Computers in Human Behavior: Artificial Humans 2 (2024) 100072

Contents lists available at ScienceDirect

Computers in Human Behavior: Artificial Humans

journal homepage: www.journals.elsevier.com/computers-in-human-behavior-artificial-humans

# Exploring people's perceptions of LLM-generated advice

Joel Wester *, Sander de Jong, Henning Pohl, Niels van Berkel

*Aalborg University, Aalborg, Denmark*

ARTICLE INFO

Original content: LLM-generated advice (Original data)

*Keywords:*
Large language models
LLM
Generative AI
Advice
User characteristics

ABSTRACT

When searching and browsing the web, more and more of the information we encounter is generated or mediated through large language models (LLMs). This can be looking for a recipe, getting help on an essay, or looking for relationship advice. Yet, there is limited understanding of how individuals perceive advice provided by these LLMs. In this paper, we explore people's perception of LLM-generated advice, and what role diverse user characteristics (i.e., personality and technology readiness) play in shaping their perception. Further, as LLM-generated advice can be difficult to distinguish from human advice, we assess the perceived creepiness of such advice. To investigate this, we run an exploratory study ($N = 91$), where participants rate advice in different styles (generated by GPT-3.5 Turbo). Notably, our findings suggest that individuals who identify as more agreeable tend to like the advice more and find it more useful. Further, individuals with higher technological insecurity are more likely to follow and find the advice more useful, and deem it more likely that a friend could have given the advice. Lastly, we see that advice given in a 'skeptical' style was rated most unpredictable, and advice given in a 'whimsical' style was rated least malicious—indicating that LLM advice styles influence user perceptions. Our results also provide an overview of people's considerations on *likelihood*, *receptiveness*, and *what advice* they are likely to seek from these digital assistants. Based on our results, we provide design takeaways for LLM-generated advice and outline future research directions to further inform the design of LLM-generated advice for support applications targeting people with diverse expectations and needs.

## 1. Introduction

The use of LLMs for everyday advice-seeking is on the rise. One reason for using LLMs, such as ChatGPT, when in need of advice is a perceived higher quality advice in comparison with traditional advice columns (Howe et al., 2023). Recently, research has focused on LLMs as advice providers that are more interactive as compared to existing solutions such as search engines, for example when looking for health-related advice (Birkun & Gautam, 2023). LLMs' ability to engage in a variety of conversational topics has led to ChatGPT and other LLMs being used for a variety of purposes. This includes increasing productivity, providing entertainment, and social interaction and support (Skjuve et al., 2023). While the popularity of using LLMs is increasing in general, people also express dissatisfaction about its performance—particularly holding true for users with low knowledge of LLMs (Kim et al., 2024). Recent efforts, such as better designs of LLM interfaces, includes Bing's AI support, which now allows users to choose between three different personality styles (Edwards, 2023).

However, how to adjust the behaviour and interface of using LLMs to varying end-users and end-user scenarios remains an open question. This is particularly relevant considering the growing likelihood that people will obtain advice on significant topics from algorithmic systems (Efendić et al., 2023). In this context, we know little about people's expectations, and how they perceive different interaction styles of LLMs. Völkel et al. recently investigated user perception of three different interaction types (extraversion, introversion, and average) as imbued in text messages provided by a conversational agent (Theres Völkel et al., 2022). The chatbot displaying an extraverted personality communicated enthusiastically and expressively (e.g., by saying "*Perfecttt*" and using emojis), in contrast to the introverted chatbot being more reserved and showing less emotions in its responses [57, p. 4]. As Völkel et al. conclude, similar to others (Chen et al., 2013; Yan & Chen, 2023), there is a need to better understand how people perceive different styles of advice depending on *who* is the perceiver. Although the role of user characteristics has been highlighted in prior research on interaction with digital agents (von der Pütten et al., 2010; Nishith Sharan & Romano, 2020; Cai et al., 2022)—the role of user characteristics in more interactive and open-ended LLM-based interactions remains underexplored.

* Corresponding author.
*E-mail addresses:* joelw@cs.aau.dk (J. Wester), sanderdj@cs.aau.dk (S. de Jong), henning@cs.aau.dk (H. Pohl), nielsvanberkel@cs.aau.dk (N. van Berkel).

https://doi.org/10.1016/j.chbah.2024.100072
Received 19 March 2024; Received in revised form 8 May 2024; Accepted 27 May 2024
Available online 7 June 2024

Jakesch et al. recently showed that AI self-presentations can come off as 'more human than a human'—indicating the relevance of exploring how LLM-generated advice can be paired with people's characteristics to better meet their expectations (Jakesch, Hancock, & Naaman, 2023).

We, therefore, set out to better understand how people perceive a range of LLM advice styles and the impact of user characteristics on their perception of these styles. An increased understanding of styles of LLM-generated advice and how it is perceived by humans, depending on who they are, can help avoid misinformed or misguided designs and lay the groundwork for more effective applications powered by LLM technology. Furthermore, meeting end-user expectations in online advice-seeking contexts can inform the design of advice-based LLMs, consequently making novel technology accessible to the public, such as ChatGPT (Jo et al., 2023). In this paper, we explore how people perceive LLM-generated advice and the possible effect of their characteristics on this perception.

To explore people's perception of LLM-generated advice, we designed topics of advice (e.g., asking for advice on living more sustainably), and manipulated the style of advice as presented by the LLM (e.g., 'caring', 'optimistic'). As aforementioned, LLMs have strong capabilities to express themselves through human-like text-based conversation. Considering recent insights on the difficulty of distinguishing between advice provided by LLMs and humans, we collected perceived creepiness ratings of the advice—computers communicating in human-like ways may cause eerie or uncanny feelings (Ciechanowski et al., 2019; Skjuve et al., 2019). Furthermore, we collected qualitative data on people's receptiveness to LLM-generated advice, their tendency to follow such advice, and what topics or concerns they would naturally seek support for from LLM-generated advice.

The results from our exploratory study suggest that user characteristics play a role in shaping perceptions of LLM-generated advice and that different advice styles further impact people's perceptions of these. Regarding perceived creepiness, we find that a 'whimsical' advice style was rated lowest on perceived malice, and the 'skeptical' advice style rated highest on perceived unpredictability. Our qualitative assessment indicates diverse user expectations and needs, where participants described LLM-generated advice as appropriate for sensitive concerns or topics. These results shed light on the effects of LLM-generated advice, indicating the benefits of structuring LLMs' responses to meet user characteristics. We discuss implications for user characteristics and LLM-generated styles of advice, and present design recommendations for LLM-generated advice.

## 2. Related work

Following recent breakthroughs in Large Language Models such as Bard and ChatGPT, users are now interacting with these models in various ways (Skjuve et al., 2023). As LLMs may influence user behaviour, it is critical to better understand human-LLM interactions and how people make sense of their experiences in interacting with this technology. More specifically, as recent research highlights, it is important to explore how LLMs and their output can be manipulated (Safdari et al., 2023).

The effects, influences, and related factors of advice-giving and advice-receiving have been researched from a variety of research domains, for example, internal medicine (Wee & Cornell, 2023), organizational psychology (Harvey & Fischer, 1997), experimental psychology (Lyn et al., 2005), and neuroscience (Biele et al., 2011; Goodyear et al., 2016). This has also been a focus in Human-Computer Interaction (HCI) research, where Waern et al. investigated people's perception of advice provided by computers in contrast to advice provided by humans (Wærn & Ramberg, 1996). Similarly, and more recently, Hertz and Wiese compared advice provided by humans and non-humans, showing that the difference between humans and non-humans matters less than the type of task and perceived expertise (Hertz & Wiese, 2019). Leib et al. provided participants with honesty-promoting and dishonesty-promoting advice given by either a human or an LLM and found that in both cases, dishonesty-promoting advice increased dishonesty, whereas honesty-promoting advice did not increase honesty (Leib et al., 2023). LLM output can appear 'more human than human' (Jakesch, Hancock, & Naaman, 2023) and presented misinformation may convincingly be presented as fact (Bender et al., 2021), further indicating the need to manage users' expectations and tailor messages to their needs. As people are increasingly likely to obtain advice from algorithms (Efendić et al., 2023), it is important to better understand users' perceptions of LLM advice. Hence, in the following, we describe related work on designs of LLMs, followed by the role of different user characteristics in the context of advice-taking.

### 2.1. Manipulating LLM output for Human-Computer Interaction

Recent research explored various ways of manipulating system output to meet user needs and demands (Theres Völkel et al., 2022), as users value features differently (Nov & Su, 2018). The underlying idea behind designing systems with human characteristics, such as personality and identity, is critical to better understand as systems get more complex. As people may increasingly treat text-based systems more like humans (e.g., treating chatbots as listeners (Lee et al., 2020)) in parallel with LLM models increasing in complexity and ability, understanding how users make sense of their experiences is increasingly important.

Prior work has focused on designing types of LLM output to better meet users' preferences and expectations. Recently, Jiang et al. argued that LLMs can be understood as social machines by inducing personality in their way of communicating (Jiang et al., 2022). The authors developed a dataset (a collection of Q&A items) based on the Big Five Personality Factors used to evaluate LLMs' personality instructed through "a specific personality in a controllable manner, capable of producing diversified behaviour" [23, p. 1], providing initial insights into how LLMs can be induced with a personality. Caron and Srivastava used the Big Five framework and explored how LLMs can display personality by consistently using language generation (Caron & Srivastava, 2022). Similarly, to see how different LLMs display different personality traits, Karra et al. used the Big Five factors to personify (quantify) traits relevant to different personalities by defining personality labels (Reddy Karra et al., 2022). Pareek et al. manipulated LLM explanations through four distinct explanation conceptualisations, namely 'consensual', 'expert', 'internal', and 'empirical' explanations, and evaluated their impact on user trust in a fact-checking context (Pareek et al., 2024). Their results show not only that explanations can increase user trust in AI explanations, but also highlight significant differences in the impact of explanation conceptualisations.

Ruane et al. explored the effects of chatbot personality on the user's experience, suggesting that people found the chatbot exhibiting high extraversion and agreeableness through 'high energy punctuation', a 'talkative nature', and 'sharing information by asking questions' as more friendly and cheery [45, p. 37]. In contrast, participants described the second chatbot that exhibited low extraversion and agreeableness through 'low energy', 'passive', and 'less interest' as very formal and less friendly [45, p. 44]. Wester et al. compared different techniques to deny user requests (Wester, Schrills, et al., 2024). Their results show that a diverting denial style, in which a chatbot steers away from the user's original request, was perceived as the most helpful and least frustrating way to deny requests. These results indicate that distinct chatbot output can be simulated through text-based interactions and highlight the necessity of considering designing distinct styles in text-based chatbots. As illustrated, it is critical to consider how personality can be perceived in interactive systems. Extensive prior work shows that (perceived) personality can be controlled by manipulating text outputs. However, how styles of LLM-generated advice in advice-giving contexts influence users' perceptions is unclear. Building on the aforementioned work that highlights the impact of natural language output on user experiences (Ruane et al., 2020; Theres Völkel et al., 2022), we set out to evaluate the effect of LLM-generated outputs in the context of receiving advice.

### 2.2. LLM-generated advice and user characteristics

Yan and Chen showed how personality traits influence peoples' ratings of recommender systems output (Yan & Chen, 2023). Others point toward emotionality, suggesting that people scoring higher on incidental anger show lower receptiveness to advice (Gino & Schweitzer, 2008). Shaw and Hepburn investigated advice-giving in telephone calls between mothers and daughters, with results indicating that advice acceptance depends on the identity of the advice recipient (i.e., being positioned as advice recipient insinuates lesser knowledge) (Shaw & Hepburn, 2013). Moreover, Feng and MacGeorge find that a person's expertise, closeness, and history strongly correlate with the receptiveness of advice (Feng & MacGeorge, 2006). Extensive prior work in HCI and Computer-Supported Cooperative Work (CSCW) has focused on how user characteristics influence the interaction with digital systems (e.g., personality or identity) (Li & Chignell, 2010; Li et al., 2023). However, bridging between end-users diverging expectations or needs and LLM-generated advice is an unresolved challenge.

A good example of highlighting this problem is presented by Li and Chignell, who investigated how blog readers tend to enjoy texts that signal similar personalities as themselves (Li & Chignell, 2010). The authors showed that people were more attracted to texts written by blog writers with similar personality traits. Interestingly, their study's results suggest that words in text may display personalities in terms of linguistic cues. Similarly, Van der Pütten et al. investigated how peoples' personality influence their interactions with virtual characters (von der Pütten et al., 2010). The authors showed that personality traits influence their evaluation of virtual characters and predict the evaluation outcome. Moreover, their results suggest that personality traits are a better predictor in users' evaluations than the actual behaviour of the virtual character (von der Pütten et al., 2010).

In the context of trust in human-agent interactions, Cerekovic et al. explored how personality influences user rapport with an agent, collecting personality ratings through a self-report survey as well as observed behaviour (behavioural cues correlated with personality traits), suggesting that people's personality is a strong predictor of rapport towards the agent (Cerekovic et al., 2017). Moreover, people rating higher on extraversion and agreeableness showed more rapport towards the agent (Cerekovic et al., 2017). Sharan and Romano investigated what effects user personality and locus of control have on peoples' trust towards AI systems (Nishith Sharan & Romano, 2020). Their results indicate significant differences in trust towards the system depending on the user's personality and that personality overrules other factors in the interaction. Moreover, participants rating high in neuroticism and extraversion showed lower agreement ratings compared to participants scoring lower on neuroticism and extraversion (Nishith Sharan & Romano, 2020).

Lastly, technology readiness has repeatedly been shown to predict users' tendencies to engage with new technology (see (Chen & Lin, 2018; Chiu & Cho, 2021; Kim et al., 2023; Metz & Wörle, 2021, pp. 165–172)). For example, Lin et al. have explored the role of technology readiness in accepting self-service technology (services without any employees involved, such as telephone voice responses (Lin & Chang, 2011; Lin & Hsieh, 2006)). Their results indicate that one's technological readiness attitude influences the perceived usefulness and ease of use of the technology. Similarly, Wang et al. recently showed how teachers' ratings of AI readiness positively predict their ratings of AI-enhanced innovations in teachers' work (Wang et al., 2023). The impact of digital technology on our character is an increasing concern in the wider field of Computer Science, especially amidst the increasing capabilities of AI. With the increasing realisation of the continuous interplay between technology use and user characteristics, we set out to assess the role of personality (Ruane et al., 2020; Theres Völkel et al., 2022) and technology readiness (Lin & Chang, 2011; Lin & Hsieh, 2006) in the context of receiving LLM-generated advice.

## 3. Study design

To evaluate how different styles of LLM-generated advice are perceived, we conducted a crowdsourcing study. We instructed ChatGPT to generate advice in accordance with the dimensions of advice style and advice topic. In addition to participants' ratings of the advice, we also elicited a range of participant characteristics to investigate how these characteristics influence perceptions towards the LLM advice.

### 3.1. LLM-generated advice

We used OpenAI's *gpt-3.5-turbo-0301* model with a set temperature of *1.0* to generate all advice. Following a template for the advice (see Appendix A), we automatically created system prompts (i.e., a template for the generated responses), user prompts (i.e., the topic of advice), and query ChatGPT for the resulting advice (i.e., the actual response of ChatGPT to the generated prompt).

We vary the generated advice across two dimensions: *style* and *topic*. Styles represent advice-giving displays, and we included five of them: *balanced*, *whimsical*, *caring*, *skeptical*, and *optimistic*. For balanced advice, we only require it to be "balanced and neutral", while we created more elaborate descriptions for the four other styles. Each of these instructions asks ChatGPT to "act like a person that …", followed by the respective description. For example, for 'optimistic', we provide the description "act like a person that sees the positive side of things, expects things to turn out well, and believes that you have the skill and ability to make good things happen".

We also varied advice topics to alleviate the effects of any one specific topic. In contrast to styles, topics are not system instructions for ChatGPT, but instead questions posed to ChatGPT. For this, we picked four areas (personal health, relationships, climate change, and career) that we deemed typical of advice-seeking. Furthermore, we created two variants of each topic, broad and specific questions, to ensure a diverse set of advice scenarios. This resulted in eight questions in total. For example, for personal health advice, we ask *"How can I lose a few pounds?"* (specific) and *"How can I improve my physical health?"* (broad).

Finally, in our system prompt, we instructed ChatGPT to return short (about 150 words) 'advice columns' and to exclude any pretext and posttext from the output. We also asked for Markdown-formatted responses, which enabled ChatGPT to output formatted text or include list elements. Subsequently, we converted the questions and returned advice to HTML snippets we can show participants (see Fig. 1). Please see Appendix A for a full overview of our prompt commands.

Throughout our prompt design process, we iterated on the requirements and instructions by trial and error, taking inspiration from open-source prompt repositories (e.g. [1]) and available literature on styles commonly used in designing interactive systems.

### 3.2. User characteristics

We posit that users' reaction to and perceptions of LLM-generated advice varies, depending on the users themselves. Hence, we include a range of scales for different user characteristics. Namely, we include validated *personality* and *technology readiness* scales. For personality, we use the BFI-2 scale (Soto & John, 2017), which covers measures of extraversion, agreeableness, conscientiousness, negative emotionality, and open-mindedness. We included technology readiness, as captured by the TRI2 scale (Parasuraman & Colby, 2015) ,[2] to see whether

[1] https://github.com/f/awesome-chatgpt-prompts.

[2] The Technology Readiness Index 2.0, is copyrighted by A. Parasuraman and Rockbridge Associates, Inc., 2014. The scale may be duplicated only with written permission from the authors.

**You asked:** How can I lose a few pounds?

**The advice you got was:**

**Balanced Advice Style**: Losing weight is a common goal for many people, but it is important to approach it in a healthy and sustainable way. In order to lose a few pounds, you can start by making some small changes to your diet and exercise routine. Consider tracking what you eat and aiming for a caloric deficit, but make sure you are still consuming enough nutrients and not depriving your body of essential fuel. Incorporating more physical activity into your daily routine, such as going for a walk after dinner or taking the stairs instead of the elevator, can also help you burn more calories. Remember, weight loss is not a one-size-fits-all solution, so listen to your body and make adjustments as needed. Additionally, consulting with a healthcare professional or registered dietitian can provide personalized guidance and support on your weight loss journey.

**Caring Advice Style**: Firstly, it's important to remember that your worth is not defined by your weight. That being said, if you want to shed some pounds, focus on making small, sustainable lifestyle changes rather than crash diets. Start by adding more fruits and vegetables to your diet, and drinking more water. Cut back on processed and sugary foods, but don't deprive yourself completely. Finding a form of exercise you enjoy, such as dancing or hiking, can also help you lose weight while improving your mood and overall health. Lastly, remember that progress takes time and be patient with yourself. Try not to obsess over the number on the scale and instead focus on how you feel. Above all, remember to love and care for your body no matter what size it is.

**Optimistic Advice Style**: Analyze your eating habits, try to consume smaller portions, and make healthier food choices. Try to incorporate more fruits and vegetables, lean proteins, and whole grains into your diet. Cook meals at home and avoid fast food as much as you can. Stay hydrated by drinking plenty of water throughout the day. Exercise regularly, make it fun, and find something that you enjoy. This can be anything from cycling, running, swimming, or dancing, to yoga or Pilates. Not only will exercise improve your physical health, but it is also an effective way to reduce stress and anxiety. Finally, stay positive and keep yourself motivated. Don't focus on the end result but rather the journey itself. Celebrate your small victories, and remind yourself why you are doing this. Remember that every healthy choice you make is a step in the right direction, and know that with time and perseverance, you will reach your goal.

**Skeptical Advice Style**: Honestly, I'm not sure there's a foolproof way to lose a few pounds. Many diets and exercise plans sound too good to be true, and that's usually because they are. It's important to remember that weight loss is a journey, not a one-time event. And it's okay if progress is slow, as long as it's sustainable. If you're looking to lose a few pounds, the most important thing is to make lifestyle changes that you can stick with long-term. This might mean finding an exercise routine that you enjoy, and incorporating more fruits and vegetables into your diet. It's also important to stay consistent, even when progress is slow. But remember, no two people are the same, and what works for one person might not work for someone else. So, while I can't give you a specific plan to follow, I encourage you to experiment with different strategies until you find what works best for you.

**Whimsical Advice Style**: Ahh, the age-old question. How can one shed the extra fluff without resorting to wearing a corset or cutting off an arm? Well my dear, there are many ways to go about it, but if you want to do it in a healthy, sustainable fashion, you may want to consider a few things. Firstly, what are you eating? Are you eating a lot of processed foods or sugar? Perhaps try incorporating more whole foods into your diet, such as fruits, veggies, and lean proteins. And don't forget to keep yourself hydrated, water is your friend! Secondly, how active are you? Try to find some activities you enjoy that get you moving. Maybe dancing, hiking, or chasing after your dog around the park. Consistency is key, so find something you love doing and stick with it. Lastly, don't forget to be kind to yourself. Losing weight can be tough, but don't beat yourself up if you slip up. Just keep moving forward and remember that progress is progress, no matter how small. Good luck on your journey, my dear!

**Fig. 1.** Examples of the various advice styles on topic: personal health, specific presented to the participants. The shown advice was generated by OpenAI's gpt-3.5-turbo-0301 model following the five advice styles (balanced, caring, optimistic, skeptical, whimsical) as a result of our system prompts (see A). Participants only saw one advice style throughout the entire study.

participants' attitudes to new technology, such as LLMs, have an impact on their perceptions thereof. This scale includes measures of optimism, innovativeness, discomfort, and insecurity.

### 3.3. Measures

For each piece of advice we presented to participants, we asked them to rate four statements on 5-point Likert scales: (1) *I like this style of advice*, (2) *I would be likely to follow this advice*, (3) *I found this advice useful*, and (4) *A friend of mine could have given this advice*. During the study, each participant only saw one style, but broad and specific versions of all four topics of advice. To control for any specific advice topic influencing participant ratings, we aggregate this data into a single rating for each of the four above statements per participant.

Furthermore, at the end of their session, we asked participants to rate the LLM-generated advice on the validated technology creepiness scale (Woźniak et al., 2021) given the often personal nature of receiving advice—as well as AI advice potentially coming off as 'more human than a human' (Jakesch, Bhat, et al., 2023). This scale measures *Implied Malice*, *Undesirability*, and *Unpredictability* on 7-point scales, forming an overall *Perceived Creepiness* rating. In addition to and complementing our quantitative measures, we asked participants to provide open-ended text responses to the following questions.

- How does receiving advice from a digital assistant affect your likelihood of following up on the advice?
- How does receiving advice from a digital assistant affect your receptiveness to the advice?
- When you are seeking advice, for which problems or questions are you more likely to reach out to a digital assistant than to a friend or family member?

### 3.4. Procedure

Participants first provided informed consent and then filled out the personality and technology readiness questionnaires. Next, participants read the following study instructions:

> In this study we investigate the use of digital assistants as advice givers. Digital assistants are systems you can have a chat or conversation with and that are trained to provide help for common

everyday tasks. For example, you can ask them for the weather, to translate a bit of text, or how to save money on your groceries. Imagine yourself in the position of seeking advice, asking the digital assistant for advice. Your task will be to rate the advice given by the digital assistant, as if it was given to you. In each task, you will see the question asked to the digital assistant, as well as the advice the digital assistant provided.

Following, we used Qualtrics' built-in randomisation function to distribute the pooled participants—they were automatically assigned to a style condition (counterbalanced across all participants) and received eight corresponding pieces of advice in randomised order. Each piece of advice was shown together with the posed question and in the format of an LLM response (see Fig. 1 for examples). For each piece of advice, participants rated the four aforementioned statements on a 5-point Likert scale. Following the completion of all tasks, participants completed the creepiness questionnaire and the three open-ended questions.

### 3.5. Participants

Using Prolific, we recruited a sample of 91 participants (39 female, 52 male). The average age of our participants is 36.2 years (SD = 11.4), ranging between 18 and 70 years of age. Of our participants, a majority reside in the United Kingdom (62%), South Africa (14%), Canada (9%), and Australia (5%). The average completion time of our survey was 22.1 min (SD = 24.5). Participants were required to speak English as their first language, and could participate in the study using any desktop device. Recruitment parameters were set to participants having a minimum number of 100 previous submissions and a 95% approval rate. Participants could participate only once in the study and were compensated using an hourly rate of £9,00.

## 4. Results

Following recent approaches to data preparation for regression analysis (Kang et al., 2022), we first assessed if our dataset met the major assumption for regression analysis (i.e., multicollinearity). In doing so, we excluded *Innovativeness (Technology Readiness)* due to the violation of the multicollinearity test. We consequently include measures of *Personality (Extraversion, Agreeableness, Conscientiousness, Negative Emotionality, Open Mindedness)* and *Technology Readiness (Optimism, Discomfort, Insecurity)*. For each scale, we first computed Cronbach's alpha scores to validate their internal consistency. We report medium to high Cronbach's alpha scores for personality ($\alpha$ =.79, 95% CI) and technology readiness ($\alpha$ =.61, 95% CI). We furthermore found high Cronbach's alpha scores for perceived creepiness ($\alpha$ =.80, 95% CI). We conducted automatic stepwise regression (R package *stats:step*) to perform sequential model selection. In total, we compute twelve models (four using only advice style as predictors, four using user characteristics, and four combining both) across four outcome variables: ('I **like** this style of advice', 'I would be likely to **follow** this advice', 'I found this advice **useful**', and 'A **friend** could have given this advice'). If a predictor is not mentioned for a specific model it is not included in the model following the aforementioned predictor selection process. We next report these results in a sequential manner.

### 4.1. Effects of advice style

For the first four models on *main effects of* Advice Style, we find that the Whimsical style predicts negative ratings of perceived likeability (see Table 1). Overall, these models suggest a low predictive power of advice style by itself for the four outcome measures. We visualise these predictors in Fig. 2. Furthermore, we ran a factorial ANOVA to investigate the effects of advice style on **Perceived Creepiness**. Our results show a significant main effect of advice style on Malice and Unpredictability (see Table 2). For **Malice**, Tukey's HSD for multiple comparisons revealed significantly lower ratings for WHIMSICAL ($M = 4.33$, $SD = 1.32$) than OPTIMISTIC ($M = 5.83$, $SD = 3.56$, $p < 0.05$) and SKEPTICAL ($M = 6.16$, $SD = 4.56$, $p < 0.01$). For **Unpredictability**, we see significantly higher ratings for SKEPTICAL ($M = 9.47$, $SD = 2.76$) than BALANCED ($M = 7.89$, $SD = 2.92$, $p < 0.01$) and CARING ($M = 7.61$, $SD = 2.65$, $p < 0.001$). See Fig. 3 for an overview of the distributions. We find no effects of advice styles on **Undesirability**.

### 4.2. Effects of user characteristics

We follow the same procedure as described in Section 4.1 to assess the effect of user characteristics by itself. For the four models on *main effects* of User Characteristics, we find that Agreeableness predicts positive ratings of perceived usefulness, and Negative Emotionality predicts negative ratings on advice given by a friend (see Table 3). The models show a relatively small predictive power of user characteristics for the four measures ($R^2$ between 0.11 and 0.14).

### 4.3. Effects of advice style, user characteristics, and interaction effects

For the four models on main and interaction effects, we found that the Whimsical style predicts negative ratings of perceived likeability, the likelihood of following advice, and perceived usefulness (see Table 4). We found a significant positive effect of Agreeableness on the perceived likeability, likelihood to follow, and usefulness of the recommendation. In contrast, we found a significant negative effect of Extraversion and Negative Emotionality on perceived usefulness. We also found that Insecurity predicts positive ratings of all four independent variables. We illustrate included predictors in Table 4. The adjusted $R^2$ values of the four models illustrated in Table 4 range from 0.22 ('Friend') to 0.43 ('Useful'). On average, these models were able to explain 32.3% of the variance in participants' ratings.

**Table 1**
Fitted linear models for main effects of **Advice Style** for each of the four measures, as determined through stepwise model selection.

| | Coefficients for selected predictors: | | | |
|---|---|---|---|---|
| | Like | Follow | Useful | Friend |
| Caring | 0.250 (0.398) | | 0.361 (0.363) | |
| Optimistic | 0.206 (0.398) | | 0.456 (0.363) | |
| Skeptical | −0.206 (0.393) | | −0.028 (0.358) | |
| Whimsical | −0.856[a] (0.398) | | −0.594 (0.363) | |
| Constant | 4.206[b] (0.282) | 4.185[b] (0.107) | 4.233[b] (0.256) | 4.071[b] (0.101) |
| $R^2$ | 0.106 | 0.000 | 0.108 | 0.000 |
| Adjusted $R^2$ | 0.064 | 0.000 | 0.066 | 0.000 |

[a] p<0.05.
[b] p<0.001.

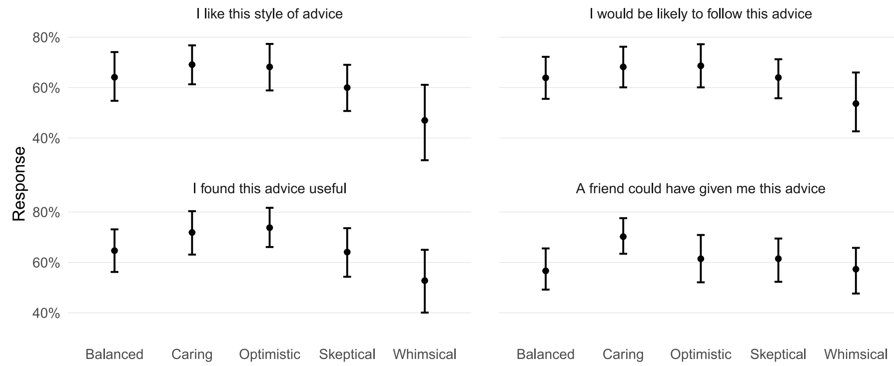


**Fig. 2.** Responses (range 1–5) for the different advice styles.

**Table 2**
Three separate two-way ANOVAs on the effects of styles on Malice, Undesirability, and Unpredictability (Perceived Creepiness).

| Measure | Df | F | *p*-value | $\eta^2$ |
|---|---|---|---|---|
| Malice | 4 | 3.672 | 0.006[a] | 0.04 |
| Undesirability | 4 | 0.431 | 0.787 | <0.01 |
| Unpredictability | 4 | 4.719 | 0.001[a] | 0.05 |

[a] $p < 0.01$.

### 4.4. Qualitative results

To better understand the underlying reasons for participants' perceptions of LLM advice, we analysed their open-ended responses. We initially familiarised ourselves with participants' responses, followed by highlighting meaningful quotes. We performed a lightweight deductive analysis across participant responses that directly correspond to the questions described in Section 3.3. In the following, we report our results and illustrate these with representative quotes of how participants responded to three questions described in Section 3.3.

#### 4.4.1. Likelihood of following digital assistant advice

Participants described themselves as likely to follow the advice when provided by a digital assistant, but that following such advice depends on different factors. For example, one participant described a tendency to follow advice that aligns with their expectations:

> "I am not concerned about where the advice comes from. I am concerned with what the advice consists of. I assess the advice to determine how relevant and helpful I think it is, and determine whether or not to act on it on this basis." (P22)

As it was seemingly important for participants that the advice aligned with their expectations, one participant described a hesitancy, harnessing parts of the digital assistant advice seen as useful:

> "I would use the advice from a digital assistant as a guide. Then I'll follow the parts of the advice that I felt applied to me and not follow the parts that I didn't think would apply to me." (P66)

However, participants also described other aspects that influence their likelihood of following the advice—where one participant mentioned the role of what others think of the advice, believing that AI echoes what is on the internet.

> "A digital assistant probably picks up advice off the internet (like AI) so it probably just repeats what other people have written about. So I probably follow if that's what others are saying." (P79)

#### 4.4.2. Receptiveness of digital assistant advice

In regards to the receptiveness of advice provided by the digital assistant, participants described themselves as rather receptive to advice provided by a digital assistant. One participant described an increased receptiveness due to the anonymous nature of interacting with a computer rather than a family member:

> "I think that I would be more receptive to advice from a digital assistant because it would have an impartial position within my life, compared to friends and family from whom I may feel a level of judgement. If I fail to take on all the advice from a digital assistant, it will not berate or criticise me for it; a human might." (P72)

A different participant described the receptiveness of digital advice as indifferent to human advice—viewing digital advice as a tool for funnelling the most relevant content. However, they also highlighted a decrease in receptiveness if the advice content displays bias:

> "It doesn't alter it too much as it still makes me think that it's advice I could find online by myself—digital assistants make it more compiled and easy to find. However, when it feels biased, then I'm more hesitant to follow the advice." (P39)

However, participants were also critical, expressing their general resistance to advice from digital assistants:

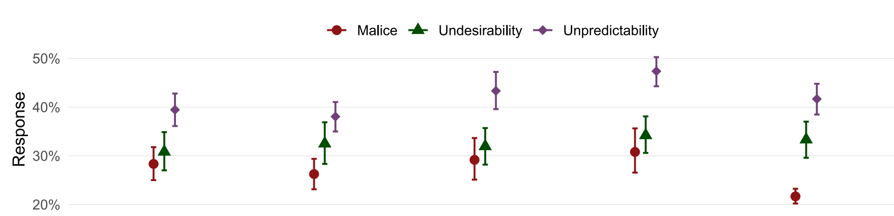


**Fig. 3.** Malice, Undesirability, and Unpredictability (Perceived Creepiness) ratings (range 1–21) for the different advice styles.

**Table 3**
Fitted linear models for main effects of **User Characteristics** for each of the four measures, as determined through stepwise model selection.

| | Coefficients for selected predictors: | | | |
|---|---|---|---|---|
| | Like | Follow | Useful | Friend |
| Agreeableness | 0.243 (0.132) | 0.210 (0.107) | 0.323[b] (0.116) | 0.165 (0.102) |
| Negative Emotionality | −0.194 (0.130) | −0.199 (0.106) | | −0.261[a] (0.102) |
| Open Mindedness | 0.210 (0.124) | 0.198 (0.102) | | |
| Optimism | 0.208 (0.137) | | 0.272[a] (0.122) | |
| Discomfort | | | | 0.177 (0.106) |
| Insecurity | 0.215 (0.131) | 0.192 (0.104) | 0.212 (0.117) | 0.202 (0.103) |
| Constant | 4.084[c] (0.121) | 4.185[c] (0.100) | 4.271[c] (0.109) | 4.071[c] (0.095) |
| $R^2$ | 0.175 | 0.159 | 0.170 | 0.149 |
| Adjusted $R^2$ | 0.126 | 0.120 | 0.141 | 0.110 |

[a] $p<0.05$.
[b] $p<0.01$.
[c] $p<0.001$.

"It feels very one size fits all, but to make it that way it has to be fairly generic, which I don't find very helpful. It's the sort of advice I could write myself without really putting much mind to it." (P14)

#### *4.4.3. Advice more likely to seek from digital assistant*

Lastly, concerning what problems or questions participants are more likely to seek advice on from a digital assistant, participants described that they were likely to use digital assistants on more delicate problems or questions:

"Anything private or embarrassing. Things about my sexual identity, insecurities and shortcomings." (P2)

"I think personal questions or ones that I don't want my friends and family members to know about (weight loss, etc.)." (P39)

In addition to the inclination to seek guidance on sensitive issues, advice that is more pragmatic was also noted for its usefulness. One participant highlighted this, contrasting it with human advice, and how human advice might also fall short in terms of emotional depth:

"I'm more likely to reach out to a digital assistant for practical advice and look to friends/family for advice relating to emotions. However, I noticed that the digital assistant in this study was very useful for advice relating to emotions surrounding human relationships. I think it was more helpful than the humans I know—but we don't always talk about emotional issues because we want advice. We often talk about them because we just need to 'unload' them and feel that someone is standing with us on our side." (P72)

In summary, participants describe a tendency to be receptive to and likely to follow the advice provided by digital assistants. However, participants also describe more skeptical viewpoints. Moreover, participants describe that they are more likely to seek more sensitive and practical advice from digital assistants.

## 5. Discussion

Perceptions of advice are complex and multi-dimensional, with factors such as individual preference, context, and advice content playing a role in the recipients' perceptions. In our exploratory study, we find that the interactions between different factors particularly determine these perceptions—we thus interpret those models that include all predictors (see Table 4). We find that advice generated by an LLM with a whimsical style resulted in lower user ratings as compared to balanced, caring, optimistic, and skeptical instructions. Furthermore, we found that our participants with high levels of agreeableness liked the advice more, were more likely to follow it, and considered it more useful. In terms of technological readiness, we find a positive effect of technological insecurity on advice ratings. Moreover, we see that whimsical LLM-generated advice elicits significantly lower ratings of implied malice and that caring and balanced LLM-generated advice elicits significantly lower ratings of unpredictability compared to skeptical LLM-generated advice. Our qualitative results suggest that one of the primary perceived uses for advice from LLMs is support on sensitive topics and practical matters.

### *5.1. The role of user characteristics in LLM advice perceptions*

Prior work has evaluated the role of user characteristics in the context of humans giving and receiving advice. For example, people with higher emotionality ratings (i.e., incidental anger) show lower receptiveness to advice (Gino & Schweitzer, 2008), while Feng and MacGeorge showed how an individual's expressivity influences receptiveness to advice (Feng & MacGeorge, 2006).

Our participants' agreeableness levels influenced their advice preferences, with those who were more agreeable being more likely to assess the presented advice more positively. Our results align with Yan and Chen, which recently showed that open-mindedness, conscientiousness, and agreeableness significantly influence participants' ratings of recommender system outputs (Yan & Chen, 2023), similar to Kunzler et al. (Künzler et al., 2020). We extend the relation between higher agreeableness ratings and increased compliance, indicating that participants with higher agreeableness ratings also show increased tendencies to like and follow LLM-generated advice. While our study did not explicitly focus on *why* those with higher agreeableness ratings rate LLM-generated advice as more likeable and likely to follow, this can partially be explained by the higher cooperative nature of these individuals (Graziano & Eisenberg, 1997). A different interpretation can be made by looking at research by De Vries et al. suggesting that individuals with higher agreeableness ratings found six out of ten types of motivational messages more motivating (de Vries et al., 2017). As we observed similar effects, those effects can be related to advice oftentimes having a motivational tone (e.g., optimistic advice), consequently influencing people's advice receptivity. Furthermore, prior research has attempted to pair user characteristics with text messages. Völkel et al. matched the chatbot's agreeableness to the user's agreeableness level, suggesting that people with higher agreeableness ratings preferred the agreeableness-imbued chatbot more, but did not show that more disagreeable individuals preferred the disagreeable chatbots (Theres

**Table 4**
Fitted linear models for each of the four measures, as determined through stepwise model selection. We use the Balanced Advice Style as the reference level.

| | *Coefficients for selected predictors:* | | | |
|---|---|---|---|---|
| | Like | Follow | Useful | Friend |
| Extraversion | | | −0.636[b] (0.209) | 0.172 (0.110) |
| Agreeableness | 0.411[b] (0.130) | 0.278[b] (0.104) | 0.443[c] (0.112) | |
| Conscientiousness | | | | 0.053 (0.098) |
| Negative Emotionality | −0.137 (0.119) | −0.176 (0.103) | −0.261[a] (0.116) | −0.229[a] (0.114) |
| Open-Mindedness | 0.123 (0.121) | −0.156 (0.157) | 0.126 (0.111) | |
| Optimism | −0.735 (0.382) | 0.127 (0.110) | 0.221 (0.117) | |
| Discomfort | −0.199 (0.123) | | −0.162 (0.108) | |
| Insecurity | 0.357[b] (0.124) | 0.280[b] (0.103) | 0.337[b] (0.109) | 0.234[a] (0.094) |
| Caring (style) | 0.399 (0.354) | 0.160 (0.312) | 0.403 (0.328) | |
| Optimistic (style) | 0.375 (0.356) | 0.252 (0.316) | 0.494 (0.313) | |
| Skeptical (style) | −0.166 (0.337) | −0.130 (0.297) | −0.082 (0.309) | |
| Whimsical (style) | −1.032[b] (0.355) | −0.675[a] (0.298) | −0.887[b] (0.318) | |
| Extraversion:Caring (style) | | | −0.019 (0.370) | |
| Extraversion:Optimistic (style) | | | 0.510 (0.290) | |
| Extraversion:Skeptical (style) | | | 1.389[c] (0.385) | |
| Extraversion:Whimsical (style) | | | 0.660[b] (0.245) | |
| Optimism:Insecurity | 0.306[a] (0.130) | | 0.155 (0.101) | |
| Negative Emotionality:Discomfort | 0.354[b] (0.132) | | 0.271[a] (0.108) | |
| Negative Emotionality:Optimism | −0.328[a] (0.136) | | −0.350[b] (0.117) | |
| Open-Mindedness:Caring (style) | | 0.665[a] (0.328) | | |
| Open-Mindedness:Optimistic (style) | | 0.318 (0.297) | | |
| Open-Mindedness:Skeptical (style) | | 0.363 (0.304) | | |
| Open-Mindedness:Whimsical (style) | | 0.764[b] (0.273) | | |
| Open-Mindedness:Optimism | 0.390[b] (0.130) | 0.216 (0.111) | 0.302* (0.115) | |
| Optimism:Caring (style) | 0.394 (0.458) | | | |
| Optimism:Optimistic (style) | 0.978[a] (0.469) | | | |
| Optimism:Skeptical (style) | 1.026[a] (0.440) | | | |
| Optimism:Whimsical (style) | 1.263[b] (0.423) | | | |
| Negative Emotionality:Open-Mindedness | 0.174 (0.123) | 0.145 (0.108) | | |
| Negative Emotionality:Insecurity | 0.194 (0.153) | | | |
| Extraversion:Open-Mindedness | | | −0.164 (0.101) | |
| Conscientiousness:Negative Emotionality | | | | −0.420[c] (0.097) |
| Extraversion:Negative Emotionality | | | | 0.201[a] (0.094) |
| Extraversion:Insecurity | | | | 0.133 (0.094) |
| Constant | 4.008[c] (0.252) | 4.262[c] (0.216) | 4.233[c] (0.239) | 4.025[c] (0.102) |
| $R^2$ | 0.516 | 0.384 | 0.553 | 0.284 |
| Adjusted $R^2$ | 0.378 | 0.261 | 0.425 | 0.224 |

[a] $p<0.05$.
[b] $p<0.01$.
[c] $p<0.001$.

Völkel & Kaya, 2021). As prior works highlight the impact of personality (Theres Völkel & Kaya, 2021) and task-fitting (Hertz & Wiese, 2019) on receptivity towards advice, our results show that technological readiness also plays a major role in this (see Section 4.2).

With the recent availability of LLM-driven interfaces, people's technological readiness can play a large role in their perception and adoption of this new technology (Lin & Hsieh, 2006). Our results suggest that those with higher insecurity ratings towards technology perceived the LLM-generated advice as more useful, liked it more, and were likelier to follow it. This has implications for how people might engage with LLM-generated advice, as higher insecurity related to AI literacy (i.e., people's understanding of AI (Long & Magerko, 2020)) could generate unrealistic expectations as LLM-generated advice may be incorrect (Bender et al., 2021; Oviedo-Trespalacios et al., 2023). People with low AI literacy may also have a hard time distinguishing between human or LLM output (Jakesch, Hancock, & Naaman, 2023) and LLM advice may even invoke dishonesty (Leib et al., 2023). The combination of high insecurity and low AI literacy may cause less efficient or even harmful interactions for a user, particularly when people seek social support from LLMs (Skjuve et al., 2023), considering that unknowingly treating unsound LLM-generated advice positively is especially harmful in such situations.

Considering the diverging needs that users might have in using LLMs for advice seeking, both our design takeaways ultimately connect to the idea of considering and designing for user personalisation. However, we envision what personalisation explicitly entails (both in general but specifically LLM-generated advice) being in the hands of the user, moving away from design assumptions that result in mismatches between user expectations and designer intentions. This idea is further supported by Suwanaposee et al., who recently showed the Barnum effect (a psychological phenomenon that makes people have more positive perceptions of digital output if framed as tailored 'specifically for you') to have little to no positive effects on user perception of (e.g., AI) recommendations (Suwanaposee et al., 2023). In the following, we outline two design takeaways for designing LLM-generated advice with an increased focus on designing for users' needs that can positively influence their use of LLMs when seeking advice.

### 5.2. Takeaways for designing LLM-generated advice

Amidst the increasing use of technology in personal settings, user perception of a technology's creepiness is receiving increased attention.

For example, Yip et al. studied how children perceive technology as creepy, suggesting that unpredictability, among other terms, is used when describing technology as creepy (Yip et al., 2019). In psychology, unpredictability has also shown to be a relevant predictor of creepiness (McAndrew & Koehnke, 2016). As such, we hypothesise that participants' increased perception of the creepiness of advice written using a skeptical style can be explained by their lack of expectation that such advice would be provided skeptically. Decreasing unpredictability (and ideally decreasing creepiness) evoked by LLM-generated advice may help avoid conversational breakdowns (Lai-Chong Law et al., 2022), which may be particularly important when users seek support on sensitive topics. As our results suggest that users favour LLM-generated advice depending on the styles, we suggest **increasing user control in determining styles of LLM advice**. This would give users increased control, with an increased chance of avoiding unpredictability that might follow given advice in a skeptical style, and choose an advice style that pertains to their needs. More concretely, this could be realised in user interfaces by providing users with a design that supports and offers not one but many ways in which LLMs can provide them with advice—contrasting Bing, which only allows users to choose between three styles (Edwards, 2023). Our results indicate that people overall rate the LLM-generated advice given in a skeptical or whimsical style as less positive (see, for example, Fig. 3). However, others might appreciate such advice—depending on the context within which they seek advice. Giving those with diverging needs the opportunity to determine the style of support they receive, whether or not that deviates from these perceptions evident from our results, is a promising way forward.

A majority of participants also indicated that they were receptive to LLM-generated advice. One reason for seeking LLM-generated advice might be due to more impartial and advisee-centred communication. This aligns with recent research on online advice communities, where Tomprou et al. investigated how people utilise such platforms, indicating that anonymity, access to multiple experts' advice, and just-in-time advice are perceived as positive factors (Tomprou et al., 2019). Skjuve et al. investigated peoples' perception of Replika, a social companion chatbot designed to converse with users through open-ended dialogue—suggesting that users appreciated the caring and non-judgmental tone Replika elicited (Skjuve et al., 2021). Furthermore, we see that whimsical advice received the lowest ratings on perceived malice, indicating an interesting duality as the whimsical was the least favoured advice style. As participants described the promise of using digital assistants for sensitive topics—designing to decrease perceived malice might be particularly relevant, as similarly suggested by Wester et al. (Wester, Pohl, et al., 2024). Given our results and the recent work on individuals' preferences regarding online support, our recommendation is, therefore, to **tailor LLM advice to meet individual user expectations**. One important aspect of tailoring such advice is to consider the context in which it is given. For example, LLM-generated advice on health-related topics might naturally be expected to be more careful than advice on what car dealership can best support you with your car-serving needs. However, this does not imply that all people want health-related advice to be provided in more careful ways. This can be exemplified by parental caregivers of autistic children describing a lack of 'open-mindedness' from human professionals or service providers (Mackintosh et al., 2012)—a feeling that might emerge if 'bad' advice is provided following the failure of meeting people's expectations. More concretely, LLM-powered applications could not only adapt to user characteristics but also to the user's situation in general—ultimately providing a closer alignment with user needs.

### 5.3. Limitations and future work

We acknowledge several limitations in our work. First, following trial and error, we iteratively assessed and compared the different styles with the generated output, consequently generating distinct LLM-generated styles of advice to be included in our study design. However, we acknowledge that people's perceptions of the included advice styles, such as WHIMSICAL, might be influenced by what advice they normally perceive as whimsical. As prompt engineering and design are receiving increased attention (e.g., Microsoft Guidance[3]), more systematic approaches to generate and assess LLM output (e.g., advice) are interesting aspects for future work.

Second, participants rated different advice presented to them in a vignette format (see Fig. 1). One positive aspect of using vignettes is that we enable systematic comparisons, i.e. all participants see the same advice for each chosen topic. However, we stress the importance of future interactive studies that adapt to participants' advice-seeking needs to increase ecological validity. We would expect additional effects if the advice was provided interactively (e.g., the advice provided in a human-chatbot interaction) and in a real-world context. We can assume that people will seek advice for various reasons and that the eight potential advice topics included in our study exclude any potential personal advice topics. However, we do include four overarching advice topics (personal health, relationships, climate change, and career) followed by broad and specific variants of those. Moreover, we assess participant ratings as an average across the eight topic scenarios.

Future work around deployments of interactive LLM-powered applications providing advice to everyday citizens would benefit from extending our exploratory work through longitudinal and in-situ investigations to increase ecological validity and capture people's behaviour when facing such advice. Furthermore, future work on LLMs in supportive roles must acknowledge the diverse needs and expectations of end-users, as for example driven by user characteristics. Consequently, we argue for an increased focus on end-user involvement in the design processes of interactive support tools to better meet user expectations across a variety of contexts.

## 6. Conclusion

User characteristics play a significant role in determining how people perceive LLM-generated advice, particularly when these are delivered in different styles. As people may turn to LLMs to seek advice, we need to better understand the effects of diverse user characteristics on perceptions of such advice. In this paper, we explored the influence of user characteristics and styles of LLM-generated advice on user perceptions. Our results suggest that those with higher ratings of agreeableness and insecurity positively predict more positive advice ratings, and that a whimsical advice style has a negative effect on advice ratings. Regarding perceived creepiness, we observe lower advice ratings of whimsical on perceived malice, and higher ratings of unpredictability on skeptical styles of advice. Our results contribute to a better understanding of the role of user characteristics in perceiving LLM-generated advice and how the advice's design impacts ratings of perceived creepiness. We thereby inform the design of LLM-powered support applications targeting people with diverse expectations and needs.

## CRediT authorship contribution statement

**Joel Wester:** Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Investigation, Conceptualization. **Sander de Jong:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Conceptualization. **Henning Pohl:** Writing – review & editing, Visualization, Supervision, Methodology, Investigation, Conceptualization. **Niels van Berkel:** Writing – review & editing, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization.

[3] https://github.com/microsoft/guidance.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

LLM-generated advice (Original data) (OSF)

## Acknowledgments

This work is supported by the Carlsberg Foundation, grant CF21-0159.

## A Prompt Template

*System prompt*

"Act as a person giving advice. You give advice to a general audience.

You are **[style description]**. You will only answer as very **[style description]**, and nothing else. Your behaviour of being **[style description]** is highly deliberate and shines through in your answers.

You provide 150-word advice columns in response to questions posed to you by the readers. You write advice that gives the person suggestions how to best answer their questions. You write advice only by using the English alphabet, and nothing else. You format your columns using Markdown. You do not present advice in listicle format. Remove any first-paragraph pretext and concluding post-text."

*User prompt*

"**[topic question]**"

Where the template variables are defined as:

**style = balanced:** "balanced and neutral"
**style = whimsical:** "whimsical and witty, and act like a person or idea that is unusual, playful, and unpredictable, rather than serious and practical"
**style = caring:** "caring and kind, and act like a person concerned when others are upset and uplifted when others are happy"
**style = skeptical:** "skeptical and doubtful, and act like a person that has the attitude of doubt or a disposition to incredulity either in general or toward a particular object"
**style = optimistic:** "optimistic and positive, and act like a person that see the positive side of things, expect things to turn out well, and believe that you have the skill and ability to make good things happen"

**topic = personal health (specific):** "How can I lose a few pounds?"
**topic = personal health (broad):** "How can I improve my physical health?"
**topic = relationships (specific):** "How can I improve the relationship with my mother?"
**topic = relationships (broad):** "How can I become more likeable?"
**topic = climate change (specific):** "What can I do to reduce my energy consumption at home?"
**topic = climate change (broad):** "What can I do about climate change?"
**topic = career (specific):** "How to get a raise?"
**topic = career (broad):** "How to achieve good work-life balance?"

## References

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can Language Models Be too Big?. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency (virtual event, Canada) (FAccT '21)* (pp. 610–623). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922.

Biele, G., Rieskamp, J., Krugel, L. K., & Heekeren, H. R. (2011). The neural basis of following advice. *PLoS Biology, 9*(6), 1–11. https://doi.org/10.1371/journal.pbio.1001089, 2011.

Birkun, A. A., & Gautam, A. (2023). Dr. Google's advice on first aid: Evaluation of the search engine's question-answering system responses to queries seeking help in health emergencies. *Prehospital and Disaster Medicine*, 1–7. https://doi.org/10.1017/S1049023X23000511, 2023.

Cai, W., Jin, Y., & Chen, L. (2022). Impacts of personal characteristics on user trust in conversational recommender systems. In *Proceedings of the 2022 CHI conference on human factors in computing systems* (p. 14). New Orleans, LA, USA: CHI '22. https://doi.org/10.1145/3491102.3517471. Article 489.

Caron, G., & Srivastava, S. (2022). Identifying and manipulating the personality traits of Language Models. *arXiv*. https://doi.org/10.48550/arXiv.2212.10276, 2022.

Cerekovic, A., Aran, O., & Gatica-Perez, D. (2017). Rapport with virtual agents: What do human social cues and personality explain? *IEEE Transactions on Affective Computing, 8*(3), 382–395. https://doi.org/10.1109/TAFFC.2016.2545650, 2017.

Chen, M.-F., & Lin, N.-P. (2018). Incorporation of health consciousness into the technology readiness and acceptance model to predict app download and usage intentions. *Internet Research, 28*(2), 351–373. https://doi.org/10.1108/IntR-03-2017-0099, 2018.

Chen, L., Wu, W., & He, L. (2013). How personality influences users' needs for recommendation diversity?. In *CHI '13 extended abstracts on human factors in computing systems (paris, France) (CHI EA '13)* (pp. 829–834). https://doi.org/10.1145/2468356.2468505

Chiu, W., & Cho, H. (2021). The role of technology readiness in individuals' intention to use health and fitness applications: A comparison between users and non-users. *Asia Pacific Journal of Marketing and Logistics, 33*(3), 807–825. https://doi.org/10.1108/APJML-09-2019-0534, 2021.

Ciechanowski, L., Przegalinska, A., Magnuski, M., & Gloor, P. (2019). In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Generation Computer Systems, 92*, 539–548. https://doi.org/10.1016/j.future.2018.01.055, 2019.

de Vries, R. A. J., Truong, K. P., Zaga, C., Li, J., & Evers, V. (2017). A word of advice: How to tailor motivational text messages based on behavior change theory to personality and gender. *Personal and Ubiquitous Computing, 21*(4), 675–687. https://doi.org/10.1007/s00779-017-1025-1, 2017.

Edwards, B. (2023). AI-powered Bing Chat gains three distinct personalities. *Ars Technica*, 2023 https://arstechnica.com/information-technology/2023/03/microsoft-equips-bing-chat-with-multiple-personalities-creative-balanced-precise/.

Efendić, E., Calseyde, P. Van de, Bahník, S., & Vranka, M. (2023). Taking algorithmic (Vs. Human) advice reveals different goals to others. *International Journal of Human-Computer Interaction*, 1–10. https://doi.org/10.1080/10447318.2023.2210886, 2023.

Feng, B., & MacGeorge, E. L. (2006). Predicting receptiveness to advice: Characteristics of the problem, the advice-giver, and the recipient. *Southern Communication Journal, 71*(1), 67–85. https://doi.org/10.1080/10417940500503548, 2006.

Gino, F., & Schweitzer, M. E. (2008). Blinded by anger or feeling the love: How emotions influence advice taking. *Journal of Applied Psychology, 93*(5), 1165–1173. https://doi.org/10.1037/0021-9010.93.5.1165, 2008.

Goodyear, K., Raja, P., Chernyak, S., Madhavan, P., Deshpande, G., & Krueger, F. (2016). Advice taking from humans and machines: An fMRI and effective connectivity study. *Frontiers in Human Neuroscience, 10*. https://doi.org/10.3389/fnhum.2016.00542, 2016.

Graziano, W. G., & Eisenberg, N. (1997). Chapter 30 - agreeableness: A dimension of personality. In *Handbook of personality psychology* (pp. 795–824). https://doi.org/10.1016/B978-012134645-4/50031-7

Harvey, N., & Fischer, I. (1997). Taking advice: Accepting help, improving judgment, and sharing responsibility. *Organizational Behavior and Human Decision Processes, 70*(2), 117–133. https://doi.org/10.1006/obhd.1997.2697, 1997.

Hertz, N., & Wiese, E. (2019). Good advice is beyond all price, but what if it comes from a machine? *Journal of Experimental Psychology: Applied, 25*(3), 386–395. https://doi.org/10.1037/xap0000205, 2019.

Howe, P. D. L., Fay, N., Morgan, S., & Hovy, E. (2023). ChatGPT's advice is perceived as better than that of professional advice columnists. *Frontiers in Psychology, 14*. https://doi.org/10.3389/fpsyg.2023.1281255, 2023.

Jakesch, M., Bhat, A., Buschek, D., Zalmanson, L., & Naaman, M. (2023a). Co-writing with opinionated Language Models affects users' views. In *Proceedings of the 2023 CHI conference on human factors in computing systems* (p. 15). Hamburg, Germany: CHI '23). https://doi.org/10.1145/3544548.3581196. Article 111.

Jakesch, M., Hancock, J. T., & Naaman, M. (2023b). Human heuristics for AI-generated language are flawed. *Proceedings of the National Academy of Sciences, 120*, 11. https://doi.org/10.1073/pnas.2208839120, 2023.

Jiang, G., Xu, M., Zhu, S.-C., Han, W., Zhang, C., & Zhu, Y. (2022). Mpi: Evaluating and inducing personality in pre-trained Language Models. https://doi.org/10.48550/ARXIV.2206.07550.

Jo, E., Epstein, D. A., Jung, H., & Kim, Y.-H. (2023). Understanding the benefits and challenges of deploying conversational AI leveraging Large Language Models for public health intervention. In *Proceedings of the 2023 CHI conference on human factors in computing systems* (p. 16). Hamburg, Germany: CHI '23). https://doi.org/10.1145/3544548.3581503. Article 18.

Kang, S., Young Park, C., Kim, A., Cha, N., & Lee, U. (2022). Understanding emotion changes in mobile experience sampling. In *Proceedings of the 2022 CHI conference on human factors in computing systems* (p. 14). New Orleans, LA, USA: CHI '22. https://doi.org/10.1145/3491102.3501944. Article 198.

Kim, S., Chow, B. C., Park, S., & Liu, H. (2023). The usage of digital health technology among older adults in Hong Kong and the role of technology readiness and eHealth literacy: Path analysis. *Journal of Medical Internet Research, 25*. https://doi.org/10.2196/41915, 2023.

Kim, Y., Lee, J., Kim, S., Park, J., & Kim, J. (2024). Understanding users' dissatisfaction with ChatGPT responses: Types, resolving tactics, and the effect of knowledge level. In *Proceedings of the 29th international conference on intelligent user interfaces (Greenville, SC, USA) (IUI '24)* (pp. 385–404). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3640543.3645148.

Künzler, F., Mishra, V., Kramer, J.-N., Kotz, D., Fleisch, E., & Kowatsch, T. (2020). Exploring the state-of-receptivity for MHealth interventions. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., 3*, 4. https://doi.org/10.1145/3369805, 2020.

Lai-Chong Law, E., Følstad, A., & Van As, N. (2022). Effects of humanlikeness and conversational breakdown on trust in chatbots for customer service. In *Nordic human-computer interaction conference (Aarhus, Denmark) (NordiCHI '22)* (p. 13). https://doi.org/10.1145/3546155.3546665. Article 56.

Lee, Y.-C., Yamashita, N., Huang, Y., & Fu, W. (2020). I hear you, I feel you": Encouraging deep self-disclosure through a chatbot. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1–12). Honolulu, HI, USA: CHI '20). https://doi.org/10.1145/3313831.3376175.

Leib, M., Köbis, N., Rilke, R. M., Hagens, M., & Irlenbusch, B. (2023). Corrupted by algorithms? How AI-generated and human-written advice shape (Dis)honesty. *The Economic Journal, 134*(658), 766–784. https://doi.org/10.1093/ej/uead056, 09 2023.

Li, J., & Chignell, M. (2010). Birds of a feather: How personality influences blog writing and reading. *International Journal of Human-Computer Studies, 68*(9), 589–602. https://doi.org/10.1016/j.ijhcs.2010.04.001, 2010.

Li, L., Zhang, Y., & Chen, L. (2023). Personalized prompt learning for explainable recommendation. *ACM Transactions on Information Systems, 41*(4), 26. https://doi.org/10.1145/3580488. Article 103 (2023.

Lin, J.-S. C., & Chang, H.-C. (2011). The role of technology readiness in self-service technology acceptance. *Managing Service Quality: International Journal, 21*(4), 424–444. https://doi.org/10.1108/09604521111146289, 2011.

Lin, J.-S. C., & Hsieh, P.-ling (2006). The role of technology readiness in customers' perception and adoption of self-service technologies. *International Journal of Service Industry Management, 17*(5), 497–517. https://doi.org/10.1108/09564230610689795, 2006.

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1–16). Honolulu, HI, USA: CHI '20. https://doi.org/10.1145/3313831.3376727.

Lyn, M. van S., & Sniezek, J. A. (2005). Factors affecting the acceptance of expert advice. *British Journal of Social Psychology, 44*(3), 443–461. https://doi.org/10.1348/014466604X17092, 2005.

Mackintosh, V. H., Goin-Kochel, R. P., & Myers, B. J. (2012). "What do you like/dislike about the treatments you're currently using?": A qualitative study of parents of children with autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities, 27*(1), 51–60. https://doi.org/10.1177/1088357611423542, 2012.

McAndrew, F. T., & Koehnke, S. S. (2016). On the nature of creepiness. *New Ideas in Psychology, 43*, 10–15. https://doi.org/10.1016/j.newideapsych.2016.03.003, 2016.

Metz, B., & Wörle, J. (2021). *Technology readiness affects the acceptance of and intention to use an automated driving system after repeated experience*. https://doi.org/10.1145/3409118.3475156

Nishith Sharan, N., & Romano, D. M. (2020). The effects of personality and locus of control on trust in humans versus artificial intelligence. *Heliyon, 6*, Article e04572. https://doi.org/10.1016/j.heliyon.2020.e04572, 2020.

Nov, O., & Su, H. (2018). Eliciting users' demand for interface features. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1–8). Montreal QC, Canada) (CHI '18). https://doi.org/10.1145/3173574.3173879.

Oviedo-Trespalacios, O., Peden, A. E., Cole-Hunter, T., Costantini, A., Haghani, M., Rod, J. E., Kelly, S., Torkamaan, H., Tariq, A., Newton, J. D. A., Gallagher, T., Steinert, S., Filtness, A., & Reniers, G. (2023). *The risks of using ChatGPT to obtain common safety-related information and advice*. https://doi.org/10.2139/ssrn.4346827. *prepr*int (2023.

Parasuraman, A., & Colby, C. L. (2015). An updated and streamlined technology readiness Index: Tri 2.0. *Journal of Service Research, 18*(1), 59–74. https://doi.org/10.1177/1094670514539730, 2015.

Pareek, S., van Berkel, N., Velloso, E., & Goncalves, J. (2024). Effect of explanation conceptualisations on trust in AI-assisted credibility assessment. *Proceedings of the ACM on Human-Computer Interaction, 8*, 1–31. CSCW2 (2024.

Reddy Karra, S., Nguyen, S., & Tulabandhula, T. (2022). AI personification: Estimating the personality of Language Models, 2022 https://doi.org/10.48550/arXiv.2204.12000.

Ruane, E., Farrell, S., & Ventresque, A. (2020). User perception of text-based chatbot personality. In *Chatbot research and design: 4th international workshop, CONVERSATIONS 2020, virtual event, november 23–24, 2020, revised selected papers* (pp. 32–47). https://doi.org/10.1007/978-3-030-68288-03

Safdari, M., Serapio-García, G., Crepy, C., Fitz, S., Romero, P., Sun, L., Abdulhai, M., Faust, A., & Matarić, M. (2023). *Personality traits in Large Language Models*. https://doi.org/10.48550/arXiv.2307.00184, 2023.

Shaw, C., & Hepburn, A. (2013). Managing the moral implications of advice in informal interaction. *Research on Language and Social Interaction, 46*(4), 344–362. https://doi.org/10.1080/08351813.2013.839095, 2013.

Skjuve, M., Følstad, A., & Brandtzaeg, P. B. (2023). The user experience of ChatGPT: Findings from a questionnaire study of early users. In *Proceedings of the 5th international conference on conversational user interfaces (Eindhoven, Netherlands) (CUI '23)*. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3571884.3597144. Article 2, 10 pages.

Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2021). My chatbot companion - a study of human-chatbot relationships. *International Journal of Human-Computer Studies, 149*(2021), Article 102601. https://doi.org/10.1016/j.ijhcs.2021.102601

Skjuve, M., Haugstveit, I. M., Følstad, A., & Brandtzaeg, P. (2019). Help! Is my chatbot falling into the uncanny valley? An empirical study of user experience in human–chatbot interaction. *Human Technology, 15*(1), 30–54. https://doi.org/10.17011/ht/urn.201902201607. Feb. 2019.

Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology, 113*(2017), 117–143. https://doi.org/10.1037/pspp0000096

Suwanaposee, P., Gutwin, C., Chen, Z., & Cockburn, A. (2023). 'Specially for you' – examining the Barnum Effect's influence on the perceived quality of system recommendations. In *Proceedings of the 2023 CHI conference on human factors in computing systems (Hamburg, Germany) (CHI '23)*. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3544548.3580656. Article 376, 11 pages.

Theres Völkel, S., & Kaya, L. (2021). Examining user preference for agreeableness in chatbots. In *Proceedings of the 3rd Conference on conversational user interfaces (bilbao (online), Spain) (CUI '21)*. https://doi.org/10.1145/3469595.3469633. Article 38, 6 pages.

Theres Völkel, S., Schoedel, R., Kaya, L., & Mayer, S. (2022). User perceptions of extraversion in chatbots after repeated use. In *Proceedings of the 2022 CHI conference on human factors in computing systems* (p. 18). New Orleans, LA, USA: CHI '22). https://doi.org/10.1145/3491102.3502058. Article 253.

Tomprou, M., Dabbish, L., Kraut, R. E., & Liu, F. (2019). Career mentoring in online communities: Seeking and receiving advice from an online community. In *Proceedings of the 2019 CHI conference on human factors in computing systems (Glasgow, Scotland UK) (CHI '19)* (pp. 1–12). https://doi.org/10.1145/3290605.3300883

von der Pütten, A. M., Krämer, N. C., & Gratch, J. (2010). How our personality shapes our interactions with virtual characters - implications for research and development. In *Intelligent virtual agents* (pp. 208–221). https://doi.org/10.1007/978-3-642-15892-623

Wang, X., Li, L., Tan, S. C., Yang, L., & Lei, J. (2023). Preparing for AI-enhanced education: Conceptualizing and empirically examining teachers' AI readiness. *Computers in Human Behavior, 146*(2023), Article 107798. https://doi.org/10.1016/j.chb.2023.107798

Wærn, Y., & Ramberg, R. (1996). People's perception of human and computer advice. *Computers in Human Behavior, 12*(1), 17–27. https://doi.org/10.1016/0747-5632(95)00016-X, 1996.

Wee, C. C., & Cornell, J. E. (2023). Conversations about obesity and weight: Good news we can use. *Annals of Internal Medicine, 0*. https://doi.org/10.7326/M23-2568, 2023.

Wester, J., Pohl, H., Hosio, S., & Berkel, N.van (2024a). This chatbot would never…": Perceived moral agency of mental health chatbots. *Proc. ACM Hum.-Comput. Interact., 8*, 28. https://doi.org/10.1145/3637410. CSCW1, Article 133 (apr 2024).

Wester, J., Schrills, T., Pohl, H., & Berkel, N.van (2024b). "As an AI language model, I cannot": Investigating LLM denials of user requests. In *Proceedings of the CHI conference on human factors in computing systems (Honolulu, HI, USA) (CHI '24)*. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3613904.3642135. Article 979, 14 pages.

Woźniak, P. W., Karolus, J., Lang, F., Eckerth, C., Schöning, J., Rogers, Y., & Niess, J. (2021). Creepy technology:what is it and how do you measure it?. In *Proceedings of the 2021 CHI conference on human factors in computing systems* (p. 13). Yokohama, Japan: CHI '21. https://doi.org/10.1145/3411764.3445299. Article 719.

Yan, D., & Chen, L. (2023). The influence of personality traits on user interaction with recommendation interfaces. *ACM Transactions on Interactive Intelligent Systems, 13*(1), 39. https://doi.org/10.1145/3558772. Article 3 (2023.

Yip, J. C., Sobel, K., Gao, X., Allison, M. H., Lim, A., Meng, L., Ofiana, R. F., Park, J., & Hiniker, A. (2019). Laughing is scary, but farting is cute: A conceptual model of children's perspectives of creepy technologies. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1–15). Glasgow, Scotland Uk: CHI '19. https://doi.org/10.1145/3290605.3300303.