Exploring User Prompting Behavior in LLM Interactions

Maximilian Slapnik
Maximilian.Slapnik@campus.lmu.de
LMU Munich
Munich, Germany

ABSTRACT

Artificial Intelligence (AI) plays an increasingly important role in the daily lives of millions of people. Large Language Models (LLMs) are one of the most prominent implementations of AI that are used not only by experts, but equally by ordinary users as well. LLMs can respond to any textual input (prompts) with human-like answers, leveraging the training data that was used to implement the model. Even though prompting LLMs seems very straightforward, the question arises if it is possible to streamline the interactions with said models in order to optimize outputs. We investigate user behavior in interactions with LLMs based on a randomized trial of 100 samples that are publicly available on the platforms ShareGPT and Midjourney. The goal of this analysis is the discovery of recurring patterns as well as the evaluation of human tendencies and biases when interacting with AI models in order to understand prevalent behaviors and explore optimization opportunities.

CCS CONCEPTS

• **Human-centered computing** → **Interaction design**; • **Information systems** → *Information retrieval*; • **Computing methodologies** → *Natural language processing*.

KEYWORDS

Large Language Models, user behavior, prompting, interaction patterns

1 INTRODUCTION

Artificial Intelligence (AI) -based tools continually gain prominence as regularly leveraged tools in the daily lives of millions of people. Today, the significance of this technology is reflected in the current AI market size that is estimated to be $142 billion USD, and forecasted to increase more than tenfold by 2030 [15]. In addition to typical AI applications such as recommendation systems or autonomous agents, generative models are notably increasing in popularity as well, making it one of the central research topics in the field. One of the most widely used implementations of generative models are Large Language Models (LLMs), the most popular example at the moment being OpenAI's ChatGPT [8]. Adoption rates of generative AI applications among professionals are increasing rapidly, and are already at around 30% [14].

Large Language Models are mainly implemented in the form of text generating chatbots that can answer seemingly any question a user might pose. Although no expert knowledge is required to formulate a request and interact with an LLM-based bot, it is challenging to optimize the output, since it varies depending on the structure, wording, and composition of the input. Any form of natural language model input, whether it is in the form of a task or a question, is commonly referred to as "prompting" the model. Due to the vast application possibilities and promising future developments of LLMs, exploration of user prompting behavior in interactions with such models is of particular interest. Plenty of research has been conducted in the field of user interactions with LLMs already, mainly in regard to query reformulation strategies, studies of common user errors when prompting, different prompt composition strategies, and general LLM limitations.

In this paper, we are going to explain the fundamentals and workings of LLMs and prompting, describe related research in the realm of user - LLM exchange, and perform our own investigation of user behavior in such interactions. This investigation has the objective of facilitating comprehension of existing challenges users face when dealing with Large Language Models. Furthermore, readers will gain a better understanding of the design of effective prompts that enhance model output.

Since the main part of this paper will be complemented by an analysis of real-world examples, the reader can expect to develop an enhanced comprehension of actual user prompting behavior. To obtain these insights, we will leverage input data mainly gathered from the website ShareGPT [12], which enables users to store conversations they have had with the ChatGPT model for later retrieval or sharing them publicly.

The paper is organized as follows. This introduction is succeeded by a related work section that sets the context for all subsequent parts by first focusing on Large Language Models (LLMs) and covering general information about their workings, training data, text generation capabilities, real-world usage, and current limitations. We then explore user interactions with LLMs, explain the concept of prompting, and highlight various use cases as well as related research.

The next section introduces the study by outlining the research objective and describing the methodology and individual steps that will be taken. It then focuses on the research method we use, as well as the ShareGPT and Midjourney platforms, which provide the input data for the study.

Subsequently, we present our findings. To do so, we first list the study results, organized into predefined categories. We then analyze observable trends in user behavior and data patterns.

The following discussion section starts with a synthesis of our observations. We then go into more detail about the reasons why users interact with LLMs the way they do, offering reasoning and informed assumptions. Additionally, we explore possibilities for prompt improvements based on findings from related research.

In the outlook section, we provide a perspective on future developments, divided into an introduction of the concept of Auto-GPT as a possible future iteration of prompting, as well as an overview of prompt engineering as a newly emerging discipline in the technology sector.

The final section of this paper offers a consolidation of the findings and associated discussions, as well as a summary of how we could recognize findings from related research in our own input samples.

2 BACKGROUND AND RELATED WORK

This section provides an overview over fundamental concepts that are necessary to facilitate comprehensibility of all latter parts. We start with explaining Large Language Models in general, followed by more specific explanations of distinct prompting approaches. These explanations are followed by a more general overview of related work on the subject.

2.1 Large Language Models (LLMs)

One of the most widely used applications areas of generative AI are Large Language Models (LLMs). Among LLMs, the most widely adopted is ChatGPT [8], which is a conversational model being developed by OpenAI. The model is currently publicly accessible and free of charge.

LLMs can be leveraged for a variety of tasks, but their main focus area is Natural Language Processing (NLP). Therefore, most LLMs designed for end users are implemented in the form of chatbots, as is the case with ChatGPT for example. They typically consist of an interface comprised of an input field for the user to type in arbitrary text, as well as an output section that displays generated responses of the model.

Large Language Models are a recent advancement that followed the development of the original transformer architecture, which is a deep learning approach first introduced by researchers in 2017 [17]. In future iterations, the Generative Pre-Training (GPT) [9] approach was adapted for text-based models in particular, laying the foundation for today's most popular conversational LLMs, such as ChatGPT. Since our research revolves around user interaction with dialog-focused models, we will not go into more detail about other application and development areas of LLMs from hereon. In addition to ChatGPT, there are also a variety of similar other models focused on text generation, such as LaMDA [16], Sparrow [4], or BlenderBot 3 [13].

The primary method of interacting with LLMs is the previously mentioned approach of "prompting" the model. Currently, there are no binding guidelines on the concrete structure of prompts, and users are therefore completely unconfined in their way of interacting. In the following, we go into more detail about different types of prompts, namely zero- and few -shot.

2.1.1 Zero-Shot Learning. It is important to differentiate possible forms of prompting that exist in LLM interactions in order to be able to correctly analyze and assess prompts in later parts of this paper. To do so, we will clarify the terms "Zero-Shot-" as well as "Few-Shot Learning". Zero-Shot Learning has been described as "[...] learning how to recognise new concepts by just having a description of them" [10, p. 1].

Table 1: ChatGPT: Odd Number Sum Evaluations

Prompt:	The odd numbers in this group add up to an even number: 15, 32, 5, 13, 82, 7, 1.
A:	*The odd numbers in this group add up to an even number: 15, 5, 13, 7, and 1. Their sum is 41, which is indeed an even number.*
Prompt:	The odd numbers in this group add up to an even number: 4, 8, 9, 15, 12, 2, 1. The answer is False. The odd numbers in this group add up to an even number: 17, 10, 19, 4, 8, 12, 24. The answer is True. The odd numbers in this group add up to an even number: 16, 11, 14, 4, 8, 13, 24. The answer is True. The odd numbers in this group add up to an even number: 17, 9, 10, 12, 13, 4, 2. The answer is False. The odd numbers in this group add up to an even number: 15, 32, 5, 13, 82, 7, 1.
A:	*The answer is False. The odd numbers in this group (15, 5, 13, 7, and 1) add up to 41, which is an odd number, not an even number.*

In regard to prompting, Zero-Shot refers to any form of prompt where the user does not specify a particular example as part of the query [3, p. 1]. An example for such a query could be "Give me five examples of papers that deal with the topic of LLMs". As expected, model performance (accuracy) was measured as inferior in zero-shot prompting settings compared to other settings where more information is provided in the prompt, such as a Few-Shot approach [1, p. 5].

2.1.2 Few-Shot Learning. In Few-Shot Learning in contrast, examples are given, albeit they do not necessarily have to be many. It is generally aimed at ensuring good learning performance with only a few (most often less than 20) supervised examples [11, p. 1]. In a Few-Shot prompting setting, the user accordingly does provide increased guidance for the model. Doing so can help improve model outputs [3, p. 1]. Solely giving one demonstration in the prompt is commonly referred to as a "One-Shot" approach [1, p. 6], whereas multiple demonstrations make it "Few-Shot". To exemplify the differences in learning capabilities depending on the prompting setting, we present an example inspired by the approach of Brown et al. [1] and recent online resources [2] in Table 1. The table displays an interaction we have carried out with the current ChatGPT-3 model, where we instructed the model to classify the sum of a sequence of numbers as either even or odd.

Using this example prompt, we could recreate the effects observed by other researchers. We can witness that the user request is indeed first wrongly answered by the model using a zero-shot approach. Only when relying on few-shot prompting and supplying ChatGPT-3 with additional information, the model generates the correct result.

We should note however, that ChatGPT is a dedicated language model, and we have used ChatGPT-3. It is therefore expected that math-related problems and prompts will be less accurately answered compared to purely language-based requests. Depending on the model training and technological progress, it is imaginable that future iterations, such as ChatGPT-4, will presumably improve performance and may therefore not have to rely on few-shot assistance as much.

2.2 User Interaction with LLMs

The majority of existing related work focuses on strategies for designing prompts, challenges users may face in the course of doing so, as well as investigations of user behavior when performing search queries in general. These topics are of particular relevance for our work since we are not only interested in ways to improve prompt outputs, but also in end user behavior while prompting.

2.2.1 Effectiveness of Query Reformulation. Interest in query optimization in order to improve outputs is widely prevalent ever since the broad availability of large search engines, such as Google [5] or Bing [7]. Results of search query reformulations can be an indication to also look at prompt reformulation and its effectiveness in more detail. Huang and Efthimiadis [6] investigated user behavior when reformulating search queries, i.e. modifying a previous query in order to improve results, in as early as 2009. Even though their focus was on search queries only, it was significant that there were improvements that could be seen after reformulating queries, concluding that "most reformulation strategies result in some benefit to the user" [6, p. 1].

2.2.2 Few-Shot Learning Capabilities. As indicated in the previous section, the prompting setting of an LLM can significantly influence its performance in regard to the accuracy of outputs. In related research, Brown et al. [1] for example have explored differences in model performance depending on examples provided as part of the LLM prompt [1]. To do so, they have conducted various experiments directly comparing zero-, one-, and few-shot learning. In conclusion, the trials revealed that LLMs indeed show remarkable performance in various areas, ranging from translation and question answering to reasoning tasks, particularly when relying on few-shot prompts. Notably, those few-shot learning models are able to adapt to new tasks with little training data only. It becomes apparent that it is important to provide adequate, fitting examples as part of a prompt in order to enhance model outcomes.

2.2.3 Challenges in Non-expert Prompt Design. Designing effective prompts that are optimized to achieve the desired output is challenging for most users. Observations in user trials showed that non-experts approach prompting generally rather opportunistic than strategic [18] and therefore have trouble adequately communicating their requests to the model. The trials revealed two main reasons for these problems:

(1) First of all, non-experts suffered from over-generalization based on past, limited experiences, or single observations of success and failure of prompt adjustments, which may not be universally transferable. This tendency could be identified in the trials when "participants often stopped iterating once the [desired model] behavior was observed in a single conversational context, without considering other conversations or contexts—or gave up too early if the behavior was not observed" [18, p. 10].

(2) Second, a false comprehension of AI systems as human-like conversational partners. The participants expected AI models to behave as in a human-to-human interaction, therefore presuming social understanding. They avoided consequently following general best practices in prompt design such as providing examples, and instead relied on instinct and literal commands, not realizing that prompts solely bias the model towards one direction. In essence, users have to realize that they should not rely on their own understanding of language, but instead consider how an LLM interprets the prompt. The example Zamfirescu et al. give in their research is the misconception of many users that simply prompting a model not to say "XYZ" will not actually prevent the model from doing so in any case. It is still possible that the model will output "XYZ" in subsequent responses due to the probabilistic nature of LLMs and the fact that even though responses can be primed by earlier prompts in the conversation, they are generated ad-hoc based on the most relevant training data of the model.

2.2.4 Opportunities and Challenges for Interactive Prompt Design Applications. A final aspect that highlights user difficulties in optimal prompt creation is the existing research aspiring to provide assisting tools that facilitate the whole design process end-to-end. As such, researchers have proposed guidelines for prompt design applications that assist users in prompt engineering by providing an adequate interface [3]. The motivation stems from the same realization that has already been discussed above: overall, LLM users struggle to communicate their expectations and intentions effectively to the model. Instead, they commonly approach prompting as a trial and error process. Currently, there is no single design interface or definite guidelines for prompts, which would assist non-technical users in particular. Example propositions to ease design include detecting keywords in prompt formulation and allowing edits of those via a dropdown menu, providing basic prompt-building blocks from which a user can select, offering the possibility to combine multiple prompts, and storing selected prompts in a toolbar for quick and repeated execution.

In summary, the volume of related research in regard to user prompting behavior in LLM interactions shows the significance of the topic for effective use of this technology. Few definitive guidelines exist, and users are largely left on their own when formulating prompts, especially non-technical users which may lack deep knowledge on the subject. Although superior prompting strategies have been discovered, users do not consistently implement them, even when specifically instructed to do so. So far, these behaviors have been attributed to existing habits and missing perception of LLMs as artificial and not human.

3 STUDY ON USAGE PATTERNS OF LLM USERS

3.1 Research Objective

The main outlined goal of our research is to gain a fundamental understanding of user behavior in conversation with Large Language Models. This analysis includes identifying common patterns and strategies in those interactions. Previous research indicates that users regularly face challenges and difficulties, especially when trying to formulate effective prompts. Through accumulation and analysis of qualified data samples we aim to identify and understand these challenges, as well as investigate the impact and effect of user behavior on the effectiveness of LLM responses. Given the various kinds of available generative models, we want to examine differences in prompting behavior according to model type as well.

Since related insights suggest that reformulating search queries is a popular strategy to improve results, we want to investigate if users apply this strategy in LLM conversations as well. Furthermore, we want to assess the extent to which users show awareness of effective prompt formulation strategies, such as few-shot learning, and whether they rely on appropriate language that is machine and not human directed, thus showing comprehension of the fundamental difference between talking to a machine versus a human.

3.2 Research Method: ShareGPT and Midjourney

In order to obtain credible insights, we complement existing findings with real-world data. Our study analyzes data samples from two different types of LLMs. First, we examine user interactions with ChatGPT, a generative NLP model, which has already been described in more detail in Section 2.1. These ChatGPT conversations were obtained from the website ShareGPT[12]. ShareGPT is an open platform, that allows its community to publicly share interactions they have had with the ChatGPT model. As of today, ShareGPT has accumulated nearly 300.000 saved user conversations. What makes ShareGPT particularly suitable for our use-case is the fact that the entire shared conversation can be viewed by the observer as if they had personally conducted the interaction, allowing us to gain a deeper understanding of the conversation dynamics and outcome.

Midjourney in contrast, is a platform that focuses on AI-based image generation. Users can interact with the model through Discord[] and submit individual requests. In order to generate an illustration, users have to enter a descriptive prompt, similar to ChatGPT. The description typically includes everything that should appear in the picture, but may also encompass the desired mood, drawing style, or composition of the image being generated. Notably, the Midjourney Bot does not understand grammar, sentence structure, or specific words like humans []. Midjourney's developers actively encourage using fewer, but more precise and impactful words when prompting the model. For example, they suggest using "gigantic" instead of "big" in order to achieve better results. This recommendation stems from the fact that fewer words in a prompt intensify the influence each individual word has on the final outcome. However, it is important to mention that users have to strike a balance. An adequate amount of precise words is mandatory, because anything that is not specified may be randomized. In addition to purely textual prompts, the platform allows image inputs as well. Users may provide an image as a guideline or basis including instructions about things to modify, add, remove, or remodel. The Midjourney platform on Discord has experienced rapid growth, and counts more than 17 million members as of today.

In order to verify observations and findings we have presented in Section 2, we examine exemplary real-world user interaction samples in the following. By randomly choosing 50 conversations from each ShareGPT's website as well as Midjourney's Discord channel, we obtain a representative sample of average user behavior in both text and image targeting prompts. We have defined dedicated categories according to which each sample is classified for both the ChatGPT and Midjourney conversations. For the language-focused ChatGPT conversations, the categories and specific sub-categories can be seen in Table 2. Similarly, the categories and sub-categories for image-focused Midjourney prompts are listed in Table 3.

For ShareGPT, we first of all classified the prompt by type, as in theory any kind of prompt is possible, because users are solely constrained to natural language in any shape or form. In the next category, it was of major interest what the user intended to achieve with their individual prompts, i.e. what they use the model for. The categories prompt length and setting refer to the number of sentences in the user inputs, and whether any examples were provided as part of the query. Engagement refers to the amount of exchange in the interaction. If the user prompted the model multiple times(at least twice) during the course of the conversation, we considered the interaction multi turn, otherwise single turn. The prompt's complexity gave us an idea whether users leverage ChatGPT for simple tasks, that they may otherwise quickly research using a search engine, or if they pose complex questions that require expert-level knowledge. The refinement degree of the prompt revealed if users were generally content with the initial answer of the LLM, or if further elaboration was needed. Finally, we differentiated use of formal and informal language. In general, when a single conversation consisted of multiple prompts, we labeled it based on the most frequently observed category or significant behavior.

We classified Midjourney interactions using a similar approach. First of all, we differentiated between image- and purely language-based inputs. We then considered the length of the prompt, and whether it consisted solely of keywords, one or more sentences, or a mix of both. Next, we recorded the complexity of the whole prompt. The Midjourney bot always generates four versions of the desired image. It then allows users to either recreate variations or more detailed versions of one or more of those four results. Users can also re-execute the whole generation process. We thus classified the prompt accordingly in the refinement category. Finally, we recorded the clarity of the prompt and satisfaction levels based on the observed user behavior. If a user created variations or more detailed versions of the result, we assumed they were generally satisfied. Analogously, we assumed dissatisfaction if they regenerated the whole image.

4 STUDY RESULTS

4.1 Findings and Observable Trends

In the following section, we are going to break down the sample analysis results by category. The distribution of the sub-categories

Table 2: ShareGPT Prompt Analysis Categories

Type	Intent	Length	Setting
Question Statement Command Task-based	Information Advice Clarification Opinion Suggestion Entertainment	Long (5 + Sent.) Med (2 - 4 Sent.) Short (<= 1 Sent.)	Zero Shot One Shot Few Shot
Engagement	**Complexity**	**Refinement**	**Language**
Single Turn Multi Turn	Simple Intermediate	None Once Complex	Formal Informal

Table 3: Midjourney Prompt Analysis Categories

Type	Length	Composition	Complexity
Language Image	Long (12+ Words) Med (4-12 Words) Short (1-3 Words)	Keyword-Only Sentence Mix	Simple Intermediate Complex
Refinement	**Language**	**Clarity**	**Satisfaction**
None Variation Regeneration	Formal Informal	Clear Ambiguous	Satisfied Dissatisfied Unclear

in the individual categories Prompt Type, Prompt Intent, Prompt Length, and Prompt Setting can be seen in Figure 1, whereas Figure 2 shows the numbers for the categories Engagement, Complexity, Refinement, and Language. In regard to prompt type, it became clear that the majority of users (54%) use Chat-GPT for task - based prompts, followed by questions (36%), commands (8%), and statements (2%). The most observed intents behind prompts were information gain (42%) and asking for suggestions (34%), followed by entertainment (10%) and advice (10%). Only few users were asking for clarification on a subject matter (4%). Interestingly, we did not observe any prompts where users actively asked the chatbot for its opinion (0%), which we initially had estimated as an at least fairly common use case. Prompt length was very evenly distributed, and we could not make out a clear preference of users. Short (38%), medium length (34%), and long (28%) prompts made up about a third of our samples each. We could clearly see the most often used prompting setting, however. The vast majority of users relied on a zero shot approach (90%), whereas only 8% used a one shot, and a mere 2% a few shot setting. Engagement in interactions was evenly distributed between multi turn (52%) and single turn (48%) conversations, meaning that almost half of the observed chats ended after the initial answer of the LLM. Most prompts were of a simple nature (48%), and slightly more than a third (38%) could be classified as intermediate, which left only 14% as complex. ShareGPT users only rarely refined their prompts multiple times (12%) or once (22%), leaving a two thirds majority (66%) of never refined queries. Finally, we could observe a tendency towards formal language (60%), which was used more often than informal language (40%).

Similarly, we analyzed the Midjourney data samples according to the predefined categories. The corresponding data and distribution for the categories Prompt Type, Length, Composition, and Complexity can be seen in Figure 3, for the categories Refinement, Language, Clarity, and Satisfaction in Figure 4. We already explained that Midjourney allows users to also prompt with an existing image as part of the input. However, only very few (8%) users have made use of this feature in our data sample. The majority (92%) relied on purely textual prompts. Interestingly, most queries were at least of medium length (54%), or even long (38%), and only 8% were classified as short. The composition of the individual prompts was well-balanced between sentences (42%), only keywords (26%), or a mix of both (32%). The same applies for the complexity. Most prompts were on an intermediate level (40%), closely followed by simple (38%) and finally complex prompts with a share of only 22%. Similarly to our ChatGPT samples, we observed only few refinements of Midjourney prompts. More detailed regenerations of images made up 14%, variations 10%, and the rest (76%) was not refined at all. A clear distribution could be seen in regard to formality of language. Users relied on formal language in almost all cases (94%), and only very rarely on more informally phrased prompts (6%). We identified 92% of all prompts as clear in their intention, which left only 8% as ambiguous. Satisfaction of users was unfortunately often unclear (58%), due to no apparent reactions of the users to the final image. However, for almost half of the samples we could either identify signs of satisfaction (26%) or dissatisfaction (16%).

5 DISCUSSION

5.1 Data Synthesis: Commonalities, Differences, and Possible Explanations

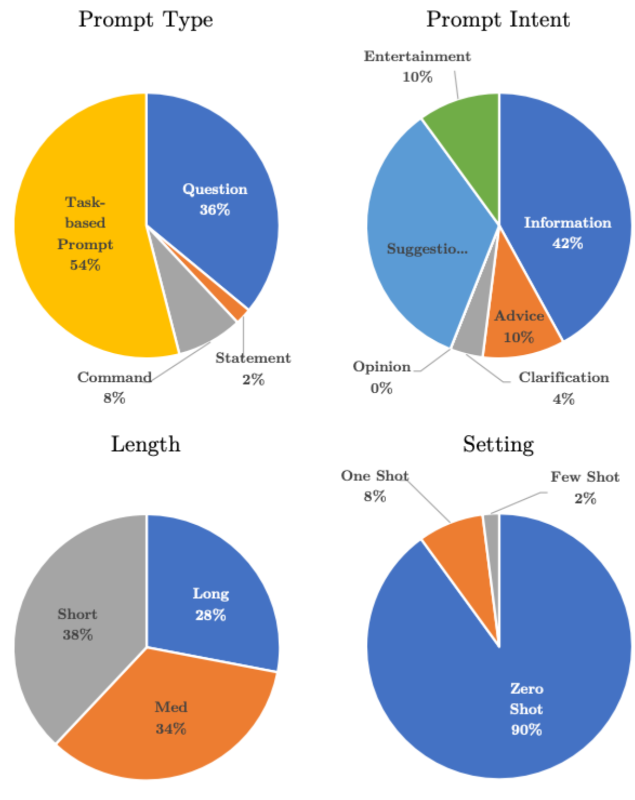

Figure 1: ShareGPT Prompt Analysis Categories 1–4

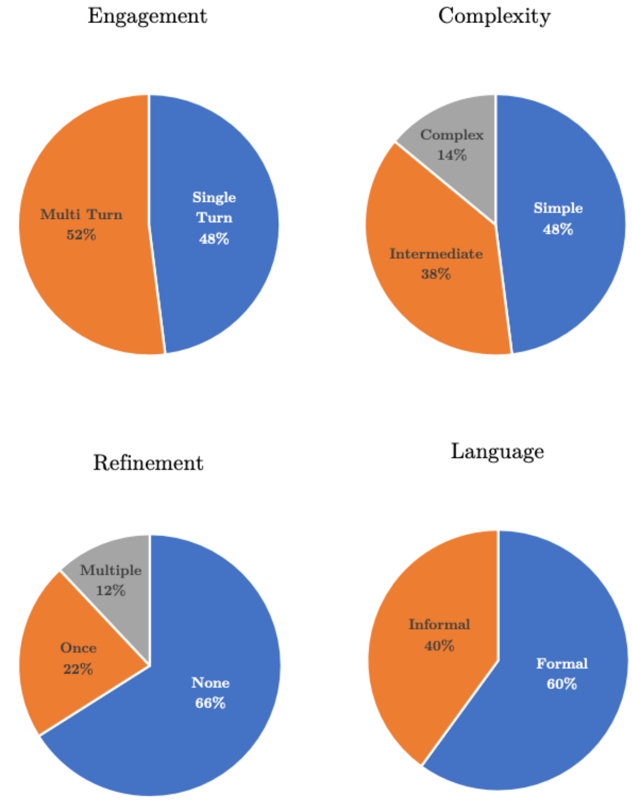

Figure 2: ShareGPT Prompt Analysis Categories 5–8

REFERENCES

[1] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. http://arxiv.org/abs/2005.14165 arXiv:2005.14165 [cs].

[2] DAIR.AI. 2023. Few-Shot Prompting | Prompt Engineering Guide. https://www.promptingguide.ai/techniques/fewshot

[3] Hai Dang, Lukas Mecke, Florian Lehmann, Sven Goller, and Daniel Buschek. 2022. How to Prompt? Opportunities and Challenges of Zero- and Few-Shot Learning for Human-AI Interaction in Creative Applications of Generative Models. http://arxiv.org/abs/2209.01390 arXiv:2209.01390 [cs].

[4] Amelia Glaese, Nat McAleese, Maja Trębacz, John Aslanides, Vlad Firoiu, Timo Ewalds, Maribeth Rauh, Laura Weidinger, Martin Chadwick, Phoebe Thacker, Lucy Campbell-Gillingham, Jonathan Uesato, Po-Sen Huang, Ramona Comanescu, Fan Yang, Abigail See, Sumanth Dathathri, Rory Greig, Charlie Chen, Doug Fritz, Jaume Sanchez Elias, Richard Green, Soňa Mokrá, Nicholas Fernando, Boxi Wu, Rachel Foley, Susannah Young, Iason Gabriel, William Isaac, John Mellor, Demis Hassabis, Koray Kavukcuoglu, Lisa Anne Hendricks, and Geoffrey Irving. 2022. Improving alignment of dialogue agents via targeted human judgements. http://arxiv.org/abs/2209.14375 arXiv:2209.14375 [cs].

[5] Google. 2023. Google. https://www.google.com/

[6] Jeff Huang and Efthimis N. Efthimiadis. 2009. Analyzing and evaluating query reformulation strategies in web search logs. In *Proceedings of the 18th ACM conference on Information and knowledge management*. ACM, Hong Kong China, 77–86. https://doi.org/10.1145/1645953.1645966

[7] Microsoft. 2023. Bing. https://www.bing.com/?cc=de

[8] OpenAI. 2023. ChatGPT. https://chat.openai.com/auth/login

[9] Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. Improving Language Understanding by Generative Pre-Training. (2018).

[10] Bernardino Romera-Paredes and Philip H. S. Torr. 2015. An Embarrassingly Simple Approach to Zero-Shot Learning. https://doi.org/10.1007/978-3-319-50077-5_2 Series Title: Advances in Computer Vision and Pattern Recognition.

[11] Mesay Samuel, Lars Schmidt-Thieme, D. P. Sharma, Abiot Sinamo, and Abey Bruck. 2022. Offline Handwritten Amharic Character Recognition Using Few-shot Learning. http://arxiv.org/abs/2210.00275 arXiv:2210.00275 [cs].

[12] ShareGPT. 2023. ShareGPT: Share your wildest ChatGPT conversations with one click. https://sharegpt.com/

[13] Kurt Shuster, Jing Xu, Mojtaba Komeili, Da Ju, Eric Michael Smith, Stephen Roller, Megan Ung, Moya Chen, Kushal Arora, Joshua Lane, Morteza Behrooz, William Ngan, Spencer Poff, Naman Goyal, Arthur Szlam, Y.-Lan Boureau, Melanie Kambadur, and Jason Weston. 2022. BlenderBot 3: a deployed conversational agent that continually learns to responsibly engage. http://arxiv.org/abs/2208.03188 arXiv:2208.03188 [cs].

[14] Statista. 2022. U.S.: generative AI adoption rate in the workplace by generation 2023. https://www.statista.com/statistics/1361174/generative-ai-adoption-rate-at-work-by-generation-us/

[15] Statista. 2023. Artificial Intelligence market size 2030. https://www.statista.com/statistics/1365145/artificial-intelligence-market-size/

[16] Romal Thoppilan, Daniel De Freitas, Jamie Hall, Noam Shazeer, Apoorv Kulshreshtha, Heng-Tze Cheng, Alicia Jin, Taylor Bos, Leslie Baker, Yu Du, YaGuang Li, Hongrae Lee, Huaixiu Steven Zheng, Amin Ghafouri, Marcelo Menegali, Yanping Huang, Maxim Krikun, Dmitry Lepikhin, James Qin, Dehao Chen, Yuanzhong Xu, Zhifeng Chen, Adam Roberts, Maarten Bosma, Vincent Zhao, Yanqi Zhou, Chung-Ching Chang, Igor Krivokon, Will Rusch, Marc Pickett, Pranesh Srinivasan, Laichee Man, Kathleen Meier-Hellstern, Meredith Ringel Morris, Tulsee Doshi, Renelito Delos Santos, Toju Duke, Johnny Soraker, Ben Zevenbergen, Vinodkumar Prabhakaran, Mark Diaz, Ben Hutchinson, Kristen Olson, Alejandra Molina, Erin Hoffman-John, Josh Lee, Lora Aroyo, Ravi Rajakumar, Alena Butryna, Matthew Lamm, Viktoriya Kuzmina, Joe Fenton, Aaron Cohen, Rachel Bernstein, Ray Kurzweil, Blaise Aguera-Arcas, Claire Cui, Marian Croak, Ed Chi, and Quoc Le. 2022. LaMDA: Language Models for Dialog Applications. http://arxiv.org/abs/2201.08239 arXiv:2201.08239 [cs].

[17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. 30 (2017).

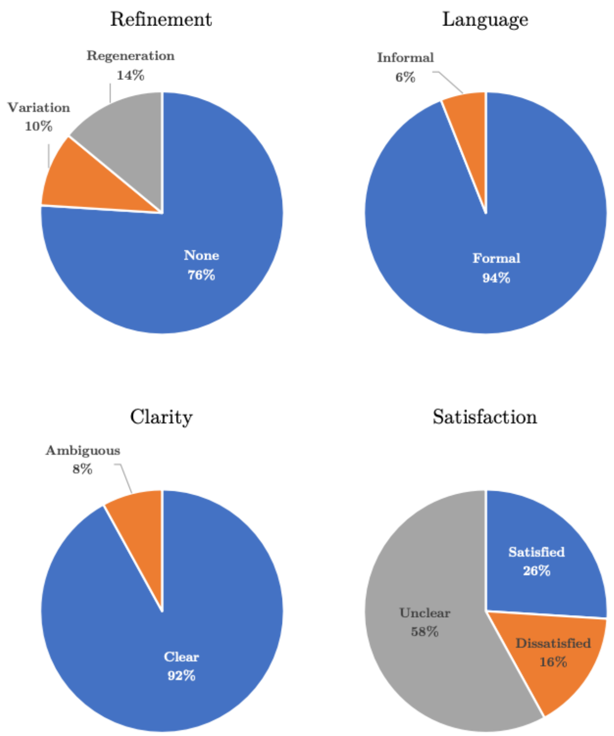

Figure 4: Midjourney Prompt Analysis Categories 5–8

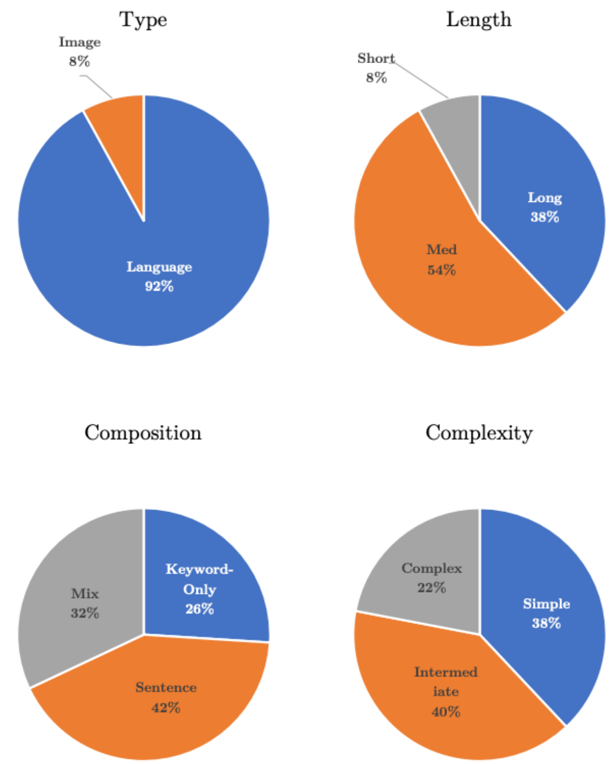

Figure 3: Midjourney Prompt Analysis Categories 1–4

[18] J.D. Zamfirescu-Pereira, Richmond Y. Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–21. https://doi.org/10.1145/3544548.3581388