023-10039-y

Cotton, D. R., Cotton, P. A., and Shipway, J. R. (2023). Chatting and cheating: ensuring academic integrity in the era of ChatGPT. *Innov. Educ. Teach. Int.* 61, 228–239. doi: 10.1080/14703297.2023.2190148

Couldry, N., and Van Dijck, J. (2015). Researching social media as if the social mattered. *Social Media Society*, 1.

Couldry, N., and Powell, A. (2014). Big data from the bottom up. *Big Data & Society*, 1. doi: 10.1177/2053951714539277

De Wet, J., Wetzelhütter, D., and Bacher, H. (2020). "Standardising the reproduction of Schwartz's two-dimensional value space using multi-dimensional scaling and goodness-of-fit test procedures." *Qual. Quant*. 55, 1155–1179. doi: 10.1007/s11135-020-01041-2

Farhat, F., Sohail, S. S., and Madsen, D. Ø. (2023). How trustworthy is ChatGPT? The case of bibliometric analyses. *Cogent Educ.* 10. doi: 10.1080/23311916.2023.2222988

Gabriel, I., Manzini, A., Keeling, G., Hendricks, L. A., Rieser, V., Iqbal, H., et al. (2024). The ethics of advanced AI assistants. *arXiv*. doi: 10.48550/arXiv.2404.16244

Gao, C. A., Howard, F. M., Markov, N. S., Dyer, E. C., Ramesh, S., Luo, Y., et al. (2023). Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *NPJ Digit. Med.* 6:75. doi: 10.1038/s41746-023-00819-6

Grover, P., Kar, A. K., and Dwivedi, Y. K. (2022). Understanding artificial intelligence adoption in operations management: insights from the review of academic literature and social media discussions. *Ann. Oper. Res.* 308, 177–213. doi: 10.1007/s10479-020-03683-9

Hansson, E., and Ellréus, O. (2023). Code correctness and quality in the era of AI code generation: Examining ChatGPT and GitHub copilot. Kalmar: Linnaeus University. oai:DiVA.org:lnu-121545

Henrickson, L., and Meroño-Peñuela, A. (2023). Prompting meaning: a hermeneutic approach to optimizing prompt engineering with ChatGPT. *AI Soc.* doi: 10.1007/s00146-023-01752-8

Hong, Z. (2023). ChatGPT for computational materials science: a perspective. *Energy Mat. Adv.* 4:0026. doi: 10.34133/energymatadv.0026

Inglehart, R., and Welzel, C. (2005). *Modernization, Cultural Change, and Democracy: The Human Development Sequence*. Cambridge: Cambridge University Press.

Jahic, I., Ebner, M., and Schön, S. (2023). Harnessing the power of artificial intelligence and ChatGPT in education – a first rapid literature review. *EdMedia+ Innov. Learn.*, 1489–1497.

Kalla, D., and Smith, N. (2023). Study and analysis of chat GPT and its impact on different fields of study. *Int. J. Innov. Sci. Res. Technol.* 8, 827–833. doi: 10.5281/zenodo.7767675

Karaali, G. (2023). Artificial intelligence, basic skills, and quantitative literacy. *Numeracy*, 16:9. doi: 10.5038/1936-4660.16.1.1438

Kim, J. Y. S., and Ng, Y. M. M. (2022). "teaching computational social science for all." PS. *Polit. Sci. Polit.* 55, 605–609. doi: 10.1017/S1049096521001815

Kritzinger, S., Pfaff, K., Barta, J., Bernhard, J., Boomgaarden, H., Eder, A., et al. (2023). Digitize! – Computational Social Science in der digitalen und sozialen Transformation. *Zeitschrift für Hochschulentwicklung* 18, 173–195. doi: 10.21240/zfhe/SH-F/11

Lai, V. D., Ngo, N. T., Veyseh, A. P. B., Man, H., Dernoncourt, F., Bui, T., et al. (2023). ChatGPT beyond English: towards a comprehensive evaluation of large language models in multilingual learning. *arXiv*. doi: 10.48550/ARXIV.2304.05613

Leitgöb, H., Prandner, D., and Wolbring, T. (2023). Editorial: Big data and machine learning in sociology. *Front. Sociol.* 8:1173155. doi: 10.3389/fsoc.2023.1173155

Malinka, K., Peresíni, M., Firc, A., Hujnák, O., and Janus, F. (2023). On the educational impact of ChatGPT: is artificial intelligence ready to obtain a university degree? Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education, 1, 47–53.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., et al. (2011). "Big data: the next frontier for innovation, competition, and productivity." Available at http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

Meyer, J. G., Urbanowicz, R. J., Martin, P. C. N., O'Connor, K., Li, R., Peng, P. C., et al. (2023). ChatGPT and large language models in academia: opportunities and challenges. *BioData Mining* 16:20. doi: 10.1186/s13040-023-00339-9

Nikolaidis, N., Flamos, K., Feitosa, D., Chatzigeorgiou, A., and Ampatzoglou, A. (2023). The end of an era: Can AI subsume software developers? Evaluating ChatGPT and Copilot Capabilities Using Leetcode Problems. doi: 10.2139/ssrn.4422122

Ouyang, S., Zhang, J. M., Harman, M., and Wang, M. (2023). LLM is like a box of chocolates: The non-determinism of ChatGPT in code generation. doi: 10.48550/arXiv.2308.02828

Pallant, J. (2023). SPSS survival manual: a step by step guide to data analysis using IBM SPSS. *Aust. N. Z. J. Public Health* 37, 597–598. doi: 10.1111/1753-6405.12166

Prandner, D., Hasengruber, K., and Forstner, M. (2021). DASH – 2020: Datasharing und Datenmanagement in den österreichischen Sozialwissenschaften. Linz: Johannes Kepler Universität.

Radford, A., Narasimhan, K., Salimans, T., and Sutskever, I. (2018). Improving language understanding by generative pre-training. Available at: https://hayate-lab.com/wp-content/uploads/2023/05/43372bfa750340059ad87ac8e538c53b.pdf

Rahman, M. M., and Watanobe, Y. (2023). ChatGPT for education and research: opportunities, threats, and strategies. *Appl. Sci.* 13:5783. doi: 10.3390/app13095783

Ray, P. P. (2023). ChatGPT: a comprehensive review on background, applications, key challenges, Bias, ethics, limitations, and future scope. *Int. Things Cyber-Phys. Syst.* 3, 121–154. doi: 10.1016/j.iotcps.2023.04.003

Resnik, D. B., and Hosseini, M. (2024). The ethics of using artificial intelligence in scientific research: new guidance needed for a new tool. *AI and Ethics*. 1–23.

Rozado, D. (2023). The political biases of ChatGPT. *Soc. Sci.* 12:148. doi: 10.3390/socsci12030148

Sada, N., Maldonado, A., and Maldonado, A. (2007). Research methods in education. Sixth edition—by Louis Cohen, Lawrence Manion, and Keith Morrison. *Br. J. Educ. Stud.* 55, 469–470. doi: 10.1111/j.1467-8527.2007.00388_4.x

Sallam, M. (2023). ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthcare* 11:887. doi: 10.3390/healthcare11060887

Salvagno, M., Taccone, F. S., and Gerli, A. G. (2023). Can artificial intelligence help for scientific writing? *Crit. Care* 27:75. doi: 10.1186/s13054-023-04380-2

Schwartz, S. H. (1992). Universals in the content and structure of values: Theory and empirical tests in 20 countries. In *Advances in experimental social psychology*. editor M. Zanna, Vol. 25 (New York: Academic Press), 1–65.

Schwartz, S. H. (1994). Are There Universal Aspects in the Structure and Contents of Human Values? *J. Soc. Issues.* 50, 19–45. doi: 10.1111/j.1540-4560.1994.tb01196.x

Schwartz, S. H. (2009). *Draft User's Manual: Proper Use of the Schwarz Value Survey, version 14 January 2009*. Compiled by Romie F. Littrell. Centre for Cross Cultural Comparisons.

Schwartz, S. H. (2012). An overview of the Schwartz theory of basic values. *Online Read. Psychol. Cult.* 2, 1–20. doi: 10.9707/2307-0919.1116

Starr, N. (2006). Special section: how to lie with statistics turns fifty. *Coll. Math. J.* 37, 244–245. doi: 10.2307/27646349

Stokel-Walker, C. (2023). ChatGPT listed as author on research papers: many scientists disapprove. *Nature* 613, 620–621. doi: 10.1038/d41586-023-00107-z

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

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. *arXiv*. doi: 10.48550/arxiv.1706.03762

Zhang, S. J., Florin, S. H., Lee, A. N., Niknafs, E., Marginean, A., Wang, A., et al. (2023). Exploring the MIT mathematics and EECS curriculum using large language models. *Am. J. Med.* 136, 725–726.

Zheng, H., and Zhan, H. (2023). ChatGPT in scientific writing: a cautionary tale. *Am. J. Med.* 136, 725–726.e6. doi: 10.1016/j.amjmed.2023.02.011