Avestia Publishing
Journal of Machine Intelligence and Data Science (JMIDS)
Volume 6, Year 2025
ISSN: 2564-3282
DOI: 10.11159/jmids.2025.008

Prompt Assessment for Human-AI Interaction: Intent, Complexity & Lay Perceptions

Marijose Páez Velázquez [1,3], Elzbieta Bobrowicz-Campos [1,2], Patrícia Arriaga [1]
[1] ISCTE-University Institute of Lisbon, School of Social Sciences, Centre for Psychological Research and Social Intervention (CIS), Department of Psychology
Av. das Forças Armadas, Lisbon, Portugal, 1649-026
mpvzz@iscte-iul.pt; elzbieta.campos@iscte-iul.pt; patricia.arriaga@iscte-iul.pt
[2] University of Coimbra, Faculty of Psychology and Education Sciences
Rua do Colégio Novo, Coimbra, Portugal, 3000-115
elzbieta@fpce.uc.pt
[3] University of Lisbon, School of Social and Political Sciences (ISCSP)
Rua Almerindo Lessa, Ajuda, Lisbon, Portugal, 1300-666
marijose.velazquez@edu.ulisboa.pt

Abstract - *Large Language Models (LLMs) are democratising access to AI for users with diverse levels of expertise, raising questions about the nature, dynamics, and effects of such interactions, particularly among lay users. Understanding how non-expert users engage with these systems is essential to inform AI literacy frameworks and responsible use guidelines, helping to reduce misinformation and address broader societal implications. To investigate these dynamics, it is first necessary to identify interaction types based on user intent, as well as prompt characteristics such as complexity, appeal, and domain familiarity, given the unprecedented flexibility of LLM use across diverse contexts. However, no categorisation of prompts with comparable complexity levels and demonstrated suitability for lay populations has yet been developed. This categorisation is essential to avoid confounds in the study of human–AI interaction. To address this gap, we applied a three-stage methodological approach. First, we generated prompts and iteratively refined their categories and complexity levels using ChatGPT, all written by the model itself. Second, we conducted a thematic qualitative analysis and curated a pre-set of 34 prompts with comparable complexity, classifying them into two main categories: a) task-oriented and b) reflexive, and two additional control categories; c) both and d) none. Third, we tested these prompts with 28 lay users from different countries through an online survey. For each prompt, participants assessed the category, perceived complexity, how interesting it was, and whether non-experts could easily understand it. Task-oriented prompts achieved a mean category confirmation rate of 62% (Max = 82%), while reflexive prompts reached 52% (Max = 71%). Complexity ratings averaged near the scale midpoint (M = 4.10), similar to interestingness (M = 4.67) and general domain (M = 4.20), indicating that prompts were neither simplistic nor overly demanding, but suitably engaging and accessible for a broad lay population. A final set of 12 prompts with at least 60% category agreement was obtained. This work can contribute to studying prompt categories among lay users of LLM-powered conversational agents, considering intent, complexity, and users' perceptions of appeal and suitability for a general audience. The final set of prompts provides a resource for advancing research in human–AI interaction, supporting future investigations into trust, emotional responses, and other key constructs in Human Computer Interaction (HCI).*

Keywords**:** Large Language Models (LLMs), human–AI Interaction, prompt-based interactions, ChatGPT, mixed methods.

This article is an extended and revised version of a paper [1] previously presented at the 12th International

Date Received: 2025-05-27
Date Revised: 2025-09-05
Date Accepted: 2025-10-26
Date Published: 2025-12-15

Conference on Multimedia and Human-Computer Interaction (MHCI 2025).

1. Introduction

Human interactions with Artificial Intelligence (AI) will continue to grow, transforming our daily lives in unprecedented ways as Large Language Models (LLMs) broaden access to AI for users from diverse backgrounds and with many levels of expertise [2], [3]. Yet, research literature on the effects of LLMs is still in its infancy [4]. This paper aims to provide a foundation for classifying types of interactions between lay users and LLMs. It addresses the need for a shared understanding of prompt categorisation based on the interaction type and complexity level. Through the clear categorisation of prompts based on their intent (type of interactions they elicit), while keeping comparable levels of complexity, we might help future research to analyse dynamics between types of interactions and users' engagement, emotional responses, trust, or other constructs, thereby improving the comparability of findings in Human-Computer Interaction (HCI).

2. Related Work

Machine Learning algorithms have evolved into highly sophisticated systems, such as Deep Learning [5], [6]. Autoregressive LLMs are a subclass of Deep Learning systems built upon the Transformer architecture, accessible through user-friendly interfaces such as ChatGPT, Claude, or Gemini [6],[7]. Namely, autoregressive LLMs are models trained on a massive amount of data to process, understand and generate human-like language. They are not only able to statistically predict the next word in a sentence, but, through attention mechanisms, they can also selectively focus on distant portions of text to achieve a more comprehensive contextual understanding [7],[8]. This has led LLMs to an unprecedented freedom of applications in many different contexts (for more information, see [6]). While the history of language models and other forms of AI is not new (see [5] for a review of this evolution), recent developments have created a paradigm shift [9]: Besides the rapid rhythm of performance improvements in AI tools [10], LLMs and other forms of generative models enable a non-technical audience to actively create many types of content (e.g., text, graphics) through simple conversational inputs (known as prompts). This ability to create novel content has come to public discourse as the relatively young term of "Generative Artificial Intelligence" (GenAI) [11]. AI systems have been integrated into entertainment, healthcare, or customer service [3], [4], [12] for years. However, the rise of LLMs introduces new interaction possibilities for lay users, making it essential to examine the types of interactions detailed in the following sections.

The first LLM with a relevant impact on public perception was OpenAI's ChatGPT [5]. Among several LLMs available in both free and paid versions, ChatGPT stands out not only as the most commonly used [13], but also as the AI system with the fastest adoption growth in the history of technology [5]. In April 2025, Sam Altman reported that 10% of the global population uses OpenAI systems, suggesting that ChatGPT had reached around 800 million users [14]. It continues to be widely studied from multiple perspectives. From a performance standpoint, researchers have compared its capabilities to human abilities on tasks such as emotional awareness [15], problem-solving [9], or rating and recommendation [7]. From the users' perspective, research in educational contexts has studied ChatGPT's acceptance, usage, learning practices and motivation among students [16], [17], [18], [19]. Industry research has explored how GPT-powered tools affect productivity and employee retention [20]. A recent study by OpenAI and MIT [21] began to explore the relationship between ChatGPT usage and the emotional well-being of its users, with particular focus on interaction modality by comparing voice and text-based conversations [22]. Despite these extensive studies, limited work has been done to address lay users' interactions with ChatGPT, particularly in terms of their initial intent and input complexity levels.

2. 1. Prompts as a key component of LLM interaction

User-friendly interactions with LLMs take place through prompts. A prompt can be defined as any text input used to elicit a conversation with AI models. Often, they work as an input-output template that allows users to mix arbitrary text with data and personalised fields [23]. While there has been an increasing interest in prompt-related research to optimize their construction and testing, such as prompt catalogues, classifications, and usage recommendations [23], [24], [25], [26], most of this work has focused on the technical aspects and not on user experience. From an HCI perspective, limited research has been done on how prompts influence users' interactions and overall experience including how users construct [27], adapt, and optimize [28] prompts. Recent studies have explored how different prompt types elicit distinct user interactions. The MIT and OpenAI [21] explored prompts categorised as personal, non-personal, or open-ended, and their relationship with interaction patterns. However, their study did not include user evaluations of these categories nor control for complexity-comparable levels across prompt types. Prompts for 'non-personal' interactions included very different cues, such as "help me brainstorm fun and educational outdoor activities for elementary school students" or "help me practice handling a difficult conversation with a co-worker by role-playing as my colleague who consistently misses project deadlines". 'Personal' condition included various cues from "Let's talk about the best show I've watched in the past few months" to "Help me reflect on the last time I was able to connect with my emotions" [22]. The 'non-personal' prompts might include cues from one's life in potentially difficult situations, like professional relationships coaching, conflict resolution practice, and problem-solving. Both categories ('personal' and 'non-personal') merged a wide variety of complexities from simple brainstorming activities to deep reflection. This might highly impact users' engagement, trust, or emotional responses.

Similarly, TUM researchers [29] provided valuable insights by analysing prompt intents and evaluating user responses, yet their research also lacked consideration of complexity levels. Other work has crafted and tested specific prompts reflecting different intents and anthropomorphic cues [30], without assessing them with participants on categorisation or complexity, limiting their applicability in more nuanced HCI investigations.

A structured categorisation of prompt types could support more targeted and comparative research. Crucially, maintaining consistent prompt complexity is essential for comparisons across interaction types in experimental settings. Prompts designed to elicit different types of interaction while keeping complexity consistent, would allow researchers to reliably compare emotional responses, behavioural patterns, or trust in AI across different groups.

2. 2. Interaction topics

Historically, research on user engagement with technology has been grounded on its instrumental value [31] as AI service chatbots have been widely used in customer service, with ongoing efforts to enhance communication style and empathetic responses to improve business outcomes [31], [32], [33], [34]. Nevertheless, a new field of research is emerging as conversational AI agents, such as LLMs, voice assistants (e.g., Alexa, Google Home), and service or social chatbots (e.g. Replika), are rapidly increasing adoption worldwide. This research field focuses on the improvements in human-like capabilities of AI technology, especially after ChatGPT's 2022 release, which marked unprecedented accuracy in language processing tasks (e.g., [7],[8],[15]) and opened up new forms of interaction that extend beyond traditional brand-consumer transactional exchanges [6]. More recently, emerging research has aimed to understand users' behaviours and meaning-making processes, including usage intent, emotional responses, trust, engagement, and action [2], [21], [31]. For example, recent research on health and AI examines LLM responses to caregivers seeking support [35] and emotional coping [36]. In professional contexts, research shows that one-third of employees think that robots would provide more unbiased feedback than managers [32]. Some users have even described conversational AI agents as a family member or friend [37], [38].

Research on LLM-powered social chatbots (e.g., Replika and Xiaolce) examines systems designed for companionship and therapeutic conversations [36]. In these conversations, users report discussing everyday activities (e.g., hobbies, sleeping habits), mental states, philosophical topics, personal worldviews, and personal problems (e.g., family conflicts, coping strategies). Nearly all participants reported self-disclosure, as they felt more comfortable sharing difficult life situations with a 'listener' perceived as non-judgmental [37]. While these interactions often focus on companion-seeking behaviour, studies show that users also engage in self-disclosing dialogues with more generalist LLMs [21]. This potentially broadens the nature of such conversations, diversifying interaction types, increasing the breadth of information exchange, and deepening vulnerability [37]. These interactions may also vary in complexity, ranging from casual and objective-specific exchanges to emotionally nuanced and reflective dialogues.

Topic patterns suggest that users interact with ChatGPT in both personal and non-personal conversations with even distribution [21]; though, heavy users tend to include more affective cues compared to casual users [22]. Conversely, research on Replika indicates that some users start with deep conversations and gradually reduce their emotional expression over time [37]. This is possibly explained by differences in the user's initial motivations [39]. Research on AI voice assistants [2] found that asking questions (20%), entertainment-jokes (12%), and searching for information (11%) are among the top 5 interactions children have with Alexa, just after playing music (40%). As these devices are integrating LLM capabilities into their architectures, such intent patterns may shift, enabling more complex and dynamic interactions. In sum, usage patterns with LLMs, AI voice assistants, and service or social chatbots highlight the need for more nuanced analysis, as the conversation type significantly affects users' emotional and psychological responses [22].

Emotional responses, trust, and perceived anthropomorphisation have also been examined [2], typically across all interaction types, without accounting for differences in complexity. Segregating by type of interaction could yield deeper insights to understand how people relate to AI agents when conversing through different intents, beyond the context (e.g. personal vs professional). Varied interactions with comparable levels of complexity might elicit similar or different relational dynamics, regardless of the AI agent's primary function or the users' intent of interaction. Accounting for complexity levels might be key to better compare users' perceptions, emotional responses, and relationship dynamics in different types of conversations. This becomes especially relevant when considering broader interactions with task-agnostic LLMs.

3. Research Question and Method

This study aims to provide a structured resource that improves the comparability of findings in LLM-based interactions by classifying types of interaction to enable controlled, systematic comparisons of factors like emotional responses, trust dynamics, or user engagement.

RQ: How can prompts be generated, curated, and classified by intent and complexity, ensuring alignment with lay users' perceptions?

The study employed a mixed-methods approach conducted in three phases for a comprehensive exploratory analysis. The first phase involved interactions with ChatGPT to co-explore possible prompts. In the second phase, qualitative research was conducted to analyse and curate these prompts according to themes and complexity levels. The third phase consisted of quantitative research with end users to assess the prompts mainly in terms of category agreement and complexity interpretation. Given ChatGPT's demonstrated performance in assisting with complex tasks, alongside its perceived human-likeness which may encourage more personal-oriented interactions, this study focused on identifying prompts that specifically elicit goal-oriented or personal conversations, excluding other applications and types of interactions from its scope.

4. First Phase: Exploring prompts with ChatGPT

Procedure: As prior studies [4], [40] suggest that ChatGPT can support researchers in generating plausible ideas, we used it as an exploratory tool to create and organize prompt examples for further analysis. All interactions were conducted using the free version of ChatGPT. We conducted two separate interactions: the first on May 3rd, 2024, using GPT-3.5, and the second on July 9th, 2024, with GPT-4o. A new chat session was started for each interaction. On each, consistent questions were asked to explore prompt categories, their characteristics, differences, and complexity levels. ChatGPT provided examples through iteration. To reduce variability in user background knowledge and ensure broader relevance, prompt examples were framed around accessible, everyday life topics that impact the general population (e.g., wellbeing practices, human development, daily habits), in line with prior research on common use cases [21], [22], [37]. This raw data was used in the qualitative analysis. Following current Best Practices on LLMs usage in research [8], we provide the list of questions sent to ChatGPT in both iterative sessions as Supplementary Material 1.

Results: GPT-3.5 and GPT-4.o proposed distinct prompt categorisations: 10 categories with 7 complexity levels and 13 categories with 3 levels. A detailed list of categories and complexity levels can be found in Supplementary Material 2. We obtained 69 preliminary prompts as examples, and selected the following categories: "informational," "comparative/contrastive," and "analytical" for their educational relevance, a widely documented use of LLMs [13]; "personal-assistance", "feedback", and "reflective" for their introspective nature, consistent with prior work describing LLMs as companion partners and tools for self-reflecting [37]; "problem-solving", given its potential applicability to real-world user needs. We excluded "opinion-based" and "experience" prompts, as ChatGPT lacks personal narratives [35]; "creative" and "entertainment" were deemed irrelevant to the study objectives; "persuasive/argumentative," present only in GPT-3.5, was excluded as it pertained more to tone than intent.

After reframing into everyday life topics, we gathered 116 prompts from three different complexities (basic, intermediate, and advanced).

5. Second Phase: Refining through qualitative approach

Procedure: Through a qualitative thematic approach [41], we interpreted and constructed themes by analysing raw responses from both GPT versions. Working with prompts around everyday life topics would reduce the potential bias of previous knowledge from participants and allow for better control when comparing complexity levels across prompts. For example, an advanced prompt on “positive psychology and resilience” could be simple to a Psychologist but complex to a Chemist. An intermediate “botanic and conservation of local plants” prompt could be relatively simple for an Ecologist, but more complex for a Psychologist.

All selected prompts were organised in a single document, keeping their original category and level of complexity. They were read and reread in detail for familiarisation purposes. After, detailed observations were made by contrasting and comparing categories and complexities. A process of labelling and recoding took place as we detected superpositions within the categories and complexities of prompts. Taking into account the characteristics of each category, intent, and theme of the prompts, we labelled them into four broader categories, while also reassigning their complexity level when needed. We selected prompts of comparable complexity, intentionally avoiding overly straightforward ones, to ensure that future research could explore more substantive and meaningful interactions with LLMs. We filtered out the prompts that were still outside of the four broader groups, too specific, too wide, repetitive, or too context-specific. We rechecked and reestablished the category or complexity if needed, and rewrote minor parts of some prompts to keep a common writing style. Again, we filtered out prompts through the previous parameters (e.g. too wide) and retained 34 for the next phase.

Results: Prompts fell into multiple categories simultaneously, for example, “informational / instructional”, “educational / technical”, “comparative / contrastive”, and “analytical” categories largely overlap in purpose. Qualitative analysis revealed that all aim to explain, clarify, or examine a specific topic. Differences often lie in the level of specificity required, or in the directive verb used, such as Describe/Explore, Explain, Compare/Contrast, or Analyse, rather than in different types of content. This overlap became evident when asking GPT for examples and clarifications, as it often provided similar prompts across these categories, reinforcing their conceptual similarity. As a result, we merged and redefined these as “task-oriented” prompts, aligning with the goal-directed use of LLMs. Verbs such as Create, Design, and Propose, which could be considered as “creative” categories, also appeared in task-based prompts, emphasising the need for careful thematic analysis.

A second group of prompts revealed a reflective intent, asking users to reflect on experiences, feelings, or behaviours. These were labelled as “personal” or “social-oriented” in previous research [21], [22], [33]. We decided to name them as “reflexive” since “personal” and “social” potentially include other topics, such as small talk, general personal assistance, and emotional support. Verbs marking these prompts included Reflect, Share, Examine, and Advice. Previous literature on self-reflection [37], [38], [39] is consistent with the designation of this category. Lastly, we identified prompts that combined both intents or belong to a different intent (as “creative” or “opinion-based”).

We retained only three levels of complexity: basic, intermediate, and advanced. GPT-3.5 initially proposed various complexity levels such as multi-step, open-ended, and context-dependent. However, these were better interpreted as prompt design features, not intrinsic complexity levels. For example, a prompt might be open-ended but still simple. “Context -dependent” complexity considers, for example, the audience’s level or area of education, and is part of the prompt design. Additionally, the “specialised/expert” level was deliberately omitted, as the study aims to establish a common ground for a general audience. The decision to use three degrees of complexity was aligned with our interactions with ChatGPT, which consistently generated examples of prompts across different categories within only three levels of complexity. Moreover, many prompts overlapped across multiple categories and complexity levels. A prompt in the category of “informational” on a basic level: "What are the key components of a balanced diet for overall well-being, and how can it impact mental health?" compared to the advanced level: "Analyse the nutritional differences between plant-based and animal-based diets, considering their impact on long-term health outcomes" illustrates how the “informational” category relates closely to the “analytical" category, by deepening the difficulty of a similar task. Similarly, a prompt from the “analytical” category, assigned to the basic complexity: "Analyse the impact of stress on physical health, discussing common stress-related ailments such as headaches and digestive issues", could be comparable in complexity with the previous one (“informational”/advanced).

The previous examples show how certain categories can be intrinsically more complex than others. After iterative refinement and thematic comparison, we selected 34 prompts from an initial pool of 116 used on the thematic analysis. These prompts were distributed across four categories: task-oriented, reflexive, both, none, and three complexity levels: basic, intermediate, advanced. This refined set of prompts aims to capture key variations in prompts’ intent and complexity, and may also serve as a resource for future research on user interaction, trust, and emotional engagement with LLMs. A complete list of prompts can be found in Supplementary Material 3.

6. Third Phase: Quantitative End User Testing

As we sought to obtain a set of prompts with different intents that could be comparable in complexity for the lay population, we conducted a quantitative study with participants using the 34 curated prompts from the previous phase to assess consistency. Data collection took place during November 2024.

The research received a positive ethical opinion from the Specialised Committee on Ethics in Psychology of ISCTE (PSI_24/2024, September 2024). Only the participants who accepted voluntarily to participate and met the inclusion criteria took part in the research; otherwise, they were redirected to the thank you message with a debriefing. No costs nor risks were associated with participating in the study. Participants could also interrupt the study or choose not to respond to questions whenever they wanted. All data was collected online through the Qualtrics software.

Participants: Using convenience sampling, the study recruited participants from multiple countries with diverse academic and professional contexts. No specific AI knowledge, technical skills, or technology expertise were required to participate; however, participants had to be at least 18 years old and possess an intermediate level of written English.

A total of 97 participants were recruited via social media platforms; of these, 21 did not meet the inclusion criteria (6 were under 18 years old and 15 reported basic-English level), 45 participants did not complete the study, and 3 were excluded due to inattentive responding (identified through control item “please check somewhat disagree for this item” [3]). As a result, the final sample comprised 28 participants (*M*age = 33.10, *SD*age = 8.44). The majority were male (57.14%), had an intermediate level of English (64.28%), and held higher education (46.42%) and masters’ degree (32.14%). Participants were mainly residents in Portugal (46.42%) and Mexico (39.28%); although, they reported other nationalities besides these countries, such as Brazilian, German, and Spanish. Professions were varied: arts, architecture, and design (21.42%); business, marketing, and finance (17.85%); while education, psychology and engineering represented 14.28% each.

Measures and procedures: The survey was organised in three parts: (1) AI prior knowledge, (2) evaluation of prompts, and (3) demographics. It also included attention check items to detect careless or automated responses.

1) AI prior knowledge was assessed using two adapted items from the familiarity factor of Körber [42], on a 5-point Likert scale (1-strongly disagree to 5-strongly agree); one question about the nature of previous interactions with LLMs (personal, professional, none, both); and frequency of use based on Vizcaino, Buman, Desroches, and Wharton [43]. We also assessed AI Literacy using the 31-item scale for non-experts by Laupichler et. al. [3], which uses a 7-point Likert format (1-strongly disagree to 7-strongly agree), and comprises 3 factors: technical understanding (14 items), critical appraisal (10 items), and practical application (7 items). Each factor was then analysed as composite variables.

2) Evaluation of prompts: For each of the 34 prompts, participants were asked to answer four questions. To ensure that the prompt content and the conversations it elicited were appropriate for lay population, they evaluated their agreement on: “I think any adult person could answer this question”. The statement “I would like to have a conversation on this topic” was also included to assess interest, as this could support more natural usage patterns in future research [22]. Both were answered using a 7-point Likert scale (1-strongly disagree to 7-strongly agree). Participants were then asked to categorise the prompts through the instruction: “Considering that ‘task-oriented’ refers to interactions asking for analysis, explanations, or customised task-assistance requests, and ‘reflexive’ involves asking for advice, guidance, or personal development assistance, select the category that best describes each prompt”. They could choose among the

previously identified 4 categories: “task-oriented”, “reflexive”, “none”, and “both”. Some prompts were included as controls, specifically those expected to fall into the “none” category (e.g., “Should freedom of speech be limited to prevent hate speech and misinformation on social media platforms?”, classified as “opinion-based” category), and the “both” category (e.g., “Share how behavioural change techniques could promote healthy habits and sustaining long-term lifestyle changes in my life", combining elements from “task-oriented” and “reflexive” categories). Finally, participants evaluated the complexity level of each prompt: “I consider that the level of complexity to answer this prompt is…” on a 7-point Likert scale (1-extremely easy to 7-extremely difficult). Prompts were presented in a randomised order across all questions.

3) Demographics included gender, years, education level, area of study, occupation, nationality, and country of residence.

Finally, criteria for detecting careless responses were included, using an attention check item (“*please check somewhat disagree for this item*”) and a bogus item (“*I consider myself among the top 10 AI researchers in the world”*) [3].

Results: Participants generally reported being familiar with ChatGPT or similar systems (n = 24) and having used them before (n = 22). Most participants (n = 20) use it for both personal and professional reasons. Participants’ average ChatGPT usage was 2.46 hours per week (SD = 3.31), with 18 participants using it for one hour or less, while two participants reported using it between 10 and 12 hours weekly. On AI Literacy, participants reported the highest scores on both Critical Appraisal (M = 5.41, SD = 1.02, range: 3.4 – 7, α = 0.89) and Practical Application (M = 5.08, SD = 1.14, range = 3 - 7, α = 0.86). The lowest reported AI Literacy was on Technical Understanding (M = 3.5, SD = 1.54, range = 1.6 - 6.8, α = 0.95), aligned with their non-technical profiles. Responses on prompts showed mean scores clustering around the midpoint of the scale, suggesting moderate perceptions across all dimensions. Prompts were generally seen as accessible to a lay population (M = 4.20, SD = 1.04), moderately interesting (M = 4.67, SD = 1.10), and marginally above average complex (M = 4.10, SD = 0.81). Prompts labelled as “task-oriented” were rated with lower complexity levels (M = 3.57) when compared to “reflexive” prompts (M = 4.59).

Prompts originally labelled as “task-oriented” had a 62% confirmation rate (i.e., participants assigned them to the same category previously identified through qualitative analysis) while “reflexive” prompts showed a lower confirmation rate of 52%. The task-oriented prompts with the highest confirmation rates were "Design a weekly meal plan for a busy individual, incorporating grocery shopping lists" and “Create a detailed itinerary for a two-week vacation in Europe” (82%), followed by “Explore practical ways to reduce plastic waste and carbon footprint in my daily life” (75%). Table 1 shows detailed information on the task-oriented prompts with the highest confirmation rates and their individual results for the other questions. Among the reflexive prompts, the highest confirmation rate was observed for the prompt “Help me find inspiration to pursue a new hobby/interest” (71%), followed by "Thinking about a challenge in my life, give me your feedback on how I handled it” (68%), and “Help me reflect on the effectiveness of my current behaviour and habits for my personal development” (64%). Table 2 shows results on all assessments for the prompts with highest confirmation rates on the reflexive category. Interestingly, prompts previously labelled as “both” were assigned by participants to the “task-oriented”, “reflexive”, and “both” categories in comparable proportions (M = 31%), supporting their classification as mixed-category prompts. An unexpected result emerged in the “none” category of prompts, which showed a low confirmation rate of 16%. The majority of participants assigned these prompts to the “reflexive” category (M = 46%).

Following user testing, a final set of 12 prompts was selected, each assessed on complexity levels and category assignment (“task-oriented” or “reflexive”) confirmed by at least 60% of user agreement. Detailed response distribution for all prompts assessed can be found in Supplementary Material 4.

Table 1. Results for task-oriented prompts with high confirmation rate

Prompt	Confirmation Rate (%)	Complexity	Interestingness	General Domain
"Analyse the impact of stress on physical health, discussing common stress-related symptoms such as headaches and digestive issues."	60.71	4.36	5.57	3.68

"Design a weekly meal plan for a busy individual, incorporating grocery shopping lists"	82.14	3.07	4.93	4.07
"Propose a customised strategy to improve sleep quality, contributing to overall well-being"	71.43	3.79	5.39	3.86
"Propose strategies for reducing sedentary behaviour during a typical workday."	64.29	2.89	4.86	4.29
“Explore practical ways to reduce plastic waste and carbon footprint in my daily life”	75.00	3.29	5.43	4.25
“Propose strategies to reduce my household expenses”	71.43	3.43	5.29	4.21
“Create a detailed itinerary for a two-week vacation in Europe”	82.14	3.29	4.00	3.32

Prompts included in these tables obtained confirmation rates above 60% during the category assessment.

Table 2. Results for reflexive prompts with high confirmation rate.

Prompt	Confirmation Rate (%)	Complexity	Interestingness	General Domain
“Advice me on how could I create a family tradition that makes us feel connected to each other”	60.71	3.89	3.82	3.89
“Help me find inspiration to pursue a new hobby / interest”	71.43	3.68	4.04	4.32
“Help me reflect on the effectiveness of my current behavior and habits for my personal development”	64.29	5.04	4.61	4.36
"(Thinking about a challenge in my life) give me your feedback on how I handled it.”	67.86	4.64	3.93	5.50
“How should I approach resolving a long-standing conflict with a family member?”	60.71	5.25	3.75	4.39

Prompts included in these tables obtained confirmation rates above 60% during the category assessment.

7. Discussion

The purpose of this research was to examine how prompts can be grouped by interaction intent and complexity levels, while also considering users’ perceptions of prompt interest and general domain familiarity. Our exploration of prompts with ChatGPT revealed differences in categorisation and complexity from both versions of GPT. Variations were expected, as these models generate responses by sampling from probability distributions rather than following fixed rules, which rarely result in identical outputs. While parameters like temperature can be configured via paid Application Programming Interface (API) to reduce output variability, reproducibility is still not guaranteed [8] as additional factors may contribute to variation (e.g. personalisation algorithms) [4], [11]. In any case, our aim was to use the free version of ChatGPT, as it reflects the access typically available to most users. As variability may limit strict replicability, we interacted with different GPT versions to test the consistency of its responses. Although the outputs were not identical, they resulted in similar categorisations and complexity assessments.

Despite lacking technical backgrounds, most participants reported being familiar with ChatGPT and notably, using it for both personal and professional purposes. This reinforces the relevance of analysing prompt-based interactions with ChatGPT in various contexts to better understand the dynamics of different intents that extend a transactional-specific interaction. The prompts selected with at least 60% of confirmation rate for their assigned category are grounded in user-oriented goals (task-oriented category) and introspection (reflexive category). They are framed around everyday topics that a lay user can relate to and were designed to go beyond overly simplistic tasks, to enable substantive studies. While task-oriented prompts were recognised, the reflexive category did not reach a strong consensus. Generative models tend to be supportive without critical inspection and lack personal narratives [35], which are required for reflexive processes. This might have limited participants’ perception of the system's capacity to foster reflection and, as a result, led to prompts not being classified as reflexive. Nevertheless, in previous research [37], participants reported that social chatbots were supportive for introspection and reflection in the same ways as in our list of prompts. Additionally, it would be valuable to test the prompts with participants from different profiles (e.g. heavy users, users of social companion chatbots, or who already use ChatGPT for these purposes) to assess the consistency of interpretations across different groups, as the reflexive category might not be easily inferred from a single prompt. Furthermore, prompts with "advise me", "reflect with me", "help me reflect" were not selected by many participants as reflexive. This might be explained by prompts containing topics related to professional life not being assessed as reflexive by participants. (e.g., "Advise me on how I could overcome imposter syndrome in professional life"), even though we included them as they potentially elicited reflexive conversations in a very important aspect of an adult life. One might think that prompts were then segregated by participants in a professional vs. personal (as leisure) logic, nevertheless, the prompt with the biggest confirmation rate in "task-oriented" included a cue for a vacation trip. Participants might have perceived prompts containing words as "professional life, or career opportunities" not as reflexive. This could be supported by previous research [27] indicating that users often have difficulties creating a verbal cue for an abstract goal. Future studies could explore how users perceive these reflexive conversations with generalist LLMs, whether they feel supported, emotionally engaged, or cognitively challenged, and if they trust the AI in any type of personal growth interactions.

Perceived complexity levels of prompts were assessed from a user's perspective, an approach that, to our knowledge, remains underexplored within the LLM context. We must not overlook the importance of measuring complexity, as its understanding will allow comparisons between types of interactions, diminishing the risk of misinterpretation (e.g., analysing the level of trust in task-oriented interaction vs. reflexive, not because of its simplicity but because of the type of interaction). Our results on perceived complexity levels indicate that prompts' corresponded to an intermediate level. Moreover, prompts labelled as "task-oriented" were generally evaluated as less complex than "reflexive" prompts. Perceived complexity may also depend on conceptual difficulty, as "reflexive" prompts seem to demand higher cognitive effort by requiring deep introspection. Notably, differences in perceived complexity also emerged among prompts within the same category that shared a similar linguistic structure (e.g., "propose strategies to…") but required additional knowledge or abstraction ability. This suggests that participants considered cognitive difficulty beyond linguistic cues. This is important, as LLM-mediated conversations may address complex topics across varied fields, even when expressed in simple terms. Complementarily, the question on general domain rating offered further insight into complexities that may arise from specialised knowledge. Through the integrated results from both questions (complexity level and general domain), prompts that remain adequate for the lay population but still differ in emotional or cognitive difficulty might be recognised.

We identified, refined and assessed prompt categories as well as complexity levels through a mixed methods approach. Our main contributions for human-AI interaction research are the novel classification of prompts, based on qualitative analysis, into "task-oriented" and "reflexive" categories, and the conceptualisation of prompt complexity as a potential variable for HCI, supported by findings from a quantitative study with lay users. As AI systems continue to evolve in human-like capabilities and natural front-end interactions, this may be particularly relevant since users may engage differently with the system depending on whether it responds to "task-oriented" or "reflexive" prompts, which, in turn, may influence participants' emotional responses, adoption, or trust in distinct ways.

On the one hand, trust is a necessary condition for (any) human-robot collaboration [44]; it is dynamic and layered [45]. Our findings can contribute through the lens of situational trust [45], which accounts for context and task difficulty. Within this layered model, "task-oriented" and "reflexive" prompts may elicit different patterns of trust calibration as users learn and adapt moment by moment.

On the other hand, trust in these AI systems may predict both emotional dependence and problematic use [22]. For example, low levels of trust or perceiving the system as indifferent to user's emotions tend to reduce engagement and emotional dependence, whereas high trust and perceiving the system as emotionally attentive seem to increase both emotional dependence and user's engagement with the system. These findings align with recent work [46] proposing a dual-dimension model of trust, in which cognitive and affective components of trust operate interactively; and where different types of LLM-based interactions might induce varying levels of cognitive and affective trust through their style and conversational framing.

In sum, findings suggest that the type and complexity of prompts may not only calibrate cognitive expectations of the system's competence but also modulate affective engagement, potentially influencing users' emotional investment and reliance on the system. In this sense, studying interactions with ChatGPT and other AI conversational agents by type/intent, their complexity, appeal, general domain and linguistic style, might offer an interdisciplinary understanding with broader theories of trust or affective engagement to help explain how certain types of interaction not only shape situational trust but also have an effect on users' cognitive and emotional responses when engaging with conversational AI.

This mixed methods study does not come without limitations. The sample size and set of prompts were small, resulting in a smaller set of prompts with high user agreement, which also constrains the range of application scenarios available for future research. Additionally, confirmation levels on categories were generally low, indicating that this area requires further exploration as the boundaries between prompt categories can be blurred, context-dependent, and difficult to infer from a single prompt. These limitations could be addressed by expanding the set of prompts, exploring linguistic cues within them, testing across larger populations with diverse profiles, as well as providing examples or framing context. Moreover, the reflexive category remains underexplored, and its operationalisation could foster further theoretical and empirical research.

8. Future Research

Recent studies are exploring the intersection of HCI with ChatGPT and other conversational AI agents, from a user's perspective. Nevertheless, limited research has been done to assess the effects of prompts on users' trust, behaviours, and emotional responses. Also, to the best of our knowledge, very few studies in HCI include interactions between end users and ChatGPT as part of the experimental design; the challenge of controlling and monitoring interactions might partially explain this.

Still, as task-agnostic, LLMs demand new approaches for studying human interactions with AI chatbots. Future research could use this set of prompts to compare users' responses across levels of complexity or to contrast them based on prompt categories. Researchers could, for example, compare trust levels in users when engaging with LLMs on "task-oriented" vs. "reflexive" interactions, or identify differences in emotional responses between groups that engage with LLMs across these intents. Additionally, future research could explore whether changing the type of interaction modifies trends or assists in recalibrating users' trust and emotional responses. Specifically, the reflexive prompts could be used to explore well-being, emotional, and interpersonal engagement.

It is important to note that, although this research differentiated prompts by type of interaction and comparable complexity levels, the list of prompts alone is insufficient. Future studies should use this list of prompts while considering prompting strategies [47] for comparable results. It is recommended that, as part of the future study design, researchers create a template that users can complete to enable personalisation, and set a frame for the output response, such as the expected length of response, or follow-up steps from the LLM, to avoid confounding aspects in the interaction. A specific prompt template should be designed in accordance with the aims of the study.

Future research could replicate this approach with a larger, more diverse sample to amplify the insights from this exploratory study, and to examine how categories, complexities, and lay perceptions vary across different populations. Furthermore, future studies could assess additional perceptions to better interpret results and prompt characteristics, for instance, by expanding the distinction between cognitive, emotive, and linguistic complexities.

At the time this research was conducted, the free version of ChatGPT was used to provide text-based outputs. Meanwhile, other companies (e.g. Meta AI, Anthropic, or Mistral AI) have also released their own LLMs; however, reproducing outputs continues to represent challenges. Paid versions of these systems and more recent open-source models, accessible via APIs, enable developers and researchers to fine-tune parameters and model weights. This offers greater control compared to public chat interfaces while improving reproducibility of findings [8]. Therefore, future studies should use LLM APIs to enhance repeatability and robustness for better studying user-AI interaction; by sharing all model settings (parameters, weights, prompts, etc.) via supplementary materials, researchers may promote validation from peers through methodological transparency and will help tracing changes in LLMs' performance and capability across time.

Additionally, recent models now provide multimodal capabilities, including interpreting and generating images, processing audio inputs, and voice-based responses. Creative prompts involving image, narrative, and music generation could be further explored, as this constitutes a growing use case for ChatGPT [48] and other LLM-based interactions; despite representing only a small share of users' intents, multimedia creation trends have nearly tripled in less than one year [49]. These extended functionalities and usage trends also open new research questions on user interaction, trust, anthropomorphisation, and the potential use of co-creative AI tools.

9. Conclusion

This work provides an empirical resource for understanding user-centered prompt categories in LLM-powered conversational agents. By systematically generating, categorising, refining and assessing prompts according to intent, complexity, appeal, and accessibility for non-experts, we offer a structured resource for future research and practical applications in human–AI interaction. The study produced a set of prompts with consistent category agreement and comparable complexity levels. Complexity, interestingness and general domain ratings were balanced, reflecting prompts that are engaging, accessible, and appropriately challenging for a general population, while also enriching future research.

The final set of prompts can support future studies on trust, affective responses, and other HCI constructs, enabling researchers and practitioners to design more inclusive and effective interactions. Importantly, the typology highlights the value of analysing interaction types that are similar in complexity and cognitive demand, allowing for a clearer understanding of the dynamics both within and between categories. Overall, this work demonstrates the importance of a user-centered approach in advancing human–AI interaction research, particularly for populations without specialised technical expertise, and provides a reproducible framework for exploring prompt design across diverse contexts.

Acknowledgements

The present paper is an extended version of the paper "Towards a Typology of Prompts for Human-AI Interaction: Mapping Intent and Complexity with Lay Users" presented in the 12th International Conference on Multimedia and Human-Computer Interaction (MHCI 2025). DOI: 10.11159/mhci25.122

ChatGPT 4.o free version was used to check grammar in some parts of this paper. No ideas, references, nor full written sections were obtained through ChatGPT.

Replicability Statement

All interactions with the ChatGPT models were conducted via web interface (chat.openai.com), with parameters set to default. Models used were the free versions of GPT-3.5 (accessed on May 3rd, 2024) and GPT-4.o (accessed on July 9th, 2024). No API was used. All Supplementary Materials can be accessed here.

ChatGPT Exploration Update

Considering that GenAI models are always changing, through self and human-on-the-loop optimisation, we conducted an extra interaction process to explore how much the model's response regarding prompt categorisation and complexity levels had changed. This was conducted after running the study with participants (May 6th, 2025). The free GPT-4.o model kept "informative", "creative", "problem-solving", and "analytical" categories; while sharing a new "interrogative", merging "educational" and "reflective" categories. Regarding complexity levels, ChatGPT kept the same three, but emphasised on style and structure. While the essential categorisation and complexity patterns remained largely unchanged, the observed differences may reflect updates in training data and, potentially, algorithmic personalisation. For example, ChatGPT mentioned an "iterative" complexity for the first time in our interactions. This update could be explained by new state-of-the-art knowledge and evolving user behaviour through iterative engagement. For more consistent analyses on GPT's dynamic capabilities [50], future studies might consider using the OpenAI API, which allows greater control over model parameters and versions, providing a more stable basis for tracking changes over time. Such longitudinal tracking could help not only to evaluate the model's capabilities for flexible, cross-domain reasoning, but to support research that anticipates critical questions regarding trust in AI and its implications for societal discourse.

Author Contributions

CRediT: **Marijose Páez Velázquez**: Conceptualisation, Methodology, Formal Analysis, Investigation, Resources, Data Curation, Writing (Original Draft, Review & Editing), Visualisation, and Project Administration. **Elzbieta Bobrowicz-Campos**: Conceptualisation, Methodology, Writing (Review), and Supervision. **Patricia Arriaga**: Conceptualisation, Methodology, and Writing (Review & Editing).

References

[1] M. P. Velázquez, E. Bobrowicz-Campos, and P. Arriaga, "Towards a Typology of Prompts for Human-AI Interaction: Mapping Intent and Complexity with Lay Users," Paris: Avestia, Aug. 2025. doi: 10.11159/mhci25.122.

[2] V. Andries and J. Robertson, "Alexa doesn't have that many feelings: Children's understanding of AI through interactions with smart speakers in their homes," *Comp. and Ed.: Art. Int.*, vol. 5, 2023, doi:10.1016/j.caeai.2023.100176

[3] M. C. Laupichler, A. Aster, N. Haverkamp, and T. Raupach, "Development of the "Scale for the assessment of non-experts' AI literacy" – An exploratory factor analysis," *Computers in Human Behavior Reports*, vol. 12, 2023, doi: 10.1016/j.chbr.2023.100338.

[4] S. S. Sohail, D. Ø. Madsen, Y. Himeur, and M. Ashraf, "Using ChatGPT to navigate ambivalent and contradictory research findings on artificial intelligence," *Frontiers in Art. Intell.*, vol. 6, Jul. 2023, doi:10.3389/frai.2023.1195797

[5] A. Oliveira, *A Inteligência Artificial Generativa*. Lisbon: Fundação Francisco Manuel dos Santos, 2025.

[6] M. Johnsen, *Developing AI Applications With Large Language Models*. 2025.

[7] J. Liu, C. Liu, P. Zhou, R. Lv, K. Zhou, and Y. Zhang, "Is ChatGPT a Good Recommender? A Preliminary Study" 2023

[8] S. Abdurahman, A. Salkhordeh Ziabari, A. K. Moore, D. M. Bartels, and M. Dehghani, "A Primer for Evaluating Large Language Models in Social-Science Research", *Advances in Methods and Practices in Psychological Science*, vol.8, no.2, 2025, doi: 10.1177/25152459251325174.

[9] G. Orrù, A. Piarulli, C. Conversano, and A. Gemignani, "Human-like problem-solving abilities in large language models using ChatGPT," *Frontiers in Artif. Intell.*, vol. 6, 2023, doi: 10.3389/frai.2023.1199350.

[10] T. Kwa, B. West, J. Becker, A. Deng, K. Garcia, M. Hasin, S. Jawhar, M. Kinniment, N. Rush, S. Von Arx, R. Bloom, T. Broadley, H. Du, B. Goodrich, N. Jurkovic, L. Miles, S. Nix, T. Lin, N. Parikh, D. Rein, L. Sato, H. Wijk, D. Ziegler, E. Barnes, L. Chan, "Measuring AI Ability to Complete Long Tasks," Mar. 2025.

[11] R. Ronge, M. Maier, and B. Rathgeber, "Towards a Definition of Generative Artificial Intelligence," *Philosophy & Technology*, vol. 38, no. 1, 2025, doi: 10.1007/s13347-025-00863-y.

[12] M. Castro, L. Lisboa, and A. Barcaui, "Beyond the Code: Understanding Professional Users' Perspectives on AI Implementation," in *Proc. 10th World Congr. on EECSS*, Avestia Publishing, Aug. 2024. doi: 10.11159/mhci24.111

[13] L. Rainie, "Close encounters of the AI kind: The increasingly human-like way people are engaging with language models," 2025. https://imaginingthedigitalfuture.org/wp-content/uploads/2025/03/ITDF-LLM-User-Report-3-12-25.pdf

[14] S. Altman and C. Anderson, "OpenAI's Sam Altman Talks ChatGPT, AI Agents and Superintelligence," YouTube. Accessed: May, 2025 [Online Video]. Available: https://www.youtube.com/watch?v=5MWT_doo68k.

[15] Z. Elyoseph, D. Hadar-Shoval, K. Asraf, and M. Lvovsky, "ChatGPT outperforms humans in emotional awareness evaluations," *Frontiers in Psychology*, vol. 14, 2023, doi: 10.3389/fpsyg.2023.1199058.

[16] J. L. Steele, "To GPT or not GPT? Empowering our students to learn with AI," *Computers and Education: Artif. Intell.*, vol. 5, 2023, doi: 10.1016/j.caeai.2023.100160.

[17] F. Hanum Siregar, B. Hasmayni, and A. H. Lubis, "The Analysis of Chat GPT Usage Impact on Learning Motivation among Scout Students," *Int. Journal of Research and Review*, vol. 10, no. 7, 2023, doi: 10.52403/ijrr.20230774.

[18] A. Habibi, M. Muhaimin, B. K. Danibao, Y. G. Wibowo, S. Wahyuni, and A. Octavia, "ChatGPT in higher education learning: Acceptance and use," *Computers and Education: Artif. Intell.*, vol. 5, 2023, doi:10.1016/j.caeai.2023.100190

[19] Y. Zheng, "ChatGPT for Teaching and Learning: An Experience from Data Science Education," in *SIGITE 2023 - Proc. of the 24th Annual Conf. on*

Information Technology Education, 2023. doi: 10.1145/3585059.3611431.
[20] E. Brynjolfsson, D. Li, and L. Raymond, "Generative Ai at Work," *SSRN El. Journal*, 2023, doi:10.2139/ssrn.4426942
[21] J. Phang, M. Lampe, L. Ahmad, S. Agarwal, C. Fang, A. Liu, V. Danry, E. Lee, S. Chan, P. Pataranutaporn, P. Maes, "Investigating Affective Use and Emotional Well-being on ChatGPT," 2025. [Online]. Available: https://cdn.openai.com/papers/15987609-5f71-433c-9972-e91131f399a1/openai-affective-use-study.pdf
[22] C. Fang, A. Liu, V. Danry, E. Lee, S. Chan, P. Pataranutaporn, P. Maes, J. Phang, M. Lampe, L. Ahmad, S. Agarwal, "How AI and Human Behaviors Shape Psychosocial Effects of Chatbot Use: A Longitudinal Controlled Study," 2025.
[23] V. Sanh, A. Webson, C. Raffel, S. Bach, L. Sutawika, Z. Alyafeai, A. Chaffin, A. Stiegler, T. Le Scao, A. Raja, M. Dey, M. Bari, C. Xu, U. Thakker, S. Sharma, E. Szczechla, T. Kim, G. Chhablani, N. Nayak, D. Datta, J. Chang, M. Jiang, H. Wang, M. Manica, S. Shen, Z. Yong, H. Pandey, M. McKenna, R. Bawden, T. Wang, T. Neeraj, J. Rozen, A. Sharma, A. Santilli, T. Fevry, J. Fries, R. Teehan, T. Bers, S. Biderman, L. Gao, T. Wolf, A. Rush A., "Multitask Prompted Training Enables Zero-Shot Task Generalization," in *ICLR 2022 - 10th Int. Conf. on Learning Representations*, 2022.
[24] J. WhitE, Q. Fu, S. Hays, M. Sandborn, C. Olea, H. Gilbert, A. Elnashar, J. Spencer-Smith, D. Schmidt, "A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT," Feb. 2023.
[25] S. Geng, S. Liu, Z. Fu, Y. Ge, and Y. Zhang, "Recommendation as Language Processing: A Unified Pretrain, Personalized Prompt & Predict Paradigm" *Proc. 16th ACM Conf. on Reco Systems*, 2022. doi: 10.1145/3523227.3546767
[26] S. K. Santu and D. Feng, "TELeR: A General Taxonomy of LLM Prompts for Benchmarking Complex Tasks" 2023.
[27] H. Subramonyam, R. Pea, C. Pondoc, M. Agrawala, and C. Seifert, "Bridging the Gulf of Envisioning: Cognitive Challenges in Prompt Based Interactions with LLMs," *Conf. on Human Factors in Computing Systems Proc.*, 2024 doi:10.1145/3613904.3642754
[28] S. Mondal, S. D. Bappon, and C. K. Roy, "Enhancing User Interaction in ChatGPT: Characterizing and Consolidating Multiple Prompts for Issue Resolution," 2024.
[29] A. Bodonhelyi, E. Bozkir, S. Yang, E. Kasneci, and G. Kasneci, "User Intent Recognition and Satisfaction with Large Language Models: A User Study with ChatGPT," 2024.
[30] L. Ibrahim, C. Akbulut, R. Elasmar, C. Rastogi, M. Kahng, M. Morris, K. McKee, V. Rieser, M. Shanahan, L. Weidinger, "Multi-turn Evaluation of Anthropomorphic Behaviours in Large Language Models," 2025, http://arxiv.org/abs/2502.07077
[31] S. Chandra, A. Shirish, and S. C. Srivastava, "To Be or Not to Be …Human? Theorizing the Role of Human-Like Competencies in Conversational Artificial Intelligence Agents," *Journal of Management Information Systems*, vol. 39, no. 4, 2022, doi: 10.1080/07421222.2022.2127441.
[32] R. P. Bagozzi, M. K. Brady, and M. H. Huang, "AI Service and Emotion," *Journal of Service Research*, vol. 25, no. 4. 2022. doi: 10.1177/10946705221118579.
[33] Y. Xu, J. Zhang, and G. Deng, "Enhancing customer satisfaction with chatbots: The influence of communication styles and consumer attachment anxiety," *Frontiers in Psychology*, vol. 13, 2022, doi: 10.3389/fpsyg.2022.902782.
[34] C. Pelau, C. Volkmann, M. Barbul, and I. Bojescu, "The Role of Attachment in Improving Consumer-AI Interactions," *Proc. of the Int. Conf. on Bus. Excellence*, vol. 17, no. 1, 2023, doi: 10.2478/picbe-2023-0097.
[35] K. Saha, Y. Jain, C. Liu, S. Kaliappan, and R. Karkar, "AI vs. Humans for Online Support: Comparing the Language of Responses from LLMs and Online Communities of Alzheimer's Disease," *ACM Transactions on Computing for Healthcare*, Jan. 2025, doi: 10.1145/3709366.
[36] K. T. Pham, A. Nabizadeh, and S. Selek, "Artificial Intelligence and Chatbots in Psychiatry," *Psychiatric Quarterly*, vol. 93, no. 1. 2022. doi: 10.1007/s11126-022-09973-8.
[37] M. Skjuve, A. Følstad, K. I. Fostervold, and P. B. Brandtzaeg, "My Chatbot Companion - a Study of Human-Chatbot Relationships," *Int. Journal of Human Computer Studies*, vol. 149, 2021, doi: 10.1016/j.ijhcs.2021.102601.
[38] A. Purington, J. G. Taft, S. Sannon, N. N. Bazarova, and S. H. Taylor, "Alexa is my new BFF: Social roles, user satisfaction, and personification of the

Amazon Echo," *Conf. on Hum Facts in Comp Sys*, 2017: 10.1145/3027063.3053246
[39] R. E. Guingrich and M. S. A. Graziano, "Chatbots as social companions: How people perceive consciousness, human likeness, and social health benefits in machines", 2023, doi: 10.1093/9780198945215.003.0011.
[40] M. Dowling and B. Lucey, "ChatGPT for (Finance) research: The Bananarama Conjecture," *Finance Research Letters*, vol. 53, 2023, doi: 10.1016/j.frl.2023.103662.
[41] V. Braun and V. Clarke, "Reflecting on reflexive thematic analysis," *Qualitative Research in Sport, Exercise and Health*, vol. 11, no. 4, pp. 589–597, Aug. 2019, doi: 10.1080/2159676X.2019.1628806.
[42] M. Körber, "Theoretical considerations and development of a questionnaire to measure trust in automation," in *Adv. in Intell. Systems and Computing*, 2019. doi: 10.1007/978-3-319-96074-6_2.
[43] M. Vizcaino, M. Buman, C. T. Desroches, and C. Wharton, "Reliability of a new measure to assess modern screen time in adults," *BMC Public Health*, vol. 19, no. 1, 2019, doi: 10.1186/s12889-019-7745-6.
[44] F. Schoeller, M. Miller, R. Salomon, and K. J. Friston, "Trust as Extended Control: Human-Machine Interactions as Active Inference," Frontiers in Systems Neuroscience, vol. 15, Oct. 2021, doi: 10.3389/fnsys.2021.669810.
[45] K. A. Hoff and M. Bashir, "Trust in automation: Integrating empirical evidence on factors that influence trust," Human Factors, vol. 57, no. 3, 2015, doi: 10.1177/0018720814547570.
[46] R. Shang, G. Hsieh, and C. Shah, "Trusting Your AI Agent Emotionally and Cognitively: Development and Validation of a Semantic Differential Scale for AI Trust" Nov. 2024. https://arxiv.org/html/2408.05354v1#Sx1
[47]S. Schulhoff, Ilie M., Balepur N., Kahadze K., Liu A., Si C., Li Y., Gupta A., Han H., Schulhoff S., Dulepet P., Vidyadhara S., Ki D., Agrawal S., Pham C., Kroiz G., Li F., Tao H., Srivastava A., Da Costa H., Gupta S., Rogers M., Goncearenco I., Sarli G., Galynker I., Peskoff D., Carpuat M., White J., Anadkat S., Hoyle A. & Resnik P., "The Prompt Report: A Systematic Survey of Prompt Engineering Techniques," Arxiv, Feb. 2025.
[48] A. Chatterji, Cunningham, T., Deming, D., Hitzig, Z., Ong, C., Shan, C. & Wadman, K. 2025, September 15. "How People Use ChatGPT", OpenAI. https://www.nber.org/system/files/working_papers/ w34255/w34255.pdf
[49] R. Appel,, McCrory, P., Tamkin, A., McCain, M., Neylon, T., Stern, M. 2025, September 15. The "Anthropic Economic Index report: Uneven geographic and enterprise AI adoption", Anthropic. https://assets.anthropic.com/m/218c82b858610fac/original/Economic-Index.pdf
[50] S. Bubeck., Chandrasekaran V., Eldan R., Gehrke J., Horvitz E., Kamar E., Lee P., Lee Y., Li Y., Lundberg S., Nori H., Palangi H., Ribeiro M., Zhang Y., "Sparks of Artificial General Intelligence: Early experiments with GPT-4," Apr. 2023.