universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„From Acknowledgments to Acceptance? Academic Beginners' Perceptions of LLMs in Social Science Research"

verfasst von / submitted by

Hannah Maria Kronschnabl

angestrebter akademischer Grad / in partial fulfillment of the requirements for the degree of

Master of Science (MSc)

Wien, 2023 / Vienna 2023

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 550
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Communication Science
Betreut von / Supervisor:	Univ.-Prof. Hajo Boomgaarden, PhD
Mitbetreut von / Co-Supervisor:	Dr. Jakob-Moritz Eberl, MA

Acknowledgments

First of all, I would like to acknowledge the use of ChatGPT (OpenAI Inc., San Francisco, CA, USA) to generate the abstract displayed in the stimulus material for the present study. After using ChatGPT, we reviewed and edited the abstract created by ChatGPT as needed and take full responsibility for the generated content. ChatGPT was not used in any other capacity while producing this thesis.

To my supervisors, Univ.-Prof. Hajo Boomgaarden and Dr. Jakob-Moritz Eberl, thank you for all your valuable feedback and advice. Without your guidance, support, and inspiration, this thesis would not have been possible.

Finally, I would like to thank my parents. Without your endless support and love, I would never have made it this far.

.

Table of Content

Introduction

"This article was written by the ChatGPT chatbot, in response to prompts from MK" (King & ChatGPT, 2023, p. 1). Lately, sentences like this can be found in preprints and published papers in different academic fields. The sentence that occurs above, for example, is from an academic publication in the social science field by Michael King, in which he wrote on Chatbots, and Plagiarism in Higher Education, and reported responses from the currently widely discussed large language model (LLM) ChatGPT to his prompts summarizing them into a research article (King & ChatGPT, 2023). As most of the texts in his research article were generated by the chatbot ChatGPT, which is an LLM combined with an easy-to-use interface (Teubner et al., 2023) and specifically optimized for generating conversational answers (Megahed et al., 2023), he listed the language model as the co-author of this academic article. Listing ChatGPT as an author and using its automatically generated content for research was also done by some researchers in the medical field (ChatGPT & Zhavoronkov, 2022; O'Connor & ChatGPT, 2023). By this time, concerns about these research practices and types of authorship began to emerge in the academic community (Stokel-Walker, 2023). Most scholars rejected the citing of ChatGPT as co-authors (Stokel-Walker, 2023) and expressed concerns about the LLM's potential to change scientific communication (Teubner et al., 2023). Publishers such as Nature, which includes all Springer Nature journals, and Elsevier echoed these concerns and responded by banning all LLMs as authors for their publications (Stokel-Walker, 2023). Based on this, the focus of this current debate around LLMs is shifting away from the act of automated independent writing of LLMs in academia, to the supporting role of LLMs with language and actual content or ideas used for research-related tasks.

According to a survey of Nature readers (Owens, 2023), ChatGPT was most commonly used in an academic context for brainstorming research ideas (27%), writing computer code (24%), and writing research manuscripts (17%). Owens (2023) states that LLMs can help with tedious or repetitive tasks and allow researchers to focus on "higher thinking". Social Science scholars such as Teubner et al. (2023) and Haensch et al. (2023) also see LLMs as a positive contribution to their field when used properly, as their language-related services can save time and improve research quality as researchers can focus on more complex tasks in their work.

Artificial Intelligence (AI) tools in general recently received increased attention in social science research. Several assisting AI tools such as Quiltbot (Kurniati & Fithriani, 2022) or Grammarly (O'Neill & Russell, 2019) used for academic writing and teaching in higher education have been studied, as well as AI perceptions and the underlying machine heuristics (Lee, 2018; Molina & Sundar, 2022; Sundar, 2020) of AI tools to explore why machines are trusted under certain circumstances but not under others. Factors influencing AI perceptions such as prior belief patterns (Sundar & Kim, 2019) or differences in roles Artificial Intelligence plays in Human-AI-Interactions (Kim et al., 2021) have also been studied. ChatGPT as a new emerging generative AI tool has already been extensively analyzed, tested, and tried out to explore its specific applications by scholars (Buriak et al., 2023; Kutela et al., 2023; Lund et al., 2023; Salvagno et al., 2023), yet there has been relatively little research on perceptions towards the tool in academia. Quality perceptions of texts produced with LLMs and generative AI's perceived usefulness for specific tasks in the academic writing process have not yet been investigated, although quality assessment occupies a very important role in academic publishing (Lindgreen et al., 2021). To fill this research gap, this study examines the perceptions towards ChatGPT in terms of quality assessment and perceived usefulness, and how these differ between ChatGPT

functioning as an editing tool or content and idea generation tool. In addition, the positivistic attitudes of academics are examined as a possible influencing factor on the perception of ChatGPT.

This study follows the practical instructions from publishers such as Nature and Elsevier, who have established new guidelines for dealing with LLMs in academic publishing (Stokel-Walker, 2023), and displays acknowledgments indicating the involvement of ChatGPT in the research process to participants in this experimental study. To gain initial insight into this under-explored field of research, social science students, who are considered academic beginners, participated in the study. This leads to the overarching research question, namely, whether acknowledgments indicating different involvements of ChatGPT in the research process, such as editing or content and idea creation, have an impact on young academics' perceptions of research quality and usefulness of ChatGPT, and whether this is influenced by their level of positivistic thinking.

Understanding Large Language Models (LLMS) and ChatGPT

The underlying AI concepts for understanding Large Language Models and chatbots such as ChatGPT are machine learning (ML) and natural language processing (NLP) (Lund et al., 2023). ML allows systems such as ChatGPT to learn from the data to improve their abilities in natural language processing (Lund et al., 2023). The term natural language processing (NLP) describes "the study of computational and mathematical modeling of various aspects of language and the development of a wide range of systems" (Joshi, 1977, p. 1242). NLP aims to model language computationally and deals with linguistic features of computation (Joshi, 1977).

Large Language Models

Language models that have made substantial strides in NLP in recent years are the so-called large language models (LLMs). These large-scale models can perform language-related tasks with high accuracy because they are trained on a very huge amount of text data (Kasneci et

al., 2023). Two of the key developments that have laid the foundation for today's LLMs are the use of transformer architecture (Vaswani et al., 2017) and the use of pretraining (Min et al., 2021). A LLM that incorporates both developments and is notable for its exceptional scope as well as the sheer volume of training data, is the Generative Pre-trained Transformer (GPT) (Lund et al., 2023). According to Lund et al. (2023, p. 572), advanced algorithms and deep learning techniques allow the GPT to "understand the context of a piece of text and generate human-like responses, making it unique among language processing tools." The research lab OpenAI developed the sophisticated generative language models GPT, GPT-2, and GPT-3, and in March 2023 the latest version GPT-4 (OpenAI, 2023). By the combination of pretraining on large datasets and fine-tuning on more task-specific input, the latest versions GPT-3 and GPT-4 can perform various NLP tasks with a high degree of accuracy (Megahed et al., 2023).

ChatGPT

The AI-powered chatbot variation of GPT-3 (technically GPT-3.5) is called ChatGPT and is fine-tuned for generating human-like responses to user inputs in a conversational context (Megahed et al., 2023). ChatGPT's generative model can perform a variety of natural language tasks, such as answering various questions, translations, computer programs, or coherent essays (Kasneci et al., 2023). The easy-to-use chat interface of the LLM and the public accessibility provided by OpenAI makes it usable for a wide mass (OpenAI, 2023). According to Teubner et al. (2023, p. 96) the public release of ChatGPT, was a "breakthrough move" that opened the technology to broader use and not just experts as was the case with previous LLMs. Thus, ChatGPT, as one of the first language models open to a broad mass of people, is of great social importance and is increasingly becoming a research topic in the academic field.

Research on ChatGPT in Academia

Studies on ChatGPT in an academic context can be divided into two distinct areas: Firstly, studies in an educational context researching the impact of ChatGPT for teaching and learning in higher education (see literature overview in Rudolph et al. (2023)), and secondly, studies about ChatGPT in scientific publishing discussing the LLM's potential impact on academic work and its role for scientific writing (Buriak et al., 2023; Gao et al., 2022; Kutela et al., 2023; Lund et al., 2023; Salvagno et al., 2023). As this paper is focussing on academics' perception of ChatGPT used for scientific publishing, the following paragraph will focus on the second research area and provide more information about ChatGPT in scientific writing.

Benefits and Challenges of ChatGPT for Scientific Writing

ChatGPT in scientific writing was studied by Gao et al. (2022) and Kutela et al. (2023) by comparing human-written to ChatGPT-written texts. The texts could be distinguished using several supervised and unsupervised text mining techniques (Kutela et al., 2023), and in the majority of cases also by human reviewers who described ChatGPT-generated content as formulaic and rather vaguely worded (Gao et al., 2022). Nearly 40% of human reviewers in Gao et al.'s (2022) study incorrectly identified ChatGPT-written abstracts as human ones. LLMs thus have great potential for scientific writing but are mostly studied not as completely autonomous acting but as assisting tools for the writing process of academic papers (Bishop, 2023; Buriak et al., 2023; Lund et al., 2023; Salvagno et al., 2023). The LLM can assist in various steps within the manuscript preparation - from creating an initial draft, organizing the literature, and generating code for the analysis to proofreading the finished paper (Salvagno et al., 2023).

First and foremost, ChatGPT can provide help during the writing process by generating an initial draft of a scientific paper (Salvango et al, 2023). Having an initial draft to start with can

help those struggling to write their first words break their mental log jams and get into writing (Buriak et al., 2023). According to Salvagno et al. (2023), ChatGPT can also be compared to AI research assistants such as "elicit.org" as it can identify references that might be missed by conventional literature research and create links between already existing ideas (Buriak et al., 2023; King & ChatGPT, 2023; Lund et al., 2023). Besides, ChatGPT can also help researchers by providing code for the statistical analyses of their data (Buriak et al., 2023; Halaweh, 2023).

At the end of the manuscript preparation, the LLM can assist the researchers with editing tasks (Salvagno et al., 2023) as it is writing with nearly perfect grammar, syntax, spelling, punctuation, vocabulary, and sentence and paragraph structure (Bishop, 2023). Additionally, it can help with formatting (Buriak et al., 2023; Lund et al., 2023), formulating complex sentences more clearly (Salvagno et al., 2023), and in a more formal manner (Bishop, 2023). This can be helpful for non-native English researchers as revising their language in manuscripts or translating English literature to better understand the content can help them to archive higher quality publications for an international research community (Jiao et al., 2023).

Summing up the challenges of using ChatGPT for text and content generation in the academic field the reliance on the existing database can be named (Buriak et al., 2023). As until now ChatGPT's database only includes information published before 2021, it is not possible to use it to write up-to-date introductions or overviews, and the insights and in-depth analyses gained will be limited in their perspectives (Buriak et al., 2023). Generating erroneous citations as well as false or inheriting biases and inaccuracies already presented in the scientific field can also be mentioned as challenging issues (Buriak et al., 2023; Sundar & Liao, 2023). Many scholars (Bishop, 2023; Buriak et al., 2023; Sundar & Liao, 2023) also point out that AI technologies can only communicate widely published ideas whereas scientific writing also requires creativity and critical thinking

to develop original concepts and ideas. Ethical considerations such as the fear that the use of the LLMs in academia could lead to more plagiarism and a lack of critical human expert minds behind scientific work are also discussed in depth in the literature (Salvagno et al., 2023).

All in all, researchers in the literature agree that human intervention is required when using LLMs for scientific writing (Buriak et al., 2023; Kutela et al., 2023). Buriak et al. (2023) and Kutela et al. (2023) suggest that ChatGPT's output should only be used as guiding material serving as an initial draft, which should be carefully reviewed for academic publishing and edited if necessary. Therefore, the present study is limited to perceptions towards ChatGPT as an assisting tool in research and not to perceptions of fully automatically generated papers by ChatGPT.

ChatGPT for Text Generation vs. Idea and Content Generation

As mentioned in the introduction, ChatGPT is most often used in research for brainstorming ideas, creating code, and helping with the actual writing of the research manuscript, such as editing or formatting (Owens, 2023). In this study, the focus lies not on computational tasks but on language-based tasks of ChatGPT in an academic context. The generating of ideas and content within the research process and the generating of linguistically correct texts in a scientific writing style are therefore examined in the presented study. Text generation by LLMs in research can save time as researchers don't have to do tedious editing tasks (Haensch et al., 2023; Teubner et al., 2023). According to Haensch et al. (2023), there should be no concern about ChatGPT's use for proofreading, as it is nothing more than consulting a human editor, which previously used to be a common practice for researchers and students.

Using ChatGPT for content and idea generation has elicited mixed opinions among scholars. According to Lund et al. (2023) and Salvagno et al. (2023), ChatGPT can assist in the literature review and also create ideas for that part. Other scholars (Bishop, 2023; Buriak et al., 2023;

Sundar & Liao, 2023) point out that LLMs' capability to generate original concepts or ideas is limited as they can only communicate already existing ideas whereas scientific writing also requires creativity and critical thinking to develop original concepts and ideas. Halaweh (2023) considers Human-AI Collaboration as the best way for idea generation, as ChatGPT can very quickly provide humans with already existing ideas, and humans can build further ideas on them. In this study, ChatGPT is compared in its function as an idea and content generating tool with its function as a text generating, editing tool. These two different language-based roles of ChatGPT for scientific writing are reflected in the two treatment groups of the experimental study presented in this paper (see Method).

Acknowledging LLMs in Academic Publishing

After the first author attributions of scientific publications to ChatGPT (ChatGPT & Zhavoronkov, 2022; King & ChatGPT, 2023; O'Connor & ChatGPT, 2023), a critical discussion started about the ways of crediting LLMs in scientific publications (Stokel-Walker, 2023). Several scholars argued that LLMs cannot be an author as they are not fulfilling authorship criteria (Hosseini et al., 2023; Pourhoseingholi et al., 2023; Yeo-Teh & Tang, 2023). Only human authors can be held fully accountable, as the language models cannot give conscious, autonomous consent and therefore could not be considered responsible for their part in scientific publications (Hosseini et al., 2023; Yeo-Teh & Tang, 2023). Some scholars argue that LLMs are only proactively responding to prompts from researchers (Hosseini et al., 2023) and therefore do not enter into a collaborative, but rather an "obligatory-commercial" relationship with them (Pourhoseingholi et al., 2023). A point most researchers emphasize in their papers is the importance of transparency for the use of LLMs in academic publishing (Hosseini et al., 2023; Kutela et al., 2023; Pourhoseingholi, 2023). According to them the use of ChatGPT has to be disclosed and

publishers should adapt their transparency policies to the vast emergence of LLMs in the academic field (Buriak et al., 2023; Hosseini et al., 2023; Pourhoseingholi et al., 2023).

Publishers such as Nature and Elsevier responded to these demands and changed their transparent and editorial policies (Stokel-Walker, 2023). In its “Guide to Authors”, Nature added that (1) LLM tools cannot be credited as authors of research papers and (2) that the use of LLMs in manuscript preparation should be documented in the acknowledgments and method section (Nature Editorials, 2023). Elsevier (2023) has taken a similar stance to Nature in its publishing guidelines, prohibiting authorship of generative AI, but allowing its use to "improve readability and language" if disclosed in an additional statement at the end of the manuscript. Following these practical instructions from publishers to disclose all ChatGPT-generated text and ideas, this paper follows the guidelines of Nature and examines acknowledgments disclosing ChatGPT’s assistance in different parts of the paper (Nature Editorials, 2023).

Perceptions of Human-AI Interaction in Academic Publishing

As mentioned earlier, due to the very recent appearance of ChatGPT, there is only limited evidence of academics’ perceptions towards ChatGPT for academic publishing. For this reason, in this section, we will draw on the literature on general perceptions towards AI and assisting AI (Molina & Sundar, 2022; Sundar, 2020; Sundar & Kim, 2019), and apply these views to ChatGPT to develop the hypotheses for this study. Since ChatGPT cannot be used as the sole author, text, and idea generator for research, but only as an assisting tool for academic publishing, we consider the perception of Human-AI Interactions (HAII) in particular.

In Human-AI Interactions Assisting AI tools are influenced by different “machine heuristics” (Sundar, 2020). Heuristics are “rules of thumb” applied by individuals to make quick decisions or judgments in complex information environments (Sherman & Corty, 1984). Based on

stereotypes and depending on whether using machine attributes for the activity in question is appropriate, the heuristic can be rather positive or negative when a chatbot is the source of interaction (Sundar, 2020). According to Molina and Sundar (2022) the "positive machine heuristics" refers to the perception that AI is more precise and accurate compared to humans, and the "negative machine heuristics" refers to the perception of a rigid AI that lacks the ability of nuanced subjective judgments that humans are capable of. The possible impact of these machine heuristics on individuals' perceptions towards ChatGPT in scientific publishing in terms of research quality and usefulness of LLM tools is described in the following sections.

Perceptions of Research Quality & AI Assistance

The quality of research contributions is nowadays evaluated by most universities using recognized lists of top journals, citation counts, or other specific article- or author-level metrics (Lindgreen et al., 2021). Certain journal publishers or academic associations have developed specific criteria for evaluating the quality of research papers and applying them to the acceptance of papers into their journals or conferences. In the social science field, the International Communication Association (ICA) quality criteria for evaluating submissions to the annual conference can be mentioned in this regard, including evaluation points on theory, clarity, innovation, method, and overall impression of the paper (International Communication Association, n.d.).

Even though the evaluation of research based on these criteria is structured according to the given points, heuristics play a role in the quality assessments of texts (Arazy et al., 2017). Arazy et al.'s (2017) research show that cognitive decision processes and heuristic principles were applied in assessing text quality dimensions such as accuracy, completeness, objectivity, and representation (Arazy et al., 2017). Similar quality dimensions are also found in the academic context when evaluating research contribution, indicating that the aforementioned

"machine heuristics" could be applied to the quality assessment of the participants in the present study. According to Sundar (2020), the assignment of positive or negative machine attributes depends on the particular task of the AI tool. Mechanical tasks are perceived as more suitable tasks for algorithms than typical human tasks that involve subjective judgments and emotional capabilities (Lee, 2018).

As mentioned earlier, the use of ChatGPT in this study differs between assisting with editing and generating ideas for the content of the research. Based on the findings of Lee (2018) and Sundar (2020) editing tasks could be perceived as mechanical tasks and positive machine attributes such as accuracy, precision, and objectivity could be triggered by participants seeing the involvement of LLMs in the editing process of a research paper. Moreover, some studies on previous AI tools used in academia, such as Quilt Bot (Kurniati & Fithriani, 2022) or Grammarly (O'Neill & Russell, 2019), already confirm that these editing tools are perceived as positive and not as a hindrance for academic research performance. The generation of ideas and actual content by generative AI tools, on the other hand, is viewed more critically by some scientists, who argue that these tools rely on the synthesis of pre-existing content and are limited in their ability to generate truly original concepts or ideas as humans do (Sundar & Liao, 2023). Therefore, idea generation for research may be perceived more as a human task that triggers negative machine heuristics, which include that machines cannot make nuanced judgments (Lee, 2018; Sundar, 2020). From these research findings, the following assumptions regarding the participants' research quality perceptions towards the generative AI tool ChatGPT can be made:

> **H1a:** Research with acknowledgments stating ChatGPT's assistance in language editing and formatting is rated with the same quality as research with acknowledgments indicating no ChatGPT involvement in the research process.

H1b: Research with acknowledgments stating ChatGPT's assistance in content and idea creation for the draft is rated with less quality as research with acknowledgments indicating no ChatGPT involvement in the research process.

Perceived Usefulness of AI Tools in Academia

As explained above, LLMs such as ChatGPT are perceived as a beneficial development and useful tool for scientific writing by many researchers (Bishop, 2023; Buriak et al., 2023; Lund et al., 2023; Salvagno et al., 2023). Analyzing the perceived usefulness of AI tools is a common practice when new technologies are emerging in the field (O'Neill & Russell, 2019; Park & Kim, 2023). Par and Kim (2023) examined factors influencing the intentions of students, faculty, and staff to use mental health content in the form of an AI chatbot. The perceived usefulness of the AI chatbot was measured alongside other factors and was found to be a significant and positive influencing factor on people's intentions to use this new technology. O'Neill and Russel (2019), on the other hand, investigated the perceived usefulness of integrating Grammarly, a grammar tool providing feedback in terms of language editing, into students' assessments. They found that the Grammarly feedback was rated with a very high perceived usefulness, compared to more traditional feedback approaches (O'Neill & Russell, 2019).

The term "perceived usefulness" derives from the Technology Acceptance Model (TAM) developed by Davis (1989), which explains how people adopt new forms of technology. Davis (1989, p. 320) describes perceived usefulness as a motivational factor that influences people's intentions to use technology and can be defined as "the degree to which a person believes that using a particular system would enhance his or her job performance". Since the present study focuses on young scholars' perception of ChatGPT to improve their academic performance in scientific writing, Davis's (1989) concept of perceived usefulness was adopted.

The perceived usefulness of AI also differs depending on the different roles it plays in human-AI Interactions. Kim et al. (2021) examined perceptions toward social and functional AI in an online experiment. Social AI is described in the study with the primary role of serving as a social companion or even friend, while functional AI is described with the primary role of performing specific functions or tasks, such as proofreading or publishing a book (Kim et al., 2021). The results of the study show that the perceived usefulness of AI has a mediating effect, indicating that functional AI is perceived to be more useful and therefore elicits more positive attitudes toward functional AI than social AI (Kim et al., 2021). Since ChatGPT can be perceived more as a functional AI that performs tasks for humans as described in the study, including proofreading or generating ideas and content, it can be assumed that the ratings of perceived usefulness of ChatGPT across all three conditions will be reported as relatively high.

To date, there is only literature on functional AI and its perceived usefulness in editing and proofreading (see O'Neill and Russel, 2019), but nothing on idea-generating AI. Therefore, there is no real insight into whether ChatGPT is also perceived differently in its usefulness when performing different functions. But as the general perception of AI is influenced by so-called "machine heuristics" that depend on the task the AI is helping with, it can also be assumed that the perceived usefulness of ChatGPT is influenced by both positive and negative machine heuristics (Molina & Sundar, 2022; Sundar, 2020). In the area of proofreading, the chatbot might trigger positive machine characteristics and is therefore perceived as very useful in this specific function, as previously shown by O'Neill and Russel (2019), while using the ChatGPT for more human tasks such as idea generation might trigger more negative machine heuristics (Lee, 2018). These considerations lead to the following hypotheses:

H2a: Research with acknowledgments stating ChatGPT's assistance in language editing and formatting is rated with higher perceived usefulness of the LLM as research with acknowledgments indicating no ChatGPT involvement in the research process,

H2b: Research with acknowledgments stating ChatGPT's assistance in content and idea creation for the draft is rated with lower perceived usefulness of the LLM as research with acknowledgments indicating no ChatGPT involvement in the research process.

The Effect of Scientific Beliefs on AI Perceptions

In addition to their task dependency, the perception of AI also strongly relies on the personal attitudes and prior beliefs of its users, as "machine heuristics" are not only triggered by a particular task type (Lee, 2018) but may be also stored in a pre-existing belief structure users already hold (Sundar & Kim, 2019). In their study, Sundar and Kim (2019) found that people who expressed stronger prior beliefs in machine heuristics had higher intentions to disclose their private information to AI applications, supporting the view of the influence of preexisting cognitive heuristics in decision-making processes (Sundar & Kim, 2019). Since prior belief structures might also have an effect on LLMs' perceptions in an academic context, prior scientific beliefs are considered influencing factors in the present study.

The field of social sciences can be characterized as a highly diverse field in which researchers hold various pre-existing scientific beliefs that differ widely between quantitative and qualitative approaches (Goertz & Mahoney, 2012) and their positivistic tendency (Park et al., 2020). According to Park et al. (2020), positivism in science is consistent with a hypothesis-driven deductive model of science, which is based on testing hypotheses and experiments by the operationalization of measures and variables. Researchers inclined toward positivism concentrate on establishing "explanatory associations or causal relationships" using quantitative approaches

(Park et al., 2020, p. 690). Researchers who take less positivistic approaches tend to rely on qualitative methods that implicitly use logic and set theory, rather than probability and statistical theory which more positivistic researchers tend to draw on (Goertz & Mahoney, 2012). Whether positivistic attitudes of academics have an impact on the perceptions towards LLMs, and thus on the effects mentioned above regarding research quality and perceived usefulness ratings of ChatGPT, is explored in the following research question:

> **RQ1:** How do the effects of acknowledgments indicating different ChatGPT involvement in the research process differ on participants' perceptions of research quality (a) and perceived usefulness towards LLMs (b) depending on their attitudes toward positivism?

Method

To answer the proposed research question and hypotheses, a between-subject online experiment with three conditions, namely (1) research acknowledging ChatGPTs involvement in language editing, (2) research acknowledging ChatGPTs involvement in the creation of content, and (3) research without ChatGPTs involvement, was conducted. The first two conditions represent the experimental groups with the treatment of different ChatGPT involvement and the last condition is the control group without ChatGPT involvement in the research process.

Pre-test and Stimulus Material

For this experiment, an entry about an academic paper on science skepticism on the academic journal's website SAGE Journals was recreated (see Appendix A). The website was designed using Canva and was intended to look like the real SAGE Journals website. The name of the journal was not displayed to avoid priming effects on participants, but the rest of the stimulus material matched the information provided on the original SAGE journal website about a research paper: Title, authors, DOI, and abstract of a research paper. For the purposes of our study,

we also included an acknowledgment section directly below the abstract. The title and the abstract were created by ChatGPT Plus, which is based on the latest GPT-4 version.

Science skepticism was chosen as the research topic displayed in the stimulus material because it is a trans-disciplinary subject, primarily studied in the social science field, but tangible to other disciplines as well. To create an abstract of a new study on science skepticism that does not yet exist, ChatGPT was provided with four abstracts of real studies on science skepticism from social science research (Rutjens & van der Lee, 2020; Rutjens et al., 2022; Scheitle & Corcoran, 2021; Većkalov et al., 2022) as a starting point, as well as some additional information such as details about the method, word length of the abstract, etc. (see all used prompts in Appendix B). Initially, the pseudonyms John B. Doe and Jane A Doe were used as author names, but after critical feedback in the pretest, they were replaced with more real-sounding names.

All participants see the same title and abstract in the same layout, the only thing that differs between the conditional groups is the acknowledgments. For the experimental conditions (Condition 1 and 2) the acknowledgments were formulated according to a given example statement for the declaration of generative AI technologies' use in the writing process from Elsevier's (2023, p.10) Guide for Authors. Derived from the literature, for the first experimental condition, reflecting ChatGPT's involvement in editing, its use was formulated to "edit language, grammar, and style as well as format the entire paper", and for the second experimental condition, reflecting ChatGPT's involvement in content and idea generation, its use was specified to "provide ideas and feedback at different stages of our manuscript draft". In both cases, the acknowledgments noted that the authors take responsibility for the content and the paper was reviewed and edited by them after using ChatGPT. The third condition, reflecting the control group, on the

other hand, only received an acknowledgment thanking all anonymous reviewers (see Appendix A).

A pretest was conducted with 30 participants to test the stimulus material. Participants had to recall what was written about ChatGPt's involvement or non-involvement in the paper to check whether the manipulation worked. Two-thirds of the participants passed the manipulation check and remembered the correct content of the acknowledgments. To make the acknowledgments more prominent and to ensure that participants in the final study would read them, the abstract was shortened, and the instructions were reworded to emphasize the importance of reading the entire paper. As mentioned earlier, the authors' names were changed from pseudonyms to real-sounding names so that the paper would be perceived as more realistic.

Procedure

The questionnaire was implemented in Qualtrics. In the beginning, all participants were informed about their rights and had to give their informed consent. Participants had to pass the filter question that they were studying or working at a university or university-related institution to participate in the study. Before the stimulus presentation, they had to answer to which discipline and academic position (Bachelor, Master, etc.) they belong to, and rate their scientific beliefs towards positivism, the moderator variable of the study (see Appendix C). Then they were randomly assigned to one of three conditions (see Appendix A). The participants were asked to look closely at the entire entry and read every piece of information carefully, they could only proceed with the survey after 30 seconds. Then the participants had to answer several items measuring the dependent variables and the manipulation check. As the manipulation check was conducted in a so-called “split-design”, people were randomly assigned to see the manipulation check first and then the dependent variables or see it the other way round (see Appendix C). At

the end of the survey, participants were asked to provide their age as well as gender and were given a debriefing. The questionnaire was incentivized - participants had the chance to participate in a lottery and win a 50 € Amazon voucher.

Data Collection and Cleaning Processes

Data collection was carried out using a snowball sampling method. The link to participate in the study was sent to students in the field of social sciences via social networks of Austrian and German universities. The priori power analysis performed by G*power with an alpha = .05, a power = .80, and a small to medium-sized effect of $f^2 = .04$, calculated a necessary sample size of 244 people for the intended study design.

In the first data collection, in which mainly social science students from the University of Vienna were recruited, a total of 1900 fully completed responses could be collected. After a closer screening of the data, it was found that many responses follow certain patterns and are more likely to come from bots than from humans. Drawing on Brainard et al., (2022), who are giving insights about their experiences with bot attacks, a series of data screening processes were performed using R version 4.2.3 to filter out the bots and ensure the quality of the data.

First, all data with a large number of identical IP addresses and repetitive, strange, or artificial-sounding comments were deleted, as this strongly indicates a bot as the source of data generation (Brainard et al., 2022). The automated bot detection application from Qualtrics was also used to filter out data records with a Recaptcha Score ≤ 0.5. Entries, where participants did not enter the lottery and did not provide their email addresses, were left in the dataset, as it was assumed that the bots were programmed for winning the voucher. Responses that contained email addresses were filtered by their domains. All datasets with domains other than ".gmail" and ".yahoo" were considered human and included in the final data. The Gmail and Yahoo email

addresses were manually filtered for human-sounding and bot-generated emails that consisted of randomly strung-together letters and numbers. At the end of the data cleaning process, outliers were deleted and only bachelor's and master's students were included since the sample of this study consists of academic beginners, leaving 216 valid cases.

As soon as the bot infiltration was detected in the first dataset, a second data collection was started to collect more data unfiltered from bot answers. This time, a captcha query and a bot filter question with a cultural reference suggested by Brainard et al. (2022) were added to the beginning of the survey ("When was Adele prime minister?"). This data was collected by recruiting German students mainly through social network groups and advertising in university courses. After data cleaning for incomplete responses and time outliers, 50 cases remained. Thus, a total of 266 valid responses could be extracted from the two data collections.

Sample

Since only participants who passed the manipulation check can be included in the analysis of this study, the final sample consists of 152 people, with most participants (73%) identifying as female, 26% as males, and 1% as other. The average age was 24.14 (SD = 3.05) and the majority of participants were master's students (61%), with the rest being bachelor's students (39%). The sample included mainly students from the social science field (78%), and 22% non-social science students.

A high discrepancy between the conditions in their size can be observed, which is due to the prior thorough data cleaning after bot infiltration. 61 participants of the final sample were assigned to Condition 1, only 38 to Condition 2, and 53 to Condition 3. To check whether the randomization among the conditional groups worked as intended in terms of gender, age, and social sciences/non-social sciences worked as intended, Fisher’s exact and Kruskal-Wallis tests were

performed. The results of Fisher's exact test showed no significant associations between the three conditions and gender (p = .325) as well as social sciences/non-social sciences (p = .828). The Kruskal-Wallis test showed no statistical differences between the three conditions in terms of age ($H(2) = 0.49$, p = .828). Therefore, randomization for age, gender, and, social science/non-social science among the three conditional groups has worked as intended. But since the group sizes of the three conditions are very different and especially the second condition is very small compared to the other ones, the control variables are still included in all subsequent calculations and analyses of this study.

Measures

The dependent variables measured in this study for the three different conditions are the *research quality (RQ)* of the displayed entry on the academic journal's website and the *perceived usefulness (PU)* of ChatGPT. Additionally, *attitudes towards positivism (ATP)* were measured as a moderator variable at the beginning of the study (see Appendix C).

Research Quality

This dependent variable assessing the participants' quality perceptions towards the displayed research was measured using the ICA's quality criteria for evaluating submissions to their annual conference (International Communication Association, n.d.). Participants were rating five items on (1) theoretical depth, (2) clarity, (3) innovation, (4) methodology, and (5) the overall impression of the research paper on a scale from 1 = *low research quality* to 5 = *high research quality*. Cronbach's alpha coefficient found that the research quality variable consisting of five items was .79, which indicates high reliability.

Perceived Usefulness

To measure the other dependent variable evaluating the participants' perceptions of the usefulness of ChatGPT, Davis's (1989) scale on perceived usefulness was adopted. In this scale, six statements on the participants' perceptions of using ChatGPT for their studies and/or research work had to be rated from 1 = *extremely unlikely* to 7 = *extremely likely*. For the perceived usefulness variable consisting of six items, a Cronbach's alpha coefficient of .88 was calculated, also indicating the high reliability of the scale used.

Attitudes towards Positivism

The moderator variable was based on a scale introduced by Steel et al. (2004) on scientists' attitudes toward positivism. Five statements had to be rated by the participants from 1 = *strongly disagree* to 5 = *strongly agree.* The statements reflect a positivism index, which means the higher the participants' ratings, the stronger their positivistic views (Steel et al., 2004). Cronbach's alpha coefficient of this index is slightly lower than Cronbach's alpha of the two dependent variables, but can still be considered as a reliable factor with a value of .66.

Control Variables

A number of control variables are included to ensure that the dependent and moderating variables are measured correctly. Demographic variables such as age and gender, as well as social sciences (1 = *social sciences*, 2 = *non-social sciences*), are included.

Manipulation Check

To ensure that participants read the stimulus material, participants had to respond *yes*, *no*, or *I don't know* to four statements about what the authors had mentioned in the acknowledgments section of the presented journal's entry (see Appendix C). The manipulation check was deemed as passed if participants agreed to information that matched their assigned condition and rejected

information that did not occur in their condition. As mentioned only 152 from 266 people managed to pass it. This high number could be related to the possibility that not all bots could be filtered out by the data cleaning process and therefore some were eliminated by the manipulation check. To examine whether the stimulus material was perceived as realistic, participants had to rate statements from 1 = *strongly disagree* to 5 = *strongly agree.* They agreed that the entry (M = 3.81, SD = 1.06) and also the research (M = 4.10, SD = 0.91) presented in the study were very similar to entries or research found on academic journal websites.

Data Analysis

The data was analyzed using R version 4.2.3. To test the hypotheses stated in this study, two multiple regression models were run for the dependent variables RQ and PU, taking into account the control variables age, gender, and social sciences. For the research question about the moderation effect of participants' attitudes towards positivism (ATP), two moderation models (one for each dependent variable) were calculated. As a robustness check, all calculations were also performed with the filtered social science sample.

Results

Descriptives and Intercorrelations

Before testing the hypotheses and research question, the descriptives and correlations were calculated for all participants. The research quality of the displayed academic entry and the perceived usefulness of ChatGPT were both rated as quite moderate with a tendency towards rather higher quality ratings (M = 3.33, SD = 0.73) and rather higher perceived usefulness (M = 5.12, SD = 1.23) across all conditions (see Table 1). The two dependent variables show only a weak positive correlation $r(152) = .26$, $p = .001$.

Table 1

Descriptive Statistics and Correlation Coefficient for the Dependent Variables

Variable	*n*	*M*	*SD*	1	2
1. Research Quality (RQ)	152	3.33	0.73	—	
2. Perceived Usefulness (PU)	152	5.12	1.23	.26**	—

** $p < .01$.

Note. RQ was rated from 1 = *low quality* to 5 = *high quality* and PU from 1 = *extremely unlikely* to 7 = *extremely likely*

Hypotheses Testing

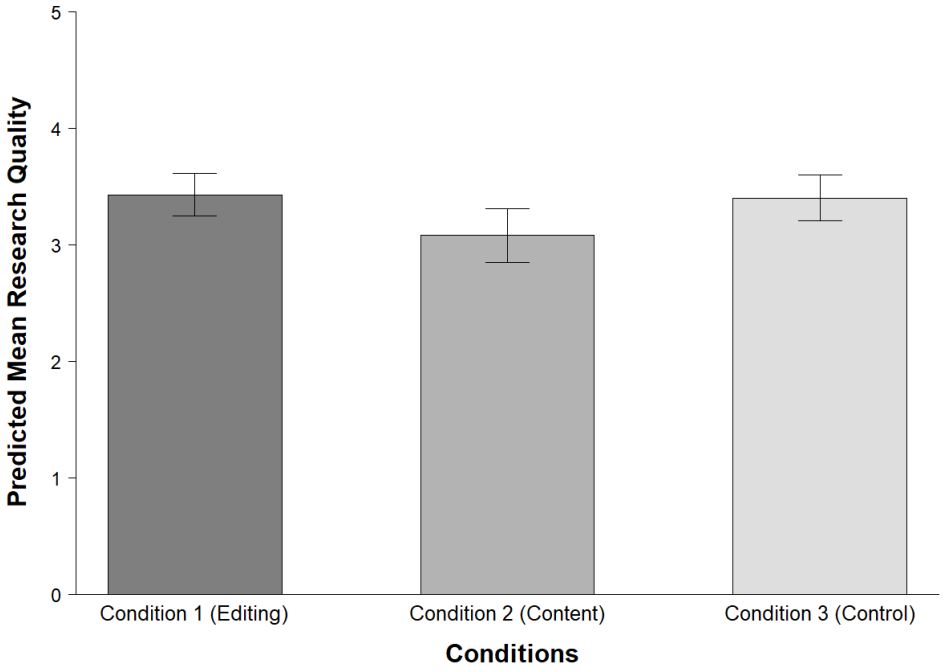
For testing the hypotheses and the research question, dummy variables for the two experimental conditions were created with the control condition (Condition 3) as a reference category. All models included the control variables age, gender, and social sciences.

Research Quality Ratings

To answer the first hypotheses, assuming that ChatGPT's use for editing has no negative impact on the quality ratings of the presented research (H1a), whereas its use for content and idea creation leads to lower quality ratings (H1b), multiple linear regression was performed for the dependent variable RQ, $F(6, 145) = 3.58$, $p = .002$, $R^2 = .09$ (see Table 2). Examining the predicted means of the regression shown in Figure 1, it can be observed that the predicted mean of participants' research quality perceptions for Condition 1, in which participants were presented with ChatGPT's assistance in editing ($M = 3.41$, 95% CI [3.23, 3.58]) was relatively equal to the predicted mean of the control condition, in which participants were not exposed to any ChatGPT involvement ($M = 3.42$, 95% CI [3.22, 3.61]). For Condition 2, in which participants were displayed to ChatGPTs assistance with content and idea creation ($M = 3.09$, 95% CI [2.87, 3.32]), the predicted mean showed lower quality perceptions compared to the control condition.

Figure 1

Bar Plots depicting Predicted Means of Research Quality Ratings per Condition



Note. Predicted means based on multiple linear regression. Error bars with 95% CI for the mean.

Confirming the previous observations of the predicted means, Table 2 shows that the results of the regression reveal no significant effect for Condition 1 (b = -0.01, $t(145)$ = -0.07, p = .946), while Condition 2 shows a significant effect (b = -0.33, $t(145)$ = -2.19, p = .030). This indicates that participants being displayed with ChatGPT's assistance in content and idea creation are estimated to be on average 0.33 units lower in their quality ratings compared to the control group, whereas participants being disclosed with ChatGPT's use for editing tasks revealed no significant differences in their quality ratings compared to the control group. There were no significant effects of the control variables and as a robustness check, the same analysis was conducted with a sample that included only social science students (N=118), yielding the same results (see Appendix D: Table 4). Therefore, H1a and H1b can both be confirmed.

Table 2

Multiple Linear Regression for the Effect of the Experimental Conditions (Condition 1, Condition 2) on Research Quality Ratings (RQ)

Variables	Estimate	*SE*	95% CI		*t*	*p*
			LL	*UL*		
Intercept	3.414	.628	2.174	4.654	5.441***	.000
Condition 1 (Editing) [a]	−.010	.133	−.272	.254	−.068	.946
Condition 2 (Content) [b]	−.327	.149	−.621	−.033	−2.194*	.030
ATP	.265	.084	.099	.431	3.156**	.002
Control Variables						
Gender	−.197	.126	−.446	.053	−1.560	.121
Age	−.015	.019	−.053	.022	−.798	.426
Social Sciences	−.131	.139	−.401	.143	−.946	.346

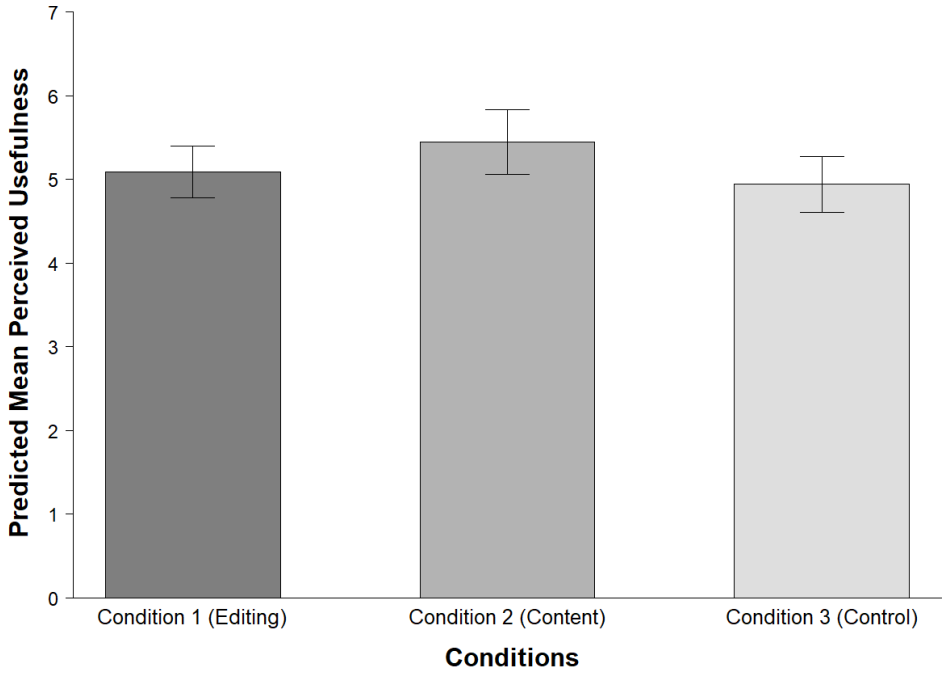
Note. R^2 adjusted = .09, N = 152; CI = confidence interval; *LL* = lower limit; *UL* = upper limit; ATP = attitudes towards positivism; model includes dummy variables for the experimental conditions: [a] 0 = Condition 3 (Control), 1 = Condition 1 (Editing). [b] 0 = Condition 3 (Control), 1 = Condition 2 (Content). *p < .05. ** p < .01 ***p < .001.

Perceived Usefulness Ratings

To answer the second hypotheses, assuming that ChatGPT would be considered more useful in editing (H2a) and less useful in idea and content creation (H2b), a multiple linear regression was performed for the dependent variable PU, $F(6, 145) = 3.11$, $p = .007$, $R^2 = .08$ (see Table 3). Examining the predicted means of the regression analysis shown in Figure 2, it appears that participants in Condition 1 (M = 5.05, 95% CI [4.75, 5.35]) showed similar predicted means of ChatGPT's usefulness perceptions as participants in the control condition (M = 4.97, 95% CI [4.65, 5.29]), whereas participants predicted means of PU in Condition 2 (M = 5.47, 95% CI [5.09, 5.84]) were higher compared to the control condition.

Figure 2

Bar Plots depicting Predicted Means of Perceived Usefulness Ratings of ChatGPT per Condition



Note. Predicted means based on multiple linear regression. Error bars with 95% CI for the mean.

Table 3 shows that the regression results for the perceived usefulness of ChatGPT were significant for Condition 2 (b = 0.50, $t(145)$ = 1.98, p = .049), but not for Condition 1 (b = 0.08, $t(145)$ = 0.37, p = .713). Therefore, participants showed no significant differences from the control group when they were displayed with ChatGPT's use for editing, but participants who saw ChatGPT's assistance in creating content and ideas are estimated to be on average 0.50 units higher in their perceived usefulness ratings compared to the control condition. There were no significant effects of the control variables and the robustness check conducted with only a social science sample (N=118) showed the same results (see Appendix D: Table 5). Since the significantly higher ratings of perceived usefulness for people being displayed with ChatGPT's function as an idea and content generator (Condition 2) reveal opposite results as assumed, H2a and H2b must be rejected.

Table 3

Multiple Linear Regression for the Effect of the Experimental Conditions on Perceived Usefulness Ratings of ChatGPT (PU)

Variables	Estimate	*SE*	95% CI		*t*	*p*
			LL	*UL*		
Intercept	3.450	1.056	1.362	5.537	3.266***	.000
Condition 1 (Editing) [a]	.082	.224	−.360	.525	.368	.713
Condition 2 (Content) [b]	.497	.251	.001	.992	1.982*	.049
ATP	.446	.141	.167	.725	3.158**	.002
Control Variables						
Gender	−.314	.212	−.733	.106	−1.479	.141
Age	.011	.032	−.052	.074	.337	.737
Social Sciences	.274	.233	−.187	.735	1.174	.243

Note. R^2 adjusted = .08, N = 152; CI = confidence interval; *LL* = lower limit; *UL* = upper limit; ATP = attitudes towards positivism; model includes dummy variables for the experimental conditions: [a] 0 = Condition 3 (Control), 1 = Condition 1 (Editing). [b] 0 = Condition 3 (Control), 1 = Condition 2 (Content).
* $p < .05$. **p < .01 ***$p < .001$.

The Effect of Attitudes Towards Positivism

To answer the research question (RQ1) about the moderating effect of participants' attitudes towards positivism (ATP), two moderation analyses for participants' ratings of research quality ($F(8, 143) = 2.82$, $p = .006$, $R^2 = .09$) and ChatGPT's perceived usefulness ($F(8, 143) = 2.35$, $p = .021$, $R^2 = .07$) as dependent variables were calculated. The results of the moderation analysis for research quality (see Appendix D: Table 6) show no significant interaction effects for the first ($b = 0.08$, 95% CI [-0.303, 0.452], $t = 0.39$, $p = .697$) and the second condition ($b = -0.19$, 95% CI [-0.666, 0.290], $t = -0.78$, $p = .438$), and the moderation analysis for participants' perceived usefulness (see Appendix D: Table 7) also reveals no significant interaction effects for

the first (b = 0.10, 95% CI [-0.535, 0.739], t = 0.32, p = .753) and the second condition (b = -0.14, 95% CI [-0.949, 0.665], t = -0.35 p = .728).

When looking at the direct effects of ATP on the two dependent variables (see Table 2 and Table 3), the results show significant positive direct effects of academics' positivism beliefs on research quality ratings (b = 0.27, t(143) = 3.16, p = .002) and perceived usefulness ratings of ChatGPT (b = 0.45, t(143) = 3.16, p = .002). Therefore, it can be said that participants with stronger attitudes towards positivism tend to rate the research quality and perceived usefulness of ChatGPT higher, regardless of condition, compared to participants with less positivist attitudes.

Discussion

The paper investigated the impact of acknowledgments indicating different involvements of ChatGPT in the research process, such as editing or content and idea creation, on young scientists' perceptions of research quality and usefulness of LLMs, and whether this is influenced by their level of positivistic thinking. Experimental evidence was found demonstrating differences in perceptions of specific language-based tasks of LLMs for scholarly publishing.

Research quality perceptions were found to decrease when ChatGPT was used for content or idea generation, but not when it was used for editing. Accordingly, the assumptions derived from the literature can be confirmed that the use of ChatGPT in more machine-like (Lee, 2018), already established AI tasks (Kurniati & Fithriani, 2022; O'Neill & Russel, 2019) such as proof-reading or grammar correction does not have a negative impact on the quality perception of academic research. ChatGPT’s assistance in more human-like tasks requiring nuanced judgments (Lee, 2018), such as an idea or content generation, on the other hand, is associated with lower research quality. These results are consistent with the skepticism expressed by some scholars (Bishop, 2023; Buriak et al., 2023; Sundar & Liao, 2023) that LLMs such as ChatGPT are

limited in developing truly original ideas. According to the results of this study regarding the quality perception of ChatGPT-assisted written texts, the literature of Arazy et al. (2017) can be agreed that text quality assessments might be guided by heuristic principles such as, in this case, "machine heuristics" (Sundar, 2020).

However, these heuristics do not seem to apply to the usefulness perceptions of ChatGPT. Contrary to our initial assumptions, the results of the present study show higher usefulness perceptions for ChatGPT in idea and content generation, but not in editing as expected. It seems that previous study results about usefulness perceptions of editing AI tools such as Grammarly (O'Neill & Russell, 2019) do not directly transfer to the usefulness perceptions of LLMs. A possible explanation could be that since the release of ChatGPT marked the first time that LLMs and their content and idea creation feature were accessible to a broad mass (Teubner et al., 2023), this new function of automatically generated content and ideas might therefore be perceived as much more useful than the older, already established editing function, which has already been performed by many other tools in the past, such as Grammarly or Quiltbot (Kurniati & Fithriani, 2022; O'Neill & Russell, 2019). These considerations are also consistent with the findings of a recent survey conducted by Owens (2023), in which most participants indicated that they use ChatGPT most often for brainstorming research ideas.

All in all, the findings of our study thus reveal a dilemma, as ChatGPT is perceived as a very useful tool for generating content and ideas in an academic context, but when it has been used and disclosed for this purpose in academic research, the research was evaluated with diminished quality. In contrast, using and disclosing ChatGPT as an editing assistance had no negative impact on the quality assessment of the research. These observations provide important practical implications for the use of LLMs in academic publishing. Language Models such as ChatGPT

can be openly acknowledged for their assistance in editing academic papers without researchers having to be concerned about sacrificing the quality assessment of their research. This can be a key advantage for non-native English speakers in academia, as they can produce higher quality papers for an international research community with the language editing support of LLMs (Jiao et al., 2023), but still, be transparent about their research process without any fear of a loss in the perception of research quality. The fact that LLMs are seen as more critical in terms of research quality when used in content and idea generation is also interesting for research practitioners, especially since language models such as ChatGPT are seen as very useful in this functional area by the young academics participating in this study and also by researchers such as Lund et al. (2023) or Salvagno et al. (2023). How and whether large language models can also be used and transparently documented in their function as content and idea-generation tools in science without reducing the quality of the presented research needs to be investigated in further studies.

While the study's findings do not show a moderating effect of participants' attitudes towards positivism for the two experimental conditions, they still indicate that pre-existing belief structures of academics likely play a role in LLM perceptions, as direct effects of the student's attitudes towards positivism were found on both dependent variables. Therefore, Sundar and Kim's (2019) view that preexisting belief structures of AI users influence perceptions towards AI tools can be agreed with. Considering that the study described in the stimulus material has a quantitative design, the effect that more positivistic individuals tend to rate the quality of the research shown in this study higher is not surprising, as positivistic beliefs are generally based on the use of quantitative approaches (Park et al., 2020). To our knowledge, there is no existing literature on the relationship between positivistic attitudes and perceptions of AI, so this study provides the first insights into the possible effect of these attitudes on AI perceptions. The fact that

no moderating effect of positivistic attitudes was found in this study does not necessarily indicate the absence of a moderating effect of ATP in general due to several limitations as the small sample size and the unevenly sized conditions in this study.

Since the planned sample size was originally 266 people for demonstrating a moderating effect, the actual sample size of this study is too small and therefore one of the major limitations of the present paper. In addition, the bot intrusion of the first data collection should be mentioned as another huge limitation, as the quality of the data may be affected by remaining bot responses in the dataset, limiting the conclusions that can be drawn from the results of this study. Due to the extensive data cleaning after the bot infiltration, the conditional groups varied considerably in size, which could also weaken the power of the study results. Although attempts were made to sort out the bots as accurately as possible and control variables were included in all analyses to counteract the effect of the different group sizes, the results must still be treated with caution, as the data quality in the present study can still be seen as a primary disadvantage. The low R-squared values of all models analyzed in this study should also be pointed out in the limitations of the study. According to Achen (1990), and Hagquist and Stenbeck (1998), R-squared is a less important measure in studies with conditional regressions in which the focus is not on the model but on single effects in the conditional groups. Another limitation of this paper, which is also mentioned in Lund et al.'s (2023) study, is that the results may only be valid at the time the study was conducted, as new language models with new features may emerge in the future, resulting in the need to revise the study from time to time. In addition, the sample consisting only of students who have limited experience in academic writing and academic work may limit the validity of the results of the present study. Therefore, individuals working in academia such as professors,

postdocs, and PhDs that have more experience in academic publishing should be recruited for future research to verify the results of the present study with a more adequate sample.

Other future research directions could include investigating ChatGPT's assistance with various tasks that are not so much language-based, but rather data-driven statistical tasks, such as assisting in generating code, as this is also mentioned as one of the current uses of ChatGPT for research-related tasks in Owens's (2023) survey on the chatbots use in academia. Other influencing variables reflecting prior belief systems that affect perceptions of AI (Sundar & Kim, 2019), such as individuals' openness to new technologies or prior experiences with LLMs, should be considered in future research in addition to the positivist attitudes examined in the present study. Since the present study focused mainly on social science academics and the evaluation of the perception of a research study in the social science field, the results are limited only to this discipline. Future research should therefore examine the perceptions of researchers in other disciplines or compare perceptions towards LLMs for academic writing among different disciplines. As the present study focuses only on the large language model ChatGPT, future research can as well investigate whether the present results on task-specific perception of LLMs are transferable or different to other LLMs such as BERT.

Conclusion

In sum, the paper is a response to the current discussions about LLMs in academia and the arising questions about how they should be credited and handled in scientific publications. Moving away from authorship to an assisting role in certain scientific language tasks (Bishop, 2023; Buriak et al., 2023; Lund et al., 2023; Salvagno et al., 2023), this paper follows the practical requirements of publishers not to list LLMs as authors, but to transparently disclose their involvement in the research process (Stokel-Walker, 2023). Given that the research subject of the

present study is based on the current practical circumstances of academic publishing, high external validity can be inferred in the present study as the findings are generalizable to real-world settings (Anderson & Bushman, 1997). The results of the online experiment revealed significant differences in quality and usefulness perceptions of LLMs based on their specific involvement in editing or content generation for the displayed research and should be interpreted with caution due to the limitations mentioned above. Nonetheless, this paper can be seen as a first step in filling research gaps regarding perceptions towards new emerging LLMs in an academic context and can be considered a valuable starting point for an evolving field of research.

References

Achen, C. H. (1990). What Does "Explained Variance" Explain?: Reply. *Political Analysis*, *2*(1), 173–184. https://doi.org/10.1093/pan/2.1.173

Anderson, C. A., & Bushman, B. J. (1997). External validity of "trivial" experiments: The case of laboratory aggression. *Review of General Psychology*, *1*(1), 19–41. https://doi.org/10.1037/1089-2680.1.1.19

Bishop, L. (2023). A Computer Wrote this Paper: What ChatGPT Means for Education, Research, and Writing. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4338981

Brainard, J., Killett, A., Houghton, J., Bunn, D., Watts, L., Mumford, S., O'Brien, S., & Lane, K. (2022). *The Wasps are Clever: Keeping Out and Finding Bot Answers in Internet Surveys Used for Health Research*. https://doi.org/10.20944/preprints202203.0243.v2

Buriak, J. M., Akinwande, D., Artzi, N., Brinker, C. J., Burrows, C., Chan, W. C. W., Chen, C., Chen, X., Chhowalla, M., Chi, L., Chueh, W., Crudden, C. M., Di Carlo, D., Glotzer, S. C., Hersam, M. C., Ho, D., Hu, T. Y., Huang, J., Javey, A., Kamat, P.V., Kim, I.D., Kotov, N.A., Lee, T.R., Lee, Y.H., Li, Y., Liz-Marzán, L.M., Mulvaney, P., Narang, P., Nordlander; P., Oklu, R., Parak, W.J., Rogach, A.L., Salanne, M., Samorì, P., Schaak, R.E, Schanze, K.S., Sekitani, T., Skrabalak, S., Sood, A.K., Voets, I.K., Wang, S., Wee, A.T.S., & Ye, J. (2023). Best Practices for Using AI When Writing Scientific Manuscripts. *ACS Nano*, *17*(5), 4091–4093. https://doi.org/10.1021/acsnano.3c01544

ChatGPT, & Zhavoronkov, A. (2022). Rapamycin in the context of Pascal's Wager: Generative pre-trained transformer perspective. *Oncoscience*, *9*, 82–84. https://doi.org/10.18632/oncoscience.571

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008

Elsevier (2023). *Engineering Applications of Artificial Intelligence. Guide for Authors*. Elsevier. https://www.elsevier.com/journals/engineering-applications-of-artificial-intelligence/0952-1976/guide-for-authors

Gao, C. A., Howard, F. M., Markov, N. S., Dyer, E. C., Ramesh, S., Luo, Y., & Pearson, A. T. (2022). Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers. *npj Digital Medicine, 6*(75). https://doi.org/10.1038/s41746-023-00819-6

Goertz, G., & Mahoney, J. (2012). *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*. Princeton University Press. https://doi.org/10.1017/S1537592713001862

Haensch, A., Ball, S., Herklotz, M., & Kreuter, F. (2023, March 9). *Seeing ChatGPT Through Students' Eyes: An Analysis of TikTok Data*. ArXiv. https://doi.org/10.48550/arXiv.2303.05349

Hagquist, C., & Stenbeck, M. (1998). Goodness of Fit in Regression Analysis – R2 and G2 Reconsidered. *Quality and Quantity, 32*(3), 229–245. https://doi.org/10.1023/A:1004328601205

Halaweh, M. (2023). ChatGPT in education: Strategies for responsible implementation. *Contemporary Educational Technology, 15*(2), ep421. https://doi.org/10.30935/cedtech/13036

Hosseini, M., Rasmussen, L. M., & Resnik, D. B. (2023). Using AI to write scholarly publications. *Accountability in Research*, 1–9. https://doi.org/10.1080/08989621.2023.2168535

International Communication Association. (n.d.). *Reviewer guidelines for conference submissions*. https://www.icahdq.org/members/group_content_view.asp?group=186109&id=633473

Jiao, W., Wang, W., Huang, J., Wang, X., & Tu, Z. (2023). *Is ChatGPT A Good Translator? Yes With GPT-4 As The Engine*. ArXiv. http://arxiv.org/abs/2301.08745

Joshi, A. K. (1977). Natural Language Processing. *Science, 253(5025)*. 1242-1249. https://doi.org/10.1126/science.253.5025.1242

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J., Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kim, J., Merrill Jr., K., & Collins, C. (2021). AI as a Friend or Assistant: The Mediating Role of Perceived Usefulness in Social AI vs. Functional AI. *Telematics and Informatics*, *64*, 101694. https://doi.org/10.1016/j.tele.2021.101694

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

Kurniati, E., & Fithriani, R. (2022). Post-Graduate Students' Perceptions of Quillbot Utilization in English Academic Writing Class. *Journal of English Language Teaching and Linguistics*, *7*(3), 437-451. https://doi.org/10.21462/jeltl.v7i3.852

Kutela, B., Msechu, K., Das, S., & Kidando, E. (2023). Chatgpt's Scientific Writings: A Case Study on Traffic Safety. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4329120

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data and Society*, *5*(1). https://doi.org/10.1177/2053951718756684

Lindgreen, A., Di Benedetto, C. A., & Brodie, R. J. (2021). Research quality: What it is, and how to achieve it. *Industrial Marketing Management*, *99*, A13–A19. https://doi.org/10.1016/j.indmarman.2021.10.009

Lund, B., Ting, W., Mannuru, N. R., Nie, B., Shimray, S., & Wang, Z. (2023). ChatGPT and a New Academic Reality: Artificial Intelligence-Written Research Papers and the Ethics of the Large Language Models in Scholarly Publishing. *SSRN Electronic Journal*.. https://doi.org/10.2139/ssrn.4389887

Megahed, F. M., Chen, Y.-J., Ferris, J. A., Knoth, S., & Jones-Farmer, L. A. (2023). *How Generative AI models such as ChatGPT can be (Mis)Used in SPC Practice, Education, and Research? An Exploratory Study*. ArXiv. http://arxiv.org/abs/2302.10916

Min, B., Ross, H., Sulem, E., Veyseh, A. P. B., Nguyen, T. H., Sainz, O., Agirre, E., Heinz, I., & Roth, D. (2021). *Recent advances in natural language processing via large pre-trained language models: A survey*. ArXiv. https://doi.org/10.48550/arXiv.2111.01243

Molina, M. D., & Sundar, S. S. (2022). Does distrust in humans predict greater trust in AI? Role of individual differences in user responses to content moderation. *New Media & Society*. https://doi.org/10.1177/14614448221103534

Nature Editorials. (2023). Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature*, *613*, 612. https://doi.org/10.1038/d41586-023-00191-1

O'Connor, S. & ChatGPT. (2023). Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse? *Nurse Education in Practice*, *66*, 103537. https://doi.org/10.1016/j.nepr.2022.103537

O'Neill, R., & Russell, A. M. T. (2019). Grammarly: Help or hindrance? Academic Learning Advisors' perceptions of an online grammar checker. *Journal of Academic Language and Learning*, *13*(1), 88-107.

OpenAI. (2023). *OpenAI GPT-4*. https://openai.com/research/gpt-4

Owens, B. (2023). How Nature readers are using ChatGPT. *Nature*, *615*(7950), 20. https://doi.org/10.1038/d41586-023-00500-8

Park, D. Y., & Kim, H. (2023). Determinants of Intentions to Use Digital Mental Healthcare Content among University Students, Faculty, and Staff: Motivation, Perceived Usefulness, Perceived Ease of Use, and Parasocial Interaction with AI Chatbot. *Sustainability*, *15*(1), 872. https://doi.org/10.3390/su15010872

Park, Y. S., Konge, L., & Artino, A. R. (2020). The Positivism Paradigm of Research: *Academic Medicine*, *95*(5), 690–694. https://doi.org/10.1097/ACM.0000000000003093

Pourhoseingholi, M. A., Hatamnejad, M. R., & Solhpour, A. (2023). Does chatGPT (or any other artificial intelligence language tool) deserve to be included in authorship list? *Gastroenterology and Hepatology from Bed to Bench*, *16*(1). https://doi.org/10.22037/ghfbb.v16i1.2747

Rudolph, J., Samson, T., & Shannon, Tan. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching*, *6*(1). https://doi.org/10.37074/jalt.2023.6.1.9

Rutjens, B. T., & van der Lee, R. (2020). Spiritual skepticism? Heterogeneous science skepticism in the Netherlands. *Public Understanding of Science, 29*(3), 335–352. https://doi.org/10.1177/0963662520908534

Rutjens, B. T., Sengupta, N., der Lee, R. van, van Koningsbruggen, G. M., Martens, J. P., Rabelo, A., & Sutton, R. M. (2022). Science Skepticism Across 24 Countries. *Social Psychological and Personality Science, 13*(1), 102–117. https://doi.org/10.1177/19485506211001329

Salvagno, M., ChatGPT, Taccone, F. S., & Gerli, A. G. (2023). Can artificial intelligence help for scientific writing? *Critical Care*, *27*(1), 75. https://doi.org/10.1186/s13054-023-04380-2

Scheitle, C. P., & Corcoran, K. E. (2021). COVID-19 Skepticism in Relation to Other Forms of Science Skepticism. *Socius: Sociological Research for a Dynamic World, 7*, https://doi.org/10.1177/23780231211049841

Sherman, S. J., & Corty, E. (1984). Cognitive heuristics. In *Handbook of social cognition, Vol 1*. (pp.189–286). Lawrence Erlbaum Associates Publishers.

Steel, B., List, P., Lach, D., & Shindler, B. (2004). The role of scientists in the environmental policy process: A case study from the American west. *Environmental Science & Policy*, *7*(1), 1–13. https://doi.org/10.1016/j.envsci.2003.10.004

Stokel-Walker, C. (2023). ChatGPT listed as author on research papers: Many scientists disapprove. *Nature*, *613*(7945), 620–621. https://doi.org/10.1038/d41586-023-00107-z

Sundar, S. S. (2020). Rise of Machine Agency: A Framework for Studying the Psychology of Human–AI Interaction (HAII). *Journal of Computer-Mediated Communication*, *25*(1), 74–88. https://doi.org/10.1093/jcmc/zmz026

Sundar, S. S., & Kim, J. (2019). Machine Heuristic: When We Trust Computers More than Humans with Our Personal Information. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–9. https://doi.org/10.1145/3290605.3300768

Sundar, S. S., & Liao, M. (2023). Calling BS on ChatGPT: Reflections on AI as a Communication Source. *Journalism & Communication Monographs*, *25*(2), 165–180. https://doi.org/10.1177/15226379231167135

Teubner, T., Flath, C. M., Weinhardt, C., van der Aalst, W., & Hinz, O. (2023). Welcome to the Era of ChatGPT et al. *Business & Information Systems Engineering*, *65*(2), 95–101. https://doi.org/10.1007/s12599-023-00795-x

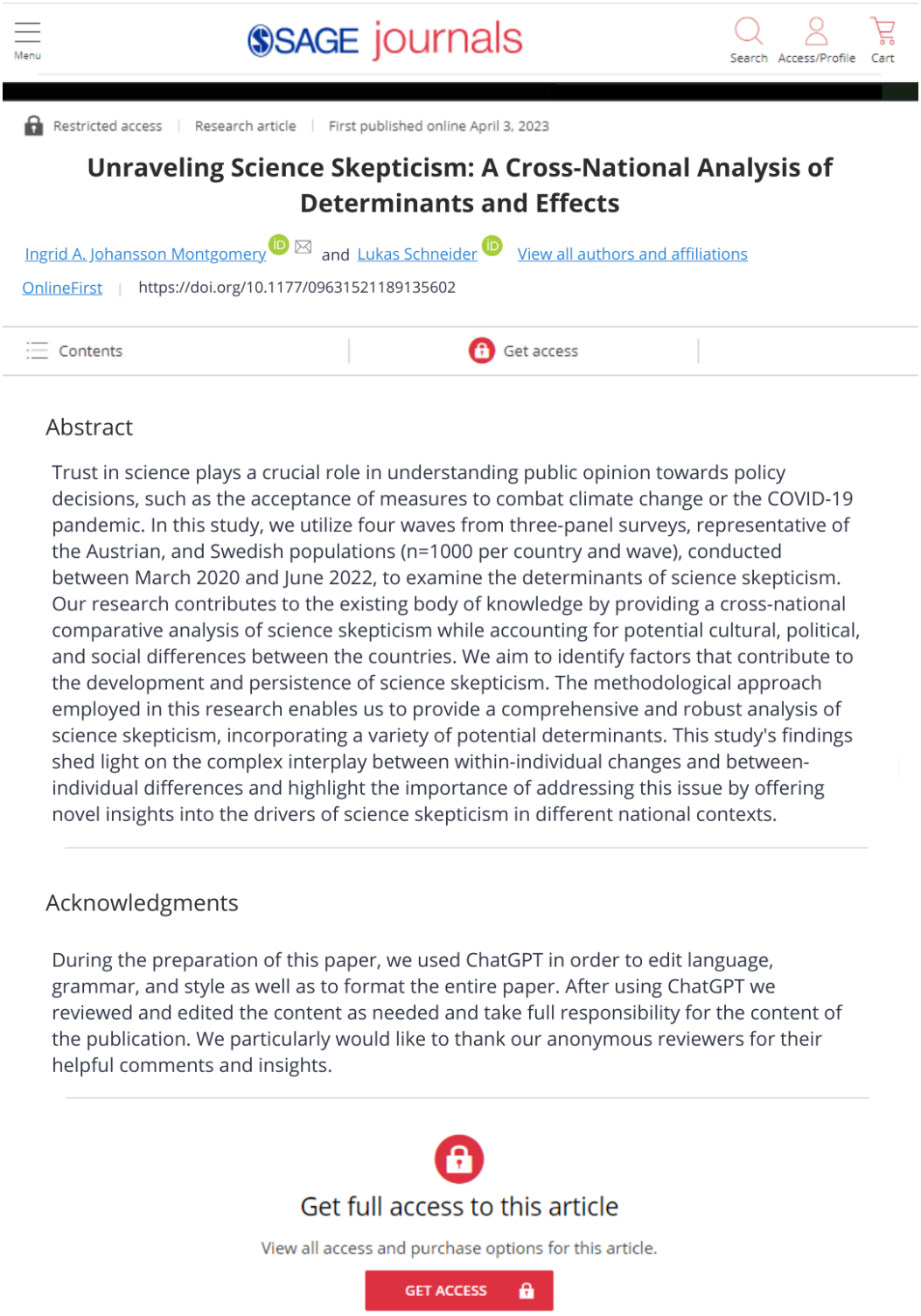
Vaswani, A., Shazeer, N. M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is All you Need. *Advances in Neural Information Processing Systems*, 5998–6008.

Većkalov, B., Zarzeczna, N., McPhetres, J., van Harreveld, F., & Rutjens, B. T. (2022). Psychological Distance to Science as a Predictor of Science Skepticism Across Domains. *Personality & Social Psychology Bulletin*, 14616722211181–1461672221118184. https://doi.org/10.1177/01461672221118184

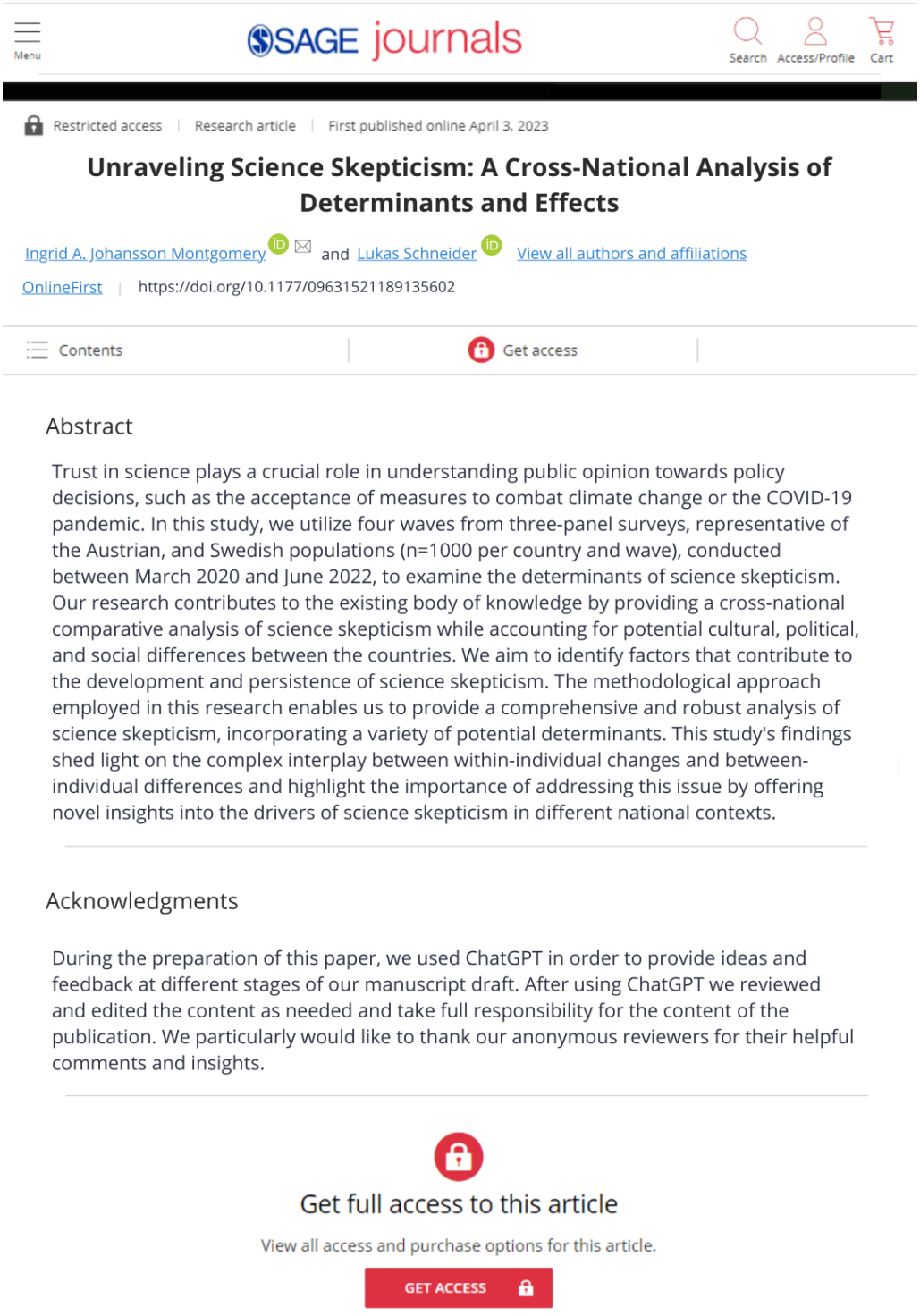
Yeo-Teh, N. S. L., & Tang, B. L. (2023). Letter to editor: NLP systems such as ChatGPT cannot be listed as an author because these cannot fulfill widely adopted authorship criteria. *Accountability in Research*, 1–3. https://doi.org/10.1080/08989621.2023.2177160

Appendix

Appendix A: Stimulus Material



Condition 1: Acknowledgments stating ChatGPT's involvement in the editing of the paper



Menu SAGE journals Search Access/Profile Cart

Restricted access | Research article | First published online April 3, 2023

Unraveling Science Skepticism: A Cross-National Analysis of Determinants and Effects

Ingrid A. Johansson Montgomery and Lukas Schneider View all authors and affiliations

OnlineFirst | https://doi.org/10.1177/09631521189135602

Contents Get access

Abstract

Trust in science plays a crucial role in understanding public opinion towards policy decisions, such as the acceptance of measures to combat climate change or the COVID-19 pandemic. In this study, we utilize four waves from three-panel surveys, representative of the Austrian, and Swedish populations (n=1000 per country and wave), conducted between March 2020 and June 2022, to examine the determinants of science skepticism. Our research contributes to the existing body of knowledge by providing a cross-national comparative analysis of science skepticism while accounting for potential cultural, political, and social differences between the countries. We aim to identify factors that contribute to the development and persistence of science skepticism. The methodological approach employed in this research enables us to provide a comprehensive and robust analysis of science skepticism, incorporating a variety of potential determinants. This study's findings shed light on the complex interplay between within-individual changes and between-individual differences and highlight the importance of addressing this issue by offering novel insights into the drivers of science skepticism in different national contexts.

Acknowledgments

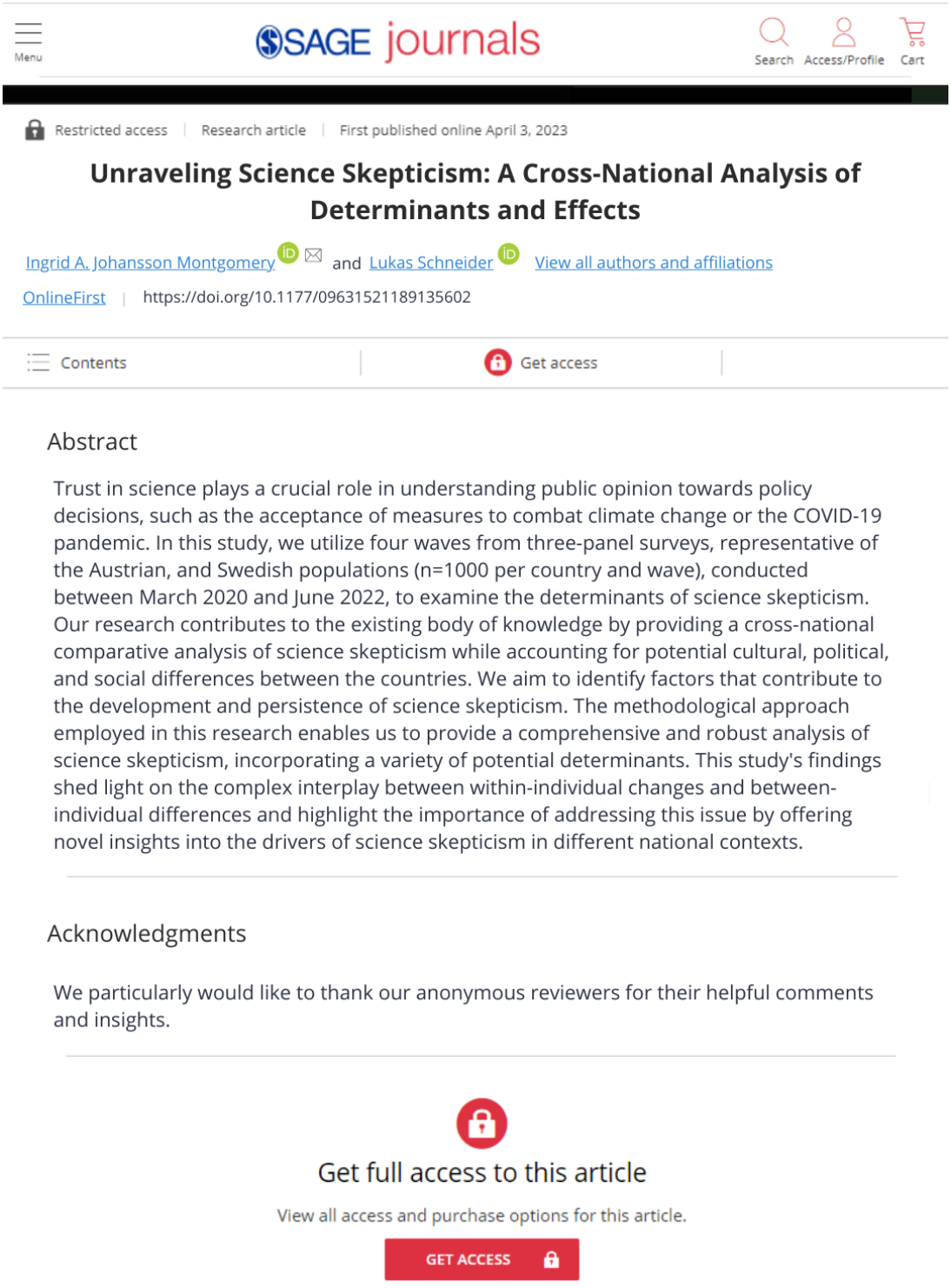
During the preparation of this paper, we used ChatGPT in order to provide ideas and feedback at different stages of our manuscript draft. After using ChatGPT we reviewed and edited the content as needed and take full responsibility for the content of the publication. We particularly would like to thank our anonymous reviewers for their helpful comments and insights.

Get full access to this article

View all access and purchase options for this article.

GET ACCESS

Condition 2: Acknowledgments stating ChatGPT's involvement in the content of the paper



Condition 3: Acknowledgments without ChatGPT's involvement in the paper

Appendix B: ChatGPT Prompts for Stimulus Material Generation

We are going to write an abstract to use as stimulus for an experimental study. The abstract is going on the topic of science scepticism from a social science perspective. I am going to feed you some information about it. Okey?

Do you know what is important in order to create a good abstract for an academic publication in the field of social sciences?

You will use academic language when writing this abstract.

Please answer Yes, sir. If you understand.

Here is an example of an abstract on the same topic.

"Efforts to understand and remedy the rejection of science are impeded by lack of insight into how it varies in degree and in kind around the world. The current work investigates science skepticism in 24 countries (N = 5,973). Results show that while some countries stand out as generally high or low in skepticism, predictors of science skepticism are relatively similar across countries. One notable effect was consistent across countries though stronger in Western, Educated, Industrialized, Rich, and Democratic (WEIRD) nations: General faith in science was predicted by spirituality, suggesting that it, more than religiosity, may be the 'enemy' of science acceptance. Climate change skepticism was mainly associated with political conservatism especially in North America. Other findings were observed across WEIRD and non-WEIRD nations: Vaccine skepticism was associated with spirituality and scientific literacy, genetic modification skepticism with scientific literacy, and evolution skepticism with religious orthodoxy. Levels of science skepticism are heterogeneous across countries, but predictors of science skepticism are heterogeneous across domains."(Rutjens et al., 2022)

Here is another example of an abstract on the same topic.

"This article presents and tests psychological distance to science (PSYDISC) as a domain-general predictor of science skepticism. Drawing on the concept of psychological distance, PSYDISC reflects the extent to which individuals perceive science as a tangible undertaking conducted by people similar to oneself (social), with effects in the here (spatial) and now (temporal), and as useful and applicable in the real world (hypothetical distance). In six studies (two preregistered; total N = 1,630) and two countries, we developed and established the factor structure and validity of a scale measuring PSYDISC. Crucially, higher PSYDISC predicted skepticism beyond established predictors, across science domains. A final study showed that PSYDISC shapes real-world behavior (COVID-19 vaccination uptake). This work thus provides a novel tool to predict science skepticism, as well as a construct that can help to further develop a unifying framework to understand science skepticism across domains."(Većkalov et al., 2022)

I have more information. Just answer with READ.

Here is another example of an abstract on the topic.

"Efforts to address the coronavirus disease 2019 (COVID-19) pandemic have encountered skepticism among the public, but COVID-19 is not the only medical or scientific issue that receives such skepticism. How does COVID-19 skepticism relate to other forms of science skepticism? Using new data from a nationally representative survey of U.S. adults, this study reveals that skepticism toward COVID-19 is

similar to patterns of skepticism toward vaccines in general and, more interestingly, skepticism toward climate change. Patterns of skepticism toward evolution and genetically modified foods are more distinct from COVID-19 skepticism. Notably, even after accounting for other forms of science skepticism, political conservatism is significantly associated with greater skepticism toward COVID-19. Finally, contrary to some media narratives, the analysis reveals few racial or ethnic differences in skepticism toward COVID-19, and the differences that do exist indicate less skepticism among Black and Asian individuals relative to White individuals." (Scheitle & Corcoran, 2021)

Here is another example of an abstract on the same topic.

"Recent work points to the heterogeneous nature of science skepticism. However, most research on science skepticism has been conducted in the United States. The current work addresses the generalisability of the knowledge acquired so far by investigating individuals from a Western European country (The Netherlands). Results indicate that various previously reported findings hold up: Mirroring North American patterns, climate change skepticism is associated with political conservatism (but only modestly), and scientific literacy does not contribute to skepticism, except about genetic modification (Study 1 only) and vaccine skepticism (Study 2 only). Results also reveal a crucial difference: Religiosity does not consistently contribute to science skepticism, except about evolution. Instead, spirituality is found to most consistently predict vaccine skepticism and low general faith in science—which in turn predicts willingness to support science. Concerns about societal impact play an additional role. These findings speak to the generalisability of previous findings, improving our understanding of science skepticism." (Rutjens & van der Lee, 2020)

I have more information. Just answer with READ.

Here are bullet points for an abstract I would like to write today.

The paper that the abstract is about is about science scepticism.
Start the abstract by saying something about the societal relevance of this research.
Then say something about how important trust in science is for democracy.
Then maybe add an example of why this is the case.
Only afterwards, say something about the method. E.g., I am using four survey waves from two panel surveys representative of the Austria, Italian, and Swedish population (n=1000 per country and wave).
The field work was done between March 2020 and October 2022.
The aim of the study is to identify determinants and effects of science skepticism.
Don't forget to say something about the contribution to the existing research in the field.
Do not mention the pandemic, COVID, or artificial intelligence.
Think of something that may sound novel and interesting.

Please draft an abstract of between 300 and 350 words and consisting of only one paragraph, using these notes and the examples of the previous abstracts for reference with regards to formatting and writing style.

Appendix C: Questionnaire

Block A: Warm-up/ Screening

1. Introduction (IN):

Dear participant,

Thank you very much for your interest and welcome to this study that is being conducted as part of a master's thesis in the Communication Science program at the University of Vienna!

Welcome to the questionnaire! To take part in the survey, you must meet the following criteria:

- Currently **studying or employed at a university or university-related teaching institution**
- Proficiency in **English**

This study aims to explore your perspectives and perceptions regarding an academic journal's web-site. The survey is **completely anonymous** and takes **approximately 5 minutes** to complete. We want to emphasize that there are **no wrong answers.** Your opinions, perspectives, and perceptions are highly valued and will contribute significantly to the study.

As a token of our appreciation, you have the opportunity to win an **Amazon voucher worth 50€**. If you wish to participate in the lottery, please **provide your email address** at the end of the survey. Please note that your **email address will be kept separate from your answers** and will not be linked to them in any way.

On the next page, you will be informed about your rights as a participant. If you have any further questions about the study, please feel free to contact Hannah Kronschnabl via email at hannah.maria.kronschnabl@univie.ac.at.

----------page break----------

2. Informed Consent

Before you start the questionnaire, we would like to inform you about your rights.

Your data will be collected and processed exclusively based on the Austrian legal regulations (§ 2f Abs 5 FOG), which are in line with the General Data Protection Regulation (GDPR). You have the following personal rights in the context of this survey:

- Participation in the survey is **voluntary**. You can cancel the questionnaire at any time without any consequences.
- Your participation is **anonymous**, your answers cannot be traced back to you.
- This also means that your personal data set is **not identifiable** to us after the survey is completed. If you would like to have information about your data after the study or if you would like to withdraw your participation, we ask you to note this in the final comment field (together with a contact address, if necessary).
- Your data will be used for **scientific purposes only**. The research does not follow any commercial interest.
- We treat all data **strictly confidential.**

If you have any questions about the content, purpose, or research ethics of this survey, please contact Hannah Kronschnabl (hannah.maria.kronschnabl@univie.ac.at).

In order for you to participate in this study, we now need your informed consent:

- I have been informed of my rights in the study and now wish to proceed.

- -page break- -

3. Are you studying or working at a university or in a university-related teaching institution?

1. Yes
2. No (Finish Survey)
3. Maybe (Specify: _____)

- -page break- -

To begin, please answer the following questions related to your experience at academic institutions.

4. As of summer 2023, which of the following groups would you say you predominantly belong to? (Single choice)

If you belong to more than one group, choose the predominant one. You can only tick one answer option.

a. Bachelor Student
b. Master Student
c. Doctoral Student/Researcher
d. Post Doctoral Researcher
e. Professor

5. What is your current discipline or field of study/research? (Single Choice)

If more than one discipline fits your study/research field, please indicate the one you feel most connected to.

[RANDOMIZE ITEMS]

a. Social Sciences (Anthropology, Communication & Media Studies, Psychology, Political Science, Sociology, etc.)
b. Natural Sciences (Biology, Chemistry, Physics, etc.)
c. Humanities (Film Studies, History, Language Studies, Literature, Philosophy, Religion, etc.)
d. Engineering and Technical Sciences
e. Medicine and Health Sciences
f. Agriculture and Environmental Sciences
g. Computer Science and Information Technology
h. Mathematics and Statistics
i. Economics, Business, and Management
j. Law and Legal Studies
k. Arts and Fine Arts
l. Physical Education and Sports Sciences
m. Education Sciences

n. Other

- -page break- -

Block B: General Attitudes toward Science (Positivism-Scale, Moderator)

6. We would like to ask you to provide some information about your general attitudes toward science. Please indicate how much you agree or disagree with each of the following statements below.

   [RANDOMIZE ITEMS]

   a. The use of the scientific method is the only certain way to determine what is true or false about the world.
   b. The advancement of knowledge is a linear process driven by key experiments.
   c. Science provides objective knowledge about the world.
   d. It is possible to eliminate values and value judgments from the interpretation of scientific data.
   e. Facts describe true states of affairs about the world.

   Matrix-Labels: (Single Choice)
   - *1 = Strongly disagree*
   - *2 = Disagree*
   - *3 = Neutral/Neither agree nor disagree*
   - *4 = Agree*
   - *5 = Strongly agree*

- -page break- -

Block C: Stimulus Presentation

7. Below you can see an entry from an academic journal's website. Please **take a close look** at the **Authors, Title, Abstract, and Acknowledgments**, and **read everything carefully** to answer the follow-up questions on the next page. **Take your time** as the button to move to the next page will only appear in **approximately 30 seconds**.

   (IN BETWEEN_DESIGN:
   RANDOMIZED PRESENTATION STIMULUS MATERIAL OF CONDITIONS 1,2 OR 3)

("Next Page"-Button appears after 30 seconds)
- -page break- -

Block D: Dependent Variables and Manipulation check

[SPLIT DESIGN: RANDOMIZATION OF BLOCKS D1 AND D2]

Block D1: Dependent Variables

8. DV1: Research Quality:

Based on the information that was just provided to you, please rate the research presented in the entry on the academic journal's website in terms of several points asked in the following questions. (Single Choice)

a. How do you evaluate the theoretical contribution of the research (paper)?

Slider-Labels:
- *1 = Lacking theoretical depth (Minimum)*
- *2*
- *3*
- *4*
- *5 = Important theoretical contribution (Maximum)*

b. How clearly do the authors present the overall argument?

Slider-Labels:
- *1 = Not clear at all (Minimum)*
- *2*
- *3*
- *4*
- *5 = Very clear (Maximum)*

c. How innovative is the research?

Slider-Labels:
- *1 = Not significant (Minimum)*
- *2*
- *3*
- *4*
- *5 = Significant (Maximum)*

d. How appropriate is the methodological approach?

Slider-Labels:
- *1 = Not appropriate at all (Minimum)*
- *2*
- *3*
- *4*
- *5 = Very appropriate (Maximum)*

e. What is the overall impression of the research?

Slider-Labels:
- *1 = Low (Minimum)*
- *2*
- *3*

- *4*
- *5 = High (Maximum)*

- -page break- -

9. DV2: Perceived Usefulness (PU)

We would like to gather your insights on how the large language model ChatGPT could potentially assist you in your studies or research work. Please evaluate the following statements.
(Single Choice)

[RANDOMIZE ITEMS]

a. Using ChatGPT in my studies/research work would enable me to accomplish tasks more quickly.
b. Using ChatGPT in my studies/research work would improve my performance in academic tasks.
c. Using ChatGPT in my studies/research work would increase my productivity.
d. Using ChatGPT in my studies/research work would enhance my effectiveness on academic tasks.
e. Using ChatGPT would make it easier to do my studies/research work.
f. I would find ChatGPT useful for my studies/research work.

Matrix-Labels:
- *1 = Extremely unlikely*
- *2 = Quite unlikely*
- *3 = Slightly unlikely*
- *4 = Neither likely nor unlikely*
- *5 = Slightly likely*
- *6 = Quite likely*
- *7 = Extremely likely*
- *I don't know [88]*

- -page break- -

Block D2: Manipulation check

Please recall the entry on the academic journal's website you read at the beginning of the survey and answer the following questions related to it.

10. What did the authors mention in the acknowledgments of the research paper?

[RANDOMIZE ITEMS]

a. They thanked the anonymous reviewers.
b. They stated that they used **ChatGPT** for formatting and language, grammar, and style editing.
c. They stated that they used **ChatGPT** for providing ideas and feedback.
d. They stated that they used **ChatGPT** for debugging and writing code.

Matrix-Labels:

- *1 = Yes*
- *2 = No*
- *I don't know [88]*

11. How much do you agree with the following statements?

[RANDOMIZE ITEMS]

a. Entries similar to this one can be found on academic journal websites.
b. Research similar to this one can be found published in academic journals.

Matrix-Labels:

- *1 = Strongly agree*
- *2 = Agree*
- *3 = Neutral/Neither agree nor disagree*
- *4 = Agree*
- *5 = Strongly agree*
- *I don't know [88]*

- -page break- -

Block E: Sociodemographics

Please answer a few general questions about yourself:

12. What is your gender? (Singe Choice)

a. Male
b. Female
c. Other
d. Prefer not to say [99]

13. How old are you?

<…>

(Under 18 years old and over 99 years >>> Cancel Survey)

- -page break- -

Block F: Debriefing

14. **Thank you for taking part in our study!**

You have greatly helped us by participating in this survey. Because you participated in this study, we can improve our understanding of science communication. The purpose of this study is to examine the **impact of specific acknowledgments** on **research quality ratings** and the **perceived usefulness**

of large language models such as ChatGPT. These acknowledgments varied depending on the treatment group and indicated **ChatGPT's involvement in editing or content creation** of the paper draft.

The entry on the academic journal's website you encountered at the beginning of this survey **was specifically designed and developed for this purpose** and has never been published. The **title and the abstract** you saw were **created by ChatGPT** and the described research was **never conducted and has not been published** in SAGE Journals. Furthermore, the **acknowledgments** themselves were **manipulated for this study.**

Please keep this information about the purpose of the study to yourself and do not share it with others, who may also participate in the study. If you have questions or comments regarding the study, please contact (hannah.maria.kronschnabl@univie.ac.at) or add them below. To participate in the lottery, please provide your email address on the following page.

Comments:
(Comment field)

- -page break- -

Block G: Lottery+ E-Mail Contact

15. **Lottery: Win a 50€ Amazon Voucher!**

We appreciate your participation in our survey and would like to offer you a chance to win an Amazon voucher worth 50€. To enter the lottery, simply **provide your email address in the space below**. Please note that your contact information will be treated **confidentially** and will only be used for the purpose of conducting the lottery. Your email address will be kept separate from your answers and will not be linked to them in any way. Once the winner is selected, all **contact data will be promptly deleted**.

By entering your email address, you agree to participate in the lottery and acknowledge that the **winner will be chosen randomly**. The winner will be **notified via email**, so please ensure that the provided email address is accurate.

We value your privacy and assure you that your contact information **will not be shared with any third parties**. It will be **securely stored** and used exclusively for the purpose stated above.

Thank you for your participation and good luck in the lottery!

Email Address: [Insert Email Address Field]

[Submit Button]

Appendix D: Tables

Table 4

Multiple Linear Regression for the Effect of the Experimental Conditions on Research Quality Ratings (RQ) including only Social Science Students

Variables	Estimate	*SE*	95% CI		*t*	*p*
			LL	*UL*		
Intercept	3.013	.646	1.732	4.293	4.661 ***	.000
Condition 1 (Editing) [a]	−.061	.145	−.349	.227	−.419	.676
Condition 2 (Content) [b]	−.476	.165	−.802	−.149	−2.888 **	.005
ATP	.213	.091	.033	.392	2.348*	.021
Control Variables						
Age	−.003	.021	−.044	.038	−.154	.878
Gender	−.077	.141	−.355	−.202	−.546	.586

Note. R^2 adjusted = .09, N = 118; CI = confidence interval; *LL* = lower limit; *UL* = upper limit; ATP = attitudes towards positivism; model includes dummy variables for the experimental conditions: [a] 0 = Condition 3 (Control), 1 = Condition 1 (Editing). [b] 0 = Condition 3 (Control), 1 = Condition 2 (Content). **p < .01 ***p < .001.

Table 5

Multiple Linear Regression for the Effect of the Experimental Conditions on the Perceived Usefulness of ChatGPT (PU) including only Social Science Students

Variables	Estimate	*SE*	95% CI		*t*	*p*
			LL	*UL*		
Intercept	3.098	1.147	.826	5.369	2.702**	.008
Condition 1 (Editing) [a]	.164	.258	−.348	.675	.633	.528
Condition 2 (Content) [b]	.592	.292	.013	1.171	2.025*	.045

ATP	.490	.161	−.172	.808	3.053**	.003
Control Variables						
Age	.015	.037	−.058	.087	.398	.691
Gender	−.127	.249	−.621	.366	−.511	.611

Note. R^2 adjusted = .08, *N* = 118; CI = confidence interval; *LL* = lower limit; *UL* = upper limit; ATP = attitudes towards positivism (mean-centered moderator variable). model includes dummy variables for the experimental conditions: [a] 0 = Condition 3 (Control), 1 = Condition 1 (Editing). [b] 0 = Condition 3 (Control), 1 = Condition 2 (Content). *p < .05 **p < .01.

Table 6

Moderation Analysis for the Effect of the Experimental Conditions on Research Quality Ratings (RQ) moderated through Attitudes towards Positivism (ATP)

Variables	Estimate	*SE*	95% CI		*t*	*p*
			LL	*UL*		
Intercept	4.323	.564	3.208	5.437	7.667 ***	.000
Condition 1 (Editing) [a]	−.017	.133	−.281	.247	−.125	.901
Condition 2 (Content) [b]	−.335	.150	−.631	−.038	−2.233 *	.027
ATP	.267	.136	−.001	.535	1.970 †	.051
Condition 1*ATP	.075	.191	−.303	.452	.390	.697
Condition 2*ATP	−.188	.242	−.666	.290	−.778	.438
Control Variables						
Age	−.015	.019	−.053	.023	−.782	.436
Gender	−.208	.129	−.462	.047	−1.609	.110
Social Sciences	−.143	.140	−.419	.133	−1.024	.308

Note. R^2 adjusted = .09, *N* = 152; CI = confidence interval; *LL* = lower limit; *UL* = upper limit; ATP is ATP = attitudes towards positivism (mean-centered moderator variable); model includes dummy variables for the experimental conditions: [a] 0 = Condition 3 (Control), 1 = Condition 1 (Editing). [b] 0 = Condition 3 (Control), 1 = Condition 2 (Content). † p < .10. * p < .05. ***p < .001.

Table 7

Moderation Analysis for the Effect of the Experimental Conditions on Perceived Usefulness Ratings of ChatGPT (PU) moderated through Attitudes towards Positivism (ATP)

Variables	Estimate	*SE*	95% CI		*t*	*p*
			LL	*UL*		
Intercept	4.975	.952	3.094	6.857	5.227 ***	.000
Condition 1 (Editing) [a]	.076	.225	−.370	.521	.335	.738
Condition 2 (Content) [b]	.492	.253	−.009	.992	1.943 †	.054
ATP	.429	.229	−.024	.881	1.873 †	.063
Condition 1*ATP	.102	.322	−.535	.739	.316	.753
Condition 2*ATP	−.142	.408	−.949	.665	−.349	.728
Control Variables						
Age	.010	.033	−.054	.075	.313	.755
Gender	−.328	.218	−.759	.102	−1.508	.134
Social Sciences	.264	.236	−.202	.730	1.121	.264

Note. R^2 adjusted = .07, N = 152; CI = confidence interval; *LL* = lower limit; *UL* = upper limit; ATP = attitudes towards positivism (mean-centered moderator variable); model includes dummy variables for the experimental conditions: [a] 0 = Condition 3 (Control), 1 = Condition 1 (Editing). [b] 0 = Condition 3 (Control), 1 = Condition 2 (Content). † $p < .10$. ***$p < .001$.

Appendix E: Abstracts in English and German

Abstract

The growing emergence of large language models (LLMs) such as ChatGPT in academia has sparked a debate in the research community about their use for scientific publishing. This study follows practical instructions of recently adapted publication guidelines to disclose any assistance of LLMs in the acknowledgment sections of research papers. The present study is the first one to explore the impact of acknowledgments indicating different involvements of ChatGPT in the research process on young scientists' perceptions of research quality and the usefulness of large language models, and whether this is influenced by their level of positivistic thinking. An experimental approach was employed to analyze students' (N = 152) perceptions of an academic journal entry showing ChatGPT's assistance in editing or generating content and ideas. The results showed that quality perceptions of research decreased when ChatGPT was used for content or idea generation, but not when it was used for editing. The perceived usefulness of ChatGPT, on the other hand, tended to be rated higher for content and idea generation. No moderating effects of positivistic attitudes could be found.

Keywords: Artificial Intelligence, Large Language Models, ChatGPT, chatbots, scientific writing, research quality, science communication

Zusammenfassung

Der wachsende Einsatz großer Sprachmodelle (LLMs) wie ChatGPT im akademischen Bereich hat in der Forschungsgemeinschaft eine Debatte über deren Verwendung in wissenschaftlichen Publikationen ausgelöst. Die vorliegende Studie folgt praktischen Anweisungen der kürzlich angepassten Publikationsrichtlinien wissenschaftlicher Verlage, die fordern jegliche Unterstützung durch LLMs in den Danksagungen von Forschungsarbeiten offenzulegen. Diese Studie ist die erste, die Danksagungen, in welchen auf eine Beteiligung von ChatGPT in verschiedenen Bereichen im Forschungsprozess hingewiesen wird, ansieht. Es wird deren Einfluss auf die Wahrnehmung junger WissenschaftlerInnen bezüglich Forschungsqualität und der Nützlichkeit von LLMs geprüft, und untersucht ob dies von deren positivistischen Denken abhängig ist. Mithilfe eines experimentellen Ansatzes wurde die Wahrnehmung der Studierenden (N = 152) eines akademischen Journal-Eintrages analysiert, der auf Unterstützung von ChatGPT zum Redigieren oder zur Generierung von Inhalten und Ideen hinweist. Die Ergebnisse zeigen, dass die Qualitätswahrnehmungen der Forschung abnahmen, wenn ChatGPT für den Inhalt oder die Ideengenerierung verwendet, nicht aber, wenn es zum Redigieren herangezogen wurde. Die wahrgenommene Nützlichkeit des Sprachmodells hingegen wurde tendenziell höher bewertet, wenn das LLM zur Generierung von Inhalte und Ideen verwendet wurde. Es konnten keine moderierenden Effekte positivistischer Einstellungen gefunden werden.

Schlagwörter: Künstliche Intelligenz, Große Sprachmodelle, ChatGPT, Chatbots, Wissenschaftliches Schreiben, Forschungsqualität, Wissenschaftskommunikation