



UiT The Arctic University of Norway

Department of Psychology, Faculty of Health Sciences

Advice from Humans and Artificial Intelligence: Can We Distinguish Them, and is One Better Than the Other?

An Investigation into the Transparency, Quality and Helpfulness of Life Advice Generated by Large Language Models (Text-Davinci-3, GPT-3.5-Turbo, GPT-4) and Online Advice (Human)

Otto J.B. Kuosmanen

Master's thesis in psychology (PSY-3900), May 2024



Foreword

Interesting times to be a student. I was inspired to embark on this project by happenstance. I had just completed a course on Python programming and was inspired to integrate programming into my research by Professor James Edgar McCutcheon. Around the time OpenAI released Chat-GPT, I was fascinated by the implications of this technological advancement. I saw this project as something relevant to the times and to my learning journey in combining psychology research with programming. Another interesting thing to note is that I aimed to only test one of the AI models but between the time I started my project, and the data collection two new models were released. This is partly why they were included as I wanted to test the latest models.

I'm grateful to Professor Tove Irene Dahl who supervised me on this novel topic and supported my unconventional approach, also big thanks for all the positivity. I want to thank all my fellow students for their support and friendship. UiT The Arctic University of Tromsø for the opportunity to study in such a calm place surrounded by nature. Last but not least I want to thank my parents, brothers and Celestyna for their love and support. It has been a wonderful experience.



OTTO KUOSMANEN



TOVE I. DAHL



Advice from Humans and Artificial Intelligence: Can We Distinguish Them, and Is One Better Than the Other?

Otto J.B. Kuosmanen

Supervisor: Tove, I. Dahl

PSY-3900

Master's Thesis in Psychology

Institute of Psychology (IPS) - Faculty of Health Sciences

UiT The Arctic University of Norway

Tromsø, May 2024

Sammendrag

Med utviklingen av kunstig intelligens (AI) har store språkmodeller (LLM) blitt tilgjengelige for alle. Folk vender seg i økende grad til AI for informasjon og råd. Det er lite kjent om kvaliteten på AI-ens utdata. Dette reiser presserende spørsmål om vi kan skille AI fra menneskelige ressurser, hvordan kvaliteten på innholdet sammenlignes, og om vi bør bry oss. Vi designet tre studier for å undersøke rådgivningsevner til AI-modeller. Identifikasjonsstudien undersøkte hvor godt folk kan skille mellom menneskelig og AI-råd. Preferansestudien evaluerte preferanseforholdet mellom menneskelig og AI-råd, og studien om rådgivningskvalitet vurderte den oppfattede kvaliteten på hver rådgivningskilde. Et web-skrapingskript ble laget med Python og brukt til å opprette en database med kvalitetsspørsmål-og-svar par hentet fra nettstedet Reddit. Alle modellene ble bedt om å svare på rådgivningsspørsmål som allerede var besvart av mennesker på Reddit. Vi fant at folk kan identifisere menneskelig råd over tilfeldighetsnivået, men ikke AI-råd. Vi fant at deltakerne foretrakk AI-råd over menneskelig råd. Når det gjelder rådgivningskvalitet, rangerte deltakerne AI-rådene betydelig høyere enn menneskelige råd. Konklusjonen er at AI-råd blandes godt med menneskelig råd og at det foretrekkes over topprangerte Reddit-råd. Disse funnene fremhever de siste forbedringene i LLM-er.

Nøkkelord: Råd, Hjelpsomhet, Oppfatning, Kunstig Intelligens, KI Deteksjon.

Abstract

The emergence of generative Artificial Intelligence (AI) has brought about the release of large language models (LLM) to the public. People are increasingly turning to AI for information and advice. Little is known about the quality of the AI output. This raises urgent questions regarding whether we can distinguish AI from human resources, how the quality of the content compares, and whether we should care. We designed three studies to investigate the advice-giving capabilities of AI models. The Identification Study investigated how well people can distinguish human and AI advice. The Preference Study evaluated the preference ratio of human and AI advice, and the Advice Quality Study assessed the perceived advice quality of each source of advice. A web-scraping script was made using Python and utilized to create a database of quality advice question-and-answer pairs extracted from the website Reddit. All models were prompted to answer the advice questions already answered by humans on Reddit. We found that people can identify human advice above the chance level, but not AI advice. We found that participants preferred AI advice over human advice. In terms of advice quality, participants rated the AI advice significantly higher than human advice. In conclusion, AI advice blends in with human advice and it is preferred over the top-rated Reddit advice. These findings highlight the recent improvements in LLM's.

Keywords: Advice, Helpfulness, Perception, Artificial Intelligence, AI Detection.

Advice from Humans and Artificial Intelligence: Can We Distinguish Them, and Is One Better Than the Other?

'Elves seldom give unguarded advice, for advice is a dangerous gift, even from the wise to the wise, and all courses may run ill.', (Tolkien, 1954, p. 84).

Seeking, and giving advice is a prevalent practice, possessing the power to shape decisions and, consequently, lives. As a phenomenon, advice is deeply integrated into our social practices and can be found in a vast amount of literature both in explicit and implicit forms (religious texts, fairytales). During our lives, we all will have experiences with giving and receiving advice, often in our daily encounters. In most problematic situations, individuals consult for advice, such as navigating relationship difficulties or making significant purchases (Decapua & Huber, 1995). The influence of advice on decision-making has been established (Meshi et al., 2012) and confirmed by a recent meta-analysis (Bailey et al., 2022), suggesting that on average, after seeking advice, people often change their decisions in ways that lean toward the advised suggestions.

The phenomenon of advice has probably been around since the dawn of man. But in a formal text-based format, advice seeking and giving was started in 1690s England, in the newspaper *Athenian Mercury* (Golia, 2021). Since then, people have been asking, giving, and reading advice given to others to this day. Now, with digitalized news sources, advice columns have mostly moved to the internet. A modern version of these newspaper columns can be found on the website Reddit. Here, an example is a thread under the name Advice

(www.reddit.com/r/Advice). It is active, with 960,000 followers (as of May 1, 2024), publishing an average of 651 advice requests per day (Kuosmanen, 2023; data retrieved May 1, 2024).

Now we are on the brink of a new change. The recent developments in Artificial Intelligence (AI) technology have led to new computational models called large language models (LLM). These models can answer complex questions, and they are now used by hundreds of millions of people (Neriya, 2024). Users can turn even more expeditiously to these models for advice than humans. However, it is not yet known what the quality of the advice these LLMs can give.

Given the influential role of advice in human decision-making, what are the implications of this AI development? Can people discern whether the source is human or AI? How does AI advice compare with human advice? Finally, does it matter? Before describing the design used to investigate these questions, we must first understand what is already known about the psychological nature of advice and how to study it, what is known about AI models, and their present use for generating advice.

The Nature of Advice

Its Etymology and Definition

What is its etymology, how is it used in everyday life, and how is it defined in research? Etymologically, the origin of the word “advice” can be traced back to the Sanskrit word “Veda” which signifies knowledge or understanding (Limberg & Locher, 2012,). This is suggested to have evolved into the Proto-Indo-European root “Weid” meaning “to see”. This term made its way into Latin as “Visum” and from here into Old French as “Vise”, which starts to represent not only seeing but also opinion and judgment. From here the term took its shape into Modern English as advice, with the addition of the prefix ad- from Latin, meaning toward (Harper, n.d.).

As suggested by the etymology, at its core, advice is about understanding, seeing, knowledge, and making a judgment to impart to others.

In modern times, a word is not limited by its etymology but by a clear and precise definition and how it is used in practice. Advice is utilized and studied across multiple fields and contexts that span from basic life advice to medical treatment decisions and governmental policies, leading to a whole spectrum of conceptualizations. We will work with the following definition, where advice is the recommendation for action in response to a problem (MacGeorge et al., 2018). According to this definition, advice is about formulating what someone else should do in response to a problem. Such a definition is clear and direct and provides a solid basis for discussing the act of giving advice.

However, this definition does not account for the advice's quality. Under this definition, advising is purely a form of communication, without necessarily requiring any of the core aspects of advice derived from the etymological origin. No understanding, knowledge, seeing, or judgment, is required. While the main purpose of seeking advice is to obtain assistance, the results of how to use or respond to the advice will depend on the quality of the advice and the ability of the receiver to act on it. Assessing the quality of advice therefore requires careful consideration. How can we determine if the advice we receive is any good?

The Study of Advice Effects

The Judge Advisor System (JAS) is a framework used to investigate the dynamics of advice-giving and exploring its impacts on decision-making (Sniezek & Buckley, 1995). It's designed to quantify the influence of advice, in an experimental context with information that can be expressed numerically. Here the decision-makers are conceptualized as judges who

decide how to incorporate advice from advisors. The experimental protocol unfolds as follows: First, the judge is asked a question, and their initial answer is recorded. For example, they might be asked to estimate the length of the Amazon River. The participant's initial response could be 5,500 kilometers, based on their knowledge and guesswork. Following this initial response, an advice answer is presented for the subject, stating that it comes from an advisor (further information about the advisor may or may not be provided depending on the purpose of the study). For instance, the guess might be followed with "advice" stating that the Amazon River may be 7,000 kilometers long. After receiving this advice, the subject is asked to provide a final answer to the original question. Influenced by the new information, the subject adjusts their estimate to 6,000 kilometers. The analysis of the results will reveal how much weight the advice has on the final answer. The weight of the advice is defined with a function that gives the proportion of the adjustment made by the subject relative to the adjustment suggested by the advisor.

$$\textit{Weight of Advice} = \frac{(\textit{Final Answer} - \textit{Initial Answer})}{(\textit{Advice Answer} - \textit{Initial Answer})}$$

$$\textit{Weight of Advice} = \frac{\textit{Adjustment made to initial answer}}{\textit{Adjustment to the initial answer suggested by the advisor}}$$

If we put in the values from the example above: Initial answer: 5,500. Advice answer: 7,000.

Final answer: 6,000.

$$\textit{Weight of Advice} = \frac{(6\ 000 - 5\ 500)}{(7000 - 5500)}$$

$$\textit{Weight of Advice} = \frac{500}{1500}$$

$$\textit{Weight of Advice} = 0.33$$

The weight of the advice is 1 if the subject adjusted their answer equal to the advised answer and 0 if the subject completely ignored the advice. In this case, the weight of the advice was 0.33, meaning that the subject instead of adjusting their answer exactly to the advised answer, adjusted their answer by 33% of the suggested adjustment. While we now know a system by which advice is mostly studied, let us have a look into some of the findings about advice and its influence on us.

The Influence of Advice

Recent research highlights the significance of advice and its influence on decision-making. A meta-analysis by Beiley et al. (2022) found that on average advice shifts decisions by 39 % as defined in the JAS. Although advice is generally influential, some aspects make people use advice less or more. For example, the same meta-analysis revealed that the perceived expertise of the advisor significantly affects how advice is utilized. When advice is framed to be given by expert advisors, the average weight of the advice increases to 55% and decreases for novice advisors to 32%. While, the expertise status of the advisor influences the weight of advice implementation, a curious finding has emerged, where even if the advisor is said to be completely incompetent, the advice they give influences the judges final estimate (Schultze et al., 2017), this finding suggest that even when knowing that advice is coming from an incompetent source it still can nudge our decisions.

In addition to the behavioral results, research has demonstrated distinct brain activations associated with receiving expert and novice advice. Specifically, the expectation of expert advice is linked to greater activity in the ventral striatum, an area that is linked with reward anticipation (Meshi et al., 2012). Findings also suggest that the advisor's expertise influences to what extent we use the advice and the judge's expertise in the subject matter. More knowledgeable judges tend to weigh advice less than judges who do not know much about the problem (Yaniv, 2004).

So, there is evidence that advice influences the decisions we make by shifting the decision towards the advised option. However, egocentric discounting may also play a role. Egocentric discounting involves not weighing others' advice as much as our own judgment. There is an argument that states that if we were fully rational, we should weigh the advice we get as much as our own when situations come by where we don't know anything about the advisor and we are not sure of the answer ourselves. This is based on the statistical principle that the average of multiple estimates is more accurate than a single estimate. The strategy of taking the average is often shown to be superior in cases that deal with numeric estimates, and it is also suggested to be a possible good solution in non-numeric problem-solving (Soll & Larrick, 2009; Yaniv & Kleinberger, 2000).

So, advice is influential, though it is unclear whether the results from the lab studies can be generalized for general advice-taking *in situ*. Nonetheless, it is assumed that the same mechanisms at play in the lab are also at play when we receive advice and make decisions in real situations. One can therefore expect that advice affects our real-life decisions, by nudging them towards the advice that was offered. Now that we know of the influential nature of advice and the variables that tweak this influence, how would we go about determining if the advice we receive is any good?

Determining the Quality of Advice

If we set out to ask for or give advice, how do we determine if the advice is any good? In other words, what determines the quality of advice? One way to determine whether the advice is of high quality is to equate quality with the objective accuracy of the advice. In this case, the question of whether the advice is good is simple: “Is it factually, correct?”. We can illustrate this with the example of the Amazon River. In this case, the advisor suggested that the Amazon River is 7,000km long, and this is indeed the case, therefore the advice is good, helpful, and of good quality. Overall, in the context of numeric advice, if the answer is correct and the judge adjusts their answer towards the advice, it will help get closer to the truth of the matter. It’s simple to say that a piece of advice is good or bad when there is a right answer. If the question has a factual numeric answer, then clear quantitative data can be stated about the correctness of the advice. For example, an advice answer can be exactly right or +10 % off.

Although, advice quality is mentioned in the JAS studies (Bailey et al., 2022), what is usually done is manipulating the perceived quality of advice by giving information about the advice source (Schultze et al., 2017). This is by giving information about the advisor’s past performance (Yaniv & Kleinberger, 2000), expertise (Birnbaum & Stegner, 1979) or confidence. An example of measuring the perceptions of advice quality can be found in a study where participants were asked how accurate and reasonable the advice is and to indicate that on a 7-point Likert scale (1=not accurate/reasonable, 7=very accurate/reasonable). The mean of these two items was taken as a measure of advice quality (Gino et al., 2012). While this approach as we just illustrated was taken from a study where the problems had true numerical solutions, a similar approach works for assessing the quality of advice in a setting where we deal with non-

numerical open-ended problems. As this is a subjective evaluation based on participants' perceptions, the objectivity of the advice is not a necessity.

In most daily situations, advice questions do not concern matters that can be resolved with numbers, nor, in most cases, will there be one correct answer that may be offered through language-based responses. People tend to ask for advice when situations are complicated and there is a lack of clarity about what should be done. For example, "What should I do? My neighbor vacuums at night." There are many possible ways to try to resolve the problem. So, how do we evaluate the quality of the advice, when there is not an objectively correct answer to the problem?

When a problem is complex and there are no simple solutions, determining the quality of advice becomes a subjective process of evaluating the advice based on advice quality perceptions. These perceptions are made by the judge, who ultimately decides how much they will rely on this advice. Jones and Burlison (1997), Goldsmith and MacGeorge (2000) and MacGeorge et al. (2004) have developed a measure that aligns with this approach. The measure consists of four items—Helpfulness, Appropriateness, Sensitivity, and Effectiveness—that participants rate using a Likert scale. The mean of the 4 items is suggested to measure advice quality. The sum score is reported to have very high internal consistency [$\alpha = 0.93$], thereby suggesting that they capture the underlying construct of perceived advice quality, at least on these terms (Bo Feng & Burlison, 2008).

Ideally, we would like to assess the consequences of advice taken, focusing on real-world outcomes. Although it is of vital importance to have standardized, valid, and reliable tools of measurement, given the currently available measures, this is not possible. Meanwhile, what we can do for now, is to investigate how people perceive the quality of any given piece of advice.

Advice Seeking

Advice seeking happens more outside of labs than in them. We have established that most people ask for advice when they encounter problematic situations (Decapua & Huber, 1995) and that advice is influential in changing initial options toward the advised suggestions (Bailey et al., 2022). The results also suggest that people rely more on advice when faced with subjective problems and when the complexity of these problems increases. When the problems are subjective, defined as lacking a single objectively correct answer, judges weigh advice more, than when the problem has a clear correct answer (Bailey et al., 2022). Also, when the problem is more complex the weight of advice increases (Gino & Moore, 2007). These findings would suggest that people would rely on advice especially when facing real-life problems, as they are subjective in their solutions and often complex.

Advice can potentially harm or benefit the person who follows it, depending on its quality. Knowing the expert status of the advisor and having insight into advice quality affects the weight with which we implement the suggested advice as would be expected (Yaniv & Kleinberger, 2000). Although bad advice still influences us, it does so to a lesser extent (Schultze et al., 2017), we should know these factors as they are important in maximizing the benefits and reducing the dangers of following advice.

Online Advice Seeking

Advice has moved to the Internet. The digital age has dramatically transformed the landscape of advice-seeking. Now it is possible to find and ask advice anytime, from anywhere in the world, if you are connected to the internet. People can do it anonymously and without being restrained to local communities (Golia, 2021). For example, many teachers are turning to social media sites (Kelly & Antonio, 2016) and online forums like Reddit for profession-related

advice (Carpenter et al., 2018). This digitalization of advice is the foundation of a revolutionary possibility, that of AI generated advice.

The emergence of the LLM, is a form of AI that can gracefully use language and answer questions, has led to a new source of advice that we know next to nothing about. Ensuring that AI-generated advice is from a recognizable source and of good quality is relevant and important for the future, as is establishing its status as a potentially ubiquitous advisor. Careful consideration of AI as an advisor needs to be conducted to maximize the benefits of this recent technology while minimizing the potential risks.

What do we know about the nature of AI today, particularly in terms of its use as a source of advice? We will start with the fundamentals of AI.

The Nature of Artificial Intelligence (AI)

AI has sparked human interest both in the past and the present. There are signs that humans have dreamed of creating machines that are intelligent for a long time. In Homer's Iliad, the master craftsman Hephaestus created machines that can think and talk (Silverman, 2022), this is just one of many myths that involve a AI-like phenomena (McCorduck, 2004).

The term Artificial intelligence itself comes from a workshop held in 1956 (Solomonoff, 1985) when it is said the field began its modern era, and the name was set in writing. At its core, AI is the idea of capturing intelligence into something outside of us, into an artificial artifact. How it's specifically defined depends largely on the subfield of AI, and the concept is changing and evolving in accordance to debate and technological advances (Russell et al., 2021). In our investigation, we will rely on a general definition of AI as the ability of a machine or computer

system to simulate and perform tasks that would normally require human intelligence (Morandín-Ahuerma, 2022).

The Imitation Game or the Turing test, was conceptualized by Alan Turing (Turing, 1950). It is a proposed benchmark to test if a machine can think. In this test, a human interrogator interacts with an unseen entity which could either be a machine or a human. They communicate using language and the interrogator must determine if the entity is human by asking questions. If the machine can convince the interrogator that it is human, then the machine is supposedly able to think.

Natural Language Processing is the branch of AI that focuses on enabling machines to process, interpret, and generate human language (Hirschberg & Manning, 2015). The Turing test has inspired the field of Natural Language Processing, as the core of the test revolves around the machine's ability to use language. For a machine to pass the test it needs to be able to seem to understand language and respond accordingly.

Artificial Intelligence, a Master of Games

Much progress has been made in the field of AI in recent years. Increases in computing power advances in algorithms, and the availability of large amounts of information have made this possible (Zhai et al., 2020).

Deep learning algorithms can reliably master different kinds of games. With only the rules of the game and self-play reinforcement loops, algorithms have been able to learn and master the games of Chess, Shogi and Go to the point of superhuman levels (Silver et al., 2018). While these were games of high complexity, that are renowned for their difficulty to (Hirschberg & Manning, n.d.) master, they are a long way from capturing the complexity of the everyday world. Recently, however, similar models have

learned more complex video games. In the game StarCraft 2, known for being one of the most challenging real-time strategy games in the world, a model was able to perform at the professional level (Vinyals et al., 2019). Also, in the game of Dota 2, AI defeated the esports world championship team (Berner et al., 2019). Mastering games are good benchmarks for AI, they pose an advanced challenge and can give insights into what is possible, but the long-term goal is to one day be able to solve advanced challenges that occur in the real world.

Language As a Game

In his book *Philosophical Investigations*, Ludwig Wittgenstein named the phrase language games (Wittgenstein, 1968). He laid forth the idea that language is not only one set of rules but consists of multiple so-called games. Wittgenstein concept of language games emphasizes the idea that meaning is not static in language, but the meaning of words is determined by their use and the meaning of words and expressions can change depending on their context. Because language use consists of multiple sets of rules that depend on the context, it makes a great challenge for AI to master. As games that consist of one set of rules, even complex ones can be mastered with AI, but creating an AI system that can recognize what language game is being played and adapt its responses is a challenge.

In the context of the Turing test, the machine must understand what language game the interrogator is playing by inferring it from the language, then adopting the rules of the game and responding accordingly. If, for example, the interrogator asks a question, the language exchange will take the form of a question-answer language game, but if the interrogator tells a joke, then the language use and response will need to change. In the context of jokes, the words often need to be interpreted non-literally, and a convincing response could involve matching the informal tone and demonstrating an understanding of the humor.

As with most AI models, the field of NLP started with rule-based algorithms or so-called expert systems. After natural language processing, AI adopted the deep learning approach and started training models based on existing text data. This resulted in improvements (Russell & Norvig, 2021), making the models more capable of adapting to the multiple different language games.

A recent paradigm shift in NLP introduced the transformer model. This drastically improved the technology. This change improved the efficiency of the LLM's by introducing self-attention mechanisms that improve contextual processing of the words in their given context, and parallel processing that improves the efficiency of processing data in the models (Vaswani et al., 2017). A prominent example is the Generative Pre-trained Transformer models (GPT).

Generative Pre-trained Transformer Models (GPT)

The first version of GPT was released in 2018; it utilizes the transformer architecture and unsupervised learning techniques (Radford et al., 2018). Unlike supervised learning, which relies on human-labeled data, unsupervised learning enables the model to autonomously discover the underlying patterns and structures within the data. GPT, specifically, is trained using a variant of unsupervised learning, called self-supervised learning, where the system generates its own labels from the input text data itself. This is achieved through tasks like predicting the next word in a sentence, which enables the model to learn language patterns and nuances without human guidance. During this self-supervised learning process, the model adjusts its parameters. These parameters are trainable weights that store the knowledge it acquires from the data. The

parameters are crucial as they determine the models behavior and increasing the number of parameters seems to be increase the quality of the models outputs (Brown et al., 2020).

Now, building on the foundational technology of the GPT, newer versions have been released. These new models are closing in on the mastery of human-like language. Chat-GPT, which was released free to the public, is a prominent case example of the new GPT models, while GPT 4 is the latest model as of 2023. These newer models have increased the number of parameters. While GPT-1 had 117 million parameters, GPT-4 now has 100 trillion. Furthermore, the amount of text data the models are trained on have also increased. While we know more about the training data on the older models for the new models much information has not disclosed (Yenduri et al., 2023). The three newest OpenAI language models up to this date are, Text-Davinci-3, GPT-3.5-Turbo and GPT-4.

Now that there are LLM's available for use, and millions of people are using them (Nerdynav, 2024), people might turn to these models for advice. Little is known about the quality of output in these models or about how people approach and utilize AI advice. It becomes important to establish the advising capabilities of these models as soon as possible and investigate more into the emerging field of AI advice.

Artificial Intelligence-Generated Advice

Experiments inspired by the Turing Test are still being run. Recently, OpenAI's Chat-GPT was tested in answering medical advice requests (Nov et al., 2023). How well could participants tell apart advice answers from doctors or AI? The results showed that the answers generated by Chat-GPT were identified as such 65% of the time (being mistaken as being from humans 35% of the time). This indicates that the advice answers were sometimes difficult to tell apart from doctors' responses. Still, they did not test the quality of the answers from either

source. The fact that they were often mistaken as answers from doctors suggests that the quality of the answers was somewhat comparable. The study concluded that Chat-GPT medical advice is weakly distinguishable from answers from medical professionals.

Striking results have emerged from two recent studies (Ayers et al., 2023; Liu et al., 2023), where GPT models were given the task of answering patients' questions regarding health issues. This can be conceptualized as medical advice. The medical advice from Chat-GPT was rated as being of higher quality than medical advice from actual doctors. While this is a striking fact, what makes it even more impressive is that the ratings were conducted by healthcare professionals. This suggests models such as Chat-GPT have advanced to a level that makes them possibly useful in the healthcare sector.

There are few experiments looking into how well LLMs can answer life advice questions and help solve basic-human problems. This is relevant because we think that as more people are starting to use these models and as people become more comfortable with these technologies, an increasing amount of people will turn to AI with advice requests. One study found that large LLM's failed to give helpful life advice. In the study they compared the best-rated online advice on Reddit, with advice generated by various AI models (Zellers et al., 2021). The best-performing model was chosen as more helpful than the human advice in only 14 % of cases. While the GPT 3 model sometimes produced harmful advice and was only chosen as more helpful in 4 % of cases.

People seem to be more affected by AI-generated advice than by advice from other people. This trend was even found where people preferred AI advice to their own judgment (Logg et al., 2019). This preference is termed algorithm appreciation. It has also been shown that the preference for algorithmic advice increases as complexity increases (Bogert et al., 2021).

While in some cases, using AI advice could be beneficial, there might be cases where it could be detrimental. After all, research shows that there are situations where algorithms outperform humans (Silver et al., 2018; Vinyals et al., 2019), but not in all domains. In the case of life advice, for example, existing research suggests that it would not be good to take all AI advice, at least that is what is suggested by Zellers et al. (2021). Even if the advice is explicitly mentioned to be worthless, there is evidence suggesting that the bad AI advice might still influence decision-making (Schultze et al., 2017). Just as human advice might be to the undiscerning advice-taker.

Design

The fast implementation of the new AI technology by individuals and companies has left in its wake a large knowledge gap. Novel and urgent questions need to be asked and answered. With hundreds of millions of individuals using Chat-GPT and other LLM's, many people might turn to it when seeking solutions or suggestions for real problems in the form of advice. People's ability to recognize it and evaluate it for quality must be investigated. Can LLM's give good advice, comparable to, or better than. human advice? This has not yet been documented in the scientific literature. If the models can perform at par with good quality online advice given by humans, then this marks an important milestone for AI. In that case, using AIs for advice would not be dangerous and could even be helpful. If the quality of the advice is bad, however, then it's important to highlight this fact and raise caution when using these models for generating advice for everyday life.

Research Questions and Experiments

The work presented in this thesis revolves around the answering of three key questions:

R1: Can people distinguish between human and AI-generated advice?; R2: Compared to human online advice popularly rated as good quality, how helpful is AI-generated advice regarded to be?; and R3: What is the quality of AI advice compared to good quality human online advice?

To answer R1: we designed the Identification Study -- an experiment where people indicate the source of advice, based on a binary choice of AI or Human. To answer R2: We designed the Preference Study-- an experiment where people choose between two pieces of advice (human and AI) and indicate which they find more helpful. This is a replication of Zellers et al. (2021). To answer R3: We designed the Advice Quality Study. —an experiment where people rate advice on a scale of advice quality.

There were two primary parts to completing this work. The first was developing the database to use in the questionnaire studies, and the second was the questionnaire studies themselves. Each will be presented separately below.

Generating the Database

As we set out to determine the quality of AI-generated advice and how well it can be distinguished from human online advice, we decided to make an advice database consisting of Advice Requests, Human advice, and AI advice. We decided to focus on the newest models of Open AI, as we wanted to have some insight into its incremental improvements and to see how far the newest models have advanced. So, we chose to gather advice from the models Text-Davinci-3, GPT-3.5-Turbo (Chat-GPT), and GPT4. The end goal was to have stimuli for our experiments in one database that could be used for multiple experiments and that could be shared with other researchers.

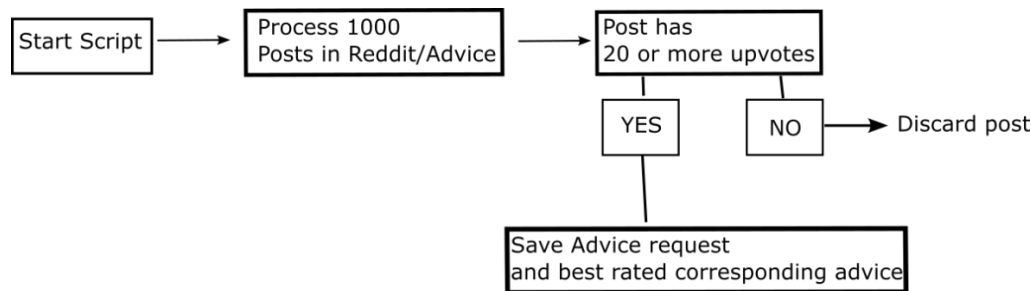
Human Advice Data

The first problem we wanted to solve was getting good quality human advice requests and advice to address them. Because we are in a digital age with so much information on the internet, we decided to use this to our advantage. Impressively many people ask for advice on the internet, so there is a massive existence of advice and advice requests available. We therefore decided to use Reddit (www.reddit.com/r/Advice) as a source. This is a site with almost a million members, and it has many active users and hundreds of new daily posts (Kuosmanen, 2023). With a good source for the human advice data, we then needed to find a way to sort based on the quality of the content. For this, we opted to use similar methods to Zellers et al. (2021), scripting a web scraping algorithm that would find the best-rated advice requests and advice.

A Python script was created that looked for advice posts on Reddit on the subreddit dedicated to advice (www.reddit.com/r/advice), all advice requests and pieces of advice given were in English) (see Figure 1). Posts on reddit are rated by users, who can either upvote or downvote posts, a cumulative rating is shown for all posts and responses. The script file processes 1000 of the newest posts. If advice posts had more than 20 upvotes they were saved along with the best-rated advice. We ran the script for 27 consecutive days. In total 27000 advice posts were processed by the script and only 325 posts matched the script criteria. In other words, only 1.2% of the posts were considered good enough to be included; this ensured that the quality of the advice requests and advice was good. Each post that passed the criteria was saved with the corresponding best-rated Reddit advice in JSON format. See appendix A for more information regarding the scripts connected to getting the human advice data.

Figure 1

Python Scrip Used to Process and Sort Advice Posts: Illustration.

***Censoring the Data***

A censor script was run on the questions. The script excluded questions that included the words: abuse, assault, rape, molest, suicide, and kill in all forms. These topics were determined unnecessarily intense for our research purposes. A sensor sweep was then run where the sensor checked for the words update and edit, as these are indications that changes have been made to the post, changes that often include information on a resolution of the advice problem or additional information that might have not been used while writing the best-rated advice answer. A minimum and maximum length of both advice request and advice answer were also set (Advice request: max 500 words, min 32 words; Advice answer: max 400 words, min 32 words). Before excluding the censored items, the database included 325 items. After the censor sweep, the database consisted of 202 advice requests.

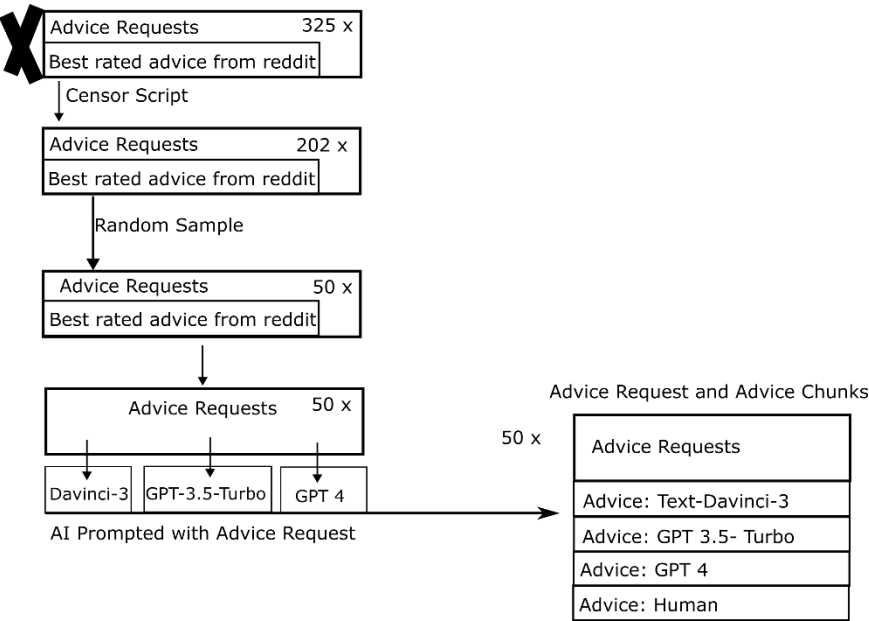
AI Advice Data

We wanted to gather advice for the requests from the AI models. We decided to limit the scope of our database, by taking a random sample of fifty advice samples from the database. This was

done with a Python script. Then we established connections to the AI models with application program interfaces (API) of the models, that were available through the OpenAI.com developer section. The AI models were tasked with responding to the fifty advice requests one at a time. Each request was processed individually using a structured prompt format. The prompt began with the tag ‘Question’: followed by ‘Advice request X’, where ‘X’ represents the sequence number of the specific advice request, from 1 to 50. This was immediately followed by ‘Advice:’ indicating where the model should insert its response. We stored all the advice given by the three models, resulting in 150 pieces of advice from the AI in total. Now the stimuli base consisted of 50 advice requests accompanied by 200 pieces of advice, 50 pieces of advice per source (50 x 4) (see Figure 2 below). For more information on the scripted procedure to gather the AI advice data see Appendix B.

Figure 2.

Script Sequence from the Advice Database to the Experiment-ready Stimuli Including the AI Advice.



Edits to the Final Stimuli.

We wanted to minimize or better avoid editing any of the answers, but still, we found some possible differences in the advice that we wanted to minimize. We found that the human answers were shorter and contained spelling errors. We, therefore, decided to take some measures to combat this, asking the AIs to write shorter answers to the questions and correcting the human spelling errors with the help of Grammarly. We also sorted and chose 15 answers from the sample of 50 that had the least difference in advice length. After we implemented this procedure, we had a sample of 15 advice requests, each paired with 4 different pieces of advice (Text-Davinci-3, GPT-3.5-Turbo, GPT-4, Human). We used this sample for the Preference Study and the Advice Quality Study. For the Identification Study, we removed 6 questions that contained self-referencing, thereby making the identification of the advice source evident. See Appendix C for information regarding pre-processing of the stimuli.

Experimental Methods

Data were collected in 3 studies where all participants received the same introductory information, and once they agreed to participate, they were randomly sent either to the questionnaire for The Identification study, the Advice Preference study, or the Advice Quality study. In that way, we are assured that the sample groups are similar in where they found the information to participate and likely more similar than different across groups. This section will begin with the information that was common to all groups, and then specify what was unique for each study.

Materials

Software.

In our project, we used the following software: Python, Qualtrics, OpenAI.

Python.

We used Python to read, structure, visualize, and analyze the data from the experiments.

The Python libraries we used included: Matplotlib, Pandas, NumPy, OpenAI, Json, Time, Seaborn, SciPy, Prawn, Datetime, Sys, Random, Docx. More information regarding the python scripts used in data analysis see Appendix J.

AI Models.

For the AI-generated advice we used the OpenAI application programming interface and the OpenAI website for access to GPT-4. Models were specified as: text-davinci-003, GPT-3.5-turbo and GPT-4. The models had not been trained on data that had been published later than 2021 September (Models - OpenAI API, n.d.).

Recruitment

Data collection began on November 25, 2023, and continued until December 16, 2023. We aimed to recruit a large and varied sample of participants. The only inclusion criterion was that participants must be over the age of 18 and be able to read English. We recruited participants from Reddit on a subreddit dedicated to research (www.Reddit.com/r/samplesize) Also, a poster was made with a link and QR code that sent potential participants randomly to one of the three studies (see Appendix E for Poster). We shared the poster on Facebook (www.facebook.com), (1) on the researcher's personal page and asked people to share it, and (2) on public Facebook groups, dedicated to research participation.

Participants

Demographics

A hub was created in Qualtrics to randomly send consenting participants to one of the three ongoing studies. In total 308 participants started the experiment, and 103 participants quit the experiment without answering any of the questions. We decided to include all participants in the analysis who had answered at least 1 of the questions, yielding 205 participants for the analysis (74 Identification Study, 66 Preference Study, 65 Advice Quality). Table 1 below shows relative parity among the groups in terms of number of participants, age, education, and home country in each study. This provides a sound foundation for making cursory comments about what can be learned from all three groups together in the final synthesis of the data.

Table 1

Participant Demographics in All Studies

| | Identification Study | Preference Study | Advice Quality Study |
|----------------------|----------------------|------------------|----------------------|
| Participation | | | |
| Started | 106 | 103 | 99 |
| Completed | 48 | 49 | 50 |
| Partially Completed | 26 | 17 | 15 |
| In Analysis | 74 | 66 | 65 |
| Gender | | | |
| Male | 28 | 26 | 23 |
| Female | 39 | 37 | 39 |
| Other | 6 | 0 | 3 |
| Missing | 1 | 3 | 0 |
| Age | | | |
| 18 – 24 | 19 | 17 | 18 |
| 25 – 34 | 31 | 31 | 25 |
| 35 – 44 | 4 | 7 | 7 |
| 45 – 54 | 12 | 8 | 11 |
| 55 – 64 | 3 | 0 | 4 |
| 65 – 74 | 4 | 3 | 0 |

| | | | |
|---------------------|----|----|----|
| 75 – 100 | 0 | 0 | 0 |
| Missing | 1 | 0 | 0 |
| Education | | | |
| (Highest) | | | |
| Primary – School | 1 | 4 | 0 |
| High – School | 13 | 16 | 11 |
| Bachelors | 36 | 20 | 39 |
| Masters | 17 | 21 | 12 |
| Professional Degree | 5 | 3 | 2 |
| Doctorate Degree | 1 | 1 | 0 |
| Other | 1 | 1 | 1 |
| Missing | 0 | 0 | 0 |
| Home Country | | | |
| Finland | 34 | 29 | 18 |
| USA | 10 | 10 | 17 |
| Norway | 5 | 4 | 3 |
| Poland | 4 | 4 | 3 |
| England | 4 | 0 | 6 |
| Canada | 0 | 4 | 3 |
| Australia | 2 | 3 | 0 |
| Spain | 1 | 2 | 1 |
| Sweden | 2 | 2 | 0 |
| Hungary | 1 | 1 | 0 |
| China | 0 | 1 | 0 |
| Greece | 0 | 1 | 0 |
| The Netherlands | 2 | 1 | 3 |
| Italy | 1 | 1 | 0 |
| Belgium | 0 | 1 | 0 |
| Cyprus | 0 | 1 | 0 |
| Ireland | 0 | 1 | 0 |
| Denmark | 1 | 0 | 0 |
| Switzerland | 1 | 0 | 0 |
| Serbia | 1 | 0 | 0 |
| Austria | 0 | 0 | 2 |
| India | 0 | 0 | 1 |
| Iceland | 0 | 0 | 1 |
| Brazil | 0 | 0 | 1 |
| Vietnam | 0 | 0 | 1 |
| Germany | 0 | 0 | 1 |
| Malaysia | 0 | 0 | 1 |
| New Zealand | 0 | 0 | 1 |
| Missing | 1 | 0 | 2 |

| Recruited via | | | |
|----------------------|----|----|----|
| Social media | 31 | 28 | 20 |
| Reddit | 27 | 27 | 33 |
| Shared | 12 | 10 | 9 |
| Other | 3 | 1 | 3 |
| Missing | 1 | 0 | 0 |

Identification Study

Materials

Items: Advice Requests and Advice. Our items for the experiment consisted of 9 advice requests from the advice database. These requests are in the domain of General Advice and cover topics ranging from: Relationship Advice, Economic Advice, and Parenting Advice. For all these requests, we had advice from the four sources (Human-Reddit, Text-Davinci-3, GPT-3.5-Turbo, and GPT-4). The Human advice length varied from 36 to 126 words ($M=82$, $SD=30$). The advice from Text-Davinci-3 varied from length 72 to 221 words ($M=159$, $SD=43$). The GPT-3.5-Turbo advice varied from 71 to 189 words ($M=117$, $SD=46$). The GPT-4 advice length varied from 55 to 160 words ($M=129$, $SD=33$). See Appendix D for all advice requests and related pieces of advice included in this study.

Procedure

The experiment was conducted using the Qualtrics XM platform (*Qualtrics XM*, n.d.). Participants would read the information letter and indicate their consent to participate. If they consented, they were sent to the Identification questionnaire. Here they first answered several demographic questions (Age, Education, Country, Gender, Source) after which they received the instructions for the study itself. For that, they were presented with a series of random Reddit advice requests from the sample accompanied by randomly chosen advice from one of the four

sources (Text-Davinci-3, GPT-3.5-Turbo, GPT-4, Human). They were then asked to indicate with a binary choice if they believed the advice to come from AI or a Human.

For example, let's say they received Advice Request 5: "I'm 15. I have a job, but I'm not saving for anything in particular yet. Might save for a car when I turn 16. Is it a waste of my money to spend it on things that don't benefit me in the long run, like video games?" With it, they receive advice, in this instant from Text-Davinci-3: "No, it is not a waste of your money to spend it on things like video games. Even if such purchases do not benefit you in the long run, they are still a form of temporary pleasure that can make you feel good and be rewarding in the present. Saving for a car or any other goal is important, but it can also be helpful to have some fun and enjoyment with your money too." After reading both the advice request and the advice, the participant had to indicate whether they believed that the advice came from a human or an AI by answering "Indicate your best guess as to whether the answer was produced by a human or AI". After indicating a choice, a new random advice request would be presented with advice from one of the four sources. This process was repeated 9 times total, after which the participant was prompted to give any feedback, they might like to share before the experiment ended.

No monetary reward was given for participation. We aimed to have the experiment last a maximum of 15 minutes on average. The average time it took for participants to fully complete the study was 13 minutes, excluding participants who took over 2 hours. The data reported from this study were the correct identification rates for each of the advice sources (Human, Text-Davinci-3, GPT-3.5-Turbo and GPT-4). See Appendix F for the information letter, consent form, and instructions used for this study.

Advice Preference Study

Materials

Items: Advice Requests and Advice. Our items for the experiment consisted of 15 advice requests from the advice database. The advice requests are in the domain of general advice. The requests covered the following topics: Relationship Advice, Economic Advice, Legal Advice, Parenting Advice, Health Advice. An example is: What to do when a parent is diagnosed with Alzheimer's disease? For all these requests we had advice from the four sources (Human-Reddit, Text-Davinci-3, GPT-3.5-Turbo, and GPT-4). The Human advice length varied from 36 to 256 words ($M=128$, $SD=67$). The advice from Text-Davinci-3 varied from length 72 to 283 words ($M=186$, $SD=57$). The GPT-3.5-Turbo advice varied from 71 to 240 words ($M=133$, $SD=53$). The GPT-4 advice length varied from 55 to 245 words ($M=142$, $SD=42$). See Appendix D for all advice requests and related pieces of advice included in this study.

Procedure

We surveyed with Qualtrics. In this survey participants would randomly receive 6 questions. We chose only 6 trials because we wanted the experiment to take less than 15 minutes and therefore keep our participants engaged without dropping out. Each question was one of the Reddit advice requests, and with it came two pieces of advice: the human Reddit advice and a randomly selected AI answer (note that randomly selected AI answers were evenly distributed so that they were approximately equally represented in the analyses). The participants were asked to choose the advice which they thought was more helpful. After the 6 trials, participants were prompted to give any feedback they might like to share before the experiment ended.

No monetary reward was given. The average time it took for participants to fully complete the study was 12 minutes, excluding participants who took over 2 hours. The data produced by this study were the preference ratios for each of the AI advice sources (Text-Davinci-3, GPT-3.5-Turbo and GPT-4) against the human advice. See Appendix H for the information letter, consent form, and instructions.

Advice Quality Study

Materials

Items: Advice Requests and Advice. Our items for the experiment consisted of 15 advice requests from the advice database. These items were identical to those in the Advice Preference Study. See Appendix D for all advice requests and related pieces of advice included in this study.

Advice Quality Scale: In the Advice Quality Study: we used a perceived advice quality measure with 4 items (helpfulness, effectiveness, appropriateness, and sensitivity) rated on a 7-point Likert type scale, where 1 and 7 were anchored by descriptors corresponding to the item (e.g., for helpfulness, 1 = Not Helpful, 7 = Very Helpful). Overall perceived advice quality was estimated using the mean of the 4 items. (Goldsmith & MacGeorge, 2000; Jones & Burleson, 1997; MacGeorge et al., 2004).

Procedure

Participants would receive a random advice request with advice from one of the sources (human or Text-Davinci-3, GPT 3.5-Turbo or GPT-4). The advice was chosen randomly so that ratings would be approximately evenly distributed. The participants were instructed to rate the advice on a 7-point Likert scale on 4 advice quality items. As we aimed to make the experiment last less than 15 minutes we repeated this procedure 6 times. No monetary reward was given.

The average time it took for participants to fully complete the study was 13 minutes. Participants who took over 2 hours to complete the form were excluded. The data produced by this study were the quality scores for each of the advice sources (human, Text-Davinci-3, GPT-3.5-Turbo and GPT-4). See appendix I for the information letter, instructions and consent form used in this study.

Results

We chose to conduct Chi-Square Goodness of Fit tests for the Identification study and the Preference study as we were working with categorical data. In the Advice Quality study, we anticipated ordinal data with a non-normal distribution, particularly due to our method of exclusively capturing the highest-quality answers. This expectation held for both human-provided advice and AI-generated ratings, resulting in a leftward skew in the data distribution. Given this deviation from normality, we decided to employ non-parametric tests to analyze the data. We, therefore, used the following: Kruskal Wallis Test, and Man-Whitney-U-test. In all analyses, the number of observations is depicted as N. See appendix J for more information on the scripts used for statistical analysis.

Identification Study

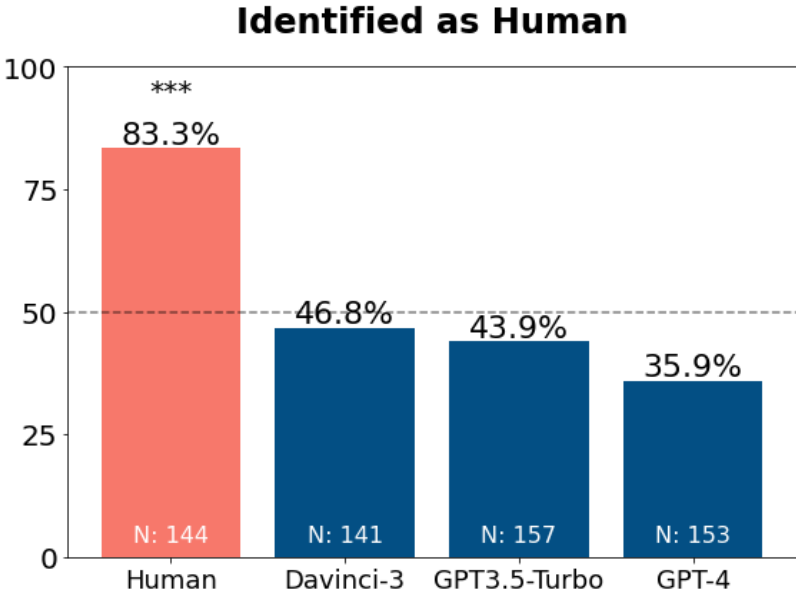
In this study, we conducted five Chi-Square tests. Investigating if participants could identify the source of the advice correctly (AI or Human) significantly above a chance level of 50%. We tested each group separately: Human, Text-Davinci-3, GPT-3.5-Turbo, GPT-4, and AI (data pooled values from all AI models). Due to multiple comparisons, we used the Bonferroni correction and adjusted the alpha level from .05 to .01 ($0.05 / 5 = 0.01$).

Analysis of participant responses indicated a statistically significant ability to correctly identify human-generated advice $\chi^2 (1, N = 144) = 34.52, p < .001$ (see Figure 4). But, when examining the aggregated AI advice, the trend toward correct identification did not reach statistical significance after the Bonferroni correction, $\chi^2 (1, N = 451) = 5.31, p = .021$, with an adjusted significance threshold of .01. Specifically, for individual AI models, participants ability to correctly identify AI-generated advice did not significantly differ from chance levels: Text-Davinci-3, $\chi^2 (1, N = 141) = 0.17, p = .676$; GPT-3.5-Turbo, $\chi^2 (1, N = 157) = 0.92, p = .336$; and GPT-4, $\chi^2 (1, N = 153) = 5.60, p = .018$. For the GPT-4 advice, there was a trend towards correct identification greater than chance, though that did not meet the adjusted significance criterion. These findings suggest that, after adjusting for multiple comparisons, participants could reliably distinguish human advice from AI-generated advice, but the ability to identify specific AI models as the source of advice more often than chance did not reach significance, indicating that only human advice was reliably identified correctly (Figure 4).

Despite the trends, none of the AI models reached significance after the correction. Indicating that only the human advice was reliably identifiable as such.

Figure 4

The Distribution of Correct Identifications Grouped by Advice Source.



Note: Depicts the percentage of identifications of the source of advice as human. The horizontal dashed line at 50% indicates the chance level for correct identification. Each bar represents the identification rate as human for each source. The three asterisks to denote a p-value less than .01, indicating statistical significance. The number of observations per group is noted as N.

Advice Preference Study

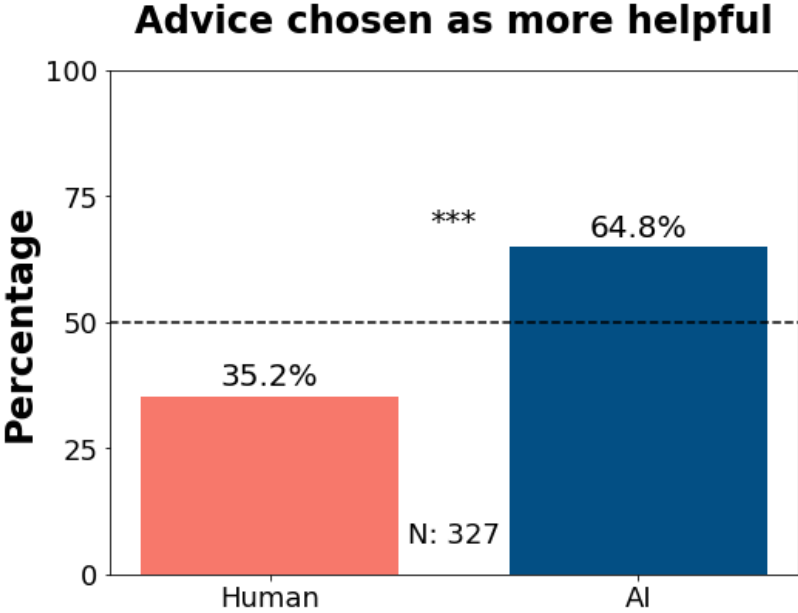
For the study, we analyzed observations from 66 participants. The data is based on 327 observations. In this study, we conducted four Chi-Square tests, investigating if participants preferred AI or Human advice in a way that significantly differed from a 50/50split. We tested each group separately regarding preference ratio: Text-Davinci-3, GPT-3.5-Turbo, GPT4, and AI (data pooled values from all AI models). Due to multiple comparisons, we used the Bonferroni correction and adjusted the alpha level from .05 to .0125 (0.05 / 4 = 0.0125).

Analysis of participant advice preferences indicated a statistically significant preference towards AI-generated advice over human advice, $\chi^2 (1, N =327) = 28.77, p < .0001$ (see Figure

5). When examining preferences for each AI model individually against human advice (see Figure 6), results demonstrated a statistically significant preference for AI advice from Text-Davinci-3, $\chi^2 (1, N=105) = 13.04, p < .001$, and GPT-3.5-Turbo, $\chi^2 (1, N =111) = 12.33, p < .001$ (see Figure 6). Although GPT-4 advice was also preferred, $\chi^2 (1, N=111) = 4.77, p = .029$, it did not meet the Bonferroni adjusted significance threshold of $p < .0125$, indicating a trend towards significance that falls short of the corrected criteria for statistical significance.

Figure 5

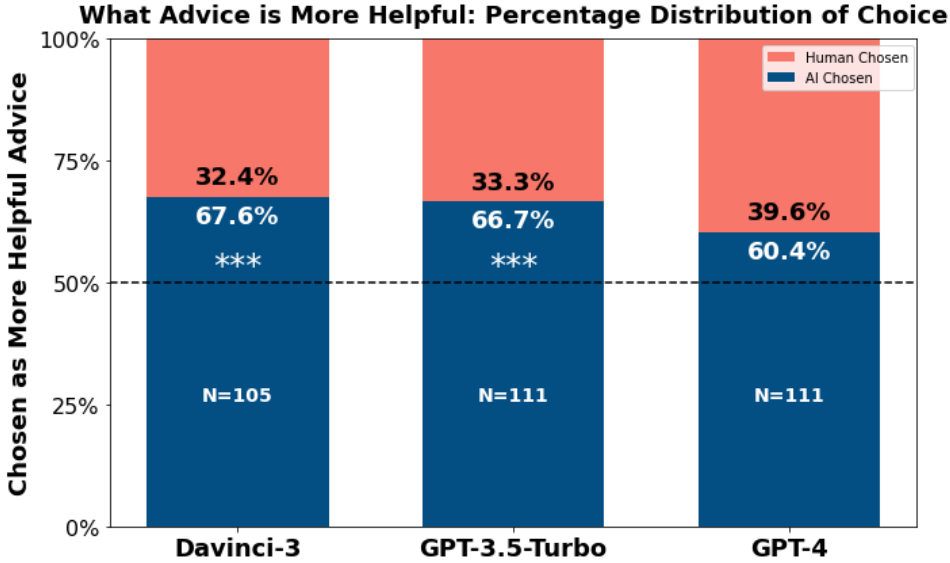
Total Distribution of Advice Preference



Note: Pooled AI advice (AI) and Human advice ratio of being chosen as more helpful than the other. Significant preference was found for AI ($p < .0125$). N is denoting the number of observations.

Figure 6

The Ratio of Advice Preference Grouped by the Source of Advice



Note: Chosen as more helpful advice, ratio of AI to Human choices for each model. Significant preference found for Text-Davinci-3 and GPT-3.5-Turbo ($p < .0125$). N is denoting the number of observations per group.

Advice Quality Study

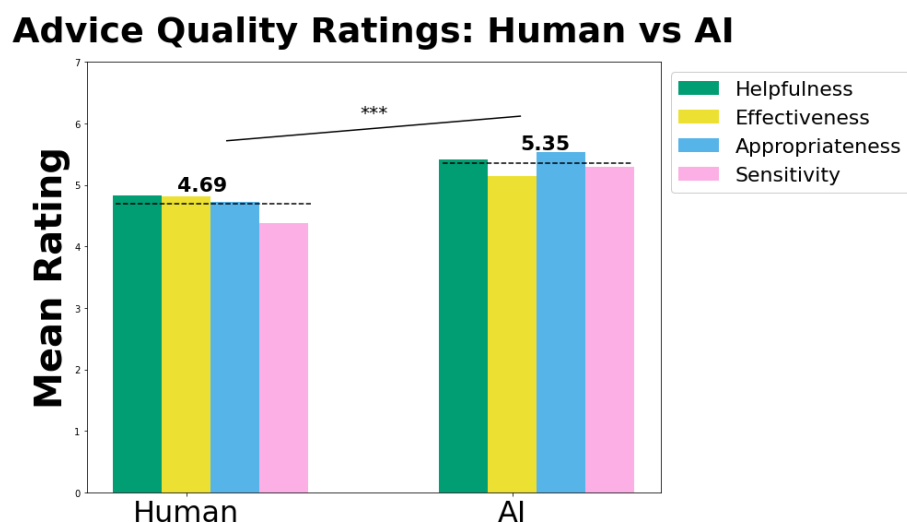
For the study, we analyzed observations from 65 participants. We compared the advice quality in the four groups (Human, Text-Davinci-3, GPT-3.5-Turbo, and GPT4). We were also interested in a general comparison between humans and AI, therefore we pooled the AI ratings to create a general AI category (AI Total). We conducted a total of 13 tests, due to multiple comparisons, we used the Bonferroni correction and adjusted the alpha level from .05 to .0125 ($0.05 / 13 = .0038$).

A Mann-Whitney U test was performed to evaluate whether advice quality differed between the human and AI, we pooled the AI ratings to create a general AI category. The results indicated that AI advice was rated significantly higher in quality than human advice, $z = -6.32$, p

< .00001 (Figure 7). As we wanted to know if these differences were found in all items in the Advice Quality scale, we performed four additional Mann-Whitney U tests. Our results indicate that AI-generated advice is rated significantly higher in terms of helpfulness ($z = -3.06$, $p = .0016$), sensitivity ($z = -4.11$, $p < .00003$), and appropriateness ($z = -3.70$, $p < .0001$). However, not in terms of effectiveness ($z = -1.70$, $p = .082$)

Figure 7

Advice Quality Ratings for Human and AI Advice (Pooled)



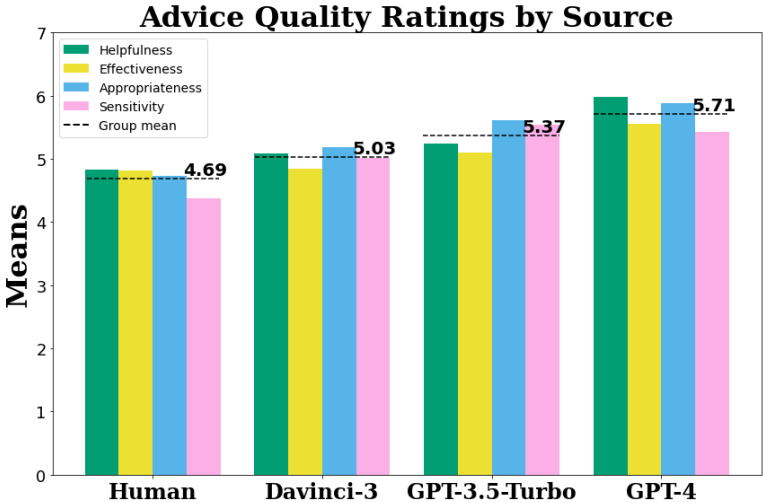
Note: The mean advice quality is shown at the top of the bars (Human = 4.69, AI = 5.35) this is the cumulative mean of all the items (Helpfulness, Effectiveness, Appropriateness and Sensitivity). A significant difference was found between Human Advice quality ratings and AI advice quality ratings ($p < .0038$).

Further analyses were conducted to compare advice quality across individual AI models against human advice using Mann-Whitney U tests. The comparison between Text-Davinci-3 and human advice did not reveal significant differences ($z = -2.73$, $p = .00549$), suggesting comparable levels of advice quality between Humans and Text-Davinci-3. However, significant differences

were observed in the comparisons of Human advice and advice from GPT-3.5-Turbo ($z = -5.21, p < .00001$) and GPT-4 ($z = -7.83, p < .00001$), indicating a marked preference for the advice quality provided by these AI models.

In addition to the Mann-Whitney U tests, we performed more exploratory tests investigating the differences between the AI models (Figure 8). A Kruskal-Wallis test was conducted to evaluate the differences in advice quality across three AI models. The test revealed a significant difference, $\chi^2(2, N = 248) = 31.57, p < .0001$, suggesting that advice quality differences are present between the models. To further investigate in what domain of advice quality these differences were, we performed additional Kruskal-Wallis tests comparing the differences of the items means across the AI models. Though Helpfulness stood out with a significant difference among the AI models $\chi^2(2, N=248) = 17.53, p = .00016$ suggesting that perceptions of helpfulness vary significantly across the different AI models. No significant differences were observed in effectiveness, $\chi^2(2, N=248) = 8.96, p = .011$, appropriateness $\chi^2(2, N=248) = 6.87, p = .032$, nor sensitivity, $\chi^2(2, N=248) = 5.90, p = .052$

Figure 8
Advice Quality ratings by all sources



Discussion

Overall, participants could not identify any AI advice at a level better than chance; only human advice was identified correctly significantly above chance levels. Nevertheless, in general, AI advice was preferred to human advice. When asked to choose the more helpful advice, participants also chose AI advice significantly more than human advice. When examining the perception of each specific model in our data, curiously, the newest model GPT-4 was not preferred to human advice, while the older models Text-Davinci-3 and GPT-3.5-Turbo were, and significantly so. Meanwhile, our results also suggest that AI advice was nevertheless perceived by participants as being of greater quality overall – perceived as being more, helpful, sensitive, and appropriate than human advice, though not as more effective. When examining the perception of each specific model in our data, we found that Text-Davinci-3 the oldest of the models tested, did not differ significantly from human advice in terms of advice quality, while the models GPT-3.5-Turbo and GPT-4 did, and significantly so. A deeper dive exploring the specific perceptions of the AI model advice quality revealed significant differences between models. These differences were found specifically in their perceived helpfulness.

AI Identification.

The AI field is growing at breathtaking speed, and after the introduction to this thesis was drafted and the research design set, more research in this area was published. How do our results compare considering those newer findings?

A study that was published during our data collection found similar results to ours. In that study, Howe et al. (2023) presented each participant with advice request, accompanied by one piece of advice from a professional advice columnist and one advice from GPT-4. Their results

indicated that participants could not identify advice from GPT-4 at levels above chance. The study included 387 US participants with a mean age of 42 years and with approximately equal split in men and women. They also had participants rate the quality of the advice, but this part of the study will be discussed later in the discussion.

On the other hand, the results of Howe et al. (2023) and the results of our study are not in line with a previous study, where medical advice from generated by Chat-GPT and medical advice written by doctors were both weakly distinguishable (Nov et al, 2023), in a US sample of 392 participants with a mean age of 47 and an equal distribution of men and women. Interesting to note, is that participants had an economic incentive to identify the source of the advice correctly. Another puzzling contradiction is that the human advice was not distinguishable in Howe et al. (2023), while it was distinguishable in our study and the study by Nov et al. (2023).

These contradictions present a complicated picture. Taken these findings, together it seems that AI is hard to distinguish, two out of the three studies found this to be true. It is a possibility that because of the economic incentive to identify the advice correctly in Nov et al. (2023) participants were paying extra attention to the task and by doing this were able to distinguish the AI advice from the human. Also, from the three studies it was the only study that dealt with medical advice. It is possible that there is a clearer distinction between doctor advice and AI advice.

Our study was the only one that found human advice easier to distinguish than AI advice. Establishing this conclusion as a fact is not that clear, as methodical differences make a comparison between the studies less convincing. In our study, we always presented participants a total of 6 advice requests accompanied each by one piece of advice that was randomly taken from the source pool. We did not state any information about how many of the answers are AI.

This made each advice source guess independent from the next. While both Nov et al. (2023) and Howe et al. (2023) did inform the participants of the number of human and AI answers, thereby making it in a sense easier to distinguish the source. For example, in the case of Howe et al (2023), two advice was presented, and participants had to indicate the AI advice. If you know which one is human, then you can infer which one is AI. The same is somewhat true for the Nov et al. (2023) study, here participants would receive ten advice requests accompanied by one advice each (randomly either AI or human), they were given the information that five out of the ten advice was written by an AI.

One other explanation seems intuitive, that advice from Reddit may be written in a more casual way than advice from professional advice columnists. This observation could explain why participants in our study were better at identifying advice when it came from human sources but not in the study of Howe et al. (2023). It suggests that the source of advice may influence its recognizability. Although it still cannot explain how participants could identify both AI and doctor advice in the study by Nov et al. (2023).

An interesting trend in our data is also worth mentioning, that we found a non-significant trend that suggests that newer models might be easier to identify as AI than older models. Why might this be? An analogy could be drawn to the chess AI engines. Now that AI is superior to humans in chess, the AI source might be easier to identify. This is plausible, even though we did not find research on the topic. An interesting idea emerges, that when AI is inferior to humans in a task it is easily identifiable. When it is on par with human output it's harder to distinguish, and then when it is superior it again becomes easier to distinguish. This is worthy of further research.

To fully understand how recognizable AI advice is to participants, more Turing type tests should be conducted. Although the evidence is not uniform, it seems that the new GPT models are hard to distinguish from human advice. As of now, all the differences between studies make clear conclusions hard to draw.

AI Advice Preference

Our research was initially inspired by the sudden leap in technology, which presented itself in the form of OpenAI's release of Chat-GPT. At the time we started our investigation, we did not find much research on the topic of AI advice. The study that stood out at the time was a study on the helpfulness of AI advice by Zellers et al. (2021), they were testing the best LLM's at the time, including GPT-3. As the human advice comparison, they used Reddit-top-rated advice drawn from Reddit. This is the approach that inspired our Advice Preference study and we opted to use a similar Python scripting method and advice comparison methodology. While our study aimed to be a replication of Zellers et al. (2021), our results were strikingly different. Contrary to Zeller's findings where GPT-3 advice was preferred only in 4 % of cases to human advice from Reddit, and even rated to be dangerous in 10 % of cases, we found a clear preference for AI advice. While we did not test the same models, we tested the newest models from OpenAI. Our results indicate that a clear and notable improvement has been made in the technology of LLMs.

One curious finding though, was that while the models Text-Davinci-3 and GPT 3.5-Turbo were preferred over Human advice, no significant preference was found for GPT-4. The fact there was less preference for the most advanced model was not expected. One possibility for this finding is that only the GPT-4 advice was restricted in length as at the time this was the only

model that we determined to successfully follow the length requirements without sacrificing quality. As of now, we have a hard time explaining this finding as it is contrary to our expectations, our results in the Advice quality study, and other research (Howe et al., 2023). While the GPT-4 advice was not significantly different from human advice when Bonferroni corrected, a clear trend was present in people preferring the GPT-4 advice that would have been significant without the correction.

Advice Quality

Before our study, we did not find any research that investigated AI advice with the advice quality measure (Goldsmith & MacGeorge, 2000; Jones & Burlison, 1997; MacGeorge et al., 2004). Our data showed AI advice to rate higher than human advice in terms of helpfulness, sensitivity, and appropriateness, but not in effectiveness. During our investigation, another study was published that confirms our findings that people perceive AI advice to be of better quality than human advice (Howe et al., 2023). However, Howe et al. (2023) used a different measure of advice quality than the measure used in our study. They gauged the participant's preferences for the advice in terms of helpfulness, balance, empathy, completeness, and simply which advice is better. They found that on all the qualities, participants preferred AI advice. Also, instead of comparing AI advice to human advice from online sources, they used professional newspaper advice columnists.

There is further corroboration of our findings from other recent studies where medical advice provided by GPT-3.5-Turbo and GPT 4 models were rated by doctors as more accurate, useful, empathic, and responsive than medical advice provided by doctors (Liu et al., 2023). Further, another study where healthcare professionals rated medical advice from both physicians and ChatGPT found that the advice from ChatGPT was rated as being of higher quality and more

empathic than the physicians' advice (Ayers et al., 2023). While both studies concerning medical advice and our study reported that AI advice was longer than human advice, Howe et al. (2023) performed the analysis on both longer format AI advice and AI advice that matched in length with that of the human Advice and found the same significant preference for AI advice, suggesting that length of the advice is not the explaining factor in this consistent finding.

Why might effectiveness not be significantly different between AI advice and human advice in our study? Perhaps even though the overall AI text quality was rated higher on other quality measures, the underlying advice is still not more effective for solving the problem. Determining the meaning of this is hard because of the subjective nature of the measure we used. How can AI advice be more helpful but not more effective than human advice? It will depend on how people understand the terms. This is an inherent weakness of the measure we used. We followed the approach of earlier studies (Goldsmith & MacGeorge, 2000; Jones & Burlison, 1997; MacGeorge et al., 2004), participants were left to decipher the meaning of these terms subjectively and no clarifying definitions of the items were made explicit to the participants. This could have been a good addition to the measure leading to more uniform and clear meanings behind the terms. Another possibility remains, to decipher the meaning of the advice quality items by examining the actual content of the advice qualitatively.

Our participants in the Advice Quality Study rate the advice from the GPT-4 to be more helpful than the other models, while in the Preference Study, the older models are preferred. When looking at the descriptives (Figure 8) we can see an increase in advice quality as the models evolve, from Text-Davinci-3, GPT-3.5-Turbo to GPT-4. This suggests that there are improvements from each AI iteration and that human advice has already been surpassed in terms of perceived advice quality. Our results suggest that it's the helpfulness rating that improves

significantly with the new iterations of models. However, this finding conflicts with our results from the preference study where Text-Davinci-3 and GPT-3.5-Turbo are preferred (based on helpfulness perceptions) over human advice at a higher rate than the GPT-4 model, which is not preferred significantly over human advice (Figure 6). This a puzzling finding that we find hard to explain.

Implications

The indistinguishability of AI from human advice is a notable finding with both concerning aspects and inspiring implications. First and foremost, this represents a significant milestone for LLMs. While the models have not been intentionally trained to deceive people into thinking that they are humans, their ability to do so marks a significant achievement in AI's journey to master language. At the same time, the ability of a non-human system to seamlessly blend into digital textual environments and garner influence is intuitively disturbing. The fact that these models can be used with such speed and that they can be automated to respond, initiate posts, and spread information that could be embedded with an agenda is alarming. While most models have certain safeguards in place to make malignant use harder, the increasing need for robust AI detection tools is evident, especially on online platforms. Knowing the source for suggestions and advice is important information for the advisee, to consider before implementing any advice. Because this is an important factor in determining the expertise of the advice and as has been shown (Birnbbaum & Stegner, 1979; Meshi et al., 2012) the perceived competence status of the advisor influences the weight of the advice.

In some sense, the competence status of the AI is what we were aiming to initially determine with our experiments. While we used the best measures, we could find we were only able to determine the perceived competence of the models in the form of advice quality

measures. This distinction matters as the consequences of implementing AI advice are still unknown while we now know that people rate AI advice to be of high quality, often higher than human advice (Ayers et al., 2023; Howe et al., 2023a; Liu et al., 2023). It has been shown that perceived advice quality affects the weight of advice implementation (Bailey et al., 2022). The findings suggest that people weigh advice more in their decisions when they perceive the advice quality to be good. Taking these findings together would suggest that people most likely will use AI advice in their decision-making. It is important, however, to note that there are inconsistencies in definitions and measures of advice quality between studies, making any cumulative conclusions hard to establish.

While the advice of the LLMs is designed to assist users objectively, we cannot escape the fact that advising on subjective matters is based on the values and judgments of the advisor. In the case of the early models, for instance, all models by OpenAI up to and including GPT- 3, an objective approach was in a sense aimed for. In these models, the output of the AI was fully based on the patterns it learned from the training data (Brown et al., 2020; Radford et al., n.d.). What it seemed to lead to is that it mirrored the language of its training data, and this data was based on individuals who present their subjective values in their writing. Thereby the model in a sense is a child who learns from all these various examples of writing and extracts its values from the texts. For example, when looking at the GPT-3 model in Zellers's study (2021), its advice was found to be dangerous in 10 % of the cases, these responses were not AI being malignant in a novel way but echoed the trends and bad quality of the content in the training data. Now the newer versions of the GPT, models are also trained with human reinforcement learning (OpenAI et al., 2024).

Companies that develop LLM's have taken the evaluation of these models partially into their own hands, by doing this they can align the models to respond in ways that represent the values they want. This will present a challenge for the companies that create AI models, to strike a factual, objective, and yet morally and culturally inclusive system. It's also hard to investigate as the OpenAI models are not open about their training data and value alignment processes. However, it seems evident that high-quality data should be used, data that is inclusive and represents not only a small subsection of the population, and that extreme caution is to be taken when making decisions about how to align the model's values based on human reinforcement learning. These are issues that require careful consideration as the alignment of AI values could have far-reaching consequences, more and more impactful as these technologies are further integrated into practical use in societies.

Limitations

We aimed to evaluate the AI-generated advice with minimal editing. However, due to clear differences in formatting patterns between AI and human advice, we were required to ask the AI models to rewrite the responses without numbered lists as we found that the AI models preferred this format to the extent where it would have been a possible confound in our research. Additionally, the length of the AI answers was substantially longer. While we tried to make the answers shorter for all models, only GPT-4 managed to do this in a way that did not clearly compromise the quality of the advice. Therefore, we only applied this form of modification to the advice from GPT-4. This inconsistency in response length adjustment introduces potential biases when comparing the models. While we deemed the advice from GPT-4 to not suffer in

quality in response to the prompting, it is still a subjective judgment that could have resulted in lesser advice quality.

While we scripted the selection of advice with the intent to capture the best advice from Reddit. There are some reasons to say that this is still not representative of the best advice that is out there. Using “upvotes” on Reddit as a quality metric may not fully capture the quality of the advice, but they may also represent the entertainment value of the advice and the advice question. This may result in not getting the best human online advice but a combination of good and entertaining advice. Thereby, the entertainment value of the advice could get in the way of trying to scrape the best quality advice from Reddit. Professional advice or advice from a friend might still be better than this advice. Another thing to note is that of the 27000 advice our web scraping script processed, only 1.2 % had more than 20 upvotes, suggesting that most of the advice requests do not garner much interaction from people. This leads to the question, what kind of advice requests do capture the attention and engagement of people?

The short scale for advice quality lacked an explanation of each of its items such that helpfulness, appropriateness, sensitivity, and effectiveness, could have been interpreted differently by our participants. We suggest developing a standardized advice quality rating scale, which would explicitly define the items for participants. As of now, studies that mention advice quality often define it in different ways, and studies that measure advice quality use their own novel scales. Also, our quality ratings were based on perceptions and not the actual use of the advice or its consequences.

We made sure that we used novel Reddit advice requests, and there was no possibility that the models had been trained on these exact requests and advice pairs, as the training data included only data from the Internet up to September 2021 (OpenAI, n.d.). It is still possible that

similar training data have been used as the Reddit advice board, and that the advice questions have been already asked and answered not in identical format but rephrased. Then, instead of coming up with a novel answer the model could have in a sense looked up the old answer and used that one. This problem, or concept, is called benchmark contamination (Yang et al., 2023). While it does not minimize our findings, it can still cloud our conclusion. Maybe instead of AI being good at giving advice, it is only good at advising on matters that have been discussed in its training data. We are unable to make this distinction.

In our identification study, 75% of the advice participants were presented with was produced by AI. As we randomly chose advice from one of our four sources, and three out of four were AI sources. This uneven ratio of AI to human Advice could have possibly biased our participant's guesses. While we have no evidence that this would result in a bias there is still a possibility that it did. In the future we recommend a 50 / 50 split of AI and human answers, to eliminate the possibility of any expectation bias.

In general, we also saw a high dropout rate in our studies. In total 106 people from 308 participants started the experiment without answering any of the questions. This makes the dropout rate of our studies 34.4% in total. We think this could have been minimized by giving an incentive to complete the study. We also decided to include everyone in the analysis who answered at least one of the questions. This approach maximized the data we could capture but could also lead to some people being included in the analysis who did not pay as careful attention to their answers as others (though we did not control for this or test it in any way).

Further Research

Further research into LLMs is crucial given their novelty and possible impact on society. There is a need for large-scale studies that incorporate diverse models and methodologies to fully

understand the capabilities and limitations of LLMs. A good starting point would be to investigate how users are currently interacting with these technologies. Initial studies could focus on profiling the user base and identifying the most common use cases. Such data would be crucial for preemptively identifying and mitigating potential misuse before it escalates into societal issues. Furthermore, guiding users and organizations on the effective utilization of LLMs. This could have a positive impact in multiple domains, such as healthcare and science, and perhaps even fuel progress through more effective and safe use.

Incorporating the findings from recent advice research into the evolving field of AI advice is critical for understanding and enhancing the utility of AI-generated advice. Current methods for evaluating the quality and impact of advice are underdeveloped. While existing measures like JAS for assessing the impact of quantitative advice are robust, they lack ecological validity and do not reflect the types of advice commonly encountered in everyday life. On the other hand, there seems to be no standard method that is designed to measure the quality of more prototypical instances of advice, that do not deal with numerical or factual questions. All instances of research investigating this kind of AI advice quality, used different measures, highlighting the fact that no consensus on how to best do this exists.

To move forward, it is essential to find a robust way to measure the perceived quality of advice, ideally, a measure that could be used in multiple studies and across a broad range of advice topics. Also, it would be important to explore ways to capture the impact of this kind of advice. Ideally, these measures would incorporate outcome-based evaluations. These evaluations would assess the effectiveness of advice by examining its real-world applications in decision-making and its consequences. This approach would provide a more direct measure of how beneficial or effective the advice is and would enable researchers to assess the real impact of the

AI advice. Another possibility would be to extend the JAS methodology to include non-numerical advice. This could theoretically be possible by utilizing the same mechanism that LLM's use to map language, by representing longer texts in multidimensional vector space. In a sense it could be possible to quantify the language used. Using semantic textual similarity (Chandrasekaran & Mago, 2022) , it could be possible to calculate the influence of non-numeric advice by giving a number to the similarity between written texts. If this were successful, it would extend the utility of JAS, allowing the measurement of more ecologically valid advice.

Further research could also investigate if there indeed is a breaking point in AI identification. Is the trend in our data for people identifying GPT-4 only a statistical anomaly, or would the newer models be easier to identify than the older models? Tests could be done on chess engines with varying levels of mastery of the game and with newer models of LLMs to see if such a breaking point emerges.

Conclusion

It seems that LLM's are closing in on the mastery of language. Our research has shown that some of the newest models of OpenAI can give advice that is close to indistinguishable from human advice and perceived to be more helpful and of higher quality than human advice. Other research seems to echo this finding. The implications of this are both exciting and concerning. If used to its full potential, AI isn't a tool to replace human intelligence; it's a powerful intelligence support. With proper use, it has the potential to aid innovations in many domains, while at the same time, caution is advised.

While studies investigating AI advice are important for the future, creating new measures of advice quality and refining already existing ones is a fundamental step that should be taken to ensure the validity of this research. As the use of these technologies increases and the quality of

the models improves, the task of scientists to explore these new technologies will become ever more important.

References

- 107 Up-to-Date ChatGPT Statistics & User Numbers [April 2024]. Retrieved April 26, 2024, from <https://nerdynav.com/chatgpt-statistics/>
- Ayers, J. W., Poliak, A., Dredze, M., Leas, E. C., Zhu, Z., Kelley, J. B., Faix, D. J., Goodman, A. M., Longhurst, C. A., Hogarth, M., & Smith, D. M. (2023). Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum. *JAMA Internal Medicine*, 183(6), 589.
<https://doi.org/10.1001/jamainternmed.2023.1838>
- Bailey, P. E., Leon, T., Ebner, N. C., Moustafa, A. A., & Weidemann, G. (2022). A meta-analysis of the weight of advice in decision-making. *Current Psychology*.
<https://doi.org/10.1007/s12144-022-03573-2>
- Birnbaum, M. H., & Stegner, S. E. (n.d.). *Source Credibility in Social Judgment: Bias, Expertise, and the Judge's Point of View*.
- Bo Feng, & Burleson, B. R. (2008). The Effects of Argument Explicitness on Responses to Advice in Supportive Interactions. *Communication Research*, 35(6), 849–874.
<https://doi.org/10.1177/0093650208324274>
- Bogert, E., Schecter, A., & Watson, R. T. (2021). Humans rely more on algorithms than social influence as a task becomes more difficult. *Scientific Reports*, 11(1), 8028.
<https://doi.org/10.1038/s41598-021-87480-9>
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020).

- Language Models are Few-Shot Learners* (arXiv:2005.14165). arXiv.
<http://arxiv.org/abs/2005.14165>
- Carpenter, J., McDade, C., & Childers, S. (2018). *Advice Seeking and Giving in the Reddit r/Teachers Online Space*.
- Chandrasekaran, D., & Mago, V. (2022). Evolution of Semantic Similarity—A Survey. *ACM Computing Surveys*, 54(2), 1–37. <https://doi.org/10.1145/3440755>
- DECAPUA, A., & HUBER, L. (1995). 'If I were you ...': Advice in American English. 14(2), 117–132. <https://doi.org/doi:10.1515/mult.1995.14.2.117>
- Gino, F., Brooks, A. W., & Schweitzer, M. E. (2012). Anxiety, advice, and the ability to discern: Feeling anxious motivates individuals to seek and use advice. *Journal of Personality and Social Psychology*, 102(3), 497–512. <https://doi.org/10.1037/a0026413>
- Gino, F., & Moore, D. A. (2007). Effects of task difficulty on use of advice. *Journal of Behavioral Decision Making*, 20(1), 21–35. <https://doi.org/10.1002/bdm.539>
- Goldsmith, D., & MacGeorge, E. (2000). The impact of politeness and relationship on perceived quality of advice about a problem. *Human Communication Research*, 26(2), 234–263. <https://doi.org/10.1111/j.1468-2958.2000.tb00757.x>
- Golia, J. (2021). *Newspaper Confessions: A History of Advice Columns in a Pre-Internet Age* (1st ed.). Oxford University Press. <https://doi.org/10.1093/oso/9780197527788.001.0001>
- Harper, D. (n.d.). Etymology of advice. Online Etymology Dictionary. Retrieved April 12, 2024, from <https://www.etymonline.com/word/advice>
- Hirschberg, J., & Manning, C. D. (n.d.). Advances in natural language processing. *ARTIFICIAL INTELLIGENCE*.

- Howe, P. D. L., Fay, N., Saletta, M., & Hovy, E. (2023b). ChatGPT's advice is perceived as better than that of professional advice columnists. *Frontiers in Psychology, 14*, 1281255. <https://doi.org/10.3389/fpsyg.2023.1281255>
- Jones, S. M., & Burleson, B. R. (1997). The Impact of Situational Variables on Helpers' Perceptions of Comforting Messages: An Attributional Analysis. *Communication Research, 24*(5), 530–555. <https://doi.org/10.1177/009365097024005004>
- Kelly, N., & Antonio, A. (2016). Teacher peer support in social network sites. *Teaching and Teacher Education, 56*, 138–149. <https://doi.org/10.1016/j.tate.2016.02.007>
- Kuosmanen, O. J. B. (2023). Reddit Data Extraction and Basic Language Analysis (Version 1.0.0) [Computer software]. Script name: advice_posts_per_day.py LINK: news.view.reddit/scripts/advice_posts_per_day.py at main · OttoKuosmanen/news.view.reddit · [GitHub](#)
- Limberg, H., & Locher, M. A. (Eds.). (2012). *Advice in Discourse*. John Benjamins. <https://www.jbe-platform.com/content/books/9789027273932>
- Liu, S., McCoy, A. B., Wright, A. P., Carew, B., Genkins, J. Z., Huang, S. S., Peterson, J. F., Steitz, B., & Wright, A. (2023). *Leveraging Large Language Models for Generating Responses to Patient Messages*. <https://doi.org/10.1101/2023.07.14.23292669>
- Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes, 151*, 90–103. <https://doi.org/10.1016/j.obhdp.2018.12.005>
- MacGeorge, E. L., Feng, B., Butler, G. L., & Budarz, S. K. (2004). Understanding Advice in Supportive Interactions.: Beyond the Facework and Message Evaluation Paradigm.

- Human Communication Research*, 30(1), 42–70. <https://doi.org/10.1111/j.1468-2958.2004.tb00724.x>
- MacGeorge, E. L., & Van Swol, L. M. (2018). *The Oxford Handbook of Advice*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190630188.001.0001>
- McCorduck, P., & Cfe, C. (2004). *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence*. CRC Press.
<https://books.google.no/books?id=r2C1DwAAQBAJ>
- Meshi, D., Biele, G., Korn, C. W., & Heekeren, H. R. (2012). How Expert Advice Influences Decision Making. *PLoS ONE*, 7(11), e49748.
<https://doi.org/10.1371/journal.pone.0049748>
- Models—OpenAI API*. (n.d.). Retrieved April 26, 2024, from
<https://platform.openai.com/docs/models>
- Morandín-Ahuerma, F. (2022). *What is Artificial Intelligence*. 3(12).
- Nov, O. (n.d.). *Putting ChatGPT's Medical Advice to the (Turing) Test*.
- OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., ... Zoph, B. (2024). *GPT-4 Technical Report* (arXiv:2303.08774). arXiv. <http://arxiv.org/abs/2303.08774>
- OpenAI, Berner, C., Brockman, G., Chan, B., Cheung, V., Dębiak, P., Dennison, C., Farhi, D., Fischer, Q., Hashme, S., Hesse, C., Józefowicz, R., Gray, S., Olsson, C., Pachocki, J., Petrov, M., Pinto, H. P. d O., Raiman, J., Salimans, T., ... Zhang, S. (2019). *Dota 2 with Large Scale Deep Reinforcement Learning* (arXiv:1912.06680). arXiv.
<http://arxiv.org/abs/1912.06680>

Qualtrics XM: The Leading Experience Management Software. (n.d.). Retrieved April 26, 2024, from <https://www.qualtrics.com/>

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (n.d.). *Language Models are Unsupervised Multitask Learners*.

Russell, S. J., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach*. Pearson. <https://books.google.no/books?id=B4xczgEACAAJ>

Schultze, T., Mojzisch, A., & Schulz-Hardt, S. (2017). On the Inability to Ignore Useless Advice: A Case for Anchoring in the Judge-Advisor-System. *Experimental Psychology*, 64(3), 170–183. <https://doi.org/10.1027/1618-3169/a000361>

SciPy -. (n.d.). Retrieved April 30, 2024, from <https://scipy.org/>

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K., & Hassabis, D. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science*, 362(6419), 1140–1144. <https://doi.org/10.1126/science.aar6404>

Silverman, K. (2022). The Androids of Hephaestus: Between Human and Machine in the Iliad. *Selected Proceedings of the Classics Graduate Student Symposia at the University of Florida*, 1. <https://doi.org/10.32473/pcgss.v1i.130426>

Sniezek, J. A., & Buckley, T. (1995). Cueing and Cognitive Conflict in Judge-Advisor Decision Making. *Organizational Behavior and Human Decision Processes*, 62(2), 159–174. <https://doi.org/10.1006/obhd.1995.1040>

Soll, J. B., & Larrick, R. P. (2009). Strategies for revising judgment: How (and how well) people use others' opinions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(3), 780–805. <https://doi.org/10.1037/a0015145>

- Solomonoff, R. J. (1985). The time scale of artificial intelligence: Reflections on social effects. *Human Systems Management*, 5(2), 149–153. <https://doi.org/10.3233/HSM-1985-5207>
- Tolkien, J. R. R. (2005). *The lord of the rings*. HarperCollins. <https://archive.org/details/tolkien-j.-the-lord-of-the-rings-harper-collins-ebooks-2010/page/84/mode/2up>
- Turing, A. M. (Alan M. (1950). Computing machinery and intelligence. In *The Philosophy of artificial intelligence* (pp. 40–66).
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (n.d.). *Attention is All you Need*.
- Vinyals, O., Babuschkin, I., Czarnecki, W. M., Mathieu, M., Dudzik, A., Chung, J., Choi, D. H., Powell, R., Ewalds, T., Georgiev, P., Oh, J., Horgan, D., Kroiss, M., Danihelka, I., Huang, A., Sifre, L., Cai, T., Agapiou, J. P., Jaderberg, M., ... Silver, D. (2019). Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature*, 575(7782), 350–354. <https://doi.org/10.1038/s41586-019-1724-z>
- Welcome to Python.org*. (n.d.). Retrieved April 30, 2024, from <https://www.python.org/>
- Wittgenstein, L. (1968). *Philosophical investigations*. Basil Blackwell.
- Yang, S., Chiang, W.-L., Zheng, L., Gonzalez, J. E., & Stoica, I. (2023). *Rethinking Benchmark and Contamination for Language Models with Rephrased Samples* (arXiv:2311.04850). arXiv. <http://arxiv.org/abs/2311.04850>
- Yaniv, I. (2004). Receiving other people's advice: Influence and benefit. *Organizational Behavior and Human Decision Processes*, 93(1), 1–13. <https://doi.org/10.1016/j.obhdp.2003.08.002>

- Yaniv, I., & Kleinberger, E. (2000). Advice Taking in Decision Making: Egocentric Discounting and Reputation Formation. *Organizational Behavior and Human Decision Processes*, 83(2), 260–281. <https://doi.org/10.1006/obhd.2000.2909>
- Yenduri, G., M, R., G, C. S., Y, S., Srivastava, G., Maddikunta, P. K. R., G, D. R., Jhaveri, R. H., B, P., Wang, W., Vasilakos, A. V., & Gadekallu, T. R. (2023). *Generative Pre-trained Transformer: A Comprehensive Review on Enabling Technologies, Potential Applications, Emerging Challenges, and Future Directions* (arXiv:2305.10435). arXiv. <http://arxiv.org/abs/2305.10435>
- Zellers, R., Holtzman, A., Clark, E., Qin, L., Farhadi, A., & Choi, Y. (2021). *TuringAdvice: A Generative and Dynamic Evaluation of Language Use* (arXiv:2004.03607). arXiv. <http://arxiv.org/abs/2004.03607>
- Zhai, Y., Yan, J., Zhang, H., & Lu, W. (2020). Tracing the evolution of AI: Conceptualization of artificial intelligence in mass media discourse. *Information Discovery and Delivery*, 48(3), 137–149. <https://doi.org/10.1108/IDD-01-2020-0007>

Appendix A

Python scripts: Web scraping Reddit

In this section, we go through the scripts that created the advice request-advice pair database used in the experiment. Code files can be found on www.github.com/OttoKuosmanen/reponame

REDDIT.py

For the advice questions and human answers, we used Reddit. A Python script was created that looked for posts in the subreddit: Advice, if posts had more than 20 likes they were saved.

Because of limitations to the Reddit API, we could only look through 1000 posts. Therefore, the script was executed 27 times, a maximum of 1 time per day, between the dates: 03.05.2023 - 04.19.2023. Each post that passed the criteria was saved with the corresponding best-rated Reddit advice in JSON format.

JSON Reader Unique.

Because this resulted in multiple cases of duplicate posts being saved, we also made a script that checked for duplicate IDs on the posts and passed on only one of them. For the next step.

Update.py

Because time passes between running the data gathering script, we included this update script. This script compiled the unique posts and updated the best advice based on the new upvotes.

Clean.py

A censor script was run on the questions: That excluded questions that included the words: abuse, assault, rape, molest, suicide, kill in all forms. These topics were determined unnecessarily intense for our research participants. The sensor also checked for the words update and edit, as these are indications that changes have been made to the post, changes that often include information on a resolution of the advice problem or additional information that might have not been used while writing the best-rated advice answer.

A minimum and maximum length of both advice request and advice answer were also set. Advice request: max 500 words, min 32 words. Advice answer: max 400 words, min 32 words. Before excluding the censored items, the database included 325 items.

```
"before": {  
  "mean length of answer": "84.16923076923077",  
  "mean length of question": "220.68307692307692",  
  "mean answer score": "317.05846153846153",  
  "std answer score": "451.9722383892137",  
  "std question": "174.4745461437599",  
  "std_answer": "72.59755648161027",  
  "max answer score": "3830",  
  "longest answer": "804",  
  "longest question": "1219",  
  "min answer score": "3",  
  "shortest answer": "31",  
  "shortest question": "10",  
  "number of items": "325"
```

After the censored items were removed the database included 202 items:


```
"after": {  
  "mean length of answer": "82.38118811881188",  
  "mean length of question": "185.81683168316832",  
  "mean answer score": "308.7821782178218",  
  "std answer score": "475.1938760973292",  
  "std question": "107.16287739208884",  
  "std_answer": "57.53800502581533",  
  "max answer score": "3830",  
  "longest answer": "346",  
  "longest question": "498",  
  "min answer score": "20",  
  "shortest answer": "32",  
  "shortest question": "38",  
  "number of items": "202"
```

Appendix B

Python Scripts: AI advice

These are the scripts that were used to generate the AI advice in response to our advice requests from Reddit. They can be found on GitHub under the file name of AI. The names scripts are the same as the bolded headers.

Model scripts were connected to the OpenAI application programming interface (API) found on the OpenAI website under the developer section. At the time of generating the AI advice, the GPT 4 API was not available, so here we used the available website interface for subscribers. For the prompt, we only included the advice request.

random_sample.py

A script was created that took a random sample of 50 from the 202 eligible question-answer pairs in the advice database we generated.

question_reader.py

We made a simple script that processed the random sample and saved the advice requests and advice separately in lists. This organization allows for easier access and manipulation in subsequent analyses.

Gpt3_davinci.py

This script was connected to the OpenAI API. We linked the advice requests from question_reader to the script. For each request, the script formats a prompt that includes the advice requests followed by a placeholder for the answer. This structured prompt is then fed into

the model. The process repeats for each advice request, effectively allowing the model to generate potential answers for all 50 distinct questions.

X = one of the actual advice requests

The exact prompt was: prompt = Question: X. Answer:

Chat_gpt3.5.py

This script was connected to the OpenAI API. We linked the advice requests from `question_reader` to the script. For each request, the script formats a prompt that includes the advice requests followed by a placeholder for the answer. This structured prompt is then fed into the model. The process repeats for each advice request, effectively allowing the model to generate potential answers for all 50 distinct questions.

The prompt is a bit different in each model because of different prompting possibilities between the models.

Here we defined 3 roles, the system, the user and the assistant.

System = You are answering an advice question

User = Question: X

Assistant = Answer:

X = one of the actual advice requests

Appendix C

Preprocessing of the items

Formatting of answers

Answers often contained a numbered list format: In GPT 4 this format was used in 49 of the 50 answers. In GPT 3.5 this format was used in 15 of the 50 answers. In GPT 3 this format was used in 2 of the 50 answers. In the human answers, this format was used in 0 of the 50 answers.

To remove this systematic difference, we prompted the same models with “Transform this text into paragraph structure” and passed in the numbered list replies that they had produced. This transformation was a success. These answers were used instead of the numbered list replies, leading to the result that no advice answers used in the experiment were in the form of a numbered list.

Again at the time, the GPT4 API was not available so we used the online interface with the same prompt as the other models.

The scripts that handled this were the following:

chat_gpt3.5_rewrite_list_paragraph_format.py

davinci_rewrite_paragraph_format.py

Length of the answers

When generating the AI answers to the advice questions. An attempt was made to match the length of answer to the corresponding human answers. By setting the max token length to

number of words used in the human answers. This resulted in an automated script that set each prompt with a custom max token setting depending on the length of the corresponding human advice answer. This did not yield a satisfactory result, as the AI answers were in most cases cut mid-sentence.

Therefore, the AI answers length was not restricted during the formation of the original advice answers. This resulted in a general trend of longer AI answers (mean = 230.47) than human answers (mean = 88.38 words). Text-Davinci-3 mean length of the answer was = 175.62, GPT3.5-Turbo mean length of the answer was = 209.86, and the mean length of the GPT4 answer was 305.93. (Sample 50). As can be seen, there was a trend for increasing answer length with model progression.

As the mean length of AI-generated answers was so much higher, we decided to try to generate shorter answers. We prompted the AI with “Write the same thing but more shortly” and passed in the long answers. This resulted in shorter answers, Text-Davinci 3 (55.70), GPT3.5-Turbo (67.36), and GPT4 (148.63) but it was clear at face value that the answer quality had dropped substantially in both Text-Davinci-3 and GPT-3.5-Turbo. The GPT 4 quality of the advice did not seem to suffer at face value. We decided to only use the shorter answers by GPT 4.

These scripts for doing this are the following:

Davinci_shorten

GPT3.5_shorten

GPT4_shorten

Prompt for GPT-4

```
messages=[
    {"role": "system", "content": "slightly shorten the
text"},
    {"role": "user", "content": question},
    {"role": "assistant", "content": "The same text but
slightly shorter: "},
]
```

Missed the censor script.

At the data mining stage, a censoring script had already been run on the advice questions. In the random sample of 50 questions taken from the 300 questions, we found that 3 questions still contained sensitive material, questions 9, 30, and 37. This was noted by GPT4's inbuilt censor system and it refused to answer these questions. One of the answers contained a case of sexual assault where the wording had been abbreviated to SA, one was related to a grooming incident, and the last one mentioned underage sexual conduct. In the end, we removed the 3 questions and were left with 47 questions as the final sample.

Even after this we also found two advice request-advice pairs that we saw fit to remove from the sample. We removed question 42 (as seen in Qualtrics). Because the question was regarding physical abuse. Removed question 36 (as seen in Qualtrics). Because the human answer was made by the same person who asked the question. Seems like there was some error in the algorithm that got the questions and answers.

Grammatical Edits

We decided not to make any major edits to the texts, at most an article or preposition was removed or added. However, we chose to remove obvious typos that would result in identification strategies based only on human errors in the text. While we think that it would be a good strategy to distinguish between human and AI advice, it was not the effect we were interested in.

We used Grammarly to locate errors in the human answers while also doing a subjective evaluation of the correctness of the grammar. We split the error into 4 types: Misspelling, wrongly used punctuation or commas, capitalization errors, and spacing errors. In total, only 9 out of 47 questions were without errors in the language. While most of the error types are self-explanatory in their name, in the misspelling category we also included changes of the article, and preposition.

Notes:

We noticed that the AI uses the American spelling of the words, this is not a surprise since OpenAI is an American company.

Grammarly is also an AI tool. The language analysis is prone to errors, as I'm not a language expert, and the language was not analysed by more than 1 person.

Categories: 1: All grammatical errors except, comma, punctuation, spacing, and capitalization errors.

| Source | Misspelling and other errors | Capitalization | Spacing errors | Punctuation or comma | Total Errors |
|--------|------------------------------|----------------|----------------|----------------------|--------------|
| | | | | | |

| | | | | | |
|----------------|----|---|----|----|----|
| Human | 46 | 8 | 10 | 35 | 69 |
| Text-Davinci-3 | 13 | 0 | 0 | 11 | 24 |
| GPT-3.5-Turbo | 6 | 0 | 0 | 9 | 15 |
| GPT-4 | 16 | 0 | 0 | 9 | 25 |

Self-referencing:

We found that people sometimes refer to themselves and their experiences when giving advice. When they do this revealing information about themselves, we call it self-referencing. We found that this occurred in 13 out of the 47 human advice answers. In 9 out of the 11 instances the following occurred: the self-referencing was a statement that the advice was or had experienced something similar. In the other three, the self-referencing was revealing personal information that resulted in them having an insightful perspective into the issue (In the case of asking advice on what to do with a stalker: Working in security management, having been homeless, experience with the opposite position). One casual dropping of self-reference, without an apparent reason. The following advice requests had human advice where self-referencing occurred.

(1,5,11,13,15,18,21,23,30,40,41,43,44)

Instances of Self Referencing for each advice source

| SOURCE | AMOUNT OF SELF-REFERENCES |
|-----------------------|----------------------------------|
| HUMAN | 13/47 |
| TEXT-DAVINCI-3 | 0/47 |
| GPT-3.5-TURBO | 0/47 |

GPT-4

| 1/47

As suspected the AI did not use self-referencing in the same way as people did while giving advice. Interesting to note is that GPT4 did use a kind of anti-self-referencing. It stated in 6 answers something in the form of I'm not a: lawyer, mental health professional, or legal expert and in 1 answer directly stated that I'm an AI.

Based on this analysis we chose to exclude any questions that included self-referencing. As that would beat the purpose of our AI identification study.

Final Edits: Minimize Length Differences Between AI and Human Advice:

We started to worry about having too many questions in the sample. Because the more questions we have the more we will need to have participants completing the experiments. To combat this, we found a way to deal with two flies in one go. Because there was also a possible confounding variable of answer length, we decided to choose 15 of the answers with the lowest difference in word count between AI and human answers. To do this we made a script that counted the word count of each advice answer. We then counted the average of the AI answers for each question, resulting in a list of averages for all answers, and then we counted the length of each human answer. We then did a differential analysis where each human answer length was contrasted with the average of the AI answers. This way we located the 15 questions with the least amount of difference between lengths across all models. This resulted in the following advice items from the sample of 50 [1,41,47,11,13,38,35,3,30,16,26,20,12,5,28]. We decided to use these advice

requests and the four pieces of corresponding advice for each request in our advice preference and advice quality study. For the identification study, we additionally removed advice that included self-referencing, leaving us with the following 9 items [3,12,16,20,26,28,35,38,47].

```
GENERAL_ANALYTICS
Top 15 items with the least differences between words counts: AI/average - Human
Question: 1: Difference: 19
Question: 41: Difference: 28
Question: 47: Difference: 30
Question: 11: Difference: 33
Question: 13: Difference: 35
Question: 38: Difference: 36
Question: 35: Difference: 44
Question: 3: Difference: 45
Question: 30: Difference: 46
Question: 16: Difference: 55
Question: 26: Difference: 55
Question: 20: Difference: 65
Question: 12: Difference: 69
Question: 5: Difference: 70
Question: 28: Difference: 72
```

After this, we decided to shift our referencing convention for the advice request and advice items. As we had 15 items, we started to refer to them as advice requests [1-15], starting from the lowest number to the highest (item 1=1, item 3=2, item 5 =3, etc).

see Appendix D for the actual items that were used in the final experiments.

Appendix: D

Advice requests and advice

These 15 advice requests and advice were used in the Preference Study and the Advice Quality study. A subsection of these was used for the Identification Study (2,5,7,8,9,10,12,13,15).

Advice request: 1

Is it weird to semi-regularly talk to your sister's boyfriend?

I'm tired of explaining this but again tonight my boyfriend said it was weird "how much" I talk to my sister's boyfriend, we'll just say K for brevity. First off, we talk like 2-3 times a week and it's a few texts that he usually leaves on delivered for days. I've known K for like 6 years at this point, even rented an apartment with him and my sister for a bit last year. And I have never, literally never thought about him in any more than a platonic way. I just don't see him that way. And the thought of doing anything w him makes my skin crawl. Yet my mother, sister, and my boyfriend have acted suspicious and insinuated I liked him multiple times over the years. I don't want to cut someone off that I genuinely enjoy talking to, but I also don't want to make my sister or boyfriend uncomfortable by making them feel like I'm doing something shady.

Possible solutions: Block him, don't block him, talk even less than we already do. But I feel the problem would remain and I'd lose a good friend

Advice Text-Davinci-3:

It may not be the most ideal situation, but there really isn't anything wrong or "weird" about talking to your sister's boyfriend as long as you are only having conversations that are appropriate and platonic in nature. If your family or your boyfriend is still suspicious of your interactions, you could try being more open about why you are talking to him and setting boundaries for your interactions. It's up to you to decide if the friendship is worth maintaining, and if it is, then be sure to communicate that to your family and boyfriend.

Advice 1 GPT-3.5-Turbo:

It is not weird to talk to your sister's boyfriend if there is a genuine friendship and connection between you two that is strictly platonic. You have known him for several years, shared an apartment together, and have a good friendship. As long as both of you are comfortable with the level of communication, there is nothing wrong with it. It is natural for your boyfriend to feel a little insecure or jealous, but if you have reassured him that it is strictly platonic and he trusts you, there should not be a problem. If you feel like you are losing a good friend by blocking or reducing communication, you can have an open and honest conversation with both your sister

and your boyfriend to address any concerns they may have and try to reach a mutual understanding.

Advice 1 GPT-4:

Conversing with your sister's boyfriend, K, can be normal, especially if you have a longstanding bond. But remember to consider the feelings of your boyfriend and sister. The key is open communication. Discuss the situation with those involved, reassuring them that their feelings are important to you. Establish clear boundaries with K and discuss these with your sister and boyfriend to prevent potential misunderstandings.

Involve your sister and boyfriend in your chats with K often, so they can see its platonic nature. Self-reflect on your feelings for K and ensure the relationship remains platonic. Balance is crucial in maintaining harmony among all parties. While it may be challenging, open dialogue and explicit boundaries can help maintain relationships without causing discomfort or suspicion.

Advice 1 Human:

No girl, just talk to him normally and make sure to mention in front of others that he's like family to you.

As in "Oh, he's my brother now" since he's your sister's bf. Or "Oh he's like my family" or "Oh, I see him as my brother" ...

It's normal to be close to a friend or family member. And obviously if he's been around for years, you'll consider him family at this point.

Same way it's normal for my family, especially grandmas to treat my bf like their grandson now at this point. He's become family. Nothing wrong with that.

And since you yourself have said you don't see him that way... then nothing wrong with it as long as you set a very clear boundary. In this case the boundary is called publicly "familyzoned"

Advice request 2:

is it unreasonable for me to ask my friend not to smoke weed when we're together?
 i Simply dislike it. I don't like the smell nor the way she behaves when she's high. also it gets kinda awkward because i don't smoke and she's by herself giggling all the time and we can't do shit cause she's high.

before she started smoking we would do a lot together but now it's just me going to her place (she never comes to mine) and watching her smoke.

is it selfish of me to ask her to change her habits when I'm around? be brutally honest.

Advice 2 Text-Davinci-3:

No, it is not selfish of you to ask your friend not to smoke weed when you're together. It is perfectly reasonable for you to make your boundaries clear in terms of what kind of behavior you are comfortable with. If your friend values your friendship, she should be willing to respect your preferences when you are together. On the other hand, it would also be reasonable for you to respect that your friend may use weed at times and to be flexible when it comes to spending time with her.

Advice 2 GPT-3.5-Turbo:

It is not selfish of you to ask your friend not to smoke weed when you're together. You have the right to set boundaries in your relationship and express what makes you feel uncomfortable. However, you should approach the conversation with care and respect for your friend's choices. You could start by explaining how you feel when she smokes weed around you, and how it affects your time together. Be honest but also try to be understanding of her perspective. You could also suggest alternative activities that you both can do together instead of smoking weed. Ultimately, it is up to your friend whether or not they want to smoke weed when you're together, but it is important that you communicate your feelings and needs in the friendship.

Advice 2 GPT-4:

Communicating your feelings and preferences concerning shared experiences to your friend is crucial, requiring open and honest dialogue. Should you need guidance initiating this conversation, consider a few points. First, aim for a relaxed setting when your friend is sober, as it makes the dialogue more effective. Be honest yet respectful when describing how her smoking weed affects your time together. Suggest mutually enjoyable, sober activities to help you reconnect and make new memories. Be open to compromises, like asking her to smoke less during your get-togethers, and be prepared for possible concessions. Lastly, reassure her of the importance of your friendship, emphasizing that your intentions are to improve shared experiences. Keep in mind that successful friendships depend on open communication and mutual understanding.

Advice 2 Human:

It's fair for you to not want to chill with her while she smokes. I would just say, "Hey it's not really fun for me just to sit at your house and watch you smoke. Maybe if you want to meet at the park before or after and do x y and z that would be more fun for me." And if she says nah then cut your losses.

Advice Request 3:

should I(f32) sue hairdresser for not refunding \$140

Long story short, I went into a salon to dye my hair. They said, "we can dye it that color, no problem, no bleach will be used." After dyeing it, my hair came out two very different colors. Then they apply something in my hair. I ask what it is and, it's bleach. I express how much I didn't want that and that and how dare they just put it on my hair without telling me even after it was agreed that bleach wouldnt be used etc. They're like, "oh you'll be fine" and then proceeded to dye my hair a second time after the bleach. I look like a clown now with different colors on my hair (orange at the root and brownish red at the body). But the worst of it all is that I have bleach in my hair and I'm upset that my hair will never be the same. I will have to keep cutting it until the damaged bleached hair is gone. I have curly hair so my curls will most likely not be the same also.

I wanted a refund for this experience. These stylist did whatever they wanted to me. Bleached my hair and dyed it twice in 1 day. Instead the stylist and the owner start yelling at me telling me they can't refund me and that I'm wasting their time and that the most they could do for me is give me some type of moisturizing treatment the next time I come in. (They are so clueless) I explain that I'm never coming back and leave with damaged hair, orange roots, and redish/brown hair, a lousy blowout, an uneven trim, and the delight of being shouted at by two extremely unprofessional ladies. Not to mention I paid \$140 for this nightmare.

I spoke to friends they're all in shock of such an experience and they warn me that in a couple of months, my hair might fall out bc of the bleach 🙄 . I am considering suing this business. What are your thoughts?

Advice 3 Text-Davinci-3:

Given the magnitude of the issue, you may have a solid case for a lawsuit against the hairdresser. You should consult a lawyer experienced in consumer protection or business law to assess the strength of your argument and advise you on the correct next steps. Even if suing is not a viable option, you may be able to pursue other legal remedies available to you, such as filing a complaint with your state's department of consumer protection or attorney general's office. In addition to potential legal recourse, you may also be able to file a complaint with the state board of cosmetology and get the hairdresser in trouble depending on the laws and regulations in your state. This could provide you with a bit of closure, even if it doesn't result in a refund. In any case, it is important to ensure that what happened to you does not happen to anyone else. Take action and keep your receipts as evidence in case you decide to pursue any type of legal remedy.

Advice 3 GPT-3.5-Turbo:

It is understandable that you feel frustrated and angry about the experience you had at the salon. However, before you decide to sue, it's important to assess the situation and consider your

options. Firstly, you could try to resolve the issue directly with the salon by writing a complaint letter or speaking to the owner/manager. Explain the situation calmly and clearly, and ask for a refund or compensation for the damage done to your hair. It's possible that they may be willing to work with you to find a resolution. If the salon is unresponsive or refuses to offer any compensation, then you might consider taking legal action. However, before you do so, it's important to understand the costs and risks involved. It might end up costing you more in legal fees than the amount you're trying to recover, and there is no guarantee of winning the case. In the end, the most important thing is to take care of your hair and take steps to minimize any damage that may have been done. Consider getting professional advice from a stylist or haircare expert on how to best care for your hair in the aftermath of this experience.

Advice 3 GPT-4:

Though I can't offer legal advice, I can suggest some steps in dealing with this. Start by recording the damage: take photos of your hair and keep any receipts or records of conversations with the salon. You might also want to get a second opinion from another haircare professional. Reach out to the salon owner to express your dissatisfaction and ask for a refund. Be very clear about how your hair was damaged and the poor service you received. If they refuse, let them know you're considering further action. If you choose to go legal, get acquainted with your local small claims court process. These courts are meant for disputes under a specific amount and the process is usually cheaper and less complicated than a typical lawsuit. You could also meet with a lawyer to evaluate your case and give advice on what to do next. Bear in mind that pursuing legal action can be costly and time-consuming, so weigh the possible benefits against the cost and effort. Finally, sharing your experience on social media or review sites could raise awareness and prevent similar situations in future.

Advice 3 Human:

Wow I'm so sorry you experienced this, that is very unprofessional. Just to calm your anxiety. I have curly hair as well, and have bleached it more than I would like to admit. You will see some breakage at the ends but I can assure you it's not going to fall out. Use a deep conditioning treatment if it seems damaged, gradually trim the ends rather than just cutting it all, there is no reason you should have to lose your length and your money. You could take it to small claims but I feel in the long run it would be worth more of your energy than it's worth. If it is a corporation call and make a complaint about the salon specifically those ladies, we don't need people like that in the industry, they can have their license revoked. I know a lot of people are against doing their hair at home, but I personally do because of an experience similar to this, and I did attend cosmetology school. Your roots are much lighter because they take bleach the fastest. Go to Sally's get a toner and developer of your shade choice I would recommend like a burgundy shade a little darker than the ends of your hair and just do the roots, it will give you a nice ombre effect. If you do decide to do a dye at home, feel free to reach out with the look you were going for and I can offer more advice and walk you through it.

Advice Request 4:

I am dying, I don't want people to know. What should I do?
I just started therapy so Reddit is not the only place for this but I am feeling alone. I don't want to tell people I have less than a year to live. I'm so fucking exhausted. Chemo is horrible. My body is failing me. I have nothing left to live for and I don't want to be around my crying family and friends. I don't want anything anymore but to just hurry up and die. Sorry I know it's morbid but it's how I feel. Can barely type this. The hospital is depressing and I feel like I've done nothing in life worth anything and I'm only in my twenties. How do I feel better again and does it even matter if I do if I'll be dead soon anyways? Idk anymore

Advice 4 Text-Davinci-3:

I'm so sorry to hear about the struggle that you are going through right now. It sounds like a very difficult and heartbreaking situation, and I can only imagine how hard it must be for you. I would encourage you to talk to a professional - either your therapist or a counsellor or support group - so that you have someone to talk to about what you're feeling and to get some help in managing this difficult time. It can be incredibly helpful to share your experience and receive support from others who understand what you're going through. Also, try to keep a sense of perspective about your life and focus on the good moments and memories. Remember that whatever time you do have left, is still precious and meaningful. You are still here for a reason and your life has meaning, no matter how long you get to live it. Try to focus on the positive things in your life and the relationships that you have with the people around you. Finally, remember that you can still make a difference in the time you have left. Think about the things that make you feel better, and the things that you love to do, and find a way to make them a part of your life. Even if it's something as small as taking a drive to a nearby park, it can make all the difference. Take care of yourself.

Advice 4 GPT-3.5-Turbo:

I'm so sorry to hear about what you are going through. Please know that you are not alone, and there are people who want to support you. It's understandable that you don't want everyone to know about your condition, but it's important to have at least a few people you can confide in and lean on during this difficult time. Consider a close friend or family member, or even a support group for others going through similar experiences. While it may be difficult to find things to look forward to right now, it's important to try to find some joy and comfort in your life. Maybe that comes in the form of spending time with loved ones, reading a good book, enjoying your favorite foods, or simply finding moments of peace and rest. Therapy can be a great tool for exploring your feelings and finding ways to cope during this time. Remember that your life matters, and you have left an impact on those who love you. Your presence and love are valuable, and it's okay to ask for help and support in this difficult time. If you feel like you're struggling to cope or have thoughts of hurting yourself, please reach out to a medical professional or crisis hotline right away. You are not alone, and there is help available.

Advice 4 GPT-4:

While I'm not a mental health professional, I can provide suggestions. However, reaching out to a qualified therapist or counselor is crucial. Start by sharing your emotions with someone you trust. Try to focus on the present moment, engage in enjoyable activities, or seek out new hobbies for pleasure and achievement. Connect with support groups or online communities managing similar challenges. Prioritize self-care, maintaining a healthy routine inclusive of sleep, nutrition, and relaxation. Mindfulness techniques like meditation or deep breathing may help manage stress and improve mood. Contemplate your legacy to find a sense of purpose. This could mean sharing your experiences, advising others or spending quality time with loved ones. Remember, a mental health professional is essential for personalized guidance and support tailored to your circumstances.

Advice 4 Human:

I understand how you feel. I was a terminal cancer patient with no options. I had cancer before but was stage 3, then 10+ years later it came back much much worse. My doctors were fantastic and called all over the country to see if they could find anyone with my specific diagnosis. They found a hospital that had done an experimental procedure 3 times. Out of those one worked, the others passed. I had 4-6 months left or try this experimental procedure which could be deadly. My family and friends were completely sympathetic but I alone had to deal with my decision either way. I decided to do the experimental procedure. I was scared shitless and consistently depressed. I was heavily medicated for anxiety and panic attacks. I couldn't get out of my own head for a long time. Fortunately I've been in remission for 2 years. The other person who had the procedure and lived has now died. I don't know how long I have months to 20+ years. I've been where you are and I could be again. It's all a mystery. Please don't give up and find just one person, even another terminal patient, to talk with. Having a horrible situation in common can be cathartic. Please feel free to message me if you would like to talk. I'm sending you healing and peaceful vibes.

Advice Request 5:

I'm 15. I have a job, but I'm not saving for anything in particular yet. Might save for a car when I turn 16. Is it a waste of my money to spend it on things that don't benefit me in the long run, like video games?

Advice 5 Text-Davinci-3:

No, it is not a waste of your money to spend it on things like video games. Even if such purchases do not benefit you in the long-run, they are still a form of temporary pleasure that can make you feel good and be rewarding in the present. Saving for a car or any other goal is important, but it can also be helpful to have some fun and enjoyment with your money too.

Advice 5 GPT-3.5-Turbo:

It's great to hear that you have a job at such a young age and are thinking about saving for the future! While it's important to save for big goals, it's also important to enjoy your hard-earned money by spending it on things that bring you joy in the present. Video games can provide a

source of entertainment and relaxation, which can be beneficial for your mental health. However, it's important to strike a balance between spending on leisure activities and saving for future needs. One idea is to allocate a certain percentage of your income towards saving and the rest towards discretionary spending. For example, you could save 50% of your income and use the other 50% for fun activities like buying video games or going out with friends. This way, you can enjoy your money now and work towards financial stability in the future. Remember, personal finance is all about balance and finding what works best for you. Keep up the good work with your job and savings goals, and don't forget to have a little fun along the way!

Advice 5 GPT-4:

It's impressive you're thinking about your finances at this stage. Balancing spending and saving is important. It's okay to spend some funds on things you enjoy, as long as you save for the future. Creating a budget can help manage your money by allocating income for saving, spending, and financial goals. Setting goals, both short-term and long-term, can keep you focused and motivated. Consider the 50/30/20 rule; 50% of income for necessities, 30% for personal spending, and 20% for savings. But remember, you can adjust these numbers according to your needs and ambitions. Set aside 'fun money' for leisure activities, so you can enjoy the present without sacrificing future plans. Learning about personal finance, like budgeting, investing, and debt management, can help shape healthy financial habits. Remember, the right balance between spending and saving depends on your financial goals and priorities. Spending on things that bring you joy is fine, as long as you're also progressing towards your financial goals.

Advice 5 Human:

Do both. Practice self-restraint and don't splurge on games. You don't need a brand new car just something used that will last as long as you can make it. I'd like to tell you that you should just have fun (what I did at your age) but odds are if you're expected to buy your own car you might not be getting financial help from your parents with college.

Advice Request 6:

am I (19) silly for giving money away to the homeless?

I (19) have moved to a city with a big homeless population and walk by homeless people everyday on the way to school, mall, whatever. i always feel super bad and today i gave 20 dollar bills to 2 people separately, 40 dollars each because i just felt bad for them. one women was on the street looking sad and disheveled and she said god bless u which made me happy as someone kinda religious. the other guy had a sign saying he was in a fire and he looked it so i gave him a twenty after debating with myself (i was gonna give a five but felt bad and figured i didn't need it). i know this seems like i'm bad with money maybe? i'm worried but i figured i would just spend that on weed so i was better off doing this. i don't know if this is reckless spending but my

parents gave me spending money and i'm worried they would disprove ig. also i heard "reckless spending" is a result of mental illness and i'm not sure if this qualifies?

Advice 6 Text-Davinci-3:

No, you are not silly for giving money away to the homeless. It shows great kindness and empathy towards those who are in a vulnerable state and who may not have access to the same resources and safety nets as you. Generosity and selflessness are admirable qualities and are important aspects of any society. Of course, if you are concerned about reckless spending, it's best to plan ahead and set a limit for how much money you will donate each month so that you do not overspend.

Advice 6 GPT-3.5-Turbo:

It's kind and generous of you to want to help those in need, but it's important to be aware of the potential risks involved when giving money directly to homeless people. There is no way to guarantee that the money you give will be used in a positive or responsible way, and it's possible that it may contribute to enabling destructive habits such as drug or alcohol abuse. It's also important to consider your own financial situation and responsibilities. While your parents have given you spending money, it's important to budget and prioritize your expenses wisely. Giving money away without considering the potential consequences may not be the most responsible use of your resources. Instead of giving money directly to homeless individuals, you may want to consider donating to a reputable charity organization that works to address homelessness and poverty. This way, you can ensure that your donation goes towards effective and sustainable solutions that help to uplift entire communities. Additionally, volunteering at a local shelter or soup kitchen can provide a more direct way to help those in need while also gaining more insight into the issue of homelessness. In terms of your concerns about recklessness, it's important to be mindful of your financial behavior and seek help if you feel like it's becoming a problem. However, giving to those in need is not necessarily a sign of mental illness - it's a noble and compassionate act. Just be sure to consider the potential risks and use your resources responsibly.

Advice 6 GPT-4:

Your desire to assist the homeless is commendable. While offering financial aid is one method, it's important to consider your own economic stability. To continue this support responsibly, consider setting a monthly budget for this cause, helping you balance charity with personal financial needs. Other ways to help include volunteering at shelters, donating necessary items like food or clothes, or supporting organizations fighting homelessness. Educating yourself on homelessness can guide your effective support. If you're apprehensive about parental approval, discuss your intentions with them. They could suggest alternative assistance methods. Remember, acts of generosity do not necessarily indicate mental health issues. But if you're worried, consult a professional. The goal is balancing aid to those in need with your financial well-being. Monitor your spending habits and consider other methods to support the homeless.

Advice 6 Human:

Here's my point of view, as someone who was on the streets for 7 years. I was always grateful regardless if it was 50 cents or 100 bucks that someone would kick down to me. So I'm sure they are grateful and really do appreciate you helping them out. 20 bucks doesn't seem like a lot, but when you're on the streets you can make it go far. Sure I'd buy alcohol, and I was mostly honest about it, (Yo, can you spare some change for alcohol and I promise you I won't buy food) was my line, but most the time I'd buy food lol. So they'll more than likely will be getting their fix or maybe they'll get food, but now when I give back, after the money leaves my hand to theirs it's really none of my business what they spend it on. I normally say "Get whatever you need" when I give people money. I personally don't think it's foolish, but at the end of the day you need to be sure you can take care of yourself. Sorry if this makes no sense, was a long day.

Advice Request 7:

My boyfriend (23m) is a dealer and won't get a legal job what do
Hello everyone, I (22f) am having a hard time with my bf (23m). We have been together for almost 6 years, as the title states he is a dealer. I have been trying to encourage him to get a job for years, and he acts like a complete man child about it, he has a temper problem when it comes to any kind of authority over him. Other than that we really don't get into disagreements so it makes this harder for me. When we started dating we only did it because it was what we had to do if we wanted food or shelter, but now he's become lazy. He just wants fast easy ways to get rich (y'all don't even know how many crypto scams I have prevented him from giving our savings to) and not put in effort for anything he wants in life. Obviously I can't just change him and make him work, but idk what to do anymore and I'm tired of it. To make things worse I'm a very introverted person and don't have any near friends I could just move in with if I decided to leave. I hate saying I'm with him only for financial reasons because I do really love him, but ultimately if I had the money I would have been gone long ago. He just keeps saying he's going to get a job as an "entrepreneur" but as I said he wants everything handed to him with no effort. Anyways what can or should I do?

Advice 7 Text-Davinci-3:

It sounds like you have been trying hard to encourage your boyfriend to get a legitimate job, and it sounds like he doesn't take the issue seriously. I think the first step is to have a serious discussion with him about his career and financial prospects. Make it clear that it is important for you both to have a stable source of income in order to plan for the future. Explain that getting a legal job is a necessary step for him to get to where he wants to be. If he is still resistant to the idea, then you need to think about what is best for your long-term goals. If leaving the relationship is something you would consider, then start putting together the resources and plan of action you would need to do so. In the meantime, it would be a good idea to talk with a counselor or a trusted friend to help you think through your options. It is okay to feel frustrated, but it might be helpful to discuss your thoughts with someone who is objective and can provide you with a different perspective.

Advice 7 GPT-3.5-Turbo:

It sounds like you are in a challenging situation. It might be helpful to have a conversation with your boyfriend about your concerns and how his behavior and choices are affecting your relationship. You can try to approach the conversation from a place of concern and love for him, but also be firm in your expectations around what you need in the relationship. It might be helpful to come up with some specific goals or actions that he can take to show that he is serious about making changes. At the same time, it's important to prioritize your own well-being and safety. If his behavior is putting you at risk or making you uncomfortable, you may need to consider leaving the relationship, even if it means finding alternative housing options. It might also be helpful to seek support from a therapist or counselor, who can help you work through your feelings and options in a safe and non-judgmental space. Ultimately, remember that you deserve to be with someone who shares your values and is committed to building a healthy and stable future together.

Advice 7 GPT-4:

Firstly, acknowledge your conflict over your boyfriend's illegal job. The key to handling this is open, honest communication. Discuss your concerns and their effects on you and the relationship calmly. Be upfront about your feelings and expectations. Setting boundaries is also critical. Define what you can and can't tolerate, and ensure you stick to these. Be a positive influence, and inspire him to consider legal entrepreneurial pursuits or additional education, aiding him in finding resources if possible. Don't neglect your own needs. Evaluate your financial and emotional status. Consider getting a job for financial independence and, if needed, the option to exit the relationship. Consult a therapist or counselor if you're stuck or if the situation isn't improving. They could offer valuable advice and a way forward. Finally, prioritize your stability and safety. If your boyfriend's decisions continue to harm you and he's unwilling to change, reassess your relationship. You deserve a healthy, supportive, and secure relationship.

Advice 7 Human:

Ya, you're an adult woman still hanging out with some kid you dated in high school. It's time to change that. It's nice that you love him and all but that's not actually a basis for an adult relationship on its own. If it was, the divorce rate wouldn't be so high. You need compatible goals, dreams and values and you two don't have that. Your relationship has a hard end date and whether that's today or a decade from now, this isn't the man you're going to decide to spend your life with. Do both of you a favor and go your separate ways. Move back in with your parents or find someplace cheap, but you need to end this.

Advice Request 8:

Someone here gonna have to explain what I said wrong?

So Prom was a few days ago and I started talking to this girl and we rly got along and we're genuinely having a good conversation and it was going well, one of her friends even came over and said they were going and she responded back to her friend "give me a few more minutes" and we continued talking for about 10 more minutes then Prom ended and I got her snap

We were snapping back and forth for a few days with just random pics then I texted her “Hey” and she texts back “Hi” and then asked her “So what type stuff u into” and then checked my phone a few minutes later and she had blocked me, ima need someone to explain what happened here because I clearly missed something, because I thought things were going pretty good and that we both liked each other

Advice 8 Text-Davinci-3:

It's difficult to say for sure exactly what happened here as we don't have all the information, but it does seem like something went wrong. It's possible that your question about what type of stuff she is into was too personal for her and made her feel uncomfortable or it could be something completely unrelated. The best thing you can do is just accept that she blocked you and move on.

Advice 7 GPT-3.5-Turbo:

Based on the information provided, it's difficult to say exactly what went wrong. However, it's possible that the way you started the conversation with the girl over SnapChat may have made her uncomfortable. "So what type stuff u into" can come across as a bit forward or even potentially creepy, especially if you don't know someone very well. Additionally, it's possible that she could have changed her mind or simply lost interest for any number of reasons.

If you want to try to salvage the relationship, it might be a good idea to reflect on how you're coming across and consider starting with more light or friendly conversation next time. In general, it's best to try to get to know someone gradually and not put too much pressure on them to share personal information or move things forward too quickly.

Advice 7 GPT4:

A sudden halt in communication can be confusing and annoying. However, it's important to realize that many factors, many unrelated to you could affect one's behavior. There could be multiple reasons. She may have misinterpreted your question or found it too vague, leading to discomfort. External factors like personal problems or stress which are unrelated to you could be influencing her. Fear of commitment or getting close can cause people to pull away when they feel a bond forming. Influence from friends could also affect her decision to block you. Lastly, sometimes, people simply lose interest or change their minds without any clear reason. Even though it's natural to want to understand her actions, try not to take it personally. You can't control others' actions, and sometimes things don't work out as expected. Focus on moving forward and establishing relationships with those who share your interests and values.

Advice 7 Human:

"So what do you like to do in your spare time?" That's like a "1."

"So what kinky shit are you willing to let me do to you?" That's like an "8."

[Unsolicited dick pic] That's about an "11."

"So what are you into?" That's about a "6.5" or a "7."

Hope this helps.

Advice Request 9:

My boyfriend of 8 months refuses to do anything most days besides playing video games
Me (25F) and my boyfriend (44M) refuses to do anything with me out in public most days and seems to think that eventually in time I'll have to be responsible for all the household tasks and if we ever have kids I'll have to be the sole provider so all he'll have to do is pay bills and go to work. He does pay a majority of the bills because he does make more money. I think that we both need to split the needs of work in the house 50/50, I've mentioned to him before that it bothers me that we don't do things together and he'll usually come back with something like "well we watched a movie together the other day" or "I'm trying to save money" even though he gets Amazon packages like everyday. Am I being unreasonable or annoying? Or is boyfriend the one who should step up?

Clarification:

I am not a mom I don't have children, I don't really have a desire to have children when America is the way it is. I'm all set. I am however am about to finish my first year of college and I work 30+ hours a week as an LNA. I don't expect this to make anyone's opinions change in the comments I just wanted to clarify I am not a mom. But I am one busy chick.

Advice 9 Text-Davinci-3:

It sounds like you are feeling very frustrated with your boyfriend's lack of engagement in activities you can do together. It is reasonable to expect that both people in a relationship should make an effort to spend time together and share in the responsibilities of the household. It does not sound like your boyfriend is making any effort to meet you halfway on this. It may be helpful to talk to him about how you are feeling and see if you can come to a resolution about how to better engage in shared activities. If he refuses to listen or make any changes, you may want to consider if the relationship is really worth the effort.

Advice 9 GPT-3.5-Turbo:

You are not being unreasonable or annoying. It is important for both of you to have a healthy and balanced relationship where you both contribute equally, not just financially but also in terms of household tasks and spending quality time together outside of the house. Your boyfriend needs to understand that being in a relationship means compromise, effort, and making time for each other. It is not fair for him to expect you to take on all the household tasks and become the sole provider if you have children in the future.

I suggest having a serious and honest conversation with your boyfriend, expressing your

concerns and expectations for the relationship, and coming up with a plan together that works for both of you. Additionally, it may be helpful to seek relationship counseling to work through any underlying issues and ensure that both of you are on the same page about the future of the relationship.

Advice 9 GPT-4:

To address imbalance in a relationship, both partners need to communicate openly and commit to a balanced dynamic. If you feel ignored while your boyfriend spends lots of time gaming, express your feelings honestly and calmly. Emphasize the importance of shared household responsibility and quality time. Establish clear expectations and boundaries around responsibilities, finances, and leisure activities. Suggest enjoyable, affordable activities to strengthen your relationship. Discuss your finances and the necessity to balance savings with expenses. Urge your boyfriend to balance his gaming hobby with time spent with you by setting gaming and non-gaming times. If issues continue despite your efforts, consider couples counseling to aid conversations and provide insights on healthier dynamics. It's vital for both partners to contribute equally in a relationship and support each other. Communication is key, and the goal should always be a mutually agreeable solution.

Advice 9 Human:

Ask yourself why he can't date anyone his age. What self-respecting 44-year-old woman would date someone like this? Imagine being the sole provider for a 44-year-old man. This isn't fucking worth it. You don't have to be with this guy, even if you love him, sometimes dating is a test to see if someone is a life partner. Also good luck having kids with his biological clock. Do you want your children's father to be in his mid-70s when they're in college?

Advice Request 10:

Ultimate Karen Mamabear, help

My 2nd grade son was involved in a scuffle at recess. Before I even heard anything of it from school, I received an email from the mother of one of the boys involved. At the end of it she threatened to file a police report. People are absolutely insane these days but I laughed and dismissed her bizarre behavior. Didn't reply. Until one night at 730, a knock on the door to reveal 2 armed patrol officers looking to interrogate my 7 year old son. This psycho really called the cops on a 2nd grader. I'm floored. I'm thinking about going with a Harassment or Restraining Order but what do you all think? If she is so mentally unstable as to call police on a 7 year old, what's she liable to do or say if she ever sees my son around school, for example, and I am not there?

Thoughts?

Advice 10 Text-Davinci-3:

Unfortunately, it seems that this mother is clearly unhinged and taking things to an extreme. The best thing to do would be to talk to the school first and make sure they are aware of the situation and that they are taking it seriously. If the school is not taking appropriate action, you may wish to consider a restraining or harassment order to protect your son. It is important to be proactive and take this seriously so that your son is protected from further harassment and intimidation from this mother. You should also address the matter with the police directly and see if they are willing to investigate the matter further.

Advice 10 GPT-3.5-Turbo:

I'm sorry to hear about your situation, and I can understand how you must be feeling. It's understandable that you would want to take action to protect your son and yourself from this person's erratic behavior. If you believe that this person is a threat to you or your son, and you feel unsafe, then seeking a restraining order may be a good idea. This can provide a legal barrier between you and this person, which may help you feel more secure. It's important to document everything that has happened so far, including the email, the visit from the police, and any other interactions you have had with this person. This documentation can help support your case if you do decide to seek legal action. It's also a good idea to speak with a lawyer, who can advise you on the best course of action. In the meantime, it may be helpful to have a conversation with your son about the importance of avoiding this person and seeking out a teacher or other authority figure if he ever feels uncomfortable or unsafe. This can help him feel empowered and give him tools to cope with the situation. Remember that you are not alone in dealing with difficult people like this. There are resources available to help you navigate this challenging situation.

Advice 10 GPT-4:

Your situation with your son is worrisome. Although I can't give legal advice, I can propose general steps. Start by discussing the incident and the other mother's response with your son's school. The staff might help mediate the situation. Keep a record of all interactions with the other mother. This could prove helpful in potential legal action or providing an account to the school. Despite provocation, stay calm and respectful towards the other mother to avoid escalation. If her behavior threatens your son or crosses boundaries, consider seeking legal guidance. Prioritize your son's feelings and experiences, encourage open dialogue about the incident, and give support. If things remain difficult, a mediation session with the other mother and a neutral third party might help. Remember to prioritize your son's well-being. By collaborating with the school, keeping composure, and possibly consulting with legal professionals, you can manage this tough situation.

Advice 10 Human:

This is a matter for school administration since it happened during recess. Make an appointment with the admin to discuss the next steps. Usually if police are involved they will contact the school. I'm surprised you haven't heard from the school. The thing is schools easily label kids

and sometimes those labels affect how they are dealt with going forward. You will want to get their take on the event. It will give you a good idea of how they viewed and recorded the event. Don't escalate. Investigate.

Advice Request 11:

My dad was diagnosed with ALS a few months ago and is starting to go downhill a bit. He'll be 75 at the end of the year, and my mom and I are trying to come up with fun things to do, so I'm reaching out! Any ideas are welcome! Thank you in advance ☺

Advice 11 Text-Davinci-3:

I'm so sorry to hear about your dad's diagnosis with ALS. It's understandable that you and your mom want to make sure he is able to still have some fun activities despite his condition.

Some ideas to consider may include:

- having a movie night together by watching his favorite movie
- trying out some at-home butterscotch fitness exercises or yoga
- taking a drive together to a local spot he loves
- having a picnic in your backyard or a nearby park
- organizing a scavenger hunt around the house (you can theme it for his favorite hobby or pop culture interests)
- setting up a virtual game night with friends and family
- taking a virtual art class together
- creating a special photo album or scrapbook of memories
- going on a virtual shopping spree
- planning a special dinner night with Mom.

I hope these ideas help you make some special memories with your dad. Wishing you and your family the very best during this difficult time.

Advice 11 GPT-3.5-Turbo:

I'm sorry to hear about your dad's diagnosis. It's great that you and your mom are looking for ways to make his remaining time enjoyable. Here are a few ideas to consider. Firstly, planning a family vacation could be a wonderful way to create lasting memories. Choose a destination that your dad has always wanted to visit and make sure it is wheelchair-friendly and accessible. This way, he can fully enjoy everything that the trip has to offer. Another idea is to host a family game night. Gather some of your dad's favorite games, whether it's cards or board games, and have a fun night together. Get some snacks, drinks, and maybe even some prizes to make it extra enjoyable. Creating a photo album or scrapbook can be a heartfelt and meaningful activity. Go through old family photos or take new ones, and compile them into an album that your dad can treasure and look through whenever he wants. If your dad enjoys movies, having a movie marathon day could be a great option. Pick some of his favorite films and spend the day watching them together. Make sure to have plenty of snacks and drinks to make it a cozy and

enjoyable experience. Lastly, consider bringing your dad's favorite restaurant to him if he can no longer go out to eat. Order takeout from his preferred place or cook his favorite meal at home to create a special dining experience. Above all, remember that the most important thing is to spend quality time with your dad and make him feel loved and supported. These activities can provide enjoyable and memorable moments, but the key is to cherish each moment you have together.

Advice 11 GPT-4:

I'm sorry to learn about your father's condition. It's admirable that you and your mom are seeking enjoyable activities for the three of you. Here are some suggestions keeping in mind his health. Movie nights could be comforting where you can watch his favorite films or discover new ones, perhaps with blankets, pillows, and snacks. Board games or puzzles could keep his mind active while providing family bonding. Storytelling sessions about past memories or favorite moments can foster deep connections. Engaging in arts, like painting, might kindle creativity and sense of achievement. Listening to music or audiobooks can be both soothing and engaging. Gentle exercises like chair yoga or tai chi can help maintain his flexibility. Consider low-impact outdoor activities, like short walks or scenic drives. Involving your father in cooking his favorite dishes can be enjoyable. Video calls with friends and family can keep him socially active. Collaborating on a memory book could create a valuable keepsake. Remember to be patient and understanding, cherishing the moments you share together.

Advice 11 Human:

My mother had ALS. Chances are, you can expect the following to happen:

Your dad is going to lose his ability to communicate (or at least to communicate without difficulty) at some point. Use this time now to get him to tell his stories, and record them so that you can write them down for his grandkids to enjoy. Your dad is going to lose his ability to swallow, and will eventually need a feeding tube. So cook what he loves, take him out to restaurants he loves, and order in food he loves, while you can. Your dad is going to lose his mobility, and navigating the world in a wheelchair (especially a power chair, which he is likely going to eventually need) is anything but easy. So if there are places he likes to go, do it now while you still can. Your dad is going to tire easily and anything you plan for his needs to account for that.

Advice Request 12:

Why am I still disgusted with my boyfriend

I've been dating my boyfriend for three years now. Honestly when we met I wasn't looking for a relationship, and I don't think he was either.

So when we first met I saw him and I felt disgusted. This overwhelming feeling of disgust...

I ignored it then, as we'd only just met, but to this day it's still there.

There's things I've grown to love about him, but still underneath this feeling of disgust is there.

Why do I feel this way?

Advice 12 Text-Davinci-3:

It is possible that this feeling of disgust comes from unresolved issues you have with yourself. For example, you may be projecting feelings of low self-worth onto this relationship, subconsciously believing that you don't deserve someone better or that you don't deserve to be in a healthy, loving relationship.

It's also possible that you are being influenced by insecurities from past relationships, if you have any. Furthermore, this feeling of disgust could be due to a traumatic experience in your past that has impacted the way you view relationships and/or intimate connections.

Our emotions can often be mysterious and difficult to decipher. It is possible that in order to get to the bottom of your feelings and make sense of them, you may need to seek the assistance of a professional. Counseling or psychotherapy can be a great way to explore and understand your feelings and identify any underlying issues that could be causing them.

Advice 12 GPT-3.5-Turbo:

It's difficult to say for sure without more information, but it's possible that you could be experiencing something called "disgust sensitivity." This is a common and natural response to things that we perceive as unsanitary, dangerous, or generally unappealing. It can be caused by a variety of factors, including personal experiences, cultural influences, and even genetics. In the case of your relationship, it's possible that something about your boyfriend triggers this response in you. It could be a physical trait, a habit or behavior, or even something about his personality. It's important to explore these feelings and try to pinpoint what specifically is causing you to feel this way. Keep in mind that feeling disgusted does not necessarily mean that your relationship is doomed. Many couples have differences or annoyances that they learn to work through and overcome. However, if the feeling persists and is impacting your ability to have a healthy relationship, it may be worth seeking the help of a therapist or counselor to explore your feelings more deeply.

Advice 12 GPT-4:

Identifying the cause of your disgust for your boyfriend without more context is difficult. Several possibilities could explain this. One is that your first impression about him still affects your emotions. Also, despite your three-year relationship, certain incompatibilities might trigger these feelings. Feelings of disgust may also act as a defense mechanism in a relationship. Unresolved issues within your relationship could also be a cause. It's important to address any conflicts or communication issues you might have. To better understand and manage your feelings, you may want to consult a mental health professional. Open conversations with your boyfriend about your feelings can improve your understanding and connection.

Advice 12 Human:

Why did you date someone you think is disgusting in the first place? Why are you still dating him? You are not doing either him or you any favors. You haven't changed in feelings towards him and honestly it's going to turn to resentment eventually, it doesn't matter if you've "grown to love some stuff..." Break up so he can find someone who truly loves and accepts him, you'll be cruel to stay. And if your only arguments to this are "but we've been together this long.." and "But I kinda love him..", that's not a good enough reason to stay

Advice Request 13:

My mom has been crying and is devastated since I moved out what do I do?

I grew up with a single mom and for a while was my only guardian before my dad came back to be part of my life. We were very close and to her it's just me in her family. I decided I wanted to move out for college with a friend and get an apartment in the city but since then my mom has been torn and never even wanted me to move out in the first place. She's lived alone since and said how much she misses me and wants me to move back in but today she called and was crying on the phone how lonely she is and how afraid she is and wants me to consider moving back in when my lease is up though I've been planning for me and my gf to then move in which again she's not fond of. I visit her each week between school and work but she's still sad. What do I do, I don't like my mom being so lonely and sad but then again I'm almost 20 and should live my own life?

Advice 13 Text-Davinci-3:

It sounds like you have a very close relationship with your mom, and it must be difficult to see her so upset. It's understandable that you want to live your own life and have your own independence, but it's important to consider your mother's feelings and emotions. One thing you could do is have a conversation with your mom about it and try to come up with a solution that is mutually beneficial. Explain to her that you still want to live your own life, but also have some way to make her feel more connected to you. Perhaps you can offer to video call her every week, or maybe you could take her out on special occasions or plan regular mother-daughter weekends. It's also possible that she just needs some extra assurance that you still love her and that you won't forget about her even though you don't live in the same house. It's important to be patient and understanding as you talk to your mom and work out a solution that works for both of you.

Advice 13 GPT-3.5-Turbo:

It is understandable that you feel conflicted about your mom's sadness but also want to live your own life. It sounds like your mom may be experiencing empty nest syndrome, which is a common feeling for parents when their children move out. You can have an open and honest conversation with your mom and express your love and concern for her, but also let her know that it's important for you to have your own life and independence. You can reassure her that you will still visit and keep in touch regularly. You can also suggest activities or hobbies for her to do that can help her feel less lonely and more engaged. Encourage her to connect with her own friends or family members for support, and to seek professional help if she feels overwhelmed. It's important to remember that you are not responsible for your mom's happiness, but you can be supportive and understanding. Ultimately, you have to make the best decision for yourself and your future, and your mom will eventually have to adjust to this new phase in both of your lives.

Advice 13 GPT-4:

Balancing your mother's well-being with your growing independence can be complex but manageable. Begin by discussing your feelings and ambitions clearly while also understanding and acknowledging her loneliness. Help her combat this by developing a wider social circle through activities or reaching out to friends. Regular contact with her can also make her feel

connected and reassure her that your stepping into independence isn't abandonment. If her feelings of sadness persist, professional counselling may be beneficial. Don't neglect your own needs, too. Remember, your future plans, including cohabitation with your girlfriend, shouldn't be compromised. Striking a balance between supporting your mother and moving ahead in life is vital. You can do this through open communication, encouraging her social interactions, and maintaining regular contact.

Advice 13 Human:

Your mom needs therapy. I can understand being lonely but the whole point of being a parent is getting your child to the point where they can be independent. Everyone has to deal with being alone at some point. Even if you lived with just her your whole life, one day she will die and then the roles are reversed. You visit her every week. I don't even hang out with certain friends that often. You shouldn't change anything you're doing because of her. She's the one who's going to have to learn to cope with her empty nest syndrome. If I were you I'd gently recommend counseling, but don't try to compromise any more than you have been. You'll just be stunting your own growth.

Advice Request 14:

Are any teen moms, dads, or parents, in general, able to advise me on caring for a baby at 16?
I got my girlfriend pregnant and she doesn't want an abortion but neither do I. Im gonna be 16 in about 2 weeks so I can get a decent amount of jobs but pretty low pay. I don't know how my mom is going to react, she could kick me out but I'm not sure. Would it be better to move in with my GF and care for the baby or stay with my mom? I also need help knowing what I need for a baby.

Advice 14 Text-Davinci-3:

It is recommended to talk to your family and your girlfriend to find the best solution. There are many support networks available for teen parents, such as parenting classes and support groups. Talk to your healthcare provider or local family doctor for referrals. There are also a variety of online resources that can provide helpful advice and tips for teen parents, such as The National Campaign to Prevent Teen and Unplanned Pregnancy and the March of Dimes. It is important to start preparing now by learning about parenting skills and understanding the responsibilities involved in caring for a baby. Make sure to look into financial support and assistance, such as the Supplemental Nutrition Assistance Program (SNAP). Additionally, it will be helpful to start gathering the supplies and items needed for a baby beforehand. Make a list of things you may need, such as clothing, diapers, baby wipes and items for the baby's room. It is also important to build a support system of family and friends who can help you through this transition. Good luck!

Advice 14 GPT-3.5-Turbo:

First of all, it's important to understand that caring for a baby is a huge responsibility that requires a lot of time, effort, and resources. Teen parenthood can be challenging, but with

support, education, and dedication, it is possible to do it successfully. Regarding whether it would be better to move in with your girlfriend or stay with your mom, there are several factors to consider. It's important to talk openly with your girlfriend and your mom about your options and their feelings. Living with your girlfriend may provide more privacy and autonomy, but it may also come with additional expenses and responsibilities. Living with your mom may offer more stability and support, but it may also require compromise and communication about your new role as a parent.

When it comes to what you need for a baby, some essential items include:

- Diapers
- Wipes
- Clothing (onesies, sleepers, socks, hats, etc.)
- Bottles and formula (if not breastfeeding)
- A car seat
- A crib or bassinet
- A stroller
- Baby bath tub
- Baby monitor

Additionally, it's important to have resources for medical care, such as a pediatrician and health insurance.

It's also important to note that there are many community resources available for young parents, such as parenting classes, support groups, and financial assistance programs. Do some research and see what resources are available in your area. Remember that it's okay to ask for help, and that being a good parent means putting the needs of your child first. Take care of yourself, your girlfriend, and your baby, and don't be afraid to reach out for support.

Advice 14 GPT-4:

Even if I can't personally relate to your teen parent situation, I can advise on how to adjust to this life change and make informed decisions. Start by having open communication with both sets of parents- this can facilitate discussions about your fears, explore options, and garner their support. They can offer insightful guidance in this journey. Consider where you will live. This depends on family dynamics and available resources. The place needs to be stable and supportive for your baby and you. Parenting is challenging, hence the need to learn about baby care from books, parenting classes, consultations, or online sources.

Financial readiness is key. Babies are costly; budget for essentials, medical costs, childcare. Consider cost-saving options like buying used items or borrowing baby gear. Regular prenatal care for your girlfriend during pregnancy is a must. This ensures the health of her and the baby, and provides many resources. Prepare the essential items like clothing, diapers, a bassinet, blankets, bottles, and a car seat. As the baby grows, think about additional necessities. Building support networks of friends, family, and healthcare professionals is vital, as well as seeking help from local community organizations or young parent support groups. While your baby should be a priority, don't abandon your education and career goals to ensure a secure future. Remember,

each situation is unique; make decisions based on your circumstances and resources. Don't shy away from asking for help from trusted adults, healthcare professionals, and other resources.

Advice 14 Human:

I'm 30 now. But when I was 17 my girlfriend and I got pregnant. We opted for adoption, my father knew of a couple who wanted but couldn't have kids. They were super awesome, well-off, amazing jobs. Needless to say the kid was going to be set, better off than I ever was and I was raised middle class. They paid for all the hospital expenses, pampered her, and let us stay at their house. It was the best option for us at the time, cause neither of us felt right getting an abortion, but knew we weren't ready to be parents. Flash forward to now, that girl and I broke up more than a decade ago, good riddance honestly. I met my NOW wife a few years after the breakup, we have two boys, and life is good. Parenting is hard, I do not regret what we did and the choices we made. The way our relationship went after the adoption was horrible. She cheated, manipulated, tormented. The true colors showed once we were out of high school and got to experience the real world. Just weigh your options man, and think about your future self. Good luck and hope you the best.

Advice Request 15:

Is it ok to walk away?

Yesterday I was wearing a pullover of my favorite baseball team. At the grocery store. Doing my thing. Minding my business.

This man, a creepy squirrely type, starts talking to me about the team. He makes zero sense. I can barely understand him and I'm trying to be polite and saying mmmhmm and walk away. Again. I don't know him. I think that's generous. But, he turns his cart around and starts following me to talk more. I end up turning again, because ... I don't care to continue this conversation because I feel he's being a little racist. But, I feel he's taken advantage of my being nice. He approaches again an aisle over and continues, at this point, I try an uh-huh and walk off.

In this situation, is it ok to walk off? I don't know this person. He's a little on the creeper side. I don't want trouble. I was just minding my business, shopping.

Advice 15 Text-Davinci-3:

Yes, it is completely ok to walk away! You were minding your own business and you can choose not to engage in conversations you aren't comfortable with. No one should ever make you feel unsafe or uncomfortable and it is important to trust your gut if you feel like something isn't right or if someone is making you feel uncomfortable. You have every right to politely excuse yourself and walk away.

Advice 15 GPT-3.5-Turbo:

Yes, it is absolutely okay to walk away in this situation. Your safety and well-being come first, and you should never feel obligated to engage with someone who is making you uncomfortable

or being inappropriate. It sounds like you made multiple attempts to politely disengage from the conversation, and that is more than enough. It's important to trust your instincts and remove yourself from any situation that doesn't feel safe or comfortable.

Advice 15 GPT-4:

Sure, it's alright to leave if you're uncomfortable. Your safety and comfort matter. If someone makes you uneasy, you're entitled to exit the situation. Trust your instincts and prioritize your health. You don't owe anyone your time or attention, especially if they cross personal boundaries. If harassment continues, seek help from store staff or security.

Advice 15 Human:

Yes, perfectly fine. Also it would be okay for you to be firm and straight up tell him to please leave you alone. Men can sometimes be persistent unless their directly told to go pound rocks.

Appendix E

Recruitment Poster



ADVICE RESEARCH



WHAT MAKES ADVICE GOOD?

- › **What makes for good advice? Let us know in this study!**
- › **Read about real problems and the kind of advice offered.**
- › **Any adult (18 or older) comfortable with English can participate.**
- › **Join our study. It only takes about as long as it takes to enjoy a good cup of coffee.**
- › **Scan the QR code or use the link to learn more and get started.**
- › **Link: https://uitpsych.qualtrics.com/jfe/form/SV_erc5kvAnFLiXgZE**



Appendix F

Advice Identification Study

Information Letter

Online Advice: Source

Giving and receiving advice is a part of everyday life. It comes up during casual conversations, in books, and can be found all around the internet. Advice can have a significant impact on people's choices and consequently their lives. In this study, you will help us understand more about the nature of advice.

Advice in today's digital age can come from various sources, including technology known as Artificial Intelligence (AI). AI refers to computer systems designed to perform tasks that typically require human intelligence, such as understanding language, making decisions, and generating content.

In this study, you will be presented with real-life advice requests that people have posted online. After reading them, you will be asked a simple question about the advice giver. The questionnaire takes about 15 minutes to complete – about as long as one might spend perusing a published advice column.

The study:

- Format:
 - This will be an online study that you can do from anywhere with a computer, smartphone, or tablet.
- Goal:
 - To understand the nature of good advice. You will help us by reading several examples of real problems people have posted on an online advice channel. The topics they touch on range from relationships, work, sex, finance, and other challenging situations. You will then read advice given concerning their problem and asked if you believe the author of the advice giver to be human or artificial intelligence (AI).
- Duration:
 - 1 x 15 minutes

How to participate:

When you have read this form press the green arrow at the bottom of the page. This will take you to the consent form. If you agree to participate, you will directly start the online questionnaire. Your participation is completely voluntary.

Who is qualified to participate:

Everyone who is over the age of 18 and comfortable reading English can participate in our study. Some questions may contain explicit language and situations that are sensitive in nature. Many questions include curse words, and some advice requests describe situations regarding relationship issues, illnesses, blackmail, sex, or other situations that some might find uncomfortable to read about. Keep that in mind when deciding whether to participate.

We take care of you and your data:

We respect your privacy and adhere to the data protection standards recommended by UiT The Arctic University of Norway and the Norwegian Agency for Shared Services in Education and Research (SIKT). Accordingly, all the data you provide to us will be anonymous. None of the data we collect can be traced back to any individuals. Note that because the data cannot be traced back to you, we are unable to honor requests to later access or delete your responses. When this study is completed, the entire study will be in Otto Kuosmanen's master's thesis published on UiT's website (<https://munin.uit.no>) Should we go on to publish this work in a peer-reviewed journal, the entire data set will also be made available on an approved open data source.

If you have any comments or questions:

If you have any comments, you will be given a chance to write them at the end of the study. If you have any questions, you can contact us by email.

Sincerely yours,
Otto J. B. Kuosmanen, MA student, oku012@uit.no
Tove I, Dahl, Professor, tove.dahl@uit.no
Department of Psychology at UiT The Arctic University of Norway



Consent Form

By taking part, you agree to the following:

- I am 18 years or older.
- I'm aware that my responses are anonymous and cannot be traced back to me.
- I understand that my participation is voluntary
- I understand the information that has been presented to me in the information letter.
- I'm aware that some questions may include explicit language and sensitive topics.

I agree to participate -- Take me to the questionnaire!

I prefer not to participate.



Instructions

Instructions:

In this exercise, you will be presented with a total of nine advice requests, each accompanied by advice. The advice is either written by a human or generated by an Artificial Intelligence (AI) algorithm. Your task is to read each question and its corresponding answer carefully, and then decide whether you believe the answer was written by a human or generated by AI.

Please take your time to consider each question and answer thoughtfully. After reviewing each one, indicate your best guess as to whether the answer was produced by a human or by AI.

Appendix G

Ethics Approval



Institutt for Psykologi
Arkiv ref.: 25/23/R1
Dato: 8. november 2023
off. § 26.4

Tove Dahl
Department of Psychology
UiT - The Arctic University of Norway

Ethical evaluation of research project

Dear Tove,

Your application concerning your research project "*A comparison of human and artificial intelligence (AI) generated advice*" has been evaluated and approved by the Department of Psychology's internal research ethics committee (IPS-REC) based on the received information.

on behalf of the Committee

Matthias Mittner
Chair of IPS-REC
—
research-ethics.ips@uit.no

Kopi sendt: John Vegard Bjørklund



Appendix H

Advice Preference Study

Information Letter

Online Advice Quality

Giving and receiving advice is a part of everyday life. It comes up during casual conversations, in books, and can be found all around the internet. Advice can have a significant impact on people's choices and consequently their lives. So, what kind of advice is helpful for the people who request it?

In this study, you will be presented with real-life advice requests that people have posted online. After reading them, you will be asked short questions about the nature of the advice given. The questionnaire takes about 15 minutes to complete – about as long as one might spend perusing a published advice column.

The study:

- Format:
 - This will be an online study that you can do from anywhere with a computer, smartphone, or tablet.
- Goal:
 - To understand the nature of good advice. You will help us by reading several examples of real problems people have posted on an online advice channel. The topics they touch on range from relationships, work, sex, finance, and other challenging situations. You will then read the advice given concerning their problem. For each problem, you will be presented with two pieces of advice, and you will be asked to rank the advice based on how helpful you think each is.
- Duration:
 - 1 x 15 minutes

How to participate:

When you have read this form press the green arrow at the bottom of the page. This will take you to the consent form. If you agree to participate, you will directly start the online questionnaire. Your participation is completely voluntary, and you can withdraw at any time without penalty.

Who is qualified to participate:

Everyone who is over the age of 18 and comfortable reading English can participate in our study. Some questions may contain explicit language and situations that are sensitive in nature. Many questions include curse words, and some advice requests describe situations regarding relationship issues, illnesses, blackmail, sex, or other situations that some might find uncomfortable to read about. Keep that in mind when deciding whether to participate.

We take care of you and your data:

We respect your privacy and adhere to the data protection standards recommended by UiT The Arctic University of Norway and the Norwegian Agency for Shared Services in Education and Research (SIKT). Accordingly, all the data you provide to us will be anonymous. None of the data we collect can be traced back to any individuals. Note that because the data cannot be traced back to you, we are unable to honor requests to later access or delete your responses. When this study is completed, the entire study will be in Otto Kuosmanen's master's thesis published on UiT's website (<https://munin.uit.no>). Should we go on to publish this work in a peer-reviewed journal, the entire data set will also be made available on an approved open data source.

If you have any comments or questions:

If you have any comments, you will be given a chance to write them at the end of the study. If you have any questions, you can contact us by email.

Sincerely yours,

Otto J. B. Kuosmanen, Master's student, oku012@uit.no
Tove I. Dahl, Professor, tove.dahl@uit.no
Department of Psychology at UiT The Arctic University of Norway



Instructions

Instructions:

In this exercise, you will be presented with six series of texts where someone is seeking advice. For each scenario, there will be two pieces of advice provided. Your task is to read both the question and the corresponding advice before making a decision. Your goal is to choose the piece of advice that you find more helpful.

Take your time to consider each situation and the advice given before making your choice.

Consent Form

By taking part, you agree to the following:

- I am 18 years or older.
- I'm aware that my responses are anonymous and cannot be traced back to me.
- I understand that my participation is voluntary
- I understand the information that has been presented to me in the information letter.
- I'm aware that some questions may include explicit language and sensitive topics.

I agree to participate -- Take me to the questionnaire!

I prefer not to participate.



Appendix I

Advice Quality Study

INFORMATION LETTER

Online Advice Quality

Giving and receiving advice is a part of everyday life. It comes up during casual conversations, in books, and can be found all around the internet. Advice can have a significant impact on people's choices and consequently their lives. So, what kind of advice is helpful for the people who request it?

In this study, you will be presented with real-life advice requests that people have posted online. After reading them, you will be asked short questions about the nature of the advice given. The questionnaire takes about 15 minutes to complete – about as long as one might spend perusing a published advice column.

The study:

- Format:
 - This will be an online study that you can do from anywhere with a computer, smartphone, or tablet.
- Goal:
 - To understand the nature of good advice. You will help us by reading several examples of real problems people have posted on an online advice channel. The topics they touch on range from relationships, work, sex, finance, and other challenging situations. You will also read the advice given concerning their problem and you will be asked to rate the advice based on helpfulness, efficiency, sensitivity, and appropriateness.
- Duration:
 - 1 x 15 minutes

How to participate:

When you have read this form press the green arrow at the bottom of the page. This will take you to the consent form. If you agree to participate, you will be sent directly to the questionnaire. Your participation is completely voluntary, and you can withdraw at any time without penalty.

Who is qualified to participate:

Everyone who is over the age of 18 and comfortable reading English can participate in our study. Some questions may contain explicit language and situations that are sensitive in nature. Many questions include curse words, and some advice requests describe situations regarding relationship issues, illnesses, blackmail, sex, or other situations that some might find uncomfortable to read about. Keep that in mind when deciding whether to participate.

We take care of you and your data:

We respect your privacy and adhere to the data protection standards recommended by UiT The Arctic University of Norway and the Norwegian Agency for Shared Services in Education and Research (SIKT). Accordingly, all the data you provide to us will be anonymous. None of the data we collect can be traced back to any individuals. Note that because the data cannot be traced back to you, we are unable to honor requests to later access or delete your responses. When this study is completed, the entire study will be in Otto Kuosmanen's master's thesis published on UiT's website (<https://munin.uit.no>) Should we go on to publish this work in a peer-reviewed journal, the entire data set will also be made available on an approved open data source.

If you have any comments or questions:

If you have any comments, you will be given a chance to write them at the end of the study. If you have any questions, you can contact us by email.

Sincerely yours,

Otto J. B. Kuosmanen, Master's student, oku012@uit.no
Tove I. Dahl, Professor, tove.dahl@uit.no
Department of Psychology at UiT The Arctic University of Norway



Instructions**Instructions:**

In this task, you will be presented with a six scenarios where someone is seeking advice. You will be presented with the advice request and the advice given for this particular problem. Your task is to read both the request and the corresponding advice and, rate the advice with scales of helpfulness, efficiency, sensitivity, and appropriateness. Take your time to consider each situation and the advice given before making your ratings.



Appendix J

Statistical Analysis Details

All the data work was conducted using the Python programming language (*Python.Org*, n.d.).

All statistical analyses were conducted with the library SciPy (*SciPy* -, n.d.). The scripts can be found on GitHub. [OttoKuosmanen/analysis \(github.com\)](https://github.com/OttoKuosmanen/analysis). Here I will briefly explain the relevant analysis scripts connected to the results of the three studies.

Identification Study

In total, we had 4 scripts that performed analysis operations on the data from our Identification Study. Below we describe the purpose of each script.

Participant_demographics.py

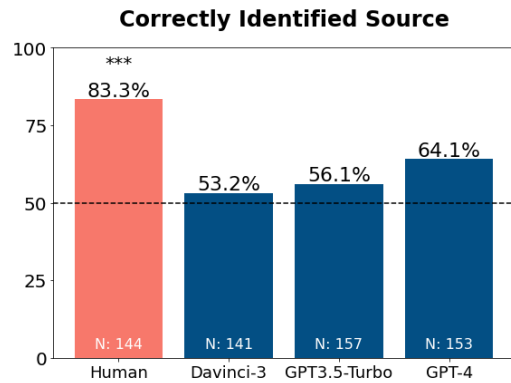
This script collected and presented us with information about the participants. The following information was processed and printed out for the researcher: Total amount of observations, Participants' completion rates, gender distribution, age distribution, education distribution, location distribution, recruitment information and average completion times.

Chi_Square_tests.py

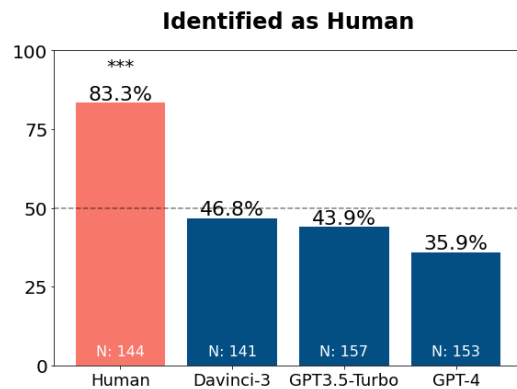
This script performed Chi-Square tests determining if the identification rates were significantly different from chance levels. Five tests were performed, one test per source (Human, text-davinci-3, gpt-3.5-Turbo, Gpt-4, AI-aggregate) against the expected ratio of 0.5, representing chance level.

Plot_id_all_models.py

This script constructed a bar graph that displays the rates of correct identification per source.

**Plot_id_all_models_identified_as_human.py**

This script constructed a bar graph that displays the rates of identifications as human per source.

***Preference Study***

In total, we had 4 scripts that performed analysis operations on the data from our Preference Study. Below we describe the purpose of each script.

Participants.py

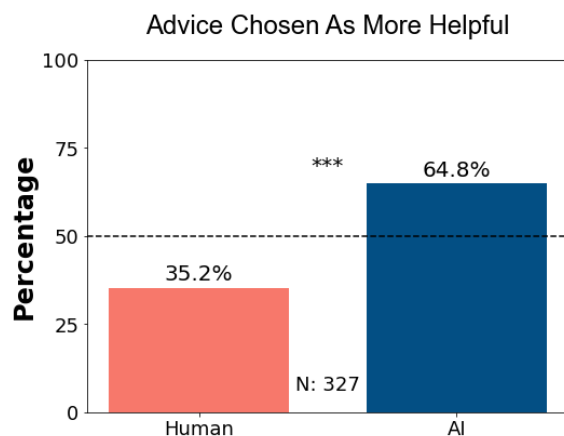
This scripts collected and presented us with information about the participants. The following information was processed and printed out for the researcher: Total amount of observations, Participants' completion rates, gender distribution, age distribution, education distribution, location distribution, recruitment information and average completion times.

Chi_Square.py

This script performed Chi-Square tests determining if the preference rates were significantly different from chance levels. Four tests were performed, one test per AI source (text-davinci-3, gpt-3.5-Turbo, Gpt-4, AI-aggregate) against the expected ratio of 0.5, representing chance level.

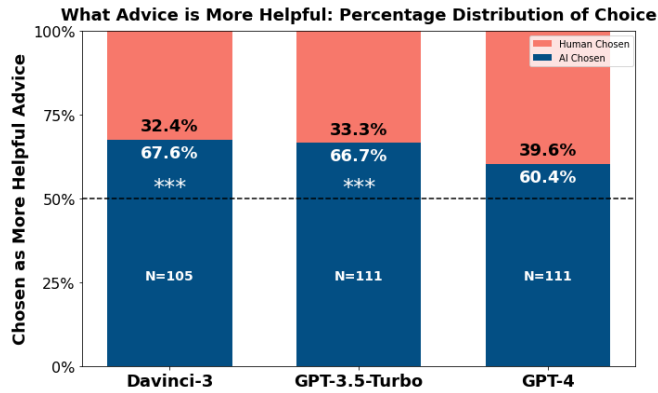
Hammer_Best.py

This script plotted a bar graph for the preference ratio of the AI aggregate and Human advice.



Smelter_Best.py

This script plotted a bar graph of the preference ratio of all AI models. Including the preference for human advice.

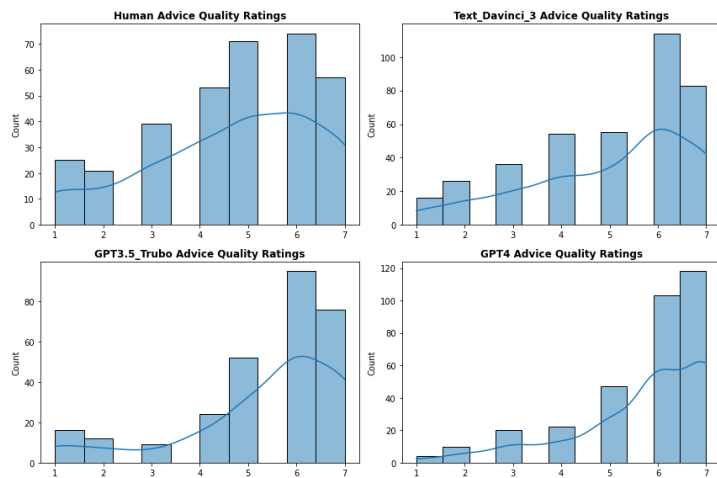


Advice Quality Study

In total, we had five scripts that performed analysis operations on the data from our Advice Quality Study. Below we describe the purpose of each script.

Normality_test.py

This script performed a histogram plot to investigate the normality distribution of our advice ratings. As expected, we found left skewed ratings indicating that all models were rated high on our scales.



Demographics.py

This scripts collected and presented us with information about the participants. The following information was processed and printed out for the researcher: Total amount of observations, Participants' completion rates, gender distribution, age distribution, education distribution, location distribution, recruitment information and average completion times.

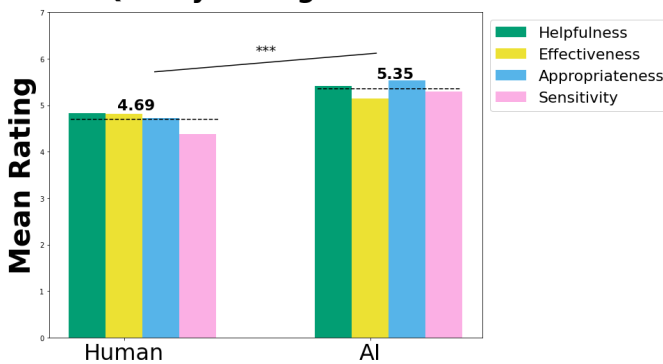
Significance_testing.py

This script conducted 13 significance tests (Man Whitney-U-tests and Kruskal Wallis tests).

AiHuman.py

This script plotted a bar graph displaying the advice ratings of the AI aggregate and human advice.

Advice Quality Ratings: Human vs AI



Gemcutter.py

This script plotted a bar graph displaying the advice ratings of the AI models and human advice.

