



AI-empowered scale development: Testing the potential of ChatGPT

Stefan Hoffmann^{a,*}, Wassili Lasarov^b, Yogesh K. Dwivedi^{c,d}

^a Kiel University, Department of Marketing, Kiel Institute for Responsible Innovation, Westring 425, 24098 Kiel, Germany

^b Audencia Business School, Marketing Department, 8 Route de la Jonelière B.P. 31222, 44312 Nantes Cedex 3, France

^c Digital Futures for Sustainable Business & Society Research Group, School of Management, Swansea University, Bay Campus, Fabian Bay, Swansea, Wales, UK

^d Symbiosis International (Deemed University), Pune, Maharashtra, India

ARTICLE INFO

Keywords:

Artificial intelligence
ChatGPT
ChatGPT-supported consumer behavior
Scale development
Validation

ABSTRACT

AI-tools such as ChatGPT can assist researchers to improve the performance of the research process. This paper examines whether researchers could apply ChatGPT to develop and empirically validate new research scales. The study describes a process how to prompt ChatGPT to assist the scale development of a new construct, using the example of the construct of perceived value of ChatGPT-supported consumer behavior. The paper reports four main empirical studies (US: $N = 148$; Australia: $N = 317$; UK: $N = 108$; Germany: $N = 51$) that have been employed to validate the newly developed scale. The first study purifies the scale. The following studies confirm the adjusted factorial validity of the reduced scale. Although the empirical data imply a simplification of the initial multi-dimensional scale, the final three-dimensional operationalization is highly reliable and valid. The paper outlines the shortcomings and several critical notes to stimulate more research and discussion in this area.

1. Introduction

A report from CSIRO (2022) revealed that 98 % of scientific disciplines had already incorporated artificial intelligence (AI) in some form. Hence, even before OpenAI released ChatGPT in November 2022, researchers in various scientific fields frequently employed AI tools or at least tested them (Dwivedi et al., 2021). AI has been extensively discussed in recent years, especially in the field of technological forecasting and social changes (Dwivedi et al., 2023a). Researchers have adopted AI-powered tools to expedite and enhance various aspects of the research process. Examples of such tools include Grammarly, Jarvis, QuillBot, ResearchRabbit, RTutor, and ShortlyAI. Furthermore, researchers have begun to use chatbots as research assistants to help them organize their thoughts, provide comments on their work, assist with coding tasks, and summarize research literature (Hutson, 2022). However, even the most advanced tools have been known to generate subpar textual content (Hutson, 2022). OpenAI's ChatGPT seemingly surpassed these limitations and holds the potential to fundamentally transform research across various domains (Casella et al., 2023; Dwivedi et al., 2023b; Kung et al., 2023; Sallam, 2023; Zhu et al., 2023). Researchers have already demonstrated ChatGPT's ability to generate plausible research ideas, test suggestions and abstracts, and to improve spelling, grammar, and general editing (Altmäe et al., 2023; Bouschery et al.,

2023; Dowling and Lucey, 2023; Gao et al., 2023; Hammad, 2023; Macdonald et al., 2023; Seghier, 2023). Numerous researchers have shared the outcomes of using ChatGPT prompts for various purposes, such as writing essays, novels, poems, or code (Biswas, 2023; Fitria, 2023; Kashefi and Mukerji, 2023; Surameery and Shakor, 2023). Given ChatGPT's potential to process vast amounts of information swiftly and accurately, it can also be employed for other research tasks, including generating hypotheses, detecting patterns, extracting information from large datasets, simplifying literature research, and much more (Rossi et al., 2024; Susarla et al., 2023; venturebeat.com, 2023).

One such task is the development of psychometric scales. Scholars develop psychometric scales to capture real-world phenomena with high accuracy. The scale development is deeply rooted in creativity, but also adheres to a strictly standardized procedure in its development and validation process. Given these characteristics of the process and considering that scale development can be quite labor intensive for researchers, it is valuable to explore how generative AI might assist researchers. Following the concept of the "jagged technological frontier" (Dell'Acqua et al., 2023), it becomes crucial to investigate which aspects of the scale development process can be more efficiently executed with generative AI and which aspects still require direct human intervention. While recognizing the significance of AI-powered tools as research assistants, researchers must also acknowledge several drawbacks,

* Corresponding author.

E-mail addresses: stefan.hoffmann@bwl.uni-kiel.de (S. Hoffmann), wlasarov@audencia.com (W. Lasarov), y.k.dwivedi@swansea.ac.uk (Y.K. Dwivedi).

<https://doi.org/10.1016/j.techfore.2024.123488>

Received 25 August 2023; Received in revised form 23 April 2024; Accepted 28 May 2024

Available online 5 June 2024

0040-1625/© 2024 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

including the potential for incorrect or nonsensical answers, biases in training data, the generation of insecure content, the risk of plagiarism, and the facilitation of the spread of “fake news” by producing seemingly credible but misleading texts that endorse false, hazardous, and counterproductive hypotheses (Dwivedi et al., 2023b; Quintans-Júnior et al., 2023; Rahimi and Abadi, 2023; Salvagno et al., 2023; Stokel-Walker, 2023). A careful evaluation of the usage of AI tools in the research process (i.e., the scale development process) is therefore required. In order to avoid misuse and to ensure a high quality of measurement scales, it is necessary to evaluate the appropriateness of ChatGPT in the scale development process. For example, would empirical validation studies confirm the sound psychometric qualities of the measures created with generative AI tools? Thus far, no study has empirically answered whether generative AI can assist researchers to improve and speed up the scale development process.

Responding to the call to test how AI can assist researchers (Dwivedi et al., 2023b; Peres et al., 2023), this paper investigates whether researchers can employ AI-powered tools in the development of new research scales. The motivation for our research is to help avoid misuse of ChatGPT or other AI-powered tools in the scale development process and to provide evidence-based guidelines for a reasonable use of such tools. As a major contribution of this research, we will not simply evaluate the outcome subjectively, but will run empirical studies to evaluate the results. To the best of our knowledge, this is the first study to use ChatGPT as a research assistant in psychometrics to develop a research scale and guide the researcher through the empirical validation process. This paper tests if and which parts of the classical scale development and validation process (e.g., Churchill, 1979; Clark and Watson, 1995; Dawis, 1987; Gerbing and Anderson, 1988; Loevinger, 1957; Simms, 2008) can be assisted with AI-powered tools. The study’s main claim is that AI can significantly aid the initial steps of the scale development process. Using the example of the construct of perceived value of ChatGPT-supported consumer behavior (perceived value of GPT-CB), the paper first of all describes a process how to prompt ChatGPT to assist the scale development of a new psychological construct. After a scale was developed with the assistance of ChatGPT, the paper reports multiple empirical studies conducted in the US, Australia, UK, and Germany, that validate the new scale. The first study purifies the scale to reduce the initial multi-dimensional suggestion to the essential dimensions that are discriminant valid. Furthermore, this study attempts to establish the nomological network (Cronbach and Meehl, 1955) of the new scale. The following studies confirm the adjusted factorial structure with different samples. Besides these methodological contributions, this paper adds to the literature by suggesting a validated scale of the perceived value of GPT-CB, which scholars might use in future research.

The paper’s objective is to stimulate discussion on the help of AI tools in the research process and to show how to create empirical evidence to enrich this discussion. We demonstrate how ChatGPT can assist researchers to develop new scales, and we report empirical studies that test whether this type of assistance is actually helpful. Building on these findings, we also outline the shortcomings and potential harms. Particularly, we discuss the ethical dimension of the use of AI-powered tools as research assistants.

2. Overview of the scale development and validation process

This paper introduces the new construct of the perceived value of ChatGPT-supported consumer behavior¹ and it develops and validates a scale to measure how consumers evaluate GPT-CB. We choose the domain of consumer behavior as an example because researchers in the field of technological forecasting and social change have recently

¹ We are thankful to an anonymous reviewer who suggested to label the construct “perceived value of GPT-CB,” as our initial notion of ChatGPT-supported consumer behavior was not specific enough.

considered how AI changes the way in which individuals consume (e.g., Aw et al., 2022; Baabdullah et al., 2022; Yeo et al., 2022). Notably, to date, the construct of the perceived value of GPT-CB does not exist in the literature. We introduce this construct to let ChatGPT develop a completely new construct. In this way, we could exclude the possibility that ChatGPT merely copies or integrates existing work. Nonetheless, recent research showed that there is a need for a scale development process in related areas such as measuring consumers’ trust toward interactions with AI service robots (Chi et al., 2021). We rely on the traditional process of creating a new scale, which distinguishes the process of scale development from the process of scale evaluation (Churchill, 1979). We test the claim that ChatGPT might be helpful in the first stage, i.e., the scale development process.

We used ChatGPT version 3.5 in May 2023. ChatGPT is an AI language model that was introduced in November 2022, that provides generated conversational responses to question prompts. The model is trained using a combination of reinforcement learning algorithms and human input on over 150 billion parameters (OpenAI, 2023). The ChatGPT-3.5 series has a trained dataset with a 570 GB equivalent size (The Conversation, 2023). However, it is not connected to the internet and has limited knowledge of the world and events after 2021. Hence, positive for our research, the database is not disturbed by the latest research and buzz about ChatGPT.

Scale development. In a traditional scale development process (e.g., Churchill, 1979; Clark and Watson, 1995; Dawis, 1987; Gerbing and Anderson, 1988), the initial steps consist of defining and conceptualizing the construct as well as creating an item universe. These steps build on an intensive literature review and often on qualitative, explorative research, including focus groups and interviews with consumers, marketers, experts, researchers, or other stakeholders. In this paper, we replace these steps with a dialog with ChatGPT, in which we prompt the AI tool to stepwise create all the output that is usually based on intensive interviews, etc.

Scale purification and evaluation. In the traditional scale development process, the initial dimensional structure and the item universe is purified and validated in a series of multiple empirical studies (Churchill, 1979). These studies typically consider factorial validity (dimensionality), the empirical validity (or prognostic validity), and the construct validity, including discriminant validity and the nomological network. In this study, we actually run these tests empirically. Assuming that some researchers might have started to use ChatGPT to develop items, this is the first study that provides empirical evidence on whether such an item-generation process is actually advisable. Based on the widely-used psychometric criteria, such as Cronbach’s alpha, average variance extracted (AVE), factor loadings, etc. (Churchill, 1979; Gerbing and Anderson, 1988; Nunnally, 1978), we restructure the dimensions and eliminate indicators in order to establish the unidimensionality and the discriminant validity of the dimensions. Put differently, we empirically purify and validate the scale suggested by ChatGPT to demonstrate ChatGPT’s potential in the scale development process.

3. ChatGPT-assisted scale development

Recent insights on the effectiveness of prompt engineering suggest that walking the AI-powered tool through the process will lead to better results than just giving one prompt and hoping for the perfect result (Terry, 2023). For example, chain-of-thought prompting (CoT) improves the reasoning ability of large language models such as ChatGPT by prompting them to generate a series of intermediate steps that lead to the final answer of a multi-step problem (Chen et al., 2023; Diao et al., 2023; Peng et al., 2023; Zhang et al., 2022). In line with these insights, we used a series of subsequent prompts instead of asking ChatGPT to develop the scale with just one prompt. We created prompts to ask ChatGPT-3.5 to conceptualize the construct, establish its dimensionality, and develop the item universe (see Appendix 1 for a full description).

3.1. Defining and conceptualizing

When starting with the prompt “please define ‘consumer behavior,’” ChatGPT answered: “Consumer behavior refers to the actions and decisions made by individuals or households when purchasing, using, evaluating, and disposing of products, services, ideas, or experiences. It includes a wide range of psychological, social, cultural, and economic factors that influence the way people search for information, perceive, select, and buy goods and services” (see [Appendix 1](#) for the full answer). Although there is no citation, the definition matches typical definitions of consumer behavior ([Hoffmann and Akbar, 2023](#); [Hoyer et al., 2017](#); [Kardes et al., 2014](#)).

The next prompt asked ChatGPT to describe its possibilities to support consumers, based on the previously delivered definition. ChatGPT answered that it could “support consumers in various ways by providing them with useful information and guidance throughout their decision-making process.” ChatGPT listed the following examples: product recommendations, informational support, comparison and evaluation, personalized assistance, and post-purchase support (see [Appendix 1](#) for full description). Based on this, ChatGPT summarized the notion of GPT-CB in one sentence with the following words: “ChatGPT-supported consumer behavior refers to the actions and decisions made by individuals or households in the process of purchasing, using, evaluating, and disposing of products or services with the assistance of a language model.”

3.2. Dimensionality, item universe, initial purification

In the next step, we employ ChatGPT to provide ideas about the dimensionality of the construct perceived value of GPT-CB. Since our scale should measure how consumers evaluate the benefits and risks of GPT-CB, we prompted ChatGPT to develop four dimensions of the potential benefits and four dimensions of the potential risks that consumers face when they use ChatGPT to assist their consumer behavior. Notably, the choice of four benefits and risks is somewhat arbitrary. Our intention was to let the AI tool create indicators for a basic set of aspects (here, four facets of benefits and four facets of risks), which will later be reduced in an open, explorative research process with the AI tool’s reflection on the possible discriminant validity of the dimensions and with the help of factor analysis in the quantitative studies. The ChatGPT output presented the following benefit dimensions: convenience, personalization, time-saving, accessibility. The risks include privacy, accuracy, security, and dependence (see [Appendix 1](#) for full description).

In the subsequent step, we asked ChatGPT to develop four potential indicators for each dimension (see [Appendix 1](#)). We asked for four indicators, as we wanted to keep the number of indications as lean as possible while our intention was to have at least three indicators per dimension after a scale purification process. As an initial purification, the next prompt was intended to stimulate ChatGPT to develop a multi-dimensional scale of which the dimensions are actually discriminant and internally consistent. Therefore, we instructed ChatGPT to change the content of the dimensions, merge them, or reduce their number, if necessary. The tool ended up with three dimensions of benefits: convenience, personalization, and accuracy. It also suggested three dimensions of risks: privacy, security, and dependence. As asked, the tool suggested three indicators per dimension, which we actually used in the empirical study.

3.3. Nomological network to validate the new scale

Since we consider a new construct, there is no validation criterion directly available. Consequently, we asked ChatGPT to develop a scale to establish the criterion validity of the scale. ChatGPT created a three-item intention scale (see [Appendix 1](#) for indicators).

Next, we prompted ChatGPT to name four constructs that could be

used to test the theoretical validity, that is the nomological network, of the scale, and to create three indicators per construct. ChatGPT suggested the constructs attitude toward technology, trust in technology, perceived usefulness of ChatGPT, and perceived ease of use of ChatGPT. It is highly plausible to use these constructs to test the nomological network. Although these constructs are well known and frequently used in the field, ChatGPT did not provide any references. ChatGPT’s problem of missing or wrong references has already been documented in the literature ([Dwivedi et al., 2023b](#); [Peres et al., 2023](#)). Obviously, the constructs suggested by ChatGPT are related, for example, to the technology acceptance model and its extensions ([Davis, 1989](#); [Slade et al., 2015](#); [Venkatesh et al., 2003](#)). The construct of trust has already been considered in relation to human confidence in AI ([Chong et al., 2022](#)) and in relation to AI-enabled consumer experience ([Ameen et al., 2021](#)). The indicators are plausible, but they only partly overlap with the established indicators.

While we have thus far mimicked the traditional scale development process with the help of ChatGPT, we now turn to a classic scale validation process. In this way, we can show how the dimensionality and items suggested by ChatGPT perform in real empirical studies.

4. Study 1 – US sample

4.1. Sample and measurement

To evaluate how the scale suggested by ChatGPT performs in an empirical scale purification and validation process, we recruited a US sample via the recruiting platform Prolific ([www.prolific.com](#)). Participants could follow the link to the questionnaire that was announced on the platform if they met the inclusion criteria. The inclusion criteria were US citizenship, English as first language, and age of 18 years or older. We controlled whether the participants were aware of ChatGPT or not. Of the 150 participants, 33 indicated that they often use ChatGPT, 75 have tried it already, and 40 have at least heard of it. However, although the survey call indicated that the study is about ChatGPT and consumer behavior, one person explicitly indicated to have never heard of ChatGPT and one person did not answer this question. Excluding those two participants, the final sample consists of $N = 148$. Of these participants, 52.7 % are female, 46.6 % are male, and 0.7 % are diverse. To be able to include the gender as a dichotomous variable in the analysis, we created a dummy with 0: male versus 1: female/diverse. The mean age is 39.3 years ($SD = 13.1$, ranging from 18 to 77 years). In sum, 58.8 % of the participants have a university degree (bachelor, master, or PhD).

Participants first answered the indicators to measure the benefits and risks of GPT-CB. Afterward, they indicated their intention to engage in ChatGPT-supported consumer behavior. Finally, they responded to indicators of the nomological network. All items were assessed using a seven-point Likert scale anchored by (1) “strongly disagree” and (7) “strongly agree.” Please consult [Appendix 1](#) for the wording of the indicators (the items used are marked with an asterisk (*)).

To control for common method variance ([Fuller et al., 2016](#); [Podsakoff et al., 2003](#)), we conducted [Harman’s \(1976\)](#) one-factor test. Ranging below the critical threshold of 50 %, the first unrotated factor explained 43.2 % of the shared variance of the indicators of all multiple-item constructs.

4.2. Factorial validity and empirical scale purification

As suggested in the traditional psychometric scale development process ([Gerbing and Anderson, 1988](#)), we initially ran an explorative factor analysis (principal component analysis (PCA) with varimax rotation), including all 18 indicators of perceived value of GPT-CB as suggested by ChatGPT. Based on the eigenvalue criterion ($\lambda > 1$), the analysis extracted three factors, which jointly extracted 75.34 % of the variance. The loading matrix is displayed in [Table 1](#).

Table 1
Loading matrix of the explorative factor analysis (Study 1 and Study 2).

	Study 1 (US)			Study 2 (Australia)		
	1	2	3	1	2	3
Eigenvalue	6.337	5.456	1.767	5.408	4.882	1.770
Extracted variance (in %)	35.21	30.31	9.82	30.05	27.12	9.83
ChatGPT provides instant assistance and information that saves me time and effort when researching products.	0.835			0.735		
I appreciate being able to get assistance from ChatGPT at any time of day, which makes shopping and decision-making more convenient for me.	0.858			0.814		
ChatGPT's availability and responsiveness makes it easier for me to shop and make purchase decisions.	0.892			0.809		
ChatGPT's personalized recommendations match my preferences and needs, and help me discover new products that I like.	0.878			0.834		
I feel that ChatGPT understands my preferences and needs better than other sources of information.	0.835			0.792		
I value the personalized service that ChatGPT provides and find it helpful when making purchase decisions.	0.882			0.856		
I am confident in the accuracy and reliability of the information provided by ChatGPT.	0.841			0.750		
ChatGPT's recommendations are unbiased and provide me with relevant and useful information.	0.729			0.634		
I trust ChatGPT's recommendations more than those of human experts.	0.731			0.636		
I am concerned about the privacy of my personal data when using ChatGPT, and I would like to have more control over how my data is used.		0.896			0.889	
I worry that my personal data might be misused or sold to third parties, and I am unsure about how secure it is.		0.938			0.899	
I feel uneasy about the amount of personal information that ChatGPT collects,		0.932			0.891	

Table 1 (continued)

	Study 1 (US)			Study 2 (Australia)		
	1	2	3	1	2	3
and I am uncertain about how it is stored and protected.						
I am concerned about the security of my personal information and transactions when using ChatGPT or related platforms.		0.952			0.918	
I worry that my personal information might be compromised or stolen by hackers.		0.897			0.835	
I feel uneasy about the safety of my financial transactions when using ChatGPT or related platforms.		0.888			0.831	
I am concerned about becoming too dependent on ChatGPT or other AI-based systems, and I worry that I might lose my critical thinking or decision-making abilities.			0.851			0.796
I am unsure if using ChatGPT has made me a better shopper and decision-maker, or if it has made me overly reliant on technology.			0.830			0.870
I would prefer to use ChatGPT as a supplement to my own decision-making, rather than relying on it entirely.						0.508

Notes. Explorative factor analysis (EFA), PCA with varimax rotation. Factor loadings <0.40 are not displayed. The three highest loading items per factor are displayed bold.

The explorative factor analysis requires that the extracted factors are interpreted by the highest loading items in order to label the factors. Usually, the researcher labels the factors. This process is not free of subjective interpretation. In this research, we asked ChatGPT to support this step by prompting it as follows: "An exploratory factor analysis over the following indicators extracted three factors. Please provide a label for factor 1, factor 2, and factor 3." The prompt further included the structured list of the indicators loading highest on the factors. ChatGPT suggested three labels and proactively provided a brief description of the factors:

- **Factor 1 Convenient and Personalized Assistance:** "This factor seems to revolve around the convenience and personalization of assistance provided by ChatGPT. Users appreciate the time-saving aspect of instant information and recommendations, which align with their preferences. The factor emphasizes the role of ChatGPT in streamlining decision-making processes."
- **Factor 2 Privacy and Security Concerns:** "Factor 2 reflects users' concerns about the privacy and security of their personal data when using ChatGPT. These concerns pertain to data collection, storage, and potential misuse. Users express unease about the security of their personal information and financial transactions when interacting with ChatGPT."

- Factor 3 Dependency and Decision-Making Apprehension:** “This factor revolves around the potential risks associated with becoming overly dependent on ChatGPT. Users worry that relying too heavily on AI-based systems might erode their critical thinking and decision-making skills. The factor highlights a preference for using ChatGPT as a supplementary tool rather than a replacement for personal decision-making.”

The team of researchers initially labeled the three dimensions as information, data security, and dependence, which shows a large overlap of the human and the AI interpretation of the dimensions.

In order to assure that the factorial structure is not dependent on the participants’ experience with ChatGPT, we split the analysis. We compared a group of participants who have heard of ChatGPT but have never used it ($n = 40$) and a group of participants who have used ChatGPT ($n = 108$) at least once. The loading matrices for both explorative factor analyses are reported in [Appendix 2, Table A.2](#). For the first group, the EFA replicated the three-dimensional structure, which explains 80.61 % of the variance. When applying the eigenvalue (> 1) criterion, the EFA extracted four factors for the second group. However, the unrotated fourth factor reached an eigenvalue of exactly 1.000. As documented in [Table A.2](#), only item 3 of the Dependence scale has a high loading on this factor, which is typically considered as not enough to extract a factor. Notably, this item has previously proved to be problematic. The item has to be excluded from the analysis with all participants and from the analysis with the first group. Thus, there is evidence that this item needs to be eliminated. Rerunning the factor analysis without this item replicated the three-dimensional structure in both groups (variance extracted group 1: 83.79 %; variance extracted

group 2: 76.12 %). In sum, the group comparison supports the three-dimensional structure in both group.

In order to keep the scale short and economic, we picked the three highest loading items per scale. In case of dependency and decision-making apprehension, only two items displayed high loadings and the third indicator had to be dropped. Hence, the final scale consists of eight items to measure three dimensions.

4.3. Convergent and discriminant validity

To further establish the convergent and discriminant validity of the short scale—and following the traditional scale validation process ([Gerbing and Anderson, 1988](#))—we ran confirmatory factor analysis (CFA) with AMOS 29.0. Since the software requires full data sets, we excluded four participants with missing values on any of the indicators of the final scale. The CFA attests a good fit to the three-dimensional, eight-item perceived value of GPT-CB scale ($\chi^2 = 20.685$; $df = 17$; $\chi^2/df = 1.217$; CFI = 0.997; RMSEA = 0.039). Items and item-to-factor loadings are shown in [Table 2](#). There is discriminant validity, as each construct’s AVE exceeds the maximum of the squared correlations with all latent variables ([Fornell and Larcker, 1981](#)).

To ensure that the three-dimensional solution is actually better than the six-dimensional solution, we also ran a CFA with the initial solution. The CFA reveals a model fit that is worse than the fit of the three-dimensional solution ($\chi^2 = 241.510$; $df = 120$; $\chi^2/df = 2.013$; CFI = 0.953; RMSEA = 0.085). In order to test the discriminant validity with the [Fornell and Larcker \(1981\)](#) test, we calculated the AVE and the highest squared correlation of a construct with all other constructs (r_{max}^2): convenience (AVE = 0.781; $r_{max}^2 = 0.760$), personalization (AVE

Table 2
Convergent and discriminant validity.

Construct Indicator	Study 1 (US)					Study 2 (Australia)					Study 3 (UK)				
	λ	α	AVE	r_{max}^2	F/L	λ	α	AVE	r_{max}^2	F/L	λ	α	AVE	r_{max}^2	F/L
Convenient and personalized assistance		0.922	0.799	0.006	Yes		0.870	0.871	0.043	Yes		0.891	0.726	0.212	Yes
ChatGPT’s personalized recommendations match my preferences and needs, and help me discover new products that I like.	0.853					0.791					0.844				
I feel that ChatGPT understands my preferences and needs better than other sources of information.	0.948					0.841					0.837				
I value the personalized service that ChatGPT provides and find it helpful when making purchase decisions.	0.878					0.862					0.875				
Security		0.973	0.927	0.236	Yes		0.938	0.838	0.119	Yes		0.965	0.901	0.212	Yes
I worry that my personal data might be misused or sold to third parties, and I am unsure about how secure it is.	0.970					0.878					0.936				
I feel uneasy about the amount of personal information that ChatGPT collects, and I am uncertain about how it is stored and protected.	0.972					0.951					0.962				
I am concerned about the security of my personal information and transactions when using ChatGPT or related platforms.	0.947					0.916					0.950				
Dependency and decision-making apprehension		0.828	0.721	0.236	Yes		0.736	0.613	0.119	Yes		0.746	0.620	0.153	Yes
I am concerned about becoming too dependent on ChatGPT or other AI-based systems, and I worry that I might lose my critical thinking or decision-making abilities.	0.909					0.858					0.705				
I am unsure if using ChatGPT has made me a better shopper and decision-maker, or if it has made me overly reliant on technology.	0.785					0.700					0.862				

Notes. Statistics based on CFA with AMOS 29.0. λ = factor loading, α = Cronbach’s alpha. AVE = average variance extracted, r_{max}^2 = highest squared correlation of this construct with all other constructs. F/L: If AVE $> r_{max}^2$ discriminant validity is given according to [Fornell and Larcker \(1981\)](#).

= 0.813, $r_{\max}^2 = 0.760$); accuracy (AVE = 0.649, $r_{\max}^2 = 0.721$); privacy (AVE = 0.492, $r_{\max}^2 = 0.955$); security (AVE = 0.857, $r_{\max}^2 = 0.955$); and dependence (AVE = 0.901, $r_{\max}^2 = 0.256$). Although the average variances extracted are sufficiently high, the solution is not good. According to the test, the dimensions of convenience, personalization, and dependence are the only ones that do not overlap with the other dimensions. There is no discriminant validity of the other initial dimensions, because the correlations between the dimensions are too high.

4.4. Criterion validity

We calculated indices (means) for the three dimensions of the perceived value of GPT-CB (convenient and personalized assistance: M = 4.527, SD = 1.399; privacy and security concerns: M = 5.093, SD = 1.603; dependency and decision-making apprehension: M = 3.956, SD = 1.631) and for the three dimensions of the three-item usage intention scale (M = 4.514, SD = 1.392, $\alpha = 0.946$). We again excluded the four individuals who failed to answer single items. Using the calculated indices, we ran an ordinary least squares (OLS) regression with usage intention as dependent variable and the three dimensions age, sex (dichotomized), and education (dichotomized) as predictors. Findings are listed in Table 3. As expected, the dimension convenient and personalized assistance influences the usage intention positively, while the dimension of privacy and security concerns dampens the usage intention, i.e., influences it negatively.

4.5. Nomological network

To evaluate whether ChatGPT suggests measurement scales that fit into a nomological network of related constructs, we used the four constructs suggested by ChatGPT. These constructs include attitude toward the technology (M = 5.779; SD = 0.840; $\alpha = 0.862$), trust in technology (M = 4.281; SD = 1.262; $\alpha = 0.870$), perceived usefulness of ChatGPT (M = 4.511; SD = 1.295; $\alpha = 0.883$), and perceived ease of use of ChatGPT (M = 5.249; SD = 1.146; $\alpha = 0.916$). The correlation of the three dimensions of the perceived value of GPT-CB with these four constructs are displayed in Table 4.

4.6. Segmentation

Finally, we tested whether different groups differ with regard to the three dimensions convenient and personalized assistance, privacy and security concerns, and dependency and decision-making apprehension (see Table 5). We found no differences with regard to sex and education. There is a marginally significant difference with regard to privacy and security concerns across different age groups. Younger consumers tend to care less about privacy and data security. However, with regard to previous knowledge of and experience with ChatGPT, we find statistically significant differences. Consumers who often use ChatGPT consider the tool to be more informative than other groups. By contrast, these consumers see less issues of privacy and data security than other groups.

4.7. Discussion of Study 1

According to this first validation study, ChatGPT can support the process of scale development. However, the initially developed factorial structure and the item universe needed refinement on basis of empirical research. Specifically, the scale purification process revealed that the initially suggested six-dimensional structure of the perceived value of GPT-CB had to be reduced to three dimensions. It is noteworthy that the traditional scale development process (Churchill, 1979; Gerbing and Anderson, 1988) also suggests to start with a broad conceptualization and a large item universe which the researcher has to reduce with the help of empirical studies in order to achieve unidimensionality and discriminant validity of the dimensions. After an empirically supported

Table 3
OLS regression, DV = usage intention.

	Study 1 (US)			Study 2 (Australia)			Study 3 (UK)			Study 4 (Germany)		
	β	t	p	β	t	p	β	t	p	β	t	p
Convenient and personalized assistance	0.776	13.721	<0.001	0.751	19.666	<0.001	0.789	13.092	<0.001	0.375	2.718	0.009
Privacy and security concerns	-0.130	-2.094	0.038	-0.102	-2.542	0.012	-0.148	-2.307	0.023	-0.335	-2.574	0.014
Dependency and decision-making apprehension	0.066	1.076	0.284	-0.036	-0.913	0.362	0.023	0.366	0.715	0.004	0.028	0.978
Age	0.179	3.179	0.002	-0.061	-1.629	0.104	0.049	0.836	0.405	0.233	1.480	0.146
Sex	-0.181	-3.259	0.001	0.106	2.766	0.006	-0.002	-0.036	0.971	0.106	0.828	0.412
Education	-0.016	-0.288	0.774	0.030	0.825	0.410	0.056	0.968	0.336	0.017	0.105	0.917
F	33.373			75.256			28.955			4.217		
R ²	0.594			0.775			0.689			0.376		
R ² adj	0.576			0.600			0.671			0.287		

Notes. OLS regression, standardized regression coefficients. DV = usage intention, sex (0: male, 1: female, diverse), education (0: other, 1: university degree). Level of significance: *** p < .001, ** p < .01, * p < .05, + p < .10.

Table 4
Nomological Network (US sample).

	Convenient and personalized assistance		Privacy and security concerns		Dependency and decision-making apprehension	
	r	p	r	p	r	p
Attitude toward the technology	0.276***	<0.001	0.014	0.867	-0.109	0.191
Trust in technology	0.679***	<0.001	-0.388***	<0.001	-0.195*	0.018
Perceived usefulness of ChatGPT	0.845***	<0.001	-0.160+	0.055	-0.047	0.574
Perceived ease of use of ChatGPT	0.642***	<0.001	-0.156+	0.061	-0.091	0.275

Notes. r = Pearson product-moment correlation. Level of significance: *** $p < .001$, ** $p < .01$, * $p < .05$, + $p < .10$. We used pairwise exclusion for missing values (N: 144–147).

Table 5
Segmentation.

	Study 1 (US)			Study 2 (Australia)			Study 3 (UK)		
	1	2	3	1	2	3	1	2	3
Sex									
Male	4.270	5.010	3.913	4.149	4.806	4.010	3.776	5.347	4.021
Female	4.752	5.165	3.994	3.948	5.290	4.134	4.046	5.397	3.819
F	4.421	0.339	0.089	1.955	8.658	0.481	1.159	0.030	0.435
p	0.037	0.561	0.766	0.163	0.004	0.489	0.284	0.863	0.511
Age									
to 30 years	4.818	4.620	3.872	4.156	4.861	4.112	3.833	5.111	3.450
31–45 years	4.511	5.289	4.110	4.129	5.090	4.105	4.507	5.217	4.111
> 45 years	4.273	5.288	3.833	3.621	5.364	3.947	3.823	5.833	4.234
F	1.650	2.706	0.444	4.428	2.557	0.272	0.390	2.386	2.132
p	0.196	0.070	0.642	0.013	0.079	0.762	0.678	0.097	0.124
Education									
No university degree	4.481	5.159	4.082	4.068	5.076	4.020	4.127	5.071	3.988
University degree	4.561	5.465	3.919	4.011	5.046	4.112	3.793	5.561	3.877
F	0.115	0.173	0.107	0.154	0.032	0.260	1.734	2.922	0.128
p	0.735	0.678	0.744	0.695	0.858	0.611	0.191	0.090	0.721
Knowledge/experience									
Heard of ChatGPT	4.598	5.525	4.321	3.161	5.582	3.858	3.515	5.522	3.800
Tried ChatGPT	4.538	5.185	3.893	4.080	5.079	4.155	4.196	5.370	4.870
Often used ChatGPT	5.729	4.364	3.667	4.431	4.771	4.090	4.313	4.938	3.781
F	4.856	5.284	1.560	22.549	6.186	0.748	4.319	0.946	0.455
p	0.009	0.006	0.214	<0.001	0.002	0.474	0.016	0.392	0.636

Notes. Mean values. Factors: (1) convenient and personalized assistance, (2) privacy and security concerns, (3) dependency and decision-making apprehension. Category female includes diverse participants.

scale purification process in Study 1, the refined scale of GPT-CB turned out to be discriminant valid and the single dimensions are internally consistent. Hence, we can conclude that ChatGPT cannot easily develop a reliable and valid scale on its own. However, ChatGPT can assist researchers to create an initial conceptualization and an item universe, which then has to be purified in the traditional scale validation process with empirical data. Notably, ChatGPT suggested a scale that is aligned with the nomological network. The theoretical embedding of the three dimensions meets our expectations. The dimension of convenient and personalized assistance is positively correlated with the four external criteria, including attitude toward the technology, trust in technology, perceived usefulness of ChatGPT, and perceived ease of use of ChatGPT. By contrast, the other two dimensions, privacy and security concerns and dependency and decision-making apprehension, which express perceived risks, are only negatively correlated with the trust in technology. This finding matches other findings in the teaching domain that show that perceived threats reduce AI readiness (Wang et al., 2023).

In sum, we conclude that, in our case study, ChatGPT produced an initial set of items for a new measurement scale, which is a valid input for the scale purification process. This indicates that ChatGPT could be used to substitute or augment the initial steps of traditional scale development, which require a sound knowledge of the construct as well as creativity. To further support this conclusion, we ran a follow-up study in which we contrast the initial eight-dimensional solution with the purified scale. Afterward, complementing the process of scale validation, the subsequent studies test whether a ChatGPT-assisted scale development will survive cross-validation studies with different

samples. Following the traditional scale validation process, the following studies aim to further establish the purified scale's validity and to demonstrate the findings' robustness and external validity. We will not use ChatGPT again in these steps, as our claim is that we apply the AI in the creation process, while we build on the typical empirical process for validation. We test the scale in different country contexts. Yet, to keep the influence of cultural variance in this initial scale development process at a low level, we selected only Western countries.

5. Follow-up study to test initial solution – US sample

5.1. Sample and measurement

The objective of this follow-up study is to test whether the scale purified by ChatGPT that we used in Study 1 is actually better than ChatGPT's initial suggestion comprising eight dimensions (four benefits dimensions and four costs dimensions) with four items each.

To be able to compare the findings of this follow-up study with the findings of Study 1, we recruited another US sample via Prolific. Participants were again US citizens with English as first language, who had not participated in Study 1. Of the 152 participants, 50 indicated that they often use ChatGPT, 78 have tried it already, and 22 have at least heard of ChatGPT. We excluded one person who indicated to have never heard of ChatGPT and one person who did not answer this question. The final sample consists of $N = 150$. Of these participants, 58.7 % are female, 37.3 % are male, and 3.3 % are diverse. One person did not provide any information. The mean age is 38.1 years ($SD = 13.5$, ranging

from 19 to 86 years). In sum, 52.7 % of the participants have a university degree (bachelor, master, or PhD).

Participants first answered the indicators to measure the benefits and risks of GPT-CB, but this time we used the indicators initially suggested by ChatGPT for eight dimensions (four benefit dimensions and four risk dimensions) with four indicators per dimension. Afterward, the participants indicated their intention to engage in ChatGPT-supported consumer behavior. Please consult [Appendix 1](#) for the wording of the indicators (the items used are marked with a plus (+)).

Harman's one-factor test indicated that the first unrotated factor accounts for 42.3 % of the variance of the indicators of the multiple-item constructs, which ranged below the critical threshold of 50 %.

5.2. Factorial validity of the initial dimensions

We ran an explorative factor analysis (PCA with varimax rotation), including all 36 indicators that were initially suggested by ChatGPT. Based on the eigenvalue criterion ($\lambda > 1$), the analysis extracted five factors, which jointly extracted 74.64 % of the variance. The loading matrix is displayed in [Appendix 3, Table A.3](#). We further ran isolated factor analysis with the indicators of the benefits. This analysis extracted two factors, which account for 75.43 % of the variance. However, the unrotated eigenvalue of the first factor (EV = 10.86, 67.88 % explained variance) is dramatically higher than the eigenvalue of the second factor (EV = 1.201, 7.55 %). Moreover, as indicated in [Table A.3](#), the rotated solution produced many cross-loadings. We ran isolated factor analysis with the indicators of risks. This analysis extracted four factors, which account for 76.53 % of the variance. However, the rotated solution again indicates several cross-loadings.

Confirming the inadequacy of the eight-dimensional solution, a CFA reports an unsatisfactory fit ($\chi^2 = 1142.392$; $df = 436$; $\chi^2/df = 2.620$; CFI = 0.834; RMSEA = 0.107). The internal consistency (Cronbach's alpha), the AVE, and the highest squared correlation of a construct with all other constructs (r_{\max}^2) are as follows: convenience ($\alpha = 0.898$; AVE = 0.690; $r_{\max}^2 = 0.933$), personalization ($\alpha = 0.913$; AVE = 0.728; $r_{\max}^2 = 0.857$), time-saving ($\alpha = 0.907$; AVE = 0.711; $r_{\max}^2 = 0.986$), accessibility ($\alpha = 0.892$; AVE = 0.690; $r_{\max}^2 = 0.986$), privacy ($\alpha = 0.912$; AVE = 0.736; $r_{\max}^2 = 0.863$), accuracy ($\alpha = 0.764$; AVE = 0.509; $r_{\max}^2 = 0.575$), security ($\alpha = 0.869$; AVE = 0.648; $r_{\max}^2 = 0.863$), and dependence ($\alpha = 0.719$; AVE = 0.437; $r_{\max}^2 = 0.212$). These figures indicate that although the internal consistency of the scales (Cronbach's alpha) and the AVE are sufficiently high, the solution is not good. According to the [Fornell and Larcker \(1981\)](#) test, dependence is the only dimension that does not overlap with the other dimensions. There is no discriminant validity of the other initial dimensions, because the correlations between the dimensions are too high.

5.3. Discussion

In sum, the analysis shows that ChatGPT's initial eight-dimensional suggestion is not acceptable. The suggested dimensions are not discriminant. The three-dimensional solution of Study 1 performs better. We can therefore conclude that it is important to prompt ChatGPT to develop the dimensionality in such a way that they are discriminant valid. Yet, as Study 1 showed, this prompt is not sufficient. A further empirical scale purification process is required too.

6. Study 2 – Australian sample

6.1. Sample and measures

Study 2 further validates the short scale of GPT-CB with a sample recruited via Prolific in Australia. The main objective of Study 2 is to rerun the scale purification process with a larger sample of at least 300 participants, because the suggestions for the ideal sample size for EFA are ambiguous. For example, some authors recommend a sample-size-to-

parameter ratio of 4:1 ([MacCallum et al., 2001](#)) or even 10:1 (e.g., [Hair et al., 2019](#)) or recommend to include a range of 100 to 200 participants for well-defined factors ([MacCallum et al., 1999](#)) and 300 participants or more if communalities are low ([Tabachnick and Fidell, 2014](#)). Inclusion criteria were Australian citizenship, English as first language, and age of 18 years or older. Of the 320 participants, 118 indicated that they often use ChatGPT, 139 have tried it already, and 60 have at least heard of ChatGPT. Three participants, who indicated that they have never heard of ChatGPT, were excluded. The final sample consists of $N = 317$. Of these participants, 52.1 % are female, 46.4 % are male, 0.9 % are diverse, and 0.6 % did not want to answer this question. The mean age is 35.9 years (SD = 12.3, ranging from 18 to 86 years). In sum, 59.6 % of the participant have a university degree (bachelor, master, or PhD).

The second study included the indicators of the three benefits dimensions, the indicators of the three risk dimensions, as well as the indicators of the usage intention as suggested by ChatGPT. The one-factor test indicated that the first unrotated factor accounts for 38.53 % of the variance of the indicators of the multiple-item constructs. In this study, we additionally added a marker item to further confirm that common method variance does not disturb the findings ([Lindell and Whitney, 2001](#)). We used one item of [Haws et al.'s \(2014\)](#) green consumption scale ("I would describe myself as environmentally responsible"), which is theoretically unrelated to the indicators of the perceived value of GPT-CB. In fact, the marker item is statistically unrelated to 15 of the included indicators, and we found only six statistically significant relationships, which are very weak ($r_{\max} = 0.15$).

6.2. Factorial validity and empirical scale purification

With this new and larger data set, we reran the EFA (PCA with varimax rotation), including all 18 indicators of the perceived value of GPT-CB as suggested by ChatGPT. Based on the eigenvalue criterion ($\lambda > 1$), the analysis extracted three factors, which jointly extracted 67.00 % of the variance. The loading matrix is displayed in [Table 1](#). The structure of the factor loadings corroborates the findings of Study 1. Hence, the analysis with the new data set confirms the reduction to three factors and scale purification.

6.3. Convergent and discriminant validity

To further establish the scale, we ran a confirmatory factor analysis, excluding five participants with missing values on some indicators. The analysis confirms a good fit to the three-dimensional, eight-item perceived value of GPT-CB scale ($\chi^2 = 12.805$; $df = 17$, $\chi^2/df = 0.753$, CFI = 1.000, RMSEA = 0.000). According to [Fornell and Larcker's \(1981\)](#) test, there is discriminant validity. Psychometric properties of the scale and the indicators are shown in [Table 2](#).

To ensure again that the three-dimensional solution is actually better than the six-dimensional solution, we ran a CFA with the initial solution. However, the identified model fit is worse than the fit of the three-dimensional solution ($\chi^2 = 208.714$; $df = 120$; $\chi^2/df = 1.739$; CFI = 0.977; RMSEA = 0.049). The [Fornell and Larcker \(1981\)](#) test confirms the findings of Study 1: convenience (AVE = 0.695; $r_{\max}^2 = 0.717$), personalization (AVE = 0.704; $r_{\max}^2 = 0.717$), accuracy (AVE = 0.592; $r_{\max}^2 = 0.604$), privacy (AVE = 0.833; $r_{\max}^2 = 0.916$), security (AVE = 0.723; $r_{\max}^2 = 0.916$), and dependence (AVE = 0.422; $r_{\max}^2 = 0.116$). The initial six dimensions lack discriminant validity, because the correlations between the dimensions are too high.

6.4. Criterion validity

To test the criterion validity, we ran an OLS regression with the dimensions convenient and personalized assistance (M = 4.034, SD = 1.273), privacy and security concerns (M = 5.058, SD = 1.472), and dependency and decision-making apprehension (M = 4.074, SD = 1.577) as predictors and the usage intention scale (M = 3.982, SD =

1.539, $\alpha = 0.947$) as dependent variable. We excluded nine participants from this analysis due to a few missing values on some indicators. As documented in [Table 3](#), the dimension of convenient and personalized assistance again fosters usage intention, while concerns about privacy and data security reduce the usage intention.

6.5. Segmentation

Again, we compared consumers' responses to the three dimensions of the perceived value of GPT-CB with regard to sex, age, education, and ChatGPT knowledge and experience. The results largely repeat the results of Study 1 ([Table 5](#)). Particularly, there is again a strong relationship between the previous knowledge of and experience with ChatGPT and how consumers evaluate the information provided by ChatGPT.

6.6. Discussion of Study 2

Study 2 supports the findings of Study 1 in another country setting and with a larger sample size. The study confirms the construct's three-dimensional structure and assures that this structure is superior to the initial six-dimensional structure.

7. Study 3 – UK sample

7.1. Sample and measures

Study 3 aims to further validate the short scale of GPT-CB with a sample recruited via Prolific in the UK. Inclusion criteria were UK citizenship, English as first language, and age of 18 years or older. Of the 121 participants, 16 indicated that they often use ChatGPT, 46 have already tried it, and 46 have at least heard of ChatGPT. After excluding 11 subjects who have never heard of ChatGPT and two subjects who did not answer this question, the final sample consists of $N = 108$ of which 52.8 % are female, 45.45 are male, 0.9 % are diverse, and 0.9 % did not want to answer this question. The mean age is 38.6 years ($SD = 11.8$, ranging from 21 to 66 years). In sum, 61.1 % of the participants have a university degree (bachelor, master, or PhD).

This study included only the items of the shortened, final scale as well as three indicators of the intention to engage in ChatGPT-supported consumer behavior. The first unrotated factor accounts for 47.0 % of the variance of the indicators of the multiple-item constructs.

7.2. Convergent and discriminant validity

For the CFA, we excluded one participant with missing values on some indicators. The analysis attests a good fit to the three-dimensional, eight-item perceived value of GPT-CB scale ($\chi^2 = 25.303$; $df = 17$; $\chi^2/df = 1.488$; CFI = 0.987; RMSEA = 0.068). According to [Fornell and Larcker's \(1981\)](#) test, there is discriminant validity. Psychometric properties of the scale and the indicators are shown in [Table 2](#).

7.3. Criterion validity

Again, we ran OLS regression with the predictors convenient and personalized assistance ($M = 3.923$, $SD = 1.289$), privacy and security concerns ($M = 5.370$, $SD = 1.463$), and dependency and decision-making apprehension ($M = 3.921$, $SD = 1.560$) and the dependent variable usage intention ($M = 3.787$, $SD = 1.541$, $\alpha = 0.947$). Given a few missing values on some indicators, two participants are excluded from this analysis. In line with Study 1 and Study 2, the dimension convenient and personalized assistance again fosters usage intention, while concerns about privacy and data security reduce the usage intention.

7.4. Segmentation

The segmentation results are largely in line with the results of Study 1 and Study 2 ([Table 5](#)). There are marginally significant differences in the dimension privacy and security concerns across consumers of different age groups and education levels. Again, there is a relationship between the previous knowledge of and experience with ChatGPT and how consumers evaluate the information provided by ChatGPT.

7.5. Discussion of Study 3

Study 3 provides further evidence that items suggested by ChatGPT can survive the empirical validation process. The study cross-validates the shortened perceived value of GPT-CB scale in another country setting, confirming the construct's three-dimensional structure found in Study 1 and Study 2. Although the participants of the studies stem from different countries, they might still be rather similar. Therefore, a slightly different group of participants could help further establish the external validity.

8. Study 4 – German sample

8.1. Sample and measures

To further test ChatGPT's potential to suggest valid measurement scales, we continue to empirically validate the purified scale. We ran another study in which we modified several sample characteristics to establish external validity. We distributed the link to the questionnaire in the classroom at a German university in a midsized city. The students are highly involved in the subject of consumer behavior, as they were taking an undergraduate class on consumer behavior. Moreover, one week prior to answering the questionnaire, they attended a guest lecture of an expert on AI and chat bots, who explained the basic functionalities as well as the advantages and disadvantages of ChatGPT. Hence, this is a younger, well-educated, and well-informed sample. Of the 51 participants, eight indicated that they often use ChatGPT, 31 have tried it already, and 12 have at least heard of ChatGPT. Of these participants, 51.0 % indicated to be female, 47.1 % indicated to be male, and 2.0 % did not want to answer this question. The mean age is 22.8 years ($SD = 2.6$, ranging from 18 to 34 years).

Similar to the third study, this study included only the items of the shortened, final scale as well as three indicators of the intention to engage in ChatGPT-supported consumer behavior. Again, all items were assessed using a seven-point Likert scale. Harman's one-factor test indicated that the first unrotated factor accounts for 35.0 % of the variance of the indicators of the multiple-item constructs, which ranged below the critical threshold of 50 %.

8.2. Convergent and discriminant validity

Given the small sample size, we did not run CFA. The EFA (PCA with varimax rotation), including the eight indicators of the perceived value of GPT-CB, extracted the three expected factors, which jointly accounted for 80.29 % of the variance. The indicators loaded highest on the expected factors: convenient and personalized assistance (eigenvalue λ : 2.315; loadings a: 0.868, 0.883, 0.850), privacy and security concerns (eigenvalue λ : 2.815; loadings a: 0.954, 0.949, 0.889), and dependency and decision-making apprehension (eigenvalue λ : 1.293; loadings a: 0.873, 0.683).

8.3. Criterion validity

We calculated the mean scores of the three dimensions of GPT-CB (convenient and personalized assistance: $M = 3.608$, $SD = 1.275$; privacy and security concerns: $M = 4.314$, $SD = 1.718$; dependency and decision-making apprehension: $M = 3.784$, $SD = 1.317$) and of the

three-item usage intention scale ($M = 3.767$, $SD = 1.229$, $\alpha = 0.862$). Owing to individual missing values on single items (intention and sex), we excluded two participants. Again, we find that convenient and personalized assistance intensifies the usage intention, while privacy and security concerns reduce such intentions (Table 3).

8.4. Discussion of Study 4

Study 4 provides further evidence that measurement items suggested by ChatGPT can be useful to develop scales. The study largely supports the short scale of the perceived value of GPT-CB. Notably, this sample consists of a younger, well-educated group of students who are knowledgeable when it comes to consumer behavior and the basic functionalities of ChatGPT.

9. General discussion

The objective of this paper is to test whether AI could support scholars in the scale development process. To the best of our knowledge, this is the first report of an AI-assisted scale development process that is subsequently evaluated in a traditional manner with empirical studies. Research has already intensively considered how AI will alter different functions and processes in decision-making of, for example, health care (Dalvi-Esfahani et al., 2023; Wang et al., 2018), the public sector (Di Vaio et al., 2022), business (Loureiro et al., 2021; Makarius et al., 2020), innovation management (Füller et al., 2022; Keding and Meissner, 2021; Trocin et al., 2021), marketing (Mustak et al., 2021; Vlačić et al., 2021), advertising and communication (Shen, 2022; Shi and Wang, 2023), consumer behavior (Aw et al., 2022; Baabdullah et al., 2022; Yeo et al., 2022), or social companionship (Chaturvedi et al., 2023). However, there is still a lack of evidence with regard to the role of AI in the research process, including the scale development process. With the help of a series of prompts, it was possible to define and conceptualize a new construct and to develop an item universe in a relatively short period of time (about 1 h). This leads to the conclusion that AI tools such as ChatGPT can support researchers in the scale development process.

However, an AI model such as ChatGPT does not include an inherent validation of the results. Therefore, we used the traditional process to evaluate the suggested scale. The purification and validation process leads to a mixed evaluation. On the negative side, the new scale had to be reduced in a purification process. Three of the initial six dimensions survived. The indicators of the other three dimensions were not discriminant and these dimensions were deleted. Remarkably, ChatGPT's initial eight-dimensional solution was even worse in terms of discriminant validity. Therefore, it is necessary to prompt ChatGPT to develop discriminant dimensions. However, this prompt is not sufficient. An empirical scale purification process is definitively needed. Scholars cannot rely directly on the ChatGPT solution if they develop scales for multi-dimensional constructs. It is important to note that scales developed in a traditional scale development process also have to be purified. On the positive side, the three final dimensions turned out to be internally consistent, discriminant valid, and logically embedded in a nomological network. The new scale performed acceptably in multiple empirical studies in the US, Australia, UK, and Germany.

The following recommendations can be extracted from the findings. First, scholars can make use of AI tools such as ChatGPT to develop new measurement scales. However, they should not expect ChatGPT to suggest a perfectly valid scale based on a single prompt.

Second, we recommend an interaction between the researcher and the AI tool during the whole scale development phase, which mimics the well-established traditional scale development process. Researchers need to assure that all necessary steps of the scale development are fulfilled. This includes the construct definition, conceptualization, dimensionality, item generation, and initial item purification. For example, we warn scholars to ask ChatGPT to only develop items. We suggest to first develop the construct conceptualization with the

dimensional structure and then to develop the item universe adjusted to the dimensions. Second, we advise scholars to prompt ChatGPT to develop discriminant scales, which they afterward purify empirically.

Third, we assume that integrating the knowledge of the humans and the input of the AI tool will produce the best output. The more expertise researchers have, the more they will benefit from sharing their knowledge in the prompting process. While there are indications in this study for positive effects of this integration (e.g., consider our instruction to disentangle benefits and risks as dimensions), we have not yet tested this assumption. We call for future research to compare scales that are developed only by ChatGPT and scales developed in teams of ChatGPT and humans to provide further support for this recommendation. In the classic scale development process, scholars also use the input of study participants via in-depth interviews and focus groups to develop new, creative items for the item universe. We assume that combining the input of such qualitative research techniques and ChatGPT will produce the most creative item universe.

Fourth, we urge scholars not to directly implement the scales that ChatGPT suggests. Our research demonstrates that ChatGPT's suggestion are a good basis. However, our results also stress that an empirical scale purification process is required. ChatGPT cannot substitute the empirical work done in a traditional scale evaluation process, which consists of a test of factorial validity, convergent and discriminant validity, and criterion validity, and the embedding in the nomological network of related constructs.

Finally, we strongly recommend scholars to critically evaluate ChatGPT's output in every step, based on their expertise in psychometrics and their substantial expertise in the field for which they develop the scale.

10. Limitations, cautious remarks, and the need for future research

Although the initial test of AI-assisted scale development yielded satisfactory outcomes, it is essential to highlight various cautious remarks and limitations associated with this study and the tested process.

The first set of limitations and remarks pertains to the *required professional competence and critical thinking ability of researchers* when utilizing ChatGPT. In the context of using AI language models like ChatGPT, critical thinking skills have become more crucial than ever. ChatGPT and other large language models generate answers based on affinities and statistical matches, making it challenging to discern correct answers or parts thereof, especially when they are presented with expert-like language. OpenAI acknowledges this challenge, stating on the ChatGPT blog that "ChatGPT sometimes writes plausible-sounding but incorrect or nonsensical answers" (OpenAI, 2023). Furthermore, researchers and individuals typically rely on bookstores, libraries, or search engines to access knowledge sources. Platforms like Google provide links and previews of text fragments, allowing users to click and explore further to find accurate information. In this case, recipients have the option to at least validate the accuracy of the content by examining the website itself (e.g., if it is a governmental website). However, with ChatGPT and similar AI systems, when users pose a question, they receive a response in the form of a few paragraphs. Although the answer may appear correct, there is a possibility that it is inaccurate (Zhang et al., 2023). For example, due to the limitations of AI systems, the generated scale may unintentionally contain biases or flawed assumptions. The AI tool might suggest questions that assume a certain cultural context or fail to capture important nuances specific to the target population, thereby developing a useless scale. Consequently, a naïve application that solely relies on AI-powered tools for scale development is not recommended. The absence of an expert human mind in scientific work, when relying solely on AI, may perpetuate such biases and inaccuracies and it may hinder scientific progress, emphasizing the importance of human expertise in conducting and ensuring the quality of scientific activity (Salvagno et al., 2023). For an ideal scale development

process, researchers seeking AI assistance should possess extensive knowledge of psychometrics and knowledge of the subject for which the scale is being developed. This enables researchers to prompt the AI tool effectively based on their expertise and to evaluate the AI's results, deciding which parts to retain, modify, or refine with subsequent prompts. It is worth noting that prolonged and exclusive reliance on ChatGPT or similar tools for scale development (or any other research purpose) may potentially erode the researcher's skills over time, as the tool performs the majority of the work. Remarkably, this dependency and decision-making apprehension dimension has even been included in the perceived value of the GPT-CB scale developed by ChatGPT. We recommend that researchers always explicitly declare whether they have used AI-tools to develop scales and, if so, they should name the respective tools. They need to check, edit, and improve the AI-supported scales and they must indicate that they take full responsibility for the final scale.

Another set of limitations and remarks pertains to the *researcher-ChatGPT interaction*, particularly the *researcher's prompt engineering competence*, which proves to be critical. Prompt engineering involves designing questions and instructions (prompts) to enhance the quality of results obtained from ChatGPT (Polak and Morgan, 2023; Liu et al., 2023). Previous research has shown that the quality and relevance of ChatGPT's output can be significantly improved by crafting well-designed prompts. Conversely, using poorly crafted prompts can lead to unsatisfactory or incorrect responses (e.g., Short and Short, 2023). In our present study, we observed that the scale development process yielded reasonable findings when researchers went beyond using a simple prompt like "Develop scale XY" and instead organized the process into distinct steps. However, it is important to note that our prompts were primarily based on our unstructured knowledge as researchers in this domain. Furthermore, OpenAI states that "ChatGPT is sensitive to tweaks to the input phrasing or attempting the same prompt multiple times. For example, given one phrasing of a question, the model can claim to not know the answer, but given a slight rephrase, can answer correctly" (openai.com/blog/chatgpt). The variation of prompts can therefore produce different results. It is crucial to test the system's sensitivity to prompt variation. To further enhance the effectiveness of prompt engineering, it is advisable to develop a series of prompts that guide ChatGPT progressively from defining and conceptualizing the construct to developing the item universe. Future research on AI-empowered scale development could focus on systematizing prompt crafting (White et al., 2023), resulting in *scale prompt patterns* that serve as templates for other researchers. It would be valuable to empirically test and compare different prompt series or patterns to determine the ones that produce the best results. Such research endeavors not only contribute to the development of scales but also provide optimized prompt patterns that enable researchers to develop their own scales effectively.

In addition to the researchers' competence in their discipline and as prompt engineers, they are also required to properly *assess the abilities and knowledge of ChatGPT*. In this study, we explicitly chose a construct that had not yet been introduced in the literature at the time of the study in order to ensure that ChatGPT would not simply reproduce already developed scales. Future research might check how ChatGPT performs when scholars prompt to develop a scale for a construct already intensively discussed in the literature. Furthermore, to test the nomological network of the new scale, we prompted ChatGPT to suggest related constructs. The suggestions provided (e.g., attitude, trust, perceived ease of use, perceived usability) are highly plausible, and a human researcher might have considered similar constructs as a first step to test the nomological network. It is important to note that ChatGPT did not refer to the existing literature in making these suggestions. This challenge is well documented (Dwivedi et al., 2023b; Peres et al., 2023). Perhaps future versions of ChatGPT or other AI tools may address this issue, improving their ability to make use of relevant literature. However, at its current state, a naïve adoption of the suggested validation constructs

may result in incorrectly specified constructs and could potentially lead to plagiarism. It is important to recognize that ChatGPT lacks deep contextual understanding and may generate biased responses. Furthermore, the current study was carried out with the assistance of ChatGPT-3.5, but newer versions of ChatGPT, such as ChatGPT-4, already demonstrate greater general intelligence and can be considered early versions of artificial general intelligence (AGI) systems (Bubeck et al., 2023). We therefore call for continuous methodological research on AI assistance in research, including the scale development process, to constantly monitor how newer versions are able to assist scholars and to assess any potential harms that may arise from such assistance.

As an avenue for future research, we suggest to test *how humans and ChatGPT can jointly develop new scales*. In the present study, we completely replaced the initial scale development steps with ChatGPT. We did not run literature reviews and qualitative interviews or focus groups in parallel to conceptualize the construct and define indicators. Since ChatGPT builds on an existing knowledge corpus, it might not be able to develop radically innovative items for constructs that are newly introduced in the literature. Above, we suggested to first define the construct conceptualization and then to develop items based on this conceptualization as a way to improve the item universe. Nonetheless, we assume that the scale development process could be further improved by integrating classic steps performed by humans (literature review, interviews, brainstorming) with the AI input. Particularly, we assume that creating an item universe that builds on the contributions of humans and AI will greatly improve the process, which is similar to other idea generation and innovation development processes that integrate AI and human ideation (Bouschery et al., 2023).

Another limitation of this study pertains to the *generalizability of the results*, which can be influenced by potential sample biases in ChatGPT and the study design. ChatGPT was trained using the "Reinforcement Learning from Human Feedback" (RLHF) approach (OpenAI.com, 2023). However, it is important to acknowledge that the individuals involved in training ChatGPT may not represent the entire world population, potentially leading to sample biases in the underlying training data (Cosuins, 2023). Notably, given the origin of OpenAI in the United States, it is possible that the algorithm has been primarily trained using data from a US American perspective. Furthermore, in the present study, the scale developed with the assistance of ChatGPT has only been empirically tested in Western cultures. Therefore, it is crucial to consider these biases both during the development process by ChatGPT and during the validation phase carried out by researchers. Future research should aim to examine the validity of such scales in different cultures, exploring whether the results of scale development processes facilitated by AI are culturally bound and potentially not inclusive. This necessitates an interdisciplinary approach that incorporates ethical considerations. Moreover, it is worth questioning whether the results of the AI-assisted scale development process are reproducible and applicable to other contexts. Thus far, our testing has focused solely on one topic, namely the perceived value of ChatGPT-supported consumer behavior. To ensure the broader usefulness of an AI-assisted scale development process, it is essential to conduct further tests encompassing diverse topics. We call for future research that follows our protocol to develop scales for different constructs and to compare ChatGPT's performance in different domains. By addressing these limitations and by conducting comprehensive research, we can enhance the reliability and applicability of AI-powered scale development methods.

It is essential to examine our results critically in light of the ongoing discussion on *research ethics*. While some authors believe that AI tools will revolutionize scientific publishing (Grimaldi and Ehrler, 2023), significant ethical concerns surround the use of ChatGPT and other AI-powered tools in academic publishing. For example, the reliance of ChatGPT on existing internet text (until 2021) raises doubts about the originality, verifiability, and accuracy of its outputs. AI or ChatGPT systems can both commit plagiarism and be programmed to avoid it by rephrasing content, though the latter practice is not acceptable in

scientific research (Salvagno et al., 2023). Therefore, such AI-powered tools have the potential to propagate “fake news” by producing scientifically plausible texts that support false, hazardous, and counterproductive hypotheses. Sole reliance on AI-generated content can therefore mislead the public and undermine trust in science (Rahimi and Abadi, 2023). For scale development, researchers must therefore be aware of already existing scales to avoid replication and plagiarism of these scales. Another topic of discussion pertains to whether AI-powered tools can serve as (co-)authors of scientific studies. ChatGPT appears to possess the necessary technical capabilities in this regard (Lee, 2023; Teubner et al., 2023; Zimmerman, 2023). To address these issues and uphold the integrity of academic publishing, it is crucial to establish revised policies and guidelines (Stokel-Walker, 2023). Journals such as *Nature* and *Science* have already revised their policies to address situations where researchers received assistance from ChatGPT or other AI-powered tools (Thorp, 2023). These journals and others explicitly state that AI chatbots cannot be considered authors of articles published in their respective journals (Koçak and Altay, 2023; Nature Editorials, 2023; Nature.com; Rahman et al., 2023). Researchers employing ChatGPT or other AI-powered tools for scale development must take these concerns into consideration.

Besides the methodological contributions of suggesting a procedure how to develop and validate measurement scales with the combined help of AI and empirical studies, this paper also adds to the literature by suggesting a validated scale of the perceived value of GPT-CB, which scholars might use in future research. The final scale, however, uses only eight items to measure three subdimensions, with only two items assigned to the third dimension. As it is generally desirable to use more items per dimension, we advise researchers to increase the potential number of items per dimension during the initial scale development process before they start the empirical purification. Nonetheless, it is noteworthy that our suggested scale is the result of the rule-based procedure of scale development and that two-item constructs are not uncommon in related fields of research, including innovation and technology management as well as marketing and consumer research. Two-item scales are used, for example, to measure affordable value innovation (Ernst et al., 2015), several subdimensions of the customer acceptance of automated technologies in service encounters (Fernandes and Oliveira, 2021), perceived value of artificial intelligence service agents (Noor et al., 2022), or the construct ‘against ethics anxiety’ of artificial intelligence (Li and Huang, 2020).

11. Conclusion

This paper has shown that ChatGPT can assist researchers in the scale development process. A series of prompts can help define and conceptualize new constructs and can help develop an item universe. In this

Appendix 1. Prompting ChatGPT

Notes. In the following, the prompts for ChatGPT are displayed italics and bold on a grey shaded background. The items selected for the empirical study are marked with an asterisk (*).

Step 1: Defining and Conceptualizing

paper, the final product was not perfect and directly usable. However, after a purification of the dimensionality and the indicators in a first empirical study, the subsequent studies evaluated the shortened scale positively. The empirical work to purify and validate the scale is definitively still needed. In sum, this is a first indication that scholars could make use of AI in the scale development process. However, we have also identified several shortcomings and potential harms in our discussion of the results. Scholars need to be careful. They need to critically evaluate and validate the suggestions and make informed decisions based on their deep understanding of the functionalities of AI language models, their domain knowledge, their expertise in psychometrics, and with the help of empirical studies. The current version of ChatGPT can be used in conjunction with human expertise to develop scales. Scholars must avoid overreliance on the automated assistance and they need to keep control. Still, more methodological research on the scale development process is needed before we can actually trust the results of an AI-supported scale development.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRedit authorship contribution statement

Stefan Hoffmann: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Wassili Lasarov:** Writing – review & editing, Writing – original draft, Conceptualization. **Yogesh K. Dwivedi:** Writing – review & editing, Writing – original draft, Conceptualization.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT-3.5 in order to develop a new measurement scale (see Manuscript and Appendix 1 for details). After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

None.

Data availability

Data will be made available on request.

Please define "consumer behavior".

Consumer behavior refers to the actions and decisions made by individuals or households when purchasing, using, evaluating, and disposing of products, services, ideas, or experiences. It includes a wide range of psychological, social, cultural, and economic factors that influence the way people search for information, perceive, select, and buy goods and services. Consumer behavior is a crucial area of study for businesses and marketers who aim to understand and predict consumer needs and preferences, identify market trends, and develop effective marketing strategies.

Based on this definition of consumer behavior, please describe how ChatGPT could support consumers.

As a language model, ChatGPT can support consumers in various ways by providing them with useful information and guidance throughout their decision-making process. Here are some examples:

- **Product recommendations:**

ChatGPT can help consumers find the right products based on their preferences and needs. By analyzing the consumer's search queries and past behavior, ChatGPT can recommend products that are most likely to meet their requirements.

- **Informational support:**

Consumers may have questions about products or services they're considering buying. ChatGPT can provide answers to their queries by providing accurate and relevant information.

- **Comparison and evaluation:**

Consumers often compare products before making a purchase decision. ChatGPT can help consumers compare and evaluate different products by providing a side-by-side comparison of their features, prices, and other relevant details.

- **Personalized assistance:**

ChatGPT can provide personalized assistance to consumers by analyzing their past behavior, preferences, and needs. Based on this information, ChatGPT can recommend products, suggest ways to improve their shopping experience, and offer personalized promotions and discounts.

- **Post-purchase support:**

ChatGPT can provide post-purchase support to consumers by answering their questions, providing assistance with returns or exchanges, and offering suggestions on how to use the product or service effectively.

Overall, ChatGPT can support consumers by providing them with relevant, accurate, and timely information and assistance throughout their purchase journey.

We will now use the term "ChatGPT-supported consumer behavior" to refer to what you have just describe. We will call it the construct of "ChatGPT-supported consumer behavior". Please give a definition of "ChatGPT-supported consumer behavior" in one sentence.

ChatGPT-supported consumer behavior refers to the actions and decisions made by individuals or households in the process of purchasing, using, evaluating, and disposing of products or services with the assistance of a language model.

Step 2: Dimensionality

Please list the four main benefits and the four main risks that might come to consumer's mind when they think about "ChatGPT supported consumer behavior". Give labels and brief descriptions per category (one sentence).

Sure! Here are the four main benefits and the four main risks that consumers may associate with ChatGPT-supported consumer behavior:

Benefits:

- **Convenience**

consumers can receive instant assistance and information on products or services they are interested in.

- **Personalization**

ChatGPT can provide personalized recommendations based on the consumer's past behavior, preferences, and needs.

- **Time-saving**

consumers can save time and effort in researching and comparing products.

- **Accessibility**

ChatGPT can provide assistance and information 24/7, without the need for human interaction.

Risks:

- **Privacy**

consumers may have concerns about the collection and use of their personal data by ChatGPT and the companies that use it.

- **Accuracy**

consumers may worry about the accuracy and reliability of the information provided by ChatGPT, especially if it is not verified or validated by a human expert.

- **Security**

consumers may be concerned about the security of their personal information and transactions when using ChatGPT or related platforms.

- **Dependence**

consumers may become overly dependent on ChatGPT or other AI-based systems, potentially leading to reduced critical thinking or decision-making abilities.

Step 3: Item universe

Next, we want to develop a questionnaire to measure the potential benefits and risks of “ChatGPT-supported consumer behavior” from the consumer perspective. Please create four statements for each of the four benefits and four risks that you mentioned. The statements should be written in the “I”-perspective of the consumer. Consumer should be able to agree or disagree on these statements. Use a reflective (not a formative) specification of the items.

Sure, here are four statements for each of the four benefits and four risks of ChatGPT-supported consumer behavior. Consumers can indicate their level of agreement or disagreement with these statements on a Likert scale:

Benefits:

Convenience

- I find it convenient to receive instant assistance and information from ChatGPT. (+)
- ChatGPT saves me time and effort when researching and comparing products. (+)
- I appreciate being able to get assistance from ChatGPT at any time of day. (+)
- ChatGPT's availability makes it easier for me to shop and make purchase decisions. (+)

Personalization

- ChatGPT provides me with personalized recommendations that match my preferences and needs. (+)
- I feel that ChatGPT understands my preferences and needs better than other sources of information. (+)
- ChatGPT's personalized recommendations have helped me discover new products that I like. (+)
- I value the personalized service that ChatGPT provides. (+)

Time-saving

- ChatGPT saves me time and effort when searching for products or services. (+)
- I find that ChatGPT's product recommendations are more relevant and useful than other sources of information. (+)
- ChatGPT helps me compare products more efficiently than I could on my own. (+)
- I appreciate the time-saving benefits of using ChatGPT when making purchase decisions. (+)

Accessibility

- I find it convenient to get assistance and information from ChatGPT without having to talk to a person. (+)
- ChatGPT's availability makes it easier for me to shop and make purchase decisions on my own schedule. (+)
- I appreciate the ease of use and accessibility of ChatGPT. (+)
- ChatGPT's accessibility has made shopping and making purchase decisions more enjoyable for me. (+)

Risks:**Privacy**

- I am concerned about the privacy of my personal data when using ChatGPT. (+)
- I worry that my personal data might be misused or sold to third parties. (+)
- I feel uneasy about the amount of personal information that ChatGPT collects. (+)
- I would prefer to use ChatGPT if I had more control over how my personal data is used. (+)

Accuracy

- I am concerned about the accuracy and reliability of the information provided by ChatGPT. (+)
- I worry that ChatGPT might provide me with incorrect or biased information. (+)
- I am unsure if ChatGPT's recommendations are better than those of human experts. (+)
- I would prefer to use ChatGPT if I was certain that its recommendations were accurate and unbiased. (+)

Security

- I am concerned about the security of my personal information and transactions when using ChatGPT or related platforms. (+)
- I worry that my personal information might be compromised or stolen by hackers. (+)
- I feel uneasy about the safety of my financial transactions when using ChatGPT or related platforms. (+)
- I would prefer to use ChatGPT if I was certain that my personal information and transactions were secure. (+)

Dependence

- I am concerned about becoming too dependent on ChatGPT or other AI-based systems. (+)
- I worry that I might lose my critical thinking or decision-making abilities if I rely too much on ChatGPT. (+)
- I am unsure if using ChatGPT has made me a better shopper and decision-maker. (+)
- I would prefer to use ChatGPT if I could still make my own decisions and think critically when needed. (+)

Step 4: Purification

We have now created a eight-dimensional operationalization of the construct "ChatGPT-supported consumer behavior". For each dimension we have four potential items. Each dimension should have a high internal consistency. And there should be a high level of discriminant validity. Please reduce the number of items to three for each scale and take care of the internal consistency and discriminant validity. If necessary reformulate items. If necessary drop a dimension.

Sure, I can help with that. Based on the original list of 4 benefits and 4 risks, here's a reduced set of 6 dimensions and 3 potential items for each:

Benefits:**Convenience**

- ChatGPT provides instant assistance and information that saves me time and effort when researching products. (*)
- I appreciate being able to get assistance from ChatGPT at any time of day, which makes shopping and decision-making more convenient for me. (*)
- ChatGPT's availability and responsiveness makes it easier for me to shop and make purchase decisions. (*)

Personalization

- ChatGPT's personalized recommendations match my preferences and needs, and help me discover new products that I like. (*)
- I feel that ChatGPT understands my preferences and needs better than other sources of information. (*)
- I value the personalized service that ChatGPT provides and find it helpful when making purchase decisions. (*)

Accuracy

- I am confident in the accuracy and reliability of the information provided by ChatGPT. (*)
- ChatGPT's recommendations are unbiased and provide me with relevant and useful information. (*)
- I trust ChatGPT's recommendations more than those of human experts. (*)

Risks:**Privacy**

- I am concerned about the privacy of my personal data when using ChatGPT, and I would like to have more control over how my data is used. (*)
- I worry that my personal data might be misused or sold to third parties, and I am unsure about how secure it is. (*)
- I feel uneasy about the amount of personal information that ChatGPT collects, and I am uncertain about how it is stored and protected. (*)

Security

- I am concerned about the security of my personal information and transactions when using ChatGPT or related platforms. (*)
- I worry that my personal information might be compromised or stolen by hackers. (*)
- I feel uneasy about the safety of my financial transactions when using ChatGPT or related platforms. (*)

Dependence

- I am concerned about becoming too dependent on ChatGPT or other AI-based systems, and I worry that I might lose my critical thinking or decision-making abilities. (*)

- I am unsure if using ChatGPT has made me a better shopper and decision-maker, or if it has made me overly reliant on technology. (*)
- I would prefer to use ChatGPT as a supplement to my own decision-making, rather than relying on it entirely. (*)

In terms of internal consistency and discriminant validity, these six dimensions should be relatively distinct and not overlap too much in terms of the concepts they measure. The three items for each dimension should also have good internal consistency, with a Cronbach's alpha of at least 0.7. It may also be helpful to validate the questionnaire through further testing and refinement.

Step 5: Validation Criteria

Please suggest a three-items scale to measure whether consumer will actually use ChatGPT in the future.

Sure, here are three potential items to measure a consumer's intention to use ChatGPT in the future:

- I am likely to use ChatGPT in the future for assistance with shopping and purchase decisions. (*)
- I would recommend ChatGPT to others who are looking for help with shopping and purchase decisions. (*)
- I plan to continue using ChatGPT in the future to assist with my shopping and purchase decisions. (*)

These items measure the consumer's likelihood of using ChatGPT in the future, as well as their willingness to recommend it to others. A high score on this scale would suggest that the consumer is likely to continue using ChatGPT and find it useful. However, it's important to note that intention to use may not always translate to actual behavior, and external factors such as availability and accessibility of the technology may also play a role in determining usage.

Please suggest four further constructs to measure the nomological network of the scale. Make suggestions of three items to measure each of these four constructs.

Sure, here are four potential constructs to measure the nomological network of the "intention to use ChatGPT" scale, along with three potential items for each:

Attitude toward technology:

- I generally enjoy using new technologies and gadgets. (*)
- I find that technology makes my life easier and more efficient. (*)
- I am open to trying out new technologies and apps. (*)

Trust in technology:

- I trust that ChatGPT will provide me with accurate and reliable information. (*)
- I feel confident that my personal information will be secure when using ChatGPT. (*)
- I believe that ChatGPT is a trustworthy source of information and advice. (*)

Perceived usefulness of ChatGPT:

- I find ChatGPT to be helpful when making purchase decisions. (*)
- I feel that ChatGPT provides me with useful information that I wouldn't be able to find on my own. (*)
- I believe that ChatGPT improves my shopping and decision-making experiences. (*)

Perceived ease of use of ChatGPT:

- I find ChatGPT to be easy to use and navigate. (*)
- I feel that ChatGPT's interface is intuitive and user-friendly. (*)
- I believe that it is easy to find the information I need using ChatGPT. (*)

These constructs are related to the "intention to use ChatGPT" scale in that they capture attitudes and perceptions that may influence the likelihood of using the technology. The items for each construct should be reliable and distinct from each other, with a Cronbach's alpha of at least 0.7.

Thank you for assisting the scale development process of the new construct "ChatGPT-supported consumer behavior". Do you have any final suggestions?

You're welcome! Here are some final suggestions to consider as you continue to refine and develop the "ChatGPT-supported consumer behavior" scale:

• Pilot test the scale:

Before using the scale in a larger study, it's important to conduct a pilot test with a small group of participants to ensure that the scale is reliable, valid, and easy to understand.

• Consider the context:

Consumer behavior is influenced by a wide range of factors, including social, cultural, and economic factors. When using the scale, it's important to consider the context in which the consumer is making their decisions, as this can impact their perceptions and attitudes toward ChatGPT.

• **Keep the scale short and simple:**

While it's important to capture as much information as possible, it's also important to keep the scale short and simple so that participants are more likely to complete it. Consider reducing the number of items for each construct to three, as we did in the previous step.

• **Use a variety of data collection methods:**

In addition to using self-report measures like surveys, consider using other data collection methods such as behavioral observation or focus groups to gain a more comprehensive understanding of how consumers use ChatGPT and what factors influence their behavior.

By keeping these suggestions in mind, you can continue to refine and develop the "ChatGPT-supported consumer behavior" scale and gain valuable insights into how consumers use technology to make purchasing decisions.

Appendix 2

Table A.2
Loading matrix for participants with and without ChatGPT experience.

	Heard of ChatGPT, but not used			-	Have used ChatGPT			
	1	2	3		1	2	3	4
Eigenvalue	6.953	5.464	2.095		5.582	5.524	1.753	1.139
Extracted variance (in %)	38.63	30.35	11.64		31.57	30.69	9.741	6.328
Convenience 1	0.881				0.786			
Convenience 2	0.912				0.807			
Convenience 3	0.947				0.867			
Personalization1	0.899				0.849			
Personalization2	0.851				0.794			
Personalization3	0.935				0.819			
Accuracy1	0.899				0.776			
Accuracy2	0.747				0.690			
Accuracy3	0.735				0.716			
Privacy1		0.749	0.408			0.938		
Privacy2		0.961				0.950		
Privacy3		0.973				0.945		
Security1		0.956				0.953		
Security2		0.911				0.901		
Security3		0.922				0.870		
Dependence1			0.796				0.842	
Dependence2			0.836				0.895	
Dependence3			0.457					0.860

Notes. Explorative factor analysis (EFA), principle component analysis (PCA) with Varimax rotation. Factor loadings <0.40 are not displayed. The three highest loading items per factor are displayed bold.

Appendix 3

Table A.3
Loading matrix of initial eight-dimensional solution.

	Joint analysis with benefits and risks					-	Only benefits		-	Only risks			
	1	2	3	4	5		1	2		1	2	3	4
Eigenvalue	11.051	6.486	2.561	2.341	1.444		6.309	5.760		5.154	2.499	2.463	2.128
Extracted variance (in %)	34.54	20.27	8.00	7.32	4.51		39.43	36.00		32.21	15.62	15.40	13.30
Convenience													
Item 1	0.787							0.789					
Item 2	0.822						0.505	0.666					
Item 3	0.762							0.827					
Item 4	0.861						0.718	0.509					
Personalization													
Item 1	0.768						0.800						
Item 2	0.739						0.860						
Item 3	0.799				-0.406		0.843						
Item 4	0.827						0.680	0.495					
Time-saving													
Item 1	0.912						0.715	0.572					
Item 2	0.827						0.795						

(continued on next page)

Table A.3 (continued)

	Joint analysis with benefits and risks					Only benefits		Only risks			
	1	2	3	4	5	1	2	1	2	3	4
Item 3	0.795					0.621	0.485				
Item 4	0.854					0.518	0.680				
Accessibility											
Item 1	0.734						0.814				
Item 2	0.861					0.539	0.705				
Item 3	0.774						0.848				
Item 4	0.846					0.720	0.483				
Privacy											
Item 1		0.871						0.875			
Item 2		0.890						0.861			
Item 3		0.883						0.885			
Item 4		0.683	0.507					0.650	0.602		
Accuracy											
Item 1		0.616			0.491			0.495		0.692	
Item 2		0.522			0.600					0.773	
Item 3		0.408								0.765	
Item 4			0.796						0.817		
Security											
Item 1		0.870						0.800			
Item 2		0.875						0.805			
Item 3		0.791						0.685		0.462	
Item 4		0.434	0.724						0.781		
Dependence											
Item 1				0.797							0.878
Item 2				0.801							0.849
Item 3				0.716						0.428	0.610
Item 4			0.749						0.784		

Notes. Explorative factor analysis (EFA), principle component analysis (PCA) with Varimax rotation. Factor loadings <0.40 are not displayed. The three highest loading items per factor are displayed bold.

References

Altmäe, S., Sola-Leyva, A., Salumets, A., 2023. Artificial intelligence in scientific writing: a friend or a foe? *Reproductive BioMedicine Online*, in press. <https://doi.org/10.1016/j.rbmo.2023.04.009>.

Ameen, N., Tarhini, A., Reppel, A., Anand, A., 2021. Customer experiences in the age of artificial intelligence. *Computers in Human Behavior* 114, 106548.

Aw, E.C.X., Tan, G.W.H., Cham, T.H., Raman, R., Ooi, K.B., 2022. Alexa, what's on my shopping list? Transforming customer experience with digital voice assistants. *Technological Forecasting and Social Change* 180, 121711.

Baabdullah, A.M., Alalwan, A.A., Algharabat, R.S., Metri, B., Rana, N.P., 2022. Virtual agents and flow experience: an empirical examination of AI-powered chatbots. *Technological Forecasting and Social Change* 181, 121772.

Biswas, S., 2023. Role of ChatGPT in computer programming. *Mesopotamian Journal of Computer Science* 2023, 8–16.

Bouschery, S.G., Blazevic, V., Piller, F.T., 2023. Augmenting human innovation teams with artificial intelligence: exploring transformer-based language models. *Journal of Product Innovation Management* 40 (2), 139–153.

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., ... & Zhang, Y., 2023. Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712*.

Casella, M., Montomoli, J., Bellini, V., Bignami, E., 2023. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. *J. Med. Syst.* 47 (1), 1–5.

Chaturvedi, R., Verma, S., Das, R., Dwivedi, Y.K., 2023. Social companionship with artificial intelligence: recent trends and future avenues. *Technological Forecasting and Social Change* 193, 122634.

Chen, J., Chen, L., Huang, H., Zhou, T., 2023. When Do You Need Chain-of-Thought Prompting for ChatGPT? *arXiv preprint arXiv:2304.03262*.

Chi, O.H., Jia, S., Li, Y., Gursoy, D., 2021. Developing a formative scale to measure consumers' trust toward interaction with artificially intelligent (AI) social robots in service delivery. *Computers in Human Behavior* 118, 106700.

Chong, L., Zhang, G., Goucher-Lambert, K., Kotovsky, K., Cagan, J., 2022. Human confidence in artificial intelligence and in themselves: the evolution and impact of confidence on adoption of AI advice. *Computers in Human Behavior* 127, 107018.

Churchill, A., 1979. A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research* 16 (1), 64–73.

Clark, L.A., Watson, D., 1995. Constructing validity: basic issues in objective scale development. *Psychological Assessment* 7 (3), 309–319.

Cosains, B., 2023. Uncovering The different types of ChatGPT bias, *Forbes*. <https://www.forbes.com/sites/forbestechcouncil/2023/03/31/uncovering-the-different-types-of-chatgpt-bias/?sh=282f52b8571b> (accessed May 20th, 2023).

Cronbach, L.J., Meehl, P.E., 1955. Construct validity in psychological tests. *Psychol. Bull.* 52 (4), 281.

CSIRO. Artificial Intelligence for Science. Adoption Trends and Future Development Pathways, Australia's National Science Agency. <https://www.csiro.au/en/research/>

[technology-space/ai/artificial-intelligence-for-science-reportCSIRO](https://www.csiro.au/en/research/technology-space/ai/artificial-intelligence-for-science-reportCSIRO) (accessed May 21th, 2023).

Dalvi-Esfahani, M., Mosharaf-Dehkordi, M., Leong, L.W., Ramayah, T., Kanaan-Jebna, A. M.J., 2023. Exploring the drivers of XAI-enhanced clinical decision support systems adoption: insights from a stimulus-organism-response perspective. *Technological Forecasting and Social Change* 195, 122768.

Davis, F.D., 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q.* 13 (3), 319–340.

Dawis, R.V., 1987. Scale construction. *J. Couns. Psychol.* 34 (4), 481–489.

Dell'Acqua, F., McFowland, E., Mollick, E.R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Kraye, L., Candelon, F., Lakhani, K.R., 2023. Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. *Harvard Business School Technology & Operations Mgt. Unit Working Paper* (24-013).

Di Vaio, A., Hassan, R., Alavoine, C., 2022. Data intelligence and analytics: a bibliometric analysis of human-artificial intelligence in public sector decision-making effectiveness. *Technological Forecasting and Social Change* 174, 121201.

Diao, S., Wang, P., Lin, Y., Zhang, T., 2023. Active prompting with chain-of-thought for large language models. *arXiv preprint arXiv:2302.12246*.

Dowling, M., Lucey, B., 2023. ChatGPT for (finance) research: the Bananarama conjecture. *Finance Research Letters* 53, 103662.

Dwivedi, Y.K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Williams, M. D., 2021. Artificial intelligence (AI): multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management* 57, 101994.

Dwivedi, Y.K., Sharma, A., Rana, N.P., Giannakis, M., Goel, P., Dutot, V., 2023a. Evolution of artificial intelligence research in technological forecasting and social change: research topics, trends, and future directions. *Technological Forecasting and Social Change* 192, 122579.

Dwivedi, Y.K., Kshetri, N., Hughes, L., Slade, E.L., Jeyaraj, A., Kar, A.K., Wright, R., 2023b. "So what if ChatGPT wrote it?" multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management* 71, 102642.

Ernst, H., Kahle, H.N., Dubiel, A., Prabhu, J., Subramaniam, M., 2015. The antecedents and consequences of affordable value innovations for emerging markets. *Journal of Product Innovation Management* 32 (1), 65–79.

Fernandes, T., Oliveira, E., 2021. Understanding consumers' acceptance of automated technologies in service encounters: drivers of digital voice assistants adoption. *J. Bus. Res.* 122, 180–191.

Fitria, T.N., 2023. Artificial intelligence (AI) technology in OpenAI ChatGPT application: a review of ChatGPT in writing English essay. *ELT Forum: Journal of English Language Teaching* 12 (1), 44–58.

Fornell, C., Larcker, D.F., 1981. Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research* 18 (1), 39–50.

- Fuller, C.M., Simmering, M.J., Atinc, G., Atinc, Y., Babin, B.J., 2016. Common methods variance detection in business research. *J. Bus. Res.* 69 (8), 3192–3198.
- Füller, J., Hutter, K., Wahl, J., Bilgram, V., Tekic, Z., 2022. How AI revolutionizes innovation management—perceptions and implementation preferences of AI-based innovators. *Technological Forecasting and Social Change* 178, 121598.
- Gao, C.A., Howard, F.M., Markov, N.S., Dyer, E.C., Ramesh, S., Luo, Y., Pearson, A.T., 2023. Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *npj Digital Medicine* 6 (1), 75.
- Gerbing, D.W., Anderson, J.C., 1988. An updated paradigm for scale development incorporating unidimensionality and its assessment. *Journal of Marketing Research* 25 (2), 186–192.
- Grimaldi, G., Ehrler, B., 2023. AI et al.: machines are about to change scientific publishing forever. *ACS Energy Letters* 8, 878–880.
- Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., 2019. *Multivariate data analysis*, 8th ed. Cengage.
- Hammad, M., 2023. The impact of artificial intelligence (AI) programs on writing scientific research. *Ann. Biomed. Eng.* 1–2.
- Harman, H.H., 1976. *Modern Factor Analysis*. University of Chicago Press, Chicago, IL.
- Haws, K.L., Winterich, K.P., Naylor, R.W., 2014. Seeing the world through GREEN-tinted glasses: green consumption values and responses to environmentally friendly products. *Journal of Consumer Psychology* 24 (3), 336–354.
- Hoffmann, S., Akbar, P., 2023. *Consumer Behavior*. Understanding Consumers—Designing Marketing Activities. Springer.
- Hoyer, W.D., MacInnis, D.J., Pieters, R., 2017. *Consumer Behavior*, 7th ed. Cengage Learning.
- Hutson, M., 2022. Could AI help you to write your next paper? *Nature* 611 (7934), 192–193.
- Kardes, F., Cronley, M., Cline, T., 2014. *Consumer Behavior*, 2nd ed. Cengage Learning.
- Kashefi, A., Mukerji, T., 2023. ChatGPT for programming numerical methods. *Journal of Machine Learning for Modeling and Computing*. <https://doi.org/10.1615/JMachLearnModelComput.2023048492>.
- Keding, C., Meissner, P., 2021. Managerial overreliance on AI-augmented decision-making processes: how the use of AI-based advisory systems shapes choice behavior in R&D investment decisions. *Technological Forecasting and Social Change* 171, 120970.
- Koçak, Z., Altay, S., 2023. Balkan medical journal policy on the use of chatbots in scientific publications. *Balkan Med. J.* 40 (3), 149–150.
- Kung, T.H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J., Tseng, V., 2023. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLoS Digital Health* 2 (2), e0000198.
- Lee, J.Y., 2023. Can an artificial intelligence chatbot be the author of a scholarly article? *Journal of Educational Evaluation for Health Professions* 20 (6). <https://doi.org/10.3352/jeehp.2023.20.6>.
- Li, J., Huang, J.S., 2020. Dimensions of artificial intelligence anxiety based on the integrated fear acquisition theory. *Technology in Society* 63, 101410.
- Lindell, M.K., Whitney, D.J., 2001. Accounting for common method variance in cross-sectional research designs. *Journal of Applied Psychology* 86 (1), 114–121.
- Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., Neubig, G., 2023. Pretrain, prompt, and predict: a systematic survey of prompting methods in natural language processing. *ACM Computing Surveys* 55 (9), 1–35.
- Loevinger, J., 1957. Objective tests as instruments of psychological theory. *Psychol. Rep.* 3 (3), 635–694.
- Loureiro, S.M.C., Guerreiro, J., Tussyadijah, I., 2021. Artificial intelligence in business: state of the art and future research agenda. *J. Bus. Res.* 129, 911–926.
- MacCallum, R.C., Widaman, K.F., Zhang, S., Hong, S., 1999. Sample size in factor analysis. *Psychological Methods* 4 (1), 84–99.
- MacCallum, R.C., Widaman, K.F., Preacher, K.J., Hong, S., 2001. Sample size in factor analysis: the role of model error. *Multivariate Behavioral Research* 36 (4), 611–637.
- Macdonald, C., Adeloye, D., Sheikh, A., Rudan, I., 2023. Can ChatGPT draft a research article? An example of population-level vaccine effectiveness analysis. *Journal of Global Health* 13 (01003). <https://doi.org/10.7189/jogh.13.01003>.
- Makarasi, E.E., Mukherjee, D., Fox, J.D., Fox, A.K., 2020. Rising with the machines: a sociotechnical framework for bringing artificial intelligence into the organization. *Journal of Business Research* 120, 262–273.
- Mustak, M., Salminen, J., Plé, L., Wirtz, J., 2021. Artificial intelligence in marketing: topic modeling, scientometric analysis, and research agenda. *J. Bus. Res.* 124, 389–404.
- Nature Editorials, 2023. Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature* 613, 612.
- Nature.com, 2023. Initial submission. *Nature*. <https://www.nature.com/nature/fo-r-authors/initial-submission>.
- Noor, N., Rao Hill, S., Troshani, I., 2022. Developing a service quality scale for artificial intelligence service agents. *European Journal of Marketing* 56 (5), 1301–1336.
- Nunnally, J.C., 1978. *Psychometric Theory*, 2nd ed. McGraw Hill.
- OpenAI, 2023. Corporate website retrieved June, 23rd. www.openai.com.
- Peng, K., Ding, L., Zhong, Q., Shen, L., Liu, X., Zhang, M., ... & Tao, D., 2023. Towards making the most of ChatGPT for machine translation. *arXiv preprint arXiv:2303.13780*.
- Peres, R., Schreier, M., Schweidel, D., Sorescu, A., 2023. On ChatGPT and beyond: how generative artificial intelligence may affect research, teaching, and practice. *International Journal of Research in Marketing* 40 (2), 269–275.
- Podsakoff, P.M., MacKenzie, S.B., Lee, J.Y., Podsakoff, N.P., 2003. Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology* 88 (5), 879.
- Polak, M.P., Morgan, D., 2023. Extracting accurate materials data from research papers with conversational language models and prompt engineering—example of ChatGPT. *arXiv preprint arXiv:2303.05352*.
- Quintans-Júnior, L.J., Gurgel, R.Q., Araújo, A.A.D.S., Correia, D., Martins-Filho, P.R., 2023. ChatGPT: the new panacea of the academic world. *Revista da Sociedade Brasileira de Medicina Tropical* 56, e0060–e2023.
- Rahimi, F., Abadi, A.T.B., 2023. ChatGPT and publication ethics. *Archives of Medical Research* 54 (3), 272–274.
- Rahman, M., Terano, H.J.R., Rahman, N., Salamzadeh, A., Rahaman, S., 2023. ChatGPT and academic research: a review and recommendations based on practical examples. *Journal of Education, Management and Development Studies*. 3 (1), 1–12.
- Rossi, S., Rossi, M., Mukkamala, R. R., Thatcher, J. B., Dwivedi, Y. K., 2024. Augmenting research methods with foundation models and generative AI. *International Journal of Information Management*, 102749, doi: <https://doi.org/https://doi.org/10.1016/j.ijinfomgt.2023.102749>.
- Sallam, M., 2023. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthcare* 11 (6) 6, 887.
- Salvagno, M., Taccone, F.S., Gerli, A.G., 2023. Can artificial intelligence help for scientific writing? *Critical Care* 27 (1), 1–5.
- Seghier, M.L., 2023. Using ChatGPT and other AI-assisted tools to improve manuscripts readability and language. *International Journal of Imaging Systems and Technology*. <https://doi.org/10.1002/ima.22902>.
- Shen, G., 2022. AI-enabled talent training for the cross-cultural news communication talent. *Technological Forecasting and Social Change* 185, 122031.
- Shi, B., Wang, H., 2023. An AI-enabled approach for improving advertising identification and promotion in social networks. *Technological Forecasting and Social Change* 188, 122269.
- Short, C.E., Short, J.C., 2023. The artificially intelligent entrepreneur: ChatGPT, prompt engineering, and entrepreneurial rhetoric creation. *J. Bus. Ventur. Insights* 19, e00388.
- Simms, L.J., 2008. Classical and modern methods of psychological scale construction. *Social and Personality Psychology Compass* 2 (1), 414–433.
- Slade, E.L., Dwivedi, Y.K., Piercy, N.C., Williams, M.D., 2015. Modeling consumers' adoption intentions of remote mobile payments in the United Kingdom: extending UTAUT with innovativeness, risk, and trust. *Psychol. Mark.* 32 (8), 860–873.
- Stokel-Walker, C., 2023. ChatGPT listed as author on research papers: many scientists disapprove. *Nature* 613 (7945), 620–621.
- Surameery, N.M.S., Shakor, M.Y., 2023. Use chat GPT to solve programming bugs. *International Journal of Information Technology & Computer Engineering* 3 (1), 17–22.
- Susarla, A., Gopal, R., Thatcher, J.B., Sarker, S., 2023. The Janus effect of generative AI: charting the path for responsible conduct of scholarly activities in information systems. *Information Systems Research* 34 (2), 399–408.
- Tabachnick, B.G., Fidell, L.S., 2014. *Using Multivariate Statistics*, 6th ed. Pearson, Boston, MA.
- Terry, O.K., 2023. I'm a student. You have no idea how much we're using ChatGPT, *The Chronicle of Higher Education*. https://www.chronicle.com/article/im-a-student-you-have-no-idea-how-much-were-using-chatgpt?cid=gen_sign_in.
- Teubner, T., Flath, C.M., Weinhardt, C., van der Aalst, W., Hinz, O., 2023. Welcome to the era of ChatGPT et al. the prospects of large language models. *Business & Information Systems Engineering* 65, 95–101.
- The Conversation, 2023. ChatGPT could be a game-changer for marketers, but it won't replace humans any time soon. <https://theconversation.com/chatgpt-could-be-a-game-changer-for-marketers-but-it-wont-replace-humans-any-time-soon-198053>. Jan 22th, 2023, retrieved June, 23rd.
- Thorp, H.H., 2023. ChatGPT is fun, but not an author. *Science* 379 (6630), 313.
- Trocin, C., Hovland, I.V., Mikalef, P., Dremel, C., 2021. How artificial intelligence affords digital innovation: a cross-case analysis of Scandinavian companies. *Technological Forecasting and Social Change* 173, 121081.
- Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D., 2003. User acceptance of information technology: toward a unified view. *MIS Q.* 27 (3), 425–478.
- Venturebeat, 2023. Harnessing the power of GPT-3 in scientific research, *Venturebeat*, <https://venturebeat.com/ai/harnessing-the-power-of-gpt-3-in-scientific-research/>, accessed May 21th, 2023.
- Vlačić, B., Corbo, L., e Silva, S.C., Dabić, M., 2021. The evolving role of artificial intelligence in marketing: a review and research agenda. *Journal of Business Research* 128, 187–203.
- Wang, X., Li, L., Tan, S.C., Yang, L., Lei, J., 2023. Preparing for AI-enhanced education: conceptualizing and empirically examining teachers' AI readiness. *Computers in Human Behavior* 146, 107798.
- Wang, Y., Kung, L., Byrd, T.A., 2018. Big data analytics: understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change* 126, 3–13.
- White, J., Hays, S., Fu, Q., Spencer-Smith, J., Schmidt, D.C., 2023. ChatGPT prompt patterns for improving code quality, refactoring, requirements elicitation, and software design. *arXiv preprint arXiv:2303.07839*.
- Yeo, S.F., Tan, C.L., Kumar, A., Tan, K.H., Wong, J.K., 2022. Investigating the impact of AI-powered technologies on Instagrammers' purchase decisions in digitalization era—a study of the fashion and apparel industry. *Technological Forecasting and Social Change* 177, 121551.
- Zhang, Z., Zhang, A., Li, M., Smola, A., 2022. Automatic chain of thought prompting in large language models. *arXiv preprint arXiv:2210.03493*.

- Zhang, Y., Li, Y., Cui, L., Cai, D., Liu, L., Fu, T., Huang, X., Zhao, E., Zhang, Y., Chen, Y., Wang, L., Luu, A.T., Bi, W., Shi, F., Shi, S., 2023. Siren's song in the AI ocean: a survey on hallucination in large language models. arXiv preprint arXiv:2309.01219.
- Zhu, J.J., Jiang, J., Yang, M., Ren, Z.J., 2023. ChatGPT and environmental research. *Environmental Science & Technology*. <https://doi.org/10.1021/acs.est.3c01818>.
- Zimmerman, A., 2023. A ghostwriter for the masses: ChatGPT and the future of writing. *Annals of Surgical Oncology* 30, 3170–3173.

Stefan Hoffmann is a Professor of Marketing at Kiel University, Kiel. He holds the Marketing chair at Kiel University and is co-founder of the Kiel Institute for Responsible Innovation (KIRI). He received both his doctorate and super-PhD (Habilitation) from Dresden University of Technology, Germany. Prof. Hoffmann's research focuses on consumer behavior, digital marketing, sustainability marketing, international marketing, and responsible innovation. He has published in several leading marketing journals, including the *Journal of the Academy of Marketing Science*, *Journal of Retailing*, *International Journal of Research in Marketing*, *Journal of Service Research*, and *Journal of Consumer Psychology*.

Wassili Lasarov is an Assistant Professor in the Department of Marketing at Audencia Business School, Nantes (France). His main research interests are consumer psychology,

sustainable consumption, digital marketing, and international marketing. He has published in several marketing and business ethics journals, including *International Journal of Research in Marketing*, *Journal of Business Ethics*, *Journal of Business Research*, and *Ecological Economics*.

Yogesh K. Dwivedi is a Professor of Digital Marketing and Innovation and Founding Director of the Digital Futures for Sustainable Business & Society Research Group at the School of Management, Swansea University, Wales, UK. In addition, he holds a Distinguished Research Professorship at the Symbiosis International (Deemed University), Pune, India. Professor Dwivedi is also currently leading the *International Journal of Information Management* as its Editor-in-Chief. His research interests are at the interface of Information Systems (IS) and Marketing, focusing on issues related to consumer adoption and diffusion of emerging digital innovations, digital government, and digital and social media marketing particularly in the context of emerging markets. Professor Dwivedi has published >500 articles in a range of leading academic journals and conferences that are widely cited (>70 thousand times as per Google Scholar). He has been named on the annual Highly Cited Researchers™ 2020, 2021 and 2022 lists from Clarivate Analytics. Professor Dwivedi is an Associate Editor of the *Journal of Business Research*, *European Journal of Marketing*, and *Government Information Quarterly*, and Senior Editor of the *Journal of Electronic Commerce Research*.