

# EINE DELPHI-STUDIE ZUM EINSATZ VON SPRACHMODELLEN IN DER MARKT- UND KONSUMENTENFORSCHUNG

MARKO SARSTEDT | LEA RAU  
INSTITUT FÜR MARKETING





# Marketing is dead – es lebe das Marketing Das ChatGPT und Co. für Brands bede

INTERVIEW

# ”ChatGPT kann ein Marketing-Team mit zehn Mitarbeitern ersetzen”

Von Arnulf Ramcke | Freitag, 10. Januar 2025

# ChatGPT im Praxistest - Auswertung offener Nennungen durch eine KI

veröffentlicht am 08.02.2023 // ak

Welche konkreten Aufgaben könnten Marktforschende an ChatGPT übertragen? Hier die Ergebnisse.

# Versteht ChatGPT meine Kund\*innen besser als ich?

Ist ChatGPT bald der neue Kund\*innen-Versteher? absatzwirtschaft Redakteur Andreas Marx setzt sich mit

Handelsblatt

Kolumne „Kreative Zerstörung“

# Silicon Sampling ist Cambridge Analytica auf Steroiden

KI-Sprachmodelle wie ChatGPT imitieren Menschen extrem gut. Das bringt bahnbrechende Veränderungen für die Meinungsforschung mit sich – birgt aber auch

planung&analyse

PRACTICAL GUIDE

# Wie dank KI die Marktforschung besser werden kann

von Kamila Zahradnickova, Lakmoos AI

# Synthetic data could be better than real data

## Machine-generated data sets have the potential in artificial intelligence, if researchers avoid fakery.

Received: 23 November 2023 | Accepted: 30 January 2024  
DOI: 10.1002/mar.21982

### Using large language models to generate silicon samples in consumer and marketing research: Challenges, opportunities, and guidelines

Marko Sarstedt<sup>1,2</sup> | Susanne J. Adler<sup>1</sup> | Lea Rau<sup>1</sup> | Bernd Schmitt<sup>3</sup>

<sup>1</sup>LMU Munich School of Management, Institute for Marketing, Ludwig-Maximilians-University Munich, Munich, Germany  
<sup>2</sup>Marketing Department, Faculty of Economics and Business Administration, Babeş-Bolyai University, Cluj-Napoca, Romania  
<sup>3</sup>Columbia Business School, Robert D. Calkins Professor of International Business, Columbia University, New York, New York, USA

**Abstract**  
Should consumer researchers employ silicon samples and artificially generated data based on large language models, such as GPT, to mimic human respondents' behavior? In this paper, we review recent research that has compared result patterns from silicon and human samples, finding that results vary considerably across different domains. Based on these results, we present specific recommendations for silicon sample use in consumer and marketing research. We argue that silicon samples hold particular promise in upstream parts of the research process such as qualitative pretesting and pilot studies, where researchers collect exploratory information to safeguard follow-up design choices. We also provide a critical assessment and recommendations for using silicon samples in main studies. Future research should address issues of silicon sample use and present future research directions.

#### SOCIAL SCIENCE

## AI and the transformation of social science research

### Careful bias management and data fidelity are key



Working Paper 23-062

## Using LLMs for Market Research

James Brand  
Ayelet Israeli  
Donald Ngwe



## AI language models cannot replace human research participants

Laetitia Harding<sup>1</sup> · William D'Alessandro<sup>2</sup> · N. G. Laskowski<sup>3</sup> · Robert Long<sup>4</sup>

Received: 15 June 2023 / Accepted: 30 June 2023 / Published online: 21 July 2023  
Author(s), under exclusive license to Springer-Verlag London Ltd., part of Springer Nature 2023

The artificial intelligence of our projects, the priority. AI's propensity to perform

### Trends in Cognitive Sciences

#### Science & Society

## Can AI language models replace human participants?

Danica Dillion,<sup>1</sup> Niket Tandon,<sup>2</sup> Yuling Gu,<sup>2</sup> and Kurt Gray<sup>1</sup>

Recent work suggests that language models such as GPT can make human-like judgments across a number of domains. We explore whether and when language models might replace human participants in psychological science. We review nascent research, provide a theoretical model, and outline caveats of it.

an 'agent of to take over nans. Gener (LLMs) like an labor, ince, where lip edit pa, and cre- er, could substitute - that we hu

#### Does GPT make human-like judgments?

We initially doubted the ability of LLMs to capture human judgments but, as we detail in Box 1, the moral judgments of GPT-3.5 were extremely well aligned with human moral judgments in our analysis ( $r = 0.95$ ; full details at <https://arxiv.org/abs/2306.08961>). Human morality is often argued to be especially difficult for language models to capture [4] and yet we found powerful alignment between GPT-3.5 and human judgments.

We emphasize that this finding is just one anecdote and we do not make any strong claims about the extent to which LLMs otherwise. Language models also might be especially good at predicting moral judgments because moral judgments heavily including the presence of an intentional agent, the causation of damage, and a vulnerable victim, features that language models may have an easy time detecting. However, the results are intriguing.

Other researchers have empirically demonstrated GPT-3's ability to simulate human participants in domains beyond moral judgments, including predictive economic choices [11], replicating

developed a framework (Box 2) that connects LLM responses to human cognition. The model emphasizes that the 'minds' of LLMs are grounded in naturalistic expressions of people. Practically speaking, LLMs may be most useful as participants when studying specific topics, when using specific tasks, at specific research stages, and when simulating specific samples.

#### Specific topics

Language model expressions may be most correlated with human expressions when there are obvious explicit features of situations that drive human judgments. With morality, these might include whether an action was intentional or not. With mind perception, these might include whether a target is described as human or a kind of animal, and with economic behavior these might include a clear payoff matrix.

Divergence from human judgment may occur in cases with competing intuitions. Within our set of moral scenarios, humans (but not GPT-3.5) condemned coaches who rooted for the opposing team, perhaps because LLMs struggle with the subtle

## AI-Human Hybrids for Marketing Research: Leveraging Large Language Models (LLMs) as Collaborators

Neeraj Arora<sup>1</sup>, ... and Yohei Nishimura

AI-LLM (large language model) hybrid approach leads to efficiency and enhanced qualitative research, they show that LLMs can assist in both data generation and analysis, generate synthetic respondents, and conduct and moderate focus groups. LLMs possess complementary skills; the human-LLM hybrid outperforms its human-only counterpart in generating themes and summaries. Evidence from a series of experiments shows that the human-LLM hybrid is more effective than either component alone in improving through few-shot learning and retrieval-augmented generation. The authors design the synthetic respondents, and obtain responses from synthetic respondents. They provide practical guidelines for using LLMs in marketing research and conclude that LLMs serve as valuable collaborators in qualitative research, surveys, consumer insights, unstructured data, RAG, in-context learning, and generative AI.



## International Journal of Information Management

International Journal of Information Management 71 (2023) 102642  
Contents lists available at ScienceDirect  
journal homepage: [www.elsevier.com/locate/ijinfomgt](http://www.elsevier.com/locate/ijinfomgt)

## Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy\*

Yogesh K. Dwivedi<sup>a,b,\*</sup>, Nir Kshetri<sup>c</sup>, Laurie Hughes<sup>a</sup>, Emma Louise Slade<sup>d</sup>, Anand Jeyaraj<sup>e</sup>, ...

\* C. Nielsen, Jr. Chair in Information Management, University of Wisconsin, ...



LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

LMU MUNICH  
SCHOOL OF  
MANAGEMENT

# Unsere Forschung

OPEN ACCESS

## Using LLMs in sensory service research: initial insights and perspectives

在感官服务研究中使用大语言模型：初步见解与展望

Monika Imschloss <sup>a</sup>, Marko Sarstedt <sup>b,c</sup>, Susanne J. Adler <sup>b</sup> and Jun Hwa Cheah <sup>d</sup>

<sup>a</sup>Institute of Management & Organization, Leuphana University of Lüneburg Lüneburg, Germany; <sup>b</sup>Institute for Marketing, LMU Munich School of Management, Ludwig-Maximilians-University Munich, Munich, Germany; <sup>c</sup>Faculty of Economics and Business Administration, Babeş-Bolyai University, Cluj-Napoca, Romania; <sup>d</sup>Norwich Business School, University of East Anglia, Norwich, UK

### ABSTRACT

Researchers have started using large language models (LLMs), such as OpenAI's GPT, to generate synthetic datasets designed to mimic human response behavior. Several studies have systematically compared LLM-generated data with human samples in order to explore LLMs' ability to mimic consumer decision-making. Extending prior findings, our research sets out to explore how GPT-4o responds to sensory information, and to evaluate its ability to grasp crossmodal correspondences as well as multisensory congruence – as commonly encountered in service settings. Our results indicate that while GPT-4o identifies and describes sensory stimuli accurately, it often fails to replicate the associative meanings and interpretations that humans derive from these stimuli, especially in stand-alone assessments. Our research therefore underscores the need for further exploration of the conditions under which LLMs reliably mirror human responses to sensory stimuli, and the implications of using LLMs in research on sensory-rich service settings.

### 摘要

研究人员已经开始使用大语言模型 (LLMs), 例如OpenAI的GPT, 来生成旨在模拟人类响应行为的合成数据集。一些研究系统地将大语言模型生成的数据与人类样本进行了比较, 以探索大语言模型模仿消费者决策的能力。延续先前的研究, 我们的研究旨在探索GPT-4o如何响应感官信息, 并评估其理解跨模态对应关系以及多感官一致性的能力。

Received: 23 November 2023 | Accepted: 30 January 2024  
DOI: 10.1002/mar.21982

REVIEW ARTICLE

## Using large language models to generate silicon samples in consumer and marketing research: Challenges, opportunities, and guidelines

Marko Sarstedt<sup>1,2</sup> | Susanne J. Adler<sup>1</sup> | Lea Rau<sup>1</sup> | Bernd Schmitt<sup>3</sup>

<sup>1</sup>LMU Munich School of Management, Institute for Marketing, Ludwig-Maximilians-University Munich, Munich, Germany  
<sup>2</sup>Marketing Department, Faculty of Economics and Business Administration, Babeş-Bolyai University, Cluj-Napoca, Romania  
<sup>3</sup>Columbia Business School, Robert D. Calkins Professor of International Business, Columbia University, New York, New York, USA

Correspondence  
Marko Sarstedt, LMU Munich School of Management, Institute for Marketing, Ludwig-Maximilians-University Munich, Munich, Germany.  
Email: sarstedt@lmu.de

**Abstract**  
Should consumer researchers employ silicon samples and artificially generated data based on large language models, such as GPT, to mimic human respondents' behavior? In this paper, we review recent research that has compared result patterns from silicon and human samples, finding that results vary considerably across different domains. Based on these results, we present specific recommendations for silicon sample use in consumer and marketing research. We argue that silicon samples hold particular promise in upstream parts of the research process such as qualitative pretesting and pilot studies, where researchers collect external information to safeguard follow-up design choices. We also provide a critical assessment and recommendations for using silicon samples in main studies. Finally, we discuss ethical issues of silicon sample use and present future research avenues.

**KEYWORDS**  
generative AI, GPT, large language models, silicon samples, synthetic datasets

**1 | INTRODUCTION**

Generative artificial intelligence (AI) is transforming academic and practical research. A particularly prominent type of generative AI is large language models (LLMs) that can process a myriad of inputs and predict the next word or a part of the next word (referred to as a token) in a sequence. The most visible outcome of this development is, arguably, the generative pre-trained transformer (GPT) model (Brown et al., 2020; OpenAI, 2023), which was made available to the general public via ChatGPT in November 2022. GPT uses large databases of text as input, trains the model by using a self-supervised language modeling objective, and employs reinforcement learning from human feedback (OpenAI, 2023). This procedure enables LLMs to mimic human response behavior (Jeon et al., 2023; Luo et al., 2022).

Psychologists and marketing researchers have started reflecting on how LLMs might impact consumer and marketing research (e.g., Peres et al., 2023). Studies in this domains emphasize LLMs' potential to improve marketing communications (e.g., content marketing campaigns and content design), deliver superior customer experience through hyperpersonalization, and enhance classic marketing research functions (Brand et al., 2023; Ooi et al., 2023; Paul et al., 2023). Researchers have also started using LLMs to substitute human participants in academic empirical research (Argyle et al., 2023; Demsky et al., 2023; Dillon et al., 2023). These studies use LLMs to generate so-called "silicon samples" (also referred to as "synthetic datasets") that seek to mimic human respondents to describe, explain, and predict human behavior.

Silicon samples have also emerged in marketing practice. For example, the startup Synthetic Users has set up a service using LLMs

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.  
© 2024 The Author(s). *Psychology & Marketing* published by Wiley Periodicals LLC.

*Psychol Mark.* 2024;1-17.  
[wileyonlinelibrary.com/journal/mar](https://onlinelibrary.wiley.com/journal/mar) | 1

Annals of Tourism Research 108 (2024) 103819

Contents lists available at ScienceDirect

Annals of Tourism Research

journal homepage: <https://www.journals.elsevier.com/annals-of-tourism-research>

RESEARCH NOTE

## The use of synthetic data in tourism

Giampaolo Viglia<sup>a,b,\*</sup>, Susanne J. Adler<sup>c</sup>, Caroline Lancelot Miltgen<sup>d</sup>, Marko Sarstedt<sup>c,e</sup>

<sup>a</sup>University of Portsmouth, Department of Marketing, Portland Street, PO1 3DE, UK  
<sup>b</sup>Department of Economics and Political Science, University of Asolo Valley, Asolo, Italy  
<sup>c</sup>Institute for Marketing, Ludwig-Maximilians-University Munich, Munich, Germany  
<sup>d</sup>Audencia Business School, Department of Marketing, 8 Route de La Jonelière, 44312 Nantes, France  
<sup>e</sup>Faculty of Economics and Business Administration, Babeş-Bolyai University, Cluj-Napoca, Romania

ARTICLE INFO

Article history:  
Received 25 May 2024  
Received in revised form 19 July 2024  
Accepted 25 July 2024  
Available online xxxxx

Associate editor: Sara Doinic

Keywords:  
Synthetic data  
Large language models  
ChatGPT  
Experimental stimuli  
Pretest  
Prompting

**Introduction**

Will a future reliant on synthetic data for tourism research lead to progress or stagnation? While synthetic data promises effectiveness and accuracy in mimicking human behavior, its inability to capture evolving real-world behaviors raises concerns. Can we truly understand and study the dynamic nature of tourism without real-world data?

Large language models utilize vast text databases to predict the next word or part of it. These artificial intelligence systems are trained using language modeling and human feedback-based reinforcement learning (Achilam et al., 2023), enabling them to mimic human responses (Grossmann et al., 2023). While OpenAI's Generative Pre-trained Transformer (GPT) has been the primary focus in tourism and hospitality (Davišević et al., 2024), researchers have also examined other models like Gemini, Bard, and Claude 3 (Wu et al., 2023).

Researchers across various disciplines have started using large language models as substitutes for human participants in empirical studies (e.g., Argyle et al., 2023; Dillon et al., 2023; Li et al., 2024), including tourism (Sop & Kurcer, 2024). These models generate synthetic data that seek to mimic human responses to describe, explain, and predict human behavior. Deontological ethics may categorically oppose such synthetic data, arguing that they fundamentally differ from human respondent data, thus violating research values and responsibilities (Ilutson & Mastin, 2023). Conversely, a utilitarian perspective would pragmatically consider the low costs and high benefits of using large language models for such predictions (Sarstedt et al., 2024). A key argument here is that large language models are a suitable means to generate large-scale data for specific research areas very quickly and at low costs (Li et al., 2024). But would synthetic data live up to this promise in tourism research?

\* Corresponding author.  
E-mail address: [giampaolo.viglia@port.ac.uk](mailto:giampaolo.viglia@port.ac.uk) (G. Viglia).

<https://doi.org/10.1016/j.annals.2024.103819>  
0160-7383/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).



# Delphi-Studie



# Auswahl der Experten

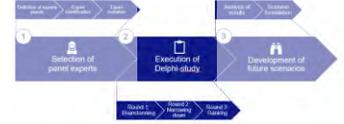
## Definition der Experten Panels



## Datenerhebungsprozess

im Zeitraum Juni 24 – Juli 24





# Aktuelle Use Cases entlang des Marktforschungsprozesses

## Einordnung der Use Cases entlang des Marktforschungsprozesses – alle Panels

Aktuelle Use Cases

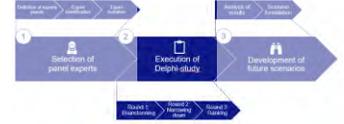
- Zusammenfassung von textbasierten Unterlagen

- Erster Entwurf von Datenerhebungsmaterialien qual. Studien
- Erster Entwurf von Datenerhebungsmaterialien quant. Studien
- Erstellung visueller Stimuli

- Transkription qual. Daten
- Qual. Datenanalyse
- Quant. Datenanalyse

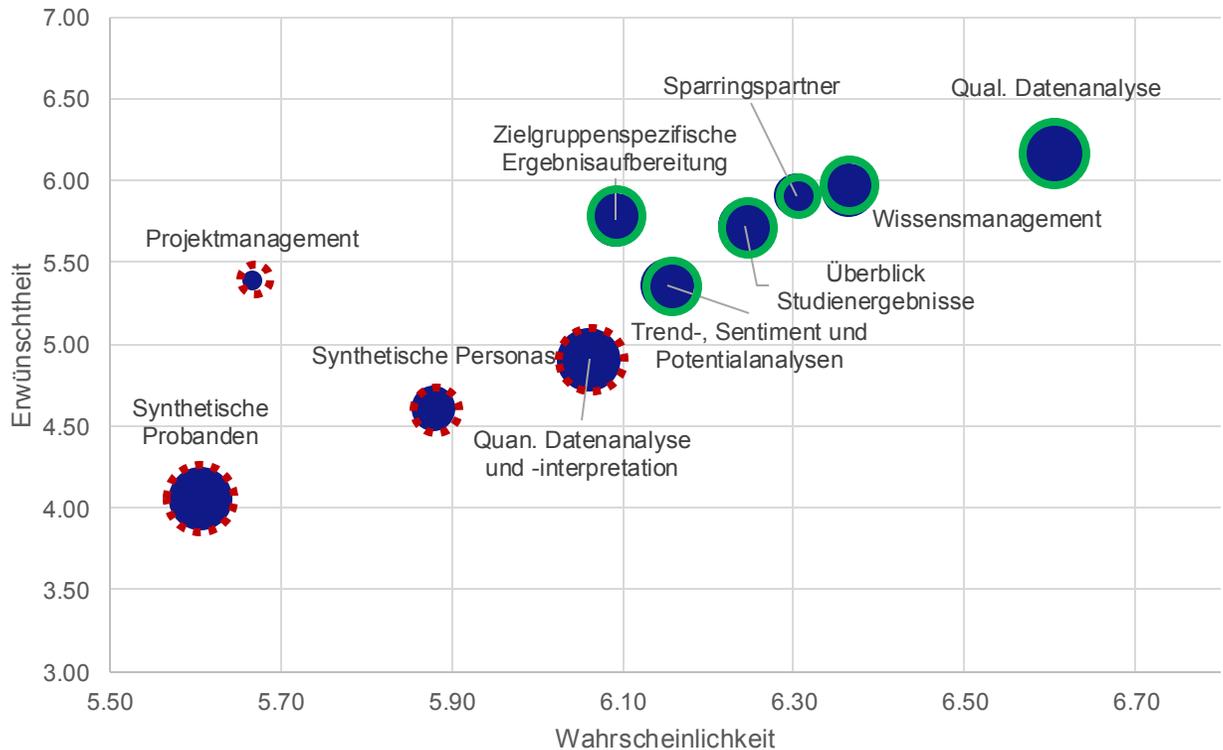
- Evaluierung und Optimierung bestehender Formulierungen
- Ausformulierung von Stichpunkten
- Übersetzungen





# Zukünftige Anwendungsbereiche von Sprachmodellen (1/2)

Einordnung der Use Cases hinsichtlich Wahrscheinlichkeit, Erwünschtheit und Brancheneinfluss – alle Panels



- ● ● Brancheneinfluss
- Konsens\* unter Expert:innen
- Dissens unter Expert:innen

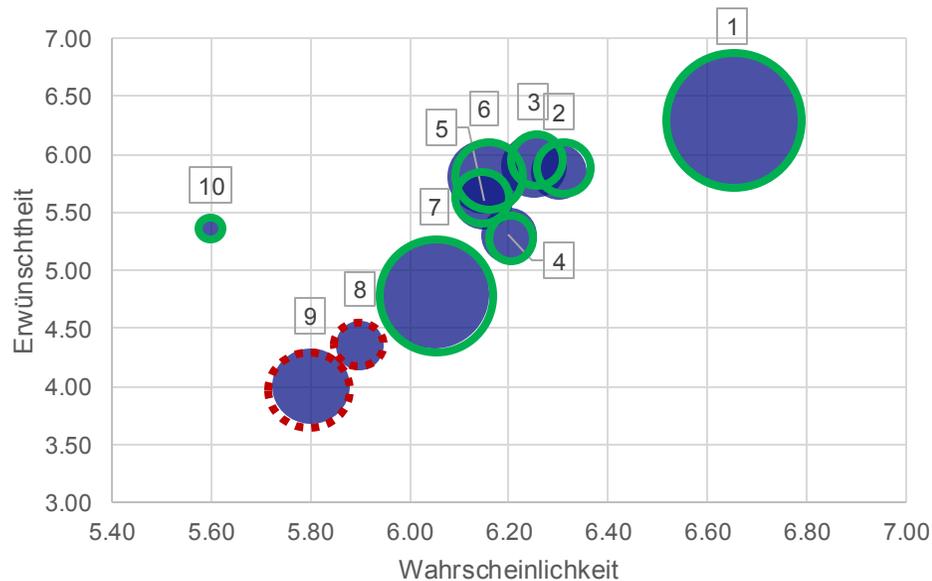
\*berechnet unter Verwendung des Interquartilsbereichs (IQR)

Nr.	Future Use Case
1	Qual. Datenanalyse
2	Wissensmanagement
3	Sparringspartner
4	Überblick Studienergebnisse
5	Trend-, Sentiment und Potentialanalysen
6	Zielgruppenspezifische Ergebnisaufbereitung
7	Quan. Datenanalyse und -interpretation
8	Synthetische Personas
9	Projektmanagement
10	Synthetische Probanden

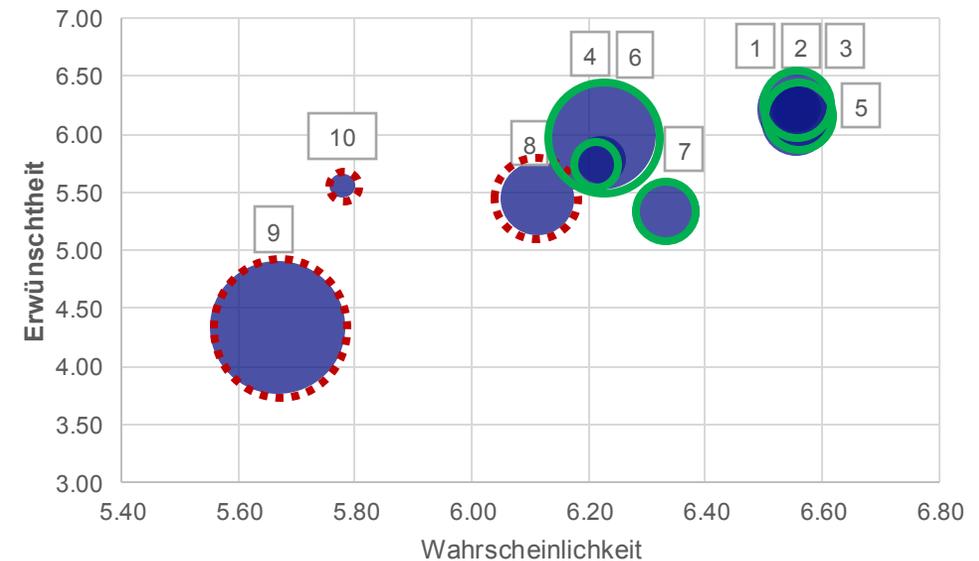
## Zukünftige Anwendungsbereiche von Sprachmodellen (2/2)

Einordnung der Use Cases hinsichtlich Wahrscheinlichkeit, Erwünschtheit und Brancheneinfluss – je Panel

Marktforschungsunternehmen (n=24)



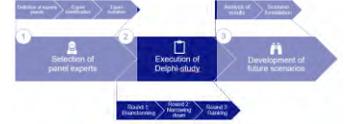
Unternehmensberatungen (n=9)



1	Qual. Datenanalyse	6	Zielgruppenspezifische Ergebnisaufbereitung
2	Sparringspartner	7	Quan. Datenanalyse und -interpretation
3	Wissensmanagement	8	Synthetische Personas
4	Trend-, Sentiment und Potentialanalysen	9	Synthetische Probanden
5	Überblick Studienergebnisse	10	Projektmanagement

- ● ● Einfluss auf Industrie
- Konsens\* unter Experten
- Dissens unter Experten

\*berechnet unter Verwendung des Interquartilsbereichs (IQR)



# Aktuelle und zukünftige Use Cases entlang des Marktforschungsprozesses

## Einordnung der Use Cases entlang des Marktforschungsprozesses – alle Panels

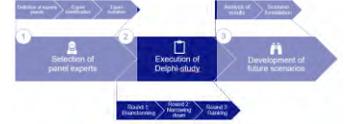
Aktuelle  
Use  
Cases

- Zusammenfassung von textbasierten Unterlagen
- Erster Entwurf von Datenerhebungsmaterialien qual. Studien
- Erster Entwurf von Datenerhebungsmaterialien quant. Studien
- Erstellung visueller Stimuli
- Transkription qual. Daten
- Qual. Datenanalyse
- Quant. Datenanalyse
- Evaluierung und Optimierung bestehender Formulierungen
- Ausformulierung von Stichpunkten
- Übersetzungen



Zukünftige  
Use  
Cases

- Projektmanagement
  - Wissensmanagement
  - Überblick Studienergebnisse
  - Sparringspartner
  - Synthetische Probanden
  - Synthetische Personas
  - Trend-, Sentiment und Potentialanalysen
  - Quant. Datenanalyse und -interpretation
  - Transkription und Qual. Datenanalyse
  - Zielgruppenspezifische Ergebnisaufbereitung
- In 10 Jahren werden Sprachmodelle zur Generierung synthetischer Probanden, die das Antwortverhalten von **(Nischen)-Kundensegmenten** imitieren (mögliche Use Cases sind z.B. Pre-Tests, Packaging-Validierungsstudien) eingesetzt.
- In 10 Jahren werden Sprachmodelle in Form von Personas (z.B. als Avatar) den realitätsnahen, **interaktiven Austausch** mit den Zielgruppen eingesetzt



# Chancen von Sprachmodellen

*Chancen von Sprachmodellen für die Markt- und Konsumentenforschung sowie Ranking nach deren Relevanz*

Nr.	Chancen
1	Vereinfachte <b>Verarbeitung und Analyse</b> großer, teilw. unstrukturierter Datenmengen
2	<b>Zeitersparnis</b> durch Übernahme repetitiver Aufgaben (erlaubt z.B. Fokus auf Konzeption des Forschungsdesigns, etc.)
3	Vereinfachte <b>zielgruppenspezifische Erstellung</b> von Reports und Ergebnispräsentationen in verschiedenen Sprachen
4	Vereinfachte Durchführung <b>internationaler Projekte</b> durch Reduzierung von Sprachbarrieren
5	<b>Sparringspartner</b> beim Entwickeln und Testen von Forschungsansätzen etc.
6	Effizienteres <b>Projektmanagement</b> (z.B. Automatisierung Rekrutings, Reminder zur Berücksichtigung übersehener Schritte, etc.)
7	Vereinfachter <b>Datenzugang</b> (z.B. mit internationalen Samples, KI-basierten Interviews, synthetischer Panels)
8	Verbesserte <b>Validität</b> durch umfangreichere Daten (z.B. mit mehr Befragten, offenen Fragestellungen, Echtzeit-Daten im Kaufprozess, Beobachtungen etc.)
9	<b>Echtzeitanalyse</b> von Trends führt zur Früherkennung negativer Stimmungen und ermöglicht zeitnahes Gegensteuern
10	<b>Konsumentenzentrierte Innovationen</b> durch greifbarere Zielgruppen und verbesserten Feedbackprozess

Kendall's  $W = 0,18$

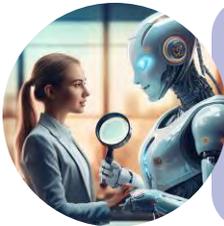


## Szenarien in Eskalationsstufen



### Szenario 1: Kollaborativer Austausch mit Sprachmodellen

- **Rollen:** Sprachmodelle als kompetente unermüdlicher Mitarbeiter; Marktforscher fungieren als qualifizierte Instruktoren
- **Synthetische Daten:** Ausschließlich für Pretests
- **Konsequenz:** höhere Effizienz und Effektivität im Alltag; Marktforscher kann sich auf strategische Aufgaben konzentrieren



### Szenario 2: Marktforscher als Manager

- **Rollen:** Sprachmodelle als primäre Wissensträger; Marktforscher als steuernder Manager mit lediglich Grundlagenwissen
- **Synthetische Daten:** Größtenteils synthetische Probanden
- **Konsequenz:** Verlust von operativem Wissen führt zu wachsende Abhängigkeit von Sprachmodellen, sinkender Bedarf an menschlichen Marktforschern



### Szenario 3: Autonomes Forschungsökosystem

- **Rollen:** Sprachmodelle als Marktforscher, die eine autonome Durchführung des gesamten Marktforschungsprozesses für explorative, deskriptive und kausale Fragestellungen ermöglichen
- **Synthetische Daten:** Synthetische Probanden ersetzen menschliche
- **Konsequenz:** Marktforscher nur als Forschungsinitiator, keine Notwendigkeit mehr für menschliche Probanden





# MARKO SARSTEDT



sarstedt@lmu.de



[linktr.ee/markosarstedt](https://linktr.ee/markosarstedt)



UNIVERSITATEA BABEȘ-BOLYAI  
BABEȘ-BOLYAI TUDOMÁNYEGYETEM  
BABEȘ-BOLYAI UNIVERSITÄT  
BABEȘ-BOLYAI UNIVERSITY  
TRADITIO ET EXCELLENTIA



Academy of  
Marketing  
Science



BUNDESVERBAND  
MARKETING  
CLUBS