

JAIR

March 2023 | No 1

Journal of Applied
Interdisciplinary Research



**Focus Issue:
Artificial Intelligence**

JAIRS

Journal of Applied
Interdisciplinary Research



JAIR

Journal of Applied
Interdisciplinary Research

PUBLISHERS

Prof. Dr. rer. nat. Peter Sperber
(President of the Deggendorf Institute of Technology – DIT)
Prof. Dr.-Ing. Andreas Grzempa
(Vice President Research and Knowledge Transfer – DIT)

praesident@th-deg.de
andreas.grzempa@th-deg.de

EDITORIAL TEAM

Prof. Dr. Michelle Cummings-Koether (principal editor)
Dr. Kristin Seffer (principal editor)
Esther Kinateder (proofreader)
Sandra Maier (typesetting)
Steffen Menzel (technical support)

michelle.cummings-koether@th-deg.de
kristin.seffer@th-deg.de
esther.kinateder@th-deg.de
sandra.maier@th-deg.de
steffen.menzel@th-deg.de

EDITORIAL BOARD

Prof. Dr. med. habil. Thiha Aung (DIT)
Prof. Dr. Georgi Chaltikyan (DIT)
Prof. Dr. Matthias Huber (DIT)
Prof. Dr. Katerina Volchek (DIT)

thiha.aung@th-deg.de
georgi.chaltikyan@th-deg.de
matthias.huber@th-deg.de
katerina.volchek@th-deg.de

DESIGN

Kathrin Weindl, Lukas Haselberger

ISSN: 2940-8199

The Journal of Applied Interdisciplinary Research (JAIR) is published as a series of the Bavarian Journal of Applied Sciences.

CONTACT US

Journal of Applied Interdisciplinary Research (JAIR)
Technische Hochschule Deggendorf
Dieter-Görlitz-Platz 1
94469 Deggendorf, Germany
Phone: ++49 (0) 991 3615-0
Fax: ++49 (0) 991 3615-297
E-Mail: jair@th-deg.de
Web: <https://jas.bayern>

Copyright © 2023
Deggendorf Institute of Technology
All rights reserved.

This is an open access journal which means that all content is freely available without charge to the user or his / her institution. Users are allowed to read, download, copy, distribute, print, search, or link to the full texts of the articles in this journal without asking for prior permission from the publisher or the author under the condition that the original publication is properly cited. This is pursuant to the BOAI definition of open access.

CONTENT

FOREWORD	6
<i>Michelle Cummings-Koether & Kristin Seffer</i>	
EDITORIAL	7
<i>Cordula Krinner</i>	
ARTICLES	
<i>Noemi Funke, Katja Stadler, Heidi Vakkuri, Anna Wagner, Marc Lunkenheimer, Alexander H. Kracklauer</i>	10
“Your conversational partner is a chatbot”—An Experimental Study on the Influence of Chatbot Disclosure and Service Outcome on Trust and Customer Retention in the Fashion Industry	
<i>Madeleine Taglinger, Stephanie Jordan, Alexander H. Kracklauer</i>	28
Acceptance of Artificially Intelligent Digital Humans in Online Shops: A Modelling Approach	
<i>Václav Stehlík, Mouzhi Ge</i>	50
TOPSIS-Based Recommender System for Big Data Visualizations	
IMPRINT	75

Foreword

Dear readers,

Welcome to our first edition of the Journal of Applied Interdisciplinary Research, short JAIR. This academic journal aims to provide a current and international overview of interdisciplinary research which is also done in an applied manner. The combination of these two types of research is a niche that has so far found little attention in academic journals and we are happy to close a previously existing gap by combining these two types of research in its own new journal. As this type of research is a growing field, it warrants its own journal. Various areas of academia are overlapping more and more, so we want to provide an opportunity for researchers to publish their interdisciplinary research in a journal dedicated to advancing this particular field, and committed to the exchange of ideas across academic disciplines.

The JAIR will be generally published annually, and each issue will be dedicated to a primary research field that will be the basis for contributions. The published issue will usually be in combination with an academic conference or exchange, but we welcome contributions from authors who want to add to the discussion, even if they did not attend the companion event.

Our first issue is dedicated to the primary field of Artificial Intelligence. The companion event to this issue was a workshop series and a conference with the title “Clashing Approaches to Artificial Intelligence – Finding New Ideas”, short AI Clash, that was held at the Deggendorf Institute of Technology in 2022. The aim of this issue is to look at different fields that AI Research can be applied to. The contributions for this first issue do just that.

We would like to take this opportunity to thank all reviewers who have contributed so much to this issue with their high-quality reviews.

In 2023, additionally a special issue of the JAIR will be published that has the primary field of Medical Informatics. And a call for contributions for the second issue to be published in 2024 will follow in the Summer of 2023.

We are dedicated to providing a valuable peer-reviewed journal with high academic standards for your research needs and look forward to your contributions. Ideas for primary topics are also welcome, but only in conjunction with an academic event.

Your JAIR editors,

Michelle J. Cummings-Koether

Kristin Seffer

Editorial

Cordula Krinner*

Welcome to the inaugural edition of the Journal of Applied Interdisciplinary Research! We are excited to present to you a collection of articles that explore the complexities and controversies surrounding the rapidly advancing field of artificial intelligence (AI).

As AI continues to make significant impacts in various sectors, from healthcare to finance to transportation, it has also sparked intense debate and disagreement. Some argue that AI has the potential to revolutionize industries and improve the quality of life for people around the world, while others caution that it could have unintended consequences and potentially threaten the very fabric of society.

In this edition, we bring together experts from a range of disciplines, including computer science, philosophy, economics, and social sciences, to examine these conflicting viewpoints and offer diverse perspectives on the future of AI.

In one article, a team of computer scientists present the latest research on the capabilities and limitations of AI, examining how it can be harnessed to solve complex problems and improve decision-making. In another, a philosopher explores the ethical implications of AI, considering questions of responsibility and accountability in a world where machines are increasingly making decisions that affect our lives.

We also feature articles that delve into the economic and social impacts of AI, examining the potential for job displacement and the need for fair and responsible deployment of these technologies.

As the field of AI continues to evolve and shape the world we live in, it is more important than ever to have a nuanced and interdisciplinary understanding of its potential and pitfalls. We hope that the articles in this edition will provide valuable insights and stimulate thought-provoking discussions on the role of AI in society.

We are grateful to all of our contributors for their valuable contributions to this edition, and we look forward to continuing to explore the complex and evolving field of AI in future issues of the Journal of Applied Interdisciplinary Research

* Deggendorf Institute of Technology, Dieter-Görlitz-Platz 1, 94469 Deggendorf

In case you have in parallel been reading the table of contents for this first edition of the Journal of Applied Interdisciplinary Research (JAIR) you might be wondering that the titles of the three articles do not quite match the wording above. The reason is quickly explained: I (Cordula Krinner, human) have not written the text above. The text in italics was generated by ChatGPT using my prompt: “Write an editorial for a new journal titled ‘Journal of Applied Interdisciplinary Research’. The journal’s first edition will be issued in February, 2023. The journal covers clashing views on artificial intelligence.”

ChatGPT (Chat Generative Pre-trained Transformer) [1] is a product by the American research laboratory OpenAI [2] and their latest modification of the GPT-3 family of large language models [3]. Released on November 30th, 2022 it has been an instant success – one million users within one week of launch date, 100 million users in January 2023 [4]. The potential applications of ChatGPT have been widely discussed, most relevant for the readership of this journal are probably the implications for academia: scientists co-authored papers jointly with ChatGPT [5,6]. Teachers and professors are quite concerned it will be used for cheating/plagiarizing on take-home exams, essays, and thesis projects [7]. First countermeasures have already been implemented, e.g. GPT Zero [8]. Incidentally, GPT Zero correctly diagnoses the ChatGPT-generated passage at the beginning: “Your text is likely to be written entirely by AI.”

Given that neither AI tools nor I have prophetic qualifications, I will not try to predict which implications ChatGPT in particular and Large Language Models in general will have on the labor market or the future of the human race. Currently, I am inclined to assume that ChatGPT will follow the Gartner Hype Cycle [9], especially considering wider negative implications, which were discussed for large language models in general in the controversial position paper by Bender et al. [10].

However, as this one limited example demonstrates quite impressively, the field of artificial intelligence has and will continue to have profound impacts on our life – at work, in mobility, in private. Thus, I am indeed very honored to contribute these opening comments for the first edition of JAIR and introduce (for real, this time), the three contributions to this journal: (1) Acceptance of artificially intelligent digital humans in online shops: a modelling approach. The authors investigate factors influencing consumer acceptance of artificially intelligent digital humans in online stores using linear regression analyses to identify determinants of intention to use digital humans. Two variables had a significant influence on behavioral intention: performance expectancy and habit; the first variable was the strongest positive predictor. (2) “Your conversational partner is a chatbot” – An Experimental Study on the Influence of Chatbot Disclosure and Service Outcome on Trust and Customer Retention in the Fashion Industry. Interestingly enough, this study shows that disclosing chatbot identity influences neither trust nor customer retention, but service outcome has an effect on both. This serves as a strong reminder that functional customer service is of paramount importance. (3) TOPSIS-based Recommender System for Big Data Visualizations. This paper analyzes state-of-the-art data visualization techniques for big data. Following this analysis, it presents a recommendation model based on selecting data features and proposes an approach to validate the applicability of the system.

As you can see, exiting and relevant research in the field of AI and clashing views on AI are not restricted to big tech companies in the US or China. I hope you find this journal edition enlightening! The editorial team of JAIR is looking forward to receiving your feedback on this issue and your contributions to future issues.

References and notes

- [1] <https://openai.com/blog/chatgpt/>
- [2] <https://openai.com/>
- [3] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33, 1877-1901. <https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfc4967418bfb8ac142f64a-Abstract.html>
- [4] <https://www.demandsage.com/chatgpt-statistics/>
- [5] <https://www.nature.com/articles/d41586-023-00107-z>
- [6] Kung, T. H., Cheatham, M., Medinilla, A., ChatGPT, Sillos, C., De Leon, L., ... & Tseng, V. (2022). Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models. *medRxiv*, 2022-12. <https://doi.org/10.1101/2022.12.19.22283643>
- [7] <https://www.dailymail.co.uk/sciencetech/article-11513127/ChatGPT-OpenAI-cheat-tests-homework.html>
- [8] <https://gptzero.me/>
- [9] <https://www.gartner.com/en/research/methodologies/gartner-hype-cycle>
- [10] Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610-623). <https://doi.org/10.1145/3442188.3445922>

“Your conversational partner is a chatbot” — An Experimental Study on the Influence of Chatbot Disclosure and Service Outcome on Trust and Customer Retention in the Fashion Industry

Noemi Funke*, Katja Stadler*, Heidi Vakkuri*, Anna Wagner*,
Marc Lunkenheimer*, Alexander H. Kracklauer*

ABSTRACT

Should companies disclose their chatbots’ nonhuman identity or not? Previous studies have found both negative and positive consumer reactions to chatbot disclosure. This experimental study explores how trust and customer retention change when the nonhuman identity of the chatbot is revealed and when different service outcomes apply in the context of the German fashion industry. The results of this experiment provide evidence that disclosing chatbot identity influences neither trust nor customer retention, but service outcome has an effect on both.

Companies should therefore focus on developing a functional customer service as chatbot failure has tremendous consequences for the volume of reliable customers and profits. The main limitation of this study is that the respondents were only shown screenshots, leaving the impact of a real interaction with chatbots undiscovered.

KEYWORDS

Chatbot, customer retention, identity disclosure, service outcome, trust

1. Introduction

“Hello, how can I help you today?” This is a chatbot ready to help customers and facilitate their experience on a website. Chatbots are sometimes difficult to distinguish from human conversational partners [1]. This presents companies with the challenge of considering whether the identity of their chatbots should be revealed to users or not. Studies have been conducted on this topic, with varying

* University of Applied Sciences Neu-Ulm, Wileystrasse 1, 89231 Neu-Ulm, Germany
<https://doi.org/10.25929/jair.v1i1.113>

results. Some previous studies have been pessimistic about the disclosure of chatbots in general. For example, Luo et al. [2] suggest that the disclosure of artificial agents has negative effects; their data reveal that exposing the identity of a chatbot before a conversation reduces purchase rates by more than 79.7%. On the other hand, Mozafari et al. [3] present a more positive perspective: When the outcome is not satisfactory and the identity of the chatbot is revealed, the effect of disclosure is positive.

This inconsistency in results merits further investigation. This study aims to replicate the concept of recent publications with a varied research design. The study focuses on the fashion industry, as there is a particular need for automated order processes due to the vast number of order processes in e-commerce and the high returns rate. Chatbots can take over simple service tasks such as changing order details. For this reason, the focus here is not on purchase abandonment rates due to a sales call, as in Luo et al. [2], but on easier-to-complete, more frequent service processes in the context of e-commerce and the associated longer-term customer loyalty. According to Mozafari et al. [3] chatbot disclosure has a negative indirect effect on customer retention through mitigated trust for services with high criticality. In cases where a chatbot fails to handle the customer's service issue, disclosing the chatbot identity not only lacks negative impact but elicits a positive effect on retention.

Part of the purpose of this study is to answer the following question:

Research Question (RQ) 1: How does disclosing chatbot identity influence customer retention? [3]

Customers nowadays expect a positive service outcome. Belanche et al. [4] point out that customers react differently to chatbot and human service failures. According to Chizhik and Zherebtsova [5] building a full-fledged chatbot that interacts with humans in a human-like manner is a very challenging and complex task. Therefore, Adam et al. [6] report that many users are still not content to interact with chatbots (e.g., because of their high failure rate), which may lead to skepticism and resistance to the technology. This in turn may discourage users from following the recommendations or requests made by a chatbot.

Previous research shows that users' behaviors differ according to whether they think they are interacting with a chatbot or a human being [7]. According to Følstad et al. [8], trust in a chatbot is due not only to the perceived characteristics of the chatbot but also and more particularly to the service context in which the chatbot is situated. It is noteworthy that the brand that provides the chatbot is crucial for trust. In other words, users are more likely to trust chatbots provided by trusted brands. From these observations, the following question emerges:

RQ2: Does service outcome moderate the effect of chatbot disclosure on trust or customer retention? [3]

Customer trust has a major influence on customer retention and the intention to buy [9]. For instance, if a customer feels angry about the service, this is likely to have a negative impact on customer retention (i.e., the customer feels the need to show aversion towards the company by no longer buying the product). This negative impact may extend to other products of the company [10]. The following research question derives from these points:

RQ3: Can the impact of trust on customer retention also be observed in this study?

This article is structured as follows: After a summary of the existing literature in the fields of chatbot disclosure, trust, service outcome, and customer retention, the research framework is laid out, along with the hypotheses. Next, this paper presents an experimental simulation of four kinds of interactions with a chatbot and investigates whether disclosure of the chatbot's identity has a positive or negative impact on customer retention in different frontline service situations. Finally, the findings are summarized, and implications are identified.

2. Theoretical background

Technological progress allows consumers to be better informed, receive targeted offers, and gain faster access to services [11]. New technologies continue to shape the ways shoppers choose channels, products, and services and how they make purchases [11].

With the current technological standard, today's consumers are demanding 24/7 service for assistance. Artificial intelligence (AI) has become a huge part of addressing this need. As a result, companies are rapidly looking to develop text-based automated conversational agents, i.e., chatbots and virtual assistants, to answer questions customers may have at any time of the day [12]. Chatbots are required not only to provide users with necessary consultancy and guidance but also to communicate in a friendly and social way with the aid of the ability to interpret natural (not pre-defined) language used by humans [7, 19, 32]. According to Chizhik and Zherebtsova [5], building a human-like chatbot is a very challenging and complex task. Despite chatbots' performance not always being error-free, chatbots are replacing human chat service agents due to the time and cost savings they offer [1].

During the COVID-19 pandemic, there has been a radical change in retail: The trend toward online shopping has drastically increased [13]. McKinsey & Company [14] report that in 2020, within eight months, e-commerce's share of fashion sales rose from 16 % to 29 % globally, which equals six regular years of growth. According to Luce [12] and Silvestri [13], specialized chatbot services are well used and will become the norm for fashion brands looking for AI-assisted product discovery, product care, and customer service. Today, customers use their smartphones to navigate on e-commerce websites, and chatting has become the most popular method of communication, especially for younger generations.

Chatbot disclosure

Machine learning endows chatbots with interaction skills for impersonating human behavior in order to meet high expectations for good customer service [15]. Sometimes this makes it challenging for users to determine whether they are interacting with a machine or a human when this information is not explicitly provided [16].

Skjuve et al. [17] find that even when the true nature of the conversational agent is unclear to the user, communication is still feasible when the transaction is easy and when expectations about the capabilities of the agent do not lead to inadequate or frustrating service outcomes. In the context of customer service, not revealing the identity of the chatbot can be useful in some situations, especially in situations where human and automated agents overlap seamlessly. Furthermore, Corti and Gillespie [18] point out that finding common ground seems to be more important in interactions with a non-disclosed chatbot with human-like behavior than in interactions with a chatbot whose behavior is not human-like.

Mozafari et al. [3] find that transparently communicating chatbot identity generates positive user reactions. When interactions with chatbots are successful, there is no significant effect on trust. However, in cases of chatbot failure, chatbot disclosure has a significant positive effect on trust. Luo et al. [2], however, suggest the opposite: the disclosure of artificial agents has negative effects due to people's subjective perceptions of machines. Mozafari et al. [3] add that the mere knowledge that users are interacting with a chatbot rather than a human causes a biased reaction. However, if the disclosure of the chatbot identity is combined with selectively presented information about the chatbot, the disclosure dilemma can be solved appropriately. Even with responses that are thoughtful, responsive, and polite, Hendriks et al. [19] find users' perceptions and evaluations of the whole process to be completely changed by chatbot impersonation. If these selected variables, which include humanity and satisfaction, are considered in terms of the overall user experience, their study shows that users still prefer to talk to a real person instead of a chatbot.

Trust in chatbots

Trust is one of the most important factors in building customer loyalty and in strengthening and retaining a relationship [20]. The previously reviewed research shows that users behave differently according to whether they think they are interacting with a chatbot or a human being. Despite promising forecasts, many chatbots built for commercial use underperformed in practice and had to be shut down, which according to industry reports was mainly necessitated by a general lack of trust in chatbots. This lack of trust also prevents widespread adoption of chatbots. This is particularly evident in human–computer interaction [7]. While chatbots mimic human behavior and even replace humans in their tasks, building trust with humans works differently from building trust with chatbots [21]. According to Komiak and Benbasat [22] trust is defined as the willingness to rely on an exchange partner, more specifically the willingness to rely on the trustee to be able to fulfill their obligations (i.e., competence), to tell the truth (i.e., integrity) and to act in the trustor’s interest (i.e., benevolence) [3].

The concept of trust is more complex in the context of AI-enabled customer service, where trust is limited not only by technology and brand but also by the purpose and process of using AI [23]. Følstad et al. [8] add that trust in a chatbot is due not only to its perceived characteristics but also, and more notably, to the service context. Building trust takes time; trust can be broken within seconds and needs a long time to be repaired. This is a dynamic process, moving from initial trust to the ongoing development of trust. To achieve the latter, a chatbot should be reliable and easy to use. The fear that AI will replace and displace jobs, along with the potential threat of AI to the existence of humanity, creates distrust and hinders the continued development of trust [23]. The findings of Ameen et al. [24] show, on the one hand, that consumer trust is an important factor to consider in AI adoption and AI experiences. On the other hand, achieving trust presents the greatest of challenges given the complexity and ambiguity of AI technology. Results also confirm that there is a positive relationship between trust and service quality. Thus, concentrating on trust, providers could offer (among other things) better service quality in terms of security, interface design, and reliability.

In cases of negative service outcomes, identifying the conversational partner as a chatbot increases trust and retention [3]. In customer service situations, a chatbot that is human-like enough to recognize a potential misunderstanding seems adequate. The ability to resolve miscommunication appears to be as effective as avoiding it.

Service outcome

When buying a product or service, consumers experience a purchase outcome that leaves them either satisfied or dissatisfied. They then try to determine what led to this result, which influences their future buying behavior. Customers are more likely to search for a responsible entity when the outcome was a failure [10]. The customer attributes responsibility for the negative service outcome to either the product, the employee, or the firm rather than to him- or herself, being convinced that it could have been avoided [10]. As a covariate, however, responsibility attribution has no effect on the relationship between chatbot disclosure, service outcome, and their interaction with trust, according to Mozafari et al. [3]. When experiencing a positive service outcome, customers tend to be less satisfied with the service provider when the conversational partner is a chatbot. A reason for this could be that the customers attribute the positive outcome to themselves and are therefore not surprised [25].

The customer’s expectations play a large role in the causal attribution of responsibility for the service outcome. Unexpected outcomes are accompanied by significantly more spontaneous attributions [26]. Attributions to the chatbot of both controllability and service stability are high when negative emotions emerge about the service outcome, meaning that service failure increases customers’ reason to complain and, therefore, the spread of negative word of mouth [27]. Dealing with negative service outcomes is

linked to anger and helplessness [28]; subsequent explanation of why an error occurred mitigates that anger and helplessness [28].

Belanche et al. [4] show that customers react differently to chatbot and human service failures. The attribution of responsibility to the agent is greater when the agent is a human employee. The agent's responsibility is assigned differently depending on the service outcome. For example, in case of a service failure, little responsibility is attributed to a chatbot, while a human agent is seen as responsible in a similar manner as for a service success. Customers may assume that chatbots are less capable of solving service issues than humans [4]. According to Mozafari et al. [3], this mistrust can be overcome by revealing the identity of the chatbot. This leads, in cases of chatbot failure, to higher levels of trust and retention. Regarding these findings, Blut et al. [1] discuss the merits of anthropomorphism: whether a chatbot should be designed to imitate human behavior or if it should be obvious to the customer when the agent is a chatbot. They conclude that anthropomorphism creates a positive effect because the human-like appearance allows the customer to act as in a human-to-human encounter, using the same social rules.

Customer retention

Even though companies are always striving to improve customer retention and satisfaction, it is ultimately always the customer who makes the decision, for various reasons, whether to stay with or leave the company [29]. For instance, if a customer feels anger with the service, this is likely to have a negative impact on customer retention. The customer feels the need to punish the company by not buying the product anymore, a decision that may extend to other products of the company [10]. Confessing the mistake helps the customer to understand that it was not due to a malicious act on the company's part but only an error, encouraging the customer to believe that the mistake will not be repeated [10].

Silitonga et al. [30] find that on e-commerce websites, the factors with the greatest effect on buyer retention are customization, contact interactivity, care, character, and trust [30]. To gain trust, companies should design their websites in a way that customers would perceive as easy to use and functional. When websites are not designed like this, trust decreases, and customers reduce their participation in online commerce with the companies. Customer trust therefore has a notable influence on customer retention and intention to buy [9]. In a mediation analysis, Mozafari et al. [3] demonstrate that a significant relationship between chatbot disclosure, service outcome, and retention exists only through the mediator of trust. Their research also shows chatbot disclosure to have a positive effect on customer retention when the service outcome is negative and when it is influenced by trust. The effect of chatbot disclosure or service outcome on retention shows no significance.

3. Research framework

For scientific progress, it is important to review research findings by replicating results from a third party. This aims to make the findings more reliable. According to Pesaran [7], there are two types of replication studies, replication in a narrow sense and replication in a wide sense. Replication in a narrow sense means checking the data of the original study for consistency and accuracy. The aim of such a study is to verify the accuracy of the data analysis performed as described in the original study. Replication in a wide sense aims to verify the research method as such. For this purpose, the original data can be used, or new data can be generated and evaluated under changed conditions (e.g., with a change in the date of the data generation, the gender of the participants, or the industries under scrutiny). The aim of such a study is to verify the findings and check whether they apply in general or just under certain conditions. This study belongs to the latter category. It replicates the study by Mozafari et al. [3] in order to verify whether their findings can be transferred to another, sufficiently different industry. If

so, these results can be seen as a first hint whether a chatbot's identity disclosure has similar effects in other industries as well.

As the current state of research shows, there is still no definitive answer to the question of whether chatbot identity should be disclosed as studies have observed contrasting results. This leaves the question whether the identity should be communicated honestly still open for discussion [17]. Accordingly, this paper focuses on the effects of service outcome and chatbot disclosure on trust and customer retention in the context of the fashion industry. Trust is one of the most important factors for customer loyalty [20]. Particularly in the case of a negative service outcome, the customer reacts with frustration [28]. Yet, if the service outcome is a failure and chatbot identity is revealed, Belanche et al. [4] say that customers attribute little responsibility to the chatbot, although they would give the chatbot more credit if it were a human being. However, successful service outcomes have no significant effect on trust according to Mozafari et al. [3], whereas when service outcomes are negative, chatbot disclosure has a positive effect on trust. To test the findings of Mozafari et al. [3] for another sector, the first hypothesis transfers this context to the online fashion industry:

Hypothesis (H) 1: If chatbot failure occurs, disclosing (vs. not disclosing) chatbot identity enhances trust in the conversational partner.

As the customer always has the last decision, either staying with or leaving the company, customer retention is an important aspect to consider [29]. A service outcome that triggers the feeling of anger can reduce retention, causing other services to be avoided in the future. For this reason, the second hypothesis addresses whether chatbot identity disclosure increases customer retention in the event of a chatbot failure in the online fashion industry, as follows:

H2: If chatbot failure occurs, chatbot disclosure (vs. non-disclosure) enhances customer retention.

Additionally, this paper considers the relation between trust and customer retention separately. Trust has a major influence on retention and the intention to buy [30]. Since research shows that trust and customer retention are contiguous, this paper tests the effect of trust on retention for the German online fashion industry. Consequently, the third and last hypothesis is formulated as follows:

H3: Trust has a positive effect on customer retention.

4. Research design

A cross-sectional online survey was conducted using a convenience sample, distributed via WhatsApp messages and e-mail as the most efficient ways to reach people. A screening question at the beginning of the survey made sure that the sample included only respondents who had interacted with a customer service feature of a homepage via online chat in the past year. Mozafari et al. [3] did not include such a screening question. However, since it would bias the results if people took part in the survey who had never experienced such customer service, it was necessary to insert a screening question at the beginning of the survey. The initial sample size was $n = 191$; after the screening question, the sample decreased to $n = 148$. Attention checks were included in the survey, as they were by Mozafari et al. [3], to make sure that people did not answer the questions randomly. Those who did not pass the attention checks ("Please tick the scale point (5) if you have read the questionnaire carefully" and "What was the service request you approached the company with?") or did not fill out the questionnaire consciously ("Have you answered the questionnaire consciously?") were excluded from further analysis, so the total final sample size was $n = 128$, with $n = 64$ male and $n = 64$ female respondents.

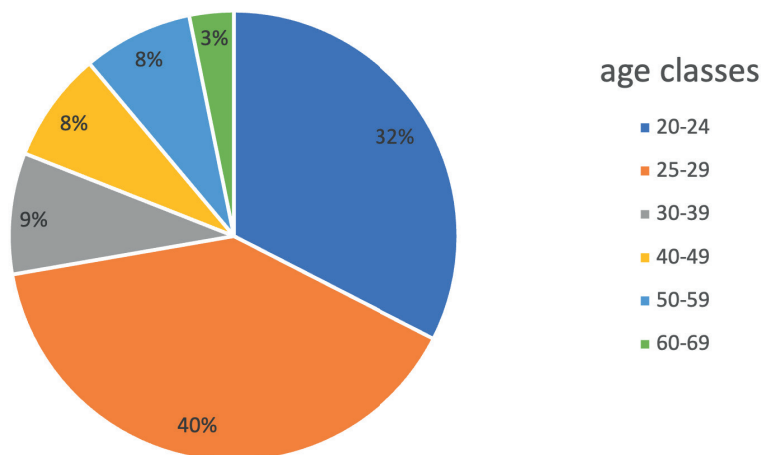


Figure 1: Age distribution in years (n = 128).

Figure 1 illustrates the age distribution of the participants. The average respondent age was 31 years. The procedure including a screening question at the beginning of the survey resulted in a relatively young sample of participants. This phenomenon can be explained by the fact that customers who have experience in dealing with chatbots are predominantly young people. A study conducted in Switzerland found that 53 % of 18–30 year-olds have already had contact with a chatbot. In the 31–50 age group, it is 42 % and of those over 50, it is only 27 % [31]. Thus, the overrepresentation of the younger age group can be explained by this. With regard to gender, equal distribution was taken into account in order to achieve isomorphism here as well.

In contrast to the study by Mozafari et al. [3], in which participants had to imagine that they were customers of an energy provider, the participants in this study had to imagine being customers of a fashion label. The scenario was the following: They, as customers, had just ordered a pair of trousers online but had used the wrong shipping address. Now, they would like to change the address for the package delivery via the customer service’s online chat. “Faulty” addresses are a well-known phenomenon in the e-commerce sector - ambiguity, spelling errors, invalid or incomplete addresses are examples for that. Each of the four groups received a different outcome for this scenario. The participants were sent four such screenshots in a row, showing the chronological development of the chat. This ensured that the participants did not already know at the beginning how the chat would proceed (see Figure 2 for a sample scenario). In the first subsample, which included n = 31 participants, the respondents experienced a successful change of the delivery address. The experience of the n = 32 members of the second subsample was a successful change of shipping address and the disclosure that the interaction was with a chatbot and not a human service employee. In the third subsample, the n = 34 respondents tried to change shipping address, which proved unsuccessful. Finally, the experience of the n = 31 members of the fourth subsample was a failed attempt to change the address and chatbot disclosure. The answers were measured with a Likert scale from 1 = strongly disagree to 7 = strongly agree. The conversations were presented in German to the German-speaking participants and translated for the purpose of this publication afterward.

	Chatbot success	Chatbot failure
Chatbot non-disclosure	n = 31	n = 34
Chatbot disclosure	n = 32	n = 31

Table 1: The distribution of respondents among groups.

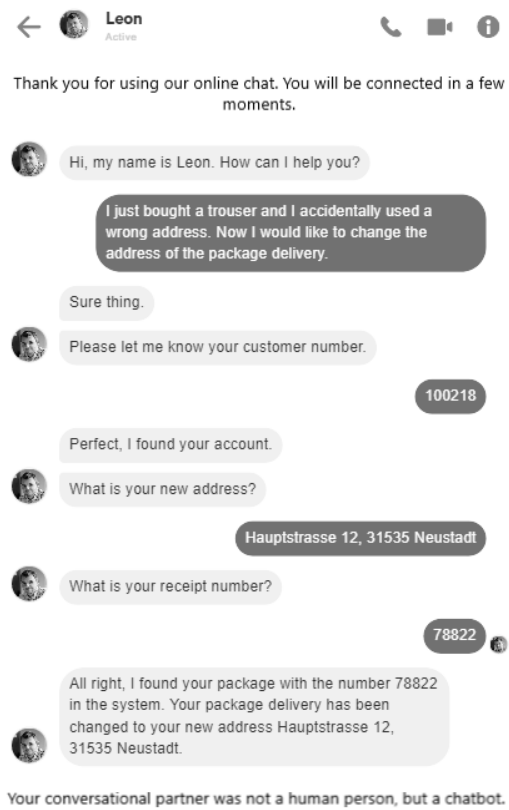
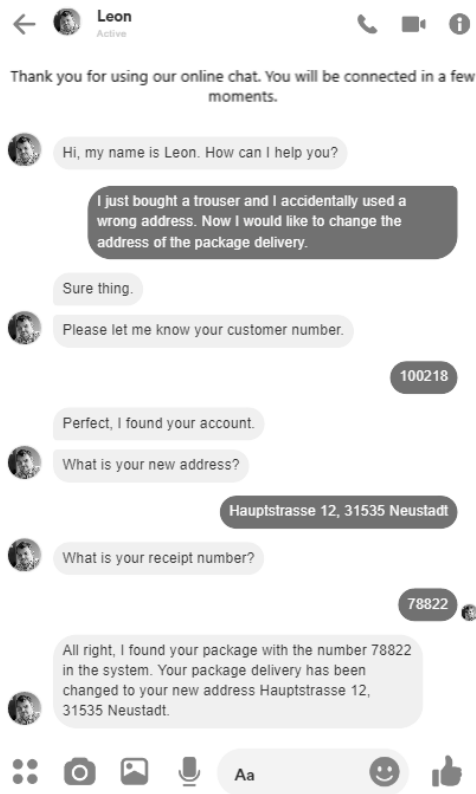


Figure 2: Sample scenario: success with non-/disclosure condition.

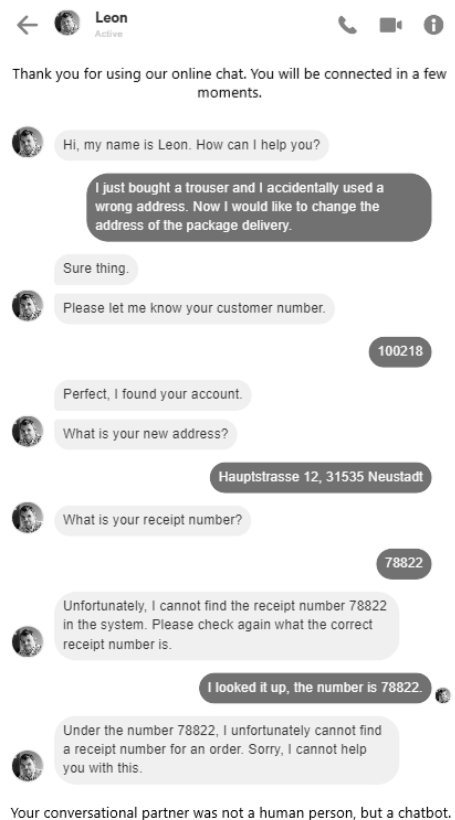


Figure 3: Sample scenario: failure with non-/disclosure condition.

5. Findings

The findings of this study provide new insights into the target phenomenon and contribute to the overall research on trust in relation to service performance and the revelation or non-revelation of the identity of a chatbot. In particular, the results show that regardless of the disclosure or non-disclosure of a chatbot’s identity, the success of a chatbot in achieving the desired service outcome leads to higher trust than the chatbot’s failure. Furthermore, analysis of the effects of service outcome and chatbot disclosure on customer loyalty shows that only service outcome has an impact. Together, these results can influence the design of corporate chatbot systems in terms of whether and under what circumstances the identity of a chatbot should be revealed.

To test the data against manipulation and for validity, this paper follows the approach of Mozafari et al. [3]. The manipulation check (“Do you think you talked to an automated chatbot or a human service employee?”) is statistically significant (see Table 5 in the Appendix). All statistically significant relationships in this paper are significant at the 95 % level of confidence. Compared to respondents who did not know the identity of their conversational partner, respondents experiencing chatbot disclosure were significantly more likely to perceive their conversational partner as a chatbot than as a human service employee ($M_{disclosed} = 1.13, SD = 0.34, M_{undisclosed} = 1.45, SD = 0.50, t = 4.22$). The validity check (“Which entity was responsible for the service outcome? Me, the customer, or Leon, the employee?”) shows whether respondents really perceived customer service failure differently from success. As this relationship is also statistically significant (see Table 6 in the Appendix), it indicates that respondents experiencing service failure attribute the responsibility for the service outcome significantly more to the employee, Leon, than the group that experienced service success ($M_{failure} = 6.25, SD = 1.12, M_{success} = 5.33, SD = 1.66, t = -3.67$).

To test reliability, a Cronbach’s alpha test was conducted, testing the internal consistency of the scales measuring the concepts of trust and customer retention. As Table 2 shows, the Cronbach’s alpha values for both concepts are above 0.7, indicating reliability [3]. However, the Cronbach’s alpha for the construct trust is above 0.9, which would hint at possible redundant values, meaning that one dimension in the concept is very similar to another dimension. A high correlation is good, but too high a correlation needs to be avoided as it would display multicollinearity. Yet, this is a matter of interpretation, as a broader concept is better for inferences but decreases reliability. Consequently, in this paper this slightly higher value of 0.91 for the construct trust is accepted. The component analysis shows the contribution of each item to the variance in the construct. The higher the value, the more valuable the information contributed by a certain item in explaining the variance within a construct. The question “Would you give the trousers back if you received them?” shows a notable low item loading with a value of 0.46 for the construct customer retention. However, leaving the variable out of the construct increases Cronbach’s alpha to 0.93, which is higher than the current value of 0.89. Consequently, the variable is not excluded from the construct.

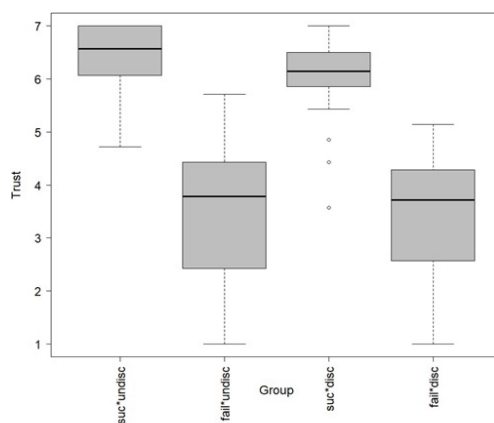


Figure 4: SEQ Figure * ARABIC 4: Boxplot of trust for each group.

<i>Construct</i>	<i>Dimension</i>	<i>Measurement</i>	<i>Component analysis</i>	<i>α</i>
Trust in the conversational partner [32]	Competence	Does the conversational partner have the necessary skills to deliver the service?	0.87	0.91
	Competence	Does the conversational partner have access to the information needed to handle my service request adequately?	0.74	
	Integrity	Is the conversational partner's conduct in response to my service request fair?	0.84	
	Integrity	Does the conversational partner have high integrity?	0.84	
	Benevolence	Is the conversational partner receptive to my service request?	0.81	
	Benevolence	Does the conversational partner make efforts to address my service request?	0.82	
	Overall trust	Is the conversational partner trustworthy overall?	0.79	
	Customer retention [32]		Would you continue being a customer of this fashion label?	
		Would you buy additional products beyond this pair of trousers from this fashion label in the future?	0.90	
		If you had to decide, would you select this fashion label again?	0.88	
		Would you return the trousers if you received them? (R)	0.46	
		Would you intend to switch to another fashion label? (R)	0.81	
		Would you plan to abandon this fashion label? (R)	0.88	

Table 2: Measures of multi-item constructs and of dimension and construct reliability.
 Notes: R = reverse scaled items; α = Cronbach's alpha. Concepts based on Mozafari et al. [3].

Figure 4 provides a descriptive view of how trust is distributed across the four groups: chatbot success with non-disclosure of chatbot identity, chatbot failure with non-disclosure, chatbot success with disclosure, and chatbot failure with disclosure. One can see that the groups that experienced chatbot success had a higher level of trust than the groups with chatbot failure, consonant with the findings of Mozafari et al. [3]. Mozafari et al. [3] did not find an effect deriving from the covariate “responsibility service outcome” while conducting an ANCOVA analysis. Consequently, this paper follows their approach and conducts an ANOVA instead [3]. The ANOVA analysis of chatbot disclosure, service outcome, and their interactions with trust indicates that only service outcome has a strong positive significant effect, with a p-value of <0.001 (see Table 7 in the Appendix). While this, however, only explains the variance, meaning that the groups differ significantly, a post-hoc test is conducted to further analyze which groups display significant differences. The t-test shows whether the difference among groups is significant by comparing means. It reveals that chatbot disclosure, compared to non-disclosure, had no significant impact on trust when the service outcome was a successful change of address ($M_{\text{success*non-disclosure}} = 6.35$, $M_{\text{success*disclosure}} = 6.05$, $p = 0.45$). Surprisingly, chatbot disclosure also had no significant effect on trust when the service outcome was a failure ($M_{\text{failure*non-disclosure}} = 3.51$, $M_{\text{failure*disclosure}} = 3.41$, $p = 0.69$). This result stands in contrast to Mozafari et al. [3]. Consequently, the significance in the ANOVA can only derive from the difference between success and failure as service outcomes. Analyzing these means shows that success had a positive and significant impact on trust compared to failure when the identity of the chatbot was not revealed ($M_{\text{non-disclosure*success}} = 6.35$, $M_{\text{non-disclosure*failure}} = 3.51$, $p < 0.001$). The same applies to the case of chatbot identity disclosure ($M_{\text{disclosure*success}} = 6.05$, $M_{\text{disclosure*failure}} = 3.41$, $p < 0.001$). The insignificance of the effect of chatbot disclosure or non-disclosure on trust is shown by comparing the groups experiencing chatbot success and chatbot failure. Independent of chatbot disclosure or non-disclosure, chatbot success as a service outcome meant higher levels of trust compared to chatbot failure (see Table 3; also see Tables 9, 10, and 11 in the Appendix for the individual trust dimensions). Therefore, the first hypothesis needs to be rejected.

	Chatbot success	Chatbot failure
Chatbot non-disclosure	M = 6.35, SD = 0.65	M = 3.51, SD = 1.24
Chatbot disclosure	M = 6.05, SD = 0.77	M = 3.41, SD = 1.13

Table 3: Mean and standard deviation for the concept trust.

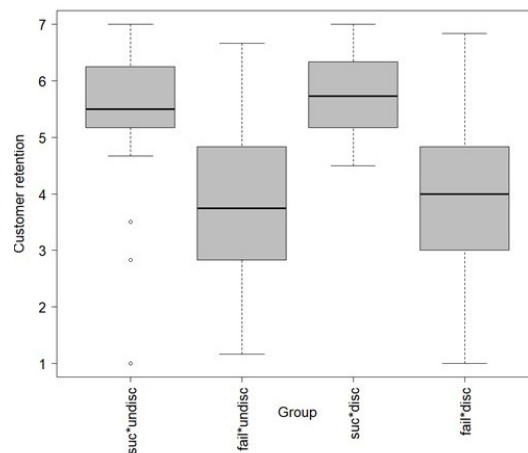


Figure 5: SEQ Figure * ARABIC 5: Boxplot of customer retention for each group.

Figure 4 provides a descriptive view of how customer retention is distributed across the four groups. One can see that the groups experiencing chatbot success had higher levels of customer retention than the groups with chatbot failure. The ANOVA analysis of chatbot disclosure, service outcome, and their interactions with customer retention indicates that only service outcome had a strong positive significant effect, with a p-value of <0.001 (see Table 8 in the Appendix). The post-hoc test shows that chatbot disclosure, compared to non-disclosure, had no significant impact on customer retention when the service outcome was a successful change of address ($M_{\text{success*non-disclosure}} = 5.46$, $M_{\text{success*disclosure}} = 5.77$, $p = 0.63$). Chatbot disclosure also had no significant effect on retention when the service outcome was failure ($M_{\text{failure*non-disclosure}} = 3.80$, $M_{\text{failure*disclosure}} = 3.88$, $p = 0.78$). Consequently, the significance in the ANOVA can only derive from the difference between success and failure as service outcomes. Analyzing these means shows that success had a positive and significant impact on customer retention compared to failure when the identity of the chatbot was not revealed ($M_{\text{non-disclosure*success}} = 5.46$, $M_{\text{non-disclosure*failure}} = 3.80$, $p < 0.001$). The same applies to the case of chatbot identity disclosure ($M_{\text{disclosure*success}} = 5.77$, $M_{\text{disclosure*failure}} = 5.88$, $p < 0.001$). The insignificance of the effect of chatbot disclosure or non-disclosure on customer retention is shown by comparing the groups that experienced chatbot success with the groups that experienced chatbot failure (see Table 4). Independent of chatbot disclosure or non-disclosure, chatbot success as a service outcome was shown to mean higher levels of customer retention compared to chatbot failure. Therefore, the second hypothesis also needs to be rejected.

	Chatbot success	Chatbot failure
Chatbot non-disclosure	M = 5.46, SD = 1.24	M = 3.80, SD = 1.30
Chatbot disclosure	M = 5.77, SD = 0.74	M = 3.88, SD = 1.37

Table 4: Mean and standard deviation for the concept customer retention.

To analyze Hypothesis 3, a regression analysis was conducted (see Table 5). Sorting out the missing data points left 128 observations. The R2 shows that 51.6 % of the variation in the data can be explained with this model, which is a mediocre prediction, leading to the conclusion that other factors also influence the relationship. The F statistic of 138.408 shows that the overall model is significant as the independent variable has a highly significant positive effect on the dependent variable, with a t-value of 11.593. Hence, trust influences customer retention positively. Increasing trust by 1 increases customer retention by 0.63. The relationship is statistically significant. Consequently, as this supports the assumption, the third hypothesis cannot be rejected.

Dependent variable:	
retention	
trust	0.630 t = 11.593***
Constant	1.682 t = 6.071***
Observations	128
R ²	0.516
Adjusted R ²	0.512
Residual Std. Error	1.033 (df = 126)
F Statistic	134.408*** (df = 1; 126)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 6: SEQ Figure * ARABIC 6: Regression table of the impact of trust on customer retention (n = 128).

Figure 6 displays the descriptive relationship between trust and customer retention. It shows the observation points and the regression as well as the confidence intervals. It clearly reveals that trust and customer retention are highly correlated with each other.

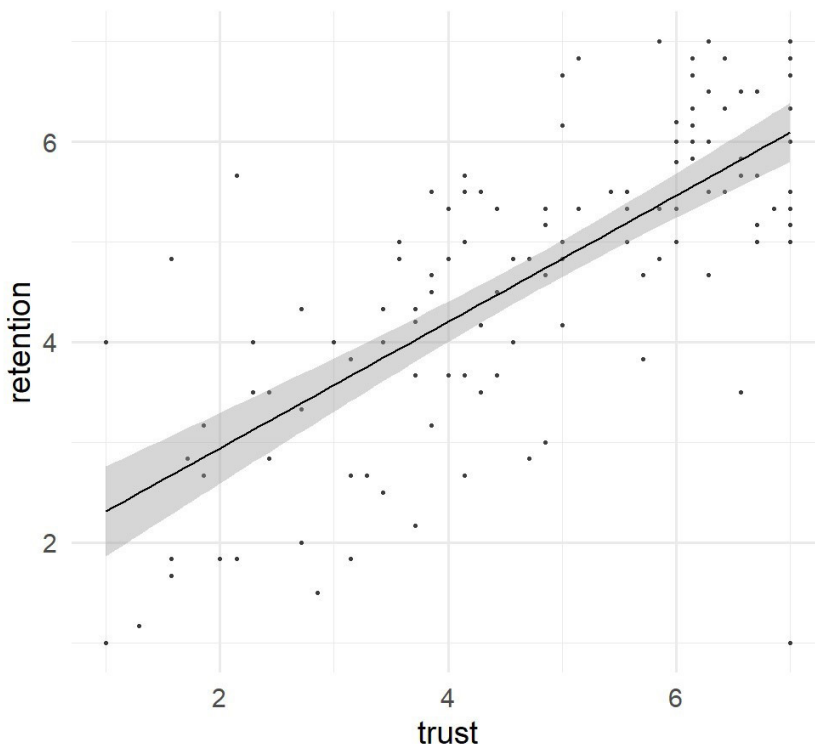


Figure 7: SEQ Figure * ARABIC 7: Regression of trust on customer retention with confidence intervals (n = 128).

6. Discussion

In summary, this paper supports the potential cross-sectoral validity of some of the findings of Mozafari et al. [3] but cannot confirm others for the online fashion industry.

Like Mozafari et al. [3], this paper shows that experiencing a positive service outcome leads to higher levels of trust compared to a negative service outcome. In general, chatbot disclosure does not have an impact on trust, neither in this paper nor in the study by Mozafari et al. [3]. Differentiating between the two service outcomes also leads to no effect of chatbot disclosure on trust when the service outcome is a success. This applies for Mozafari et al. [3] as well. However, in the present study, chatbot disclosure in cases of failure as a service outcome still does not influence trust. This result stands in contrast to Mozafari et al. [3]. To sum up, while chatbot disclosure becomes important in the case of chatbot failure in the European energy sector, for online fashion industry, only service outcome has an impact on trust, leaving disclosure and non-disclosure out of the equation.

In analyzing whether customer retention is influenced by service outcome or chatbot disclosure, an effect is found only for service outcome. Looking at the groups experiencing chatbot success and failure separately shows no impact of chatbot disclosure on customer retention in conjunction with either of the service outcomes. This leads to the conclusion that service outcome is the important factor and not chatbot disclosure, which is generally in line with current research [10, 3]. Nonetheless, this paper also stands in contrast to new research findings as it finds, unlike Mozafari et al. [3], that disclosure is also

not important in cases of chatbot failure. While looking at the relationship between trust and customer retention, which was already discovered in previous research [9], this paper reports evidence that trust does also influence customer retention in the online fashion industry.

Although the results of this paper contribute to the current state of literature by challenging a few recent findings, there are some limitations of this study that need to be addressed. As the respondents were quite young, findings might differ for an older population less familiar with digital technology; in that case, chatbot disclosure in combination with service outcome might have a greater effect. Additionally, a comparison of male and female respondents might be illuminating. It should be noted that as with Mozafari et al. [3], respondents were shown only screenshots, leaving the impact of a real interaction with chatbots undiscovered. In addition, testing other constructs of trust and customer retention would be conceivable. Beyond this, it would be interesting to see whether the same results apply in another industry or country, since the scope of this study is on the German sector only. It would be also interesting to replicate the study of Luo et al. [2] concerning negative effects of chatbot disclosure and purchase rates in the fashion sector; however, this would need another research design. From a quantitative point of view, looking at the regression model again shows that it can only predict about 50 % of the variance, leading to the conclusion that other factors influence the relationship as well. Future research should address these issues.

7. Conclusion and implications

Finally, to answer the original research questions, this paper shows that disclosing chatbot identity influences neither trust nor customer retention. It reveals that service outcome not only moderates the effect of chatbot disclosure on trust and customer retention — it is the sole factor having an impact. Lastly, this paper also extends the widely discussed influence of trust on customer retention to the online fashion industry.

Contributing to existing research are some implications deriving from this research. Consumers are generally quite critical when it comes to new technology. However, this study shows that service outcome is more important for trust in a brand and customer retention than is revealing chatbot identity. Brands offering services via chatbots should keep the critical attitude of consumers in mind. However, they should focus on developing functional customer service as chatbot failure has tremendous consequences for levels of trust and customer retention, which impact the brand directly through decreased volumes of reliable customers and profits.

Conflict of interest statement

The authors declare that there is no conflict of interest in connection with the present work.

References and notes

- [1] Blut, M., Wang, C., Wunderlich, N. V.; Brock, C. (2021): Understanding anthropomorphism in service provision: a meta-analysis of physical robots, chatbots, and other AI. In: *Journal of the Academy of Marketing Science*, pp. 632–658.
- [2] Luo, X., Tong, S., Fang, Z.; Qu, Z. (2019): Frontiers: Machines vs. Humans: The Impact of Artificial Intelligence Chatbot Disclosure on Customer Purchases. In: *Marketing Science*, vol. 38 (6), pp. 937-947.
- [3] Mozafari, N., Weiger, W. H.; Hammerschmidt, M. (2021b): Trust me, I'm a bot – repercussions of chatbot disclosure in different service frontline settings. In: *Journal of Service Management*, vol. 33 (2), pp. 221-245.
- [4] Belanche, D., Casaló, L. V., Flavián, C.; Schepers, J. (2020): Robots or frontline employees? Exploring customers' attributions of responsibility and stability after service failure or success. In: *Journal of Service Management*, pp. 267–289.
- [5] Chizhik, A.; Zhrebtsova, Y. (2020): Challenges of Building an Intelligent Chatbot, *Proceedings of the International Conference "Internet and Modern Society" (IMS-2020)*.
- [6] Adam, M., Wessel, M.; Benlian, A. (2021): AI-based chatbots in customer service and their effects on user compliance, in: *Electronic Markets*, vol. 31, (2), pp. 427–445.
- [7] Pesaran, H. (2003): Introducing a replication section. In: *Journal of Applied Econometrics* (1), pp. 111.
- [8] Følstad, A., Nordheim, C. B.; Bjørkli, C. A. (2019): What Makes Users Trust a Chatbot for Customer Service? An Exploratory Interview Study. In: *Internet Science: 5th International Conference*, vol. 2019, pp. 194–208.
- [9] Ranaweera, C.; Prabhu, J. (2003): On the relative importance of customer satisfaction and trust as determinants of customer retention and positive word of mouth. In: *Journal of Targeting, Measurement and Analysis for Marketing*, vol. 12, pp. 82-90.
- [10] Weiner, B. (2000): Attributional Thoughts about Customer Behaviour. In: *Journal of Consumer Research*, vol. 27 (3), pp. 382-387.
- [11] Grewal, D., Roggeveen, A. L.; Nordfält, J. (2017): The Future of Retailing. In: *Journal of Retailing* (1), p. 1–6.
- [12] Luce, L. (2019): *Artificial Intelligence for Fashion. How AI is Revolutionizing the Fashion Industry*, New York: Apress.
- [13] Silvestri, B. (2020): The Future of Fashion: How the Quest for Digitization and the Use of Artificial Intelligence and Extended Reality Will Reshape the Fashion Industry After COVID-19. In: *ZoneModa Journal*, 10 (2), pp. 61-73.
- [14] McKinsey & Company (2020): *The State of Fashion 2021*. Available at: <https://www.mckinsey.com/~media/mckinsey/industries/retail/our%20insights/state%20of%20fashion/2021/the-state-of-fashion-2021-vf.pdf> [Accessed: 23 August 2022].
- [15] Hyken, S. (2017): AI And Chatbots Are Transforming The Customer Experience [Online]. Available at <https://www.forbes.com/sites/shephyken/2017/07/15/ai-and-chatbots-are-transforming-the-customer-experience/?sh=71bbe5ad41f7> [Accessed: 19 December 2021].
- [16] Candello, H., Pinhanez, C.; Figueiredo, F. (2017): Typefaces and the Perception of Humanness in Natural Language Chatbots, CHI '17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 3476–3487.
- [17] Skjuve, M., Haugstveit, I. M., Følstad, A.; Brandtzaeg, P. B. (2019): Help! Is my chatbot falling into the uncanny valley? An empirical study of user experience in human-chatbot interaction. In: *Human Technology*, pp. 30–54.
- [18] Corti, K.; Gillespie, A. (2016): Co-constructing intersubjectivity with artificial conversational agents: People are more likely to initiate repairs of misunderstandings with agents represented as human. In: *Computers in Human Behavior*, vol. 58, pp. 431–442.

- [19] Hendriks, F., Ou, C. X.J., Khodabandeh Amiri, A.; Bockting, S. (2020): The Power of Computer-Mediated Communication Theories in Explaining the Effect of Chatbot Introduction on User Experience, Proceedings of the 53 Hawaii International Conference on System Sciences.
- [20] Rai, A. K.; Srivastava, M. (2014): Customer Loyalty: Concept, Context and Character. New York: McGraw Hill.
- [21] Chen, J. V., Thi Le, H.; Tran, S. T. T. (2021): Understanding automated conversational agent as a decision aid: matching agent's conversation with customer's shopping task. In: Internet Research (4), pp. 1376–1404.
- [22] Komiak, S.X., Benbasat, I. (2004), Understanding customer trust in agent-mediated electronic commerce, web-mediated electronic commerce, and traditional commerce. In: Information Technology and Management, Vol. 5 No. ½, pp. 181-207.
- [23] Siau, K.; Wang, W. (2018): Building Trust in Artificial Intelligence, Machine Learning, and Robotics. In: Cutter Business Technology Journal, vol. 31 (2), pp. 47-53.
- [24] Ameen, N., Tarhini, A., Reppel, A.; Anand, A. (2021): Customer experiences in the age of artificial intelligence. In: Computers in human behavior, vol. 114.
- [25] Jin, R. (2010): Causal Attributions on Service Outcomes by Self-Service Technology Users, Purdue University.
- [26] Kanazawa, S. (1992): Outcome or Expectancy? Antecedent of Spontaneous Causal Attribution. In: Personality and Social Psychology Bulletin, vol. 18 (6), pp. 659–668.
- [27] van Vaerenbergh, Y., Orsingher, C., Vermeir, I.; Larivière, B. (2014): A Meta-Analysis of Relationships Linking Service Failure Attributions to Customer Outcomes. In: Journal of Service Research, pp. 381–398.
- [28] Gelbrich, K. (2010): Anger, frustration, and helplessness after service failure: coping strategies and effective informational support. In: Journal of the Academy of Marketing Science, pp. 567–585.
- [29] Kumar, V., Bhagwat, Y.; Zhang, X. (2015): Regaining “Lost” Customers: The Predictive Power of First-Lifetime Behavior, the Reason for Defection, and the Nature of the Win-Back Offer. In: Journal of Marketing, pp. 34–55.
- [30] Silitonga, K. A. A., Ikhsan, R. B.; Fakhrorazi, A. (2020): Drivers of buyer retention in e-commerce: The role of transaction characteristics and trust. In: Management Science Letters, pp. 3485–3494.
- [31] Pidas AG (2018): Chatbot-Studie – Die digitalen Helfer im Praxistest. Available at: https://cdn2.hubspot.net/hubfs/5893787/PIDAS_ZHAW_Chatbot_Studie_FINAL.pdf
- [32] Bhattacharjee, B. (2002): Individual Trust in Online Firms: Scale Development and Initial Test. In: Journal of Management Information Systems, pp. 211–241.

Appendix

Test-statistic	Df	p-value	Alternative hypothesis	Mean in Group 1	Mean in Group 2
4.221	126	4.613e-05 ***	Two-sided	1.446	1.127

Table 5: Manipulation check.

Test-statistic	Df	p-value	Alternative hypothesis	Mean in Group 1	Mean in Group 2
-3.665	126	0.0003631 ***	Two-sided	5.333	6.246

Table 6: Validity check.

	Df	Sum Sq	Mean Sq	F-value	Pr (> F)
Service outcome	1	239.46	239.46	246.933	< 2e-16 ***
Chatbot disclosure	1	1.25	1.25	1.293	0.258
Service outcome* disclosure	1	0.33	0.33	0.341	0.560
Residuals	124	120.25	0.97		

Table 7: ANOVA analysis of service outcome, chatbot disclosure, and their interaction with trust (significance codes: 0 '***', 0.001 '**', 0.1 '*', 0.05 '.', 0.1 ' ', 1).

	Df	Sum Sq	Mean Sq	F-value	Pr (> F)
Service outcome	1	101.07	101.07	71.518	6.39e-14 ***
Chatbot disclosure	1	1.18	1.18	0.833	0.363
Service outcome* disclosure	1	0.38	0.38	0.272	0.603
Residuals	124	175.24	1.41		

Table 8: ANOVA analysis of service outcome, chatbot disclosure, and their interaction with customer retention (significance codes: 0 '***', 0.001 '**', 0.1 '*', 0.05 '.', 0.1 ' ', 1).

	Chatbot success	Chatbot failure
Chatbot non-disclosure	M = 6.61, SD = 0.66	M = 3.56, SD = 1.89
Chatbot disclosure	M = 6.45, SD = 0.73	M = 3.18, SD = 1.53

Table 9: Mean and standard deviation for the trust dimension competence.

	Chatbot success	Chatbot failure
Chatbot non-disclosure	M = 6.13, SD = 0.93	M = 3.56, SD = 1.44
Chatbot disclosure	M = 5.83, SD = 1.04	M = 3.66, SD = 1.49

Table 10: Mean and standard deviation for the trust dimension integrity.

	Chatbot success	Chatbot failure
Chatbot non-disclosure	M = 6.65, SD = 0.62	M = 3.21, SD = 1.67
Chatbot disclosure	M = 6.23, SD = 1.00	M = 3.32, SD = 1.54

Table 11: Mean and standard deviation for the trust dimension benevolence.

Acceptance of Artificially Intelligent Digital Humans in Online Shops: A Modelling Approach

Madeleine Taglinger*, Stephanie Jordan*, Alexander H. Kracklauer*

ABSTRACT

The UTAUT2 model is used to investigate the factors that influence consumer acceptance of artificially intelligent digital humans in online stores. Digital humans can be defined as a digital avatar that can mimic a full range of human behaviors (Ward, Boom, and Majenburg 2022). Six simple linear regression analyses are conducted to identify the determinants of intention to use digital humans. In the final multiple regression model, which includes the influences of six independent latent variables and three control variables (gender, age, and experience) on behavioral intention, statistically significant influences are identified for two variables: performance expectancy and habit. The results show that there is a tendency to accept the use of digital humans in online stores. Performance expectancy emerges as the strongest positive predictor of behavioral intention. In addition, hedonic motivation shows a positive influence on behavioral intention in the simple regression analysis, while the multiple regression results show a minimal negative correlation. The results may provide important insights into the adoption of innovative digital human technologies.

KEYWORDS

Digital human, innovation, artificial intelligence, UTAUT2, online shopping

1. Introduction

Changes in the global economic and political landscape, combined with the ongoing COVID-19 pandemic, have created a need for rapid digital innovation in retail (Lim 2021, p. 103). Online retailers face the challenge of matching the product advice, brand loyalty, and communication of brick-and-mortar retail. To achieve this, online retailers must be able to create interactions with customers and provide a more natural and engaging customer experience (Denner 2021). It therefore makes sense for retailers

* University of Applied Sciences Neu-Ulm, Wileystrasse 1, 89231 Neu-Ulm, Germany
<https://doi.org/10.25929/jair.v1i1.127>

to look at the applications of artificial intelligence (AI) and associated opportunities and challenges (Denner 2021). While a few years ago it was enough to provide a flawless customer experience, today companies know that their role is to enhance the human experience (Ward 2020). At the same time, the need of consumers for social interaction is increasing (Ward 2020).

Digital humans could be a solution for a more interactive, personalized, and modern shopping experience. A digital human can be defined as a digital avatar that can mimic a full range of human body language. Supported by AI, they can interpret the customer's input and return both the facts that consumers need and appropriate nonverbal responses (Ward, Boom, and Majenburg 2022). Digital humans combine the natural language processing abilities of chatbots with emotional intelligence. They use tones of voice, body language, and facial expressions to transmit empathy and kindness. A digital human can bridge the digital divide by offering the best of both worlds (AI Forum of New Zealand 2019; Futurside 2022), employing both conversational AI and machine learning (NTT DATA Business Solutions AG 2022).

A digital human can provide faster response times with less effort, freeing staff to address more complex tasks, and can provide personalized and consistent care at scale (UneeQ 2020). For customers, this means a significant improvement in the online experience, with personalized recommendations and interactions that feel empathetic, friendly, and trustworthy (Mills and Liu 2020, p. 3).

Digital humans are already being deployed in some industries, such as healthcare, financial services, retail, automotive, real estate, telecommunications, and technology (Futurside 2022). Since the ability to create digital humans is still in its infancy, there are many research gaps in the literature. In particular, there has been little research into consumer acceptance of digital humans. As a result, online retailers are hesitant to invest in the technology.

This study fills this research gap by determining which factors influence the acceptance of AI-supported digital humans. Six hypotheses are tested, based on the UTAUT2 model of Venkatesh and Bala.

The remainder of the paper is structured as follows. In the second section, the theoretical background of UTAUT2 is discussed and the hypotheses are outlined. In the third section, the research methodology is presented, with a detailed discussion of the rationale behind the data collection methods. In the fourth section, the results are presented and analyzed. The paper concludes with a discussion of the limitations of the study and possible future extensions.

2. Theoretical background and hypotheses

To uncover the factors influencing consumer acceptance of AI-powered digital humans, the research model and the hypotheses of this study were developed based on the UTAUT2 model (Venkatesh et al. 2003; Venkatesh, Thong, and Xu 2012). Five out of the seven original UTAUT2 constructs – behavioral intention, performance expectancy, effort expectancy, social influence, hedonic motivation, and habit – and one extended construct – trust – were adapted to the context of digital humans, as illustrated in Figure 1. The integration features of the UTAUT2 model make it well-suited to understanding the adoption and use of AI technologies. The extension of the model developed here, designed to explain the use of technologies in consumer markets, is appropriate for studying the adoption and use of specific applications of AI, like digital humans, in online purchasing situations.

Behavioral intention

Behavioral intention describes the extent to which an individual intends to use a particular technology (Fishbein and Ajzen 1975, p. 228). In acceptance research in the field of AI technologies (Gursoy et al. 2019, p. 169; Lu, Cai, and Gursoy 2019, p. 43), acceptance is operationalized as a hypothetical variable

based on behavioral intention. According to Venkatesh et al. (2003, p. 427), forecasts of the actual use behavior of these systems can be derived based on consumers' behavioral intentions.

Performance expectancy

The performance expectancy reflects the expected individual utility of a new technology for the user (Venkatesh et al. 2003). The perceived benefits of a technology can motivate potential users to adopt it. Transferring the variable into the context of online shopping with the assistance of digital humans, performance expectancy (PE) means the degree to which a consumer expects to experience a performance advantage from using digital humans. This leads to the following hypotheses:

Hypothesis 1 – Performance expectancy (PE) positively influences the behavioral intention to use digital humans.

Effort expectancy

The effort expectancy of a technology is the extent to which users perceive it to be easy to learn and use. If users believe it is easy, they are more likely to use it (Venkatesh et al. 2003). In addition to performance expectancy, effort expectancy has also been shown to be a significant positive predictor of intention to use in previous acceptance studies of AI-based technologies (Schwendener 2018, p. 55). Therefore, the following hypothesis is proposed:

Hypothesis 2 – Effort expectancy (EE) positively influences behavioral intention to use digital humans.

Social influence

Users often rely on opinions and experiences from their social environment when evaluating new technologies (Venkatesh et al. 2003, pp. 451–453). For the purposes of this study, social influence is the extent to which consumers perceive that influential people can lead them to believe that they should use digital humans in online stores. Acceptance studies on AI technologies have confirmed a positive correlation between social influence and intention to use (Schwendener 2018, p. 55). This leads to the following hypothesis:

Hypothesis 3 – Social influence (SI) positively influences behavioral intention to use digital humans.

Hedonic motivation

Hedonic motivation is the fun or pleasure derived from using technology (Venkatesh, Thong, and Xu 2012). In the context of this study, hedonic motivation is defined as the extent to which a consumer perceives the use of digital humans during the customer journey as fun, entertaining, and enjoyable. This leads to the following hypothesis:

Hypothesis 4 – Hedonic motivation (HM) positively influences the behavioral intention to use digital humans.

Habit

Habit is the extent to which an individual believes that their behavior is a result of experience (Venkatesh, Thong, and Xu 2012). When looking beyond the initial acceptance of a technology, habit has proven to be an important factor in the willingness to use technology and integrate it into one's daily life (Kim, Malhotra, and Narasimhan 2005; Limayem, Hirt, and Cheung 2007; Venkatesh, Thong, and Xu 2012). This leads to the following hypothesis:

Hypothesis 5 – Habit (HT) positively influences behavioral intention to use digital humans.

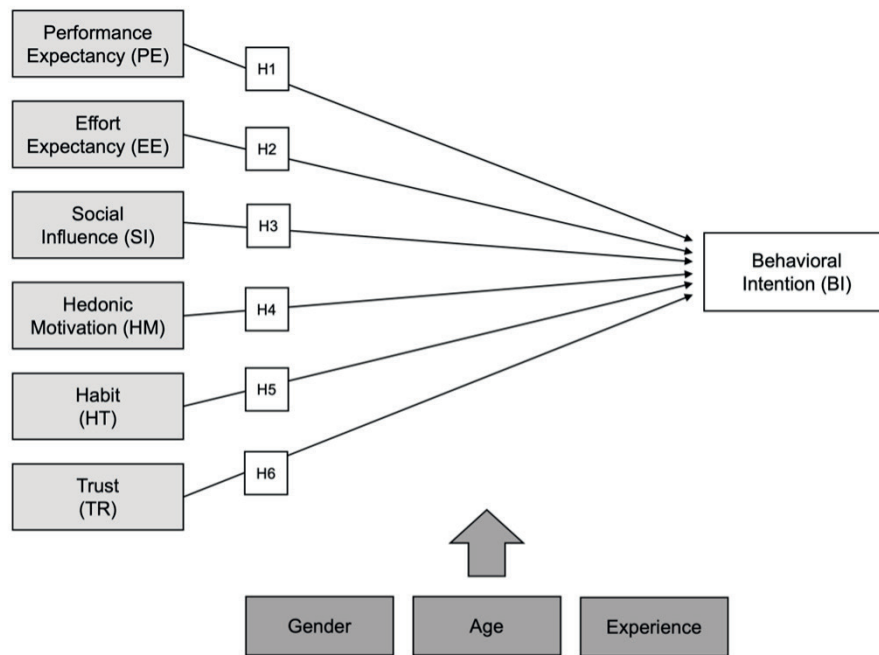


Figure 1: Research model.

Source: Own representation based on Venkatesh et al. (2003, p. 447) and Venkatesh, Thong and Xu (2012, p. 160).

Trust

Previous research has found that trust seems to affect potential users of AI technologies in addition to the UTAUT2 variables. Research by Mills and Liu (2020) draws on technology trust theory and explores the role of social presence, anthropomorphism, and privacy in determining people's trust and willingness to interact with digital humans. Ganesa, John, and Mane (2020) investigated the behavioral intention to use AI chatbots among telecom customers and extended the UTAUT2 model with the trust factor to quantify its effect on behavioral intention and user behavior, finding a positive relationship. This leads to the following hypothesis:

Hypothesis 6 – Trust (TR) positively influences the behavioral intention of using digital humans.

Control variables

In the original UTAUT2 model, in addition to the seven main determinants, moderating effects on **age**, **gender**, and **experience** were also taken into account. To avoid neglecting their influence in the present study, they were included as control variables. This procedure is in line with similar consumer research on technology acceptance. Since digital humans are a recent innovation, experience with chatbots is surveyed and included.

3. Research methodology

Venkatesh et al. (2003, p. 437) provided a quantitative survey methodology for the evaluation of the UTAUT model, which can be adapted for the purposes of this study. In contrast to a qualitative survey, a quantitative survey allows a high degree of standardization, allowing for greater comparability of the results. In addition, quantitative surveys allow the research model to be tested directly and without major preparation (Homburg 2017, p. 267). Compared to other data collection methods, such as telephone or written questionnaires, the online survey offers two main advantages: respondents can be recruited

quickly and easily, and a higher reach can be achieved through distribution on the internet. The latter advantage is often questioned, as it cannot be ensured that a specific target group will be reached. This problem of self-selection must therefore be considered when interpreting the results (Homburg 2017, pp. 269–70).

Data collection

Within the framework of an empirical cross-sectional study, an online survey was designed according to the chosen quantitative research approach and conducted in Germany from 30 June to 7 July 2022, employing convenience and snowball sampling. The online questionnaire was distributed via WhatsApp, email, and social media. The survey was conducted anonymously, and participants were requested to share the questionnaire with their social contacts.

Questionnaire rationale

Based on the literature review and the proposed research model, an online questionnaire was created, divided into four main sections:

1. Introduction and background information on AI-based digital humans
2. Experiences with chatbots and AI-based digital humans
3. Perceived acceptance of AI-based digital humans in online shops
4. Sociodemographic information of the survey participants

In the first section of the questionnaire, respondents received an explanation of AI-powered digital humans. In particular, a picture of Telekom's digital assistant Selena was shown, as well as a picture of the in-store digital assistant Kiri used at Vodafone stores in New Zealand, as shown in Figure 2. Participants were also provided with a link to a video showing a digital human in action at this point in the questionnaire.



Figure 2: Screenshot from the video of the digital human “Kiri” used by Vodafone New Zealand.
Source: <https://news.vodafone.co.nz/article/vodafone-reveals-identity-its-digital-assistant>

In the second part of the questionnaire, participants were asked about their prior experience with chatbots. The use of digital humans is currently not a widespread practice, so experience with chatbots was used as a substitute. In two further questions, the respondents were asked about their level of knowledge of digital humans.

The third and main part of the questionnaire addressed the perceived acceptance of digital humans in online stores and its influencing factors. The modified UTAUT2 model was used to capture six theoretical constructs (behavioral intention, performance expectancy, effort expectancy, social influence, hedonic motivation, habit, and trust) according to the definitions given in Section 2. Since these theoretical constructs are not directly observable variables, a reflective measurement model was applied. This involves the use of several directly measurable indicators to measure a theoretical construct (Kroeber-Riel and Weinberg 2003, p. 21). Discrete rating scales are usually employed to measure these indicators, and in particular, a Likert scale is often used to measure the attitudes of individuals (Homburg 2017, p. 314).

To operationalize the six constructs, a total of 22 indication- and application-specific items were adapted to the context of AI-based digital humans in online shops and were measured using a seven-point Likert scale ranging from “I strongly disagree” (1) to “I strongly agree” (7).

The final section of the survey collected sociodemographic information such as the age, gender, employment status, and educational status of participants. To ensure the validity of the questionnaire, it was tested in advance on test subjects to check the comprehensibility of the questions as well as the formal and technical correctness of the survey process.

Analysis strategy

The collected data were analyzed using the IBM software SPSS. Only complete data sets were considered in the analysis. The data include descriptive statistics on sociodemographic data, knowledge and experience values, and acceptance indicators of digital humans. The measurement instruments used were tested for internal consistency using reliability ratios and descriptive statistics of items. Linear and multiple regression analyses were used to test the relationships between the variables and the hypotheses derived from the proposed adapted UTAUT2 model. In the regression analysis, the model quality and the significance of regression coefficients were tested. A confidence level of 95% was used in all tests for statistical significance ($\alpha = 0.05$).

4. Analysis and results

In total, the online survey received 224 impressions, resulting in a final sample of 174 respondents with a dropout rate of 22.3 %. Table 1 shows the sociodemographic statistics of the sample with absolute and relative frequencies.

As shown in Table 1, the majority (56.9 %) of the 174 survey participants were female. The average age of the respondents was 30.2 years, with the majority (72.4 %) belonging to the young age group (15–29 years). The older age group (≥ 30 years) made up 27.6 % of the sample. The choice of age groups is based on the acceptance study by Monard et. al (2018, p. 16). The results show that the participants aged 20–30 had the most experience with chatbots and that the older age groups (> 30 years) were reluctant to use chatbots.

The division into the two age groups was intended to verify whether the reluctance of the older age group to use digital humans also applies at this early stage of the introduction of a technology. Regarding the level of education, the group with a university or university of applied sciences degree dominated (43.1 %). More than a third of the respondents (38.0 %) stated that they had graduated from a secondary school, whereas only 1.1 % did not have a degree. The remaining 17.8 % of the participants said that

they had completed an apprenticeship. Table 2 summarizes the results and provides the absolute and relative frequencies regarding the level of knowledge.

Table 2: Previous experience and knowledge of digital humans.

<i>Item</i>	<i>Category</i>	<i>Frequency</i>	<i>%</i>
Experience with chatbots	Yes	114	65.5 %
	No	60	34.5 %
The term “digital human”	Yes	77	44.3 %
	No	97	55.7 %
Usage of digital humans	Yes	43	24.7 %

Source: Own research, 2022, n = 174.

More than half of the respondents (65.5%) have already used a chatbot. Regarding prior knowledge about digital humans, it was found that less than half (44.3%) of the respondents knew the term digital human before describing it, while 55.7% of the respondents did not recognize it. In addition, 43 of the 174 respondents (24.7%) confirmed that they had already spoken to a digital human, while most respondents had never used digital humans.

Looking at these results in conjunction with the sociodemographic data, it can be seen that 32 of the 43 respondents who had already interacted with a digital human belonged to the younger age group (15–29 years). Of these 43 respondents, 22 were female and 21 were male. These results initially suggest that there are age-specific differences in the sample regarding previous use of digital humans, but no gender-specific differences. Thus, it could be concluded that age plays a role in the adaptation decision of potential users of digital humans in online stores.

Before testing the proposed research model and the hypotheses based on it, the collected constructs were evaluated in terms of their suitability for further statistical analysis based on the reliability and descriptive statistics of the scales. A scale is considered sufficiently reliable when the Cronbach’s alpha reliability coefficient reaches a value of at least 0.70. For further validation, the minimum corrected item-total correlation (r_{IS}) was recorded, which reflects the correlation of an item with the scale. This value is usually referred to as discriminatory power. According to Hair et al. (1998, p. 118), items should have a discriminatory power of at least 0.30 to be considered sufficiently reliable. The results of this analysis are shown in Table 3.

A preliminary test was conducted to assess the validity of the statistical procedures. The test for multicollinearity examined the correlations between the six latent variables. In the context of multiple regression, multicollinearity is an excessive correlation of two or more causal variables with each other. The correlation matrix is a suitable tool to test for the presence of multicollinearity. According to Field (2018, p. 402), correlation values above 0.8 between two independent variables are an indicator of multicollinearity. As shown in Table 4, all predictors correlated moderately to the behavioral intention, but none of the correlations between the predictors are above 0.8.

Table 3: Descriptive statistics and tests for reliability.

<i>Construct</i>	<i>Items</i>	α_c / SB^a	r_{IS}	<i>M</i>	<i>SD</i>
Behavioral intention	3	0.92	0.82 – 0.86	3.69	1.71
Performance expectancy	4	0.91	0.75 – 0.83	4.33	1.55
Effort expectancy	4	0.91	0.78 – 0.85	4.94	1.39
Social influence	3	0.96	0.90 – 0.92	3.20	1.53
Hedonic motivation	3	0.92	0.79 – 0.86	4.25	1.59
Habit	3	0.93	0.82 – 0.90	4.25	1.63
Trust	2	0.67 ^a	0.50	4.10	1.72

* α_c = Cronbach’s alpha; SB = Spearman–Brown coefficient; r_{IS} = minimum of corrected item-total correlation; M = mean; SD = standard deviation

Source: Own research, 2022, n = 174.

Table 4: Correlation matrix.

	BI	PE	EE	SI	HM	HT	TR
BI	1.00	0.779**	0.528**	0.570**	0.620**	0.818**	0.678**
PE		1.00	0.520**	0.516**	0.662**	0.760**	0.719**
EE			1.00	0.336**	0.467**	0.503**	0.501**
SI				1.00	0.439**	0.580**	0.478**
HM					1.00	0.701**	0.624**
HT						1.00	0.681**
TR							1.00

BI = behavioral intention, PE = performance expectancy, EE = effort expectancy, SI = social influence, HM = hedonic motivation, HT = habit, TR = trust; [** p < 0.01]

Source: Own research, 2022, n = 174.

Hypothesis testing

As illustrated in Table 5, in the simple regression analyses, all of the UTAUT2 variables **performance expectancy** ($\beta = 0.78$), **effort expectancy** ($\beta = 0.53$), **social influence** ($\beta = 0.57$), **hedonic motivation** ($\beta = 0.62$), and **habit** ($\beta = 0.82$), as well as the additional variable **trust** ($\beta = 0.68$) are significant determinants of the behavioral intention of AI-based digital humans ($p \leq 0.001$). As indicated in the correlation matrix (Table 4), the linear regressions also reflect highly significant influences of the variables **performance expectancy** and **habit** on behavioral intention. Based on these results, all six hypotheses can be confirmed.

Table 5: Simple linear regressions.

<i>Predictors</i> ^a	<i>R</i> ²	<i>b</i>	<i>SE</i>	β	<i>p</i>
Performance expectancy	0.61	0.92	0.06	0.78	0.000***
Effort expectancy	0.28	0.67	0.08	0.53	0.000***
Social influence	0.33	0.61	0.07	0.57	0.000***
Hedonic motivation	0.38	0.66	0.06	0.62	0.000***
Habit	0.67	0.86	0.05	0.82	0.000***
Trust	0.46	0.76	0.06	0.68	0.000***

^a criterion = behavioral intention; R^2 = coefficient of determination;
b = unstandardized coefficients; *SE* = coefficients std. error; β = standardized coefficients; *p* = statistical significance [* $p < 0.05$ ** $p \leq 0.01$ *** $p < 0.001$]

Source: Own research, 2022, $n = 174$.

Furthermore, due to the limited evidence base for the proposed relationship in the modified UTAUT2 model, a three-step hierarchical multiple regression was performed. Table 6 illustrates the three steps of the hierarchical multiple regression, indicating the model quality as well as the changes in the coefficient of determination (ΔR^2) and the degrees of freedom (ΔF) when including additional variables. Furthermore, the significance of the change (*p*) was calculated to determine whether the additional variance (ΔR^2) could contribute to a significant improvement in the model. Model 1 showed that the control variables of gender, age, and experience explained only 7.7 % of the variance in the behavioral intention of using AI-powered digital humans and thus did not contribute significantly to the variance explanation of the criterion behavioral intention ($\Delta F_{3,170} = 5.81$; $p = 0.001$).

Adding the UTAUT2 predictors improved the variance explanation of behavioral intention in model 2 to 74.1 % ($\Delta F_{5,165} = 88.11$; $p = 0.000$). Thus, the predictors of performance expectancy, effort expectancy, social intention, hedonic motivation, and habit explained 66.4 % of additional variance in the criterion relative to the control variables. Including the additional variable trust in model 3 resulted in a very slightly increased additional variance explanation of the behavioral intention of only 0.3 %.

Thus, the additional predictor trust could not significantly explain more variance of the criterion than the control variables and the UTAUT2 variables ($\Delta F_{1,1164} = 2.25$; $p = 0.135$).

Table 6: Model quality of multiple linear regression.

<i>Model</i>	<i>Predictors</i> ^a	<i>adj. R²</i>	ΔR^2	ΔF (<i>df1</i> , <i>df2</i>)	<i>p</i>
M1 ^b	Control variables	0.077		5.81 (3,170)	0.001***
M2 ^c	Control and UTAUT2 variables	0.741	0.660	88.11 (5,165)	0.000***
M3 ^d	Control, UTAUT2 and additional variables	0.743	0.003	2.25 (1,164)	0.135

^a criterion = behavioral intention; ^b Step 1: predictors (gender, age, experience); ^c Step 2: predictors (performance expectancy, effort expectancy, social influence, hedonic motivation, habit); ^d Step 3: predictor (trust); *adj. R²* = adjusted corrected coefficient of determination; ΔR^2 = changes in the coefficient of determination; ΔF = changes in the degrees of freedom ; *p* = significance of the change [*** *p* = 0.001]

Source: Own research, 2022, n = 174.

Table 7: Multiple linear regression.

<i>Predictors</i> ^a	<i>b</i>	<i>SE</i>	β	<i>p</i>	<i>T</i>	<i>VIF</i>
Gender	0.51	0.13	0.02	0.687	0.94	1.07
Age ^b	0.01	0.01	0.08	0.049	0.83	1.20
Experience ^c	0.31	0.14	0.09	0.027*	0.85	1.17
PE	0.34	0.08	0.29	0.000***	0.32	3.12
EE	0.10	0.06	0.08	0.114	0.61	1.64
SI	0.08	0.05	0.08	0.133	0.61	1.65
HM	-0.02	0.06	-0.02	0.722	0.43	2.34
HT	0.49	0.07	0.47	0.000***	0.30	3.35
TR	0.10	0.07	0.09	0.135	0.40	2.50

^a criterion = behavioral intention (b= -1,53); ^b reference category = female; ^c reference category = no experience; *b* = unstandardized coefficients; *SE* = coefficients std. error; β = standardized coefficients; *p* = significance [* *p* < 0.05 *** *p* < 0.001]; *T* = tolerance; *VIF* = variance inflation factor

Source: Own research, 2022, n = 174.

The final multiple regression model (model 3) thus explained a total of 74.3 % of the total variance of the criterion behavioral intention ($F_{9,499} = 56.52$; $p = 0.000$). Table 7 summarizes the results of the three-stage multiple regression analysis, reporting only the outcome measures from the last regression model (model 3) with the highest variance explanation ($R^2 = 74.3\%$).

In the hierarchical multiple regression, among the primary UTAUT2 predictors, performance expectancy (step 2, $\beta = 0.29$; $p = 0.000$) and habit (step 2, $\beta = 0.47$; $p = 0.000$) were also found to be significant independent determinants of behavioral intention. This implies that hypotheses 1 and 5 can be confirmed even when controlling for the other predictors. Contrary to hypotheses 2, 3, 4, and 6, the other predictors in the multiple regression model had no significant effect on behavioral intention ($p > 0.05$). That is, controlling for all other predictors included in the regression model, no significant relationship between the predictor's effort expectancy (step 2, $\beta = 0.08$; $p = 0.114$), social influence (step 2, $\beta = 0.08$; $p = 0.133$), hedonic motivation (step 2, $\beta = -0.02$; $p = 0.722$), or trust (step 3, $\beta = 0.09$; $p = 0.135$) and the criterion behavioral intention could be confirmed.

Regarding the control variables, significant results were found for the control variable experience (step 1, $b = 0.31$; $p = 0.027$). Accordingly, respondents with no experience with chatbots (reference category) seemed to have a higher behavioral intention than participants with experience. For the control variable age, a barely significant value (step 1, $b = 0.01$; $p = 0.049$) was observed. However, since this value is extremely close to the significance level $\alpha = 0.05$, the result could not be classified as statistically significant. Finally, the control variable gender did not have a significant influence on behavioral intention.

5. Discussion

The research question was motivated by the need to understand what factors influence the acceptance of AI-based digital humans in online stores. In this study, an extended UTAUT2 model was tested with respect to the acceptance of digital humans in online stores.

Overall, the acceptance of AI digital humans in the observed population was moderate (mean BI = 3.69). In six simple linear regression analyses, the UTAUT2 variables (performance expectancy, effort expectancy, social influence, hedonic motivation, habit) and the additional variable (trust) were found to be effective predictors of behavioral intention toward digital humans in online stores. Based on this, all hypotheses (1–6) could be confirmed, implying that the constructs are suitable for predicting the acceptance of digital humans.

The hierarchical multiple regression produced significantly different results. In the final multiple regression model (model 3), which accounted for the influences of the six independent latent variables as well as the three control variables (gender, age, and experience) on behavioral intention, statistically significant influences were found for the performance expectancy, habit, and experience variables. None of the other constructs (effort expectancy, social influence, hedonic motivation, and trust) had a statistically significant effect on behavioral intention.

In contrast to the results of the simple linear regressions, only hypotheses 1 and 5 were confirmed in the hierarchical multiple regression analysis. The results of this study show a statistically significant correlation between performance expectancy and behavioral intention. Performance expectancy ($\beta = 0.29$; $p = 0.000$) proved to be the strongest positive predictor of behavioral intention to accept digital humans, and the level of performance expectancy of the sample was moderate (mean PE = 4.33) on average. Although the participants did not think that digital humans could increase their productivity in online stores, they still showed general agreement with the use of digital humans in online stores. Based on the obtained results, hypothesis 1 is accepted.

Moreover, habit ($\beta = 0.47$; $p = 0.000$) was confirmed as a significant positive predictor of acceptance in addition to performance expectancy. These results are consistent with the findings of Ganesa, John and Mane (2020) regarding the acceptance of AI chatbots by telecommunications customers. Habit was also perceived to be in the medium range in this population (mean HT = 4.25), with endorsement of the use of digital humans in online stores receiving the highest level of agreement. According to the results, hypothesis 3 is accepted due to the significant and positive relationship between habit and behavioral intention.

In contrast, the relevance of the other three primary UTAUT2 factors of effort expectancy, social influence, and hedonic motivation appeared to be secondary when all variables were considered together. Overall, the participants largely believed that they possessed the skills (mean EE = 4.94) to use innovative technologies such as digital humans in online stores. Effort expectancy ($\beta = 0.08$; $p = 0.114$) showed a small positive beta coefficient, but the required significance threshold was not reached, so hypothesis 2 could not be confirmed.

The predictor of social influence was rather low in this population (mean SI = 3.20). This suggests that the customization decision of digital humans in online stores is not strongly influenced by the opinions and attitudes of people close to the participants. Furthermore, as with effort expectancy, there was a minimal positive beta coefficient for social influence ($\beta = 0.08$; $p = 0.133$). Due to this lack of significance, hypothesis 3 was also not confirmed.

The predictor of hedonic motivation was in the middle range (mean HM = 4.25). The population's opinion that digital humans in online stores can be fun and entertaining was the strongest. During the study, it was found that hedonic motivation showed a positive influence on behavioral intention in the simple linear regression, while the multiple linear regression showed a minimal negative correlation. However, hedonic motivation did not significantly influence the behavioral intention to use digital humans in online stores, so hypothesis 4 could not be confirmed. Furthermore, when considering all predictors at once, no direct significant influence of the additional variable trust ($\beta = 0.09$; $p = 0.135$) could be found. Trust was also in the medium range for this population ($M_{TR} = 4.10$). A differentiated picture emerged. Although the respondents were convinced that digital humans will offer the best deals, they did not trust in the technology behind AI-based digital humans.

The multiple regression included the control variables and yielded the following results. Within the sample, a higher acceptance of digital humans was found among participants without experience ($b = 0.42$; $p = 0.031$). Since this innovative technology is still sporadic in the European market, experience among participants with chatbots was used instead of experience with digital humans. Furthermore, although a significant value for age was found ($b = 0.01$; $p = 0.049$), this value was too close to the significance level of $\alpha = 0.05$. For this reason, age was not considered statistically significant. Thus, no influence of age on digital human acceptance was found. In addition, no significant influence of gender was found as a predictor of acceptance. Therefore, future research should examine the moderation effects of these variables as suggested by Venkatesh et al. (2003, pp. 467–469), which would have required a larger sample than the 174 subjects in the present study.

Overall, for the proposed research model of the study, 74.3% of the variance in behavioral intention could be explained by variation in the independent variables, namely performance expectancy, effort expectancy, social influence, hedonic motivation, habit, and trust. Since the majority of the hypotheses could not be confirmed within the fitted and modified UTAUT2 model, this is an indication that there are other important determinants that influence the adoption of digital humans in online stores but were not considered in the tested model.

6. Conclusion

The primary objective of this study was to investigate consumer acceptance of AI digital humans using the UTAUT2 model to identify the factors influencing acceptance. Since there are no findings to date on factors in the acceptance of digital humans, this study can be seen as having an exploratory character in addition to testing the hypotheses that have been formulated. As the results of the research show, there is a tendency to accept the use of digital humans in online stores. Performance expectancy and habit were found to be relevant and statistically significant determinants of the behavioral intention of digital humans in online shops.

The results of this study can provide important insights into consumer acceptance of innovative digital humans. Many brands are already preparing for conversational commerce, which will bring such fundamental change that retailers need to start familiarizing themselves with the innovative technologies involved.

Limitations

Although this study has made significant findings regarding consumer acceptance of digital humans, there are a few limitations to their generalizability. The study relies on a limited number of participants. Because of a lack in access to a sampling frame, this study had to rely on a non-probability sample. Due to the random sampling method used, the results cannot be generalized to the entire population, but they reflect the reality of emerging technology markets with a relatively high technology sensitivity within a younger population. It is to be expected that older age groups do not currently use digital humans in online stores to the same extent as younger age groups. A larger sample could provide better insights into the impact of the control variable age. In future research, the results of this study should be plausibilized by methods other than UTAUT2. This should primarily focus on uncovering further determinants that may influence the acceptance of digital humans in online stores. Future research could support an ordinal regression approach and compare the results to those linear models in this study. In addition, the use of digital humans is currently not widespread. Therefore, most participants had no personal experience with digital humans and had to rely on descriptions.

Conflicts of interest statement

The authors declare that there is no conflict of interests regarding the publication of this paper.

References and notes

- [1] AI Forum of New Zealand. 2019. Artificial Intelligence in Action: Digital Humans Case Study. <https://aiforum.org.nz/knowledgehub/artificial-intelligence-in-action-digital-humans-case-study>, accessed September 2022.
- [2] Denner, E. 2021. Are you listening? Using voice as an immersive customer experience. <https://www.bearingpoint.com/en/our-success/insights/using-voice-as-an-immersive-customer-experience>, accessed September 2022.
- [3] Fishbein, M., and Ajzen, I. 1975. Belief, attitude, intention and behavior: An introduction to theory and research. In: Massachusetts: Addison-Wesley.
- [4] Futurside 2022. We will Live In a World of Digital Humans. <https://futurside.com/we-will-live-in-a-world-of-digital-humans>, accessed August 2022.
- [5] Ganesa, D. P., John, S., and Mane, D. A. S. 2020. Behavioural Intention of AI-Chatbots by Telecom Customers – UTAUT2 Perspective with Trust. Paper presented at the International Conference on Marketing, Technology and Society.
- [6] Gursoy, D., Chi, O. H., Lu, L., and Nunkoo, R. 2019. Consumers acceptance of artificially intelligent (AI) device use in service delivery. *International Journal of Information Management*, 49, 157–169.
- [7] Homburg, C. 2017. *Marketingmanagement: Strategie-Instrumente-Umsetzung-Unternehmensführung* (Vol. 6): Springer-Verlag.
- [8] Kim, S. S., Malhotra, N. K., and Narasimhan, S. 2005. Research note—two competing perspectives on automatic use: A theoretical and empirical comparison. *Information systems research*, 16(4), 418–432.
- [9] Kroeber-Riel, W., and Weinberg, P. 2003. *Konsumentenverhalten* (Vol. 8). München.
- [10] Lim, W. M. 2021. History, lessons, and ways forward from the COVID-19 pandemic. Paper presented at the International Journal of Quality Innovation.
- [11] Limayem, M., Hirt, S., and Cheung, C. 2007. How habit limits the predictive power of intention: The case of information systems continuance. *MIS quarterly*, 31(4), 705–737.
- [12] Lu, L., Cai, R., and Gursoy, D. 2019. Developing and validating a service robot integration willingness scale. *International Journal of Hospitality Management*, 80, 36–51.
- [13] Mills, A. M., and Liu, L. F. 2020. Trust in Digital Humans: ACIS 2020 Proceedings.
- [14] Monard, F., Uebersax, H.-P., Mousser, J., Furchheim, P., Müller, S., and Hannich, F. 2020. Chatbot-Studie. Die digitalen Helfer im Praxistest. https://cdn2.hubspot.net/hubfs/5893787/PIDAS_ZHAW_Chatbot_Studie_FINAL.pdf, accessed September 2022.
- [15] NTT DATA Business Solutions AG. 2022. Wie Conversational AI Mitarbeitende und Kunden glücklich macht. <https://nttdata-solutions.com/de/innovationen/conversational-ai/>, accessed August 2022.
- [16] Schwendener, S. 2018. *Technologie-Akzeptanz von Chatbots: eine Anwendung des UTAUT-Modells*. School of Management and Law, Winterthur, Zürich.
- [17] UneeQ. 2020. AI chatbots are a good start, but now's the time to stand out from the crowd. <https://digitalhumans.com/blog/what-are-ai-chatbots/>, accessed May 2022.
- [18] Corti, K.; Gillespie, A. (2016): Co-constructing intersubjectivity with artificial conversational agents: People are more likely to initiate repairs of misunderstandings with agents represented as human. In: *Computers in Human Behavior*, vol. 58, pp. 431–442.
- [19] UneeQ. 2021. What are Digital Humans? A giant leap in brand & customer experience. <https://cdn2.hubspot.net/hubfs/4954505/What%20are%20digital%20humans%20UneeQ%20eBook%20v2.pdf>, accessed May 2022.

- [20] Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. 2003. User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478.
- [21] Venkatesh, V., Thong, J. Y. L., and Xu, X. 2012. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. *MIS Quarterly*, 36(1), 157–178.
- [22] Ward, S. 2020. Corona lockdown? Time to embrace a digital human - AI technology elevating the human experience. <https://www2.deloitte.com/be/en/pages/technology-media-and-telecommunications/articles/corona-lockdown-time-to-embrace-a-digital-human.html>, accessed May 2022.
- [23] Ward, S., Boom, M., and Majenburg, M. 2022. Digital Human - Elevating the Digital Human experience. <https://www2.deloitte.com/nl/nl/pages/customer-and-marketing/articles/digital-human.html>, accessed May 2022.

Appendix I: Survey

Welcome,

Within the framework of my Master's thesis at the Neu-Ulm University of Applied Sciences, I am investigating the acceptance of digital humans in the sales processes of online shops.

With only about 8 minutes of your time, you can make a valuable contribution to my investigation and ensure valid results.

First, Digital Humans are described in general and illustrated with a short video. Subsequently, I ask you to answer the questions based on your personal opinion.

The survey is anonymous and your data will be treated confidentially.

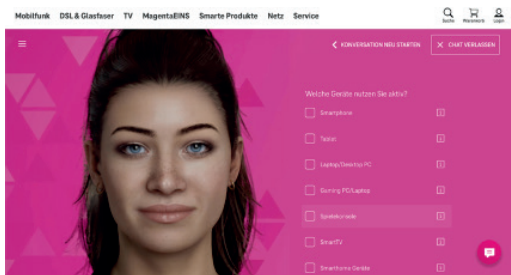
Thank you for your support!

Digital Humans are a combination of Artificial Intelligence (AI) and human conversation. They can be understood as a continuation of chatbots, with a human identity, appearance, and emotions. Behind Digital Humans is an AI platform that determines behavior, expressions, and language in real-time. This allows natural conversations to occur as they would in real life. Through verbal and non-verbal communication, they can realistically replicate natural human interaction on a large scale.

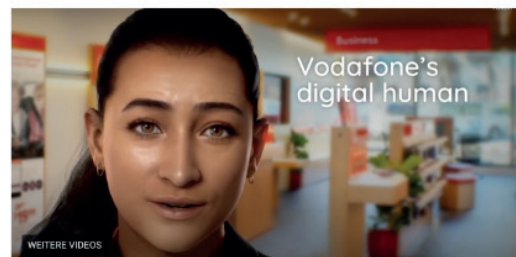
Today, they exist both in brick-and-mortar retail and online, where they can virtually advise a large number of customers 24/7 simultaneously. Digital Humans are already working for some of the biggest brands in the world such as Vodafone, Telekom, and BMW.

Examples from practice:

The digital assistant Selena helps you to find out in just a few steps what the necessary bandwidth is for your needs. This enables it to find the best individual internet tariff for you.



Digital assistant Kiki lives in Vodafone stores and helps customers manage their tariffs. She welcomes you with a smile and guides you through the entire transaction..



Click [here](#) for more information and to see the Digital Human Sophie in action in a short video.

1. Have you ever used a chatbot on a website (e.g. an online shop)?
 Yes No
2. Did you know what Digital Humans were before you read the description?
 Yes No
3. Have you ever had a conversation with a digital human?
 Yes No
4. Which personal device do you prefer for interacting with Digital Humans?
 Laptop
 Smartphone
 Smartwatch
 Tablet
5. What is your attitude towards different forms of interaction with digital humans?

The ability to talk and communicate with a digital human is important to me.

Do not agree at all (1)	(2)	(3)	(4)	(5)	(6)	Completely agree (7)
----------------------------	-----	-----	-----	-----	-----	-------------------------

The ability to type commands into a keyboard to interact with a digital human is important to me.

Do not agree at all (1)	(2)	(3)	(4)	(5)	(6)	Completely agree (7)
----------------------------	-----	-----	-----	-----	-----	-------------------------

The ability to interact with a digital human through gestures is important to me.

Do not agree at all (1)	(2)	(3)	(4)	(5)	(6)	Completely agree (7)
----------------------------	-----	-----	-----	-----	-----	-------------------------

6. Where would you most like to have a digital human in online shops to support you?

<input type="checkbox"/> More likely to welcome you in online shops	<input type="checkbox"/> More likely when offering discounts
<input type="checkbox"/> More likely when searching for a specific product	<input type="checkbox"/> More likely to list nearby stores
<input type="checkbox"/> More likely when making recommendations	<input type="checkbox"/> More likely to track your delivery
<input type="checkbox"/> More likely when advising on additional products	<input type="checkbox"/> More likely in customer support
<input type="checkbox"/> More likely when advising on higher value products	<input type="checkbox"/> More likely to handle returns

7. Please indicate how strongly you agree with the following statements:

BI1. I intend to use digital humans in online shops in the future.

Do not agree at all (1)	(2)	(3)	(4)	(5)	(6)	Completely agree (7)
----------------------------	-----	-----	-----	-----	-----	-------------------------

BI2. It is very likely that I will use digital humans in online shops, in my daily life.

Do not agree at all (1)	(2)	(3)	(4)	(5)	(6)	Completely agree (7)
----------------------------	-----	-----	-----	-----	-----	-------------------------

BI3. I plan to use digital humans in online shops frequently.

Do not agree at all (1)	(2)	(3)	(4)	(5)	(6)	Completely agree (7)
----------------------------	-----	-----	-----	-----	-----	-------------------------

8. Please indicate how strongly you agree with the following statements:

PE1. Digital humans in online shops are useful.

Do not agree at all (1)	(2)	(3)	(4)	(5)	(6)	Completely agree (7)
----------------------------	-----	-----	-----	-----	-----	-------------------------

PE2. Using digital humans in online shops increases my chances of achieving things that are important to me.

Do not agree at all (1)	(2)	(3)	(4)	(5)	(6)	Completely agree (7)
----------------------------	-----	-----	-----	-----	-----	-------------------------

PE3. Using digital humans in online shops helps me accomplish things more quickly.

Do not agree at all (1)	(2)	(3)	(4)	(5)	(6)	Completely agree (7)
----------------------------	-----	-----	-----	-----	-----	-------------------------

PE4. Using digital humans increases my productivity.

Do not agree at all (1)	(2)	(3)	(4)	(5)	(6)	Completely agree (7)
----------------------------	-----	-----	-----	-----	-----	-------------------------

9. Please indicate how strongly you agree with the following statements:

EE1. Learning how to use digital humans in online shops is easy for me.

Do not agree at all (1)	(2)	(3)	(4)	(5)	(6)	Completely agree (7)
----------------------------	-----	-----	-----	-----	-----	-------------------------

EE2. The use of digital humans in online shops is clear and understandable.

Do not agree at all (1)	(2)	(3)	(4)	(5)	(6)	Completely agree (7)
----------------------------	-----	-----	-----	-----	-----	-------------------------

EE3. I consider digital humans in online shops quite easy to me.

Do not agree at all (1)	(2)	(3)	(4)	(5)	(6)	Completely agree (7)
----------------------------	-----	-----	-----	-----	-----	-------------------------

EE4. It is easy for me to become skillful at using digital humans in online shops.

Do not agree at all (1)	(2)	(3)	(4)	(5)	(6)	Completely agree (7)
----------------------------	-----	-----	-----	-----	-----	-------------------------

10. Please indicate how strongly you agree with the following statements:

SI1. People who are important to me think I should use digital humans in online shops.

Do not agree at all (1)	(2)	(3)	(4)	(5)	(6)	Completely agree (7)
----------------------------	-----	-----	-----	-----	-----	-------------------------

SI2. People who influence my behavior think that I should use digital humans in online shops.

Do not agree at all (1)	(2)	(3)	(4)	(5)	(6)	Completely agree (7)
----------------------------	-----	-----	-----	-----	-----	-------------------------

SI3. People whose opinions that I value prefer that I use digital humans in online shops.

Do not agree at all (1)	(2)	(3)	(4)	(5)	(6)	Completely agree (7)
----------------------------	-----	-----	-----	-----	-----	-------------------------

11. Please indicate how strongly you agree with the following statements:

HM1. I think using digital humans in online shops is fun.

Do not agree at all (1)	(2)	(3)	(4)	(5)	(6)	Completely agree (7)
----------------------------	-----	-----	-----	-----	-----	-------------------------

HM2. I think using digital humans in online shops is enjoyable.

Do not agree at all (1)	(2)	(3)	(4)	(5)	(6)	Completely agree (7)
----------------------------	-----	-----	-----	-----	-----	-------------------------

HM3. I think using digital humans in online shops is entertaining.

Do not agree at all (1)	(2)	(3)	(4)	(5)	(6)	Completely agree (7)
----------------------------	-----	-----	-----	-----	-----	-------------------------

12. Please indicate how strongly you agree with the following statements:

HT1. The use of digital humans in online shops could become a habit for me.

Do not agree at all (1)	(2)	(3)	(4)	(5)	(6)	Completely agree (7)
----------------------------	-----	-----	-----	-----	-----	-------------------------

HT2. I am in favor to use digital humans in online shops.

Do not agree at all (1)	(2)	(3)	(4)	(5)	(6)	Completely agree (7)
----------------------------	-----	-----	-----	-----	-----	-------------------------

HT3. Using digital human in online shops could become natural to me.

Do not agree at all (1)	(2)	(3)	(4)	(5)	(6)	Completely agree (7)
----------------------------	-----	-----	-----	-----	-----	-------------------------

13. Please indicate how strongly you agree with the following statements:

TR1. I am convinced that digital humans in online shops are used to provide customers with the best offerings.

Do not agree at all (1)	(2)	(3)	(4)	(5)	(6)	Completely agree (7)
----------------------------	-----	-----	-----	-----	-----	-------------------------

TR2. I trust in digital humans.

Do not agree at all (1)	(2)	(3)	(4)	(5)	(6)	Completely agree (7)
----------------------------	-----	-----	-----	-----	-----	-------------------------

14. What is your gender?

Female

Male

Divers

Please enter your age.

What is your highest level of education?

No degree

Secondary level 1 school

Secondary school degree

University entrance qualification/ qualification for entrance to Universities of Applied Sciences

Apprenticeship

University/ University of Applied Sciences degree

Please select your current employment status.

Currently not working

Pupil

Trainee

Student

Full-time employee

Part-time employee

Civil servant

Self-employed, freelancer, farmer

Pensioner

Appendix II: Table A1 Measurement instruments

<i>Construction</i>	<i>Definition</i>	<i>Measurement Instruments</i>
Behavioral intention (BI)	The degree to which an individual intends to use digital humans in online shops.	BI1. I intend to use digital humans in online shops in the future. BI2. It is very likely that I will use digital humans in online shops, in my daily life. BI3. I plan to use digital humans in online shops frequently.
Performance expectancy (PE)	The degree to which using digital humans in online shops will provide benefits to consumers in performing certain activities.	PE1. Digital humans in online shops are useful. PE2. Using digital humans in online shops increases my chances of achieving things that are important to me. PE3. Using digital humans in online shops helps me accomplish things more quickly. PE4. Using digital humans increases my productivity
Effort expectancy (EE)	The degree of ease/effort associated with consumers' use of digital humans in online shops.	EE1. Learning how to use digital humans in online shops is easy for me. EE2. The use of digital humans in online shops is clear and understandable. EE3. I consider digital humans in online shops quite easy to me. EE4. It is easy for me to become skillful at using digital humans in online shops.
Social influence (SI)	The degree to which an individual perceives that important others believe he or she should use digital humans in online shops.	SI1. People who are important to me think I should use digital humans in online shops. SI2. People who influence my behavior think that I should use digital humans in online shops. SI3. People whose opinions that I value prefer that I use digital humans in online shops.
Hedonic motivation (HM)	The pleasure or enjoyment derived from using digital humans in online shops.	HM1. I think using digital humans in online shops is fun. HM2. I think using digital humans in online shops is enjoyable. HM3. I think using digital humans in online shops is entertaining.
Habit (HT)	The extent to which people tend to perform behaviors automatically because of learning.	HT1. The use of digital humans in online shops could become a habit for me. HT2. I am in favor to use digital humans in online shops. HT3. Using digital human in online shops could become natural to me.
Trust (TR)	The degree to which people believe that digital humans in online shops works for their best interest.	TR1. I am convinced that digital humans in online shops are used to provide customers with the best offerings. TR2. I trust in digital humans.

Source: Adapted from Ha et al. (2019), Venkatesh et al. (2003) and Venkatesh, Thong & Xu (2012).

TOPSIS-Based Recommender System for Big Data Visualizations

Václav Stehlík*, Mouzhi Ge**

ABSTRACT

Big data analytics can enable effective data interpretation that leads to highquality decision making in organizations. Since the amount of data nowadays has significantly increased in various industries, it thus motivates the research of visualizations for interdisciplinary and collaborative domains. Although there is a large collection of visualization techniques, it is very time-consuming to choose proper visualization techniques for a specific dataset.

This paper therefore aims to analyze the typical and state-of-the-art data visualization techniques for big data. Differing from traditional visualizations such as line chart or bar chart, this paper focuses on reviewing a set of modern visualizations for big data in terms of analyzing their advantages and disadvantages. To facilitate choosing a proper visualization, a recommender model based on TOPSIS is further proposed. In order to validate the proposed model, a prototype of the big data visualization recommender system has been implemented to validate the applicability of the system.

KEYWORDS

Data visualization, big data, data visualization recommendation, TOPSIS

1. Introduction

Big data has received increasing attention in recent years, as organizations and cities are dealing with tremendous amounts of data with high complexity and velocity [1]. These data are fast moving and changing in value, meaning and format. They also can originate from various sources, such as social networks, unstructured data from different devices or raw feeds from sensors. Although there is no unified definition for big data [2], big data can be broadly described as data of substantial size

* Masaryk University, Botanická 68A, 60200 Brno, Czech Republic

** Deggendorf Institute of Technology, Dieter-Görlitz-Platz 1, 94469 Deggendorf
<https://doi.org/10.25929/jair.v1i1.114>

possibly originating from different sources in distinct data format presentations while requiring high data processing speed [3]. One of the prominent approaches of describing big data is the V-model which presents a set of terms starting on letter 'V' that represents various aspects of big data. Initially, three main Vs were introduced by the e-commerce sector by David Laney [4]. Those three Vs were consisting of volume, variety and velocity. Volume of big data refers to the size of data being created from all the sources such as medical data, space images, traffic sensors, satellite imagery, banking transactions, financial market data, etc. [5]. Variety describes heterogeneity of data. Heterogeneity may refer to differences in terms of data types, such as text and images, and data format, or different data sources that generate data such as open data and sensor data. Velocity refers to how quickly data are generated and moved in data storage.

More organizations are focused on interdisciplinary data analytics and intend to use visualization tools to gain insights and leverage it to improve information sharing, increase return of investment, reduce burden of IT, and enable more efficient data driven decision-making [6]. Nowadays, there is a plethora of tools such as Tableau or FusionCharts that offer comprehensive sets of visualizations and help companies to visualize data. However, it requires extensive knowledge and skill sets to select a proper visualization for insightful analysis [7]. Deceptive and inappropriate usages of visualization may hurt the business development, some cases have been well studied such as [8] and [9]. Therefore, if the visualization of data is not selected properly, the previous steps in the big data life cycle such as collecting data, cleansing data, performing analysis, building and validation of models may become invalid.

To help organizations understand and select the proper visualization, the goal of this paper is to first provide a comparative study for various visualizations and then present an approach to recommend the most suitable visualization for specific needs and constraints. Reviewing the visualization is focused on detailed description of advantages, disadvantages, tips and connection to big data characteristics of each visualization. Based on the review of visualization techniques, a set of criteria that may influence the choice of the visualization is selected. Afterwards these criteria are used to build a tool that can help users select a visualization and learn more about how to properly use and take advantage of it.

The contribution of this paper is three-fold. First, the state-of-the-art visualizations are reviewed based on their advantages and disadvantages. This may help organizations, especially non-IT organizations to understand which visualizations can be used. Second, a recommendation model based TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) is proposed to facilitate the visualization selection. This will lead to the efficient selection for interdisciplinary data analytics. Third, a prototype system is implemented with open sources.

The rest of the paper is organized as follows. Section 2 reviews a variety of typical visualizations for big data in terms of summarizing advantages and disadvantages. Based on the reviews, section 3 proposes a TOPSIS-based model that helps users to select the proper visualization according to a specific context. Further, Section 4 implements system prototype to validate the applicability of the proposed model. Finally, Section 5 concludes the paper and outlines limitations as well as future research.

2. Typical Visualizations for Big Data

Data visualization is a representation of information derived from data in a pleasant graphical form that facilitates aspects of communication and insight extraction. This term can be further differentiated into information, scientific and infographic visualization [10]. Scientific visualizations usually have to comply with conventions and expected rules to convey spatial information and closely present real-world physically based objects and models [11]. Information visualization is typically used to represent abstract non-physically based data [11]. Also, infographics usually combine elements of statistical information with data visualization in order to set a narrative or polemic [10].

Visualization can serve as an easily approachable medium of valuable information that allows the user to observe trends, patterns, outliers, relationships, and performance. The most proficient visualizations leverage highlighting, filtering and interactive aspects to remove noise and help the user to focus on the message and important parts [12]. In this section, we will review 10 typical visualization techniques as well as their advantages and disadvantages for big data.

2.1. Treemap

Treemap visualization is used to display large amounts of hierarchical data. It utilizes a root rectangle that is subdivided into smaller rectangles that correspond to the hierarchical structure of the dataset as it can be seen in figure 1. Treemaps are well-known and commonly used on data dashboards, mainly because of their ability to display relationships and hierarchies in data. However, a complex visualization method, Treemap can be hard for beginners to comprehend [13]. Thus, one needs to evaluate the efforts of using treemap visualisation.

Advantages

Treemaps are based on shape volume estimation which is calculated from one or more data factors. Therefore every change in data requires total repaint of the whole visualization for the visible level of hierarchy. Levels that are higher up the hierarchy do not require the action of repainting because they are not visible in the current view [3]. The method allows usage of two measures (data factors) where the first one is size and the second one is color. Size is used for shape volume calculation and color helps with grouping of the shapes. The size measure used for volume calculation must be of computable data type [3]. All of this decreases utilization in regards to data variety.



Figure 1: Treemap depicting stock market from finviz.com [14].

Disadvantages

Velocity is poorly handled by treemaps [3]. Due to its nature, only a snapshot of one moment is displayed and the time component is not present at all. Overview of the development over the course of time would require comparison with a completely new treemap. However, such practises would be cumbersome and the process of making an insight would be less than optimal.

2.2. Sunburst

Sunburst chart, also known as ring chart, multi-level pie chart, belt chart or radial treemap [15]. It is often used to visualize hierarchical data. The origins of this radial space-filling visualization method go back to year 2000 and the research of John Stasko [16]. The radial layout is comprised of a central origin around which rings are stacked in an outward direction [15]. Each ring represents a level of hierarchy. The central origin is labeled as root node. The nodes of the last level, which are located in the most outer parts of the chart, are labeled as leaf nodes or leaves [17]. Rings are sliced up and divided based on their hierarchical relationship to the parent slice. The angle of each slice is either divided equally under its parent node or can be made proportional to a value [15].

Advantages

Advantages of sunburst lie in the ability to show the whole picture of the hierarchy. Sunburst is able to illustrate the different depths of hierarchy and connections between groups. Treemap only provides limited view of the whole hierarchy and requires travelling down or up to view different levels. Another point where sunburst shines is the ability to indicate the depth of each branch. Thus, it is possible to quickly identify outliers and accumulation of values across branches. This results in an additional insight and it also contributes to a better depiction of the overall message.

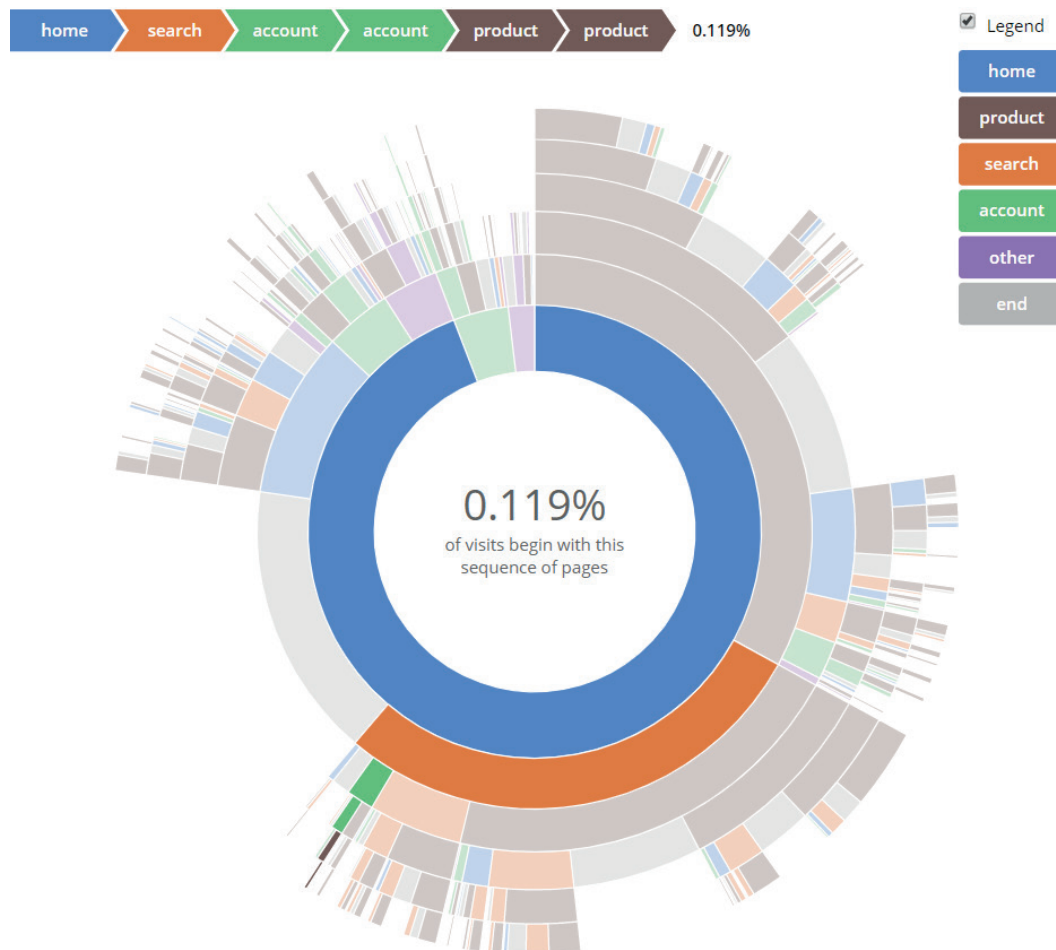


Figure 2: Flow of user actions on website [18].

Disadvantages

One of the ways to describe the data represented on sunburst is to use a legend. Unfortunately, this solution does not scale well with the increasing amount of categories. Therefore, one can employ another commonly used technique for information explanations by labels. Another challenge rises – depending on the size of the node and the length of the label text, it is highly probable that the label will not fit into the visual representation of the node. This may happen if the node is located far away from the origin where high fragmentation of the groups may occur on the left half of the most outer ring. Ignoring this issue will result in cluttering and the chart may even become unreadable.

2.3. Sankey diagram

Sankey diagrams display flows and their quantities in proportion to one another [19]. Directed arrows, called links, connect several entities, called nodes. The width of links indicates the proportional quantity of the flow. The larger the width, the higher the quantity of the flow [20]. Links can be combined together or split through their paths on each stage of a process. Color can be used to divide the diagram into different categories or to show the transition from one node to another [19]. The flow in Sankey diagrams always connects at least two nodes (processes). Thus it shows not only flow values but also information about the structure and distribution of the defined system [21]. Sankey diagrams are suitable choices for displaying movement of energy, money (figure 3), data in life cycle assessments [22] or other critical resources. Further, a Sankey diagram can be used to show the flow of any isolated system process [19].

Advantages

The Sankey diagram is useful for tracking resources inside a system. It allows us to follow the path of resource from the start to end. One can observe the state when entering a certain system and the final output at the end with ability to analyze progress at each node (stage) along the path [24]. It is possible to use interactive aspects and clearly separate single aspect within the system. Therefore, the complex system becomes simplified and optimizations can be made for a given aspect in order to improve efficiency [25]. The width of the individual links and flows provides interesting insights that may reveal extreme values, inconsistencies and other developments. This presents an opportunity that can be used in order to improve efficiency or gain benefits in previously neglected areas [25].

Disadvantages

Sankey diagrams may be hard for people who are not familiarized with data visualizations [25]. Issues may arise when comparing flows and links with similar widths which will complicate orientation [25]. The nodes require appropriate spacing. Otherwise, cluttering and overlapping issues may occur. Therefore, a Sankey diagram should be considered when the space is not an issue [25]. Incorrect placement may also lead to an unnecessary number of crossings between links [24].

2.4. Parallel Coordinates Plot

Parallel coordinates plot or parallel plot is a technique used for plotting and comparison of multi-dimensional numerical data. Parallel coordinates plot consists of several vertical axes where each represents one variable. These axes are placed parallel to each other. Individual data values are depicted as points on axes connected via lines between neighbouring axes and spanning across all dimensions. If placing all the connecting lines of a single object together, we will get a series that is representing this object across all the dimensions. These series of lines can be used to great benefit when it comes to improvements that increase interactivity. Parallel coordinates are helpful for comparing values of different objects across multiple attributes simultaneously.

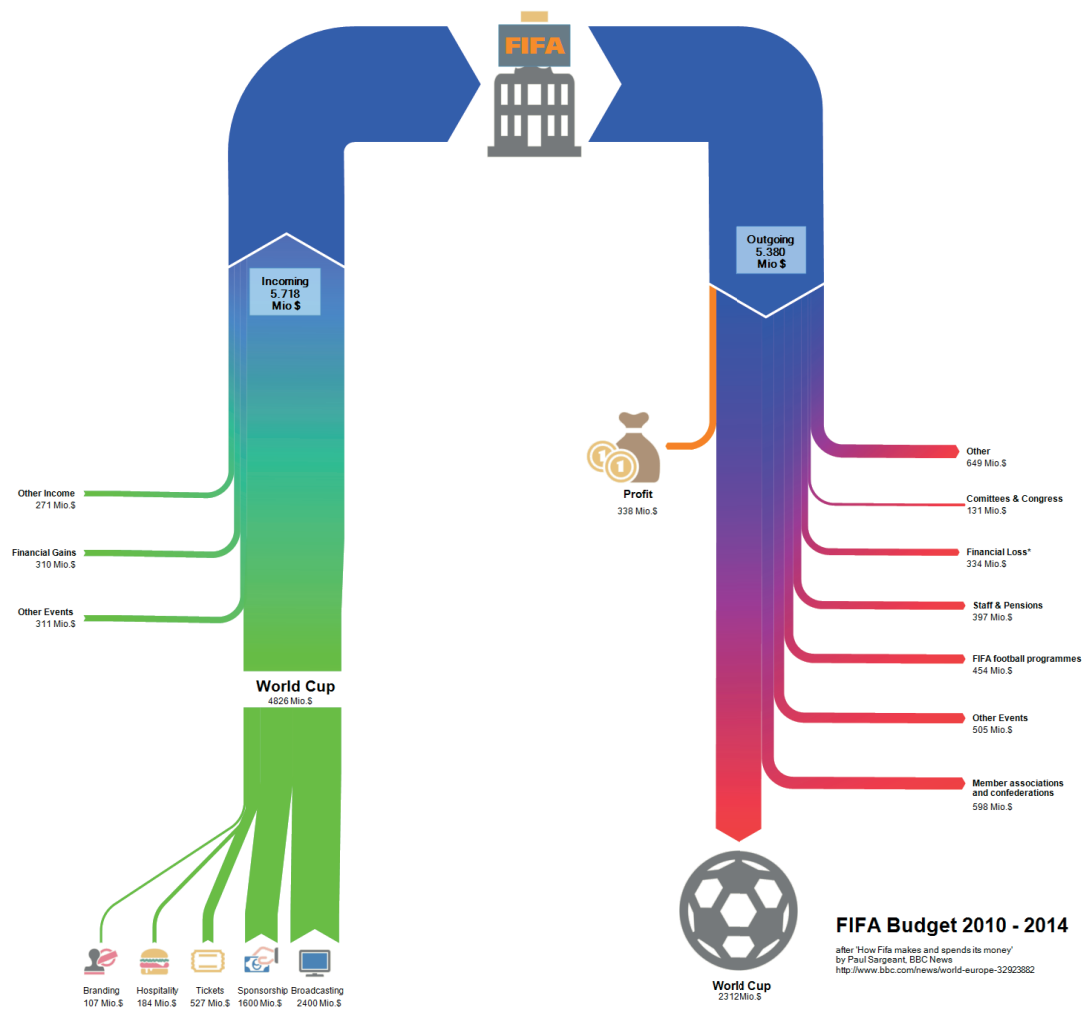


Figure 3: Revenue and outgoings of FIFA from 2010 to 2014 [23].

Advantages

Parallel coordinates plot has several variations that can be suited to specific needs and provide better insights. For example, regarding scaling [27], plot can use one scale for all dimensions which can be normalized or transformed to the desired range. It is also possible to have different scales on each axis as each dimension usually works with a different unit of measurement. Also, the axes can be rearranged. This may be helpful because optimization of the axis order may decrease clutter in the plot. The goal should be to minimize the amount of crossing lines which also improves readability of the plot. The order of axes can also have impact on how user understands the significance of the data because relationships between adjacent dimensions are easier to perceive [28]. Due to the nature of line plots, parallel coordinates plot often suffers from overlapping lines and cluttering that is making the figure unreadable. In order to neutralize this problem, highlighting a specific line with an additional tooltip or label may prove to be a priceless workaround offering the user much better experience when working with the plot. Furthermore, brushing [26] is a form of highlighting where all series based on a specified condition are highlighted 4. In addition, parallel coordinates plot has a 3D variant. In multi-relational 3D parallel coordinates the axes are placed on a circle with a focus axis in the center, allowing a simultaneous analysis between the focus variable and all other variables [29], as shown in figure 5.

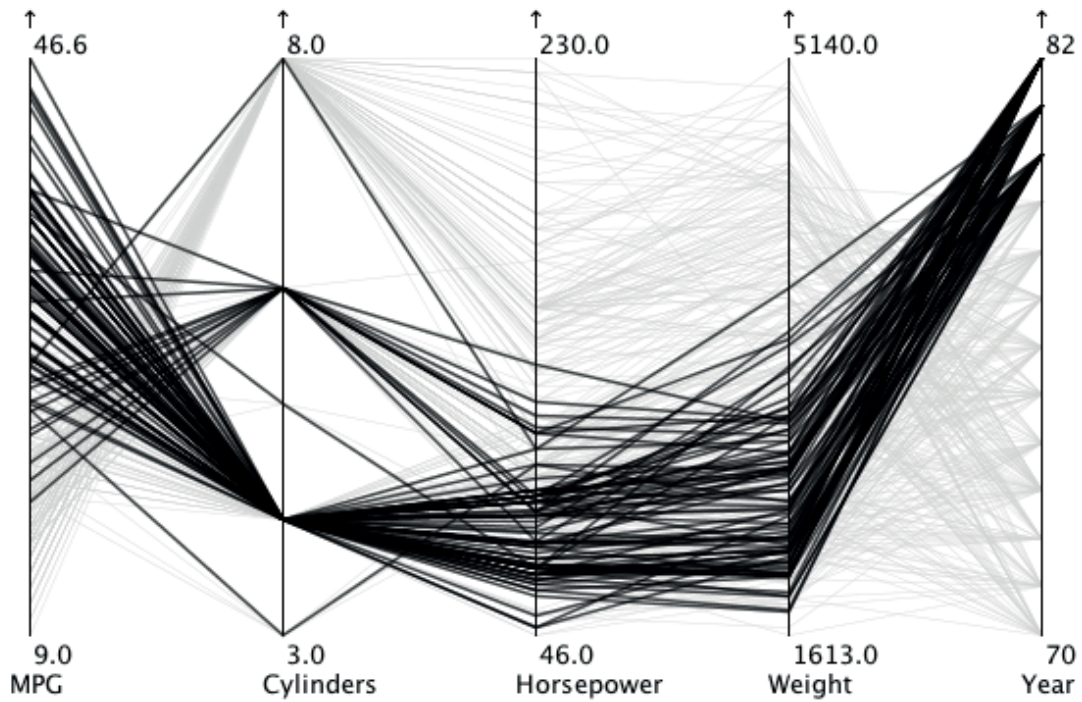


Figure 4: Brushed lines for years 1980 to 1982 on 'year' axis [26].

Disadvantages

Meanwhile, there are several disadvantages for parallel coordinates plot. First, it may display too many objects. This may lead to a large amount of overlapping lines which causes a cluttered and unreadable chart. Usage of axis reordering, highlighting, brushing and other interactive techniques to minimize this issue are highly recommended. Thus, it is hard to make a quick judgement. On the other hand, axis reordering and use of different scales for each axis may complicate interpretations [31]. Further, it may be interesting to perceive lines to be trends or value changes when using the line plots [31]. The purpose of lines is to connect points on the axes and make the relationship between these dimensions more obvious. There are more sources that cause cluttering issues rather than the large amount of lines. One of them is the amount of axes that depends on the width of the plot.

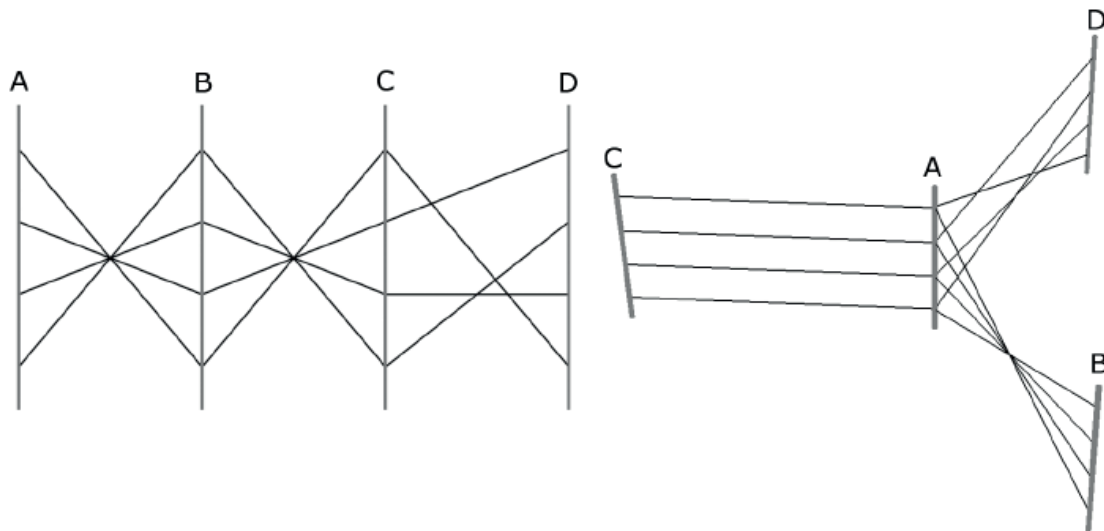


Figure 5: Comparison of parallel plots in 2D variant on the left and 3D variant on the right. [30].

2.5. Radar chart

This visualization method is often seen under many names, such as spider chart, polar chart, star pilots [32]. The aim of the method is to display multiple quantitative variables in a two-dimensional chart. Variables are represented by an array of axes where all start from a center, as shown in figure 6. The arrangement of the axes is radial and maintains equal distance between axes and the length of the axes are also usually the same. The values are placed on respective axis and the resulting set of points can be connected to form a polygon [32]. This polygon represents specific series as individual lines in the parallel coordinates plot with the arguable advantage being that a polygon is more memorable [33].

Advantages

Radar charts are very useful for multivariate analysis and the finding of extreme values. It also provides a quick general overview when certain factors are more prevalent than others. It is possible to display multiple series in the same chart and compare them. It is highly recommended to keep the number of series up to three, otherwise the chart becomes cluttered and will suffer from occlusion. When more series are to be displayed it is advisable to place each series into its own chart and group these charts in multiples, placing them into rows and columns. Then the user obtains the benefit of a compact overview for a larger number of series and the layout is also mobile-friendly [34]. It may also have a different scale for one or more variables. However, the scale should be properly indicated to the user since it will be assumed and expected that all the scales are the same.

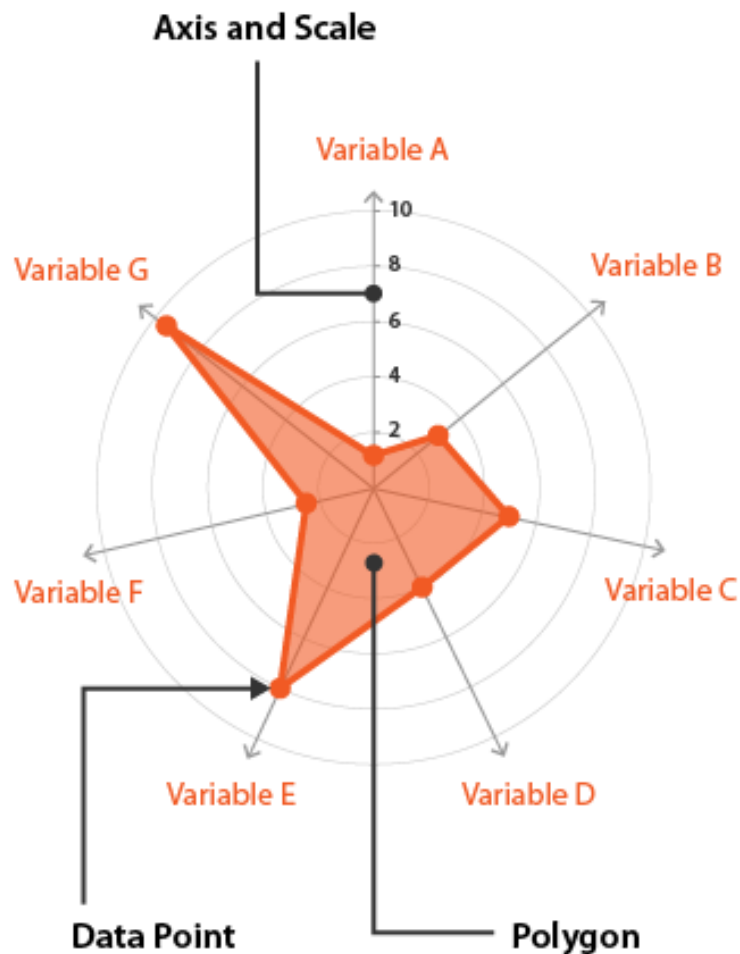


Figure 6: Anatomy of radar chart [32].

Disadvantages

When it comes to radar chart, it is not advised to use the chart as a tool to make trade-off decisions [35]. Similar to the number of displayed series, the large amount of axes will make the chart unreadable and hard to work with [32]. Also, the values on circular layouts are harder to compare. It is much more easier to compare values on bar plot or in other visualizations where values are laid in the same direction of vertical or horizontal axis [34]. Other plots may support order and ranking of variables whereas radar chart does not provide that information clearly [34]. Consider accompanying radar chart with a different visualization entity that supports ranking. When multiple series are displayed the chart will suffer from occlusion. This problem may be reduced by using transparent colors instead of opaque. However, transparent colors introduce tinting, which is hard to read without proper legend. Therefore, variant with no fill color of polygon seems to be the best option at handling occlusion [36].

The visualization does not deal well with data volume. It is not recommended to display more than three series in the same chart. The proportions of the usual shape of the chart makes it possible to stack radar charts horizontally and vertically into grids and comply to the responsive design of modern websites. The variety is restricted to qualitative data. However, the method gains points for decent handling of multi-dimensional data. The velocity handling is poor. The method only provides a snapshot of one moment. In order to observe development over time for new data, a completely new image would have to be composed. Finding patterns and comparing development between two or more completely separate entities may be quite difficult.

2.6. Choropleth

Choropleth is a very popular map-based visualization method widely used to visualize the geographical distribution of data. The main purpose is to display distribution of data and provide an easy way of comparison between geographical areas. In choropleth, the map is divided into regions and areas that are colored, shaded or patterned in relation to a data variable. This relation is usually visible in the form of color progression for each area of the map. This makes it possible to present data values for each geographical area and to observe variation and patterns [37]. In the classed version of choropleth, the variable is classified into several categories where each class has its own color with labels that indicate the range of values it coverage. The recommended amount of used classes is between 3 and 7. There are cases when only two classes are used, namely the political maps which often have only two parties. The more classes we use, the less generalization. However, more classes also means lower readability and higher risk of mistakes [38]. The values are mapped to a gradient consisting of a scale from low to high and each value is represented by a unique color, as shown in figure 7.

Advantages

The choropleth is easy to implement and widely used. It does not take that much time to adjust and use the method to their benefits. Also, it works well with enumeration units such as regions, provinces, or countries, which are suitable for geographical data that rely heavily on these spatial units [38]. Further, choropleth map works well with both numeric and categorical data.

Disadvantages

Choropleth does not offer a proper way to read or compare exact values. Multivariate visualization and display of correlation between values is not supported. Choropleth can be used when observing just one variable. Consider using different methods such as dot plot or scatter plot instead [40]. There is a perception bias towards larger areas. Choropleth maps show us how much geographical area is affected [40]. One has to be careful about making assumptions without enough information and context. For example, choropleths of visualizing voting results may color the map in red and blue depending on which party won in the area. Unfortunately, users are prone to make false assumptions connecting the size of the area to the number of votes.

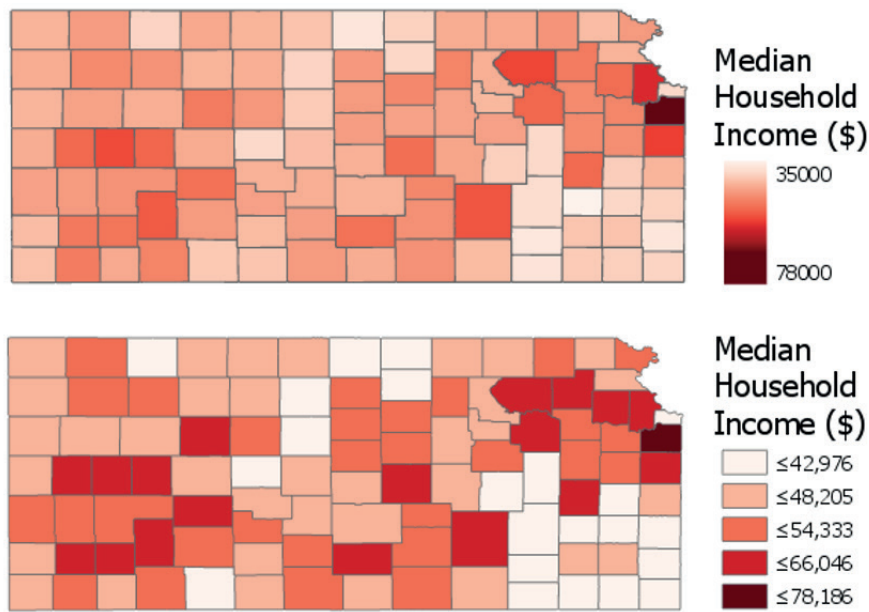


Figure 7: Comparison of color palette for classed and unclassed choropleth [39].

2.7. Streamgraph

Streamgraph is also called stream chart or ThemeRiver [41]. It is a variation of stacked area graphs used to visualize time series data [3]. The main difference is that streamgraph displays values along a central baseline [41]. The graph consists of several individual streams that form layers, an example of streamgraph can be seen in figure 8. These layers are placed within Cartesian axes where the y-axis represents the values of the series and the x-axis indicates time information [42]. The size of each layer is proportionally determined by the value of the respective time series at given time point [41]. Each layer is usually colored to distinguish easily between them and shading may be utilized to describe quantitative characteristics [41].

Advantages

The main advantage of the method is that it achieves two goals that are exclusionary in line-based charts where time series are plotted independently of each other [3, 42]. The first goal lies in the visualization of multiple time series in one chart. The streamgraph is beneficial when used to observe events, patterns and trends that are happening periodically or within specific time frames [41]. The evolution of values is portrayed in an aesthetically attractive way. Layers are usually placed in a manner that the resulting shape resembles a river [41]. This resemblance is useful because it provides an engaging way to illustrate the occurrence of events and trends [3]. Also, the streamgraph provides an efficient way to compare individual layers to the total sum. On the other hand, a layer-to-layer comparison is very difficult [42].

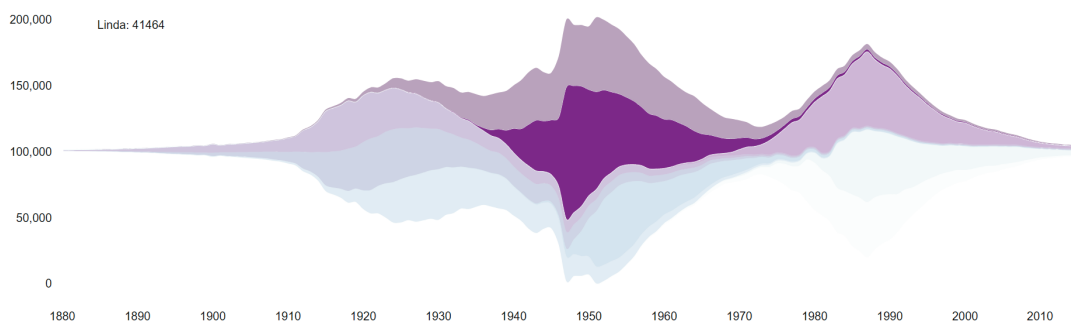


Figure 8: Baby name frequencies in the US from 1880 to 2015 [43].

Disadvantages

The streamgraph is not intended to work with negative values and will not yield proper results in these cases [44]. It is recommended to normalize values and work with compatible range [44]. This issue is caused by the fact that the vertical axis is used to convey size of layers and their stacking [45]. Since is is not applicable to multivariate analysis, the streamgraph works only with a single data dimension [3]. Several factors have an undesirable impact on the aesthetics as described in [42]. Layers may be distorted, placed inadequately or contain too much wiggle. Wiggle is a metric that indicates the degree of visual fluctuation within the graph, as shown in figure 9. The higher it is, the more complex the whole streamgraph is. Optimizations can be made in the baseline selection and layer reordering to minimize these issues. Further, larger amount of series might cause cluttering and readability issues [41]. Use tooltips and highlighting to mitigate this issue. In addition, series with smaller values are prone to be overshadowed by bigger ones and will result in barely noticeable layers. [41]. Use tooltips and highlighting to mitigate this issue.

2.8. Heatmap

The heatmap is a suitable selection when one is looking for a way to examine the correlation of multivariate data [46]. It is often used in a form of a matrix where the value of cells is indicated by their color and shading [46]. Usually, rows portray one category consisting of subcategories and the same applies to columns [46].

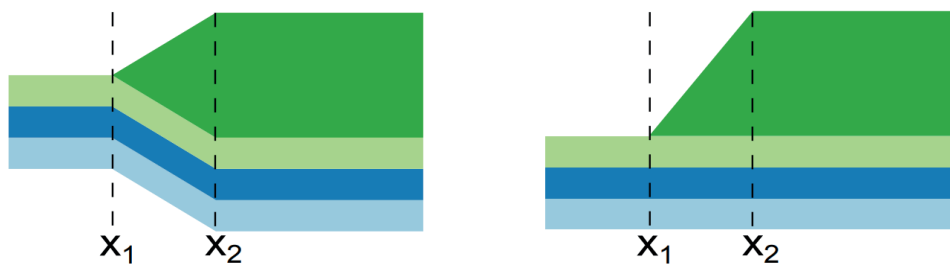


Figure 9: Illustration of different levels of wiggle when selecting baseline with a different algorithm. The left graph contains wiggle in all visible layers whereas the other graph suffers from wiggle only in the green colored layer [42].

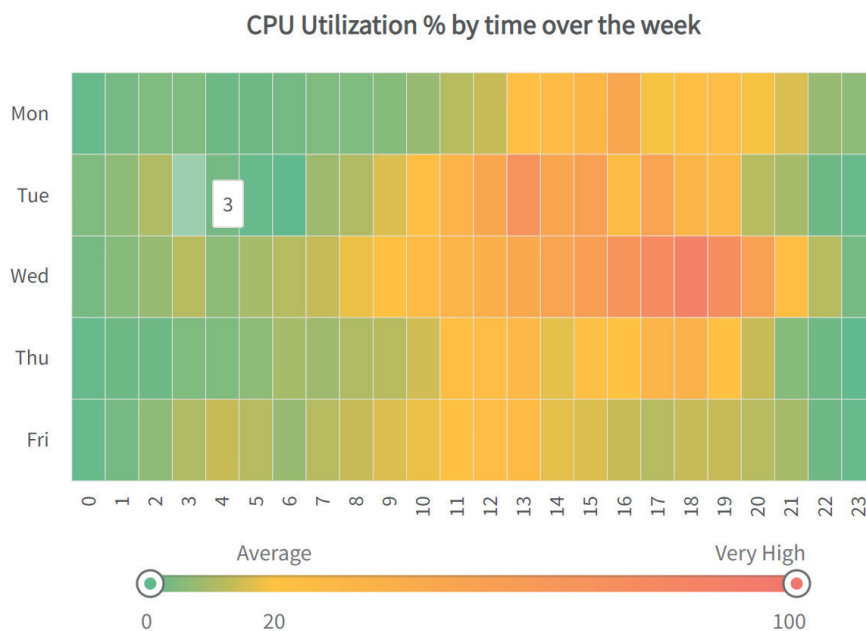


Figure 10: CPU utilization of 23 computers in a network during the working days [47].

Advantages

The heatmap is able to handle both numerical and categorical data [46], as shown in figures 10 and 11. Also, heatmap is useful for showing multivariate data, variance across multiple variables, patterns, general overview and development over time [48]. This method also leverages interactive elements, tooltips and a legend to provide more detailed information. For example, the heatmap in figure 10 provides an option to use the slider in order to highlight cells within a restricted range while other cells go temporarily blank. Furthermore, heatmaps can be used to display the result of hierarchical clustering [49], which is a statistical method that builds a structure which puts similar entities into close proximity. Another alternative is dendrogram. Both heatmap and dendrogram can be placed together in a way that they mirror each other [49].

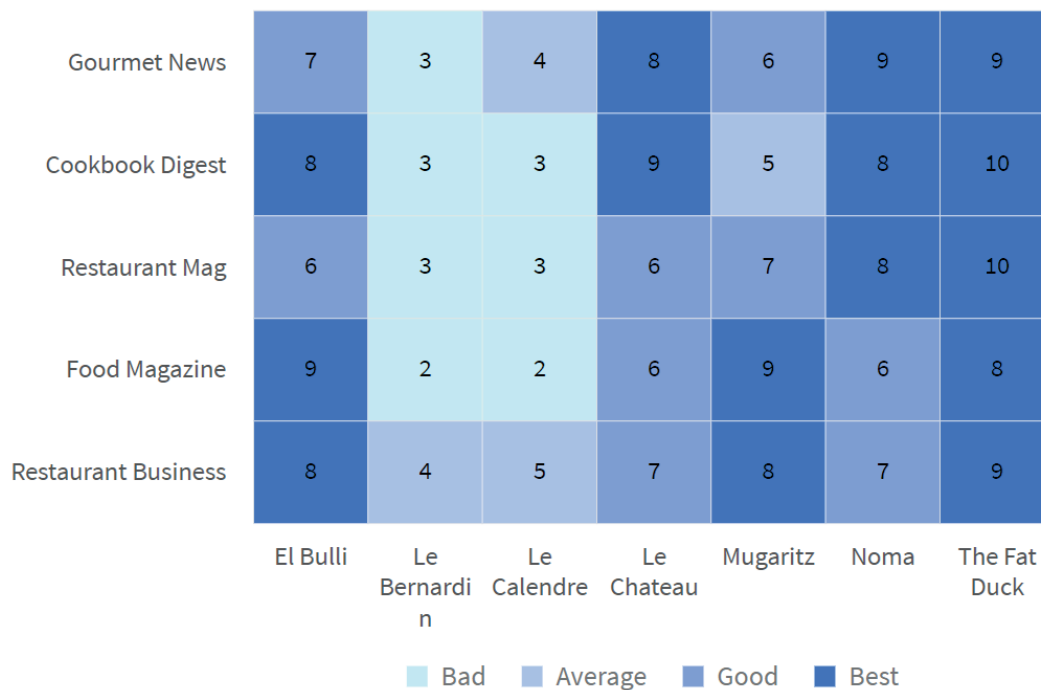


Figure 11: Reviews of restaurant with categorized rating [47].

Disadvantages

It is not advisable to use heatmaps for precision critical analysis. The usefulness of this method lies in its ability to provide a overview. This is caused primarily because of the reliance on color and shading. It can be hard to quickly differentiate certain breakpoints and accurately make conclusions even when given a legend or other supportive tools [46, 50]. Data containing a wide range of values may cause issues with the mapping of values to color scale. It will most likely be necessary to normalize the data [49]. color palette is very important [49]. One needs to be cautious of using more than two distinct hues in the color palette.

2.9. Violin Plot

The violin plot is a visualization method that is frequently used to display the distribution of data. Violin plot combines parts of two similar distribution methods, density plot and box plot (box and whisker plot) [51]. The shape usually consists of a box plot that is enveloped in two density curves (kernel density estimate – KDE) that mirror each other [52]. The anatomy of a box plot can be seen in figure 12.

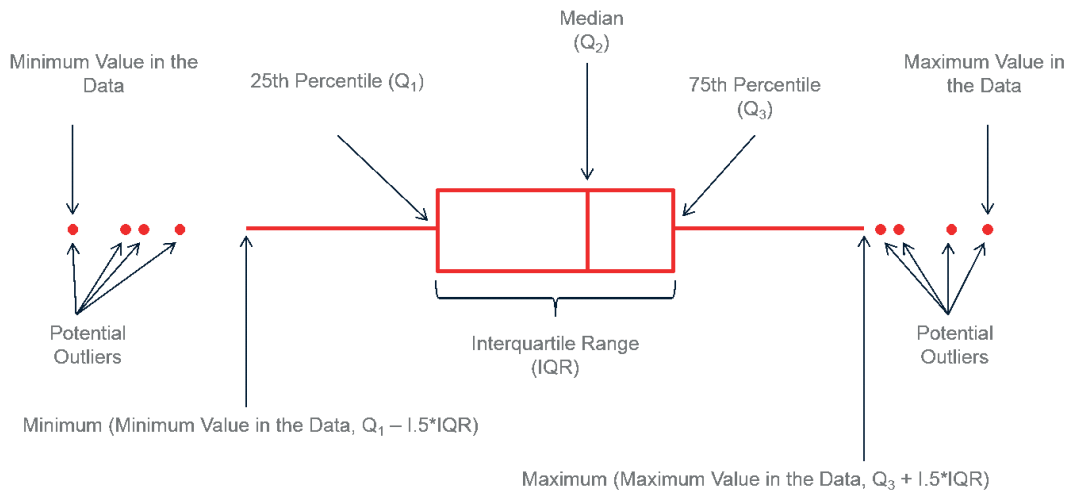


Figure 12: Anatomy of a box plot that may be used in a violin plot [53].

The white dot inside the box plot represents the data median [53]. The thicker line represents the interquartile range [53]. The thin lines extending from top and bottom of the thick line are called whiskers and their purpose is to depict variability outside of the upper and lower quartile [53]. The endings of whiskers represent the upper and lower extreme values (maximum and minimum). Anything beyond is supposed to be considered as an outlier of the dataset [53]. A numerical variable is depicted in several groups where each group has the form of a violin where the width corresponds to the density of data points in the range of the dataset [54].

Advantages

The violin plot provides an efficient way of simultaneously comparing both ranking and distribution of multiple groups [54]. Unlike box plots, it is able to handle multi-modal distributions [51, 55]. It is possible to reorder groups. A frequently used example is to use sorting by median which makes the

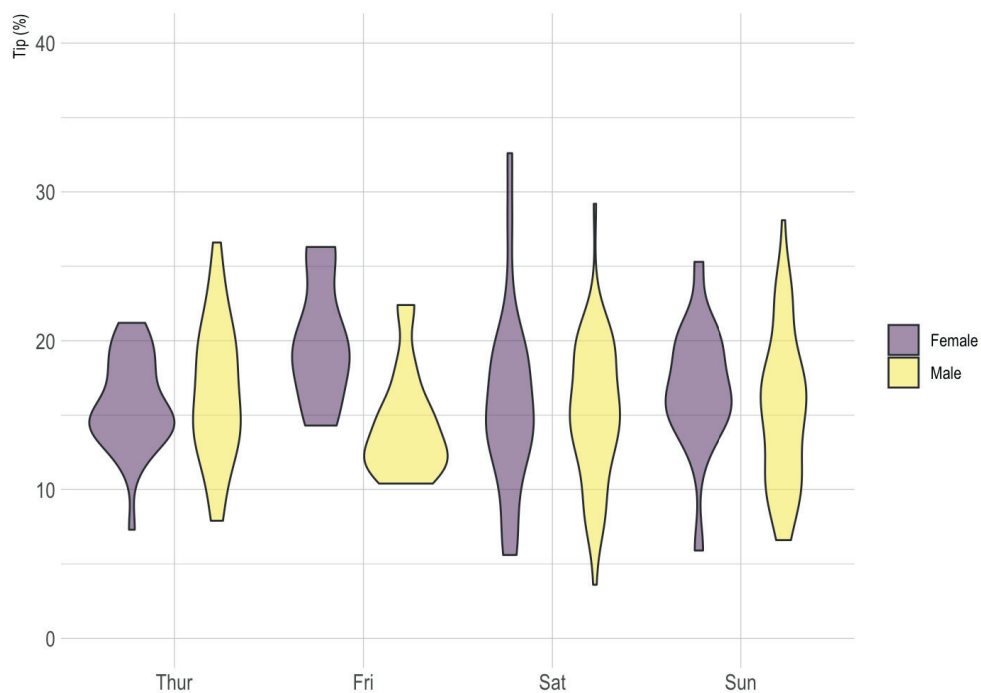


Figure 13: Distribution of restaurant tip during the week with the group distinction between male and female sex [54].

ranking of groups more evident [56]. In addition to grouping, it is also possible to use split violins. The left and right KDE of violin will no longer be mirrored, instead each of them will represent a different group [55, 57]. Likewise, if box plots are used, they should be separated into two individual entities corresponding to the KDE.

Violins can be placed in both vertical or horizontal orientation. This is handy in situations where text labels are too long or when the amount of groups is high and the cumulative length of all labels also causes overlapping issues [56]. Layout development in the area of mobile devices has taught us that when more free space is required, expanding in the vertical axis is the less troublesome way to go [56]. Also, violins are usually displayed with box plots as a part of them. This combination is assumed to be default, however, different objects can be used instead [56]. As shown in figure 14, it is also feasible to overlay the violin with a rug plot, jittered strip plot or swarm plot [56].

Disadvantages

There are also several disadvantages for the violin plot. For example, replacing the box plot as the inner part of a violin can introduce issues with noise and readability [56]. Different entities may provide better insight when the amount of data points is small, but with the increasing size of the dataset it may prove to be distracting and not helpful. If space constraints are an issue or the amount of groups is high, consider using the ridgeline plot [58]. The ridgeline plot saves space and is very useful when the data contains a clear pattern [58, 56]. Similarly, using only the box plot also saves space and it may be a preferable method if statistical summary is of the top priority [56].

2.10. Bullet Graph

The bullet graph was developed and popularised by Stephen Few [59]. The motivation for its introduction was to replace popular meters and gauges used in dashboards that were lacking in certain areas [60]. The bullet graph takes inspiration from traditional bar graph while offering several noticeable upgrades [61]. The bullet graph bar usually consists of five main parts as shown in figure 15. (1) Text label identifies the variable and unit of measurement. (2) Quantitative scale consists of tick marks that convey equal intervals of measure along the bar. (3) Feature measure is a bar indicating reached performance, the primary data. Given its importance, the bar should be visually emphasized. If the quantitative scale starts from value greater than zero, then the featured measure should be displayed in the form of a symbol, preferably a dot or X, instead of a bar. (4) Comparative measure is usually represented in the form of short but very noticeable line that is placed perpendicularly to the feature measure. The purpose of this object is to serve as a target against which the feature measure is compared. The comparative measure should remain visible when surpassed by the feature measure. (5) Qualitative scale is enveloping feature and comparative measures. It is divided into several shaded segments that represents qualitative ranges.

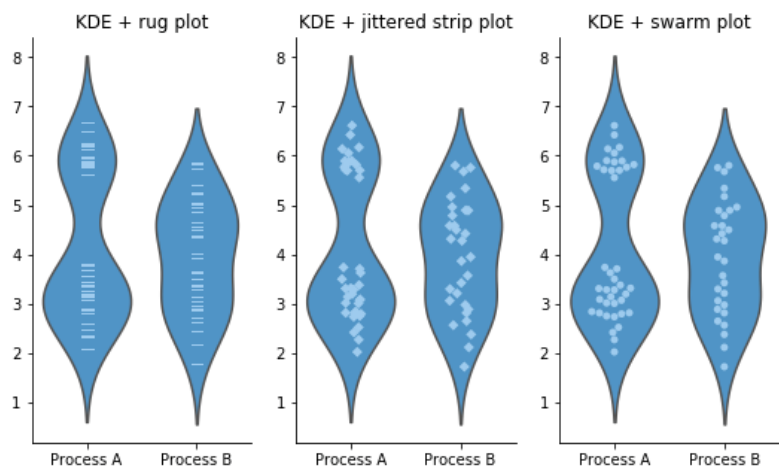


Figure 14: Different layout combinations of violins [56].

Advantages

The bullet graph provides intuitive depiction of performance (gain, loss) against a target value. There are several ways of using color and text labels to convey such information, an example can be seen in figure 16. The main advantage lies in space efficiency which is also the main factor behind utilization in dashboards [60, 62]. Depending on the situation and needs, it is possible to switch between vertical and horizontal orientation of the graph [59, 63]. The shading and hue of qualitative scale can be changed or flipped to indicate whether the state is favorable or highlight specific aspects [64]. Keep in mind though, that qualitative ranges should not be the primary element of the graph. Also this visualization is capable of incorporating future target and predictions [64]. One way to achieve this is to split the bar of feature measure into two distinct color shades. This makes it possible to track both the current progression and the estimated final result.

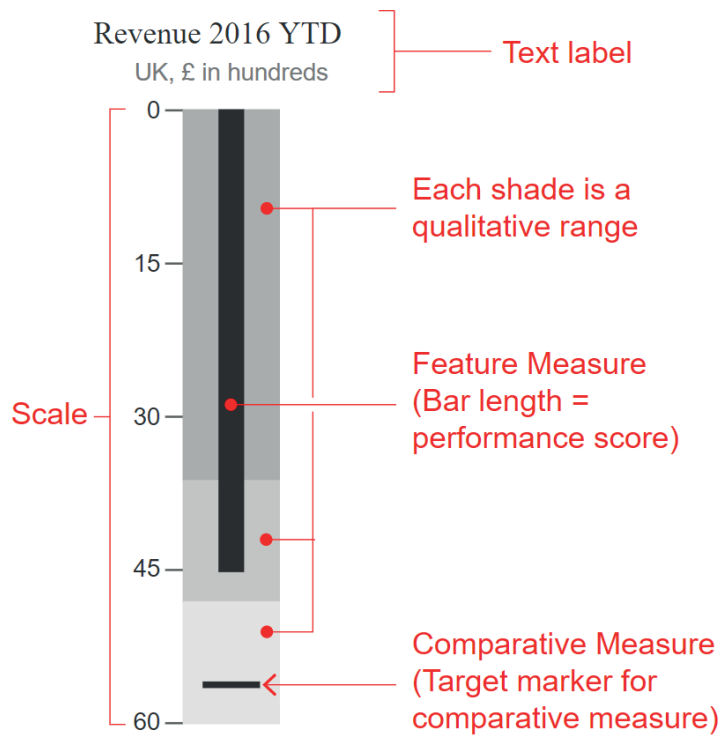


Figure 15: Anatomy of bullet graph [61].

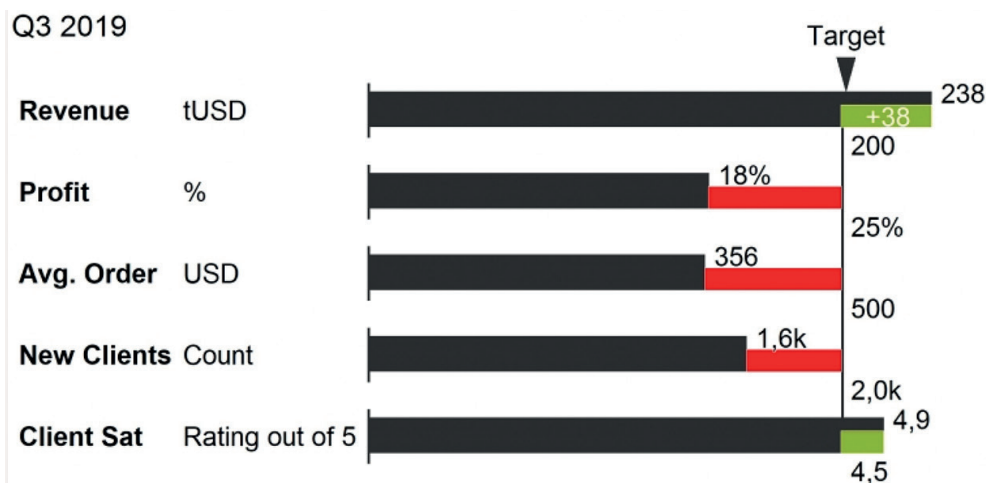


Figure 16: Different visual form of bullet graph displaying highlighted positive or negative variation from target [65].

Disadvantages

In the bullet graph, inappropriate color selection and configuration of qualitative scale can quickly make the graph hard to read and confuse inexperienced users [64]. Also, it is hard to identify the variation of the data value to the target [66]. The method indicates whether the target is reached, however, the variation itself and the amount may not be as quickly recognizable. It is possible to add another bar that would use specific colors to provide a fast indicator of positive or negative development [66], as shown in figure 16.

3. Recommendation Model

Multiple-criteria decision making (MCDM) is a decision tool that allows to analyze complex problems and evaluate the benefits and costs of all alternatives [67]. It also serves as a communication tool and a platform where different requirements can be specified and appointed with preferences and importance. During the selection of appropriate criteria several qualities should be addressed. Some basic qualities are completeness, redundancy and operationality [68]. Completeness refers to the importance of including all relevant criteria that may have impact on the specified goal/problem [68]. Redundancy recommends to remove unimportant or duplicate criteria [68]. Operationality requires the criterion to be clearly defined in order to make an assessment [68]. Furthermore, it must be possible to judge all alternatives against this criterion [68]. Due to the general nature of choosing the most suitable method, the list of criteria contains quite a large amount of items with a wide range of importance. The large amount of criteria can be structured into big data characteristics (Table 1), data features (Table 2), use scenarios (Table 3) and other characteristics (Table 4).

3.1. TOPSIS

The foundational method used in this model is TOPSIS, which is the abbreviation for Technique for Order of Preference by Similarity to Ideal Solution. The main advantage of TOPSIS is scalability – the procedure retains the same amount of steps regardless of the amount of criteria [69]. Another factor that had impact on the choice was its simplicity. The recommendation model is intended to be used by a web application which will be used by a wide range of users looking for advice. Many MCDM

Big data characteristics

Volume	Amount of records that can be processed and displayed. The criterion is influenced by the ability to highlight or filter desired data and visual elements.
Variety	Amount of attributes (dimensions) that can be visualized and handling
Velocity	Ability to provide interactive experience and handling of incoming and changing data.

Table 1: Classification of criteria in big data characteristics

methods include matrix calculations involving pair-wise criteria comparison [70] that may complicate the process beyond the understanding of a typical user and increase the overall time required to complete

the process over the tolerable limit. Luckily, TOPSIS does not rely on pair-wise comparison [70] and mentioned user unfriendliness can be avoided.

<i>Data features</i>	
Categorical	The method is capable to operate with data of categorical format.
Numerical	The method is capable to operate with data of numerical format.
Mixed num. & cat.	The method operates with a combination of both categorical and numerical data.
Geospatial	The method is capable of processing geospatial data. That may mean geographical coordinates or other data of specific meaning such country names, city names, country codes, etc.
Temporal	The method is capable to operate with time series and data conveying evolution over time. The criterion can be further differentiated by the focus on single or multiple series.
Network	The method is able to handle datasets with notable connections and relationships between individual nodes of data.
Hierarchical	The method is able to handle datasets where data is clustered into aggregated organisation or system.

Table 2: Classification of criteria in data features

TOPSIS was proposed by Huang and Yoon in 1981 [71]. It has been used in a wide range of areas such as supply and management logistics, engineering, marketing management, manufacturing systems and more [69]. TOPSIS begins with the construction of a decision matrix, as shown in equation 1, where each row represents an alternative and each column belongs to a criterion. Alternatives are rated against all criteria and the matrix is filled with initial scores. Two common rating scales are often used in MCDM: relative and ordinal [67]. In relative scale the alternative receives a value conveying the rank order of preference in comparison to other alternatives (e.g. "1 = best", "2 = second best" etc.). In ordinal scale the alternative receives a value based on how a particular interest is satisfied (e.g. a rating from 1 to 5 where "1" is equal to "very bad" and "5" is equal to "excellent").

$$DM = \begin{matrix} & C_1 & C_2 & \dots & C_n \\ L_1 & x_{11} & x_{12} & \dots & x_{1n} \\ L_2 & x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ L_m & x_{m1} & x_{m2} & \dots & x_{mn} \end{matrix} \quad (1)$$

<i>Use scenarios</i>	
Distribution	The method is capable of displaying frequency of data spread over an interval or group.
Flow/ Sequence	The method is suitable for displaying movement, flow or sequence of data.
Performance/ Ranking	The method is suitable for displaying reached performance, statistical overviews and provides an intuitive way of ranking records across multiple dimensions.
Composition	The method provides detailed insight about individual parts of a bigger overarching whole.
Comparison	The method provides a meaningful and comfortable way to compare individual records or attributes/dimensions of a single record.
Process modeling	The method provides a mechanism suitable for the modeling of a process. These mechanisms are usually in the form of branching/flow, conditions and time constraints.
Relationships	The method can be used to display correlation, connections or other forms of relationship within the data.

Table 3: Classification of criteria in use scenarios

The process is followed by normalization of the decision matrix. This operation will transform various criteria into non-dimensional attributes allowing comparison across all criteria [72], as shown in equation 2.

$$NDM = R_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^m x_{ij}^2}} \tag{2}$$

The next step introduces weights. Usually, criteria have different levels of importance which is the reason why many MCDM methods share this step of prioritizing of criteria. Each rating is multiplied by the attribute weight resulting in weighted normalized decision matrix, as shown in equation 3.

$$V = V_{ij} = W_j \times R_{ij} \tag{3}$$

Afterwards, both positive and negative ideal solutions are determined in equation 4. Beneficial and non-beneficial attributes have to be differentiated as their positive and negative impact is inverted.

$$\begin{aligned}
 &PIS = A^+ = V_1^+, V_2^+, \dots, V_n^+, \\
 \text{where: } &V_j^+ = \{(max(V_{ij}) \text{ if } f \in J); (min(V_{ij}) \text{ if } f \in J')\} \\
 &NIS = A^- = V_1^-, V_2^-, \dots, V_n^-, \\
 \text{where: } &V_j^- = \{(min(V_{ij}) \text{ if } f \in J); (max(V_{ij}) \text{ if } f \in J')\}
 \end{aligned} \tag{4}$$

Visualization method characteristics

Space efficiency	Degree to which the method utilizes space in a responsible manner and its ability to be responsive and scalable across various display devices. There are various interactive techniques that may greatly improve the amount of accessible information on the screen.
Range/variation	The method offers visual information about variation between upper and lower bounds.
Outliers	The method provides an intuitive way to examine data values that are vastly different from usual expectations and patterns, whether it is an individual data point or aggregated group.
Negative values	The method is capable of processing negative values without interfering with positive values. Certain methods can display negative values at the cost of confusion and high risk of misleading the interpretation.
Relative proportions	The method utilizes relative proportions, usually in combination with additional tooltip information, instead of the exact value depiction.

Table 4: Classification of criteria in other characteristics

The procedure continues with the calculation of the separation distance of each alternative to the positive and negative solution, as shown in equation 5.

$$\begin{aligned}
 S^+ &= \sqrt{\sum_{j=1}^n (V_j^+ - V_{ij})^2} & i = 1, \dots, m \\
 S^- &= \sqrt{\sum_{j=1}^n (V_j^- - V_{ij})^2} & i = 1, \dots, m
 \end{aligned}
 \tag{5}$$

Based on the separation distance calculations the relative closeness of each alternative to the ideal solutions is determined as shown in equation 6.

$$C_i = S_i^- / (S_i^+ + S_i^-), \quad 0 \leq C_i \leq 1
 \tag{6}$$

Finally, the relative closeness indicates the ranking of the model, where the higher the value, the better the alternative.

3.2. Rating for decision matrix

In order to perform a ranking calculation based on the TOPSIS method, the initial rating for alternatives and criteria has to be provided in order to construct the decision matrix. The rating is estimated based on the literature review and results are listed in figures 17 and 18. The rating score has a form of abstract value depicting the estimated ability of the visualization to fulfil the criteria. Due to this approach all criteria are considered to be beneficial. There is no need to transform different units into the same

range. Therefore, the rating is given within the range of 0 (poor) to 1 (good) with steps of 0.25 and the requirement of normalization is already satisfied.

3.3. Weights

The next step integrates weights into the decision matrix. Weights are obtained from a combination of basic data analysis of a data sample and survey that user fills in to indicate preference to certain criteria. The weights are translated into a 5 mark range similar to the previously described rating. Some criteria allow only weights of 0 or 1 (true/false).

3.4. Calculation of the ranking

The calculation itself is executed via Scikit-Criteria library [73]. This open source library offers a collection of MCDM methods packed in a scientific Python solution. The library provides an option to use TOPSIS and retrieve variables of the calculation such as the ideal solution, not ideal solution, closeness and the ranking itself. The calculation requires to be supplied with the aforementioned rating matrix, weight list and specification of the criteria impacts.

Method / Criteria	Volume	Variety	Velocity	Time series	One TS	multiple TS	Categoric	Numeric	Mixed Num Cat	Geospatial	Hierarchical	Network	Distribution
bullet graph	1	0.75	0.75	0.5	1	0	0	0	1	0	0	0	0
choropleth	1	0.25	0.75	0	0	0	0	0	1	1	0.1	0	0.5
heatmap	1	1	1	1	0	1	1	1	1	0	0.5	1	0.5
parallel coordinates	0.75	1	0.75	0	0	0	1	0	1	0	0	0	0.25
radar chart	0	0.75	0	0	0	0	1	0	1	0	0	0	0
sankey diagram	1	1	0.5	0	0	0	1	0	1	0	1	1	0.5
streamgraph	1	0	1	1	0	1	0	1	1	0	0.75	0	0.25
treemap	1	0.25	0	0	0	0	1	0	1	0	1	0	0.5
violin plot	1	0.25	0.5	1	1	1	0	1	1	0	0	0	1
sunburst	1	0.25	0.5	0	0	0	1	0	1	0	1	0	0.25

Figure 17: First part of the rating for decision matrix.

Method / Criteria	Flow (Sequence)	Performance (Ranking)	Composition	Space efficient	Range (Variation)	Outliers	Process modeling	Relationships	Comparison	Negative values	Number of variables	Relative proportions
bullet graph	0	1	0	1	1	0.25	0	0	1	1	0.75	0
choropleth	0	0.5	0.25	0.5	0	0.75	0	0	1	1	0.25	0
heatmap	0	0.75	0	1	0	1	0	1	1	0.75	1	0
parallel coordinates	0	1	0	0.25	0	1	0	1	1	1	1	0
radar chart	0	1	0	1	0	1	0	1	0.5	0.25	0.75	0
sankey diagram	1	0	0.5	0	0	0.5	1	0.5	0.5	0	1	1
streamgraph	0	0	0.5	1	0	1	0	0	0.25	0	0	1
treemap	0.25	0.25	1	1	0	0.75	0	0.25	0.75	0	0.25	1
violin plot	0	1	0	0.75	1	1	0	0	0.25	0	0.5	0.5
sunburst	0.5	0.25	1	0	0	0.5	0	0	0.25	0	0.25	0.75

Figure 18: Second part of the rating for decision matrix.

4. System Prototype

We have implemented the algorithm and the system that serves to visualize the recommendation process for the users¹. In the system, the top navigation is comprised of the step description in the form of a ring with an icon inside and a label underneath. Each step can be highlighted with the accent color indicating the overall progress of the form and the process, respectively. The bottom navigation can be comprised of a left-side aligned button with function of going a step backward or a right-side aligned counterpart for going to the next step. The elements of top navigation are clickable and can be used to traverse the multi-step form just as well as the bottom navigation. It is possible to go backwards and forwards, also, the new result can be generated quickly without going through the whole multi-form again. The multi-step form content includes these three steps: dataset upload, criteria and result.

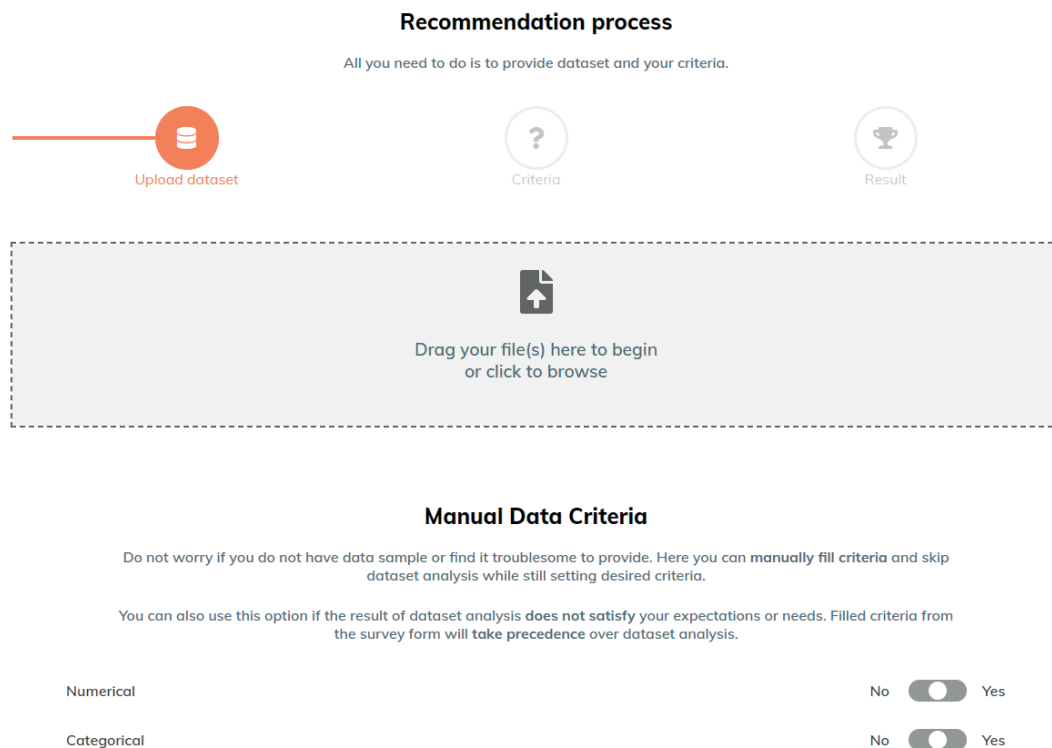


Figure 19: Showcase of the first step in the recommendation process.

The first step, labeled as dataset upload (figure 19), aims to retrieve information about criteria that are closely tied to data properties. These criteria include: numerical, categorical, mixed numerical & categorical, geospatial, temporal, one time series, multiple time series. The first option involves uploading a sample of dataset for analysis. An estimation is conducted and the result is displayed to the user. The second option offers the user a possibility to specify criteria manually. The specification is done via toggle buttons. Both options can be combined, where the manual option enables the user to overwrite result of dataset analysis.

The second step, labeled as criteria, is comprised of several lists of criteria and the user can specify his/her interest in desired criteria. Criteria are either rated in toggle style of true/false options or in a 5 likert scale. The value of 1 signifies zero interest whereas the value of 5 means the criteria is of utmost importance. The last step, labeled as result in figure 20, delivers the result of the recommendation process. The result is presented in a customized three-card layout that is often found on websites with comparison of prices and features of plans, packages or similar paid offerings. The layout depicts the top three choices of the recommendation. Each card starts with an icon of visualization, a text label of

¹ The source code of the system can be downloaded at <https://is.muni.cz/go/n8ngxg>.

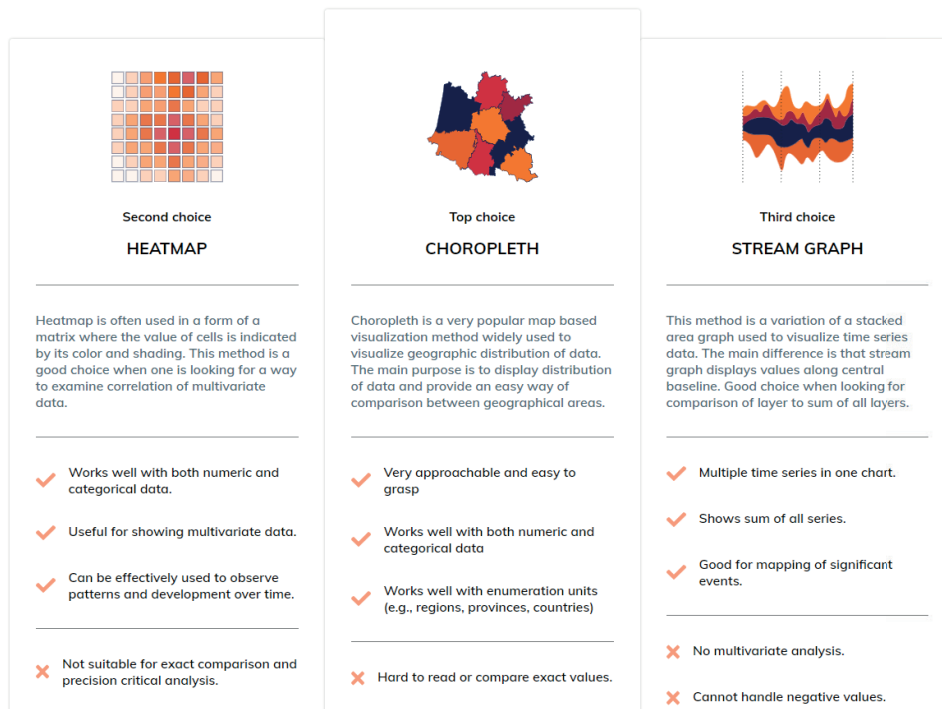


Figure 20: Showcase of the result view in the recommendation process.

the rank order and a visualization name. A short description of the visualization is placed in the middle section. Afterwards, a list of three selected positives and negatives can be found. The last element of each card is a button that links to a detailed description of the visualization.

5. Conclusion

The paper has reviewed and compared different up-to-date visualization techniques for big data. Following the selected visualizations, a set of criteria has been established for the recommendation model that leverages the MCDM approach of TOPSIS to derive suitable visualizations. Based on the proposed model, a system prototype has been implemented. The prototype has shown that the recommendation model can be implemented in practice and help to select data visualization efficiently. The paper has argued that effectively visualizing data is a complex task, while it offers valuable insights for interdisciplinary research. Although a plethora of visualizations have been proposed and each visualization has various advantages, disadvantages and suitable application domains, if the data characteristics and visualization usage are not well understood, the visualization may even complicate the decision-making process instead of helping to obtain insights.

As future works, the recommendation model can be extended to further investigate available possibilities for creation of visualizations in the context of related applications such as Tableau, FusionCharts, Google Analytics, or Sisense. This can be accomplished by enriching each description of visualization. Another future work is to conduct further experimental evaluations by comparing with other recommendation approaches to evidence the effectiveness of the proposed model.

Declaration

This paper is based on the master thesis of “Developing Effective Big Data Analytics by Leveraging Visualization Techniques” from the first author under the supervision of the corresponding author.

References and notes

- [1] M. Ge, V. Dohnal, Quality management in big data, *Informatics* 5 (2) (2018) 19.
- [2] A. Alsaig, V. Alagar, O. Ormandjieva, A critical analysis of the v-model of big data, in: 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), IEEE, 2018, pp. 1809–1813.
- [3] E. Y. Gorodov, V. V. Gubarev, Analytical review of data visualization methods in application to big data, *Journal of Electrical and Computer Engineering* 2013.
- [4] D. Laney, 3d data management, META group research note 6 (70) (2001) 1.
- [5] M. F. Uddin, N. Gupta, et al., Seven v's of big data understanding big data to extract value, in: Proceedings of the 2014 zone 1 conference of the American Society for Engineering Education, IEEE, 2014, pp. 1–5.
- [6] L. Wang, G. Wang, C. A. Alexander, Big data and visualization: methods, challenges and technology progress, *Digital Technologies* 1 (1) (2015) 33–38.
- [7] G. Andrienko, N. Andrienko, S. Drucker, J.-D. Fekete, D. Fisher, S. Idreos, T. Kraska, G. Li, K.-L. Ma, J. Mackinlay, A. Oulasvirta, T. Schreck, H. Schumann, M. Stonebraker, D. Auber, N. Bikakis, P. Chrysanthis, M. Sharaf, G. Papastefanatos, Big data visualization and analytics: Future research challenges and emerging applications, in: 3rd Intl. Workshop on Big Data Visual Exploration Analytics (BigVis 2020), 2020.
- [8] M. Lebid, Misleading statistics examples – discover the potential for misuse of statistics & data in the digital age (2018). URL: <https://www.datapine.com/blog/misleading-statistics-and-data>
- [9] R. McCready, 5 ways writers use misleading graphs to manipulate you (2020). URL: <https://venngage.com/blog/misleading-graphs>
- [10] A. Zoss, Data visualization (2019). URL: <https://guides.library.duke.edu/datavis>
- [11] H. R. Nagel, Scientific visualization versus information visualization, in: Workshop on state-of-the-art in scientific and parallel computing, Sweden, 2006, pp. 8–9.
- [12] TABLEAU SOFTWARE, Data visualization beginner's guide. URL: <https://www.tableau.com/learn/articles/data-visualization>
- [13] P. Laubheimer, Treemaps (2019). URL: <https://www.nngroup.com/articles/treemaps>
- [14] K. Robert, Visualization research, part i (2015) URL: <https://eagereyes.org/basics/visualization-research-engineering>
- [15] S. Ribecca, Sunburst diagram. URL: https://datavizcatalogue.com/methods/sunburst_diagram
- [16] J. Stasko, E. Zhang, Focus + context display and navigation techniques for enhancing radial, space-filling hierarchy visualizations, in: IEEE Symposium on Information Visualization 2000. INFOVIS 2000. Proceedings, IEEE, 2000, pp. 57–65.
- [17] Y. Holtz, Sunburst (2018). URL: <https://www.data-to-viz.com/graph/sunburst>
- [18] K. Roden, Sequences sunburst (2020) URL: <https://bl.ocks.org/kerryrodden/7090426>
- [19] S. Ribecca, Sankey diagram. URL: https://datavizcatalogue.com/methods/sankey_diagram
- [20] Google Developers, Sankey diagram (2020). URL: <https://developers.google.com/chart/interactive/docs/gallery/sankey>
- [21] ifu hamburg, What is a sankey diagram? (2020). URL: <https://www.ifu.com/en/e-sankey/sankey-diagram>
- [22] M. Schmidt, The sankey diagram in energy and material flow management: Part i: History, *Journal of industrial ecology* 12 (1) (2008) 82–94.
- [23] Phineas, Fifa accounts – my version (2015) URL: <http://www.sankey-diagrams.com/fifa-accounts-my-version>
- [24] Y. Holtz, Sankey diagram (2018). URL: <https://www.data-to-viz.com/graph/sankey>

- [25] FusionCharts, Sankey diagram (2020). URL: <https://www.fusioncharts.com/resources/chart-primers/sankey-diagram>
- [26] R. Kosara, Parallel coordinates (2010). URL: <https://eagereyes.org/techniques/parallel-coordinates>
- [27] Y. Holtz, Parallel coordinates plot (2018). URL: <https://www.data-to-viz.com/graph/parallel>
- [28] S. Rebecca, Parallel coordinates plot. URL: https://datavizcatalogue.com/methods/parallel_coordinates
- [29] J. Johansson, C. Forsell, M. Lind, M. Cooper, Perceiving patterns in parallel coordinates: determining thresholds for identification of relationships, *Information Visualization* 7 (2) (2008) 152–162.
- [30] H. R. Sankey, The thermal efficiency of steam-engines 125 (1896) (1896) 182–212
URL: <https://doi.org/10.1680/imotp.1896.19564>
- [31] Z. Gemignani, Better know a visualization (2010). URL: <https://www.juiceanalytics.com/writing/writing/parallel-coordinates>
- [32] S. Rebecca, Radar chart. URL: https://datavizcatalogue.com/methods/radar_chart
- [33] M. Akmanalp, Mis-employing radar charts to distinguish multidimensional data (2018)
URL: <https://towardsdatascience.com/plotting-in-many-dimensions-382fbd7fe76e>
- [34] Y. Holtz, The radar chart and its caveats (2018). URL: <https://www.data-to-viz.com/caveat/spider>
- [35] FusionCharts, Radar chart (2020). URL: <https://www.fusioncharts.com/resources/chart-primers/radar-chart>
- [36] G. Odds, A critique of radar charts (2011). URL: <https://blog.scottlogic.com/2011/09/23/a-critique-of-radar-charts>
- [37] S. Rebecca, Choropleth map. URL: <https://datavizcatalogue.com/methods/choropleth>
- [38] Axis Maps, Choropleth maps. URL: <https://www.axismaps.com/guide/univariate/choropleth>
- [39] C. Anderson, Making choropleth maps. URL: <https://www.e-education.psu.edu/geog486/node/671>
- [40] L. C. Rost, Choropleth maps (2018). URL: <https://blog.datawrapper.de/choroplethmaps>
- [41] S. Rebecca, Stream graph. URL: https://datavizcatalogue.com/methods/stream_graph
- [42] M. Di Bartolomeo, Y. Hu, There is more to streamgraphs than movies: Better aesthetics via ordering and lassoing, in: *Computer Graphics Forum*, Vol. 35, Wiley Online Library, 2016, pp. 341–350.
- [43] Y. Holtz, Streamgraph (2018). URL: <https://www.data-to-viz.com/graph/streamgraph>
- [44] P. Toffanin, Streamgraph in r [final] (2017). URL: <https://paolotoffanin.wordpress.com/2017/10/03/streamgraph-in-r-final>
- [45] A. Kirk, Stream graph. URL: <http://seeingdata.org/taketime/inside-the-chart-stream-graph>
- [46] S. Rebecca, Heatmap (matrix). URL: <https://datavizcatalogue.com/methods/heatmap>
- [47] FusionCharts, Heat map chart (2020). URL: <https://www.fusioncharts.com/resources/chart-primers/heat-map-chart>
- [48] V. Kesavan, What are heatmaps and what is a good way to use one? URL: <http://daydreamingnumbers.com/blog/what-are-heatmaps-and-what-is-a-good-way-to-use-one>
- [49] Y. Holtz, Heatmap (2018). URL: <https://www.data-to-viz.com/graph/heatmap>
- [50] A. Kirk, Heat map. URL: <http://seeingdata.org/taketime/inside-the-chart-heat-map>
- [51] S. Rebecca, Violin plot. URL: https://datavizcatalogue.com/methods/violin_plot
- [52] L. Ngo, 5 reasons you should use a violin graph (2018). URL: <https://blog.bioturing.com/2018/05/16/5-reasons-you-should-use-a-violin-graph>
- [55] D. Coleman, Box plot with minitab (2015). URL: <https://www.leansigmacorporation.com/box-plot-with-minitab>

- [54] Y. Holtz, Violin plot (2018). URL: <https://www.data-to-viz.com/graph/violin>
- [55] J. Carron, Violin plots 101 (2016). URL: <https://mode.com/blog/violin-plot-examples>
- [56] M. Yi, A complete guide to violin plots (2019). URL: <https://chartio.com/learn/charts/violin-plot-complete-guide>
- [57] H. Pampiglione, Violin plots (2019). URL: <https://ion.icaew.com/data-analytics/b/blogs/posts/violin-plots>
- [58] Y. Holtz, Ridgeline plot (2018). URL: <https://www.data-to-viz.com/graph/ridgeline>
- [59] T. Bodawala, Bullet charts – what is it and how to use it (2019). URL: <https://www.atyantik.com/bullet-charts-what-is-it-and-how-to-use-i>
- [60] FusionCharts, Bullet graph (2020). URL: <https://www.fusioncharts.com/resources/chart-primers/bullet-graph>
- [61] S. Rebecca, Bullet graph. URL: https://datavizcatalogue.com/methods/bullet_graph
- [62] S. Evergreen, Bullet graph (2014). URL: https://www.betterevaluation.org/en/evaluation-options/bullet_chart
- [63] B. Singh, 5 use cases for bullet charts (2017). URL: <http://visualbi.com/blogs/business-intelligence/dashboards/bullet-charts-use-cases>
- [64] S. Few, Bullet graph design specification, Perceptual Edge-White Paper 5.
- [65] Chart-me, C14h bullet graph columns bars. URL: <https://chart-me.com/portfolio-item/c14h-bullet-graph-columns-bars>
- [66] M. Wolff, Bullet graphs. URL: <https://www.chartisan.com/en/bullet-graphs>
- [67] Natural Resources Leadership Institute, Multi-criteria decision analysis. URL: <https://projects.ncsu.edu/nrli/decision-making/MCDA>
- [68] J. S. Dodgson, M. Spackman, A. Pearman, L. D. Phillips, Multi-criteria analysis.
- [69] M. Velasquez, P. T. Hester, An analysis of multi-criteria decision making methods, *International journal of operations research* 10 (2) (2013) 56–66.
- [70] P. Bajec, D. Tuljak-Suban, A framework for detecting the proper multicriteria decision-making method taking into account the characteristics of third-party logistics, the requirements of managers, and the type of input data, in: *Application of Decision Science in Business and Management*, IntechOpen, 2019.
- [71] C.-L. Hwang, K. Yoon, *Multiple attribute decision making—methods and applications* springer-verlag berlin heidelberg, New York.
- [72] P. Mishra, Topsis - a multi-criteria decision making approach (2013). URL: <https://www.slideshare.net/pranavmishra22/topsis-a-multicriteria-decision-making-approach>
- [73] J. B. Cabral, N. A. Luczywo, J. L. Zanazzi, Scikit-criteria: Colección de métodos de análisis multi-criterio integrado al stack científico de Python, in: *XLV Jornadas Argentinas de Informática e Investigación Operativa (45JAIO)- XIV Simposio Argentino de Investigación Operativa (SIO) (Buenos Aires, 2016)*, 2016, pp. 59–66. URL: <http://45jaiio.sadio.org.ar/sites/default/files/Sio-23.pdf>

JAIR

Journal of Applied
Interdisciplinary Research



Journal of Applied Interdisciplinary Research
Focus Issue: Artificial Intelligence (March 2023)

Principal Editors:
Prof. Dr. rer.nat. Peter Sperber & Prof. Dr.-Ing. Andreas Grzempa

The Journal of Applied Interdisciplinary Research (JAIR)
is a series of the Bavarian Journal of Applied Sciences. JAIR is an
English-language journal that is only published online
under <https://jas.bayern>.

Deggendorf Institute of Technology, 2023